Application of data mining methods to establish systems for early warning and proactive control in food supply chain networks

Yuan Li

Thesis committee

Thesis supervisors

Prof. ir. A.J.M. Beulens, Professor of Information Technology, Wageningen University

Prof. dr. ir. J.G.A.J. van der Vorst, Professor of Logistics and Operations Research, Wageningen University

Other members

Prof. dr. E. Hagdorn-van der Meijden, VU University Amsterdam

Prof. dr. ir. J.L. Top, VU University Amsterdam

Prof. dr. J. van Hillegersberg, University of Twente

Prof. dr. ir. A.G.J.M. Oude Lansink, Wageningen University

This research was conducted under the auspices of the Mansholt Graduate School of Social Sciences.

Application of data mining methods to establish systems for early warning and proactive control in food supply chain networks

Yuan Li

Thesis
submitted in fulfillment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. dr. M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 21 May 2010
at 11 a.m. in the Aula.

Yuan Li
Application of data mining methods to establish systems for early warning and proactive control in food supply chain networks

Thesis, Wageningen University, Wageningen, NL (2010)
With references, with summaries in Dutch and English

ISBN 978-90-8585-638-2

Dedicated to my wife Zhuangyi

Acknowledgements

Throughout the years I spent in Wageningen University, I received tremendous help and support from numerous people. It would not have been possible to write this PhD thesis without their help.

First of all, I wish to express my warm and sincere thanks to Prof. Adrie Beulens, as well as to Kees Bink from Nutreco. Without them, this fantastic research would not be possible. This project was defined by them, and they should be regarded as the 'grandfather' of the system designed in this research. I want to thank Adrie Beulens for his broad knowledge and profound insights when guiding me in this research. It is my pleasure to thank Kees Bink for his tremendous help on looking for real cases and on introducing me to the food production industry.

I would like to express my sincere gratitude to Prof. Jack van der Vorst. His precise comments and enormous mental support ensured that this research is always on the track. I am extremely grateful to Mark Kramer for his guidance and support on different aspects of my research, from the thesis writing, to the arrangement of many practical issues, such as expert validation. His logical and meticulous way of thinking has been of great value for me. I always feel very lucky to work together with him.

This research benefited a lot from the communication and discussion with many knowledgeable and open minded people. Among those people, I would like to especially thank Wil Hennen, Floor Verdenius, Nel Wognum, Cor Verdouw, and Rob Hartog. They gave precious suggestions on various aspects of this research, from data analysis, to system design and validation. Those suggestions effectively kept me away from many pitfalls that may hinder this research.

Several real cases from two Dutch food companies, Nutreco and VION, were used in this research. During the case analysis, many people in those two companies gave me a lot of help. In particular, I would like to thank Derk Oorburg, Lourens Heres, Jaco Eissen, Gerton Peeters, and Ronny van Willigen. They supplied me with the data needed for my research, and also explained to me the business context of those data.

I would like to thank all the colleagues at the Information Technology Group, with whom I have discussed about the research as well as daily life. In particular, I would like to thank Gert Jan Hofstede for his constant and generous encouragement. He always has the magic to make people comfortable. I miss the cozy conversations with him very much. I also want to thank Ria Slootman for her assistance and support on many specific but important matters, especially when I was not at Wageningen. My gratitude also goes to Ayalew Kassahun, Dik Kettnis, Gerard Moerland, Maarten Zijp, Sjoukje Osinga, and Huub Scholten. Thanks for the help you offered to me in the past years. I enjoyed a lot the talks with you during the lunch time and coffee breaks.

I also wish to thank other PhD students with whom I had very nice talk. Among them, I would like to thank my former roommate, Sebastian Meijer, for releasing me from the tough thinking on my research every now and then. I would like to thank Jelena Vlajic for sharing her thought and joyful stories with me. I also would like to thank Guangqian Peng for the pleasant discussion on a wide range of topics.

The support coming from my family is indispensible throughout this study. I want to thank my parents for their endless belief on me throughout all the years I spent in study.

At the end, I dedicate this book to my wife Zhuangyi Wang. With her endless support, I am fearless to the obstacles encountered in this research.

Purmerend, March 2010

Yuan Li

Table of content

List of abbreviations

DM	Data Mining
DOR	Design Oriented Research
DSS	Decision Support Systems
ES	Expert Systems
EWPC	Early Warning and Proactive Control
FSCN	Food Supply Chain Networks
KB	Knowledge Base
KBDSS	Knowledge-Based Decision Support Systems
KBS	Knowledge Based Systems
KE	Knowledge Engineering
KM	Knowledge Management
KPI	Key Performance Indicators
OE	Ontology Engineering
SCM	Supply Chain Management
TA	Template Approach

Glossary

Data Mining

The process of extracting valid, previously unknown, comprehensible and actionable information from large databases and using it to make crucial business decisions (Simoudis 1996).

Decision Support System

Computer program applications that analyze existing data and present results to users in order to support their decision making (Power 1997)

Design oriented research

The type of research that aims at revealing new knowledge through designing an artefact or prototype (Fallman 2007)

Deviation

The situation that the value of a performance indicator goes out of the prescribed range

Early Warning System

Based on the knowledge of relations between deviations in performance of food supply networks and determinant factors, an early warning system in food supply networks predicts as early as possible potential deviations for decision makers in food supply networks by monitoring measurable determinant factors.

Expert System

An expert system contains facts, heuristics, and inference procedures to mimic the problem solving processes of experts.

Expert validation

A method to obtain expert opinions on whether designed system fulfills its intended purpose

Food Supply Chain Networks

A FSCN is composed of multiple food supply chains that are interconnected with each other and are joined into a network.

Framework

A framework for a type of systems describes the necessary components as well as the relations between those components in that type of systems

Knowledge Base

A special kind of database for knowledge management, providing the means for the computerized collection, organization, and retrieval of knowledge (Akerkar and Sajja 2009)

Knowledge Based Decision Support System

A type of Decision Support Systems that includes an Expert System as one of its main components(Klein and Methlie 1995).

Knowledge Based System

A system normally contains a Knowledge Base for storing knowledge and an Inference Mechanism to retrieve relevant knowledge from Knowledge Base for a query from users (Klein and Methlie 1995)

Knowledge Engineering

Knowledge Engineering aims at modelling various aspects of domain experts' problem solving expertise, and hence producing Knowledge-Based Systems to help non-experts in dealing with problems (Schreiber, Akkermans et al. 2000)

Knowledge Management

Knowledge Management comprises of a set of practices used by organizations to identify, select, organize, disseminate, and transfer important information and expertise that are part of the organization's memory and that typically reside within the organization in an unstructured manner (Herschel and Jones 2005)

Key Performance Indicators

A set of measures focusing on those aspects of organizational performance that are the most critical for the current and future success of the organization (Parmenter 2007)

Manifest knowledge

Manifest knowledge represents the objects and their properties (e.g. about operations, staff, production means, environmental indicators, and performance indicators) and those relations between them that people can perceive without need for data analysis or inference.

Inferred knowledge

Different from Manifest knowledge, people have to analyze data or employ domain experts in order to arrive at Inferred knowledge.

Ontology

An ontology is a formal description of domain knowledge. It consists of concepts, their properties and relations, structured according to human conceptualization (Gruber 1995).

Ontology Engineering

Ontology Engineering deals with organizing domain knowledge by formal descriptions of the concepts with their properties and relations. The aim is to improve knowledge sharing by structuring the concepts according to human understanding (Antoniou and Harmelen 2004).

Proactive Control System

Based on the knowledge of relations between deviations in food supply networks and determinant factors, proactive control systems in food supply networks propose appropriate actions in order to prevent potential deviations which have been flagged by the early warning system. The potential actions could be discarding, taking corrective measures, or adapting subsequent processes in order to make amends.

Supply Chain Management

The planning and control of all business processes to deliver prescribed products at the right time and right place, and also to satisfy requirements from stakeholders

Template Approach

A Template approach provides wizards that support users in executing a specific task.

Chapter 1. Introduction

1.1 Introduction

Food products often have problems concerning quality and safety (Roth, Tsay et al. 2008), as can be seen from the high numbers of product recall announcements found almost weekly in newspapers or websites. As a result, European consumers are highly conscious of food quality and safety (Poppe and Kjærnes 2003). This concern has been strengthened by a series of food safety crises (Bovine Spongiform Encephalopathy (BSE), dioxin crisis, classical swine fever, Foot and Mouth Disease (FMD), etc.). The European Commission carried out a research project and found that 11% of all food products that are checked by the European Union (EU) do not comply with the demands of EU legislation (Beulens 2003). Problems in food quality and safety have brought about not only huge losses to the food companies involved, but also risks to the health of consumers.

Food companies are troubled not only by quality and safety problems, but also by logistics problems. Such problems occur in the flow of goods and resources, such as energy and people. The consequences include bad delivery performances and stock outs in retail outlets. Since reliable and responsive delivery has become increasingly important in recent years, the resolution of these problems has received increased management attention as well.

Mockler (1972) defined management control as the efforts being made to "assure that all corporate resources are being used in the most effective and efficient way possible in achieving corporate objectives". There are different ways that managers can deal with an encountered problem. The simplest way is to ignore the encountered problems and take no action at all. However, the consequence is that managers are then faced with all the losses and the adverse impacts on consumers. Or managers can take some measures to correct the detrimental effect of the problems, like abandoning problematic food products. They can also adapt the operations in their supply chains to prevent upcoming problems. Different control measures have different outcomes and impact. For the benefit of the food industry as well as consumers, it is important that management control is effective and cost efficient and thus prevents logistics, food quality and food safety problems.

There are a number of aspects that make it difficult to control problems efficiently and effectively. Food products are handled in multiple stages of the food supply chain before they reach consumers. Each food supply chain contains multiple stages of operations, from raw material processing to final product delivery. Those supply chains interconnect with each other at certain stages and form a network. This sort of network is called a Food Supply Chain Network (FSCN). Interconnections at certain stages allow resources and information to be exchanged between stakeholders (such as farmers, transporters, retailers) from different supply chains (Pérez and Martínez 2007). In order to efficiently and effectively control problems in FSCN, managers need extensive information, not only about the operations in each stage, but also about the ingredients and raw materials involved in each stage.

Nowadays, large amounts of data are recorded in FSCN without being used to improve control. The new European Food Law on traceability states that all movement of product and steps within the production process should be recorded. In order to comply with such requirements, current Information Systems in FSCN have been extended considerably with the information gathered from various resources (Meulen 2008). One of the most important resources is the monitoring system; this is the system that provides information about operations in FSCN. Monitoring systems in FSCN continuously collect information from each stage about the input materials, about how those input materials are processed, and about the results obtained (Petersen, Knura-Deszczka et al. 2002). New techniques have been invented to automatically collect data, such as RFID (Hsu, Chen et al. 2008). Such information resources provide possibilities to detect problems in operations, as well as in attributes of food products. By consulting this sort of information resources, it is possible to find out the causes of such problems. For example, when the average muscle thickness of a herd of pigs goes below the normal range of other pigs, it will be helpful to identify the cause by looking into the information about feed, feeding method, etc.

Data Mining (DM) has become a popular research area in the last decade (Fayyad, Piatetsky-Shapiro et al. 1996). It provides various methods for investigating existing data on relations between attributes in production and the quality problems in the output. One of the propositions in this research is that Data Mining methods should be made accessible to managers in FSCN. We believe that it is beneficial for their decision making if they could employ the power of Data Mining methods to analyze recorded data. They will have the potential to quickly identify which operational factors actually caused problems in the final product. They will also be able to predict which product will have similar problems. However, due to the complexity in using Data Mining methods, managers often cannot use them. There is a need for a decision support system that can enable managers to use DM methods on the extensive data collected in the supply chain.

This research aims to enable managers to use DM methods on existing data resources for predicting and preventing problems in food production. In this thesis, we developed an Early Warning and Proactive Control (EWPC) system to help managers forecast upcoming problems, and take action to prevent predicted problems from happening. The EWPC system is a Knowledge-Based Decision Support System (KBDSS). It incorporates the necessary knowledge for using Data Mining methods. Managers need to know which DM method is suitable for their specific problem situation, as well as how to apply a DM method to the available data. The system should direct them through all the steps that are required to explore causal relations, to predict upcoming problems, and to evaluate different control measures for proactive control. The system should also store the knowledge obtained from those Data Mining steps, so that such knowledge can be reused for early warning and proactive control when the problem situation recurs. Therefore, this system should allow different managers to share their obtained knowledge.

By predicting and proactively preventing problems in food quality, food safety, and logistics, an Early Warning and Proactive Control system is valuable not only to science, but also to business and society. With such a system, managers in FSCN can achieve better performance through proactively dealing with problems encountered in daily operations. Consumers can also benefit from better controlled food quality and food safety and improved product availability.

1.2 Theoretical framework

From an extended literature review we identified a number of areas that are relevant for the design of EWPC systems in FSCN. We will discuss each of these areas in more detail in the following subsections:

- Systems thinking, as it specifies how to look at the FSCN;
- Supply Chain management, since it provides the context for EWPC system;
- Decision Support Systems (DSS), because current practice in DSS design gives references on designing the architecture and components of intelligent systems for problem investigation
- Knowledge Based Systems, because the advances in this area offer methodologies and tools to develop facilities for knowledge sharing; and,
- Data Mining, because it provides various methods for data analysis in this area.

1.2.1 System view on FSCN

Systems thinking is used by social sciences as a method of understanding real-world phenomena (Jackson 1993). In systems thinking, the reality is regarded to be constructed of systems and their environments. A system is a structured set of functional units together with the relationships between them (Wilson 1993). Those functional units work together and fulfill a certain function in the environment. A system can be a subsystem of a wider system, and at the same time, contain multiple subsystems. A food supply chain can be seen as a system which is composed of multiple subsystems (organizations) together with the relationships between them (Van der Vorst 2000).

De Leeuw (2002) uses a systems approach to structure a management context which describes an arbitrary control situation for decision-making purposes. He states that organizations consist of three aspects: the object system, the control system, and the information system, together with their common environment (see Figure 1-1). The object system refers to the primary transformation processes. In a food supply chain, all the operations, like production and transportation, are treated as an Object System. The task of information systems is to register the relevant internal and external data and convert them to control information. The control system aims at realizing a certain system output by adjusting control variables (such as transport time). It takes decisions based on available information. The system output is measured by Key Performance Indicators (KPI). KPI is a set of measures focusing on those aspects of organizational performance that are the most critical for the current and future success of the organization (Parmenter 2007). In the following section we are going to map this management paradigm on the management of FSCN.

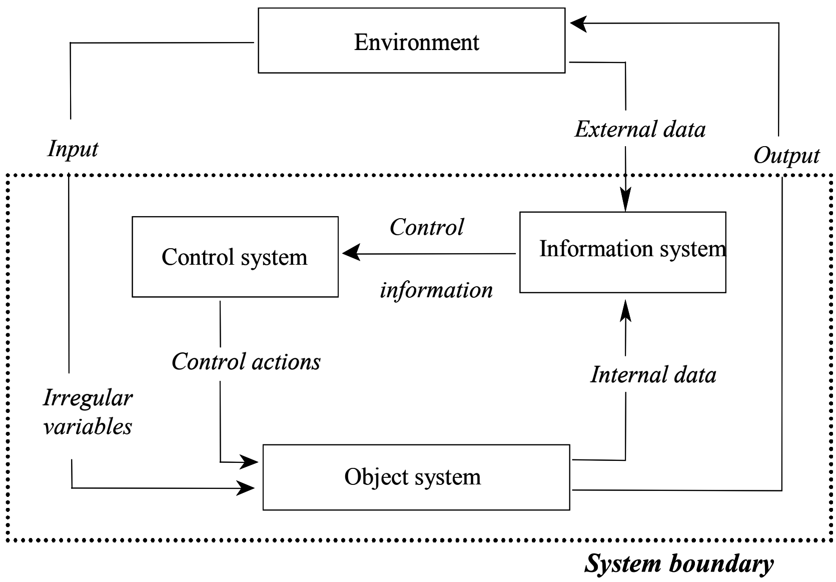


Figure 1-1: The management paradigm

<u>Object System</u>

A FSCN is composed of multiple food supply chains that are interconnected with each other and are joined into a network. The object system of this thesis is the FSCN.

A supply chain can be segmented as a sequence of processes and material flows that aim to meet final customer requirements (Van der Vorst, Beulens et al. 2005). The production of food products normally starts at farms and includes different stages—such as processors, packers, distributors, transporters, and retailers—before reaching consumers. In each stage, there are different actors involved, such as raw material (such as feed) suppliers, operators, and transporters. Figure 1-2 shows an example of a food supply chain.

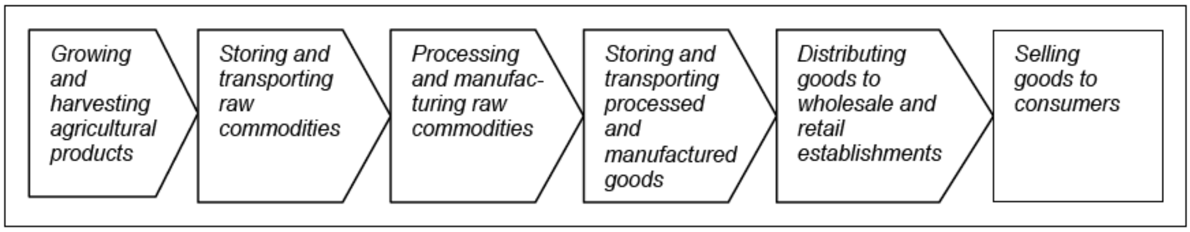


Figure 1-2: An example of a food supply chain (Levinson 2009)

Figure 1-3 depicts a generic supply chain at the organization level within the context of a complete supply chain network. Each processor is positioned in a network layer and belongs to at least one supply chain: i.e. it usually has multiple suppliers and customers at the same time and over time. Other actors in the network influence the performance of the chain.

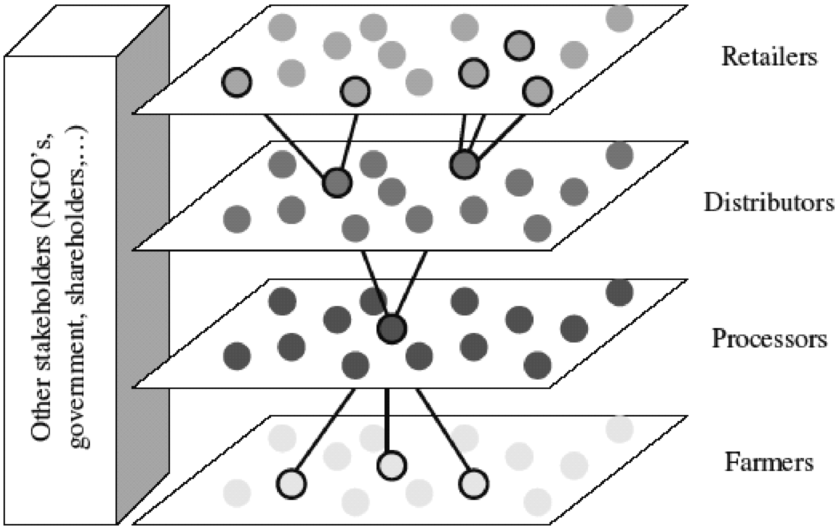


Figure 1-3: Schematic diagram of a supply chain from the perspective of the processor (bold flows) within the total FSCN (Van der Vorst, Beulens et al. 2005)

Such a process network allows managers to share resources and information with each other. If arranged well, managers can get easier access to technical know-how and financial support, and enhance the overall competitiveness of FSCN.

<u>Control system</u>

Decisions made for management control may differ in aspects such as time horizon, frequency, and level of details (Anthony 1998). Three levels of logistical management have been distinguished based on the natures of the decision being made: strategic control, tactical control, and operational control (Anthony 1998; Van der Vorst 2000).

- Strategic control is the design of strategies to achieve goals at the organizational level, such as production site selection, merger and acquisition. Strategies are normally made with longer time horizon, e.g. annually, or multi-annually.

- Tactical control is concerned with objectives for the coming weeks or months, such as arrangement of people and resources to meet actual demand.

- Operational control is concerned with daily operations - how to operate facilities to fulfill order requirements, and at the same time to make the most profit.

In this research, the emphasis is put on the operational control, i.e. to improve the day-to-day operations. In current FSCN, there are monitoring systems that continuously collect information about different aspects of food production, such as primary inputs, output products, and operations.

The key question is: based on such information resources, can we change something in the control system to improve the performance of daily operations in food supply chains?

In order to improve the performance of FSCN many Key Performance Indicators (KPI) are defined and put into practice. A KPI allows managers to quantify the performance of FSCN. Currently there is quite some research on KPI for FSCN (Ondersteijn, Wijnands et al. 2006; Aramyan 2007). KPIs are defined and categorized to measure different aspects of FSCN, such as food quality, and logistic performance. To arrive at acceptable performance, those KPIs should be controlled within a reasonable range. Deviations from reasonable ranges will incur losses to the stakeholders in FSCN.

Definition of deviation: *A deviation is the situation in which the value of a performance indicator goes out of the prescribed range.*

In an EWPC system, deviations in the KPIs that relate to daily operations are the target to be controlled. Discrepancies between the prescriptions and the actual process outcomes should be controlled to meet the standards. For example, the weight and dimension of food products, or the delivery time, should all be managed within an acceptable range. To prevent deviations on the performance of FSCN, it is important to have effective and efficient control at an earlier stage.

Information system

Information systems provide storage for the information collected by monitoring facilities in food supply chains. To realize effective control of operations, managers need sufficient information on those operations, e.g. its performance history, current status, actors involved, etc. Such information comes from information systems. Current traceability and transparency efforts in food supply chains provide tremendous resources for such information. Traceability is 'the ability to document and trace forward and backward a product (batch) and its history through the whole, or part, of a production chain from harvest through transport, storage, processing, distribution and sales' (Van der Vorst 2004). Traceability consists of tracking the course of a product through the chain network while it is being produced, or tracing back the history of a product (Hofstede, Beulens et al. 2004). According to Moe (1998), the internal traceability in the production step could give cause-and-effect indications when a product does not conform to standards. These indications and available data will be helpful for us to investigate the possibilities to control the deviations.

In order to improve the performance of supply chains, van der Vorst (2000) identifies transparency as a generic redesign principle for information systems in supply chains. The notion of transparency takes a wider stance than just traceability. Hofstede (2004) states that transparency of a netchain is the extent to which all the netchain's stakeholders have a shared understanding of, and access to, the product-related information that they request, without loss, noise, delay and distortion. Many researchers have been actively involved during the last decade in the area of FSCN and in particular in the area of transparency systems, tracking and tracing systems, and associated Information

Infrastructures (Dorp 2004; Jedermann and Lang 2007; Kelepouris, Pramatari et al. 2007). These systems are necessary to:

- Manage and control performance of FSCN processes (in particular logistic processes);
- Inform and report with integrity about these processes, products and resources used;
- Perform and control emergency and recall processes when necessary.

These systems are to be developed, implemented and maintained in an FSCN context involving many participants. All these developments will provide considerable resources for Early Warning and Proactive Control.

Requirement for effective control

De Leeuw (2000) indicates that the conditions for effective control are:

1. Objectives of control – e.g. which KPI is the target, what is the acceptable value range of this KPI?
2. Information about the state of the object system to be controlled – e.g. what stages exist in the production processes, how did the operations and quality of input material change in certain stages
3. Information about the environment – e.g. what constraints have been posed by the outside environment, e.g. new legislation on food ingredients
4. Capacity of information processing. There should be enough information processing capacities to process information on the environment and system state.
5. Sufficient decision alternatives – e.g. what are the available resources (people and facilities), and the decision varieties that are supported by those resources, such as extra capacity for transportation
6. A model of the object system to be controlled. This model relates the variables mentioned in points 1 through 4. A model bearing the relations between the control objective and factors (e.g. quality of input materials, operations) in production process supports managers in choosing decision alternatives.

For managers in FSCN to achieve early warning and proactive control, the focus should be on the model of the object system. As discussed above, currently there are many KPIs defined to measure different aspects of FSCN, such as food quality and logistical performance. There are also monitoring systems that continuously collect data about the status of operations as well as quality of food products at different stages. The capacity of information processing has considerably improved due to the advances in information systems in FSCN. Information about the environment and decision alternatives is part of the expertise of managers in FSCN. So what managers need is the sixth requirement: a model of the object system.

In our research, we aim at designing EWPC systems that will enable managers to build models for analyzing encountered problems. To deal with the complexity in the processes in FSCN, we have to choose a proper abstraction level for the modelling. The details below the chosen level do not need to be taken into account during modelling. This can be achieved with a multi-input-multi-output (MIMO) approach (Jansen 1998) (see Figure 1-4). The MIMO approach splits up the complete production chain into several production processes. Each of the processes can be described as a black box with multiple inputs and multiple outputs. The outputs of one process serve as inputs of other processes. Therefore, a FSCN can be seen as a network of connected MIMO models. For each production process, inputs include the supply of input materials, states of production means, energy, recipes, control decisions, and procedures. Inside the black box a transformation occurs. Outputs may include primary products, co- and by-products, used production means and waste. For the modelling in EWPC systems, the contents inside the black box should not be the focus for investigation. Instead, the focus should be the inputs and outputs.

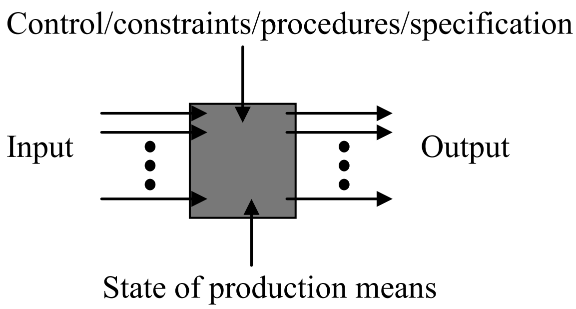


Figure 1-4: MIMO model

The transformation of input materials into output materials takes place within particular control constraints. The input materials should comply with predefined requirements. The production should also be executed according to such constraints. When all the constraints are satisfied, the properties of output materials are assumed to have variation within a specified range. Deviations from constraints will cause deviations in performance of outputs. For example, if a herd of pigs have a much higher fat thickness than others, the output pork products are expected to contain more fat than the products from other herds. In order to deliver food products as required, it is necessary to control the variations in food production within specified ranges.

1.2.2 Supply Chain management

Supply Chain Management (SCM) is an approach that focuses on business processes (Trienekens and Beulens 2001). SCM is the planning and control of all business processes to deliver prescribed products at the right time and right place, and also to satisfy requirements from stakeholders (Van der Vorst 2000). A business process is a structured, measured set of activities designed to produce a specified output for a particular customer or market (Davenport 1993). The 'planning' in the definition of SCM refers to the designing of processes in a supply chain at the strategic level (e.g. what should be the processes involved, who should be the stakeholders), as well as tactical and operational level. The 'control' refers to the identification of improvement options for processes in a supply chain in case a problem occurs: to improve the effectiveness and efficiency of communication between different processes, to decrease quality problems by enforcing stricter checks at each production stage, etc. Managers in food supply chains are responsible for the planning and control of all the stages involved in order to deliver food products at the required time and place with pre-specified quality standards.

The focus of this thesis is at the operational level of Supply Chain Management. We aim at building EWPC systems that deal with deviations in production processes of FSCN. Based on the knowledge of relations between deviations and determinant factors, EWPC systems should predict as early as possible potential deviations for decision makers by monitoring the operations in FSCN. Such systems should also propose appropriate actions in order to prevent those potential deviations. The potential actions could be discarding, taking corrective measures, or adapting subsequent processes in order to make amends. The timeliness of prediction and control is very important. The early warning signal should be released early enough to allow managers to take corrective actions. To effectively prevent an upcoming problem, it is important to find the cause of such problems. Therefore, EWPC systems should be able to identify the origin of encountered problems.

The current practice of process control in food SCM has a fixed problem as the target, such as salmonella infection (Attenborough and Matthews 2000; Rose, Mariani et al. 2003). Specific models have been built by researchers or specialists in the problem area. Managers use this sort of fixed models to deal with similar problems in their own supply chains. However, when the encountered problem is different from that being modelled, managers can no longer apply the model. The EWPC systems we aim for in this thesis should not have such restrictions. They should help managers to achieve flexible process control. As long as there are historical data recorded for the encountered problems and for the related production processes, managers can use EWPC systems to explore the causal factors. Once this causal relation is obtained, managers can effectively deal with encountered problems.

The efforts towards transparency provide resources for flexible process control. To accomplish flexible process control, it is necessary to be informed in time about the current status of the system and changing conditions. Such information allows managers to observe what happened and adjust the control mechanism accordingly. In order to model the relations between encountered problems and determinant factors, historical data about the production processes should also be stored. The fact that SCM focuses on the whole supply chain structure instead of individual stages urges the information systems in supply chains to satisfy those requirements. Monitoring systems in FSCN continuously collect data about daily operations as well as products in all the processes and store such data in information systems. Such information allows people to explore what the determinant factors are. It also provides a description of the current situation in FSCN and changes in inputs (materials, resources, etc.).

The findings from the operational level could contribute to decisions at the strategic level. For example, if the feed from a particular supplier consistently caused performance deviations in the output food products, the executives may decide to sign a contract with another supplier. Therefore, EWPC systems are potentially valuable in strategic control as well.

To accomplish flexible control for problems in FSCN, there should be modelling facilities for managers to model the relation between inputs and performance deviations. For this, we look for solutions from advances in Decision Support Systems.

1.2.3 Decision Support Systems

Decision support systems (DSS) are computer program applications that analyze existing data and present results to users in order to support their decision making (Power and Sharda 2007). Simple DSS could be just comparing facilities on some key indicators, like mean value and standard deviation. Normally DSS contains built-in models which describe the relationship between inputs and outputs, such as a model to predict revenue based on the stock level (Mannino and Walter 2006). DSS has been used in different areas, like finance (Ince and Trafalis 2006), marketing (Jonker, Piersma et al. 2006), and agriculture (Recio, Ibanez et al. 2005). In FSCN management, DSSs are often designed and used for different purposes, such as detecting missing products (Papakiriakopoulos, Pramatari et al. 2009), vehicle routing (Mendoza, Medaglia et al. 2009), and inventory control (Shang, Tadikamalla et al. 2008).

In order to accomplish flexible control in FSCN, it is necessary to design a new type of DSS. The capacity of conventional DSS is limited by the pre-built models inside the system. For example, in a DSS for prediction of microbial decay (Vaikousi, Biliaderis et al. 2008), relations between microbial growth and ingredients, as well as water activity and pH value are modelled. When managers are faced with similar situations, they can use this model to predict microbial problems. But if a manufacturer uses a new ingredient in its food production, and such ingredient has not been captured in the model, managers need a new model. In FSCN, all kinds of new problems in food quality and/or food safety occur frequently. Therefore, managers need a new type of DSS that enables them to build new models for encountered problems. With ample information being recorded in Information Systems, this type of DSS can improve management control of processes in FSCN.

Two areas can contribute to this new type of DSS: Knowledge Based Systems and Data Mining. Flexible Process Control requires the capacity to handle the rapidly changing information about the process. Knowledge Based Systems (KBS) can contribute to this sort of capacity because they are able to quickly process information and easily incorporate new knowledge (Lee 2000). One of the key components in KBS, a Knowledge Base, can be used to store built models. With KBS, users can consult the Knowledge Base for possible causes for the encountered deviation, and also for different control measures to limit the adverse impact. Data Mining is capable of deriving novel knowledge from existing data sets (Fayyad, Piatetsky-Shapiro et al. 1996). By applying DM methods on existing data, it is possible to build new models for encountered new problems. Such new models can therefore also be put into a Knowledge Base.

1.2.4 Knowledge Based Systems

Knowledge Based Systems (KBS) are systems based on Artificial Intelligence techniques. The major components of KBS are *Knowledge Base* and *Inference Mechanism* (Okeefe and Preece 1996). KBS obtain information from users regarding the encountered problems and search the Knowledge Base for relevant solutions. The inference mechanism allows KBS to mimic the actions of acknowledged experts. It selects relevant knowledge from the Knowledge Base and combines such knowledge to gather useful information for the users. New knowledge in the domain area can be easily incorporated into the Knowledge Base after being converted into a suitable format. KBS has been applied in various areas, such as medical diagnosis (Wolff 2006), the steel industry (Naylor, Griffiths et al. 2001), and agriculture (Prasad, Ranjan et al. 2006).

Knowledge base is one of the key components in KBS (Okeefe and Preece 1996). But when this research was started, we had no existing knowledge base for causes and remedies for problems in FSCN. Therefore, we had to design a knowledge base to store the knowledge about problems in FSCN, such as their symptoms, causes, and remedies. Considering the fact that new problems occur very often in FSCN, it is beneficial if the knowledge base can be extended by different managers.

KBS is limited by the knowledge in the knowledge base (Gopalakrishnan and McCoy 2008). If the encountered problem is out of scope of the Knowledge Base, then KBS cannot provide users with sufficient answers. Therefore, there is a need for a system with the capacity to diagnose a new problem for which there is as yet no existing knowledge.

1.2.5 Data Mining

Data Mining is the process of extracting valid, previously unknown, comprehensible and actionable information from large databases and using it to make crucial business decisions (Simoudis 1996). There are two approaches: verification driven, the aim of which is to validate a hypothesis postulated by a user, or discovery driven, which is the automatic discovery of information by the use of appropriate tools. The underlying principles of the Science Method, being the cycle observation-hypothesis-experiment, fit well with the processes of Data Mining. The observation-hypothesis step means to first observe the phenomena in practice and then generate a hypothesis towards the underlying principles. The hypothesis-experiment steps means to design and implement experiments to verify the correctness and validity of the generated hypothesis.

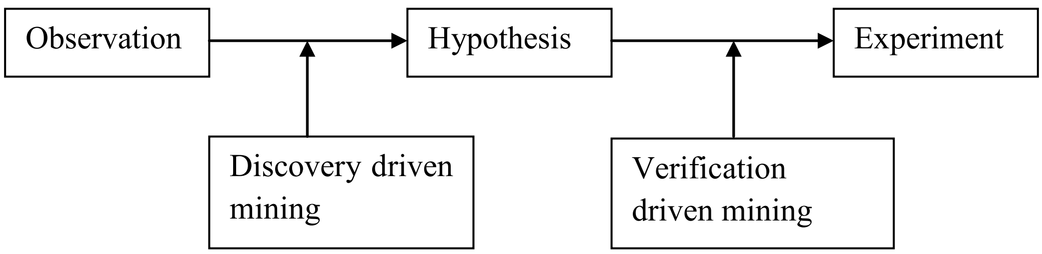


Figure 1-5: Data Mining fits into scientific environment

As shown in Figure 1-5, the discovery driven mining fits the observation-hypothesis step; and the verification driven mining fits the hypothesis-experiment step (Crawford and Crawford 1996). If a manager has no idea about the cause of an encountered problem, he can use DM methods to explore possible causes and hence get a hypothesis. If the manager has some potential causal factors in mind, then he can use DM methods to verify whether that hypothesis is correct or not. Currently, various methodology supports have been provided (Adriaans and Zantinge 1996; Fayyad, Piatesky-

Shapiro et al. 1996; Chapman, Clinton et al. 2000), and Data Mining has been successfully applied in a wide variety of contexts and applications, such as financial analysis (Ziarko, Golan et al. 1993), space data analysis (Fayyad, Djorgovski et al. 1996), and agriculture (Verdenius 2004).

In the DM area, there is no existing research on the applicability of DM methods for EWPC. The applicability of DM methods has always been an important topic. The applicability of Data Mining techniques is as important as their technical performance (Verdenius 2004). In the DM research area, much effort has been expended on the applicability of DM methods (Brodley 1992; Rajan and Saravanan 2008). But existing research on DM applicability is not tailor-made for EWPC in FSCN. The result cannot be directly used in EWPC system design. This research takes the specific requirements from EWPC in FSCN into consideration. For EWPC purposes, we have to analyze the fitness of Data Mining techniques (or combination of techniques) for modeling relations between (attributes of) processes and performance in FSCN. Based on that, we aim to design an expert system that presents managers with suggestions of proper DM methods for their specific problem situation.

Although Data Mining methods have been applied in various places in FSCN, we could not find any research about enabling managers to use DM methods. All those applications of DM methods are executed by people who are knowledgeable in Data Mining. In our research, we investigate how to guide non-experts in DM to execute the right steps in applying DM methods for EWPC. In order to enable managers, as non-experts in Data Mining, to apply DM methods on their data sets, there are two issues to be solved. Firstly, how to enable managers to select a suitable DM method for their problem situation. Secondly, once a suitable DM method is chosen, how to enable managers to use that DM method correctly.

1.3 Research objectives

As discussed in section 1.2.3, we need a type of system that can deal with new problems in FSCN. In this research, we have two major objectives:

Objective 1.

To design a framework for an EWPC system that is able to facilitate the following aspects:

- analyze relations between problems and causes
- predict upcoming problems
- suggest control actions to prevent upcoming problems
- use existing data bases in FSCN
- support non-expert users in applying DM methods
- have an extendable knowledge base

The framework should describe the necessary components as well as the relations between those components in EWPC systems.

Objective 2.

To build a prototype system based on the framework to enable managers in FSCN, as non-expert in DM, to use DM methods for Early Warning and Proactive Control on the supply chain level.

Early warning

To realize early warning, managers have to be informed about defects and to be aware of possible deviations when they are just beyond the time horizon. Consequently, they may reduce risks and improve the performance of food supply chains. The system should enable managers to quickly find the relations between the deviations in performance indicators and the determinant factors. It should also enable managers to predict upcoming problems before their occurrence as early as possible based on the information from monitoring systems.

Proactive Control

Proactive control means adapting the control system to take repair actions (such as reparation of intermediate processes, or adaptive actions by subsequent processes) early enough to prevent any further deviations downstream in the supply chain. Recognizing these deviations and taking appropriate actions can prevent losses due to quality problems or even improve performance by using optimal parameter settings. In order to accomplish proactive control, managers need to know what actions can be applied in order to prevent deviations from happening. In other words, they have to find the relations between the actions and observed and expected deviations. Therefore, research is needed on what kinds of facilities are capable of accomplishing the support needed for early warning and proactive control.

This research intends to find appropriate solutions for achieving flexible process control for FSCN management. We are looking for a facility that can enable managers to analyze relations between (attributes of) processes and performance in FSCN. In particular, this project aims at a system that can help managers to model the object system at hand using a variety of Data Mining techniques. Such models will show the relations between problems and their causes in actual FSCN. Managers can use these models to predict upcoming problems and prevent them at an early stage.

By means of a repository of modelled relations, the system can be applied to subsequent problems in similar areas. Moreover, this research will give insight into the applicability of various Data Mining techniques that can contribute to the analysis of this kind of FSCN.

1.3.1 Scientific relevance

As mentioned in the theoretical framework, we want to provide a scientific contribution to specific areas: (1) In SCM, we present an approach for flexible process control in FSCN. (2) In DSS, we design a new type of DSS architecture which can deal with new types of problem situations. (3) In Knowledge Based System, we propose a Knowledge Base structure for managers to accumulate and share obtained knowledge. (4) In DM, we investigate the fitness of Data Mining techniques (or combination of techniques) for modelling the relations between (attributes of) processes and performance in FSCN. We also research what kind of support should be provided to managers in order to enable them to correctly use DM methods for EWPC.

For SCM: an approach for flexible process control in FSCN

Data analysis has often been used in solving problems in FSCN (Liao, Chen et al. 2009). However, normally data analysis is executed by DM or statistical experts, not by managers in the real-life food industry. The fact that new problems occur frequently in FSCN implies that this type of problem-solving is not flexible enough. This research intends to realize flexible process control by offering managers a tool, instead of the result, to deal with encountered problems. In that way we extend the knowledge on SCM regarding efficient and effective control of food supply chains.

<u>For DSS: A new type of DSS that can deal with new problems</u>

Conventional DSSs depend on fixed models, which incorporate the knowledge on the causal factors for particular kinds of problems. This research aims at a new type of DSS, which can deal with new problems that are out of the scope of existing models, as long as relevant data is available. The architecture of this new type of DSS contributes to the design of DSS for dealing with unexpected problems.

<u>For Knowledge Based System: a Knowledge Base to enable knowledge sharing</u>

The knowledge obtained by managers from data analysis is beneficial for other managers with similar problems. By sharing acquired knowledge, managers will have more information to support their decision making. In this research, we study what types of knowledge are necessary for EWPC in FSCN. We design a Knowledge Base structure that is suitable for storing and sharing those types of knowledge. Such a structure is a reference for other areas where there is a need to deal with various kinds of problems, and an extendable Knowledge Base is beneficial for efficient control.

<u>For DM: Applicability of DM method & Support for non-expert users</u>

As discussed in section 1.2.5, for this research there are two important topics in the area of Data Mining: applicability of DM methods and supporting non-expert users in Data Mining. With overwhelming data recorded already, this project tries to find applicable Data Mining for early warning and proactive control in FSCN. Moreover, we investigate what kinds of support are needed for non-expert users. Such knowledge has a reference value for other domain areas where it is beneficial for non-experts to use DM methods.

1.4 Research questions

Before the system is designed, it is important to know what functions this system should accomplish in order to enable managers to accomplish EWPC in FSCN. In addition to functional requirements, there may be other concerns, such as on the ease of use, and time allowed. So the first research question is:

RQ1: What are the requirements for EWPC system design considering current practice of FSCN management?

The framework should specify what the key components of EWPC systems are, including the relations between those components. Furthermore, it should explain how managers can use the system (based on this framework) to accomplish EWPC in FSCN.

RQ2: What components should be included in the EWPC systems, and how should those components cooperate to enable managers to achieve EWPC in FSCN?

As discussed in section 1.2.5, it is necessary to research the applicability of DM methods for EWPC in FSCN in order to let the system automatically suggest a proper DM method.

RQ3: What Data Mining methods are available and applicable for EWPC in FSCN?

As discussed in section 1.2.5, research is needed on how to enable managers in FSCN, as non-experts in DM, to use DM methods for EWPC.

RQ4: What support needs to be provided to managers in order to enable them to use Data Mining methods for EWPC?

EWPC systems need a particular kind of Knowledge Base (see section 1.2.4). This type of Knowledge Base should be able to accommodate different types of knowledge obtained by managers in the process of EWPC. Moreover, this Knowledge Base should be easy to use and easy to extend by managers in FSCN. We have to research what structure is suitable for this type of knowledge base.

RQ5: What kind of structure is suitable for the Knowledge Base in EWPC systems?

After the system design, it is necessary to check whether this system is correctly designed, and also whether it can accomplish its objectives.

RQ6: What is the validity of the designed framework and prototype system?

1.5 Research method

This section explains how we executed this research. We first explain the research methods being used in this design oriented research. Then we introduce the sources of data being used. At the end, we present the research method for each research question.

1.5.1 Design oriented research

This research is a Design Oriented Research. First and foremost, our research questions are answered by designing and evaluating a prototype EWPC system. Design oriented research is the type of research that aims at revealing new knowledge by designing an artefact or prototype (Hevner, March et al. 2004; Fallman 2007). As with research in natural science, gaining scientific knowledge is the main objective. Contrary to natural research, the way to gain knowledge is not by experimenting in lab settings, but by designing a system or prototype (March and Smith 1995). Design oriented research is especially needed when "the type of knowledge could not be attainable if design – bringing forth of an artefact (e.g. research prototype) – had not been a vital part of that research process" (Fallman 2007).

In our research, we design a prototype instead of a fully implemented system. In design oriented research, the main contribution is the 'knowledge that comes from studying the system in use or the process of system design' (Fallman 2007). The developed artefact is the means to achieve the main contribution. This implies that the developed product does not need to include all the functions that a final system would have. Contrary to that, researchers in design oriented research often work with sketches and prototypes.

A design process is always iterative and incremental. As mentioned by Van Aken (2004), the typical way to design is to use multiple cases. The knowledge obtained from one case can be refined in the next case. The designing of a prototype EWPC system benefits from real case experiences. Since the artefact developed in design oriented research is supposed to be placed in the real world, it is necessary to take the reality of problem settings and potential users into consideration. Yin (1994) introduced three criteria for determining the appropriate research methods for a particular research setting. According to Yin, case analysis is the preferred strategy when

- the research deals with operational links needing to be traced over time, rather than mere frequencies or incidence;
- the investigator has little control over events (unlike in an experiment);
- the focus is on a continuously changing phenomenon within some real-life context.

The situation we are facing in this research is in accordance with those conditions. In this research we are trying to analyze the causal links in an FSCN that are too complex for other research strategies (such as survey, etc.). We as researchers have no control over the everyday operations in real-life FSCN in food companies. Furthermore, changes always occur in some stages of FSCN, from the primary input to the final output product. Therefore, we use real-life cases as well as available data from those cases in this research. In addition to case analysis, we also base our research on literature reviews from different areas, such as SCM, DM, and DSS.

In order to ensure the validity of findings obtained from this research, it is necessary to evaluate the designed prototype system. Hevner et al. (2004) emphasize that in design oriented research it is necessary to rigorously evaluate the designed artefact with a well-executed evaluation method. According to Van Aken (Van Aken 2004), design oriented research in management is more complex than in other disciplines like medicine or engineering. Management tends to have more varieties in context, and the number of cases that can be used to generalize obtained knowledge is much lower. Moreover, it is often difficult to get the acquired knowledge materialized and verified. Therefore, in this research we use Expert Validation as a method to evaluate the product of this research. Expert Validation can be executed with a few cases. Furthermore, it is a method for obtaining evaluation results from multiple experts that come from different areas (science and business management). Therefore, we can obtain comprehensive evaluation results, not only on the validity of the designed artefact, but also on the correctness of design procedure.

1.5.2 The use of practical cases

Two food companies (Nutreco and VION) are used as the source of cases in this project. Nutreco is an international food company present at various stages of the fish, poultry and pork production chains. A major proportion of salmon and poultry products are put on the market through the company's own marketing and distribution channels under the company's own labels. Nutreco's leading position in the feed and food chains brings major responsibilities with respect to producers, distributors, retailers, consumers, the society and the environment. These responsibilities mainly concern food quality, in which food safety is an essential element.

Nutreco implemented an information system for food quality and related information throughout its company. This information system consists of four pillars: Certified Quality, Monitoring, Tracking & Tracing and Risk Management. In Tracking & Tracing it gathers and stores the information generated by all activities, from raw materials purchasing through to delivery of final products; either feed or food. This way Nutreco can provide, at the "push-of-a-button", the details of a product's history in its chains.

VION is a global meat company with the focus on pig supply chains. They have agreements with many pig farms and collect pigs from them at regular intervals. With several slaughter houses at different locations, VION collects extensive information about each pig that was slaughtered. The information being recorded includes not only farm performance of slaughtered pigs, but also the meat quality, and residual levels of medicine.

During our system design we used real cases from those food companies. From those cases we studied what kind of system would help managers to implement EWPC in FSCN. During case analysis, we communicated with managers from those companies about the problems they encountered, the relevant data sets, and the objectives they wanted to achieve. By applying different DM methods to those cases, we accumulated knowledge on the applicability of those methods as well as on the generic processes of applying those methods for EWPC. Moreover, we categorized the types of knowledge obtained from problem investigation in order to design a proper structure for Knowledge Base. All this knowledge contributes to the design of framework and prototype for the EWPC system in FSCN.

1.5.3 Mapping between research question and methods

Based on case analysis, expert input, and literature review, we obtain answers to the different research questions. Table 1-1 shows how we arrive at answers to each research question.

Table 1-1: research method for each research question

Research questions	Research methods
1. What are the requirements for EWPC system design considering current practice of FSCN management?	• Interviews with managers in food companies • Literature reviews on DSS, DM, SCM • Experience from practical cases
2. What components should be included in the EWPC systems, and how should those components cooperate to enable managers to achieve EWPC in FSCN?	• Literature review on model-based DSS and KBS • Design and test the components in real cases
3. What Data Mining methods are available and applicable for EWPC in FSCN?	• Literature review on Data Mining methods and applications • Select and apply Data Mining methods in practical cases
4. What support needs to be provided to managers in order to enable them to use Data Mining methods for	• Literature review on Data Mining, Knowledge Discovery from Database (KDD) process

EWPC?	• Accumulate experiences on using Data Mining methods to analyze real cases
5. What kind of structure is suitable for the Knowledge Base in EWPC systems?	• Literature reviews on Knowledge Management and Ontology Engineering • Collect knowledge (regarding problems in FSCN) obtained from case analysis • Design and test a Knowledge Base
6. What is the validity of the designed framework and prototype system?	• Expert Validation

1.6 Outline of this book

The structure of this book is illustrated by Figure 1-6. It consists of three main parts: theory, design, and validation.

Theory

The first part, Chapter 1 and Chapter 2, is theory. This chapter introduced the background of this research and the research settings. The research objectives and research questions were specified, and the research methods were presented. Chapter 2 defines what an EWPC system is and proposes a conceptual architecture for such systems. It also presents the general design process for EWPC systems. A real case is used to illustrate the feasibility of applying DM methods to problems in FSCN.

Design

The second part is about the design of the prototype EWPC system. It contains four chapters, from Chapter 3 to Chapter 6. Chapter 3 presents the requirements of EWPC systems. It also describes the framework for EWPC systems and how the framework is designed. The framework design is a very important step in system design. The framework gives a detailed description of the components in EWPC system and how those components work together to support managers in decision making.

The applicability of Data Mining is described in Chapter 4. This chapter presents the requirements on DM methods and evaluates the suitability of various DM methods for those requirements. The requirements on DM methods are derived from the requirements on EWPC systems presented in Chapter 3. Chapter 3 and Chapter 4 lay a foundation for components design which is described in Chapter 5 and Chapter 6.

Chapter 5 explains the designing of Template approaches and an Expert System for Data Mining method selection. Template approaches guide managers through all the steps in the usage process (as introduced in Chapter 3) where DM methods need to be applied. They are based on the generic process of applying DM methods to EWPC in FSCN. The generic process is obtained from an analysis of real cases. An *Expert System* for DM methods selection finds suitable DM methods for managers' problem situations. It is based on the applicability of DM methods which is presented in Chapter 4.

Chapter 6 explains how we design a structure for Knowledge Base in EWPC system. Knowledge Base enables managers to accumulate and share their obtained knowledge in data analysis. The structure is designed according to the types of knowledge obtained from case analysis. Template approaches, Expert System, and Knowledge base together compose a prototype system for Early Warning and Proactive Control.

<u>Validation</u>

The last part includes a comprehensive evaluation of the prototype system design, as well as a conclusion on this research. Chapter 7 presents the results of expert validation. A group of experts from different domains was invited to evaluate the prototype system design. The evaluation covers various aspects of system design, not only the system itself, but also its context, methodology, major components, etc. Chapter 8 discusses the theoretical and methodological contributions of this research. The limitations and directions for further research are also highlighted.

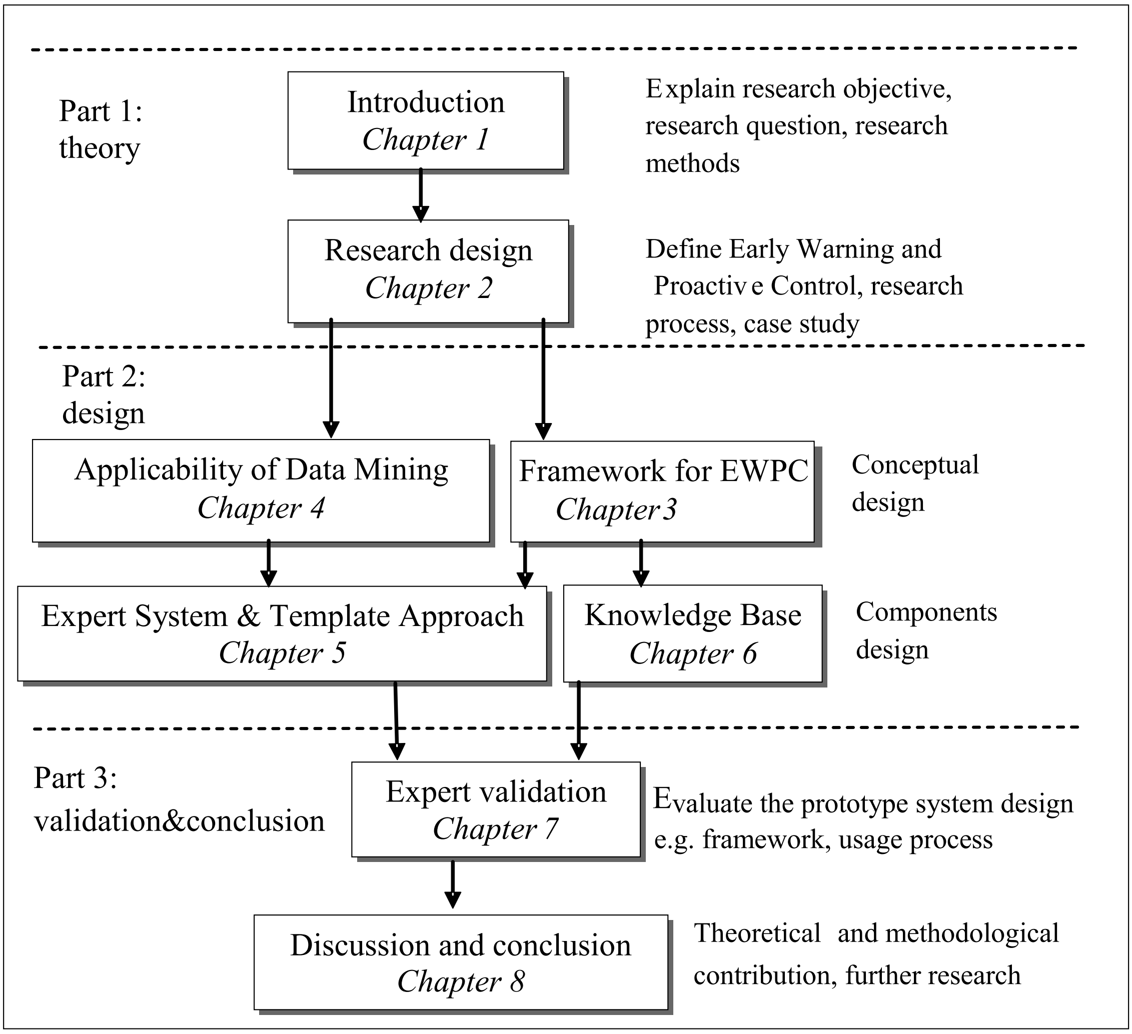


Figure 1-6: Structure of this book

1.7 Overlap and variations in different chapters

Since most of the chapters (Chapter 2 to Chapter 6) in this book are either published or in the process of being published as separate papers, there is some overlap, especially with respect to the explanation of the concept of Early Warning and Proactive Control. In each chapter, we make some effort to explain what an EWPC system is. Another overlap concerns the framework. Some chapters only deal with part of the system design, like the KB design in Chapter 6. Therefore, we have to introduce the framework (which is first mentioned in Chapter 3) in order to explain the position of KB in the EWPC system. The same overlap exists for the usage process. Chapter 5 explains how we provide support on Data Mining. We also repeat the usage process there (first introduced in Chapter 3) in order to help users understand the context of Data Mining.

The reader might also notice variations regarding the design process of EWPC system in Chapter 2 and Chapter 3. As the research goes deeper, more knowledge is obtained. The design process gets refined along the way. The general design process in Chapter 2 explains at a high level how we execute case studies, as well as how case studies contribute to the system design. This general design process has been further specified in Chapter 3. The design procedure for the framework is a further specification of the general design procedure. It explains how we arrive at the framework based on case studies and literature review.

Chapter 2. Using Data Mining to improve operations management in food supply networks[1]

Abstract

Operations management in food supply networks is especially complex due to the variability of primary inputs and the perishability of products. In order to prevent problems in food quality and operations, early warning and proactive control systems are required. This chapter describes a method to identify causal factors for these problems employing Data Mining techniques. A product of the research will be a reusable knowledge base of important types of variables and relations. In the case study, we investigated the applicability of Data Mining by applying three Data Mining techniques in the chicken chain of a large Dutch food company. Relations discovered in this case were confirmed by experts and practice, which demonstrates that Data Mining could be used to help to improve the efficiency and effectiveness of operations management in food supply networks.

Keywords: Data Mining, knowledge base, early warning, proactive control

2.1 Introduction

European consumers are sophisticated and highly conscious of food quality and safety (Beulens 2005). Recall announcements can be found in newspapers almost weekly due to problems in operations. Operations management in food supply networks is especially complex because of the variability of primary inputs and the perishability of products. To ensure consistent quality of food products and to prevent food safety problems, it is beneficial to set up early warning and proactive control systems in food supply networks.

BOX I

A case for early warning and proactive control in operations management

December 15, 2005, a Dutch dairy firm had to recall packets of chocolate milk after children turned sick from drinking the milk. The problem was caused by traces of cleaning substance left after cleaning the pipes in the factory. With an early warning system the problem could have been detected before the packets were actually sold and consumed. A proactive control system could even have prevented using the pipes that were not properly cleaned.

In current food supply networks, large amounts of data about business operations and transactions are recorded into ICT systems every day (Petersen, Knura-Deszczka et al. 2002). Apart from their day-to-day operational use, the information implicitly present in these data is valuable as a basis for

[1] Li, Y., M. R. Kramer, A. J. M. Beulens and J. G. A. J. v. d. Vorst (2006). Using data mining to improve operations management in food supply networks. International agri-food chains and networks: management and organization. J. Bijman, S. W. F. Omta, J. H. Trienekens, J. H. M. Wijnands and E. F. M. Wubben, Wageningen Academic Publishers: 163-177.

implementing early warning and proactive control. Many unknown causal relations exist between the recorded data about inputs, controls and production means, and operational performance of processes inside food supply networks. Knowledge about these relations provides the possibility to improve operations management by monitoring and proactively influencing the corresponding determinant factors. So both from a scientific view and from a practical perspective, it is an interesting topic how to get value from recorded data.

In our research project, we intend to design a generic knowledge based framework to set up early warning and proactive control systems in food supply networks. This chapter presents the general research plan and describes a first case study. Ultimately, we will use Data Mining for discovering causal relations from recorded data, as well as for monitoring determinant factors during actual operations. Here, we present the feasibility of this approach in a practical case study. Therefore we investigated the applicability of three Data Mining techniques: decision tree, neural networks, and nearest neighbor methods. The data we used were retrieved from the food supply networks of Nutreco, a large Dutch food company.

In the next section, we introduce the key concepts of our problem domain and some terminology of the techniques employed. Then, section 2.3 presents the design of the research project of which this study is an element. Section 2.4 shows the results of a first case study in employing Data Mining to find causal relations in actual data. In the last section we discuss the meaning of this case study and give directions for further research.

2.2 Review of relevant concepts

In this section we introduce the concepts of food supply network, control theory, early warning and proactive control systems, and Data Mining.

2.2.1 Control Systems and Performance

Operations management in food supply networks is about the way organizations produce food products and services. Slack et al. (Slack, Chambers et al. 1998) distinguish five categories of objectives for operations management: quality, speed, dependability, flexibility, and cost. In order to facilitate measuring performance of food supply chains, many authors provided frameworks for constructing *performance measurement systems* for food supply chains (Ni and Gunasekaran 1998; Beamon 1999; Lohman, Fortuin et al. 2004). A performance measurement system associates measurable *performance indicators* with important processes and intermediate products in the supply network. For control purposes, the variation of values of certain performance indicators should be restricted. The lower and upper acceptable values for a performance indicator are called its *control limits*.

We define *deviations* as any variations that exceed the control limits of performance indicators; a deviation exceeds the normal variation that is inherent to the process. In food supply networks, various kinds of deviations may occur, either from the specifications of product attributes, or from the specifications of logistical processes (e.g. time restrictions, excessive use of raw materials or resources). Deviations may cause defective products which may pose a threat to the environment, to the public involved, or to the chain participants.

There are two key mechanisms in control theory to keep variations between limits: feedback and feedforward control (Goodwin, Graebe et al. 2001). With *feedback control*, a system maintains its homeostatic action through adjusting the system according to monitored output properties, irrespective of the origins of the input variations that disturb the system. With *feedforward control*, the process disturbance is forecasted before it actually occurs. The forward cycle anticipates an expected situation based on earlier measurements in the process. A pure feedforward system makes no attempt to monitor and feed back the actual values of controlled variables. Rather it monitors the state of variables that are known to affect performance, and feed forward corresponding control actions to counter the disturbing actions encountered. We define *determinant factors* as those variables that play an important role in causing any deviations in controlled variables. Thus, a feedforward control system monitors determinant factors to forecast and prevent deviations.

In food supply networks, causes of deviations may reside in many aspects, for example, mistakes by operators, or biological variation of primary products. If feedback control is applied for deviations in food supply networks, production can only be influenced after losses have already been incurred. To prevent deviations from happening, it is necessary to realize feedforward control for deviations in food supply networks.

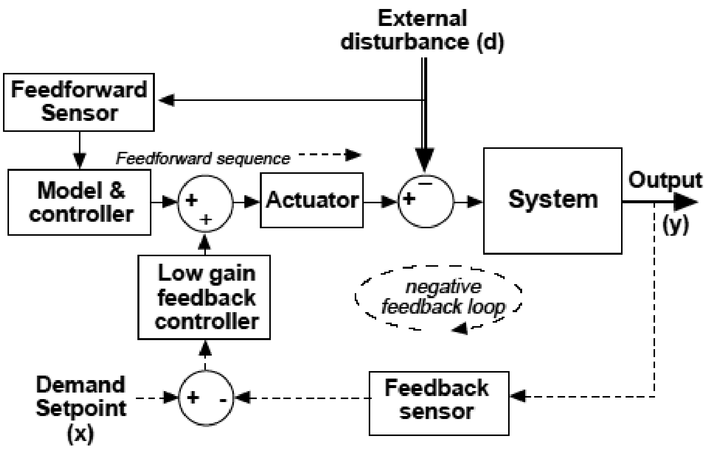


Figure 2-1. Feedback plus feedforward system(Fowler 1999)

It follows that, in order to function effectively, feedforward systems require some form of system model, a set of prediction algorithms, and a comprehensive monitoring facility to detect changes in system inputs (Fowler 1999). Counteractions can then be computed and fed forward to neutralize the effects of external disturbances and keep the system output within bounds.

2.2.2 Food Supply Networks

According to Santoso et al.(Santoso, Ahmed et al. 2005), a supply chain is a network of suppliers, manufacturing plants, warehouses, and distribution channels organized to acquire raw materials, convert these raw materials to finished products, and distribute these products to customers. Van der Vorst et al. (2005) extend this to a food supply network, referring to an interconnected system with a large variety of complex relationships such as alliances, horizontal and vertical cooperation, forward and backward integration in supply chains.

Figure 2-2 depicts a supply chain network with accompanying monitoring and control systems. Processes can be seen at different levels, depending on the kind of problems considered. They may be as large as complete farms, factories or warehouses, or as small as one individual activity. The monitoring system gathers performance data from processes and their inputs and outputs. The control system can influence settings of the processes involved. Together, the monitoring system and control system manage the whole series of processes.

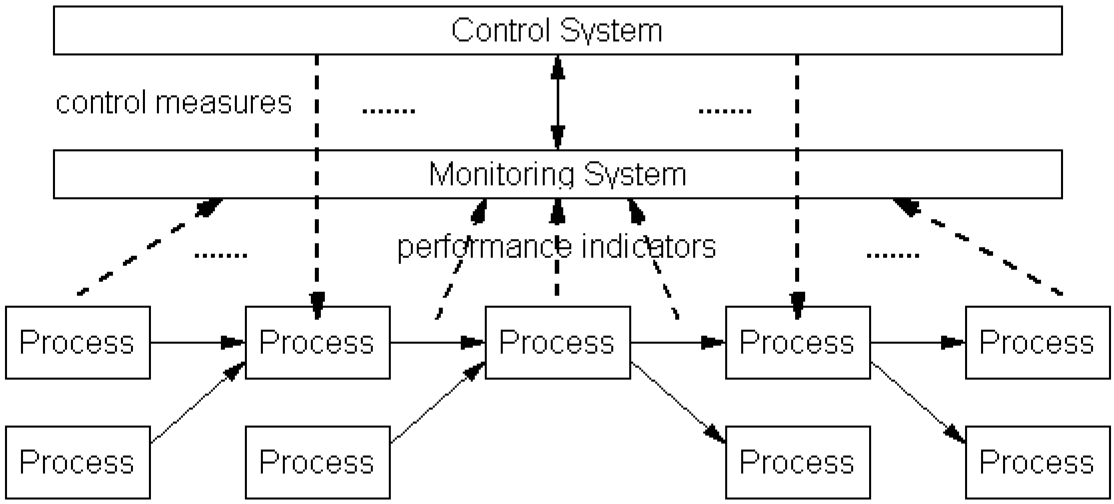


Figure 2-2. An example supply chain network

There is much theory on supply chains and networks in general. However, food supply networks are specific in a number of aspects. For example:

- food products are subject to perishability, so special requirements are imposed on transport and storage;
- quantity and quality of process yield vary due to biological variations, seasonality, random factors connected with weather, pests, and other biological hazards;
- strict requirements on homogenous output of processes are harder to maintain due to larger variability of inputs.

Moreover, recent developments in food supply networks make this kind of management even more complicated. For example, product assortments have exploded, product life cycles have decreased and consumer demands on product freshness and food safety have become more unpredictable (Van der Vorst, Beulens et al. 2005). Against this background, it becomes a challenging problem how to improve operations management in order to effectively control operational performance, specifically food quality.

2.2.3 Early Warning and Proactive Control

The principle of feedforward control forms the foundation of early warning systems. Early warning systems are well known in natural sciences. These systems, based on historical monitoring, local observation or computer modelling, predict and help to prevent or reduce the impact of natural disasters. They are typically used to monitor potential disasters relating to meteorology (e.g. floods, fire and droughts) (Grijsen, Snoeker et al. 1992; Kumar 1998), geology (e.g. earthquakes and volcanoes) or technology (e.g. nuclear safety). Early warning is being extended to other application areas as well. For example, Costello and Ewen (Costello, Ewen et al. 2003) presented a prototype sensor system for the early detection of microbially linked spoilage in stored wheat grain. Further, the importance of early warning systems is recognized by researchers and professionals in other disciplines to include social, economic and cultural factors. However, to our knowledge there is currently no existing early warning system that is particularly designed to improve operations management in food supply networks.

Our definition for *early warning systems in food supply networks* is given as follows: Based on the knowledge of relations between deviations in performance of food supply networks and determinant factors, an early warning system in food supply networks predicts as early as possible potential deviations for decision makers in food supply networks by monitoring measurable determinant factors.

We would like to define the point where determinant factors are monitored. In analogy to Hazard Analysis and Critical Control Points (HACCP; (NACMCF 1997)), we introduce the idea of a *Deviation Analysis Critical Warning Point* (DACWP): the point where an early warning signal about a potential deviation is generated and sent to decision maker before the deviation occurs. It can be a location, practice or procedure where the potential cause for deviation is monitored. As with HACCP, there are two important requirements for DACWP:

- it is possible to measure at this point
- the time consumed by measurement should be short enough to take appropriate action

Figure 2-3 gives a functional architecture for early warning systems in FSCN:

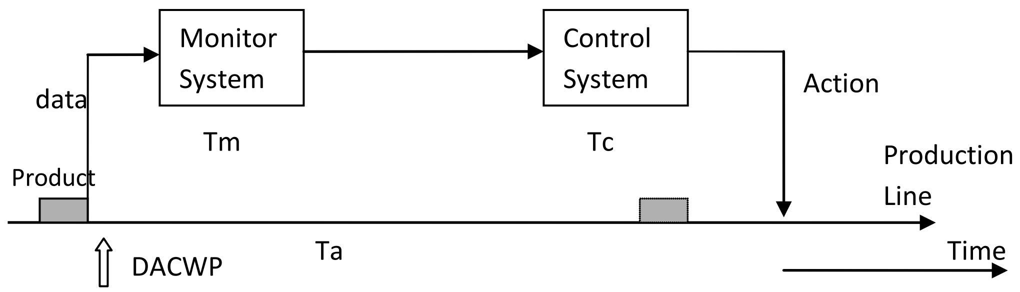


Figure 2-3. Architecture for early warning system

At the critical warning point (DACWP), the monitor system gathers the data about a specific determinant factor, and feeds it to the control system. The control system determines whether to take appropriate action or not. Each of these two steps will take some time, so to realize early warning the total time between DACWP and the warning should be shorter than the time for the product to travel from DACWP to the point where action should take place:

Mathematically, we can define this as Tm + Tc < Ta, where:

- Tm is the time period used by the monitor system
- Tc is the time period used by the control system
- Ta is the time period for the product flow from DACWP to the point where action should take place

This is only a functional architecture to illustrate our idea of early warning system in food supply networks. In practice deviations may have multiple determinant factors, so that we have multiple DACWP to monitor.

After a deviation in performance is early warned, another question appears: how to efficiently and effectively prevent or diminish this deviation? Fortunately with the implementation of early warning system, we already solved the first crucial obstacle. The set of causal relations discovered for setting up early warning systems is a valuable asset. It provides decision makers with information on control points in food supply networks for specific (early warned) deviations. With the knowledge of decision varieties in those control points and their consequences, decision makers could devise a series of actions to proactively control deviations in food supply networks. We define a *proactive control system in food supply networks* as follows: Based on the knowledge of relations between deviations in food supply networks and determinant factors, proactive control systems in food supply networks propose appropriate actions in order to prevent potential deviations which have been flagged by the early warning system. The potential actions could be discarding, taking corrective measures, or adapting subsequent processes in order to make amends.

From these definitions we can see that an indispensable component to both early warning and proactive control systems is the set of causal relations between deviations in food supply networks and determinant factors. Ultimately, expert domain knowledge is required to verify those causal relations. However, current advances of ICT systems provide the opportunities to largely automatically disentangle causal relations from huge amounts of data.

2.2.4 Data Mining

The application of early warning and proactive control requires predictive models of the object system (i.e. the controlled food supply network). However, to construct such a model, normally would require detailed insight into the processes involved. An alternative approach is to infer a model from available data with the help of Data Mining. Data Mining (DM) is the process of extracting valid, previously unknown, comprehensible and actionable information from large databases and using it to make crucial business decisions (Simoudis 1996). Data Mining has two high-level objectives: prediction of unknown or future values of selected variables, and knowledge discovery (Weiss and Indurkhya 1998). The goals of prediction and discovery are achieved by using the following primary Data Mining tasks (Chen, Han et al. 1996; Fayyad, Piatesky-Shapiro et al. 1996): classification, regression, clustering, summarization, change and deviation detection, and dependency modelling. These tasks coincide with requirements for early warning and proactive control. In practical operation, early warning and proactive control systems in food supply networks have to find infrequent problems in large amounts of operational data (prediction). In our approach, the underlying model of the early warning and proactive control systems is based on causal relations discovered in historical data (knowledge discovery). Table 2-1 presents a preliminary comparison of some widely used Data Mining techniques.

Table 2-1. Overview of some Data Mining techniques

Name	Features
Nearest Neighbors (Goodwin, Graebe et al. 2001)	Mainly used for prediction, to classify unknown instances into subgroups that have a number of properties in common. Speed is slow relative to other algorithms.
Decision Trees (Breiman, Friedman et al. 1984)	Mainly used for prediction, to classify unknown instances into subgroups which have a number of properties in common. Can be very sensitive to minute fluctuations in a data set.
Association rule (Agrawal, Mannila et al. 1996)	Mainly used for knowledge discovery, to find rules such as 'A implies B' in a data set. Results do not necessarily cover the whole data set; they only represent some aspects of dataset.
Neural Networks (Beale and Jackson 1990)	Mainly used for prediction, to learn (input-output) patterns within data sets, those patterns have a number of properties in common. How to determine the structure of the network depends on the miner's experience.

There are two approaches for Data Mining: verification driven, the aim of which is to verify a hypothesis postulated by a user, or discovery driven, which is the automatic discovery of information by the use of appropriate tools. These two aspects of Data Mining fit well with the observation-hypothesis-experiment cycle of this research. The discovery driven Data Mining works well for the observation-hypothesis phase, and the verification driven Data Mining works well for the hypothesis-experiment phase (Crawford and Crawford 1996).

2.3 Research design

This research follows the inductive/deductive research cycle in which literatures (about food supply networks, process control, system design, and Data Mining) and multiple case studies are used. In the induction phase, we induce from several cases a systematic framework for applying Data Mining to set up early warning and proactive control systems in food supply networks. In the deduction phase, we apply this framework to other cases to verify its validity. If the result is not satisfactory, we iterate these two phases again with the problem encountered. By this inductive/deductive cycle we gradually develop a knowledge based framework for setting up early warning and proactive control systems in food supply networks.

2.3.1 Knowledge base

We could benefit from case studies in several aspects. The major one is the knowledge base. In order to implement early warning and proactive control in food supply network, it is necessary to obtain the following information for every process we intend to control:

- Performance indicators, to describe some aspects of the process's performance and to facilitate monitoring
- Control limits for all performance indicators to be able to detect deviations
- Data availability (i.e. whether there is enough data to characterize operations and performance) and the time required to gather data
- Decision variety (i.e. what are the available decisions) and influence of decisions on subsequent processes

These kinds of information should be kept in the knowledge base. The knowledge base will be helpful not only for managers to make decisions, but also for researchers to discover new relations in subsequent case studies. By means of the knowledge base we could realize extendable early warning and proactive control systems. Our knowledge base will contain

- categories of determinant factors,
- categories of deviations,
- categories of relations between determinant factors and performance,
- and specific cases of causal relations in food supply networks.

Relations retrieved by Data Mining along the way will be categorized and recorded into the knowledge base, together with related determinant factors. Other benefits include the experience in building up these systems, and the merit of various Data Mining techniques in discovering causal relations. Both knowledge and experiences can be utilized by researchers and practitioners who intend to efficiently and effectively manage food supply networks.

2.3.2 Design process

Our process for designing early warning and proactive control systems is represented in Figure 2-4. The process is repeated for each case study; one cycle of the design process consists of these five steps:

1. Problem description: With the help of domain experts, descriptions of the problems (or opportunities) and explanations of the variables in the operational database are obtained. The operational database is designed to record data associated with operations and product quality performance. With this information, we specify the requirements of stakeholders involved in food supply networks.
2. Analysis and matching: Based on descriptions of problems and the database, we analyze data and, if possible, match them to existing categories in the knowledge base.
3. Apply DM methods to find relations: From the knowledge base, we find corresponding Data Mining techniques that are suitable for these types of data and specific problems (or opportunities). Then we apply these techniques and analyze their results to find relations in these new data. If we find new relations, they will be added to the knowledge base, so the knowledge base grows in the process.
4. Design control systems: Based on knowledge of those relations, we investigate the possibility of monitoring corresponding determinant factors and predicting potential deviations in the actual food supply networks, and proactively control deviations with appropriate decision varieties.

5. Evaluate early warning and proactive control systems (EW&PC systems): At the end of a cycle, we evaluate the performance of the early warning and proactive control systems. By interviewing domain experts we investigate whether they regard the effects of these systems as valid in the application domain.

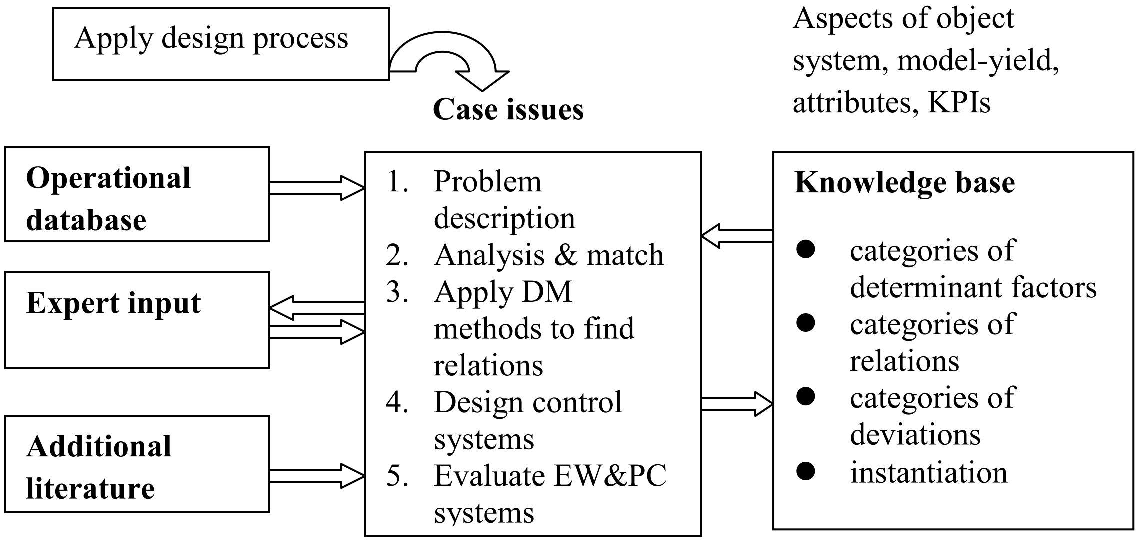


Figure 2-4. Design process of Early Warning & Proactive Control system for one case study

There is a reciprocal relation between domain experts and case studies: first domain experts provide background knowledge on specific problems, corresponding operational databases, and hypotheses of possible relations, etc.; then after the case study, they could benefit from relations found in the design process.

2.4 A Case Study

Here we present an overview of the results of the first case study we have conducted. This case comes from one of Nutreco's food supply networks; it deals with breeding chickens for meat. This supply network was suffering from a high rate of chicken's Death-On-Arrival (DOA). Currently, no early warning or proactive control system is available in this company. However, the information system of Nutreco contains detailed reliable data associated with the processes in the network. These data are gathered from various sources and cover several stages in Nutreco's chicken supply network (see Figure 2-5). In this case study, we applied Data Mining techniques to those data. This case study focused on two aspects:

1. identify causal relations for DOA by analyzing data associated with relevant processes; these causal relations are indispensable for setting up early warning and proactive control system
2. compare the merit and shortcomings of several Data Mining techniques when applied for discovering causal relations from data

Further case studies on other data will be conducted as part of ongoing research.

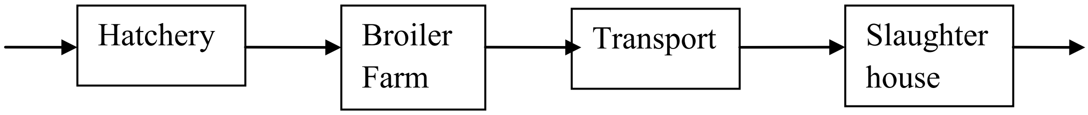


Figure 2-5. Stages in Nutreco's chicken supply networks, as far as covered in this case study

2.4.1 Methods

The Data Mining techniques we employed include decision tree, neural networks, and nearest neighbor methods.

The *Decision tree* method is widely used for classification and regression. A decision tree is a flow-chart-like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and leaf nodes represent classes or class distributions. In order to classify an unknown sample, the attribute values of the sample are tested according to the decision tree starting from the root until one of the leaves. To build decision trees, a Data Mining algorithm recursively inspects the available data set to find decisions that optimally split the data into distinguished subsets. An important property of this technique is that its functioning is easily understood.

Neural networks are used as prediction tools. A neural network consists of simulated neurons and interconnections between them. Each connection has an associated strength, called its connection weight. The neural network has to learn the connection weights from a set of inputs with corresponding optimal outputs. It learns a pattern between input and output by modifying the connection weights between neurons. After learning, a network is used to predict the output for new input of a process.

Nearest-neighbor methods classify a new case based on its 'distance' with stored earlier samples. They assign a new case the same class label as its nearest neighbor(s). In order to get accurate classification, researchers tried several ways to modify the weights of each factors (Wettschereck, Aha et al. 1997; Goodwin, Graebe et al. 2001). In this case, we look at each factor's weight as assigned by this technique, in order to find out which factors are important.

2.4.2 Results

In this section, we focus on methodological aspects of applying DM techniques to find determinant factors and relations that can be included in a knowledge base for the early warning and proactive control systems. We will not go into details regarding the specific determinant factors and relations that we identified for the DOA performance.

<u>Decision tree:</u>

Figure 2-6 shows the result obtained through the decision tree technique. Each node in the tree compares the number of flocks with DOA (in black) to those without DOA (in gray). Take the node in the black circle for example. It indicates that if condemnation (PRCAFK) is less than 0.010587, then almost no flocks have DOA (whereas with PRCAFK above 0.010587 significantly more flocks have DOA). The absolute number of instances in each leaf is not shown here, to reduce space. In practical use, we have to take the number of instances into consideration as well.

This graph shows a number of decision rules to determine whether a flock has a higher chance of DOA (Death on arrival). For example, as the arrows indicate, if condemnation (PRCAFK) is larger than 0.01, breed (KKRAS) is Cobb, transportation time (TITRN) is larger than 1.84, the amount of DOA is relatively large. Some care is needed when interpreting the results. For example, condemnation (PRCAFK) has a very high correlation with DOA. However, it is not a determinant factor because both are dependent on other factors.

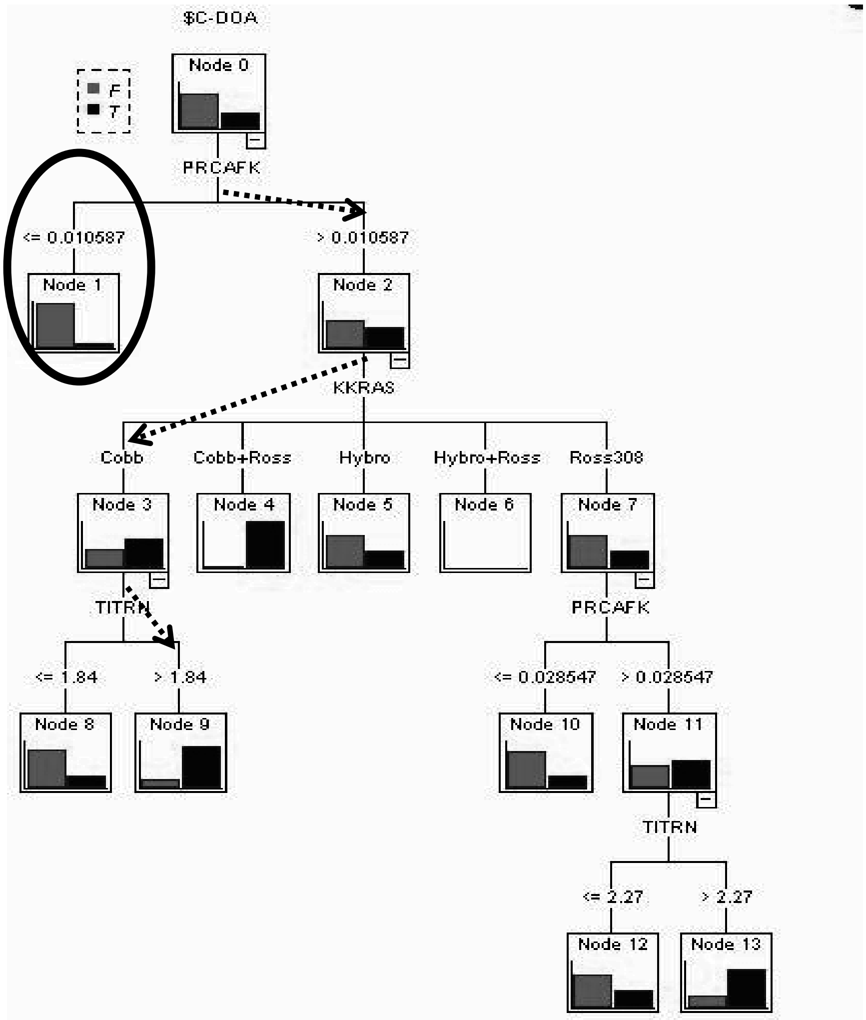


Figure 2-6. Result obtained through decision tree

This example of a decision tree shows that the decision tree method could discover relations between determinant factors. These relations provide suggestions for operations management. For example, from this decision tree we can find that when the breed ‘Cobb’ is subjected to long transportation times, DOA will increase. So managers are suggested to limiting the transport time for ‘Cobb’. Sometimes there is only a small set of instances within one leaf, so the decision tree method does not only find large groups of objects having some properties in common, but also tiny groups, that might correspond to exceptional patterns. This feature is especially helpful when dealing with a quality problem that is insignificant in quantity but potentially has serious impact.

Neural networks:

Although neural networks are mainly used as prediction tools, data obtained during training the network may provide additional insights. The effect that each of the network inputs has on the network output provides feedback as to which input channels (i.e. determinant factors) are most significant. With sensitivity analysis, we retrieve valuable information from the learned network structure. Sensitivity analysis is a method for extracting cause and effect relations between the inputs and outputs of the network. There are several methods to do sensitivity analysis with neural networks (Yao 2003), of which we have used the variable perturbation method. The first run of sensitivity analysis shows that condemnation has a strong correlation with DOA. As discussed above, this result is obvious but useless for control purpose. So we deleted this factor and analyzed again. The graph in Figure 2-7 shows the results of this analysis.

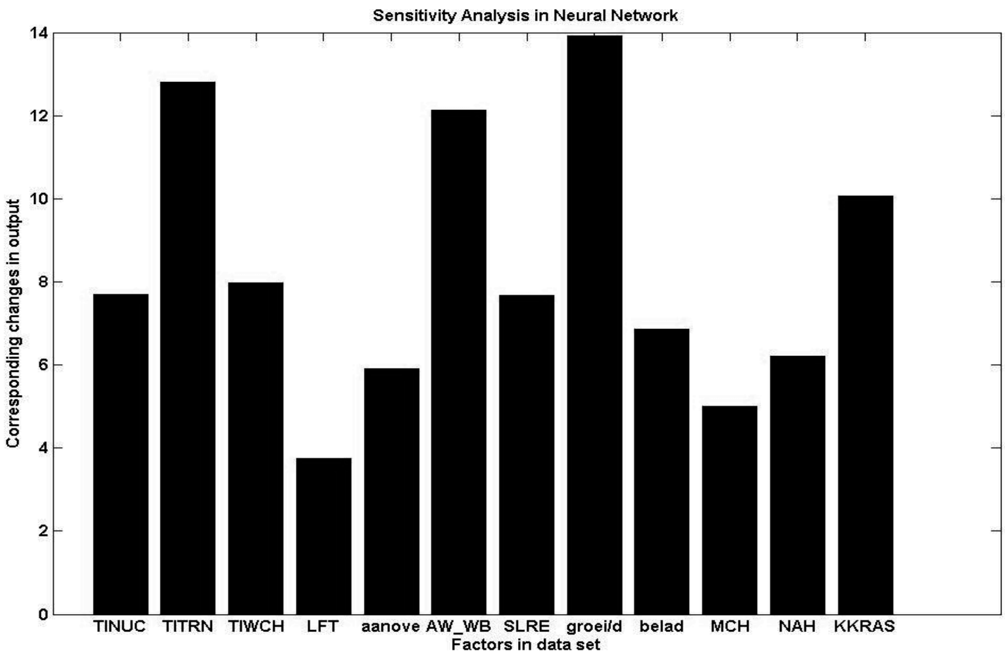


Figure 2-7. Result from sensitivity analysis on neural networks

The magnitude of each bar represents the corresponding change in output (DOA) after applying some change in that specific input factor. This graph shows that chicken's growth (groei/d), transportation time (TITRN), and breed (KKRAS) play important roles in determining DOA. This result illustrates the ability of this technique to give an overview of the influence of each factor in the data set. This information could direct further analysis about how these factors affect performance indicators. However, contrary to the decision tree method, this method can not discover the relation between influencing factors.

Nearest neighbors:

Normally nearest neighbors methods are used for prediction instead of exploring. However, we investigated the weights of factors involved in order to find the importance of each factor in predicting DOA. The factor weighting algorithms we used here is IB4 (Goodwin, Graebe et al. 2001). The results obtained show that this method has similar representation power as neural networks:

1. This method gives an overview of the importance of each factor.

2. This method also can not discover the relation between influencing factors.

Figure 2-8 shows the result obtained with this technique. Most factors' relative importance is similar with those obtained through the neural network. However, some factors' relative importance allocated by this technique is different from those assigned by the neural network. For example, chicken's breed (KKRAS) plays an important role in the result obtained through the neural network, but it is relatively unimportant in the result of the nearest neighbors method. So it is necessary to compare the results of different DM techniques. In this case, the importance of KKRAS has already been demonstrated by two techniques.

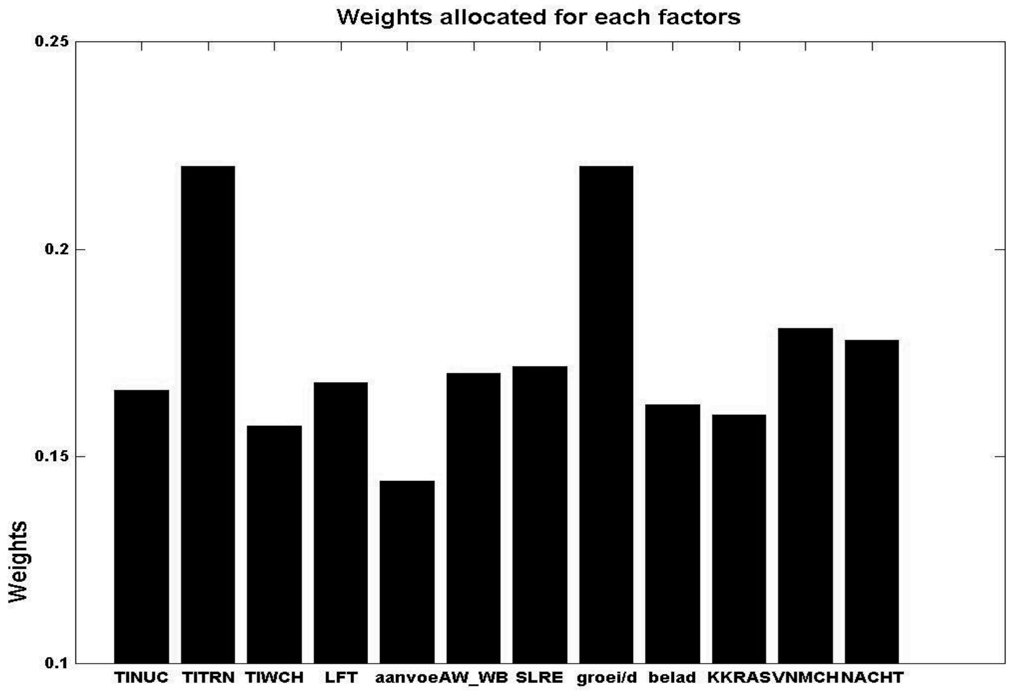


Figure 2-8. Result from Nearest Neighbors

By applying Data Mining techniques, we found a number of important relations in the data set. Most of our findings were confirmed by experts in the company, either because the relations were apparent from domain knowledge, or because the experts recognized the relations after reviewing them. The former confirms the capability of Data Mining techniques to discover domain knowledge. The latter means that the Data Mining techniques employed are capable of finding new relations that enhance or confirm expert knowledge.

Some relations found turned out to be causal relations actually. Relations found in the earliest part of the data set predicted the effects of measures effectuated in the later parts of the data set. Those relations pointed out possible measures to be taken to decrease losses and improve performance in the operations management of food supply networks.

2.5 Conclusions

This chapter presents the first results in an ongoing research project aimed at constructing early warning and proactive control systems in food supply networks. Such systems can help domain experts improve the efficiency and effectiveness of operations management in food supply networks. In this ongoing research project, we will conduct multiple case studies to gradually build up a knowledge base that can be used for constructing these early warning and proactive control systems. The knowledge base will contain generic information about operations management in food supply networks, especially about types of determinant factors and their influence on performance indicators.

The first case study has illustrated that Data Mining techniques can help to efficiently find relations between performance indicators and variables related with operational processes in food supply networks. Still, there is a need to improve current Data Mining techniques to predict which relations are actually causal relations, i.e. that some attributes of products may be influenced by changing attributes of designated processes.

Although we gained some promising experience by the case study, we have not yet acquired systematic insight into the typology of relations that may exist in food supply networks and related Data Mining techniques which are suitable to discover them. This knowledge is also crucial for implementation of early warning and proactive control systems in food supply networks. We are conducting further research in this aspect.

Chapter 3. A Framework for Early Warning and Proactive Control Systems in Food Supply Chain Networks[2]

Abstract

It is inherent to food supply chain networks that performance deviations occur occasionally due to variations in product quality and quantity. To reduce losses, one wants to be informed about such deviations as soon as possible, preferably even before they occur. Then it is possible to take actions to prevent or reduce negative consequences.

In practice, extensive operational data is recorded, forming a valuable source for early warning and proactive control systems, i.e. decision support systems for prediction and prevention of such performance problems. Data Mining methods are ideal for analyzing such data sources and extracting useable information from them. In this chapter, we present a novel framework for early warning and proactive control systems that combine expert knowledge and Data Mining methods to exploit recorded data. We discuss the implementation of a prototype system and the experiences from a case study regarding the applicability of the framework.

Keywords: Early warning; Proactive control; Data Mining; Knowledge Management

3.1 Introduction

A Dutch food company encountered a problem in their chicken supply chain; too many chickens died during transport to the slaughter house. This problem is known as Death-On-Arrival (DOA). The company continuously incurred extra costs because those dead chickens could not be used in the slaughter process and had to be disposed. Managers in this supply chain were not sure about the cause of this problem. However, the monitoring system in this supply chain recorded data associated with various factors (operational, environmental, etc.) and properties of chickens at various stages of the supply chain. So why not take advantage of those recorded data to try to predict and prevent DOA?

The DOA problem is an example of a type of problems encountered in Food Supply Chain Networks (FSCN). In our research, we aim at building Early Warning and Proactive Control (EW&PC) systems to tackle such problems. Such systems are knowledge-based, data- and model-driven decision support systems (DSS) that are designed to assist managers in prediction and mitigation of problems associated with food products in FSCN. With these systems, managers can find causes for problems related to performance indicators of food products. The system analyses

[2] Li, Y., M. R. Kramer, A. J. M. Beulens and J. G. A. J. v. d. Vorst. A Framework for Early Warning and Proactive Control Systems in Food Supply Chain Networks, Computers In Industry, forthcoming

existing data, information, and knowledge available in FSCN and uses Data Mining (DM) to generate decision support models for prediction of potential problems. Managers can in turn use these models in combination with their expertise in FSCN for decision making.

Due to characteristics of FSCN, EW&PC systems need a different architecture from traditional DSSs. In FSCN we have to deal with various types of problems, such as variability of quality and quantity of supply, shelf life constraints for raw materials, intermediates and finished product, variable process yield in quantity and quality due to biological variations, seasonality, random factors connected with weather, pests, other biological hazards (Van der Vorst, Beulens et al. 2005). In many cases, encountered performance problems are new to managers. The causes of such problems then need to be explored followed by the generation and evaluation of decision alternatives. Traditional architectures of DSSs (e.g. (Recio, Rubio et al. 2003; Lin and Hsieh 2004)), which are based on fixed models, are less suitable for EW&PC in FSCN. If encountered problems have not been included in those fixed models, managers can get little help.

In this chapter, we present a novel framework for EW&PC systems in FSCN. Such systems act as tools to solve a wide range of problems in FSCN based on available data. To our knowledge, such a framework has not been reported in literature before. We obtained our framework through several research steps (see Figure 3-1). First, we formalized our ideas about EW&PC systems, and searched for relevant knowledge from literature on FSCN, DM, and DSS. Then we identified generic steps for realizing EW&PC with DM. In parallel, we conducted several case studies in the food industry, of which we use the DOA case to illustrate our findings. From case studies and DSS literature we derived various requirements on these systems. After investigating these requirements, we designed a framework to fulfill those requirements. This framework shows the constituents of the system and their mutual relations. Alongside, we designed the processes for managers to use such systems.

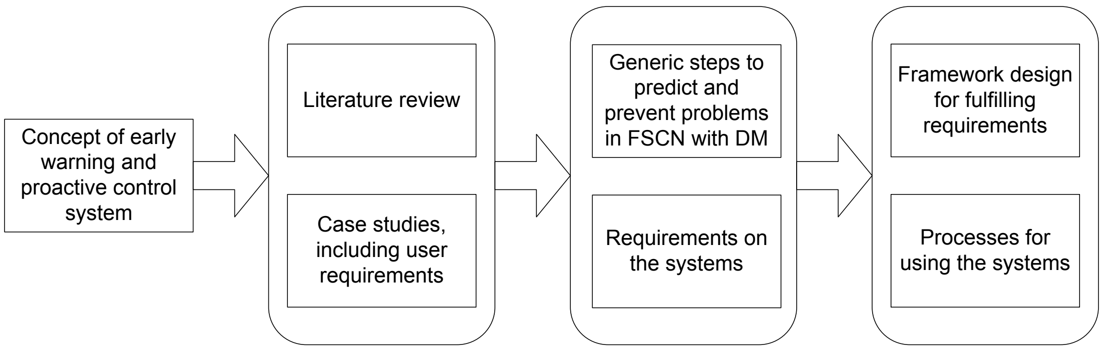


Figure 3-1: research steps to obtain the framework and processes

To illustrate the process of EW&PC we use the DOA case. This case originates from a Dutch food company that was suffering from a high rate of chicken's Death-On-Arrival (DOA). The manager had to investigate the stages of the supply chain and available measured factors to find potential causal factors for the DOA problem. In the DOA case, the chain has four stages, from hatchery to slaughter house (see Figure 3-2).

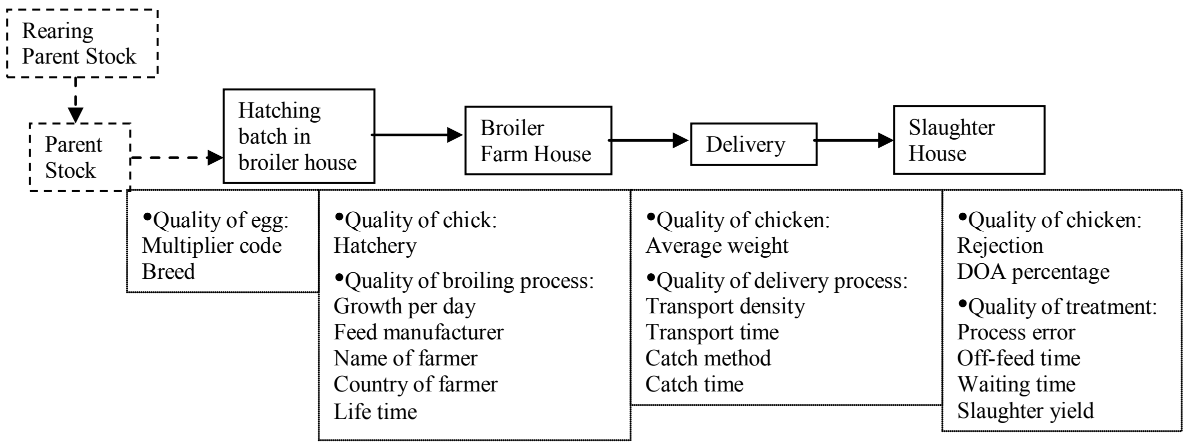


Figure 3-2: various stages and associated factors in chicken supply chain

The remainder of the chapter is organized as follows. Section 3.2 puts generic steps to predict and prevent problems by use of DM methods into context, and discusses different kinds of requirements on EW&PC system. Section 3.3 presents the framework. The components of the framework and the relations between components are explained in detail. Section 3.4 describes the processes for using this framework for EW&PC, with example application in the DOA case. Section 3.5 discusses the merits of this framework and points out issues for further research.

3.2 Early Warning and Proactive Control Systems

3.2.1 Definition, position

Early Warning and Proactive Control systems in Food Supply Chain Networks (from now on simply referred to as EW&PC systems) are meant to support decision makers in FSCN to prevent performance deviations. We define the function of EW&PC systems as: "EW&PC systems predict as early as possible potential performance deviations in FSCN by monitoring measurable determinant factors, and propose appropriate actions to prevent or reduce the forecasted deviations".

A deviation is the situation that the value of a performance indicator goes out of its normal accepted range. The EW&PC systems base their prediction on knowledge of relations between deviations in performance of FSCN and determinant factors for those deviations. Determinant factors can be operational factors (e.g. transport time) or inherent properties of products (e.g. genetics). The potential actions can be discarding products, taking corrective measures, or adapting succeeding production processes in order to make amends. Furthermore, managers can insert obtained knowledge about deviations and determinant factors into the system. Such knowledge is then beneficial for other users that have similar problems.

There are various kinds of methodologies available for building DSS (Klein and Methlie 1995; Turban 1995). However, based on the survey by Blair et al. (Blair, Debenham et al. 1997), those methodologies provide little help for the whole building process of practical DSS. As a result, we set up the design of EW&PC systems by using suitable components of different methods and tailor them for our situation. We incorporated results from our case studies in the design process as well, as we will discuss later.

Research in EW&PC systems is related to various other research areas. In order to supply managers with flexible support in building their own models for the problems they encounter in FSCN, we take advantage of research in Data Mining (Fayyad, Piatesky-Shapiro et al. 1996), Knowledge Management (Schreiber, Akkermans et al. 2000), and Ontology Engineering (Antoniou and Harmelen 2004). Combining these research areas enables managing both structured knowledge (e.g. data in information systems) and unstructured knowledge (e.g. the usage of a particular DM method).

Recent advances in operational information systems in FSCN yield new potential sources of structured knowledge. Those information systems keep records of information associated with food quality, operations, environment, etc. (Van der Vorst, Beulens et al. 2005). The development in Business Intelligence, which mainly concerns Data Warehousing and On Line Analytical Processing (OLAP) (Herschel and Jones 2005), results in facilities to store those records and query or manipulate (e.g. merge) it on specific dimensions. In addition, research in DM provides versatile methods that can be used to obtain knowledge from recorded data (Fayyad, Piatesky-Shapiro et al. 1996). With DM methods, managers can quickly identify causes of encountered problems by analyzing recorded data in FSCN. Some DM methods can also be used for prediction and decision evaluation. The idea of letting knowledgeable business users apply DM methods for decision making is already mentioned by other researchers (Ozden Gur Ali and Wallace 1997; Shaw, Subramaniam et al. 2001; Shim, Warkentin et al. 2002). We extend this idea to the design of EW&PC systems that complements domain knowledge of managers (as non-experts in DM) with the potential of DM methods.

Knowledge obtained by managers in analyzing their data sets is regarded a valuable reference for dealing with similar problems. Current research in Knowledge Management and Ontology Engineering provide us with methodologies and tools to store and manage such knowledge in an accessible and extendible way (Schreiber, Akkermans et al. 2000; Antoniou and Harmelen 2004). So we also take advantage of research in those research areas to incorporate identified knowledge on problems in FSCN into EW&PC systems. Then managers with similar problems in FSCN have a quick reference.

3.2.2 Generic steps to predict and prevent problems by means of DM

From the case studies we conducted, we discovered that there are several generic steps that need to be executed in employing DM methods for EW&PC (see Figure 3-3). These steps not only help to clarify requirements on EW&PC systems, but also contribute to building the framework.

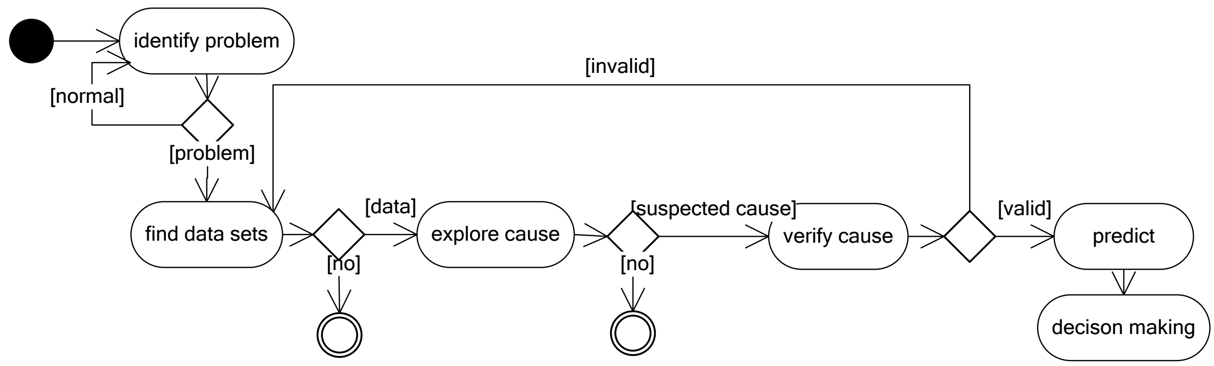


Figure 3-3: generic steps for early warning and proactive control

First, a manager identifies a problem in his FSCN and formulates it in a quantitative way. In the DOA case, for example, the death percentage of a flock of chickens is used to determine whether that flock has a DOA problem. Subsequently, the manager identifies relevant data sets for the problem among available data sets from the supply chain. Via analyzing the collected data sets, the manager discovers potential causes. For example, the transport time was identified as a suspected cause in the DOA case. Next the manager verifies the cause by modifying the corresponding real process (like transportation in the DOA case) in some instances and observing the results. If those modifications solve the problem, the manager can predict future problems of this type by timely monitoring the causal factor. If prediction warns about a potential deviation, the manager can choose appropriate decisions to proactively control problem instances (e.g. using optimal routes for transportation). The manager stops if he cannot find relevant data for his problem or data analysis does not show suspected causes.

The generic procedure in Figure 3-3 can be regarded as a special case of the generic model for decision making of Simon (1987). The difference mainly resides in the problem analysis phase, where in our approach causal factor exploration for problems is one of the crucial tasks.

Exploring the cause of a problem can be expedited by using previous knowledge about the FSCN. For example, if the manager in the DOA case somehow knows that the potential cause of DOA for some type of chickens is the transport time, he can directly go on from problem identification to verifying this cause. Accommodating such knowledge has to be taken into account in the requirements on this system.

3.2.3 Requirements on early warning and proactive control systems

From DSS literature we gathered general requirements on DSS (Zachary 1986; Winterhalder and Kleinschmidt 2000). Here we specialize and expand these requirements to the context of EW&PC systems in FSCN. From case studies in FSCN, we discovered several common case features, such as problems that need to be solved quickly, focus on performance indicators signaling the existence of the problem, various stages involved in the network structure, and multiple data sets recorded at different places of the supply chain network. What we learned from case studies is not only the specific requirements and structures of each separate case study. At least as important are requirements on a general level and generic structures of information involved.

Based on Glinz's (2005) classification, we distinguish requirements encompassing *performance requirements* (such as speed and memory usage), *specific quality requirements* (concerned with properties of usage and management, such as reliability, usability, and maintainability), and *functional requirements* (that describe what functions the system should perform).

Performance requirements on EW&PC systems are mainly concerned with ensuring that use of the system can be integrated in the business decision process. That translates into requirements with respect to the time needed to use such systems. It is important to take timely actions to prevent losses in FSCN. This requires that the time needed for searching for acceptable outcomes should be shorter than the time available for devising and executing actions. This observation has two implications for system design. First, time consuming analysis should be avoided in the system. Second, the system should always find acceptable solutions within the allotted time frame. We should eliminate time consuming tasks from the system, such as long time searching for optimal parameter settings when acceptable parameter settings are sufficient.

Specific quality requirements deal with the usage of such systems. It should be easy for managers to use. As mentioned by Mintzberg (1973), managers have little idle time, and their activities are characterized by brevity, variety, and discontinuity. These features require that the use of EW&PC systems should fit to the situation at hand. We have to provide managers with assistance during their interaction with the system. The system should help them to understand the steps to be executed, the techniques used, and how to interpret the results. The user interface should be augmented with explanations for the terminologies of the system. Since this system expects managers to use DM methods for data analysis and modeling, it is mandatory to provide them with enough help in such steps. DM expert knowledge on using DM methods should be captured in wizards and templates.

Functional requirements are more specific. We have identified six functional requirements on EW&PC systems in FSCN:

F1. Support problem recognition and expression

For conventional DSS, Klein and Methlie (2000) describe this function as important. Problems in FSCN do not naturally present themselves in quantitative ways. For example, based on expert knowledge, the occurrence of DOA is defined as more than 0.5% of a flock being dead upon arrival at the slaughter house. By formulating the problem in such a quantitative way, we can analyze the problem with DM methods. As shown in Section 3.2.2, this is the first generic step to be executed for EW&PC. So it is preferable for such a system to utilize expert knowledge and to facilitate formulating problems in a quantitative way.

F2. Support data disclosure

As shown in Figure 3-3, finding data is the second generic step. In the DOA case we identified various data bases available in different stages of FSCN. In order to apply DM methods for quantitative analysis, it is necessary to combine data sets from relevant data bases and transform them into a required format. However, when data sets have different kinds of meta-data, this precludes such combination. So it is important to give managers suggestions on what kinds of data are needed for solving a particular problem and for making decisions. The manager should get clear and easy to operate guidance on distinguishing which data sets can be combined and how to arrange them into appropriate formats for data analysis. Klein and Methlie (2000) also point out that DSS should provide information about useful data for solving the problem.

F3. Support analysis and reasoning, and provide model building facilities

One of the essential steps for EW&PC is exploring the potential causes for problems in FSCN through data analysis (see Figure 3-3). Compared to conventional investigation, data analysis is a faster way to find out causes for problems. With the help of quantitative analysis, EW&PC systems should facilitate a manager to quickly identify suspected causes for problems. From analysis of recorded data, we concluded that long transport time is an important factor for DOA in some breeds of chickens. The fact that our systems search for previously unknown potential causes for problems is a major difference between our systems and conventional DSSs.

F4. Support prediction of problems

Zachery (1986) claims that one of the contributions of DSS is to facilitate users to predict future events. Predicting potential problems is also a generic step for EW&PC. In order to give time for taking actions to deal with a problem, it is necessary to predict the problem as early as possible.

F5. Support evaluation of solutions

Decision evaluation is another generic functional requirement for conventional DSS (Holsapple and Whinston 2000; Winterhalder and Kleinschmidt 2000). As shown in Section 3.2.2, evaluating various decisions for proactive control is a generic step to be fulfilled. This task requires not only expertise in interpreting results from DM methods, but also domain knowledge on potential influences of different decisions. The EW&PC system should support the manager to utilize a model built during analysis and his domain knowledge, to experiment with different decisions.

F6. Support storage, retrieval, and organization of relevant resources (data, information, and knowledge)

Zachery (1986) discusses the importance of information control techniques for decision making support. In the DOA case, if knowledge resources show that the transport time is a potential cause of DOA, it is possible to skip the step of exploration of causes. The EW&PC system should indicate that prior knowledge on a problem exists and disclose that knowledge. It is up to the user to decide whether to use this previous knowledge or search for other causes. It should also be able to accommodate new knowledge that is obtained during problem solving and decision making. Further, the system should be able to update existing models in the system with new knowledge.

Because of the rapid development in DM research, new DM methods appear continuously (Han and Kamber 2001; Larose 2005). This requires the EW&PC system to be extendable with new DM methods and the associated knowledge.

3.3 Framework design

From the requirements presented above, we started the development of a framework for EW&PC systems. The architecture of this framework is depicted in Figure 3-4.

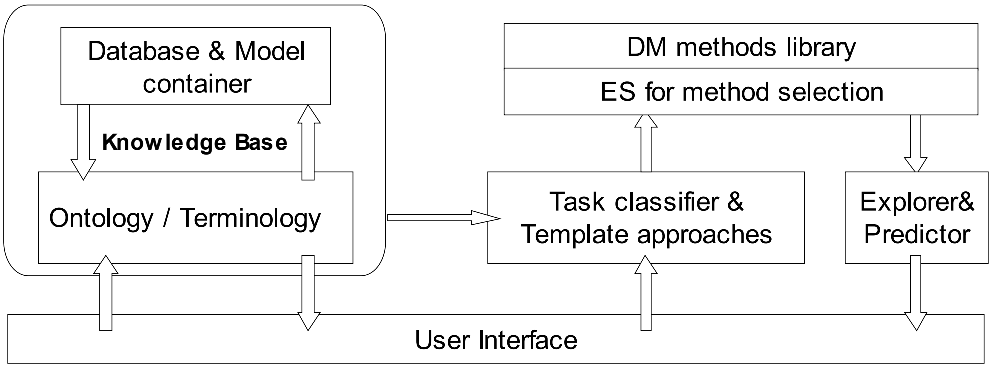


Figure 3-4: framework for early warning system in Food Supply Chain Networks

This framework is based on traditional frameworks for building DSSs (Sprague 1980; Sprague and Carlson 1982; Beulens and Van Nunen 1988; Holsapple and Whinston 2000). However, there are differences between EW&PC systems and traditional DSSs. Besides difference in target problems, there is also a difference in the way of achieving objectives. Traditional DSSs store models built by system designers and ask users for information required by those models. EW&PC systems build decision support models from existing data sets indicated by the user. Therefore, these systems are less strict in the information they require from users. So we expand the traditional framework with new components (Template approaches, Expert System for method selection, etc.) to provide sufficient and flexible assistance to build decision support models for different kinds of problems in FSCN. This framework helps managers to accomplish the generic steps shown in Figure 3-3, and complies with the various requirements discussed above.

In the subsections below, we describe the components of the framework in more detail and relate them to the generic processes and requirements for EW&PC systems. The discussion is organized around the major components of the framework: *Knowledge Base*, *Task classifier and Template approaches*, *DM methods library with Expert System (ES) for method selection*, *Explorer and Predictor*, and *User Interface*. These components interact with each other and with users (i.e. managers) to solve various problems in FSCN. Example implementation of this framework and its components can be found at the Internet address http://www.inf.wur.nl under “Staff” > “Yuan Li” > “EWPC prototype”. The processes for using this framework are discussed in next section.

3.3.1 Knowledge Base

The requirements indicate that the system should provide facilities for storage, retrieval, and organization of relevant resources (data, information, and knowledge) for solving problems in FSCN. From literature on problem solving and decision making (Simon, Dantzig et al. 1987; Winterhalder and Kleinschmidt 2000) we learn that at least two kinds of knowledge are needed for expert problem solving and decision making: (1) knowledge on a specific domain, and (2) knowledge on inference and analysis. The first aspect is taken care of in the *Knowledge Base*. It incorporates knowledge on previously encountered problems in FSCN and their causes. For the second aspect we use other components of the framework (*Template approaches* and *ES for method selection*), described below. The *Knowledge Base* is a very essential part of EW&PC systems. It stores information on previously encountered problems in FSCN, such as specific performance indicators with normal and abnormal ranges. Such information contributes to fulfilling the functional requirement of problem recognition and expression. The *Knowledge Base* also contains information about problem solutions for easy reference by later users (causal factors, causal relations, methods used to identify those factors and relations). The requirement of information storage and organization is satisfied because the *Knowledge Base* accommodates novel knowledge obtained by users in analyzing problems. Two links connecting *User Interface* and *Knowledge Base* indicate that users can both obtain information from the *Knowledge Base* and extend the *Knowledge Base* with their expert knowledge. The functions of the *Knowledge Base* are realized by two major constituents: the *Database and Model container,* and the *Ontology and terminology* component.

3.3.2 Database and Model container

One of the requirements on EW&PC systems is that they help users to find relevant databases and organize them. Furthermore, different models describing relations between problems in FSCN and determinant factors will be built based on those databases. The database stores data sets that were obtained from FSCN. These data sets contain analyzed data about various aspects at different stages in FSCN regarding quality of food products and performance of operations. The *Database container* should be flexible enough to accommodate different data sources. In the DOA case, the managers collected data from different databases and merged them into one data file for analysis. All formats of data from different databases should be manageable in the *Database container*. The *Model container* collects models that have been built using those databases. These models are intended for prediction of potential problems and decision making to take appropriate actions. The *Model container* should be able to accept different forms of models that result from various DM methods. These functions of the *Database and Model container* comply with literature on DSS design; e.g. Sprague (Sprague 1980) points out the importance of databases and models for decision making. The *Database and Model container* should also be easily extendable in order to quickly assemble new data sets and models. The links with the *Ontology and terminology* component help to accomplish this objective.

3.3.3 Ontology and terminology

The knowledge stored in *Knowledge Base* is a valuable reference for managers in FSCN with similar problems. In order to help users to select applicable knowledge, the system should include an *Ontology* for the *Knowledge Base*.

An ontology is a formal description of domain knowledge. It consists of concepts, their properties and relations, structured according to human conceptualization. As defined by Gruber (Gruber 1995), "An ontology is an explicit specification of a conceptualization." Borst (Borst 1997) expands the definition into "An ontology is a formal specification of a shared conceptualization". According to Uschold et al.(Uschold and Gruninger 1996), there are three kinds of application of ontology: communication between people and between organizations, interoperability between systems, and systems engineering (improving system reliability, facilitating re-use and making knowledge explicit). The component *Ontology* describes the structure and hierarchy of concepts in the databases and models in the *Database and Model container* in an understandable, extendable, and easy to access way. With this component, users can reason on the semantics level instead of data level. They can compare their cases with cases stored in the *Knowledge Base* to look for similar cases and relevant knowledge about those cases. Because the comparison is supported by reasoning on a semantic level rather than data level, this can substantially save management time. The *Terminology* provides explanation for terms from DM and FSCN, and defines relations between those terms to support automatic machine processing. The *Ontology and terminology* component ensures that new databases and models can be incorporated at appropriate places in the *Database and Model container* quickly and easily. It also enables the user to quickly browse the content of the *Knowledge Base* and search for relevant models for his own case. Different kinds of methodologies and tools for designing this component can be found in literature on Ontology Engineering (Corcho, Fernandez-Lopez et al. 2003; Pinto and Martins 2004).

3.3.4 Task classifier and Template approaches

As discussed in Section 3.2.2, there are several generic steps for EW&PC (see Figure 3-3). Based on those generic steps and the DOA case at hand, we defined different kinds of tasks that a user may have to perform (see Table 3-1). We added the tasks "Search for cause" and "Extend knowledge base" because knowledge contained in the *Knowledge Base* improves the decision process. The *Task classifier* helps users to quickly identify their task type. After that type has been identified, the user can obtain suggestions on how to perform some of the tasks listed below.

Table 3-1: different types of tasks in EW&PC systems

Task	Description	Remarks
Identify problem	Identify a problem that arises in FSCN and formulate it quantitatively.	For quantitative analysis of a problem, it is necessary to use a format that is appropriate for the analysis techniques.
Search for cause	Search in Knowledge Base for existing knowledge on potential causes for this problem.	Knowledge on causes of problems can speed up quantitative analysis. So this task should be executed before other tasks that are meant for quantitative analysis.
Find data	Find relevant data sets that can be used for exploring potential causes and combine the data sets.	If the Knowledge Base does not contain relevant knowledge, the user himself has to find relevant data sets for quantitative analysis of the problem.
Explore cause	Explore potential causal factors for the problem.	The user can apply appropriate DM methods to explore causal factors from relevant data.
Verify cause	Verify the causal factor by manipulating (improving) it in practice and observing real-world effects.	When potential causal factors have been found, either from the Knowledge Base or from quantitative analysis, the user has to verify those causal factors.
Early warning; Decision making	Predict problems that are likely to occur. Evaluate alternative decisions in order to choose the best solution.	This task stands for the steps 'predict' and 'decision making' in Figure 3-3. The system predicts potential problems and proposes actions based on the current situation in FSCN.
Extend knowledge base	Extend the Knowledge Base with obtained knowledge on causal factors and applicable DM methods.	The user can add knowledge obtained in earlier steps to the Knowledge Base for later reference when dealing with similar problems in FSCN. Knowledge could also contain decisions made to solve the problem.

In Figure 3-4, there is an arrow from *User Interface* to *Task classifier and Template approaches*. Through this link, the *User Interface* interacts with the *Task classifier and Template* approaches to support a manager in FSCN with looking for causes of his problem, either by searching the knowledge base or by doing quantitative analysis.

Another link connects the *Knowledge Base* to the *Task classifier and Template approaches*. It indicates that the *Task classifier and Template approaches* obtains information from the *Knowledge Base*. The information could relate to potential causal factors for a problem, or to appropriate methods that are applicable for such problems.

The tasks listed above can and need to be enhanced by corresponding *Template approaches* to expedite these tasks and support adherence to precedence constraints. The *Template approaches* component provides wizards that support users in executing these tasks. Generally speaking, we can classify template approaches into the following groups according to the kind of knowledge incorporated in the templates:

a. Problem formulation and resource collection:

Since this system has to support users in problem identification and formulation, as well as collecting associated resources (data, information, and knowledge) for problems, we designed *Template approaches* to help them to quickly understand and execute these tasks. These templates are independent of knowledge about the Knowledge Base and DM methods.

b. Searching and extending the knowledge base:

After the knowledge base has been constructed, it is important to give users examples of using the knowledge base. The use could be either searching the knowledge base for existing knowledge, or extending it with newly obtained knowledge. Template approaches for searching and extending the knowledge base could reduce time for learning to operate the knowledge base.

c. Quantitative analysis:

In order to achieve EW&PC, users may have to go through several steps, such as exploring causal factors and prediction. Extensive knowledge on DM is necessary to execute these steps correctly, so we need to insulate users from the details of DM methods. Although there are DM software systems (e.g. SPSS Clementine (SPSS 2003), SAS Enterprise Miner (SAS Institute 2004)) that incorporate procedural knowledge for executing particular DM methods, these systems are still aimed at DM experts. To guarantee appropriate usage of DM methods, the user should be experienced in formulating objectives, checking usability and consistency between data and preconditions of particular methods, and interpreting obtained results, etc (Bockenholt, Both et al. 1989). So, we have to design template approaches to incorporate that kind of knowledge and hence help users to correctly choose appropriate DM methods and use them to identify knowledge from recorded data.

d. Decision variety assessment:

In order to evaluate different possible decisions, users need to apply their insight in available decisions and expected effects. Users need to know how to employ constructed models for decision making, and how to interpret results of DM methods. Combining these two kinds of knowledge with their expert knowledge on decision varieties enables users to select the best solution for their problems in FSCN.

3.3.5 DM methods library with ES for method selection

In order to fulfill the functional requirement of supporting problem analysis and modeling, our systems should help users in selecting appropriate DM methods for different kinds of quantitative analysis. Research in DM generates more and more methods for different kinds of purposes (Adriaans and Zantinge 1996; Fayyad, Piatesky-Shapiro et al. 1996). As different methods have different strong features (Alexandros and Melanie 2001), the system should support a proper choice of DM method. The *DM methods library* stores comprehensive information about all kinds of DM methods that can be used for EW&PC. The information about a DM method includes its function, model format, and requirements on data set. Since new DM methods emerge rapidly, this library is extendable for new methods.

From literature on DM as well as case studies we recognize that the choice of DM method not only depends on DM function, but also on data requirements and extendibility (Li, Kramer et al. 2006). To facilitate users in selecting DM methods, the *DM methods library* is augmented with an *Expert System (ES) for method selection*.

Research in automatic DM method selection (Brodley 1992; Craw, Sleeman et al. 1992; Alexandros and Melanie 2001; Verdenius 2005) provides us with enough scientific insights in building *ES for method selection*. With such a component, a user can select proper methods according to his functional requirements and data characteristics. The expert system contains facts, heuristics, and inference procedures to mimic the method selection processes of DM experts. After obtaining the required information from users, this component can give suggestions on which DM methods to use and explain its reasoning.

Figure 3-4 shows a link from *Task classifier and Template approaches* to *DM library with ES for method selection*. This link indicates that a task involving method selection makes use of the *ES for method selection*.

3.3.6 Explorer and Predictor

Users employ the *Explorer* component in this framework to analyze collected data sets in order to find potential causal factors for the problems in FSCN. This task requires that the user selects appropriate DM methods for analysis. So in Figure 3-4 there is a link between *ES for method selection* and *Explorer*. The outcome of *Explorer* is a set of candidate causal factors for the problems at hand. Such knowledge needs to be verified by the user. After verification it can be used for future prediction of the same type of problems.

The *Predictor* combines stored causal relations with current data to warn for problems that are about to occur. The *Predictor* is used for decision evaluation as well. Users can employ models built previously to compare results of different available decisions and choose the best one.

Causal factor exploration and prediction need DM methods with a high level of prediction accuracy. Different kinds of data sets require different DM methods to arrive at high accuracy (Michie, Spiegelhalter et al. 1994). So the *Explorer and Predictor* is linked to the *DM methods library with ES for method selection* to support the choice of appropriate DM methods. The *Explorer and Predictor* sends the results of exploration and prediction to the *User Interface* to be shown to the user.

3.3.7 User Interface

The *User Interface* is the component responsible for communication between users and the system. Both input and output may use a variety of formats. With the help of various template approaches, the user interface supports users in executing each step in EW&PC.

The *User Interface* should at least incorporate the following functionalities:

- browsing the knowledge base or extending it with expert knowledge or new findings, e.g. from quantitative analysis;
- describing a problem, preparing relevant resources for solving the problem, and supplying required information to the system;
- exploring causal factors for problems in FSCN by means of DM methods and domain knowledge, including managerial judgment;
- inspecting available decision varieties and prediction results.

As discussed in Section 3.2.3, one of the requirements on this system is ease of use. So these interfaces should be clear in structure, provide adequate help, and direct the manager with readily understandable questions in order to reduce time needed by the manager.

3.4 Processes for using the system

Now that we have all components of the system, we can start to discuss how to apply them for EW&PC. Figure 3-5 shows the processes for using the system and Table 3-2 is used to explain the content of the steps for the DOA case. The sequence of processes is a refinement of the steps shown in Figure 3-3. To take into account the components of the framework, we added new processes between “identify problem” and “find data sets”, and refined “explore cause” and “verify cause” into more detailed processes. Numbers in Figure 3-5 correspond to the steps listed below. The sub numbering of steps 5 and 6 emphasizes the fact that these are refinements of the generic steps in Figure 3-3.

We illustrate the processes in Table 3-2 by the actual steps in the DOA case. In the data set for the case there are 63 attributes and 1357 records available. Each record corresponds to one flock of chicken. The data consists of records from December 2004 to July 2005 extracted from the operational data of the company.

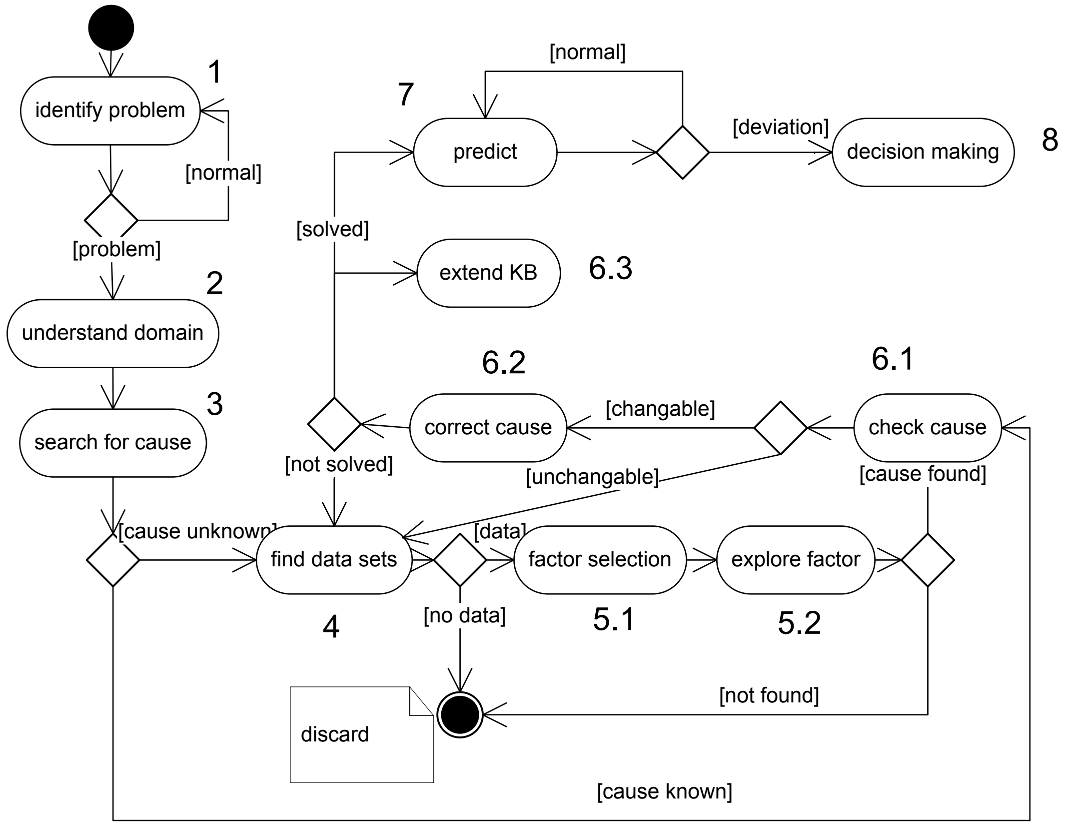


Figure 3-5: processes for using the framework; see text for numbering

1) *Problem identification and formulation.*
   The process starts when a manager discovers (or is informed about) a problem in his FSCN, e.g. regarding food quality. He formulates the problem in a quantitative way. This step requires his domain knowledge on the normal and abnormal values of performance indicators related to the problem. *Template approaches* for problem formulation can help the manager to carry out this step correctly and quickly.

2) *Domain understanding.*
   This step is a prerequisite for almost all other steps. In this step, the manager checks his knowledge on his FSCN and complements it if necessary. This knowledge relates both to different stages in FSCN, and to operations, people, and resources involved in each stage. In each stage, there are different factors that potentially cause problems in food products, such as operational errors (e.g. wrong usage of detergents), or environmental deviations (e.g. high temperature). Knowledge about those factors is needed for identifying causal factors. After this step, the manager knows the stages that exist, as well as the kinds of operations and resources used in each stage.

3) *Search KB for causal factors.*
   After the problem has been formulated and the domain understood, the *Task classifier* directs the manager to search the *Knowledge Base* for existing knowledge on possible causes of the problem. If such information is present in the *Knowledge Base*, managers may skip directly to

step 6.1 to check the causes found. *Template approaches* for browsing the *Knowledge Base* can help managers to execute this step quickly.

Table 3-2: summary of each step in the DOA case study

Step	Practice	Result
1. Problem identification and formulation	*Template approaches* for problem formulation ask which variable indicates the problem and its normal range.	DOA is defined as death percentage upon arrival at the slaughter house is above 0.5%
2. Domain understanding	The manager finds stages involved in his FSCN and the associated factors in each stage.	See Figure 3-2 for the stages and example factors.
3. Search KB for causal factors	The manager uses the *Knowledge Base* to look for previous knowledge on the DOA problem.	Since the problem was new, no previous knowledge was present yet.
4. Finding data sets	The manager combines the breed data from the hatcheries, the transport data from the delivery process, and the data on DOA and rejection obtained at the slaughter house.	Combing data from several operational data bases resulted in a data set with 63 fields.
5.1 Preliminary factor selection	The manager selects variables that might have influence on DOA.	Ten variables were selected from various stages.[3]
5.2 Causal factor exploration	The *ES for method selection* suggests using the DM method “Decision Tree”.	The outcome indicates two causal factors: “Transport time” and “Breed”. One particular breed is vulnerable to long transport time.
6.1 Check changeability of cause	The managers consider whether the causal factors found can be controlled in reality.	The transport time can be controlled directly, whereas the breed of chicken can be controlled by not accepting some breeds.
6.2 Correct the cause	The manager decides to temporarily stop accepting the vulnerable breed from farms too far from slaughter house.	The DOA occurrence drops below the acceptable threshold.

[3] from the broiler house (breed of chicken, location of broiler house), the farm house (age of chicken, average weight, growth rate), the transportation stage (catch method, transport time, density of chicken), and the slaughter house (waiting time, time without feed)

6.3 Extend KB	*Template approaches* for extending the *Knowledge Base* guide the manager to incorporate the obtained relation.	The manager adds a rule to the *Knowledge Base* that indicates the observed influence of breed and transport time on DOA.
7. Prediction	The manager uses the *Predictor* to monitor potential DOA. The *ES for method selection* suggests using the DM method "Neural Network" for prediction.	The chosen prediction method predicts the possibility of upcoming DOA based on the breed of chicken and their transport time.
8. Decision making	The manager searches the *Knowledge Base* for control measures. *Template approaches for decision making* help to forecast the effect of different control measures.	When DOA is predicted based on breed and transport time, the delivery is assigned high priority at the slaughter house.

4) *Finding data sets.*
In order to quantitatively explore causal factors for a problem, it is necessary to find data sets that contain relevant factors for the problem. If such data sets are not available, the manager can only proceed after improving data collection in his FSCN.

We expect the manager to understand the meaning of attributes in available data sets and relations between those attributes and factors. With that understanding it is possible to combine different data sets into one database, provided that the meta-data of those data sets does not preclude such combination (e.g. when no matching key fields are available).

Template approaches can help managers in two ways in this step: searching relevant data sets, and storing combined data sets into the *Database container*.

5) *Explore causes.*
Here we distinguish *Preliminary factor selection* and *Causal factor exploration*.

5.1) *Preliminary factor selection.*
The database resulting from the previous step may contain much more factors than DM methods could handle. Large amounts of unrelated factors not only waste time in analysis, but also may cause DM methods to generate wrong results (Liu and Motoda 1998). In this step, the manager goes through the database and uses his expert knowledge to select variables that are likely to relate to the problem.

5.2) *Causal factor exploration.*
In this step, the manager uses the *Explorer* to quantitatively analyze selected variables to find potentially causal factors for the problem. *ES for method selection* can help to quickly find suitable DM methods from the *DM methods library*. *Template approaches* can guide the execution steps of a selected DM method, such as setting parameters and interpreting results.

Results indicate the suspect factors and their normal ranges, and how those factors could influence the problem. Managers should use their expert knowledge to judge the results of analysis.

6) *Verify causal factors and extend knowledge base.*
This step is divided into *Check cause*, *Correct cause*, and *Extend KB*.

6.1) *Check changeability of cause.*
In this step, the manager checks the validity of potential causal factors obtained from either the *Knowledge Base* or quantitative analysis. The ultimate way to verify hypothesized causal factors is to change them in practice and observe the results. However, it is possible to do so only when the selected factor is changeable. If the factor is not changeable, or changing would be impractical (too costly, hazardous, etc.), the manager has to continue the search for other causal factors.

6.2) *Correct the cause.*
When a potential causal factor is practically changeable, the manager can verify the causal factor by modifying it in the actual FSCN. If observed results do not confirm the causal factor, the manager can change certain settings, such as choosing another DM method (with *ES for method selection*) or selecting even more factors in previous steps (with *Template Approaches*) and do analysis again.

6.3) *Extend KB.*
If observed results confirm validity of the causal factor, the manager can extend the *Model container* with the model built for identifying causal factors, together with the knowledge obtained about those factors. *Template approaches* for extending the *Knowledge Base* will guide the manager to accomplish this task.

7) *Prediction.*
When a causal factor has been found, it can be monitored continuously in the actual FSCN. So this task requires the monitoring system to record data on identified causal factors. An appropriate DM method is used as *Predictor* to find potential problems in the recorded data. *ES for method selection* helps in finding proper DM methods to predict problems based on recorded data. *Template approaches* will guide in applying a chosen DM method for prediction. When the *Predictor* finds that a deviation has occurred or is about to occur, it signals a warning to the user.

8) *Decision making.*
After obtaining a warning signal about a potential problem, the manager considers different decisions for changing production process parameters to prevent or repair the predicted problem. By inspecting the *Knowledge Base*, the manager finds the decision variety of possible control measures for proactive control. *Template approaches* for decision variety assessment can help managers in interpreting prediction results of DM methods and fine tuning the production process parameters. The template approaches help the managers to investigate the forecasted effects of changing parameters. In the end, managers should use their domain knowledge to make a choice from available options.

3.5 Conclusion and discussion

Starting from generic steps for early warning and proactive control, we found specific requirements for Early Warning and Proactive Control systems in Food Supply Chain Networks. To fulfill these requirements, we designed a framework for such systems and the processes for using the framework. We implemented a prototype EW&PC system and applied it to a new case. In section 5.1) we summarize the evaluation results of the framework and prototype.

As discussed in the introduction of Section 3.3, the design of this framework differs from the architecture of traditional DSSs because of special requirements posed in our context. In section 3.5.2 we highlight what is new in our framework. In section 3.5.3 we discuss why and how it may be applied in other domains as well. We conclude by indicating issues for further research in section 3.5.4.

3.5.1 Evaluation

After the prototype system was built, we applied the system and processes in a number of settings. Firstly, we asked students to use the prototype system incorporating data from the DOA case. The purpose of this test was primarily to find potential flaws in the prototype and process descriptions. Secondly the system was applied to a new case study. Thirdly, we invited a group of experts to evaluate the system design in a number of expert validation sessions.

The student test indicated which parts of the system were difficult to understand for non-expert users. Moreover, a number of technical flaws in the prototype implementation emerged. The concepts and processes for using the system did not have to be changed, however. After corrections to the prototype implementation and the texts in the template approaches, the system has not been changed anymore. This is the system we used for further evaluation.

To test the sequence of steps and the components in the usage processes, we applied the prototype system in a new case from a global meat company. This test strictly followed the processes specified in Section 3.4. This case study supports that the designed processes actually provide correct and usable guidance for EW&PC.

In the expert validation sessions we presented the prototype system and a description of the processes to a number of experts. The domains of these experts cover Management in Food Supply Chain Networks, Data Mining, Decision Support Systems, and Knowledge Engineering. After exercising with the system, these participants were asked to evaluate the framework and system, as well as the design process. Results show that our prototype system in general satisfies the formulated requirements. Furthermore, the design process is regarded as correct and reusable.

Our confidence on the practical value of this system is further strengthened by a demonstration session in a Dutch food company. We presented the framework and prototype system to 12 senior managers. In the discussion afterwards, those managers gave high credit to the idea of EW&PC system and the potential value in various sectors of FSCN management (e.g. feed control, meat quality optimization, risk analysis). They had a very good impression of the prototype system. As a consequence, they started to consider organizing new research or projects for implementing EW&PC systems in their domain areas.

3.5.2 Novelty

An important difference between our framework and conventional DSSs is the degree of flexibility of underlying models and approaches. Conventional DSSs intend to solve specific types of problems, such as prediction of agricultural drought (Kumar 1998), or determining telemarketing operations (Ahn and Ezawa 1997), etc.. Our framework enables managers to investigate new types of deviations as they turn up. By analyzing available data collected about food products and associated processes over the FSCN, EW&PC systems help managers to investigate all kinds of potential problems that may occur or have occurred in FSCN.

The foundation of the difference is in the requirement on information. According to Holsapple and Whinston (2000), "A DSS is constrained by the knowledge it possesses. It cannot process knowledge it does not have. Its knowledge at any moment may or may not be sufficient to respond to a decision maker's requests." Because conventional DSSs aim at specific kinds of problems, they already incorporate the knowledge on the causal factors of that kind of problems. For example, in order to predict agricultural drought, Kumar (1998) models relations between causal factors and drought. The causal factors in this case are predefined variables, such as amount and distribution of rainfall, and number of days delayed in crop sowing. Those models are a fixed part of the system. Therefore, the DSS prescribes the kinds of information from users, in this instance rainfall and crop sowing. Contrary to that, our framework does not make use of fixed models. Instead, models are derived from available data, so our framework does not put restrictions on the kinds of information. As long as users have relevant data for encountered problems, they can use our approach to explore causal factors. This is especially beneficial since different FSCN have different kinds of critical performance indicators and information sources.

Although DSSs exist that employ DM methods for analysis and prediction (Kim and Street 2004; Van Gestel, Baesens et al. 2006; Chen, Kuo et al. 2007), such systems rely on professionals in Data Mining or Statistics to execute these tasks. Since EW&PC systems are intended to be used by managers, our framework is designed for non-expert users in DM. The framework contains components like *ES for method selection* and *Template approaches* to support managers, who normally are not experts in DM. Thus, managers can analyze existing data without expertise in DM methods.

A final novelty of this framework is that it provides a mechanism to combine Business Intelligence with Knowledge Management. Business Intelligence focuses on well structured data and information. However, practical problem solving also relies on non-structured information, represented by e.g. documents and procedures. Research in Knowledge Management and Ontology Engineering provides us with facilities to capture, model, store, and organize that knowledge. As a result, several researchers already proposed the idea of enhancing Business Intelligence with Knowledge Management (Haimila 2001; Nemati, Steiger et al. 2002; Herschel and Jones 2005); Cody et. al. (Cody, Kreulen et al. 2002) also implemented a prototype system. Our framework contributes to implementing such a combination. The *Knowledge Base* accommodates the knowledge obtained by managers from data analysis, and organizes it in an appropriate way for perusal by other managers. Furthermore, the *Template approaches* capture the expert knowledge on how to execute the steps for early warning and proactive control, such as data analysis and decision assessment. Those template approaches, together with the *DM library and ES for method selection*, enable managers, with no knowledge on DM, to use DM methods to analyze data sets obtained from Business Intelligence facilities, and to take appropriate action.

3.5.3 Reusability

The context of building the framework presented here is EW&PC in the domain of FSCN. However, we believe that the strategy we used here is applicable to other application domains, as well. The DM methods we employ are helpful for EW&PC in general. For example, in a steel company, managers can use a similar strategy to deal with imperfections on the surface of the produced steel (Cser and Naude 2004). They can apply DM methods to analyze recorded data from various production stages to find potential causes of imperfections. They can use these findings to predict and proactively control such problems.

Next to the strategy, most of the components in this framework are generic enough to be reused as they are. Although the components do not need changes, small changes in their context might be needed, e.g. for *Template approaches*.

Some components can be easily incorporated into other systems based on DM. The knowledge incorporated in the *ES for method selection* is specific for EW&PC, but not for FSCN. So this component can be reused for EW&PC in other application domains. Both the *Explorer* and the *Predictor* depend only on data sets and selected DM methods for their operation. The choice of data sets and DM method are domain dependent, of course, but that is covered by other components of the framework. So to reuse the *Explorer* or *Predictor*, no changes to these components themselves are necessary.

The *Template approaches* for guiding Data Mining need only moderate change before they can be used in other applications, because the *generic* procedure in Section 3.2.2 is independent of application domain. There are several generic steps in using DM methods (Fayyad, Piatetsky-Shapiro et al. 1996). Although there are various DM methods, only a few steps are method-specific, such as parameter setting and result interpretation. Other steps, like data collection, data cleaning, and data transformation, are independent of DM methods. Moreover, they are very similar across different application domains. As a result, template approaches related to DM methods can be easily adapted to be used in other application domains.

3.5.4 Further research

We have shown that the presented framework and usage processes are applicable in principle. The prototype system provided for 'proof of feasibility', but still needs to be extended in several aspects, such as polishing user interfaces and presentation, refining template approaches, and extending the KB content. However, the framework and usage processes do not have to change to incorporate such functionality. For practical use in reality, additional development is needed.

Besides those development opportunities, there are some issues that deserve further research. Firstly, it is a challenge to visualize domain knowledge in a natural way (such as Figure 3-2) for non-experts in ontologies. Secondly, research is needed on how to enable non-experts to correctly extend an ontology. Currently, without expert involvement, extending an ontology might compromise the structure of the ontology. Results of these two issues could enhance the steps *domain understanding* and *extending KB* of the usage processes in Section 3.4. A third issue is the reliability of prediction by DM methods. Different DM methods give different prediction accuracies. So research is needed on how to combine such outcomes. Finally, further case studies are needed to evaluate our framework and usage processes in more detail.

Chapter 4. Applying Data Mining for Early Warning in Food Supply Networks[4]

Abstract

In food supply networks, quality of end products is a critical issue. The quality of food products depends in a complex way on many factors. In order to effectively control food quality, our research aims at implementing early warning and proactive control systems in food supply networks. To exploit the large amounts of operational data collected throughout such a network, we employ Data Mining in various settings. This chapter investigates the requirements on Data Mining posed by early warning in food supply networks, and maps those requirements to available Data Mining methods.

4.1 Introduction

Food quality problems in food supply networks form a critical issue for both consumers and food companies. However, in recent years, food quality crises occurred frequently all over the world. A recent case is dioxin contamination in pork in Belgium, the Netherlands, and Germany. In order to effectively control food quality, we need early warning systems to predict potential problems and give suggestions for proactive control.

The primary source of information on food supply networks is expert knowledge. However, expert knowledge is not always sufficient to deal with new quality problems in a direct way. This is partly due to the complexity of food supply networks. Further, food products and food processing procedures show inherent uncertainty and variability. Recent developments in information systems of food supply networks provide us with possibilities to discover valuable information about quality problems from recorded data. We deal with these problems with the help of a powerful quantitative method – Data Mining.

Data Mining has been successfully applied in many areas, such as biology, finance, and marketing. However, the uptake of this technique in food supply networks has not matched the amount of applications in business. Historically food supply networks were less automated than other businesses. However, in recent years, food industry began to build information systems to collect data about various stages of food supply networks. These information systems provide us with opportunities to employ Data Mining techniques to discover interesting relations for food quality problems. Because of inherent properties of food products, it is not straightforward to adopt applications from other areas to the field of food supply networks.

[4] Li, Y., M. R. Kramer, A. J. M. Beulens and J. G. A. J. v. d. Vorst (2006). Applying Data Mining for Early Warning in Food Supply Networks. 18th Belgium-Netherlands Conference on Artificial Intelligence, Namur, Belgium.

In our research, we are aiming at employing Data Mining techniques to construct early warning systems in food supply networks. Such an early warning system will adaptively identify new problems in food quality, aid domain experts in discovering possible causes for these problems, and monitor those causal factors to predict potential food quality problems. We anticipate taking even a step further towards proactive control to provide measures to prevent or correct food quality problems.

4.2 Food Supply Networks

According to Santoso et al. (Santoso, Ahmed et al. 2005), a supply chain is a network of suppliers, manufacturing plants, warehouses, and distribution channels organized to acquire raw materials, convert these raw materials to finished products, and distribute these products to customers. Van der Vorst et al. (Van der Vorst, Beulens et al. 2005) extend this to a food supply network, referring to an interconnected system with a large variety of complex relationships such as alliances, horizontal and vertical cooperation, forward and backward integration in supply chains.

Figure 4-1 depicts a supply chain network with accompanying monitoring and control systems. Processes can be seen at different levels, depending on the kind of problems considered. They may be as large as complete farms, factories or warehouses, or as small as one individual activity. The monitoring system gathers performance data of processes and their inputs and outputs. The control system can influence settings of the processes involved. Together, the monitoring system and control system manage the whole series of processes.

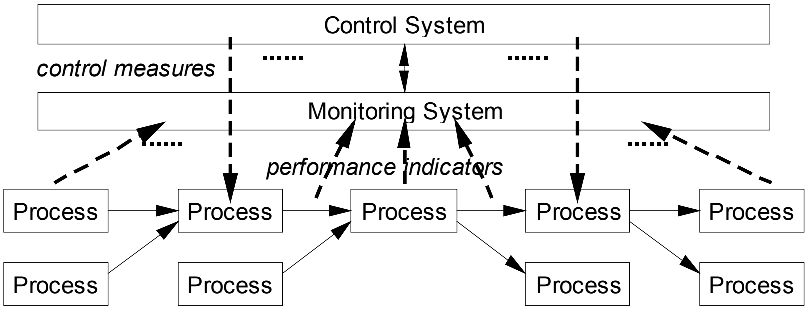


Figure 4-1: An example supply network

Food supply networks have specific characteristics compared to other supply networks (Van der Vorst, Beulens et al. 2005). A food supply network includes multiple stages, global sourcing, variety of sources, leading to a complex network structure. Moreover, food is an inhomogeneous material, with a wide uncertainty and variability in quality and quantity of supply, as well as yield of production processes.

Current research on performance measurement systems in food supply networks provides ways to quantify quality attributes of food products (Van der Vorst 2005). Each factor influencing a quality attribute has a certain effect. For instance, temperature is a important factor for many quality attributes of food products. In order to model the effect of those factors, we need quantitative models, especially when variability and uncertainty characterize the food quality attributes.

Next to uncertainty and variability, theoretical understanding of food quality problems is scarce. As a result, one has to rely on other sources of knowledge. Current advances in information systems in food supply networks provide us with another way to deal with this kind of problems: to induce knowledge from data.

4.2.1 Information Systems in Food Supply Networks

In food supply networks, the use of information systems has increased over the last decades, on the level of the network as well as on the level of participating companies. One reason is that the need for transparency in food supply networks has increased (Hofstede, Schepers et al. 2004). Transparency implies that extensive information associated with food products should be recorded along the supply network, e.g. regarding production data, product identification, and product and process properties. Several modern logistical and management trends such as E-commerce, scanning, total quality management, and HACCP also generate bulk data.

Due to these advances of information systems in food supply networks, large amounts of data about food production and processing are recorded every day. Apart from their original purpose, the information implicitly present in these data is valuable as a basis for implementing early warning and proactive control. First, abnormal change in these data will give indications for potential problems. Second, many unknown causal relations may exist between recorded data about inputs, controls and production means, and operational performance of processes inside food supply networks. Knowledge about these relations provides a possibility to prevent problems by monitoring and proactively influencing the corresponding determinant factors. So it is worthwhile to employ Data Mining methods to adaptively identify new problems and discover causal relations from recorded data.

4.3 Framework for Early Warning Systems

Early warning systems are well known in natural sciences. These systems, based on historical monitoring, local observation, or computer modeling, predict natural disasters, such as floods (Grijsen, Snoeker et al. 1992) or earthquakes (Wu, Chung et al. 1999), and help to prevent or reduce their impact. In food technology, Costello et al. (Costello, Ewen et al. 2003) presented a prototype sensor system for the early detection of microbially linked spoilage in stored wheat grain.

The early warning systems we intend to build should not only predict potential food quality problems, but also help to identify relations between determinant factors and quality attributes of food products. Ultimately, the knowledge about these relations and the decision varieties associated with these factors will enable proactive control to prevent those problems.

To achieve our objectives of early warning, we designed a framework for early warning systems in food supply networks (see Figure 4-2). The main distinguishing attribute of our approach is the aim for adaptivity. Other early warning systems are designed only for predicting specific predefined problems. Due to uncertainty in food supply networks, we do not have predefined knowledge about all kinds of problems that could occur.

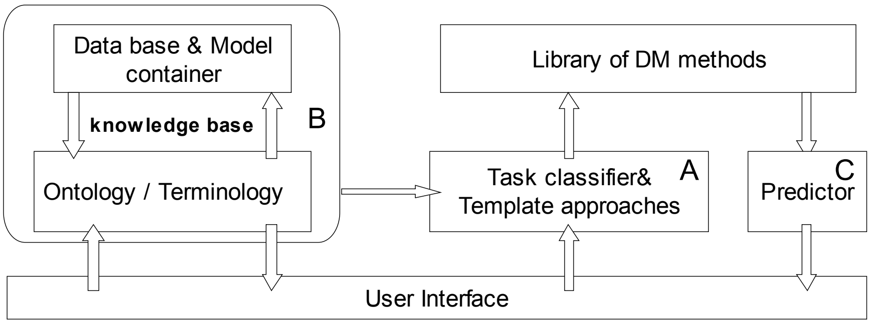


Figure 4-2: Framework for early warning system in food supply networks (letters A, B, C for reference in section 4.4)

In our framework, users with new problems follow a template approach, which will guide them to find causes of the problem, select appropriate Data Mining methods, and instantiate a new predictor. The knowledge base provides valuable references for all these steps. This knowledge base will be continuously extended with new cases and domain knowledge from experts. So it will be helpful not only for us to construct early warning systems, but also for other stakeholders to deal with similar problems.

4.4 Requirements on Data Mining

4.4.1 Requirements Imposed by Early Warning in Food Supply Networks

An effective early warning system helps to identify potential problems, and to analyze their characteristics and causes. An early warning system according to the framework of Figure 4-2 operates in an iterative way. After problem detection, it is necessary to explore determinant factors for the problem detected. Then we try to describe the structure and form of relations between factors and performance measures in an appropriate way. After this, we can update the model in our knowledge base with obtained new knowledge. Of course, domain experts may also make their contribution to the model. With the updated model, we could again monitor the determinant factors in food production and predict potential problems.

In order to realize early warning in food supply networks, we defined the following essential functional requirements. These requirements were informally obtained by generalizing existing early warning systems discussed above and by considering the characteristics of food supply networks.

1. *Prediction*: The primary purpose of early warning is to forecast potential problems as early as possible. So early warning systems should be able to describe future states of performance measures based on monitoring of historical and current determinant factors. (C in Figure 4-2)

2. *Problem detection*: In operational use, early warning systems should be able to detect new problems in food quality when they occur, and describe the characteristics of those problems. (A in Figure 4-2)

3. *Finding determinant factors*: After identifying a problem (either automatically or by expert input), it is necessary to investigate what are determinant factors for this problem, and how those determinant factors influence the performance measures involved. (A+B in Figure 4-2)
4. The first three requirements above relate to processes, whereas the next three relate to representation.
5. *Complex structure representation*: Due to the complexity of food supply networks, long chains of causal relations between factors may exist. Moreover these chains may branch and join at certain points. As a result, our system has to be able to represent those complex causal relations. (B in Figure 4-2)
6. *Different representation forms*: Many different kinds of relations between factors may exist. Relations between food quality problems and determinant factors may appear in different forms. Sometimes we find quantitative mathematical formulas directly relating one or more determinant factors to a performance measure. In other instances the relation takes the form of a conditional probability distribution. For some cases we do not have enough detailed knowledge and we can only give a model as a black box. So early warning systems require the possibility to show different representation forms. (B in Figure 4-2)
7. *New knowledge incorporation*: Our early warning systems should be capable of incorporating new knowledge when it is discovered, either from newly detected problems or by expert input (e.g. due to the continuous development in food supply networks). (B+C in Figure 4-2)

It is not necessary that one technique satisfies all requirements at once, because different techniques could be used for different steps in our framework.

4.4.2 Contribution of Data Mining to Fulfill the Requirements

In this section we look at how to employ Data Mining to fulfill the requirements discussed above. Our focus is mainly on the functional aspects, which means we look at what are the functions of Data Mining, and how these functions can be used to fulfill the requirements of early warning systems. Definitions of Data Mining functions are gathered from Fayyad (Fayyad, Piatesky-Shapiro et al. 1996) and Freitas (Freitas and Rodrigues 2006). Below, we provide tables that present suitable Data Mining methods for each function, to facilitate selecting appropriate methods for specific tasks.

Prediction. Prediction of food quality problem based on historical and current data of determinant factors is one of the principle requirements imposed by early warning. And fortunately one of the merits of Data Mining techniques is the prediction power. There are many success stories on application of Data Mining methods for prediction of food quality in food supply networks (Vallejo-Cordoba, Arteaga et al. 1995; Rousu, Flander et al. 2003).

Problem Detection. Signals that trigger early warning usually appear as anomalies in monitored performance measures first: deviations from the established norm and expected behavior. Deviation detection from various types of data is a prominent function of Data Mining (Freitas and Rodrigues 2006). Literature shows the application of several Data Mining techniques to identify problems in food production. One example is the use of neural networks in an X-ray system to identify contaminants in packaged food products (Patel, Davies et al. 1995).

Finding Determinant Factors. In order to identify determinant factors for food quality problems, ultimately the knowledge of domain experts is indispensable. However, Data Mining may help to quickly find candidate factors. Applicable Data Mining functions include causation modeling, factor selection, and to some extent classification and regression.

Causation models describe the causal relations between determinant factors and performance measures. Currently there are two methods available: Bayesian causal discovery (Pearl 2000) and constraint based causal discovery (Silverstein, Brin et al. 2000). However, both methods are only at an experimental stage. The number of successful applications of these two techniques is limited. See Freedman (Freedman 2004) for causes of this problem.

Besides causation modeling, various factor selection methods for Data Mining purposes could help to quickly find a set of potentially relevant factors to focus on, and determine the relative importance of each factor. Note however, that factor selection is not regarded as a Data Mining function, but rather as a preprocessing step. There are many methods for selecting relevant factors (Liu and Motoda 1998). In general, factor selection methods can either use a filter approach or a wrapper approach. A filter approach selects relevant factors independently of the Data Mining techniques used for prediction. But a wrapper approach can only be used in combination with a specific Data Mining technique (Kohavi and John 1997).

Some Data Mining methods for classification and regression can also help to find relevant factors. For example, neural networks can be used for classification. With sensitivity analysis on neural networks (Yao 2003), we can find which factors are influencing the outcomes of the neural network.

No matter what techniques are selected for causal relation discovery, the result should be checked by domain experts before being applied in practice. Only with the interaction between domain experts and Data Mining we can find valuable relations.

Complex Structure Representation. In order to represent interactions that span multiple stages of a food supply network, we need methods that are able to describe relations between factors from various stages. For this purpose, we may apply dependency modeling and causation modeling.

Dependency modeling describes significant dependencies among variables (Fayyad, Piatetsky-Shapiro et al. 1996). It is different from causation modeling in that causal relations are intuitively stronger than dependencies. The latter only indicates a correlation between determinant factors and performance measures, but there is no causality semantics in this relationship. Causation models not only show correlations, but also indicate that those determinant factors actually cause the observed effects.

One of the candidate methods for dependency modeling is Bayesian networks. This method builds a graphical network to describe the complex structure. Variables are represented as nodes, and dependencies between variables are represented as links. Variability in variables is described by conditional probability distributions, specifying the probability for each variable given the values of the ones linked to it.

There are many ways to combine the requirements imposed, functions of Data Mining, Data Mining methods, and some other aspects. Listing all possible combinations would yield a multi-dimensional table of all aspects. Here, we present the most important combinations of two dimensions in Table 4-1 and Table 4-2. The requirements of describing different kinds of relations and novel relation incorporation will be dealt with in following subsections.

In Table 4-1, we summarize the use of Data Mining for different functional requirements as reported in literature. In Table 4-2, we compare some commonly used Data Mining methods against the Data Mining functions mentioned above. We will use such tables for technique selection. For each non-empty cell we did find applications in literature. We use 'valid' to represent the fact that corresponding applications have been reported in literature (e.g. [nn] shows the application of classification methods for deviation detection). Due to space limitations, we cannot give all references here. Cells stating 'helpful' indicate that the technique may yield candidate relations but does not derive determinant factors by itself. See also the discussion above.

Table 4-1: Functions of DM vs. requirements from early warning system

Function of data mining	Requirements imposed by early warning system			
	1.Predict	**2.Detect problem**	**3.Find determinant factors**	**4.Describe complex structure**
Deviation detection		Valid		
Factor selection *			Helpful	
Classification	Valid		Helpful	
Regression	Valid		Helpful	
Dependence model	Valid		Helpful	Valid
Causal model	Valid		Valid	Valid

* Factor selection methods are usually regarded as a pre-processing step for Data Mining rather than a separate Data Mining function.

Table 4-2: Applicability of DM methods for specific functions

DM Function	**DM methods**				
	Decision tree	**Neural network**	**Bayesian network**	**Association rule**	**Nearest neighbors**
Deviation detection	Valid	Valid	Valid	Valid	Valid
Classification	Valid	Valid	Valid		Valid
Regression	Valid	Valid	Valid		Valid
Dependence model			Valid	Valid	
Causal model			Valid		

Different Representation Forms. Table 4-3 shows relevant representation forms and corresponding Data Mining methods.

Selection of appropriate representation forms depends on the purpose of representation, the data characteristics, and the knowledge on the relation. For example, if the purpose is to represent causal relations to users of an early warning system, then we should choose more understandable representation forms, such as decision tree or association rule.

New Knowledge Incorporation. As discussed above, there are two kinds of novel relations that need to be incorporated: relations discovered from new problems, and relations obtained from domain experts. The requirement to incorporate such new relations can be fulfilled easily by some of the Data Mining methods, such as association rules and Bayesian networks. For example, in a case study to predict DOA (Death On Arrival) in a chicken supply network, we obtained knowledge that DOA will increase with transportation density. It is easy to add this knowledge as an association rule to a rule set, or as a node and an arc to a Bayesian network. Table 4-3 also provides an overview of capabilities of different methods for incorporating new knowledge into models constructed with these methods.

Table 4-3: Representation forms of DM methods and extensibility of corresponding models

Data Mining method	**Representation form (Fayyad, Piatetsky-Shapiro et al. 1996)**	**Novel knowledge incorporation**
Decision trees	Decision trees	Easy
Association rules	Rules	Easy
Neural networks	Linear or Nonlinear model	Difficult
Nearest neighbors	Example-base methods	Difficult
Bayesian networks	Probabilistic graphical dependency model	Easy

4.4.3 Discussion on Technique Selection

There are many aspects for technique selection. Function, representation form, and capability for novel knowledge incorporation are important aspects. The data format is another aspect. Various kinds of performance measures are available for food quality; some are quantitative, such as body weight; some are qualitative, such as objective evaluation of meat color. The data format (nominal, numerical) has to be taken into consideration as well when selecting a technique.

However, the quality of a model also depends on how well the model class is able to represent patterns in the data set. Some research on automatic technique selection has already been conducted. Three main lines of interest have been found: heuristic expert rules (Kodratoff and Moustakis 1994), meta-learning (Alexandros and Melanie 2001), and landmarking (Pfahringer, Bensusan et al. 2000). Verdenius (Verdenius 2005) has used data class boundary characteristics for selecting techniques. He distinguishes orthogonal and non-orthogonal (linear and nonlinear) class boundaries. Decision trees are especially suitable for orthogonal instead of non-orthogonal class boundaries, while neural networks are also good at non-orthogonal class boundaries.

4.5 Conclusion

This chapter explains why and how Data Mining can be helpful in building up early warning systems in food supply networks. The construction and functioning of early warning systems will inevitably require the involvement of domain experts. However, with the help of Data Mining, we can complement their knowledge with new, interesting relations.

We investigated the functional requirements for Data Mining in food supply networks, and presented an overview of applicable Data Mining methods for those requirements. This overview forms a starting point for technique selection for specific applications of Data Mining. Ongoing research will refine the framework and requirements introduced in this chapter.

As part of our research, we use a number of case studies to investigate the applicability of Data Mining methods. The first case study has been reported elsewhere (Li, Kramer et al. 2006); the next one is in progress. In subsequent steps in our research, we will use the information gained by these case studies to build a knowledge base for early warning in food supply networks. Data Mining technique selection will be one of the components of this knowledge base.

Appendix 4.1: Table 4-2 with complete reference

Table 4-2. Applicability of DM methods for specific functions

DM Function	DM methods: Decision tree	Neural network	Bayesian network	Association rule	Nearest neighbors
Deviation detection	Valid (Fayyad, Djorgovski et al. 1996)	Valid (Patel, Davies et al. 1995)	Valid (Agarwal 2005)	Valid (Balderasy, Berzal et al. 2005)	Valid (Knorr, Ng et al. 2000)
Classification	Valid (Verdenius 2004)	Valid (Gestal, Gomez-Carracedo et al. 2004)	Valid (Gorte and Stein 1998)		Valid (Soeria-Atmadja, Zorzet et al. 2004)
Regression	Valid (Lobell, Ortiz-Monasterio et al. 2005)	Valid (Millan, Roa et al. 2001)	Valid (Roos, Wettig et al. 2005)		Valid (Goulermas, Howard et al. 2005)
Dependence model			Valid (Heckerman and Breese 1996)	Valid (Smyth and Goodman 1992)	
Causal model			Valid (Pearl 2000)		

Chapter 5. Guiding Users of Early Warning and Proactive Control Systems in Food Supply Chain Networks[5]

Abstract

Managers in Food Supply Chain Networks often encounter quality problems in production. In order to diminish hazards and losses, it is important to obtain warnings about such problems as early as possible, and to control them proactively. Recorded data from monitoring systems is a valuable source of knowledge for realizing such early warning and proactive control. We designed a system to guide managers, as non-experts in Data Mining, in analyzing recorded data. The system contains an Expert System for Data Mining method selection and template approaches for applying Data Mining methods. With this system, managers can explore causes for encountered problems, predict upcoming problems, and support corrective actions.

Keywords: expert system, Data Mining, early warning, proactive control, process support

5.1 Motivation

In Food Supply Chain Networks (FSCN) it is crucial to deal with quality deviations, not only because of losses they might cause, but also due to potential health threats. To actively predict and prevent quality deviations regarding food products in FSCN, early warning and proactive control systems are needed. Such systems can detect problems as soon as possible, and predict problems that are about to occur. Consequently, managers can react as early as possible to problems in FSCN.

Variability in quality of primary inputs, uncertainty of influential factors, and complexity of network structure make it difficult to manage problems in FSCN. Currently, information systems in FSCN provide managers with potential data resources for solving those problems. Additionally, research in Data Mining (DM) generates versatile methods for various tasks, such as factor selection, causal modelling, and prediction. By applying DM methods to analyze available data sets, managers can explore causes for encountered problems, predict upcoming problems, and experiment with different remedies to counteract potential hazards or losses. However, since managers are usually non-experts in DM, they need guidance on applying DM methods. Unfortunately, systems to supply such guidance are not available yet.

In this chapter, we outline a prototype system to guide managers in FSCN through the steps of correctly configuring the DM process and executing it. The functions involved are factor selection, exploring causal factors, and problem prediction. Successfully accomplishing these functions enables managers to proactively predict and prevent problems in FSCN. The main components of

[5] Li, Y., M. R. Kramer, A. J. M. Beulens and J. G. A. J. v. d. Vorst (2008). Guiding Users of Early Warning and Proactive Control Systems in Food Supply Chain Networks. Proceedings of 8th International Conference on Management in AgriFood Chains and Networks, Ede, The Netherlands.

this prototype system are an Expert System for DM method selection and template approaches for various steps in applying DM.

5.2 Context

5.2.1 Early Warning and Proactive Control

Food Supply Chain Networks are featured with complex structures and multiple stages. In each stage, there are various operational and environmental factors involved. Due to uncertainties in those factors and variations in product quality, it is a common feature of FSCN that various performance deviations occur occasionally.

Managers in FSCN, especially operational managers, are busy with various types of tasks (e.g. operations, scheduling, administration) everyday, and have little time for problem investigation. Their time is highly fragmented, so the time allocated for each task is limited (Mintzberg 1973). Although they have both domain knowledge on the problem area and the ability to control operations in FSCN, they need assistance facilities to investigate encountered problems within reasonable time.

In our research we design systems to help managers to deal with problems in FSCN. Early warning and proactive control systems in FSCN are knowledge-based, data- and model-driven decision support systems that are designed for managers to predict and prevent problems associated with food products in FSCN (Li, Kramer et al. 2006a). They enable managers in FSCN to use DM methods for analyzing existing data sets.

Such data analysis includes several phases. One phase is to explore determinant factors for encountered deviations and to build models that describe relations between deviations in FSCN and determinant factors. Another phase is to use monitored values of those determinant factors to predict upcoming problems as early as possible. After analysis, managers have to evaluate different control actions aimed at preventing problems. The potential actions could be discarding products, taking corrective measures, or adapting succeeding processes in order to make amends. Early warning and proactive control systems also provide facilities for managers to easily incorporate obtained knowledge and to quickly browse existing knowledge in such systems. The knowledge obtained by managers through data analysis is beneficial for other users with similar problems.

We will use a chicken supply chain case to illustrate the concept. This chicken supply chain has various stages, from hatchery to slaughter house. The monitoring systems keep recorded data on properties of chickens and various factors (operational, environmental, etc.) covering a period of several months. In this FSCN, there was a problem that too many chickens arrived dead at the slaughter house. So this problem is named Death On Arrival (DOA). Through applying DM methods on recorded data, managers can build causal models to explore determinant factors for DOA. Such models will supply a warning signal about upcoming DOA based on status of determinant factors. Managers can also employ causal models to evaluate different counteractive measures. The knowledge obtained by managers can be stored into a knowledge base for reference by other managers.

Figure 5-1 shows the processes for Early Warning and Proactive Control (EW&PC) in FSCN. We use the DOA case to illustrate how managers follow these processes to accomplish EW&PC. The three processes that need support on DM are marked in grey.

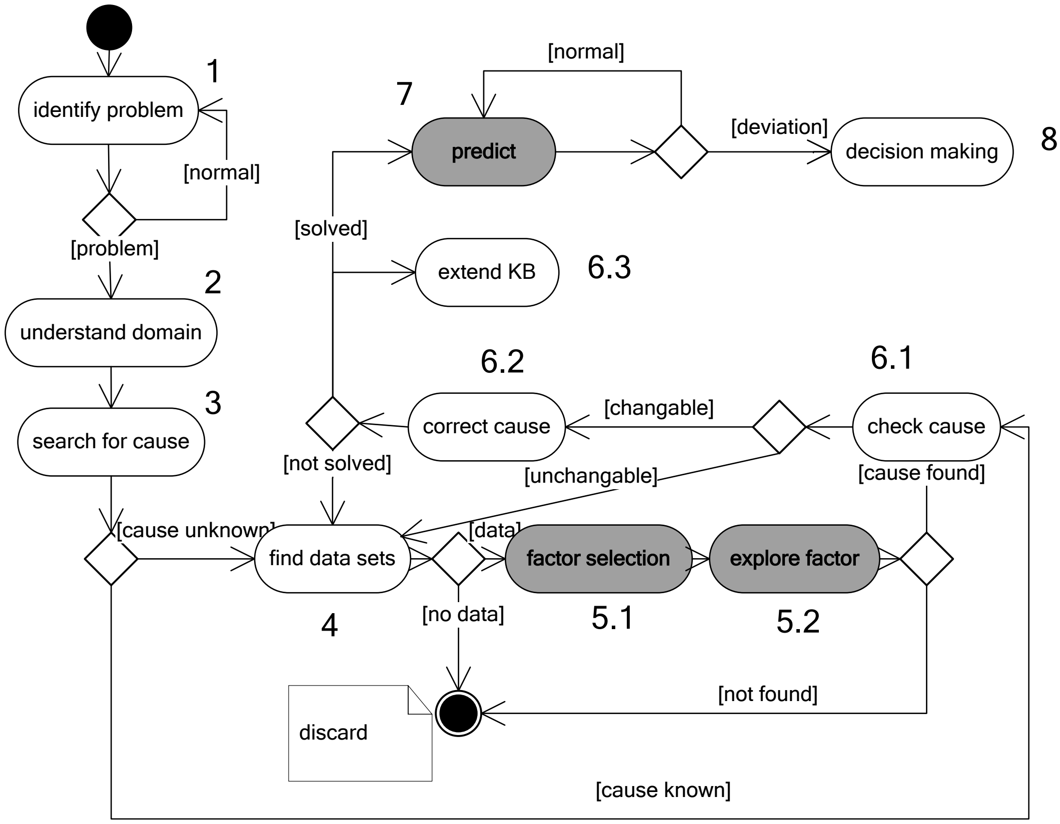


Figure 5-1: processes for early warning and proactive control in FSCN; processes marked in grey need support on DM

The whole procedure starts when a manager discovers (or is informed about) a problem in his FSCN and formulates it in a quantitative way. For example, in the DOA case the manager formulates DOA for some flock of chickens as "the death percentage upon arrival at the slaughter house is larger than 0.5%". Then the manager checks and complements his knowledge on each stage of the FSCN, as well as the kinds of operations and resources used in each stage. After that, the manager searches the Knowledge Base for existing knowledge on possible causes of the problem. If such information is available, the manager can go to step 6.1 to verify the found cause. Otherwise, he has to find relevant data sets and combine them for quantitative analysis. In the DOA case, the manager had to combine breed data from the hatcheries, transport data from the delivery process, and data on DOA and rejection obtained from the slaughter house.

The quantitative analysis starts with factor selection, which means that the user makes a preliminarily selection of a limited amount of factors that might have influence on the problem. In the DOA case, the manager can use DM methods to select about ten variables that might influence DOA, such as catch method, breed, etc. After factor selection, the manager uses DM methods to explore potential causal factors for the problem. For example, in the DOA case, one of the DM methods, Decision Trees, indicated that the factor ‘Transport time’ is a potential causal factor for one particular breed of chickens ‘Cobb’.

To check the correctness of hypothesized causal factors, the manager has to change them in practice and observe the results. For example, in the DOA case, the manager can change the transport time for ‘Cobb’ chickens. If observed results confirm the validity of the hypothesized causal factor, the manager can be warned about upcoming problems before their occurrence by monitoring those causal factors, e.g. transport time for ‘Cobb’ chicken in the DOA case. Timely recognition of potential problems enables the manager to take proactive actions that diminish potential losses. He can also insert the relation between ‘Transport time’, ‘Cobb’ and DOA into the Knowledge Base for later reference in similar cases.

5.2.2 Data Mining

Data Mining is a powerful technique to extract valid, previously unknown, comprehensible and actionable information from databases and use it to make crucial business decisions (Simoudis 1996). Applying DM methods involves multiple steps, as shown in Figure 5-2 as arrow-linked white boxes. They are generalized from literature in Knowledge Discovery in Database (KDD) process models (Fayyad, Piatesky-Shapiro et al. 1996; Kurgan and Musilek 2006). The grey boxes in Figure 5-2 indicate where we support the process.

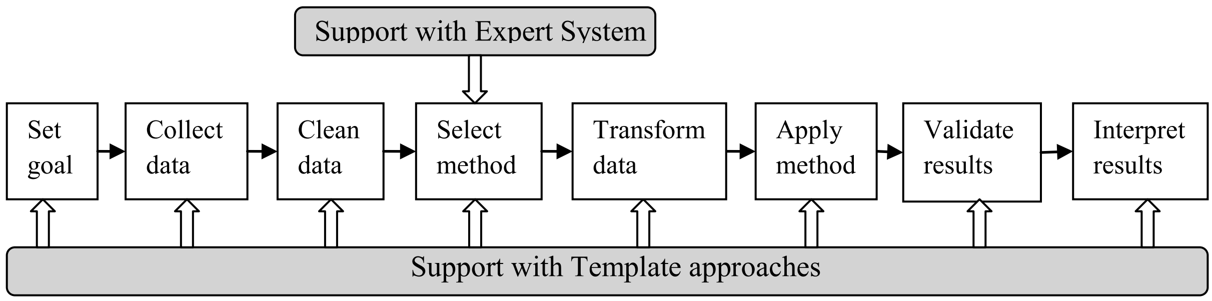


Figure 5-2: Relation between expert system, template approaches, and KDD process

The KDD process starts with learning the goals and collecting relevant data, then data need to be cleaned to get rid of noise, outliers, etc. After that, an appropriate DM method has to be found and if necessary, some transformation of data format may be needed. Then set proper parameters and apply the selected DM method. Obtained results need to be validated before they can be interpreted for practical use. These steps are iterative in the sense that when certain steps get invalid results, some previous steps need to be redone with modified settings (i.e. models or model parameters). In order to ensure the accuracy and reliability of modeling results, it is necessary to provide scientific and technical guidance on how to carry out various steps in the modeling work (Scholten, Kassahun et al. 2007). Figure 5-1 shows that in EW&PC system there are three processes (Factor selection, Explore factor, Prediction) that need DM. So it is necessary to guide managers, as non-experts in DM, to quickly and correctly select appropriate DM methods and use them to analyze existing data.

As we explained in (Li, Kramer et al. 2006b), before building systems to facilitate DM usage in a certain area, it is essential to analyze two topics first. One is the set of functional and nonfunctional requirements on DM methods and applicability of various DM methods – what DM methods are applicable for solving those problems. The other covers the generic steps – how to use certain DM methods to solve practical problems. Next, we review current research on each of those two topics.

Method selection support – what to use?

Managers in FSCN normally do not have knowledge on Data Mining. Therefore, they need assistance in determining which DM methods to use. Because their time is generally fragmented and limited, the required assistance should be simple and easy to use. Researchers in DM have been continuously working on automating method selection – how to select the most appropriate DM method for users. In literature we found one actual expert system for Machine Learning method selection: CONSULTANT (Craw, Sleeman et al. 1992). This system cover 10 methods, most of which are about First Order Logic (FOL). However, as shown in our previous work (Li, Kramer et al. 2006b), in order to guide managers, EW&PC systems in FSCN have specific requirements, not only on function, but also on representation form and extendibility of DM methods. For example, EW&PC systems require DM methods to be able to identify deviations, to explore causal factors for encountered deviations, and to predict upcoming problems based on status of causal factors. But in CONSULTANT, following common practice in conventional DM areas, functions are categorized into classification, clustering etc. Such categorization is meaningless to managers in FSCN, because the categorization is too technical in nature and not related to the problem domain. As a result, we have to use our own selection criteria to select applicable DM methods for early warning and proactive control, and to model these methods.

Besides this physical (developed) system, there is much research on the way of method selection, such as meta-learning (Alexandros and Melanie 2001), landmarking (Pfahringer, Bensusan et al. 2000), and guarded method selection (Verdenius 2005). A common drawback of these approaches is that none of them provides a systematic mapping (or relation) between the characteristics of data sets and various DM methods.

Usage support – how to use?

Next to method selection, managers also need to know how to use those methods in practice. Currently there are research efforts to design systems that help users accomplishing the KDD process. Wirth et al. (1997) designed a system that contains a user-guidance module for DM processes. This module guides users through a stepwise refinement of a high-level DM process in order to help users construct the best plan. The obtained plan is compiled into scripts for execution. Bernstein et al. (2005) claim that even with well specified goals it is very difficult to discern the one best plan, so they designed a system to rank various valid plans. Users can combine their objectives, background knowledge, etc. to select the most appropriate plan.

However, we can not use such systems directly in EW&PC. Those systems are in general not built to be used by non-experts in DM. They assume that users already have considerable knowledge on different DM methods. Some guidance that is needed by non-experts in DM can not be found in those systems. An example is outlier handling, i.e. removing or modifying abnormal values from recorded data, because those abnormal values might impair the quality of data analysis.

In our research, we design template approaches to provide managers with support on DM usage. Figure 5-2 shows the relations between the Expert System for methods selection, Template approaches, and DM usage processes. The Expert System supports managers in selecting appropriate DM methods. The template approaches support managers in various steps of applying DM methods. Although there are research results and even systems on each of these two aspects, as far as we know, there is no system that endeavors in both of these two branches yet. The prototype system in this chapter is a first step to combine the power of these two aspects in order to provide managers, as non-expert in DM, with comprehensive and easy-to-follow support on employing appropriate DM methods to solve problems in FSCN.

5.3 Designing Template approaches and Expert System

5.3.1 Template approaches

We design template approaches based on literature in KDD processes and case studies. Literature in KDD provides general guidelines on the processes in using DM methods for knowledge discovery (Fayyad, Piatesky-Shapiro et al. 1996; Kurgan and Musilek 2006). Case studies in FSCN provide specific knowledge resources for the design of template approaches. During case studies, we kept records of all the steps that we took in applying DM methods on recorded data. Then we organized those records according to the three processes discussed in the previous section. After that we generalized from those records three template approaches for factor selection, exploring causal factors, and prediction, respectively.

From the general principles of the KDD process and similarities in the inferred template approaches, we generalized a generic process for applying DM methods for early warning and proactive control depicted in Figure 5-3. This generic process needs to be validated in further case studies.

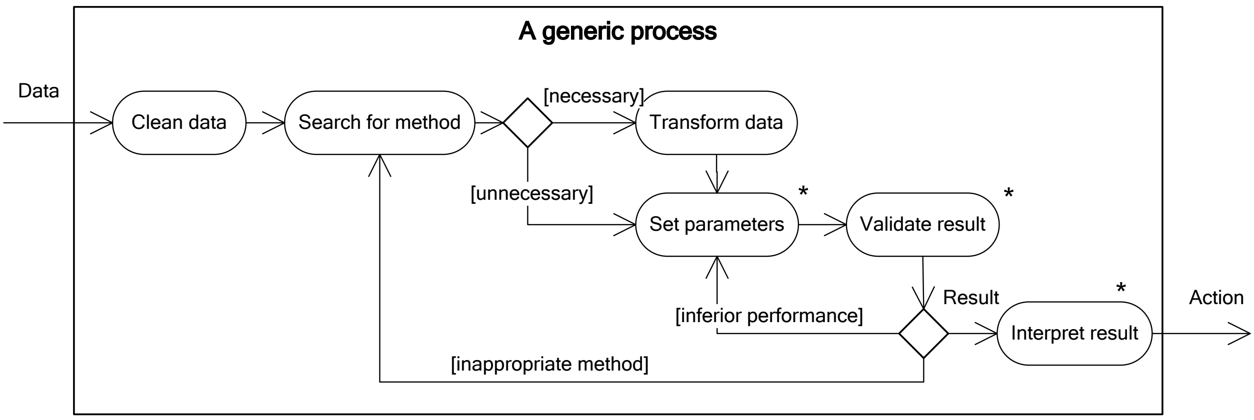


Figure 5-3: a generic process for applying DM methods in EW&PC systems; the steps marked with a star are method-specific

The generic process starts with the incoming data set. First, one has to clean the data set in order to get rid of missing values, outliers, etc. Then some suitable DM method should be selected. If necessary, data should be transformed into a proper format for the selected method. Before actually applying the DM method, it is necessary to set correct parameters for arriving at reliable results. Results from application of the DM method have to be validated. If they are invalid, there could be several reasons. If parameters have been set inappropriately, we should adjust the parameters and run the method again. If an inappropriate DM method has been selected, we can try other DM methods. As a last step, we interpret validated results to arrive at actionable knowledge.

Among those steps, there are three method-specific steps, as shown with a star in Figure 5-3. This implies that, for each supported method, there should be comprehensive and detailed information on how to set parameters, validate and interpret results.

5.3.2 Expert System for method selection

In order to make the EW&PC system useable by non-experts in DM, the necessary knowledge about decision making in method selection has to be put into the Expert System. To start the construction process for the Expert System, we reviewed literature on applicable quantitative methods for EW&PC as well as on existing research in method selection. Next, we represent the obtained knowledge with various formal models provided by the knowledge management and modelling methodology CommonKADS (Schreiber, Akkermans et al. 2000). These models serve as a bridge between conceptual design and physical implementation. After knowledge specification, we implement the system with a Graphical User Interface (GUI). The two sub-sections below provide details on these steps.

Knowledge acquisition and specification

From literature on Machine Learning, Data Mining, and statistics, we find methods that have the potential to contribute to certain functions for EW&PC. After experimenting in cases in FSCN, we come to a list of methods for Expert System as shown in Table 5-1.

Table 5-1: DM methods incorporated in ES for method selection

	DM method	Reference
1	Decision Trees	Quinlan (1992)
2	Neural Networks	Haykin (1999)
3	Bayesian Networks	Cooper & Herskovits (1993)
4	Nearest Neighbours	Aha (2001)
5	PC	Spirtes et al. (2000)
6	MIM	Edwards (2000)
7	Minimum Description Length	Hansen & Yu (2001)
8	Principal Component Analysis	Lattin et al. (2003)
9	CATPCA	Meulman et al. (2004)

As discussed in section 5.2, current research on categorizations of DM method properties is not suitable for EW&PC. In our research (Li, Kramer et al. 2006b) we derived five criteria for DM method selection: functionality; format of dependent variables; format of independent; kind of models; extendability. A suitable method will be suggested to users only when it satisfies all five criteria.

The first criterion to be considered for DM method selection is the functionality. There are several functional requirements on DM methods from EW&PC system: *problem identification, causal factor exploration,* and *prediction*. Different DM methods can be used for fulfilling different functional requirements (Li, Kramer et al. 2006b). For example, Neural Networks can be used for pattern recognition as well as prediction, but not for finding causal factors. It is necessary to specify the functions of each DM method in EW&PC systems in order to choose appropriate DM methods for certain functional requirements.

There are two criteria on DM methods regarding the models they generate. Firstly, models generated by different DM methods have different forms. For example, models built with 'Bayesian Networks' have a network form, while models built with 'Decision Trees' are in the form of a tree. If the manager knows beforehand that the relation between certain factors is in certain form, he can choose a method that can generate that model form. Secondly, models are different in their ability to adapt to changes. This causes DM methods to be different in extendibility to new knowledge. For example, it is easier to extend models built with 'Bayesian Networks' than those built with 'Neural Networks'.

There are also criteria relating to the characteristics of data sets: the format of dependent and independent variables. A dependent variable can take the format of binary, nominal, ordinal, numerical, or ratio. It is necessary to choose methods that are able to handle the format of dependent variables. For independent variables we encounter the same kinds of formats. However, since independent variables are treated as a group, we only distinguish them into categorical (if all of them are binary, nominal or ordinal), numerical (if all of them are numerical or ratio), and mixture (otherwise). Some DM methods can only handle independent variables of the numerical group, such 'PCA'. Some other DM methods can handle mixture types of variables, such as 'Decision Trees'.

System implementation

The system is composed of five major components: *User Interface*, *Inference Mechanism*, *Suggestion Tool*, *Domain Knowledge*, and *Blackboard*. The user of the system enters the case specification through the *User Interface* component. Generated results are sent back to the *User Interface* for the user to choose from. The *Inference mechanism* searches for suitable DM methods according to the entered case specification. If no suitable method has been found, the *Suggestion Tool* looks for possible transformations that can yield applicable DM methods. Both the *Inference mechanism* and the *Suggestion Tool* inspect the *Domain Knowledge* for facts and rules about different DM methods and their functions, representation forms, etc. The *Blackboard* is used to store intermediate outcomes during the working of the other components.

To use the Expert System for DM method selection, the user inputs his requirements and properties of data sets by choosing from a set of predefined options. Figure 5-4 shows a prototype implementation of the *User Interface*. The text pane at the right side of the screen dynamically provides explanations to the options in each question. The system selects all applicable DM methods from the DM library and uses criteria to check its appropriateness for the entered choices. If a DM method meets all criteria, then the system displays a description of the method in the text box at the bottom left corner. If no suitable method can be found, the system will use the same textbox to ask the user to change the settings for a new search. So this text box is dynamic in nature. If the user does not know what to change, he can ask the system for a list of suggestions, and then select an appropriate one for his case and try it out. The suggestions are also shown in the text box at the bottom left of the screen.

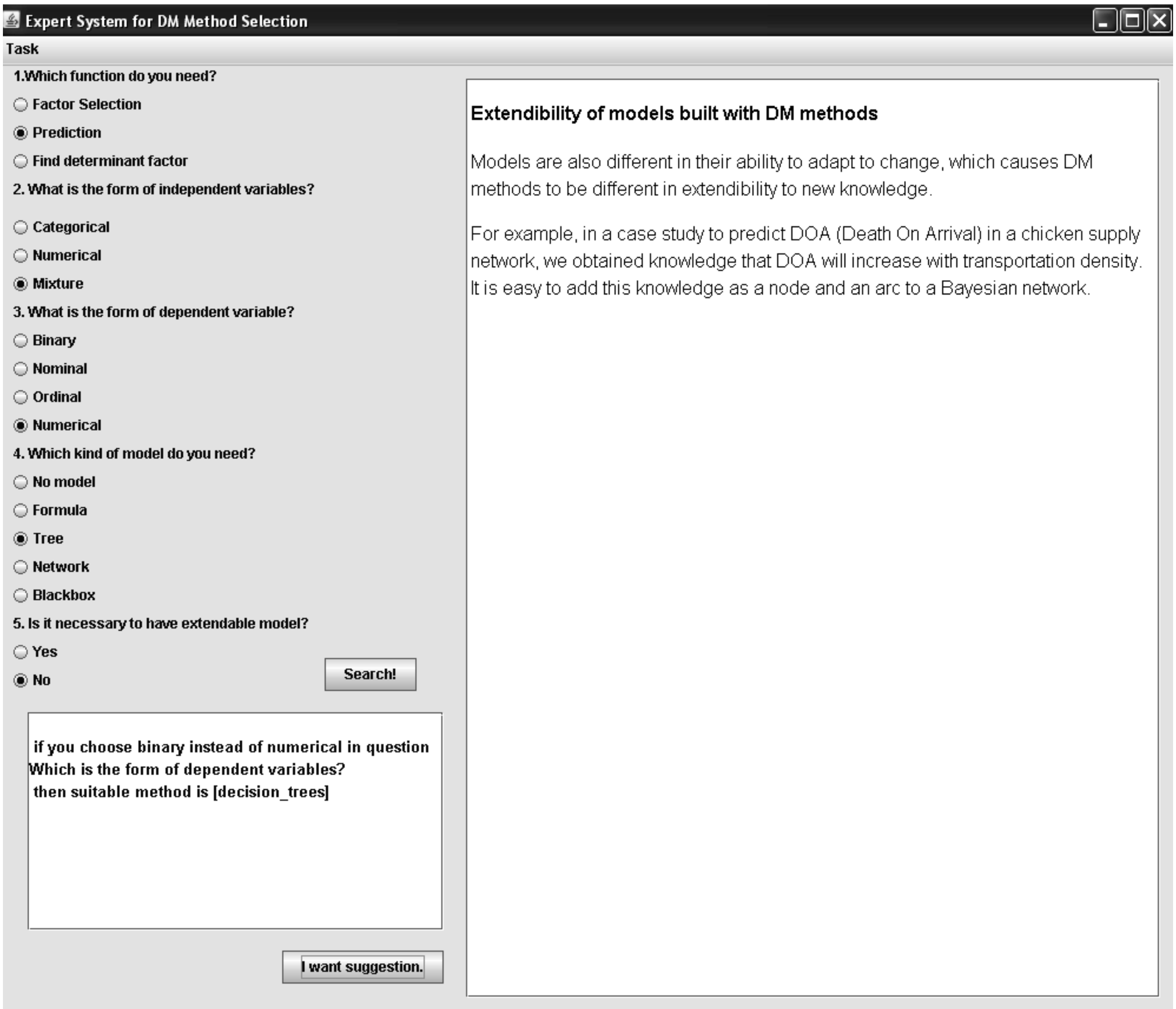


Figure 5-4: prototype User Interface of Expert System for method selection

5.4 Usage of combined Expert System and Template approaches

We designed a prototype system that incorporates both the template approaches and the expert system for method selection. Figure 5-5 gives an impression of how the system operates.

The generic process for using DM methods in EW&PC systems (as in Figure 5-3) is shown in the bottom part the user interface. It shows the procedure for using this tool. Each step of the generic process is facilitated by a separate tabbed page in the software ('Clean Data' has been split into 'Missing Values' and 'Outliers'):

1. Browse Data

This system allows managers to access databases from both the local machine and networks. Managers can inspect parts of data sets to verify his choice.

2. Missing Values

A template approach guides managers in choosing appropriate strategies to deal with records with missing values. They can either remove those records or fill in the missing cells, either based on domain knowledge or by applying built-in estimation methods.

3. Outliers

The system guesses the formats for the variables in the data set. After managers have confirmed or overridden these suggestions, suspected outliers are shown in red in a table of all data. Managers may choose different strategies to deal with outliers, either delete such records, or change values according to domain knowledge. The template approach in the right pane helps to carry out this task.

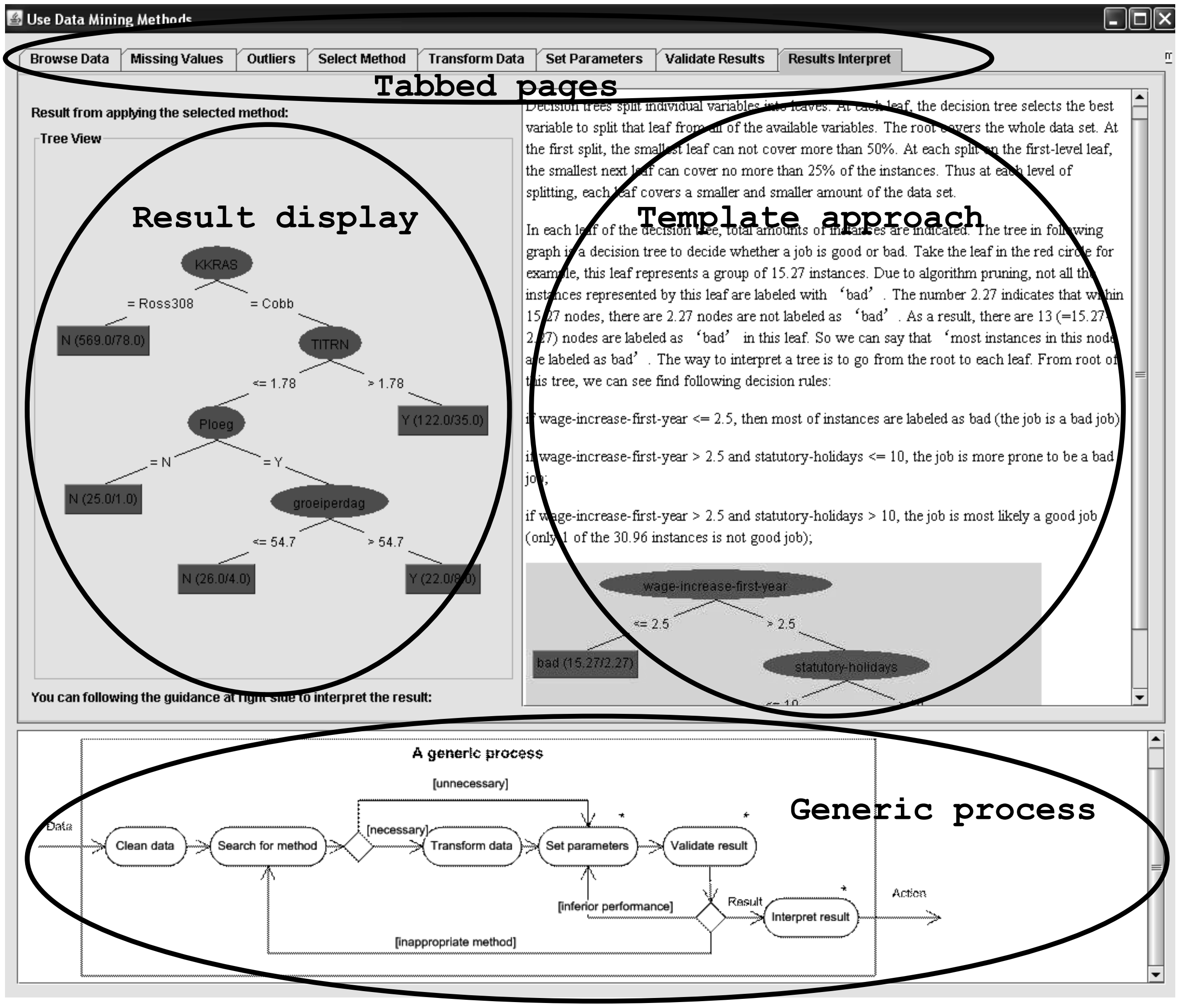


Figure 5-5: user interface of the prototype system

4. Select Method
Managers can use the Expert System for method selection (as shown in Figure 5-4) to select an appropriate DM method for the selected data set.

5. Transform Data
If the selected DM method needs a different format for some data, a template approach guides managers to transform those data to an appropriate format.

6. Set Parameters
A template approach directs Managers set appropriated parameters before running the algorithm on their data set.

7. Validate Results
Results obtained from running the DM method need to be evaluated for its validity and reliability. If the template approach finds results that seem invalid or unreliable, it makes suggestions on changes in previous steps.

8. Interpret Results
Managers have to interpret the results obtained from applying DM methods. As shown in Figure 5-5, the template approach at the right side of this page explains how to interpret the obtained decision tree.

The generic process, together with template approaches at the right side of each page, provides managers with guidance on each task in Data Mining.

5.5 Conclusion

Early Warning and Proactive Control systems intend to enable managers to predict and prevent problems. We presented the design of a system for guiding users for EW&PC in FSCN, and a prototype implementation. This system contains an Expert System for DM method selection, and template approaches for various steps in the KDD process. With these two components, this system enables managers to use appropriate exploratory methods to identify relations among food quality problems and potential influential factors (such as operational factors, environmental factors). When managers need to predict upcoming problems based on the information from monitoring systems in FSCN, the system helps selecting appropriate prediction methods and guides managers in applying those methods. Further research will focus on expert validation of this system.

To use such EW&PC systems effectively, managers have to deal with the problem of combining data sets. Our focus in this chapter is on supporting DM method application for factor selection, casual factor exploration, and prediction. However, as indicated in Figure 5-1, before managers analyze data, they have to find relevant data sets and if necessary, combine them for analysis. It is not always clear how to combine data sets. Data in different data sets that look similar could have different semantics precluding them to be joined directly. Sometimes additional domain knowledge (e.g. business domain, food production processes, data management) is needed to know whether data sets can be combined without problems. Due to the broad categories of discrepancies that can prevent data set combination, further research for practical guidance on this aspect is needed. A promising strategy is to employ Ontology Engineering for this purpose.

Chapter 6. Applying Knowledge Engineering and Ontology Engineering to construct a Knowledge Base for Early Warning and Proactive Control[6]

Abstract

Early Warning and Proactive Control systems intend to help managers predict and prevent problems in Food Supply Chain Networks. Since those problems are of various types and have broad scopes, an easy-to-access and easy-to-extend Knowledge Base is needed. This chapter presents a structure for such Knowledge Base. The structure is designed with combined techniques from Ontology Engineering and Knowledge Engineering. It contains an ontology and a rule base with inference mechanism. A prototype system, filled with information from case studies, shows that the designed structure supports consulting the Knowledge Base on relevant knowledge for encountered problems, and extending the Knowledge Base with obtained novel knowledge.

Keywords: knowledge based systems, ontology, rule base, food supply chain networks

6.1 Introduction

In Food Supply Chain Networks (FSCN), due to complexity in structure, variability in food quality and production, and uncertainty in influential factors, managers are facing various types of problems covering multiple network stages. Moreover, causes for those problems may originate from various sources, such as operations, environment, or product inherent quality. It is impossible for any single manager to possess knowledge of all types of problems in FSCN. Furthermore, new problems and new causes for problems occur frequently. This requires managers to have well-established procedures and facilities for problem investigation. Since problems in FSCN might cause hazards to human health and considerable losses to food industry and shareholders, those procedures and facilities should be efficient and effective.

In our research we aim at developing Early Warning and Proactive Control (EW&PC) systems to help managers in FSCN to predict, prevent, and correct problems. An important additional function of EW&PC systems is to support managers sharing obtained knowledge from data analysis, and using knowledge added by other managers. By comparing his own case with similar cases from FSCN, a manager can quickly locate potential causes and counter-measures for encountered problems. Whenever a manager encounters a problem for which no previous knowledge exists, the EW&PC system helps the manager to investigate the problem and store the obtained knowledge for later reference.

[6] Li, Y., M. R. Kramer, A. J. M. Beulens and J. G. A. J. v. d. Vorst (2008). Applying Knowledge Engineering and Ontology Engineering to construct a Knowledge Base for Early Warning and Proactive Control. Proceedings of 2nd International Conference on Knowledge Generation, Communication and Management, Orlando, USA.

To meet the requirements of knowledge sharing, a Knowledge Base (KB) for problems in FSCN is indispensible (Klein and Methlie 1995). A KB accommodates obtained knowledge from managers and enables managers to retrieve relevant knowledge for their cases efficiently. Due to various types of problems and variability of causes of problems, extensive knowledge will be put into such KB. Therefore, the KB should be designed in an easy-to-access and easy-to-extend way.

We apply results from current research and development in Ontology Engineering and Knowledge Management to build a structure for our KB. We separate knowledge we obtained in case studies into two types, namely manifest knowledge and inferred knowledge. Based on such separation, we build an ontology to model manifest knowledge, and construct a Knowledge-Based System to capture inferred knowledge. Furthermore, a prototype Knowledge Base has been built to link the Knowledge-Based System with ontology. Example results show that this structure is able to help managers. When managers come across problems in FSCN, the system helps them to specify the problem and obtain knowledge on causes and remedies. Managers can also quickly and correctly extend the KB when they obtain novel knowledge.

6.2 Context

This section first explains the concept of Early Warning and Proactive Control systems in FSCN, introducing a Knowledge Base as an essential component. Then we review existing research and development in areas that contribute to designing the KB: Knowledge Engineering and Ontology Engineering.

6.2.1 Early Warning and Proactive Control Systems

Early Warning and Proactive Control systems in FSCN are knowledge-based, data- and model-driven decision support systems that are designed for managers to predict and prevent problems associated with food products in FSCN. Managers in FSCN use EW&PC systems to deal with encountered problems by analyzing existing data sets. With quantitative methods managers explore causes for encounter problems, predict upcoming problems based on monitored status of those causal factors, and at the end evaluate different control actions to prevent problems, e.g. discarding products, taking corrective measures, or adapting succeeding processes in order to make amends. The knowledge obtained through data analysis is beneficial for other users that have similar problems. Therefore, EW&PC systems contain a Knowledge Base. Knowledge obtained during data analysis can be easily incorporated into the Knowledge Base. Users can share their knowledge through this KB.

Figure 6-1 shows a framework we designed for early warning and proactive control systems in FSCN (Li, Kramer et al. 2006b).

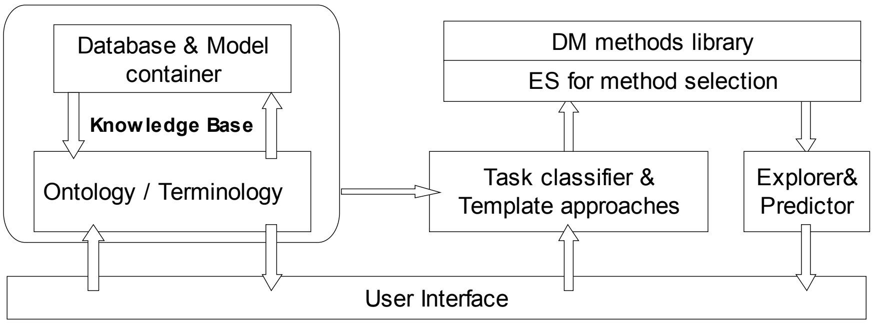


Figure 6-1: framework for early warning and proactive control systems in FSCN

Users of such a system will be guided by *Task classifiers* to search for applicable causal relations from the *Knowledge Base*. If no applicable relation is found, users can choose appropriate DM methods to explore causal relation from recorded data. *Template approaches* will guide users in each step of applying DM methods. The *Expert System (ES) for method selection* suggests applicable DM methods from *DM library*. After obtaining causal relations, users can predict potential problems and consider different control measures to reduce or prevent losses due to problems in FSCN. Users can also store obtained knowledge about causal relations and control measures into the *Knowledge Base* to make the knowledge available for other users with similar problems in FSCN. In order to provide users with relevant knowledge, the *Knowledge Base* should enable users to browse, navigate through the KB, and compare their cases against the KB. Correctness of stored knowledge should be guaranteed, so the *Knowledge Base* has to store obtained knowledge correctly and in a comprehensible way, has to ensure that entered knowledge truly represents what users intend, and has to prevent typing mistakes and inconsistencies in naming.

However, two problems have to be dealt with in building a KB. First, the databases being analyzed by different managers are normally collected in different FSCN contexts (e.g. with different processes and different operations involved). Even if they refer to the same object in FSCN, the exact meaning of variables in different databases could be totally different. It is difficult for one manager to correctly understand semantics of databases that come from other managers. Second, the models built by managers through data analysis could have different forms. Those different model forms make it difficult for managers to quickly understand and locate relevant models for encountered problem. This problem is further complicated by the first problem that variables used by one manager in modeling might not be easily understood by others.

Those difficulties impede managers to share understanding of knowledge in a KB, and hence weaken their potential for knowledge storing and knowledge accessing. In order to tackle those problems, we employ techniques from Knowledge Engineering and Ontology Engineering.

6.2.2 Knowledge Engineering

Knowledge Engineering aims at modelling various aspects of domain experts' problem solving expertise, and hence producing Knowledge-Based Systems (KBS) to help non-experts in dealing with problems (Schreiber, Akkermans et al. 2000). It intends to model not only the domain knowledge of experts, but also the inference procedures applied by those experts in problem solving. Knowledge Engineering has some overlap with Knowledge Management, which comprises of "a set of practices used by organizations to identify, select, organize, disseminate, and transfer important information and expertise that are part of the organization's memory and that typically reside within the organization in an unstructured manner" (Turban, Leidner et al. 2008). Current research and development in Knowledge Engineering yields various methodologies to build knowledge-based systems, of which we use CommonKADS (Schreiber, Akkermans et al. 2000). Knowledge Engineering techniques have been applied in various areas, such as hospital management (Moreno, Aguilar et al. 2001), weight distribution on ferryboat (Shaalan, Rizk et al. 2004), and classifier design (Sigut, Pineiro et al. 2007).

6.2.3 Ontology Engineering

Ontology Engineering deals with organizing domain knowledge by formal descriptions of the concepts with their properties and relations. The aim is to improve knowledge sharing by structuring the concepts according to human understanding (Gruber 1995; Uschold and Gruninger 1996; Borst 1997). The resulting structures are called ontologies. As defined by Neches et al. (1991), "An ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary." With ontologies we can reuse knowledge in a knowledge base. The ontology in a knowledge base plays the role of its backbone (Mizoguchi and Ikeda 1996). When people use a knowledge base for problem solving, they share assumptions and requirements on the problem domain. An ontology captures the underlying conceptualization of those assumptions and requirements.

A variety of activities of Knowledge Management can be supported by ontologies, such as knowledge retrieval, storing, sharing, and dissemination. There are many methodologies, tools, and environment for Ontology Engineering. However, currently there is no definition and standardization of methodologies that drive the development of ontologies (Gómez-Pérez 1994). For a comprehensive comparison of those methodologies and tools, please refer to Corcho et al. (2003).

6.3 Structure Development

6.3.1 Case studies

The knowledge base we build is based on three case studies in FSCN.

DOA case

This case is about a chicken supply chain. This supply chain has various stages, from hatchery to slaughter house. The monitoring system in this supply chain records data associated with various factors (operational, environmental, etc.) and properties of chickens in this supply chain. When we started the case study, there was a problem in this supply chain that too many chickens arrived dead at the slaughter house. This problem is called Death On Arrival (DOA). Managers involved in the supply chain were not sure about the cause of this problem, particularly because the level of DOA varied considerably between flocks and in time. After applying DM methods to analyze recorded data, relations were found between genetic factors, transport time, and DOA (Li, Kramer et al.

2006a). When the knowledge base is operational, knowledge obtained can be stored for later reference.

Management farm case

This case comes from a pig supply chain of a Dutch food company. In each stage of the breeding chain, the company recorded information about various attributes of pigs and the way they were managed. Slaughter house data of the pigs are also recorded, such as their meat percentage, muscle thickness, etc. For research purposes such performance indicators were actually recorded at an individual level. Through analyzing recorded data, we identified relations between inherent attributes of pigs (e.g. sex, genetics), operational factors, and performance indicators. Those relations contribute ample resources for ontology construction.

MAF case

This is a second pig supply chain case. Contrary to the previous case, this one concentrates on feeding regimes. Most data relate to the fattening farm and aggregated slaughter data. In each of the growth stages, the farms recorded information about feed and genetics. Through analyzing those recorded data, we identified relations between feeding regimes, genetics and meat quality. These relations supplement relations found in the second case.

6.3.2 Manifest knowledge and inferred knowledge

When organizing the Knowledge Base for EW&PC, we find that knowledge consists of two distinguishable kinds of knowledge. On the one hand there is domain knowledge on various stages in FSCN, involved entities, and their attributes and relations. On the other hand, there is knowledge on the influence between those stages, entities, and attributes. Such knowledge is normally obtained from either data analysis or investigation.

Therefore, we categorize the knowledge in the knowledge base into two types: manifest knowledge and inferred knowledge. Manifest knowledge represents the objects and their properties (e.g. about operations, staff, production means, environmental indicators, and performance indicators) and those relations between them that people can perceive without need for data analysis or inference. For example, there is a relation between ‘Chicken’ and ‘Genetics’ indicating the breed of chickens. Different people share the same perception upon such relations. This kind of knowledge is quite stable in the sense that it is not subject to influence from outside factors.

Unlike manifest knowledge, to identify inferred knowledge, people have to analyze data or employ domain experts. For example, the knowledge that ‘long transport time causes more DOA to chickens of breed Cobb’ is derived from data analysis. Inferred knowledge is a result of and contributes to the process of problem solving. Due to the fact that different data analysis methods may yield different outcomes, people may have varied perceptions of inferred knowledge. Such knowledge may need adaptation when it occurs in a different situation.

Ontology Engineering is suitable for modeling manifest knowledge. The predefined terms in an ontology (e.g. class, instance, properties) provide a way to capture features of objects and static relations between them. The class hierarchy in an ontology makes it possible to model the structure associated with semantics of databases in FSCN. With an ontology, users can share understanding of entities in FSCN during consulting or extending the *Knowledge Base*. They can correctly navigate to the relevant knowledge in the *Knowledge Base*. But ontologies are not well suited for representing inferred knowledge. It is difficult for different people to reach consensus on inferred knowledge. Moreover, inferred knowledge is more likely to change than manifest knowledge.

In stead of using an ontology, we model inferred knowledge with facilities provided by Knowledge Engineering. In systems built with Knowledge Engineering techniques, one of the essential components is a rule base. Rules in the rule base capture expert knowledge in problem solving. Since normally that kind of knowledge is obtained indirectly (e.g. from data analysis or expert interviews), rules provide a more mature mechanism to model inferred knowledge. Modification of rules will not harm the consistency of a built ontology. Furthermore, rules may contribute to extra functionality, such as inference for problem investigation.

As a result, we first employ Ontology Engineering techniques to build an ontology that describes manifest knowledge in the Knowledge Base. The ontology captures such aspects as the classification of various stages in FSCN and relations between different processing stages. Then we use Knowledge Engineering techniques to model inferred knowledge. Figure 6-2 gives an example of this idea. 'Chicken' and 'Transport' are two objects in the ontology for FSCN. From data analysis and expert interviewing, we found a rule 'transport time is longer then 2.4 hours and chickens have genetics Cobb is a cause for DOA rate larger than 0.5%'. This rule is based on the objects in the ontology. It describes the inferred relations between the two objects 'Chicken' and 'Transport'.

Modeling manifest knowledge and inferred knowledge with Ontology Engineering and Knowledge Engineering techniques separately has a number of advantages. Firstly, knowledge is easy to access. The rule base makes it possible to use inference mechanisms to search all relevant rules for causes that explain some observed symptom. Such causal knowledge contributes to what is called 'diagnose' type of applications in CommonKADS. The existence of different types of rules enables a reasoning process containing multiple steps for problem solving, thereby providing users with more extensive knowledge. The ontology, which is the foundation for the rule base, prevents users to misunderstand the rules in rule base, and enables them to navigate through the class hierarchy for exploring relevant knowledge.

Secondly, the separation of manifest knowledge and inferred knowledge facilitates Knowledge Base designing. The clear distinction between knowledge of different types helps designers in Knowledge Modeling. Also, designers can take advantage of developments in both Ontology Engineering and Knowledge Engineering to ease the design process. In addition, such a separation makes it possible for designers to build a rule base using existing ontologies.

Thirdly, such a structure makes a Knowledge Base easy to extend and maintain. When users have generated different rules, they can incorporate them into a rule base instead of changing the ontology. The ontology helps users to find correct terms that constitute target rules. Therefore, added rules can truly reflect what users intended to add. Changing rules in the rule base does not affect other rules, whereas changes to the ontology might influence any existing rules. Designers can even put restrictions on the rule base in order to ensure consistency of new rules incorporated.

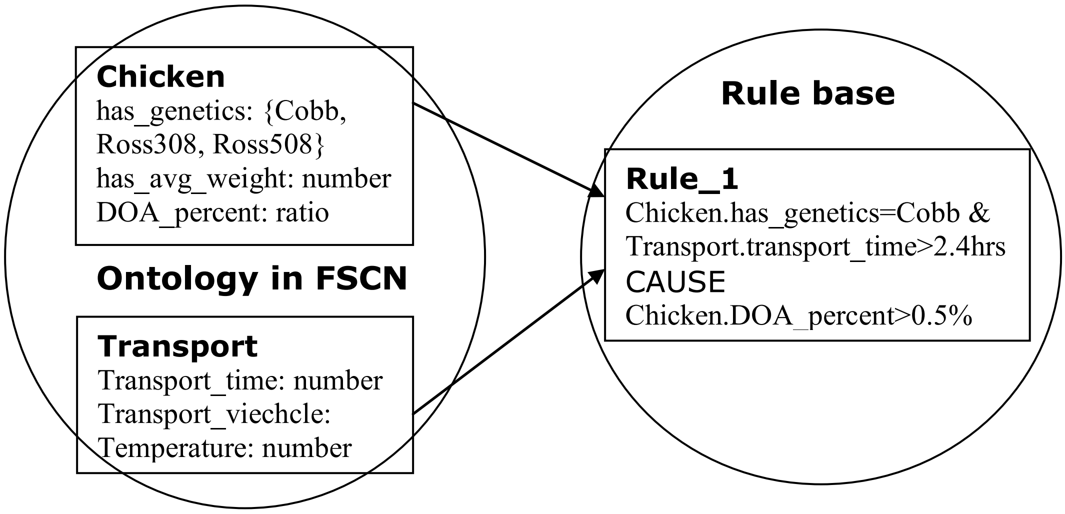


Figure 6-2: illustration of relations between rule base and ontology

To build ontologies for manifest knowledge, we mainly follow the process provided by METHONTOLOGY, one of the most mature approaches for ontology construction (Corcho, Fernandez-Lopez et al. 2003). METHONTOLOGY prescribes a set of activities for building ontologies, from objectives specification, through knowledge acquisition, conceptualization, implementation, to evaluation.

To model inferred knowledge, we use the methodology CommonKADS (Schreiber, Akkermans et al. 2000) to build a Knowledge-Based System. CommonKADS provides methods to analyze knowledge-intensive tasks and processes, and supports the development of Knowledge-Based Systems that support selected parts of the business process. Contrary to simple 'if-then-else' rules, CommonKADS enables designers to define different types of rules in order to capture different kinds of knowledge in a problem domain. Furthermore, CommonKADS provides a suite of models to capture different aspects (inference steps, tasks, etc.) of expert problem solving process. The inference structure and task models in CommonKADS provide a way to utilize constructed rule bases, thereby they facility the creation of systems to help users solve problems.

6.3.3 Ontology

In this section we present an ontology build from our case studies. We developed this ontology with OWL (Web Ontology Language) in Protégé 3.3 beta (Knublauch, Fergerson et al. 2004). Besides sufficient expressive power, OWL has well-defined syntax and formal semantics and hence provides efficient reasoning support (Antoniou and Harmelen 2004). Protégé comes with an IDE for ontology construction (with tabs, widgets, menus, etc.) developed in Java. It is open source so its Java APIs can be easily used in other applications. Furthermore, Protégé is backed by a large community of active users and developers, and the feedback from this community is valuable for its further development.

Figure 6-3 shows the major class hierarchy of the ontology we build for the Knowledge Base in EW&PC systems (‘class’, ‘instance’, ‘property’ are terms for types of components in an ontology). Relations between classes can be either subclass relations or property relations. The subclass relation, denoted by a solid arrow, means that the class is a specialization of the other class. The property relation, denoted by a dashed arrow, indicates that a property of one class takes instances of another class as its value set. The name of the property is shown by a term beside the arrow.

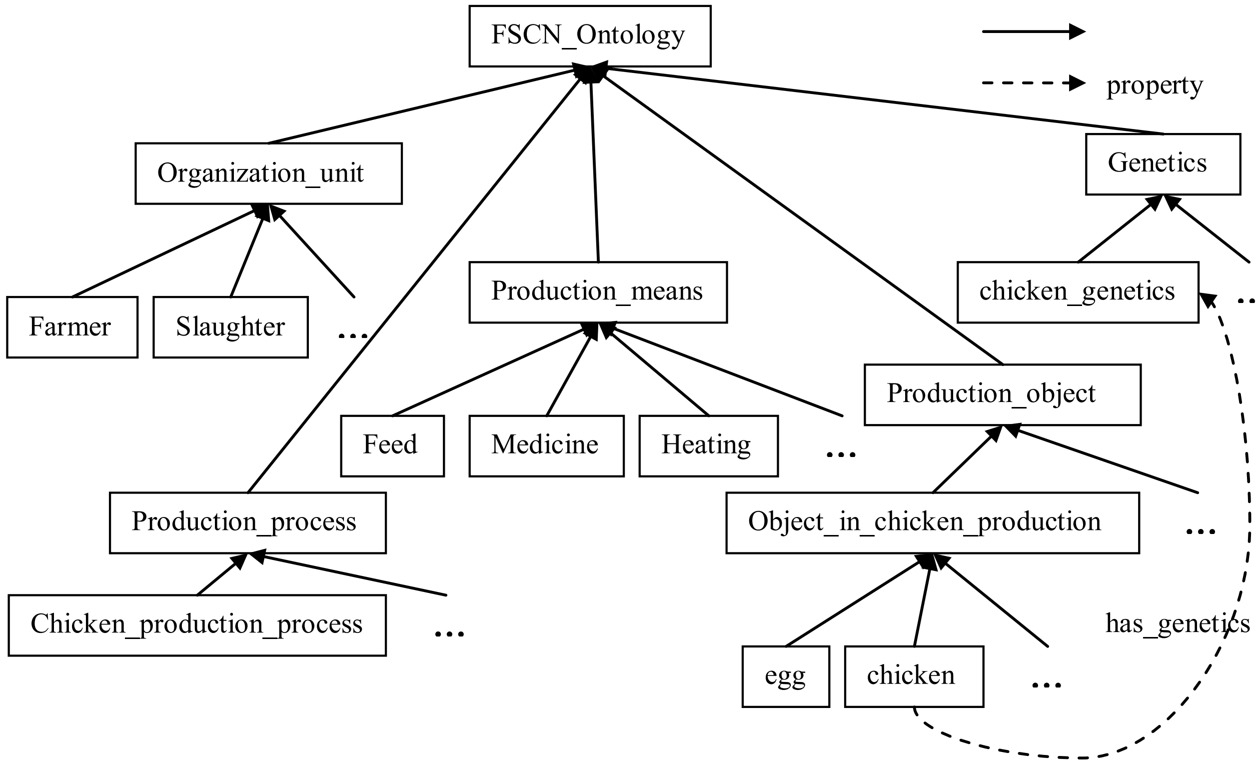


Figure 6-3: part of the class hierarchy in the ontology built for the knowledge base

The ontology in Figure 6-3 has the following major classes:

- *Organization_unit*: Organization units represent various kinds of human organizations in FSCN, such as *farmer*, *feed producer*, *slaughter house*, and *production manager*.
- *Production_process*: This is the class for various processes in food production. The processes in chicken production include: *hatchery*, *breeding*, *transport*, and *slaughter*. Pig production has its own types of processes.
- *Production_means*: This class represents facilities used in food production along the FSCN, such as *feed*, *heating equipment*, *machines*, and *medicine*.
- *Production_object*: Production objects represent food products being produced. Since we have cases in a chicken supply chain and pig supply chains, we introduced *objects_in_chicken_production* and *objects_in_pig_production* as two major subclasses. In the chicken supply chain, production objects include *chicken* and *egg*.

- *Genetics*: This class represents genetics of animals in FSCN. It has two major subclasses, *chicken_genetics* and *pig_genetics*.

Each class in the ontology contains some properties. For example, from the case study we can define following properties for class *chicken*: *has_body_weight*, *has_genetics*, *has_transport_ time* (from farm to slaughter house), and *DOA_percent*. The property *has_genetics* is shown by a dashed arrow in Figure 6-3, to illustrate that its value set consists of the instances of class *chicken_genetics*.

6.3.4 Knowledge-Based System

In order to model inferred knowledge in problem solving, we apply CommonKADS to build the Knowledge-Based System. This system is used to infer required knowledge from user input and internal rules. Figure 6-4 shows the inference structure of this system. This particular inference structure is a variation on the 'diagnose' template from CommonKADS.

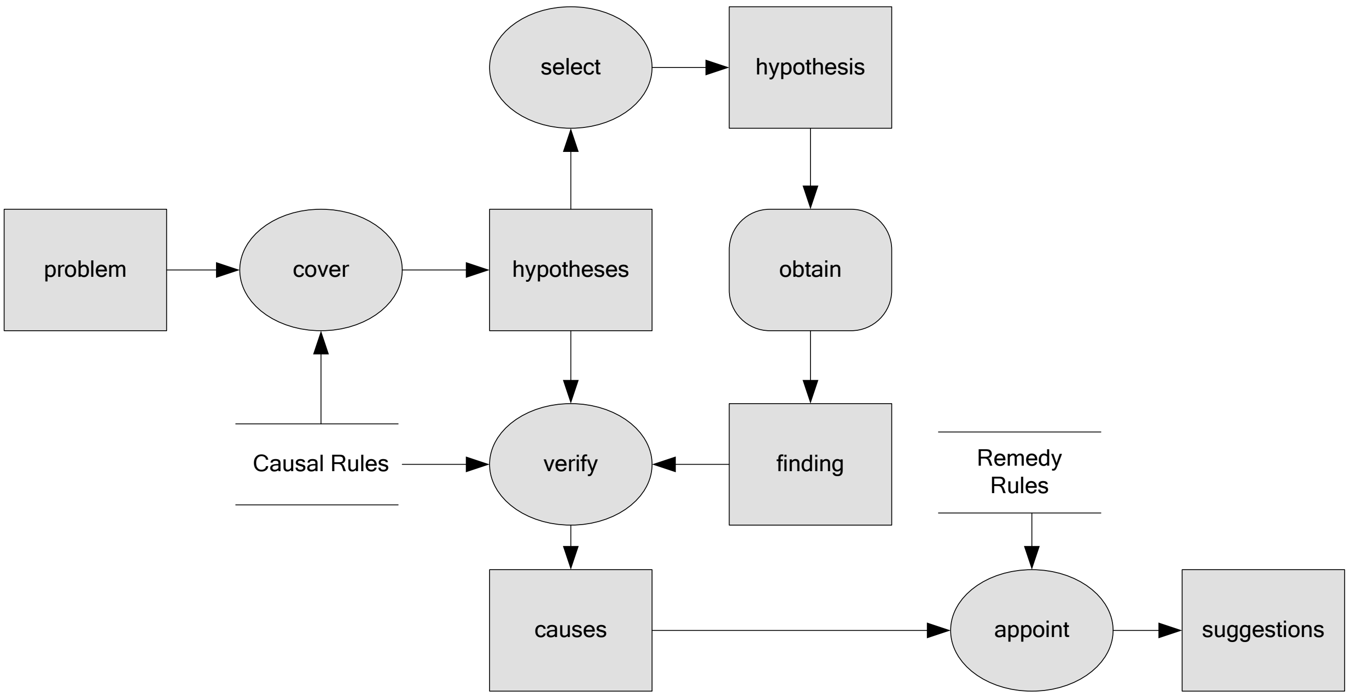


Figure 6-4: inference structure for Knowledge-Based System for EW&PC systems

Square boxes represent objects in the Knowledge-Based System, such as *problem*, *cause*, *suggestion*. Ovals represent functions within the system. Each function has inputs and outputs, e.g. the function *cover* takes the specified *problem* and *causal rules* as inputs, and outputs a set of known *hypotheses* about the causes of the problem. The rounded box represents a function that needs information from outside of the system. The function *obtain* poses questions to the user and collects his answers into the system. Items within parallel lines represent types of rules in the rule base. In this system, there are two types of rules, *causal rules* and *remedy rules*.

- *Causal rules* capture knowledge about causal factors of problems in FSCN. For example, the causal rule "*chicken.has_genetics=Cobb & transport.transport_time>2.4hrs CAUSE* *chicken.DOA_percent>0.5%*" means that if chickens with genetics 'Cobb' have been transported for longer that 2.4 hours, we expect their DOA percent to be larger than 0.5%.

- *Remedy rules* contain possible solutions for the identified causes. For example, the remedy rule "*transport.transport_time>2.4hrs HAS_REMEDY transport.departure_time<=6am*" suggests to transport before rush hours.

To start the inference procedure, the user should specify the *problem* encountered. Then the function *cover* looks into *causal rules* for *hypotheses* on possible causes. For each potential cause in the generated hypotheses, the system asks the user to input its status in the case at hand. With *hypotheses* and obtained *findings*, the system can *verify* the validity of hypothesized *causes*. With verified *causes* and *remedy rules*, the system *appoints* all available *suggestions* on how to solve this problem.

6.4 Application

To illustrate how the presented structure helps achieving the requirements specified in section 6.2.1, we have built prototypes user interfaces to link the ontology and the Knowledge-Based System. These prototypes have been built using Java and Protégé APIs.

6.4.1 Accessing the Knowledge Base

When managers encounter problems in FSCN, they can specify the problem by choosing corresponding class, property, and value from the ontology. The sequence of dialog windows in Figure 6-5 shows the interfaces for choosing class, property, and value.

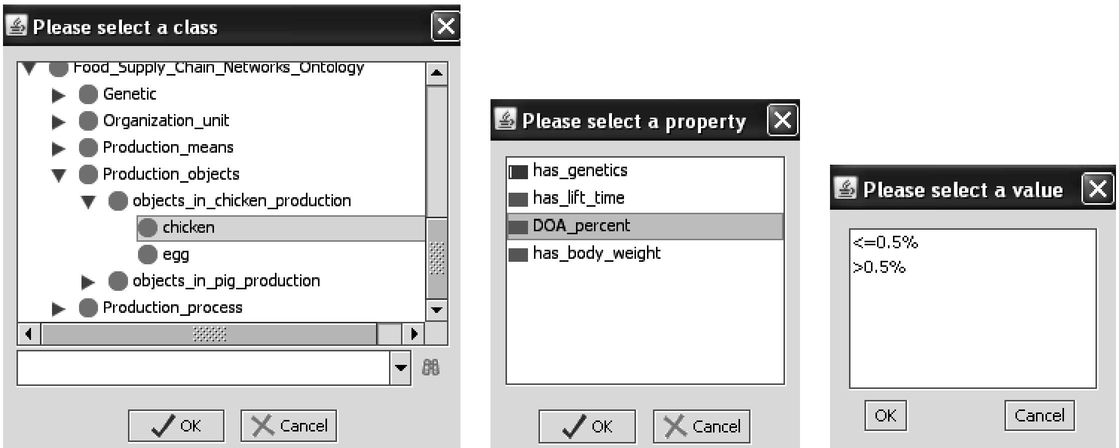


Figure 6-5: Interfaces for selecting class, property, and value

First, users have to select the class denoting the domain of the problem. As shown in the first dialog window, the class *chicken* is selected. Then the relevant property should be selected: in the second window, property *DOA_percent* of class *chicken* is selected from the four available properties. After that, its value set is shown in the third window. In this example the value set is enumerated as *"<=0.5%"* or *">0.5%"*. If *">0.5%"* is selected for the value of *chicken.DOA_percent*, the problem has been fully specified as *chicken.DOA_percent>0.5%*. The specified problem will be shown in a text field of the user interface, as illustrated at the left hand side of Figure 6-6.

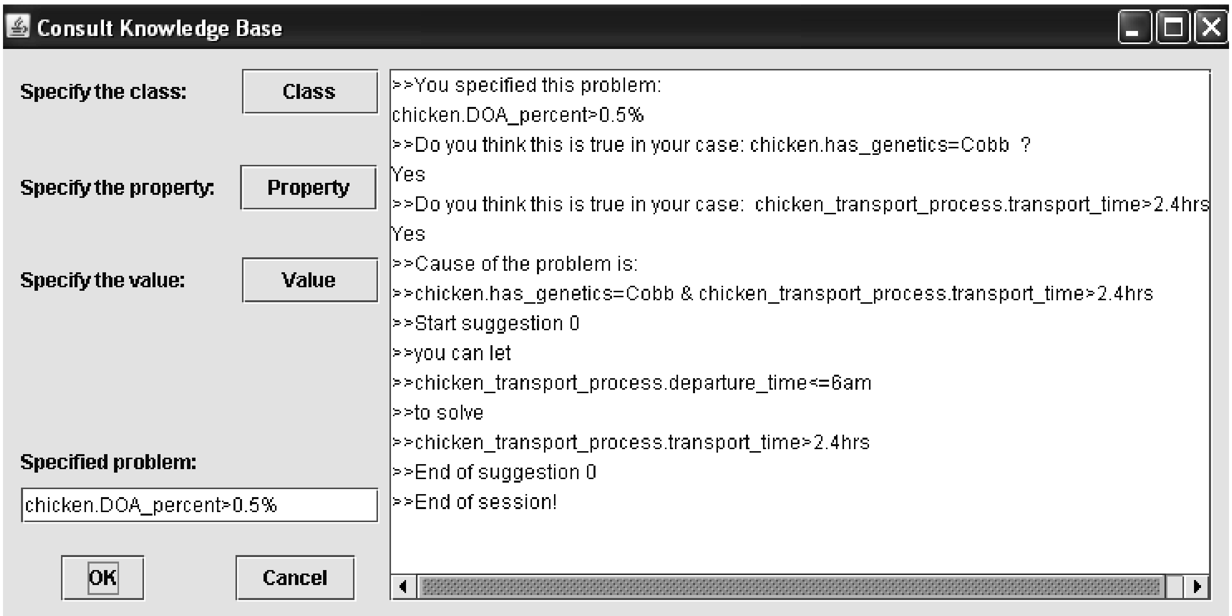


Figure 6-6: user interface for consulting the Knowledge Base

The system searches the rule base for rules that apply to the specified problem. All constituent clauses of those rules are presented to the user, who has to answer whether they are satisfied in his case. The presented clauses correspond to hypothesized causes for the specified problem. Based on the answers that the user gives, the system determines causes for the given problem and suggests solutions. As shown in the right pane of the user interface, the system reports that the cause for the given problem is *chicken.has_genetics=Cobb & chicken_transport_process.transport_time>2.4hrs*. It also makes suggestions for possible solutions, in this example to start transportation before rush hours.

6.4.2 Extending the Knowledge Base

The structure designed in this chapter eases extending the Knowledge Base with obtained knowledge. Mostly, extending the Knowledge Base means adding rules to the rule base. Each rule in the rule base consists of items that directly correspond to objects (classes, properties, instances) in the ontology. Therefore, a user can simply select relevant objects from the ontology to construct a rule. Furthermore, the ontology facilitates that users extending the rule base share an understanding about objects and terminology in FSCN. Therefore, extending the Knowledge Base can be achieved by a few mouse clicks, eliminating the chance for typing mistakes. As shown in Figure 6-7, the user interface for extending KB enables users to add, edit, or delete rules from the rule base. By choosing corresponding classes, properties, and values, users can generate front part and rear part of a rule. The selection interfaces are the same as those shown in Figure 6-5. Rule types (either "CAUSE" or "HAS_REMEDY" type) can also be selected from a list. After the user confirms the constructed rule, it will be added to the rule base as shown in the right pane of the interface.

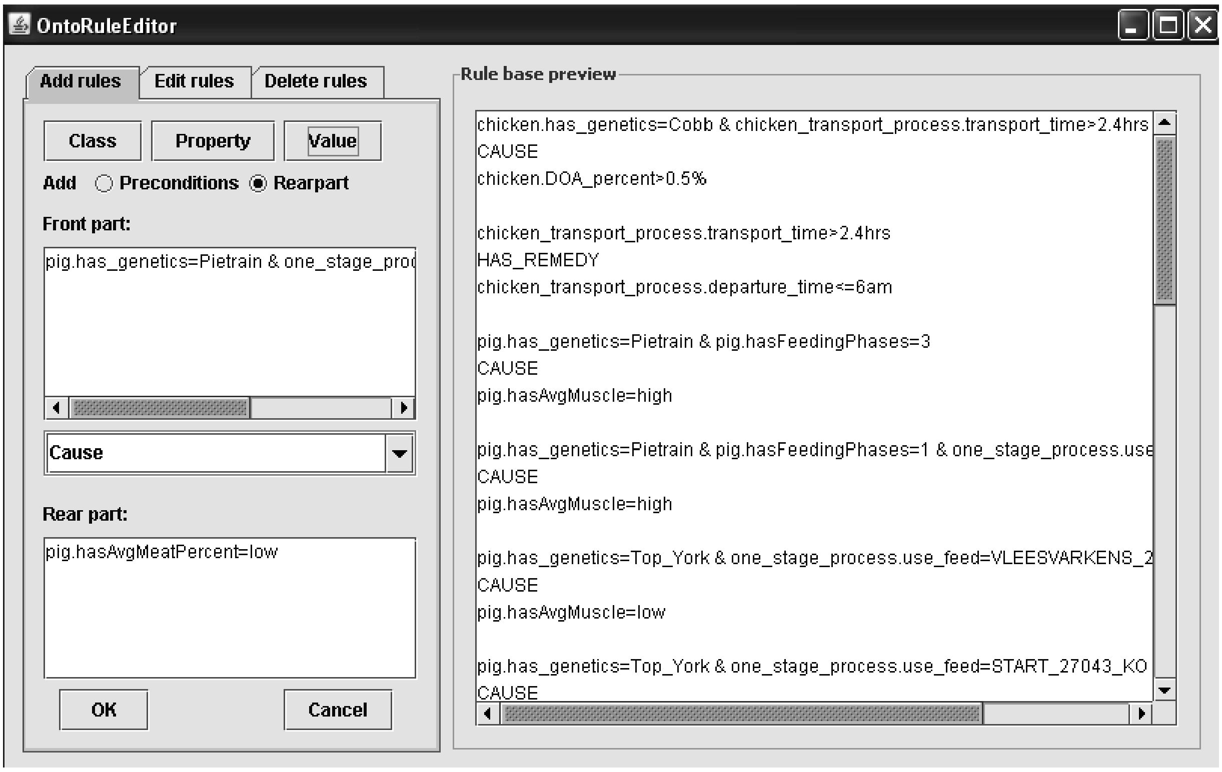


Figure 6-7: user interface for extending the Knowledge Base

Sometimes, it is necessary to add items to the ontology as well. Extending the Knowledge Base in this sense can be realized by using the Protégé IDE. Because such additions are less frequent than changes to the rule base, we did not design a special purpose user interface for this part in our prototype.

6.5 Conclusion and Discussion

The aim of our research is contributing to solving problems in FSCN by designing EW&PC systems. One of the essential components of such systems is a Knowledge Base that enables knowledge reuse and knowledge sharing among managers in FSCN. In order to facilitate designing a Knowledge Base, we propose a structure for such Knowledge Base. The structure separates manifest knowledge from inferred knowledge, and represents them with ontology and rule base respectively. As a result, designers and users can take advantage of current developments and advances in the areas of Ontology Engineering and Knowledge Engineering. To test whether this structure supports accessing and extending the Knowledge Base, we have built a prototype based on this structure. We filled the ontology and rule base with knowledge obtained from case studies in FSCN, and experimented with changing the KB contents. Results show that it is possible to obtain relevant knowledge from this KB for encountered problems in FSCN, and to extend this KB with obtained knowledge.

This structure is a first step towards a comprehensive Knowledge Base for EW&PC systems in FSCN. The prototype supports extending the Knowledge Base, but changing existing information in the Knowledge Base is not yet well-supported. Changing existing rules is relatively straightforward. However, when the user wants to change items in the ontology, consistency and correctness of the rules may become jeopardized. Further research is needed on how modification in the ontology influences rules in the rule base. The target is to devise automatic routines for changing the rule base when the ontology is changed.

Within the current design, there are still also some interesting issues to research. For example, how to maintain the consistency of the rule base when new rules are added? Duplication of rules is easy to detect and solve. For conflicting rules there could be several solutions, however. Should the system retain the conflict and leave it to the user to choose, or should some mechanism prevent conflicting rules to be added at all?

In the larger context of EW&PC systems for FSCN, the Knowledge Base presented here is one of the components of a general framework for such systems. As Figure 6-1 shows, a complete EW&PC system would contain other components as well. However, without the Knowledge Base it would be impossible to share findings between managers in FSCN by means of EW&PC systems. For every new problem encountered in FSCN, managers would have to supply information about objects and rules again.

Chapter 7. Validation of the prototype system and design process

Abstract

In previous chapters, a prototype Early Warning and Proactive Control system to predict and prevent problems in Food Supply Chain Networks has been designed. This chapter presents the assessment of a prototype implementation as well as the design process. Expert validation is employed as the main validation method. We categorize the system design into several aspects and operationalize Key Performance Indicators for each of these design aspects. In order to collect accurate and comprehensive feedback from experts, both questionnaires and plenary discussions have been conducted. Results show that the system design has generally accomplished its intended goals, but further research is needed on specific components to arrive at a mature system.

7.1 Introduction

Food Supply Chain Networks (FSCN) are characterized by complexity in structure, variability in primary inputs and operations, and uncertainty in influential factors (e.g. weather, pest). Managers in FSCN often face problems of different types, such as disease infection or quality deviation of final products. Causes for those problems may originate from various sources in FSCN, such as operations, environment, or inherent product quality. In order to diminish the losses or hazards that might be incurred by those problems, Early Warning and Proactive Control (EWPC) systems are needed to help managers predicting and preventing problems.

Current developments in Information Systems construction in FSCN provide managers with potential resources for dealing with such problems. Monitoring systems in FSCN keep records of various stages of FSCN, from primary production to final product delivery. In each stage of FSCN, recorded data also covers different kinds of factors related to food products, such as operational factors and environmental factors. By analyzing data recorded along the FSCN, managers can identify possible causes for encountered problems, and also predict upcoming problems or take actions to prevent upcoming problems based on actual status of causal factors. However, since managers in FSCN are normally non-experts in Data Mining (DM) or statistics, such a vision is not yet realized in practice.

In our research, we have built a prototype EWPC system to guide managers through data analysis for predicting and preventing problems in FSCN. This system facilitates managers to employ appropriate DM methods to analyze recorded data in FSCN. Such analysis allows managers to explore potential causal factors for encountered problems, and to predict upcoming problems based on existing status of those causal factors. This system also enables managers to store obtained knowledge about encountered problems into a Knowledge Base. Therefore, when other managers encounter similar problems, they can use such knowledge for reference.

In previous chapters we presented the designed prototype system for Early Warning and Proactive Control systems in FSCN. In this chapter, we explain how we evaluated this prototype system and present the main findings.

7.2 Aspects of design

Before we decide on which evaluation method to use and which KPIs to check, we first have to look at what we intend to evaluate. This system design contains several stages. We first conceptually designed a framework based on the requirements specified from case studies (in Chapter 3). After that we designed specific components (from Chapter 4 to Chapter 6). In order to arrive at a comprehensive and detailed evaluation on the system, we categorized our system design into the following aspects. This categorization is according to the scientific areas used in designing the design.

A. *General system design: Framework, usage process, and prototype system*

The framework for EWPC systems in FSCN depicts necessary components in the system and the relations between those components. Since problem investigation and problem solving with DM methods involves various processes, we designed usage processes to help managers execute the correct sequence of processes. Based on the framework, we built the prototype system for EWPC.

B. *Specific design aspects relating to DM: Template approaches and Expert system for method selection*

In order to help managers in data analysis, we generalized the processes for using DM methods in EWPC based on our experiences in case studies and literatures in DM. Then we designed template approaches to guide managers through those processes (in Chapter 5). Since selection of DM method requires considerable DM knowledge, we assessed the applicability of various DM methods for EWPC (in Chapter 4) and designed an expert system to help managers selecting appropriate DM methods for their tasks (in Chapter 5).

C. *Specific design aspects relating to the Knowledge Engineering: Knowledge base*

In order to facilitate managers accessing and extending the knowledge base, we designed a novel structure for such a knowledge base (in Chapter 6). This structure separates static knowledge from inferred knowledge, and represents these with ontology and rule bases respectively. Based on those rule bases, we built a Knowledge-Based System to help managers to access relevant knowledge for problems encountered in their FSCN.

D. *Employed Methodology*

This design oriented research followed a design methodology. This methodology could be a good reference for researchers when they intend to build similar systems in their own areas. The situation we faced in FSCN is quite common in other areas, such as steel and automobile industry. In those areas, various production processes are also combined into a complex structure where unexpected quality deviations also happen now and then. The information about operations and products/semi-products is also recorded regularly, which provides a promising resource for investigating and solving production problems.

7.3 Literature review

Validation of the designed prototype system is the last step in the design process (see Figure 2-4 in Chapter 2). Before we start evaluating the prototype system, we select an evaluation method which suits our situation and needs.

7.3.1 Existing evaluation methods

In design oriented research, evaluation is to check whether the prototype fits the design goals and satisfies the expectations of the stakeholders (Verschuren and Hartog 2005). Evaluation can be divided into two facets: verification and validation. *Verification* is to verify whether the system has been built right, and *validation* is to check whether the right system has been built (Chitra and Ahmad 2004). Therefore, the evaluation of our system design should cover both the correctness of the system design process and the value of the system to managers in dealing with problems in FSCN. We check the correctness of each stage in our design process, from requirement specification to system implementation. We also measure the extent to which the designed prototype system helps managers investigating and solving problems in FSCN. Furthermore, from a scientific point of view, we look at the value of the methodology being employed in designing this system, and the value of the major components in this system to researchers with similar problem areas.

In literature we find different approaches to evaluate designed systems, most importantly: conceptual validation (Rao, Owen et al. 1998), operational validation (Rao, Owen et al. 1998), empirical validation (Windrum 2007), face validation (Harrison 1991), and expert validation (Wolfert 2002). *Conceptual validation* intends to test the soundness of the underlying theory. Since we also intended to obtain feedback from users on the actual usage of the system, conceptual validation is not sufficient to serve our purpose. *Operational validation* tests whether the system could be used as required. *Empirical validation* tests the value of the system in solving real problems. Both operational validation and empirical validation requires a large number of cases for statistical evidence. This is not yet feasible in this research. Moreover, we also intend to test the correctness of the employed design method and whether it can be used in different circumstances. The opinions from experts in system designing are needed. *Face validation* confronts experts with outcomes of the system and asks their opinion on the validity of those outcomes.

Expert validation is a method to obtain expert opinions on the properties of an artefact (e.g. a model or a system), as well as on the design process leading to that artefact. The idea behind expert validation is to ensure a shared understanding about the validity of an artefact and the design methodology. Expert validation shares some features with face validation, conceptual validation, and operational validation. In expert validation, experts are confronted with the underlying theories and assumptions for the design. They can also apply the designed artefact to deal with a target problem. After that, experts can scrutinize the outcomes and give their comments on the properties to be evaluated.

Expert validation resembles a well known technique - Delphi. Delphi is a technique for aiding forecasting and decision making (Rowe and Wright 1999). It has been widely used in different areas, such as needs assessment, policy determination, and resource utilization. Its aim is to elicit reliable and stable opinions from a group of experts through controlled processes which consist of several rounds of questionnaires. This technique allows experts to express their opinions anonymously, so as to reduce social pressure. Several rounds of questionnaires are organized. After each round, experts are informed about opinions of other participants. Therefore, reliable expert opinions are obtained. At the end, those opinions are grouped together to calculate statistical indicators, e.g. median, quartiles.

Similarly to Delphi, expert validation also requires a group of experts who have relevant knowledge on the system design. The objectives and the structure of the designed system are presented to the experts. After that, questionnaires are given to them for eliciting their judgment. But there are some differences with Delphi:

- There is a session for experts to use the system by themselves on a real case. Eason (1988) emphasizes that during prototype system evaluation, it is better to let evaluators have 'hands on' experience within a task scenario. The problem situation encountered in the case is explained. Then experts practice with the system to analyze the data from the case. After experts walk through all the procedures in using the system, they obtain a more concrete conclusion.

- Experts were allowed to communicate directly in plenary discussion. Experts being invited may come from various scientific areas that are related to the system design. By allowing them to communicate, we allow experts to better comprehend the potential impact of the designed system.

7.3.2 KPIs in system validation

Literature in system validation provides different kinds of performance indicators (McCall, Richards et al. 1977; Nielsen 1993; ISO 2001). As indicated by Keen and Sol (1978), the effectiveness of decision support systems can be expressed using a combination of three 'U's: usefulness, usability, and usage.

<u>*Usefulness*</u> means whether the functions of the system in principle can fulfill what is required. It has the same meaning as the 'utility' defined by Nielsen (1993). Two specific KPIs need to be satisfied here: correctness and reliability. *Correctness* indicates the extent to which a program fulfills its specifications (McCall, Richards et al. 1977). *Reliability* indicates whether the system can be expected to perform its intended functions satisfactorily (McCall, Richards et al. 1977).

<u>*Usability*</u> represents whether a system can be used to achieve goals effectively, efficiently, and with satisfaction (Nielsen 1993). Generally speaking, usability is the question about how well users can use those functions. It indicates the ease of using the software. The system should not demand too much effort in learning and using it. Therefore, usability is further specified into three KPIs: learnability, operability, and efficiency. *Learnability* indicates whether it is easy for users to learn how to operate the system (Nielsen 1993). Users can quickly start using the system to solve problems if the system is easy to learn. *Operability* represents the effort of users for operation and operation control (ISO 2001). A system should not have components or procedures that might prevent users from operating it. *Efficiency* requires a system to fulfill its purpose without waste of

resources (either time or storage) (McCall, Richards et al. 1977). Reaction time of the system should be as short as possible.

<u>*Usage*</u> expresses the ease of using a system in a variety of contexts (Houten 2007). To cope with different contexts, a system should be flexible, understandable, and maintainable. *Flexibility* requires the system to facilitate the incorporation of changes required by the operating environment (McCall, Richards et al. 1977). *Understandability* represents whether a system requires much user effort in recognizing the logical concept and its applicability (ISO 2001). *Maintainability* indicates the effort required to locate and fix a fault in the program. In order to be maintainable, a system should be simple, concise, modular, and self-descriptive (ISO 2001). Scientists are especially interested in this aspect because it indicates the generality of the system design. It is preferable if the employed design methodology and constructed components and structure could be reusable in other situations.

Based on the analysis above, we arrive at the hierarchy of KPIs for evaluating our system design presented in Figure 7-1.

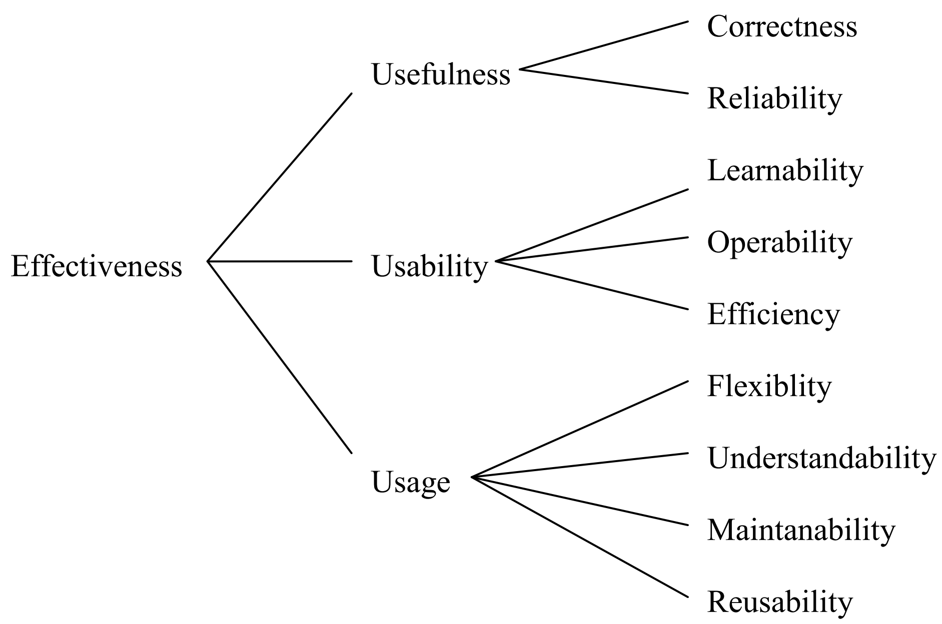


Figure 7-1: Hierarchy of KPIs for evaluating the designed system

7.4 Assign KPIs to design aspects

Since we presented different aspects of design, we have to assign appropriate performance indicators to each of those aspects and operationalize their meanings to facilitate correct evaluation. Table 7-1 summarizes the KPIs we assign to each design aspect. If a cell has a '+' symbol, the corresponding KPI is applicable for the design aspect. The specific meaning of each KPI in each design aspect is given in the following list. If a cell has a '–' symbol, the KPI is less important for that aspect based on the requirements from Chapter 3.

Table 7-1: assign KPIs to different design aspects

			Design aspect			
			Framework, usage process, and prototype system	Template approaches and Expert System for method selection	Knowledge base	Employed methodology
KPI	Usefulness	Correctness	+	+	+	+
		Reliability	+	+	+	–
	Usability	Learnability	+	+	+	–
		Operability	+	+	+	+
		Efficiency	–	+	+	–
	Usage	Flexiblity	+	–	–	–
		Understandability	–	+	+	+
		Maintanability	+	–	–	–
		Reusability	+	+	+	+

- Framework, usage process, and prototype system: we intended to design a framework which can satisfy both the functional and non-functional requirements (in Chapter 3). The prototype built upon this framework should also have other features (such as learnability, and operability) in order to facilitate implementation in reality.
    - + Correctness: whether the prototype system satisfies all the specifications defined in the framework
    - + Reliability: whether the prototype system can enable managers to accomplish EWPC in FSCN
    - + Learnability: whether it is easy for users to learn how to operate the prototype system
    - + Operability: whether it is easy for users to operate the prototype system
    - + Flexibility: whether the system can handle changes coming from the outside environment
    - + Maintainability: whether a change to one of the components, will necessitate changes to other components
    - + Reusability: whether the framework structure and the usage process are reusable in other application areas (provided there is a need for EWPC there)
- Template approaches and Expert System for methods selection: the Template approaches and Expert System should contain correct guidance from the Data Mining point of view. Furthermore, those two components should provide non-expert users with sufficient guidance.
    - + Correctness: from the Data Mining point of view, whether the knowledge incorporated in the template approaches and expert system is correct
    - + Reliability: whether the expert system and template approaches enable managers to apply Data Mining methods for EWPC
    - + Learnability: whether it is easy for users to learn how to operate the expert system and how to follow the template approaches
    - + Operability: whether it is easy for users to follow the template approaches and operate the expert system
    - + Efficiency: whether the expert system reacts fast enough to the queries from users
    - + Understandability: whether it is easy for users to comprehend the guidance provided by Template approaches and Expert System?

- Reusability: whether the Template approaches and the Expert System are reusable for EWPC in other application areas

- Knowledge base: the structure of this Knowledge Base should be capable of accumulating generated novel knowledge by managers and organizing it in an easy to access way.
  - Correctness: whether the structure of the Knowledge Base is capable of accommodating novel knowledge from managers and allowing managers to explore relevant knowledge
  - Reliability: whether this Knowledge Base enables non-expert users to accumulate and share the knowledge obtained during EWPC
  - Learnability: whether it is easy for users to learn how to access and extend the Knowledge Base
  - Operability: whether it is easy for users to explore the relevant knowledge and to extend Knowledge Base with obtained knowledge
  - Efficiency: whether the Knowledge Base reacts fast enough to the query from users
  - Understandability: whether it is easy for users to understand the mechanism of the Knowledge Base
  - Reusability: whether the Knowledge Base is reusable for EWPC in other application areas
- Employed methodology: this methodology should be clearly presented and well-defined. It should not be too complex to understand. Furthermore, it should be reusable by other researchers for building similar systems in their domain areas.
  - Correctness: whether this methodology provides correct support for system designers
  - Operability: whether it is easy for researchers to follow this methodology
  - Understandability: whether it is easy for researchers from other design areas to understand the design process as well as the individual design steps in this methodology
  - Reusability: whether the methodology is clear enough to be reused by other researchers who intend to build similar systems

7.5 Expert Validation

Before we employed expert validation to comprehensively evaluate the designed prototype system, we conducted a student test. The purpose of the student test was twofold: on the one hand to find potential flaws in the implemented components, and on the other hand to verify whether this prototype system can be easily understood and operated by non-experts in Data Mining.

In the student test session, a number of students were invited to a computer room. First, they followed our presentation on the objectives of this system and the way to operate it. They were also introduced to a real case in FSCN (see Appendix 7.2 for the task scenario). We explained the situation of this case, and the problem managers encountered in this case. After that, students operated the prototype system to deal with the given problem. The data set collected in this FSCN was provided for exploring causal factors, prediction, and decision making. Since one of the purposes of student test is to identify flaws in implemented components, we organized a stepwise questionnaire. After each step of operation, they filled in the questions prepared for that step.

Since the questionnaire for the student test contains many detailed questions for each step of using this system, it is impractical and not meaningful to show the specific results here. Instead, we provide a summary of the feedback obtained and the way we used it to improve the system (see Appendix 7.1). The student test helped us to identify parts of the system that were still difficult for non-expert users to understand.

After the student test, we improved the prototype system according to the obtained results. Then we executed the Expert validation to investigate the extent that those performance criteria are satisfied.

7.5.1 Expert selection

Experts from practical and scientific domains have different interests when they evaluate the system. To managers in practical problem solving, the system is treated as a black box. They are more interested in the learnability, operability, and reliability of the system. Experts in scientific domains focused on correctness, flexibility, and reusability. Table 7-2 shows the KPIs most interesting to experts from different areas.

Table 7-2: the relation between KPIs and different domains

	Practical area	Scientific area
Usefulness	Reliability	Correctness, Reliability
Usability	Learnability, Operability, Efficiency	Efficiency
Usage	Understandability	Flexibility, Maintainability, Reusability

Therefore, we invited experts from scientific community as well as from practice of FSCN management.

From the scientific side, we asked scientists to evaluate the correctness and reliability of the system design. Data Mining experts evaluated the Template approaches and Expert System. We asked experts in Knowledge Engineering and Ontology Engineering to evaluate the Knowledge Base. Furthermore, the methodology we employed and the framework we designed were evaluated by experts in Decision Support Systems for its scientific value.

From the practical side, we invited managers in FSCN to evaluate the reliability, operability, and understandability of the system. Managers in FSCN are interested in how can this system help them in solving problems. The system should enable them to explore potential causes and remedies, and to share obtained knowledge.

According to this plan, we invited a number of experts from FSCN management, Decision Support Systems, Data Mining, and Knowledge Engineering. Ten experts finally participated in this expert validation:

- 4 experts from Management of problems in FSCN (E1, E2, E3, E4)
- 2 experts from Decision Support Systems (E5, E6)
- 2 experts from Data Mining, (E7, E8)
- 2 experts from Knowledge Engineering and Ontology Engineering (E9, E10)

In order not to influence the evaluation results, none of these experts had been exposed to this prototype system before.

7.5.2 Steps in expert validation

In this section, we present the steps taken in expert validation. We also explain the structure of the questionnaire.

Step 1: Presentation on system design and case description. We first explained the management situation this system intends to serve, the design objectives, design steps, and the prototype system. Then we introduced a real case: the problem managers encountered in this case, the available data, etc. (see Appendix 7.2). This real case is an example application scenario for our system.

Step 2: Experts operated the prototype system to deal with the problem in the given case. With this step, experts evaluated the practical usage of our system.

Step 3: Experts filled in the questionnaire. We designed a questionnaire (see Appendix 7.3) based on target performance criteria and experts groups. The questionnaire covered the usefulness, usability, and usage of various design aspects as specified in previous sections. There are mainly two parts in this questionnaire:

Part 1: Closed questions. Closed questions with fixed formats (scoring, etc.) serve to quantitatively check the extent of fulfilling performance criteria. The different design aspects to be evaluated (framework and system, DM, KM, methodology) define the major sections in the questionnaire. The specific closed questions in each major section serve to cover the KPIs that have been assigned to that aspect. There are also two questions (question 1 and 2) about the context of this system design. These two questions are used to verify the assumptions on which the system design is based. For each closed question, we add a space for experts to explain their choices.

Part 2: Open questions. Open questions serve to collect comments and suggestions from experts. These open questions are designed to complement the closed questions. The purpose is to evaluate the potential impact of this system, and to explore opportunities for improving the system and methodology. Open questions allow experts to comprehensively explain their views.

These two different ways strengthen our confidence in collecting comprehensive feedback from experts.

Step 4: Plenary discussion. Since there are two types of experts (scientific and domain application), and considering the limited time, we separated experts into two groups. One group consisted of experts from the technical aspects: Data Mining, Knowledge Engineering, and Decision Support Systems. The other group consisted of experts from FSCN management. This allowed experts to understand each other more easily.

With all gathered feedback, we analyzed whether the designed prototype system has the potential to help managers in FSCN, and opportunities for improvement.

7.6 Validation results

7.6.1 Results from questionnaires

Table 7-3 shows the results of the questionnaire. Each row in this table represents a question in the questionnaire. The columns represent scores given by different experts. Scores range from 1 (strongly disagree) through 5 (strongly agree). The rightmost column indicates the weighted average value for each question. As discussed in section 7.5.1, experts from different domains have different emphasis during the evaluation. For each question, the yellow (shaded) color indicates the intended experts. Considering the fact that experts from one area may have opinions on other aspects, the marks from all experts are presented. However, the average values in the rightmost column are only based on the shaded cells.

Table 7-3: results of questionnaire

Questions	Experts										Weighted average
	FSCN	FSCN	FSCN	FSCN	DSS	DSS	DM	DM	KE	KE	
	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	
Context of this prototype system											
Q1:Value to management	5	5	5	4	4	4	4	5	4	5	4.8
Q2:Data availability	3	4	5		3		2	4			4.0
Framework Structure, usage process, and prototype system											
Q3.1:Problem expression	3	4	4	2	4	4	4	3	2	3	3.3
Q3.2:Data disclosure	1	4	5	4	4	4		4	3	3	3.6
Q3.3:Cause exploration	3	4	5	4	4	4	1	3	4	4	3.6
Q3.4:Prediction	3	4		4	4	3	3	4	4	4	3.7
Q3.5:Decision evaluation	4			3		5	5	3		3	3.8
Q3.6:Knowledge store&retrieve	2			2		4	5	4	3	4	3.4
Q4:Correctness	3	3	5	4	3	4	3	4		4	3.6
Q5:Easy to operate	3	3	4	4	4	3	3	4	4	3	3.5
Q6:Reusability	4	4	5	5	3	5	4	5	3	5	4.2
Template approaches and Expert System for methods selection:											
Q7:Correctness							4	3			3.5
Q8:Easy to operate	3	3	5	3	4	4	3	2	4	4	3.5
Knowledge base											
Q9:Ontology Correctness	4	3	4	3		4	3	4		4	3.6
Q10:Easy to operate	2		3	4	5	3	4	2	3	3	3.3
Employed methodology											
Q11:Easy to understand	5	3	4	4	5	5	3	5		5	5.0
Q12:Reusability	4	4	5	4	5	5	5	5	4	4	5.0
Overall system											
Q13:Impression	3	3	5	4	3	4	4	4	4	3	3.7

From the assigned meaning of values 1 to 5, we define the mapping in Figure 7-2 between value ranges and their indications. Scores in the range 3.5 to 5 mean that experts agree with the statement. Scores below 2.5 indicate disagreement, whereas scores between 2.5 and 3.5 indicate a neutral opinion.

Figure 7-2: Meaning of ranges in average score

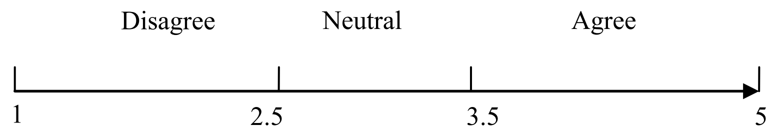


Table 7-3 shows that:

- (Q1—Q2) Experts from FSCN management acknowledged (score 4.8 in Q1) the value of Early Warning and Proactive Control (EWPC) systems. With score 4.0 in Q2, they also agreed with the context of EWPC systems, indicating that enough data is already available in real FSCN.
- (Q3) Regarding fulfilling the functional requirements, experts from all areas generally agreed that this system is able to accomplish the intended functional requirements. Some experts did not answer all sub-questions because they did not finish some steps. But the available scores indicate that experts acknowledged the value of this system for decision making. Two neutral scores (3.3 and 3.4) are given to the functions 'problem expression' and 'knowledge storage and retrieval'. The neutral score for 'problem expression' is a little bit surprising. But reactions to the support for 'knowledge storage and retrieval' are as expected. Some experts had difficulty in understanding the steps for composing a rule. Such situation has also been reflected in question 10. The '1' in column E7 is caused by the fact that this expert unfortunately overlooked the guidance we provided (as shown by the explanation written by this expert in the questionnaire).
- (Q4—Q6) With score 3.6 in Q4, experts from scientific areas (DSS, DM, and KE) acknowledged the correctness of the framework we designed for EWPC systems. Experts from FSCN management generally agreed that this prototype system is easy to use: score 3.5 in Q5. Reusability of the framework received high credit (4.2 in Q6) from experts in scientific areas, indicating the value of the framework to other researchers.
- (Q7—Q8) The score 3.5 in Q7 shows that DM experts acknowledged the correctness of the incorporated DM knowledge in this system. Experts from non DM areas regarded the template approaches as clearly written and easy to follow: score 3.5 in Q8.
- (Q9—Q10) With score 3.6 in Q9, experts from both FSCN management and KE regarded the structure and content of the ontology as correct for the purpose of the Knowledge Base. The neutral score 3.3 shows that experts outside KE areas were not satisfied with the knowledge storage capability of this system. This has been discussed above at the second bullet.
- (Q11—Q12) Regarding the methodology for building this system, experts in DSS gave high credits (5 in both Q11 and Q12) to both the understandability and reusability of our methodology.
- (Q13) On an overall level, experts are satisfied with this system, as shown by the score 3.7 in Q13.

7.6.2 Results from group discussion:

The results from the plenary discussion are organized according to the different design aspects:

Result 1: about framework and prototype system

The prototype EWPC system was considered potentially fulfilling most of its functional requirements. Experts from DSS and DM areas expressed that they were impressed by the combination of concepts from Artificial Intelligence and Knowledge Engineering. It creates a new tool for managers in FSCN to approach encountered problems. The system enables managers to investigate potential causes for new problems in FSCN. The scope of target problems is only limited by availability of relevant data for the problem. Moreover, the system can make problems in relevant data more explicit, which improves the data collection. This system can also help to bring research and practice together by making knowledge and problems explicit. Managers can use this system to verify their hypotheses about encountered problems. On a general level, experts from FSCN management thought such a system will have huge impact on FSCN management.

For further improvement, experts from FSCN management commented that the *user interface* needs major improvement, especially the part for guiding decision making. Managers should understand what happens in the system so that they may trust the system. They suggested making a summary of all the steps being taken, and all the decisions being made and the reasons. So at the end, managers can get a report to see what really happened. To make the user interface more friendly, experts from DSS pointed out that it is better to implement error-preventing ('fool-proof' in experts terminology) mechanism in the system to prevent managers from making mistakes. Those experts also suggested building a more uniform navigation frame for users to quickly getting used to it. We can also make a 'click demo' which could provide guidance to managers on the screen when they walk through all the steps.

On the functional aspect, experts from FSCN management suggested to equip our system with on-line processing capability, because real-time data is collected continuously. Furthermore, the data reading ability should be improved to handle data from various resources. Therefore, this system will be more widely applicable and easier to use.

A strong point of the framework design is reusability. Experts from DSS regarded this framework as very generic and reusable in other application domain with complex chain structures. Experts mentioned that the ontology might be an obstacle that prevents the system from being reused in another application domain. This point is discussed in Result 3.

Result 2: about Template approaches and Expert system

Data Mining experts generally acknowledged the correctness of the DM knowledge in Template approaches and expert system. Although there are some aspects that need further improvement (such as incorporating more DM methods), they did not spot any obvious flaws from Data Mining side.

Experts who are not familiar with Data Mining gave good credit to most of the Template approaches as well as the Expert system. The process guidance, together with guidance for individual steps (such as handle missing value, outlier, etc.), was not difficult to be understood. The expert system was also easy to operate. According to some experts, the parameter settings of Data Mining methods still need improvement. For example, we could provide adaptable default values for those parameters.

Result 3: about Knowledge Base

Experts from FSCN management generally accepted the knowledge stored in current Knowledge Base. They did not claim that any of the objects or relations is wrongly defined. However, they pointed out that the current knowledge content in the Knowledge Base is not ample enough for real application. For example, it does not contain any knowledge on the fishery sector. Furthermore, they commented that the user interface and ontology (which has been mentioned in Question 5) need further improvement. The operation of the Knowledge Base was not very easy for non-experts in Knowledge Engineering, because they are not familiar with the presentation of the *ontology*. In order to improve the usability of the Knowledge Base, it is important to present the ontology in a way that is easier for managers to understand. Experts from FSCN management prefer to have a certain level of overview for the ontology.

Result 4: about methodology

Experts from DSS acknowledged this methodology as generic, reusable, and valuable. They gave high credit to the requirement generation phase. They also pointed out some problems and suggestions:

- Use component review next to literature review. The idea of component review is to look at existing systems for applicable components that can be used in our system with minor change, like ontology, or Data Mining facilities.
- Incorporate user interface design at an earlier phase. If we could collect managers' expectation on the interface at an earlier stage, it will be easier to make it user friendlier.

7.6.3 Summary of Results

The summarized results of the expert validation are shown in Table 7-4. Based on the comments given in group discussion (Result 1 to Result 4), as well as on the scores given in the questionnaires (Q1 to Q 13), we categorize the performance into three levels: good, average, and unsatisfactory.

Good

If there is no negative comment during group discussion and the average score is above or equal to 3.5, the performance is categorized as good. The rectangle in the table means the performance of a particular design aspect is good for the corresponding KPI.

Average

If experts raised some issues for improvement, but the average score is above or equal to 3.5, the performance is categorized as average. The circle in the table means average performance.

Unsatisfactory

If there are some insufficiencies detected by experts, or the average score is lower than 3.5, the performance is categorized as unsatisfactory. The triangle indicates that the performance is unsatisfactory.

Table 7-4: summary of validation result

	Framework, usage process, and prototype system	**Template approaches and Expert System for method selection**	**Knowledge base**	**Employed methodology**
Usefulness	■	■	■	○
Evidence	*Q3.1-3.2, Q4, Result 1*	*Q3.3-Q3.5, Q7, Result 2*	*Q9, Result 3*	*Q13, Result 4*
Usability	○	○	▲	■
Evidence	*Q5, Result 1*	*Q8, Result 2*	*Q10, Result 3*	*Q11, Result 4*
Usage	■	■	○	■
Evidence	*Q6, Result 1*	*Q3.3-Q3.5, Result 2*	*Q3.6, Result 3*	*Q12, Result 4*

■ Good ○ Average ▲ Unsatisfactory

Since the system we designed is a prototype, we expected the usability would be an issue. Since the user interface of this prototype is not yet well engineered, we expected the usability of all design aspects (except methodology) to be average, especially the usability of the Knowledge Base. As discussed at end of Chapter 6, at the moment it is difficult for non-expert users (in Ontology Engineering area) to extend the ontology we build in the current KB.

Our expectations are roughly met by the results of expert validation. As Table 7-4 shows, the majority of those performances are good. The KPI 'usability' is not fully satisfied by most design aspects, especially the usability of the Knowledge Base. This is mainly due to the user interface design. Considering the fact that this is a prototype system, we consider such result as acceptable. The usefulness of the design methodology is average, which is a little bit lower than we expected. This is due to the reason mentioned in Result 4, the need for user interface design at an earlier stage. The usage of KB is also average. This is due to fact that the user interface for the KB is still a little bit difficult to understand.

On an overall level, we conclude the system design has accomplished its original expectation.

7.6.4 Further support

Our confidence on the practical value of this system is further strengthened by a demonstration session in a Dutch food company. We illustrated to 12 interested managers the principle of EWPC systems and how to use the prototype system to solve problems in FSCN. In the discussion session afterwards, those managers gave high credit to the prototype EWPC system and the potential value in various sectors of FSCN management (e.g. feed control, meat quality optimization, risk analysis). As a consequence, they started to consider organizing new research or projects for implementing EWPC systems in their domain areas.

7.7 Conclusion and Discussion

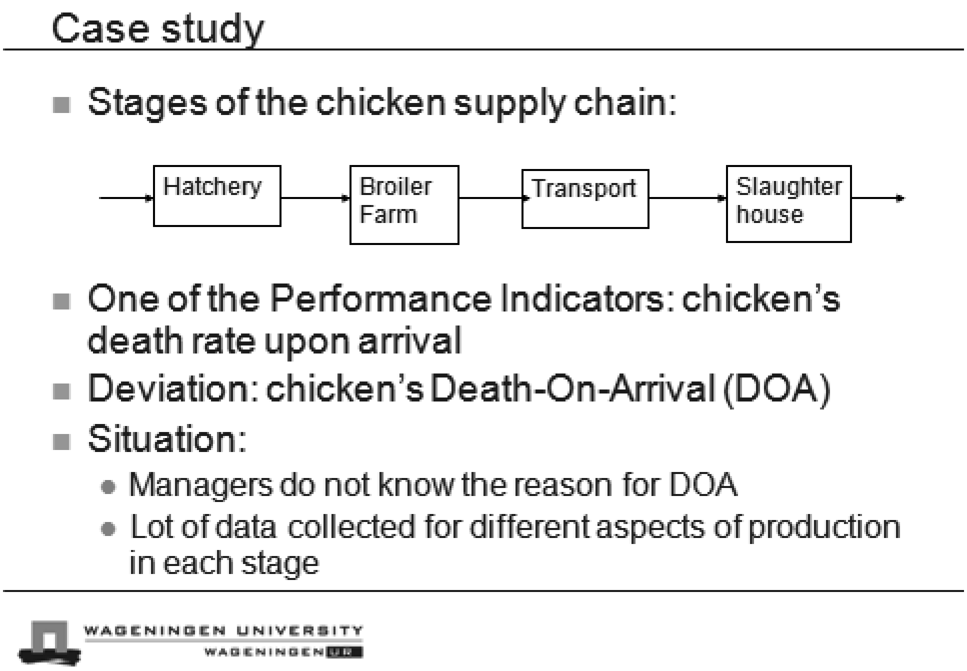
This chapter evaluates the designing context and employed methodology as well as the prototype EWPC system with its components. We started by investigating what design aspects to evaluate and which KPIs are associated to each aspect. The prototype system was first evaluated by students to eliminate flaws in system functioning. A group of experts evaluated the system design as well as the prototype system scientifically. Expert validation covered all the design aspects we planned to evaluate. Results show that the design and prototype system generally accomplish their intended purpose. The value of such a system to the FSCN management has been confirmed by the experts. Furthermore, the design methodology is regarded as correct and reusable. The validation results also indicate several directions for improvement of the system. We will further summarize them in Chapter 8.

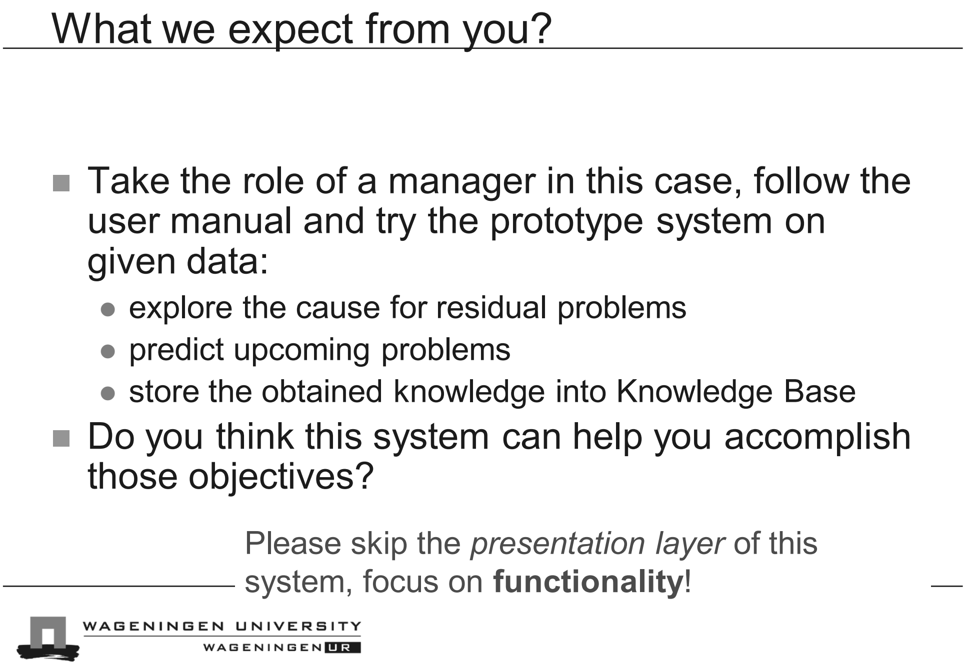
Appendix 7.1: Summary and reactions to student test

Step	Feedback	Actions
1. Problem identification and formulation	This step is easy to follow.	None.
2. Domain understanding	This step is easy to understand.	None.
3. Search KB for causal factors	Students encountered problems in understanding the terms from ontology engineering. Therefore, they were not sure about what they should do.	Provided an illustrative example in the user interface for composing problems. Designed a mechanism to formulate rule base in a human-readable way
4. Collecting data sets	Since data set was already provided to them, no feedback in this step.	None.
5.1 Preliminary factor selection	Some parts of expert system were difficult to understand.	The guidance for using expert system was clarified on what users need to do
5.2 Causal factor exploration	Similar to previous step, the problem with expert system.	See actions for previous step.
6.1 Check cause	Since they can not check any factor in practice, this step was skipped.	None.
6.2 Correct cause	Since they can not change any factor in practice, this step was skipped.	None.
6.3 Extend KB	It is difficult to compose those terms to express the obtained knowledge.	Added illustrative examples to explain terms in ontology Added a separate component for composing a condition of a rule
7 Prediction	It is difficult to understand the result of prediction.	Got rid of less relevant contents from the template approach for prediction
8 Decision making	It is difficult to understand what to do and why do it in that way.	The guidance for this step was rewritten to sharpen the type of support this system could provide

Appendix 7.2: Task scenario

Here we give a Task Scenario. Following two slides were used to explain the case situation and expected input from participants.





Following steps are taken from the user manual. In the user manual being used in expert validation, there are more details for executing each step.

Step 1: Problem identification and formulation.

This step makes you understand the nature of the problem encountered. It requires your knowledge on the normal and abnormal values of performance indicators related to the problem. In this step you are supposed to answer following questions:

1. what is the Performance Indicator (PI) that represents your problem?
2. what is the normal value range of this PI?
3. what is the current problematic value of this PI?

Step 2: Domain understanding.

In this step, managers check their knowledge on the FSCN and complement it if necessary. In each stage of FSCN, there are different factors that potentially cause problems in food products, such as operational errors (e.g. wrong usage of detergents), or environmental deviations (e.g. high temperature). After this step, you should know generally the stages that exist in this FSCN, as well as the kinds of operations and resources used in each stage.

Step 3: Search KB for causal factors.

Search the *Knowledge Base* for existing knowledge on possible causes of the problem. Choose appropriate Class, Property, and Value to compose your problem.

- Class represents the domain you are working with.
- Property represents the problem variable.
- Value represents the possible value set of the property.

System will ask some questions in order to find the realistic situation in your FSCN. Answer those questions with the knowledge you obtained in previous step.

Step 4: Collecting data sets.

In order to quantitatively further explore causal factors for a problem in FSCN, it is necessary to find data sets that contain relevant factors for the problem.

You should understand the meaning of attributes in available data sets and relations between those attributes and factors in FSCN.

Step 5: Explore causes.

Here we distinguish *Preliminary factor selection* and *Causal factor exploration*.

Step 5.1: Preliminary factor selection.

The database obtained in the previous step may contain much more factors than DM methods can handle. Large amounts of irrelevant factors for encountered problem not only waste time in analysis, but also may cause DM methods to generate wrong results. The rule of thumb is that when your variable amount is much larger than 30, it is suggested to do factor selection. There are two ways of doing factor selection.

- One is going through the database and using your expert knowledge to select variables that are likely to relate to the problem.

- The other is to use Data Mining methods to analyze collected data. *Template approaches* can guide you through each step of using DM method for factor selection, such as setting parameters and interpreting results. Result ranks variables according to their relevance to the problem. You will use this approach in this step.

Step 5.2: Causal factor exploration.

Use Data Mining methods to analyze the selected variables further to find potentially causal factors for the problem.

Please read in the data from local drive. After data has been correctly read in, follow the template approaches at the right side of the screen. Then you will be able to explore the causal factors for the DOA problem.

Step 6: Verify causal factors and extend knowledge base.

This step is divided into *Check cause*, *Correct cause*, and *Extend KB*.

Step 6.1: Check cause.

The ultimate way to validate the potential causal factors is to change them in practice and observe the results. This step checks whether those causal factors can be influenced by control measures. If a causal factor can be influenced by control measure, then we continue to next step.

Step 6.2: Correct cause.

Verify the causal factor by modifying it (either through taking control measures or through changing it directly) in the actual FSCN. If observed results do not confirm the causal factor, managers can go to previous steps and analyze again.

Step 6.3: Extend KB.

If observed results confirm validity of the causal factor, you can extend the *Knowledge Base* with the model built for identifying causal factors, together with the knowledge obtained about those factors.

The rule base contains all the rules being accumulated. Each rule in the rule base has three parts: front part, rule type, and rear part, as illustrated in the figure above. Your task here is to add a rule (which you obtained from data analysis in previous steps) to the rule base with the help of user interface for KB. At end of the rule base, you can see the rule you added.

Step 7: Prediction.

When a causal factor has been found, it can be monitored continuously in the actual FSCN. Based on the value of those causal factors, managers can predict whether this flock of chicken will have DOA problem.

Prepare data for prediction. You need 2 data files for prediction: one contains historical data for Data Mining method to learn, the other contains the instances to be predicted.

Please read in your 1st data file. After read has been correctly read in (check it in preview panel), click ‘Next’ and follow the template approaches at the right side of the screen. After having followed all the steps in the generic process (2nd data for prediction will be required in due time), you will get the prediction results as following figure shows.

Please interpret the prediction result with the help of Template Approach at the right side of the window.

Step 8: Decision making.

If some instances are predicted to be problematic, managers should consider proper control measures for changing production process parameters to prevent or repair the predicted problem. They can fine tune the value of certain causal factor, and then predict to see whether this value can make the prediction result normal. They should combine the results of prediction with his domain knowledge on possible decision varieties and their pros and cons to take correct actions.

In the data file, try different values on the causal factor and do prediction again to look at its result. After having followed all the steps in the generic process, you will see whether your change to the causal factor can prevent those flocks from DOA problem.

Appendix 7.3: Questionnaire for expert validation

Name:____________

Below you find a number of statements that can be graded on a scale ranging from 1 to 5. You can type your choice into the text box at the left side of the statement. If you like, please write your explanation to your choice in the text box under 'Explanation'.
At the end of the questionnaire, there are some open questions. Please answer them quickly and impulsively, because they will be discussed later on. You may want to change them later.

Questions about the context of this prototype system: (for experts in FSCN)

1) Early Warning and Proactive Control (EW&PC) system adds value to management of Food Supply Chain Networks (FSCN).

☐ Strongly disagree 1 2 3 4 5 Strongly agree

Explanation:

☐

2) Current FSCNs contain enough data for applying EW&PC system.

☐ Strongly disagree 1 2 3 4 5 Strongly agree

Explanation:

☐

Framework Structure, usage process, and prototype system: (for all experts)

3) This prototype system can be expected to realize the functional requirements as specified in following table.
Please type your choice into **each cell** of this table:

Strongly disagree 1 2 3 4 5 Strongly agree

	problem expression	data disclosure	Exploring casual factors	Prediction	Control measure evaluation	Store & retrieve knowledge
Satisfy specification						

Explanation:

☐

4) The framework designed as such is correct and logical given the requirements for EW&PC systems.

☐ Strongly disagree 1 2 3 4 5 Strongly agree

Explanation:

☐

5) It is easy to learn how to operate this prototype system.

☐ Strongly disagree 1 2 3 4 5 Strongly agree

Explanation:

6) This framework can be reused in another area that has similar situation as FSCN. The situation can be specified as: complex production processes with multiple steps, problems needs to be solved, and relevant data recorded along the production line.

☐ Strongly disagree 1 2 3 4 5 Strongly agree

Explanation:

Template approaches and Expert System for methods selection:

Question 7 and 8 is about following steps: Handle missing value, Handle outliers, Select method, Transform data, Set parameters, Validate results, and Interpret results.

For experts in DM

7) The Data Mining knowledge incorporated in this system is correct.

☐ Strongly disagree 1 2 3 4 5 Strongly agree

Explanation:

For non-experts in DM

8) The Template approaches provide clear instructions about how to use Data Mining methods. It is easy to follow the instructions of the template approaches.

☐ Strongly disagree 1 2 3 4 5 Strongly agree

Explanation:

Knowledge base: (for all experts)

9) The ontology contains correct structure and content for the purpose of the Knowledge Base.

☐ Strongly disagree 1 2 3 4 5 Strongly agree

Explanation:

10) It is easy to operate this Knowledge Base for both accessing and extending. There are enough facilities for me to represent target knowledge.

☐ Strongly disagree 1 2 3 4 5 Strongly agree

Explanation:

Employed methodology: (for all experts)

Question 11 and 12 is about the following steps: Case study, Literature review, Requirement generation, Framework design, Prototype implementation, Expert validation.

11) It is easy to understand the steps of the methodology for developing this prototype system.
☐ Strongly disagree 1 2 3 4 5 Strongly agree

Explanation:

12) This methodology can be reused if I intend to build a similar system.
☐ Strongly disagree 1 2 3 4 5 Strongly agree

Explanation:

Overall impression: (for all experts)

13) How would you grade the usability and value of the overall system?
☐ Very bad 1 2 3 4 5 Very good ()

Explanation:

Open questions:

You can input your answer into the text box below each question. All those questions will be discussion later on in plenary discussion session.

a) What is your opinion on the potential impacts of this system on management of FSCN?

b) Is there any EW&PC stage that still needs guidance but not being provided by this system?

c) Which components or usage processes of this system still need improvement, and what kind of improvement?

d) What might be the problem if this framework is reused in another application domain?

e) What is your suggestion for further improvement of this prototype system?

f) What tasks still have to done before this system can be implemented in reality?

g) What is your opinion upon the value of the employed methodology? Is there any improvement needed?

Chapter 8. Conclusion and discussions

The idea of using Data Mining (DM) to contribute to Early Warning and Proactive Control in FSCN was the starting point of this thesis. The first chapter introduced the research settings for this study, such as research objectives and questions, and research methods. Then we explained the concept of Early Warning and Proactive Control, and introduced the research process together with the results of a pilot case study in Chapter 2. In Chapter 3 we presented the framework that we designed for Early Warning and Proactive Control systems in FSCN. The applicability of DM methods for Early Warning and Proactive Control were investigated in Chapter 4. Based on that analysis, one major component in the framework, namely *Expert System for DM method selection & Template approaches*, was designed in Chapter 5. Chapter 6 presents a structure for the other major component in the framework: the *Knowledge Base*. After the prototype system was completed, we conducted expert validation to evaluate different aspects of our system design. The results of this validation are presented in Chapter 7. Here in the final chapter we draw the major conclusions from this research. It summarizes the answers to the research questions. We discuss the theoretical and methodological contributions of this research, as well as the limitations and further research. At the end we discuss the managerial implications.

8.1 Overview of our design process

This research aimed at designing an EWPC system to realize flexible process control for FSCN management. Such systems can enable managers to analyze relations between (attributes of) processes and performance in FSCN using existing data. In particular, they help managers to model relations between problems and their causes by means of a variety of Data Mining techniques. These models can be used to predict upcoming problems and prevent them at an early stage. By means of a repository of modelled relations, the system can be applied to subsequent similar problems. As specified in Chapter 1, we had two major objectives in this research:

Objective 1.

To design a framework for an EWPC system that is able to facilitate the following aspects:

- analyze relations between problems and causes
- predict upcoming problems
- suggest control actions to prevent upcoming problems
- use existing data bases in FSCN
- support non-expert users in applying DM methods
- have an extendable knowledge base

The framework should describe the necessary components as well as the relations between those components in EWPC systems.

Objective 2.

To build a prototype system based on the framework to enable managers in FSCN, as non-experts in DM, to use DM methods for Early Warning and Proactive Control on the supply chain level.

When looking back we can describe the design process in the following major steps:

1) Concept initiation
Our definition of *early warning systems in food supply chain networks* is as follows:

> Based on the knowledge of relations between deviations in performance of food supply networks and determinant factors, an early warning system in food supply networks predicts as early as possible potential deviations for decision makers in food supply networks by monitoring measurable determinant factors.

We define a *proactive control system in food supply chain networks* as follows:

> Based on the knowledge of relations between deviations in food supply networks and determinant factors, proactive control systems in food supply networks propose appropriate actions in order to prevent potential deviations which have been flagged by the early warning system. The potential actions could be discarding, taking corrective measures, or adapting subsequent processes in order to make amends.

For details on literature review and the explanation for the above definitions, please refer to Chapter 2.

2) Requirement generation
From case studies as well as literature on KDD (Knowledge Discovery in Databases) we summarized the generic steps that are needed to predict and prevent problems in FSCN by means of DM. Such generic steps helped us to derive the requirements for EWPC systems. This system should be able to support problem expression, data disclosure, data modelling, problem prediction, decision evaluation, and also knowledge sharing. Since the target users of this system are managers in FSCN, this system should be easy to use for non-experts in DM. Managers are less willing to spend much time on using a system. Furthermore, this system intends to predict and prevent problems in FSCN. So the system itself should be time efficient.

3) Framework and usage process design
Based on the literature on Decision Support Systems and Knowledge Based Systems as well as exploratory cases, we designed a framework for EWPC systems in FSCN. There are two main kinds of component in the framework. On one hand, there should be components to enable managers to use Data Mining methods for predicting and preventing upcoming problems. On the other hand, there should be components for managers to share knowledge obtained from the early warning and proactive control processes. The framework introduced in Chapter 3 gives a conceptual design of those requisite components. We also specified the usage process for this framework. The usage process explains how users can employ components to accomplish EWPC in FSCN.

4) Components implementation

After the framework design, we implemented each kind of component. All those components were assembled together into a prototype EWPC system.

The first kind of component (guiding DM usage) was implemented based on two aspects. One is the experience we acquired when applying DM methods for EWPC in a real-life FSCN. We learned in practice what type of knowledge is required for guiding non-expert users to employ the power of DM. The other aspect is the literature on Knowledge Discovery in Databases (KDD) and automatic method selection (see Chapter 4). We studied the support we could obtain from existing research on supporting DM. Details about this design can be found in Chapter 5.

The other kind of component (enabling knowledge sharing) was implemented in two steps. The first step is to accumulate the type of knowledge to be shared. By analyzing the real cases, we obtained knowledge on the causes and remedies for problems in FSCN. After that, we looked into literature on Knowledge Management and Ontology Engineering for suitable architectures for organizing such knowledge. The architecture should be able to organize the target knowledge so that it is easy to access, and also to present the knowledge so that it is easy to understand. Details about this design can be found in Chapter 6.

5) Evaluation
Because of the various design aspects we intended to evaluate, and the different KPIs we intended to check (see Chapter 7 for details), we chose Expert Validation as a method to evaluate the system design. We evaluated not only the value of this prototype system itself, but also the method we used in designing this system. We invited a group of experts from different areas (FSCN management, Data Mining, etc.) and explained to them the design principle of this system. Then we presented them with a simplified case (based on a real scenario) and asked them to use the prototype system to deal with the problem in this case. After that, we provided them with a questionnaire. This questionnaire covered the various KPIs defined for different design aspects. At the end, experts had a plenary discussion in which they elaborated their opinions on this system. Validation results indicated the merit of our design and also the areas for further improvement.

8.2 Answers to research questions

Here we present answers to the research questions as posed in the first chapter.

Answer to research question 1: What are the requirements for EWPC system design considering current practice of FSCN management?

The requirements for EWPC systems are derived from the literature on DSS, as well as cases in FSCN. From the DSS literature we gathered general requirements of DSS. Then we specialized and expanded these requirements to the context of FSCN management with the experience obtained from real cases. For the details about how those requirements are derived, refer to Chapter 3.

We distinguished three types of requirement:

- Performance requirements on EWPC systems are mainly concerned with integrating the systems into the business decision process. That translates into requirements with respect to the time needed to use such systems. The systems should not require time-consuming analysis.

Within the allotted time frame, the systems should always find acceptable solutions. Time-consuming tasks should be avoided, such as lengthy searches for optimal parameter settings when acceptable parameter settings are sufficient.

- Specific quality requirements deal with the usage of such systems. The systems should provide managers with sufficient assistance and help them to understand the steps to be executed, the techniques used, and how to interpret the results. Managers need guidance on how to use DM methods for data analysis and modeling. Such knowledge should be captured in wizards and templates.

- Functional requirements are more specific. We identified six functional requirements for EWPC systems in FSCN:

F1. Support problem recognition and expression

Problems need to be formulated in quantitative ways before they can be analyzed by DM methods. However, problems in FSCN do not naturally present themselves in quantitative ways. It is preferable for EWPC systems to utilize expert knowledge and to help the quantitative formulation of problems.

F2. Support data disclosure

The databases in real FSCN are often scattered at different stages of FSCN. In order to apply DM methods for quantitative analysis, it is necessary to combine data sets from relevant databases and convert them into a required format. EWPC systems should give clear and easy-to-operate guidance on distinguishing which data sets can be combined and how to arrange them into appropriate formats for data analysis.

F3. Support analysis and reasoning, and provide model building facilities

Compared to conventional investigation, data analysis is a faster way to find out the causes of problems. However, managers are normally non-experts in quantitative modeling. In order to help managers to quickly identify potential causes for encountered problems, EWPC systems should enable them to correctly utilize suitable DM methods for quantitative modelling.

F4. Support prediction of problems

Predicting potential problems is also a generic step for EWPC. In order to allow time for measures to be taken to deal with a problem, it is necessary to predict the problem as early as possible.

F5. Support evaluation of solutions

Evaluation of different solutions requires not only expertise in interpreting results from DM methods, but also domain knowledge on the potential influences of different decisions. The EWPC system should help the manager to utilize a model built during analysis as well as his domain knowledge, and to experiment with different decisions.

F6. Support storage, retrieval, and organization of relevant resources (data, information, and knowledge)

The EWPC system should indicate if prior knowledge on a problem exists and disclose that knowledge to users if it exists. It should also be able to accommodate new knowledge obtained during problem-solving and decision-making. Because of the rapid development in DM research, new DM methods appear continuously. This requires the EWPC system to be extendable with new DM methods and the associated knowledge.

Answer to research question 2: What components should be included in the EWPC systems, and how should those components cooperate to enable managers to achieve EWPC in FSCN?

The framework introduced in Chapter 3 gives a conceptual design of the components required.

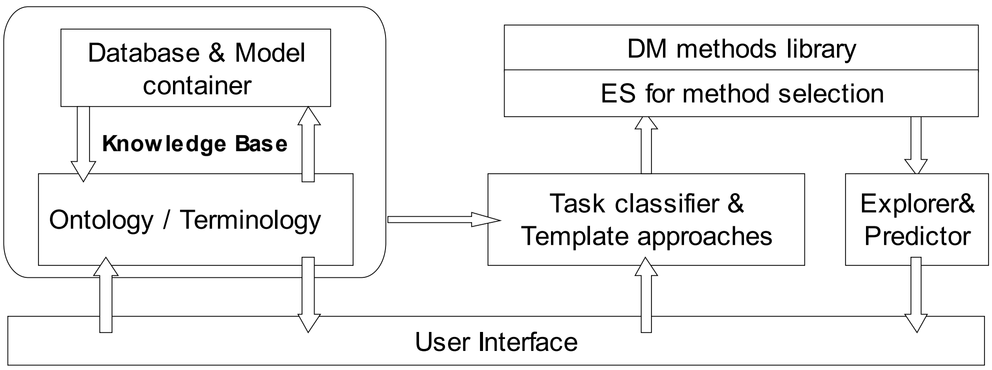


Figure 8-1: Framework for early warning system in Food Supply Chain Networks

<u>Task classifier:</u>

The framework contains a *Task classifier* which can direct managers to follow the correct processes when they intend to deal with an encountered problem. *Task classifier* helps users to quickly identify their task type. After the type has been identified, the user can obtain suggestions to perform some of the tasks listed below.

Table 8-1: Different types of task in EWPC systems

Task	**Description**
Identify problem	Identify a problem that arises in FSCN and formulate it quantitatively.
Search for cause	Search in *Knowledge Base* for existing knowledge on potential causes for this problem.
Find data	Find relevant data sets that can be used for exploring potential causes and combine the data sets.

Explore cause	Explore potential causal factors for the problem.
Verify cause	Verify the causal factor by manipulating (improving) it in practice and observing real-world effects.
Early warning	Predict problems that are likely to occur.
Decision making	Evaluate alternative decisions in order to choose the best solution.
Extend knowledge base	Extend the *Knowledge Base* with the knowledge obtained on causal factors and applicable DM methods.

The tasks listed above have been enhanced by corresponding *Template approaches* to expedite these tasks and support adherence to precedence sequences. The *Template approaches* component provides wizards that support users in these tasks.

Knowledge Base:

Knowledge Base is a very essential part. It satisfies the information needs from managers in FSCN. When managers encounter a problem, they need information on the problem, such as its causes and suitable control measures. *Knowledge Base* stores information on previously encountered problems in FSCN, such as the associated performance indicators with normal and abnormal value ranges. It also contains information about problem solutions (e.g. causal factors, causal relations) for easy reference by other users. Two main components of *Knowledge Base* are: *Database & Model container*, and *Ontology / Terminology*.

The *Database & Model container* stores metadata about the data sets obtained from FSCN, and the models that have been built using those data sets. The data sets contain data about the quality of food products and performance of operations at different stages of FSCN. The models are intended for prediction of potential problems and decision evaluation.

The *Ontology/ terminology* helps users to select applicable knowledge. An ontology is a formal description of domain knowledge. It consists of concepts, their properties and relations, structured according to human conceptualization. The *Ontology / Terminology* describes the structure and hierarchy of concepts in the databases and models in the *Database & Model container* in an understandable, extendable, and accessible way. With *Ontology / Terminology*, users can compare their cases with those stored in the *Knowledge Base* to look for similar cases and relevant knowledge about those cases. Substantial management time can be saved by reasoning on a semantic level rather than a data level. This component also ensures that new databases and models can be incorporated quickly and easily at appropriate places in the *Database & Model container*.

DM methods library and Expert System:

The *DM methods library* stores comprehensive information about all kinds of DM methods that can be used for EWPC. The information about a DM method includes its function, model format, and requirements on data sets. The *Expert System* contains facts, heuristics, and inference procedures to mimic the method selection processes of DM experts. After obtaining the required information (such as purpose of data analysis, format of available data) from the user, this component can give suggestions on which DM methods to use and explain its reasoning.

Explorer and Predictor:
The *Explorer* component allows users to explore potential causal factors for the problems in FSCN. The outcome of *Explorer* is a set of candidate causal factors for the problems at hand. Such knowledge needs to be verified by the user. After verification it can be used for future prediction of the same type of problems.

The *Predictor* warns for problems that are about to occur in FSCN based on the current data and acquired causal relations found in *Explorer*. The *Predictor* is used for decision evaluation as well. Users can employ models built previously to compare the results of different available decisions and choose the best one.

In addition to the specification on those components, we also defined the steps that are needed for using the system, as well as the correct sequence between those steps. For details about the usage process, please refer to section 3.4 in Chapter 3.

Answer to research question 3: What Data Mining methods are available and applicable for EWPC in FSCN?

To answer this question, we have to look at what kind of functional requirements have been identified for EWPC in FSCN. Based on the requirements at the system level, we derived the following required capabilities on the DM method level.

1) prediction
2) problem detection
3) finding determinant factors
4) representing complex structure
5) different representation forms
6) extendable with new knowledge

Different DM methods can contribute to different functional requirements for EWPC systems. The first four functional requirements (prediction, problem detection, finding determinant factors, and representing complex structure), deal with functions of DM methods. For details about those requirements, please refer to Chapter 4. Since in the DM area functions (such as classification, regression) are categorized differently from what we defined for EWPC, we provided a mapping between these two kinds of functions (see Table 4-1 in Chapter 4). For example, both classification and regression are helpful for finding determinant factors for encountered problems. After that, we selected a list of widely used DM methods (Decision Trees, Neural Networks, Bayesian Networks, Association Rules, Nearest Neighbours) and identified which method can accomplish which DM function (see Table 4-2 in Chapter 4). For example, Decision Trees can be used for classification and regression.

There are two functional requirements (different representation forms, and extendable with new knowledge) that relate to the representation form of DM methods. We provided another mapping between representation forms of those DM methods and their extensibility for new knowledge (see

Table 4-3 in Chapter 4). For example, Decision Trees, with a representation form of a decision tree, can easily be extended with new knowledge.

Answer to research question 4: What support needs to be provided to managers in order to enable them to use Data Mining methods for EWPC?

There are two kinds of support that need to be provided in order to enable managers to use DM methods.

One is how to find a proper DM method. This can be supported with an *Expert System for DM method selection* and a *library of DM methods*. Managers can get suggestions on which DM method is suitable after they specify their case situation and data set characteristics to the Expert System. Then they can select a suggested DM method from the DM method library.

The other is how to use the DM method found for EWPC. This can be supported with *Template approaches*. *Template approaches* capture the generic steps that are needed in applying Data Mining methods. They serve as a roadmap for managers during data analysis. Each step of applying DM methods is also supported by a particular template approach. Such a template approach tells users how to execute the particular step, what performance indicator to look at, and what to do if a particular situation occurs. For details of those components, please refer to Chapter 5.

Answer to research question 5: What kind of structure is suitable for the Knowledge Base in EWPC systems?

For knowledge sharing, we designed a knowledge base for managers to accumulate their obtained knowledge. Here we summarize the structure of this Knowledge Base on a high level. Chapter 6 explains in detail the structure we designed for the Knowledge Base.

The knowledge base consists of two main parts: a *rule base* and an *inference structure*. A *rule base* allows managers to store obtained knowledge. An *inference structure* allows managers to quickly identify relevant knowledge. Those two components together enable managers to easily share obtained knowledge.

The *rule base* allows managers to specify what kind of causal relation and/or remedies they obtained. There are two types of rules: causal rules for the causes of problems, and remedy rules for possible control measures for the causes. Each rule is composed of terms from an ontology we designed for FSCN. The ontology ensures that other managers can easily identify relevant knowledge for their encountered problem and can also correctly understand the knowledge found. Figure 8-2 gives an example of this idea. ‘Chicken’ and ‘Transport’ are two objects in the ontology for FSCN. The rule ‘transport time is longer than 2.4 hours and chickens have genetics Cobb is a cause for DOA rate larger than 0.5%’ is based on the objects and their properties in the ontology. It describes the inferred relations between the two objects ‘Chicken’ and ‘Transport’.

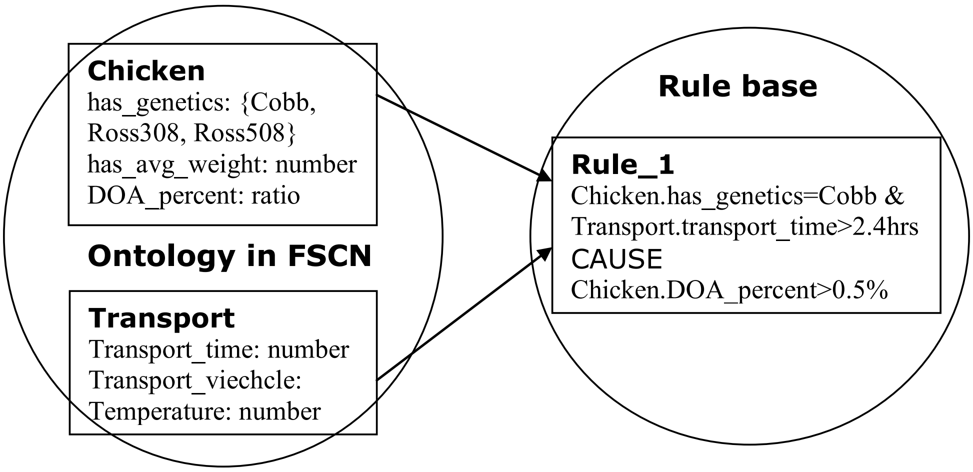


Figure 8-2: Illustration of relations between rule base and ontology

In order to help managers to quickly identify applicable knowledge for their problems, we built an *inference structure*. It can infer required knowledge from user input and internally stored rules with an automatic inference mechanism. Such a mechanism takes users' problem specification as input, and generates hypotheses on the problem by consulting the rules in a rule base. Then it asks users for verification of the hypothesized situation. If the situation in FSCN confirms the hypothesis, a causal rule will be shown to users. The causes in the causal rule will be automatically used to search for applicable remedy rules.

Answer to research question 6: What is the validity of the designed framework and prototype system?

In order to systematically evaluate the validity of the designed framework and prototype system, we first explored the appropriate KPIs for the framework and prototype system, as well as for its major components (Template approaches and Expert System for DM method selection, Knowledge Base) and design methodology. Then we employed Expert Validation to evaluate the performance of those different design aspects on those KPIs. The detailed procedure of Expert Validation can be found in Chapter 7.

The results of expert validation came from two parts: one is the scores from the questionnaire; the other is the group discussion. Table 8-2 summarizes the performance of each design aspect on each KPI (Usefulness, Usability, and Usage). We categorize the performance into three levels: good, average, and unsatisfactory.

Good: If there is no negative comment during group discussion and the average score is above or equal to 3.5, the performance is categorized as good. The (green) rectangle in the table means the performance of a particular design aspect is good for the corresponding KPI.

Average: If experts raised some issues for improvement, but the average score is above or equal to 3.5, the performance is categorized as average. The (yellow) circle in the following table represents average performance.

Unsatisfactory: If there are some insufficiencies detected by experts, or the average score is lower than 3.5, the performance is categorized as unsatisfactory. The (red) triangle indicates that the performance is unsatisfactory.

Table 8-2: Summary of validation results

	Framework, usage process, and prototype system	Template approaches and Expert System for method selection	Knowledge base	Employed methodology
Usefulness	■	■	■	●
Usability	●	●	▲	■
Usage	■	■	●	■

■ Good ● Average ▲ Unsatisfactory

The majority of these performances are rated 'good'. The KPI 'usability' is not fully satisfied by most design aspects, especially the usability of the Knowledge Base. This is mainly because the user interface design is not yet well engineered in this prototype system. The usefulness of the design methodology is average, because experts suggested including iterative design (a cyclic design process in which changes and refinements are made after each round of testing) in the methodology. The usage of KB is also average, because the way the ontology is presented does not ensure easy understanding by non-experts in Ontology Engineering.

Our expectations are roughly met by the results of expert validation. Since the system is still a prototype, we expected the usability of all design aspects (except methodology) to be average, especially the usability of the Knowledge Base. On an overall level, we conclude the system design has accomplished its original expectation.

8.3 Theoretical and methodological contributions

Here we explain the theoretical and methodological contributions of this research to several areas: Supply Chain Management, Decision Support Systems, Data Mining, Knowledge Engineering, and system design.

For *Supply Chain Management*, our designed EWPC system presents a new means to monitor and control the processes. In the current practice of Supply Chain Management, managers often use the models developed by specialists in data analysis for process control. Such models are built for specific problems. If an encountered problem is out of the scope of existing models, managers cannot use those models anymore. Instead, they have to use other ways to deal with new problems, such as field investigation. EWPC system offers a tool that can be used to gain knowledge on causes and remedies for encountered problems. Such knowledge is important for accomplishing flexible process control in FSCN. Users can employ DM methods to explore causal factors for their encountered problems, to predict upcoming similar problems, and to evaluate different control measures.

For *Decision Support Systems*, EWPC systems provide a novel architecture for supporting decision making. Traditional DSSs have predefined target problems and pre-developed models for the relation between target problems and their causal factors. But the architecture in EWPC systems allows users to deal with non-predefined problems, as long as there are relevant data recorded. We supply users with a tool to build models for the encountered problems. Therefore, the scope of problems that can be solved is not restricted beforehand. This architecture is beneficial for application domains in which new problems occur every now and then, when there are extensive data recorded about the problems and relevant operational factors.

For *Data Mining*, this study extends the existing research on the applicability of DM methods by creating a bridge between the real-life needs and the DM area. As far as we know, current research on the applicability of DM methods only use the terminologies defined from a DM method point of view. For example, Decision Tree can be used for classification. But to apply DM in reality, researchers also need to think from the point of view of real life (or the application domain): what does real life require of DM methods? Our research focuses on Early Warning and Proactive Control in FSCN. It takes real life (about managerial scenario, business process, data availability, etc.) into consideration. Based on that, we defined the functional and non-functional requirements of DM methods. With several case studies and literature reviews, we arrived at the applicability of DM methods for EWPC in FSCN. Therefore, our research not only extends the applicability of DM methods to another domain, but also serves as an example of how to investigate the applicability of DM methods for a particular domain.

For *Knowledge Engineering*, this study creates a suitable Knowledge Base structure for sharing knowledge among non-expert users by linking Knowledge Management and Ontology Engineering. We analyzed the knowledge accumulated in case studies and differentiated the concept of *manifest knowledge* and *inferred knowledge*. Then we represented those two types of knowledge with Ontology and Rules base respectively. Therefore, techniques from Ontology Engineering and Knowledge Management could be used to build and maintain the Knowledge Base of such a structure. This way of combining Ontology Engineering and Knowledge Management ensures that non-expert users share the same understanding of the knowledge stored in the Knowledge Base.

8.4 Discussion and further research

The prototype EWPC system designed in this research demonstrated the feasibility of letting managers employ the power of DM methods in FSCN management. Looking back at this design oriented research, we have some observations on this research.

Limitations of this research

First, the result of Expert Validation shows that the usability of the prototype system still needs improvement. The presentation of the ontology in Knowledge Base is not user-friendly enough.

Non-experts in Ontology Engineering have some difficulties understanding certain parts of the ontology, and they do not know how to update it.

Second, the designed prototype system did not provide sufficient support for managers to join different data sets. When the operational data sets are stored in different stages of FSCN, data joining turned out to be important in the practical application of EWPC systems. For instance, in order to investigate a performance problem in a chicken supply chain, managers need not only slaughter data, but also transport and feeding information. The more information they obtain, the more chance they have of finding causal factors. However, a slaughter house may only have data sets about the performance of slaughtered chickens, such as body weight and diseases. But these do not contain information about the feeding pattern of those chickens, which are collected at farms. Data sets from different stages need to be linked together in order to be meaningful for data analysis.

Third, the cases being studied in this research all came from the meat sector. We have not looked into other sectors, such as vegetable, dairy production. From our knowledge on those sectors, we expect comparable food quality problems. However, further research is needed on this aspect.

Further research

Based on those limitations, we suggest the following directions for further research:

The first direction is to look for a suitable way to present the ontology in Knowledge Base for managers. Managers need to understand the ontology before they can use it for formulating problems, or even update it with more terms from their management practice. Such research is also scientifically important for other areas, such as Knowledge Engineering and Ontology Engineering, as it will make it easier for non-expert users (in Knowledge Engineering and Ontology Engineering) to actually employ the artefacts built in those areas.

The second point is the research on how to help data joining. In this research, data joining was not taken as the research objective. But we were aware of the difficulty in designing guidance facilities for managers to correctly join their data sets. There are several reasons for this difficulty. Sometimes there is no key to link different data sets. Moreover, the semantic meanings of variables in two data sets may be different. To make it more difficult, we cannot foresee what kinds of data managers intend to join. We encourage further research on generalizing the kinds of data sets to be joined in FSCN. Such a typology is necessary for designing guidance for data joining. There is also some promising research from the area of data integration (Beneventano, Orsini et al. 2009; Dong, Halevy et al. 2009) and web semantics (Gracia, d'Aquin et al. 2009; Hoehndorf, Bacher et al. 2009).

The third direction is to experiment with this prototype system in more cases in FSCN, especially from other sectors than meat production. The goal of a research study is to obtain generally applicable knowledge. By taking more cases into the system design, more confidence will be obtained in the scientific validity of the findings obtained. Moreover, it would be interesting to investigate the differences between animal sectors and other sectors, and to research the implications of these differences for the EWPC system.

The fourth point is to look at the possibility of incorporating statistical methods into EWPC systems. We intentionally only included DM methods in our research, as this is sufficient to establish the feasibility of EWPC systems in FSCN. If DM methods can work in the system, we expect statistical methods to work as well. There are various statistical methods, like logistic regression and time series analysis, all of which have different functions, for example prediction and anomaly detection. It is to be expected that some statistical methods could accomplish some of the functional requirements of EWPC in FSCN. In order to complement the power of the EWPC system, further research is recommended on:

- which statistical methods have the potential for EWPC in FSCN,
- how to enable managers to use statistical methods for EWPC in FSCN.

8.5 Managerial implications

This research has several implications for management of FSCN in practice. Supply chain managers, food quality managers, and information technology managers can all benefit from this research.

Supply chain managers

Supply chain managers are responsible for the integrated planning, co-ordination and control of business processes and activities in the supply chain to deliver superior consumer value at less cost to the supply chain as a whole whilst satisfying requirements of other stakeholders in the supply chain (Van der Vorst 2000). In order to be effective and make this system work for supply chain managers in practice, they have to take into account the factors for effective proactive control (as specified in Chapter 1). Basically it means that they have to ensure:

1) clear objectives and measurable KPIs;
2) that information (databases) on the environment and system state is available and usable;
3) sufficient information processing capability, in which data joining is one of the critical issues;
4) to find and define decision alternatives;
5) to define a model for the object system to be controlled.

Defining the model is the most difficult of these points. That is where the research presented in this thesis provides support and new opportunities. The framework for EWPC makes DM methods available to supply chain managers for finding these models from new data.

Food quality managers

Food quality managers in FSCN are mainly concerned with how to utilize available resources (people, facilities, information, etc.) to ensure consistent and up to standard quality in final food products. In case there is any deviation in food quality, they have to investigate the causes of such deviation and also devise approaches to prevent similar deviations from happening in the future.

The system designed in this research could be used by food quality managers in problem investigation and prevention. Managers can explore causal factors from existing data when a problem occurs in FSCN but there is no existing knowledge or even hypothesis on the cause. Data analysis is complementary to other investigation methods (e.g. field investigation). By analyzing the collected data about daily operations on the whole chain level, managers can quickly get an indication about what might be the potential causal factors. They could verify such hypotheses in practice by visiting and investigating the related stages in FSCN, such as factories or farms. With this system, managers can also verify the hypotheses on causes of a problem in FSCN. Since food quality managers are involved in the actual practice of FSCN, they might have hypotheses on the causes when a problem occurs. If the manager analyzes historical data and finds the same causal factor, then he/she has more confidence in the validity of a hypothesis.

Information technology managers

Information technology managers in FSCN design and implement information systems to support management planning and control. The monitoring systems in FSCN collect and store information about the daily operations at different stages. Information technology managers are concerned with how to correctly obtain data from different stages of FSCN and organize it in such a way that other managers can easily understand it and find out what they need.

For information technology managers, EWPC systems can enforce correct and continuous data collection mechanisms. High-quality data is a necessary condition for obtaining correct results from data analysis. But in practice, data quality is always an issue. Missing values, outliers, and wrong information can easily be found in various data sets. In EWPC systems, the components for handling outliers and missing values enable managers to easily spot problems in collected data. Information technology managers will gain insights into where the data problems come from and improve the data collection mechanism accordingly.

Summary

Food quality problems in Food Supply Chain Networks (FSCN) have not only brought losses to the food industry, but also risks to the health of consumers. In current FSCN, Information Systems are widely used. Those information systems contain the data about various aspects of food production (e.g. primary inputs, operations) in different stages of FSCN. By applying Data Mining (DM) methods on those data sets, managers can identify the causes of encountered new problems, and also predict and prevent those problems. However, managers are often non-experts in the DM area. In this research, a framework for Early Warning and Proactive Control (EWPC) systems has been designed, and a prototype system according to this framework has been implemented. Such systems can enable managers to employ the power of DM methods to predict and prevent encountered problems. Moreover, such systems enable managers to accumulate the knowledge they obtain from data analysis into a Knowledge Base, so that other managers can use it when they encounter similar types of problems. In this research, we have two major objectives:

Objective 1.

To design a framework for EWPC systems to facilitate the following aspects:

- analyze relations between problems and causes
- predict upcoming problems
- suggest control actions to prevent upcoming problems
- use existing databases in FSCN
- support non-expert users in applying DM methods
- have an extendable knowledge base

The framework should describe the necessary components as well as the relations between those components in EWPC systems.

Objective 2.

To build a prototype system based on the framework to enable managers in FSCN, as non-experts in DM, to use DM methods for Early Warning and Proactive Control on the supply chain level.

Research questions

In order to realize those objectives, six research questions were formulated.

1. What are the requirements for EWPC system design considering current practice of FSCN management?

2. What components should be included in the EWPC systems, and how should those components cooperate to enable managers to achieve EWPC in FSCN?

3. What Data Mining methods are available and applicable for EWPC in FSCN?

4. What support needs to be provided to managers in order to enable them to use Data Mining methods for EWPC?

5. What kind of structure is suitable for the Knowledge Base in EWPC systems?

6. What is the validity of the designed framework and prototype system?

To answer these questions, we used both literature review and case analysis. We studied the literature from areas such as Decision Support Systems, Data Mining, Supply Chain Management, Ontology Engineering, and Knowledge Engineering. The cases we analyzed came from two food companies. From the cases in those companies we studied what kind of system would enable managers to realize EWPC in FSCN. During case analysis, we communicated with managers in those companies about the problems they encountered, the relevant data sets, and the objectives they wanted to achieve. The data sets obtained from those cases normally have more than ten fields and millions of records. By applying different DM methods on those cases, we accumulated knowledge on the applicability of those methods as well as on the generic processes of applying those methods for EWPC. Moreover, we categorized the types of knowledge obtained from problem investigation in order to design a proper structure for Knowledge Base.

Regarding the first research question, *what are the requirements for EWPC system design considering current practice of FSCN management?* our study distinguished three types of requirements: performance requirements concerning the time needed for using this system, specific quality requirements concerning the sufficiency and comprehensibility of the assistance this system can offer, and functional requirements. There are six functional requirements:

1) facilitate quantitatively formulating problems;
2) guiding data joining and data preparation;
3) guiding managers in using DM methods for quantitative modeling;
4) predict the problem as early as possible;
5) support evaluating different control measures;
6) provide relevant knowledge for encountered problems, and accommodate new knowledge obtained during problem solving and decision making.

Regarding the second research question, *what components should be included in the EWPC systems, and how should those components cooperate to enable managers to achieve EWPC in FSCN?* our study defined the following major components for the framework:

- Task classifier and Template Approaches: to direct managers to follow the correct processes when they intend to deal with the encountered problem. Task classifier helps users to quickly identify their task type: identifying a problem, finding relevant data, exploring potential causal factors for the problem, predicting upcoming problems, evaluate alternative control measures, and consulting the Knowledge Base. Each task is supported by a corresponding Template approach.
- Knowledge Base: stores information (e.g. causal factors, causal relations) on previously encountered problems in FSCN for easy reference by other users.

- DM methods library and Expert System: the DM methods library stores information (function, model format, and requirements on data sets) about the DM methods that can be used for EWPC. The Expert System gives suggestions on which DM methods to use and explain its reasoning.
- Explorer and Predictor: the Explorer component allows users to explore potential causal factors for the problems in FSCN. The Predictor warns about problems that are about to occur in FSCN. It is used for decision evaluation as well. Users can employ models built previously to compare results of different available decisions and choose the best one.

In addition to the specification of those components, we also defined the steps that are needed for using the system, as well as the correct sequence between those steps.

Regarding the third research question: w*hat Data Mining methods are available and applicable for EWPC in FSCN?* our study identified six requirements on the DM method level: prediction, problem detection, finding determinant factors, representing complex structure, different representation forms, and extendable with new knowledge. The first four functional requirements deal with functions of DM methods. In the DM area functions are categorized differently (such as classification, regression). Our study provided a mapping between these two kinds of functions. After that, we selected a list of widely used DM methods and identified which method can accomplish which DM function. The last two functional requirements relate to the representation form of DM methods. Our study provided another mapping between representation forms of those DM methods and their extensibility for new knowledge.

Regarding the fourth research question: *what support needs to be provided to managers in order to enable them to use Data Mining methods for EWPC?* our study found two aspects of support that are needed for enabling managers to use DM methods. One is how to find a proper DM method. This can be supported with an Expert System for DM method selection and a DM methods library. Managers can get suggestions on which DM method is proper after they specify their case situation and data set characteristics to the Expert System. The other is how to use the DM method found for EWPC. This is supported with Template approaches for data analysis. Those template approaches tell users how to execute the particular step, what performance indicator to look at, and what to do if a particular situation occurs.

Regarding the fifth research question: *what kind of structure is suitable for the Knowledge Base in EWPC systems?* our study defined a structure with two parts: a rule base and an inference structure. A rule base allows managers to store obtained knowledge. It allows managers to specify what kind of causal relation and/or remedies have been obtained. A rule base should contain an ontology that guarantees the consistent semantic meaning of terms in each rule. An inference structure allows managers to quickly identify relevant knowledge. It communicates with users, and uses the inference mechanism to find out applicable knowledge.

Regarding the sixth research question: *What is the validity of the designed framework and prototype system?* our study first assigned appropriate Key Performance Indicators (KPI) for different design aspects of the system (e.g. framework, design methodology), then employed Expert Validation to evaluate the performance of each design aspect on each KPI. Our expectations are roughly met by the results of expert validation. Expert Validation also brought forward that experts in FSCN management put special importance on the potential of the developed system.

The main contribution of this research is that it integrated different scientific areas into one decision support architecture. This architecture presents a new means to monitor and control the processes for Supply Chain Management. Managers with EWPC can handle new problems with existing data resources. The scope of problems that can be solved is not restricted beforehand. For Data Mining, this study extends the existing research on applicability of DM methods by creating a bridge between the reality needs and the DM area. For Knowledge Engineering, this study creates a suitable Knowledge Base structure for sharing knowledge among non-experts users by linking Knowledge Management with Ontology Engineering.

As far as the impact of this research for supply chain managers is concerned, we advise them to ensure that the requirements for effective proactive control are fulfilled. The framework presented in this thesis supports supply chain managers by providing them with usable DM methods for obtaining new insights through modelling and application of new data. For food quality managers in FSCN, the implication is that a EWPC system can be used to explore causal factors when a problem occurs. Food quality managers can also verify the hypothesis on the causes with the EWPC system. By using this system together with other problem investigation strategies, such as field investigation, managers can improve the efficiency and effectiveness of problem solving. For information technology managers, we advise them to use such a system to enforce correct and continuous data collection mechanism. In this system, the facilities for handling outliers and missing values can enable managers to easily identify problems in collected data. FSCN managers from practice recognize the potential of the system and knowledge stored in it for improving decision support by making Data Mining applicable for non-experts.

Samenvatting

Problemen met voedselkwaliteit in FSCN (*Food Supply Chain Networks* ofwel netwerken van voedingsmiddelenketens) hebben niet alleen geleid tot verliezen in de voedselindustrie, maar ook gevaar opgeleverd voor de gezondheid van consumenten. In huidige FSCN zijn informatiesystemen wijdverbreid. Die informatiesystemen bevatten de gegevens over verscheidene aspecten van voedselproductie (bijv. primaire productie, bewerkingen) in verschillende stadia van de keten. Door methoden van *Data Mining* (DM) toe te passen op die gegevensverzamelingen kunnen managers oorzaken vaststellen van nieuwe problemen die zij tegenkomen en ook zulke problemen voorspellen en voorkomen. Maar managers zijn vaak geen expert op het gebied van DM. In dit onderzoek is een conceptueel raamwerk ontworpen voor EWPC-systemen voor vroegtijdige probleemdetectie en proactieve besturing (*Early Warning and Proactive Control*), en is een prototype-systeem volgens dit raamwerk geïmplementeerd. Zulke systemen kunnen managers in staat stellen om de kracht van DM-methoden te benutten om gedetecteerde problemen te voorspellen en te voorkomen. Bovendien stellen zulke systemen managers in staat om kennis die zij verkrijgen uit gegevensanalyse te verzamelen in een kennisbank (*Knowledge Base*), zodat andere managers die kunnen gebruiken wanneer zij soortgelijke problemen vaststellen. In dit onderzoek hebben we twee hoofddoelstellingen:

<u>Doelstelling 1.</u>

Een conceptueel raamwerk ontwerpen voor een EWPC-systeem dat de volgende aspecten kan faciliteren:

- analyseren van verbanden tussen problemen en oorzaken
- aanstaande problemen voorspellen
- suggesties geven voor besturingsmaatregelen om aanstaande problemen te voorkomen
- bestaande databases in FSCN gebruiken
- ondersteunen van niet-experts bij het gebruik van DM-methoden
- een uitbreidbare kennisbank (*Knowledge Base*) bevatten

Het raamwerk moet zowel de benodigde componenten beschrijven als de verbanden tussen die componenten in EWPC-systemen.

<u>Doelstelling 2.</u>

Een prototype-systeem bouwen op basis van het conceptuele raamwerk om managers in FSCN, als niet-experts in DM, in staat te stellen om DM-methoden te gebruiken voor vroegtijdige probleemdetectie en proactieve besturing (*Early Warning and Proactive Control*: EWPC) in netwerken van voedingsmiddelenketens.

<u>Onderzoeksvragen.</u>

Om deze doelstellingen te realiseren zijn zes onderzoeksvragen geformuleerd.

1. Wat zijn de vereisten voor het ontwerp van EWPC-systemen met het oog op de huidige praktijk van FSCN management?

2. Welke componenten moeten deel uitmaken van de EWPC-systemen, en hoe moeten die componenten samenwerken om het managers mogelijk te maken EWPC in FSCN te bereiken?
3. Welke methoden van *Data Mining* zijn beschikbaar en toepasbaar voor EWPC in FSCN?
4. Welke ondersteuning moet aan managers geleverd worden om hen in staat te stellen DM-methoden te gebruiken voor EPWC?
5. Wat voor soort structuur is geschikt voor de kennisbank (*Knowledge Base*) in EWPC-systemen?
6. Wat is de validiteit van het ontworpen conceptuele raamwerk en prototype-systeem?

Om deze vragen te beantwoorden gebruikten we zowel literatuurstudie als analyse van cases. We bestudeerden literatuur op gebieden als *Decision Support Systems*, *Data Mining*, *Supply Chain Management*, and *Knowledge Engineering*. De cases die we analyseerden kwamen van twee voedingsbedrijven. Uit de cases van die bedrijven bestudeerden we wat voor systeem het managers mogelijk zou maken om EPWC in FSCN te realiseren. Tijdens analyse van de cases overlegden we met managers in die bedrijven over de gedetecteerde problemen, de relevante gegevens-verzamelingen, en de doelstellingen die zij wilden bereiken. De gegevensverzamelingen uit die cases bevatten meer dan tien kenmerken en miljoenen gegevenselementen. Door verschillende DM-methoden toe te passen op die cases bouwden we kennis op, zowel over de toepasbaarheid van die methoden als over de generieke processen om die methoden toe te passen voor EWPC. Bovendien rubriceerden we de typen kennis verkregen uit de probleemverkenning om een juiste structuur voor de kennisbank (*Knowledge Base*) te ontwerpen.

Ten aanzien van de eerste onderzoeksvraag “Wat zijn de vereisten voor het ontwerp van EWPC-systemen met het oog op de huidige praktijk van FSCN management?”onderscheidde onze studie drie typen vereisten: prestatievereisten met betrekking tot benodigde tijd om het systeem te gebruiken, specifieke kwaliteitsvereisten met betrekking tot toereikendheid en begrijpelijkheid van de assistentie die het systeem kan bieden, en functievereisten. Er zijn zes functievereisten:

1) faciliteren van het kwantitatief formuleren van problemen;
2) begeleiden in het combineren van gegevens (*data joining*) en bruikbaar maken van gegevens (*data preparation*);
3) managers leiden in het gebruik van DM-methoden voor kwantitatieve modellering;
4) problemen zo vroeg mogelijk voorspellen;
5) ondersteunen van het evalueren van verschillende besturingsmaatregelen;
6) relevante kennis leveren voor tegengekomen problemen, en nieuwe kennis onderbrengen die verkregen is tijdens het oplossen van problemen en het nemen van beslissingen.

Ten aanzien van de tweede onderzoeksvraag “Welke componenten moeten deel uitmaken van de EWPC-systemen, en hoe moeten die componenten samenwerken om het managers mogelijk te maken EWPC in FSCN te bereiken?” definieerde onze studie de volgende hoofdcomponenten voor het conceptuele raamwerk:

- *Task classifier* en *Template Approaches*: om managers de weg te wijzen om de correcte processen te doorlopen wanneer zij gedetecteerde problemen willen aanpakken. *Task classifier* helpt gebruikers om snel hun type taak te herkennen: probleem herkennen, relevante gegevens zoeken, mogelijke oorzakelijke factoren voor het probleem verkennen, aanstaande problemen voorspellen, alternatieve besturingsmaatregelen evalueren, kennisbank raadplegen. Elke taak wordt ondersteund door een overeenkomstige *Template approach*.

- *Knowledge Base*: slaat informatie op (bijv. oorzakelijke factoren en oorzakelijke verbanden) over eerder tegengekomen problemen in FSCN zodat andere gebruikers deze eenvoudig kunnen raadplegen.
- *DM methods library* en *Expert System*: de *DM methods library* slaat informatie op (functie, model vorm, en vereisten aan gegevensverzamelingen) over de DM-methoden die voor EWPC gebruikt kunnen worden. Het *Expert System* geeft suggesties welke DM-methoden te gebruiken en legt uit hoe het redeneert.
- *Explorer* en *Predictor*: de component *Explorer* stelt gebruikers in staat om mogelijke oorzakelijke factoren voor de problemen in FSCN te verkennen. De *Predictor* waarschuwt voor problemen die op het punt staan in FSCN te gebeuren. De *Predictor* wordt tevens gebruikt om beslissingen te evalueren. Gebruikers kunnen eerder gebouwde modellen inzetten om resultaten van verschillende beschikbare beslissingen te vergelijken en de beste te kiezen.

In aanvulling op de specificatie van deze componenten, hebben we ook de stappen gedefinieerd die nodig zijn om het systeem te gebruiken en ook de correcte opeenvolging tussen die stappen.

Ten aanzien van de derde onderzoeksvraag “Welke methoden van *Data Mining* zijn beschikbaar en toepasbaar voor EWPC in FSCN?” onderkent onze studie zes vereisten op het niveau van de DM-methoden: voorspellen, vaststellen van problemen, bepalende factoren zoeken, complexe structuren weergeven, verschillende representatievormen, en uitbreidbaarheid met nieuwe kennis. De eerste vier functievereisten gaan over functies van DM-methoden. Op het gebied van DM worden functies anders gerubriceerd (zoals classificatie, regressie). Onze studie leverde een afbeelding op tussen deze twee soorten functies. Daarna selecteerden we een lijst van veelgebruikte DM-methoden en stelden vast welke methode welke functie tot stand kan brengen. De laatste twee functievereisten hebben betrekking op de representatievormen van die DM-methoden en hun uitbreidbaarheid met nieuwe kennis.

Ten aanzien van de vierde onderzoeksvraag “Welke ondersteuning moet aan managers geleverd worden om hen in staat te stellen DM-methoden te gebruiken voor EPWC?” vond onze studie twee aspecten van benodigde ondersteuning voor managers om DM-methoden te gebruiken. De ene is om geschikte DM-methoden te vinden. Dit wordt ondersteund met een *Expert System for DM method selection* en een *DM methods library*. Managers kunnen suggesties krijgen over welke DM-methode juist is, nadat zij een specificatie van de situatie van de case en de eigenschappen van de gegevensverzameling aan het expertsysteem bekendgemaakt hebben. De andere is hoe de gevonden DM-methode te gebruiken. Dit wordt ondersteund met *Template approaches for data analysis*. Die *template approaches* vertellen de gebruiker hoe men een bepaalde stap moet uitvoeren, welke prestatie-indicator men moet bekijken, en wat men moet doen als zich een bepaalde situatie voordoet.

Ten aanzien van de vijfde onderzoeksvraag “Wat voor soort structuur is geschikt voor de kennisbank (*Knowledge Base*) in EWPC-systemen?”definieerde onze studie een structuur met twee onderdelen: een *rule base* (lijst afleidingsregels) en een *inference structure* (structuur om gevolgtrekkingen te maken). De *rule base* stelt managers in staat om verkregen kennis op te slaan. Deze stelt managers in staat om te specificeren welk soort oorzakelijke verbanden en/of remedie zijn verkregen. Een *rule base* moet een ontologie bevatten die de consequente betekenis van termen in de regels garandeert. Een *inference structure* stelt managers in staat om snel relevante kennis te vinden. Deze legt verbinding met de gebruikers en gebruikt het inferentie-mechanisme om toepasbare kennis te achterhalen.

Ten aanzien van de zesde onderzoeksvraag “Wat is de validiteit van het ontworpen conceptuele raamwerk en prototype-systeem?” kende onze studie eerst geschikte KPI’s (*Key Performance Indicators*; kritische prestatie-indicatoren) toe voor verschillende ontwerpaspecten van het systeem (bijv. raamwerk, ontwerpmethodologie), en gebruikte toen *Expert Validation* om de prestaties van elk ontwerpaspect op elke KPI te evalueren. De uitkomsten van de *Expert Validation* kwamen in grote lijnen overeen met onze verwachtingen. Uit de *Expert Validation* komt ook naar voren dat de experts in FSCN management de potentie van het ontwikkelde systeem van bijzonder belang vinden.

De belangrijkste bijdrage van dit onderzoek is dat het verschillende wetenschapsgebieden integreerde in één architectuur voor beslissingsondersteuning (*Decision Support*). Deze architectuur presenteert een nieuwe manier om processen voor *Supply Chain Management* te bewaken en besturen. Managers met EWPC kunnen nieuwe problemen met reeds beschikbare gegevens-bestanden aanpakken. De reikwijdte van problemen die opgelost kunnen worden is niet vooraf beperkt. Voor *Data Mining* vormt deze studie een uitbreiding op bestaand onderzoek naar toepasbaarheid van DM-methoden door een brug te slaan tussen de vragen uit de realiteit en het gebied van *Data Mining*. Voor *Knowledge Engineering* schept deze studie een geschikte structuur voor een kennisbank (*Knowledge Base*) om kennis te delen tussen niet-experts door een verbinding te leggen tussen *Knowledge Management* en *Ontology Engineering*.

Waar het gaat over de impact van dit onderzoek voor supply chain managers, adviseren we hen om ervoor te zorgen dat aan de vereisten voor effectieve besturing wordt voldaan. Het in dit proefschrift gepresenteerde raamwerk ondersteunt supply chain managers door hen te voorzien van bruikbare DM-methoden om nieuwe inzichten te verkrijgen via modelvorming en gebruik van nieuwe gegevens. Voor kwaliteitsmanagers in FSCN is de implicatie dat een EWPC-systeem gebruikt kan worden om oorzakelijke factoren te verkennen wanneer een probleem optreedt. Kwaliteitsmanagers kunnen met het EWPC-systeem ook hypothesen over de oorzaken verifiëren. Door dit systeem te gebruiken samen met andere probleemverkenningsstrategieën, zoals veld-onderzoek, kunnen managers de efficiëntie en effectiviteit van het oplossen van problemen verbeteren. Informatiemanagers adviseren we om een dergelijk systeem te gebruiken om correcte en continue mechanismen voor het verzamelen van gegevens af te dwingen. De faciliteiten in dit systeem om uitschieters en ontbrekende waarden af te handelen kan managers in staat stellen om op eenvoudige wijze problemen in verzamelde gegevens te onderkennen. FSCN managers uit de praktijk onderkennen de potentie van het systeem en de daarin opgeslagen kennis om beslissingsondersteuning te verbeteren door Data Mining voor niet-experts toepasbaar te maken.

Reference

Adriaans, P. and D. Zantinge (1996). Data Mining. Harlow (UK), Addison-Wesley.

Agarwal, D. (2005). An Empirical Bayes Approach to Detect Anomalies in Dynamic Multidimensional Arrays. International Workshop on Data Mining Methods for Anomaly Detection, Chicago, Illinois, USA.

Agrawal, R., H. Mannila, R. Srikant, H. Toivonen and A. I. Verkamo (1996). Fast discovery of association rules. Advances in knowledge discovery and data mining. U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy, MIT Press**:** 307-328.

Ahn, J.-H. and K. J. Ezawa (1997). "Decision support for real-time telemarketing operations through Bayesian network learning." Decision Support Systems **21**(1): 17-27.

Akerkar, R. and P. Sajja, Eds. (2009). Knowledge-Based Systems, Jones & Bartlett Publishers.

Alexandros, K. and H. Melanie (2001). "Model Selection via Meta-learning: a Comparative Study." International Journal of Artificial Intelligence Tools **10**(4).

Aliferis, C., E. Chao and G. F. Cooper (1993). "Data explorer: a prototype expert system for statistical analysis." Proceedings of Annual Symposium on Computer Application in Medical Care: 389-93.

Anthony, R. N. (1998). Management control system, McGraw-Hill.

Antoniou, G. and F. v. Harmelen (2004). A Semantic Web Primer, The MIT press.

Aramyan, L. H. (2007). Measuring Supply Chain Performance in the Agri-Food Sector, Wageningen University. **PhD thesis**.

Attenborough, M. and K. R. Matthews (2000). "Food safety through the meat supply chain." Journal Of Applied Microbiology **88**: 144S-148S.

Balderasy, M.-A., F. Berzal, J.-C. Cubero, E. Eismany and N. Marın (2005). Discovering hidden association rules. International Workshop on Data Mining Methods for Anomaly Detection, Chicago, Illinois, USA.

Beale, R. and T. Jackson (1990). Neural Computing: An Introduction. Bristol, Adam Hilger.

Beamon, B. M. (1999). "Measuring supply chain performance." International Journal of Operations & Production Management **19**(3-4): 275-292.

Beneventano, D., M. Orsini, L. Po, A. Sala and S. Sorrentino (2009). An Ontology-Based Data Integration System for Data and Multimedia Sources. IEEE International Conference on Semantic Computing, Berkeley, CA, USA.

Bernstein, A., F. Provost and S. Hill (2005). "Toward intelligent assistance for a data mining process: An ontology-based approach for cost-sensitive classification." IEEE Transactions On Knowledge And Data Engineering **17**(4): 503-518.

Beulens, A. J. M. (2003). Transparency Requirements in Supply Chains and Networks: Yet another challenge for the Business and ICT Community. . Herausforderungen der Wirtschaftsinformatick in der Informationsgesellschaft, Leipzig, Wissenschaftsverlag Edition am Gutenbergplatz Leipzig.

Beulens, A. J. M. (2005). "Food safety and transparency in food chains and networks: relationships and challenges." Food Control **16**: 481-486.

Beulens, A. J. M. and J. A. E. E. Van Nunen (1988). "The use of expert system technology in DSS." Decision Support Systems **4**(4): 421.

Blair, A., J. Debenham and J. Edwards (1997). "A comparative study of methodologies for designing IDSSs." European Journal of Operational Research **103**(2): 277.

Bockenholt, I., M. Both and W. Gaul (1989). "A knowledge-based system for supporting data analysis problems." Decision Support Systems **5**(4): 345-354.

Borst, W. N. (1997). Construction of engineering ontologies for knowledge sharing and reuse. CTIT. Enschede, University of Twente, The Netherlands. **PhD thesis:** 227.

Breiman, L., J. H. Friedman, R. A. Olshen and C. J. Stone (1984). Classification and regression trees California, Wadsworth International Group.
Brodley, C. E. (1992). Dynamic Automatic Model Selection. COINS Technical Report 92-30, University of Massachusetts.
Chapman, P., J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer and R. Wirth. (2000). "CRISP-DM 1.0, Step-by-step data mining guide." from www.crisp-dm.org.
Chen, M. S., J. W. Han and P. S. Yu (1996). "Data mining: An overview from a database perspective." IEEE Transactions On Knowledge And Data Engineering **8**(6): 866-883.
Chen, S.-H., T.-W. Kuo and P. P. Wang (2007). Computational Intelligence in Economics and Finance : Volume II. Berlin Heidelberg, Springer-Verlag GmbH.
Chitra, D. and A.-B. Ahmad (2004). OAV-VVT Expert, An active system for Verification and Validation of Knowledge Base Based on Ex-OAV KB. ICSE & INCOSE 2004, Las Vegas, Nevada, USA.
Cody, W. F., J. T. Kreulen, V. Krishna and W. S. Spangler (2002). "The integration of business intelligence and knowledge management." IBM Systems Journal archive **41**(4): 697 - 713
Corcho, O., M. Fernandez-Lopez and A. Gomez-Perez (2003). "Methodologies, tools and languages for building ontologies. Where is their meeting point?" Data & Knowledge Engineering **46**(1): 41-64.
Costello, B., R. J. Ewen, H. Gunson, N. M. Ratcliffe, P. S. Sivanand and P. T. N. Spencer-Phillips (2003). "A prototype sensor system for the early detection of microbially linked spoilage in stored wheat grain." Measurement Science & Technology **14**(4): 397-409.
Craw, S., D. Sleeman, N. Graner, M. Rissakis and S. Sharma (1992). Consultant: Providing Advice for the Machine Learning Toolbox. Research and Development in Expert Systems IX: Proceedings of Expert Systems 92, 12th Annual Technical Conference of the British Computer Society Specialist Group on Expert Systems, Cambridge University Press**:** 5-23.
Crawford, J. and F. Crawford (1996). Data Mining in a Scientific Environment. AUUG 96 & Asia Pacific World Wide Web 2nd Joint Conference, Australia.
Cser, L. and J. Naude (2004). "Fine-tuning of the rolling mill using an Internet-based on-line application-processing interface." Proceedings Of The Institution Of Mechanical Engineers Part B-Journal Of Engineering Manufacture **218**(6): 631-640.
Davenport, T. H. (1993). Process Innovation; reengineering work through information technology, Harvard Business School Press.
Dong, X. L., A. Halevy and C. Yu (2009). "Data integration with uncertainty " The VLDB Journal **18**(2): 469-500.
Dorp, C. A. v. (2004). Reference modelling for Tracking and Tracing, Wageningen University. **Phd Thesis**.
Eason, K. (1988). Information technology and organisational change. Bristol, PA, USA Taylor & Francis/Hemisphere.
Edwards, D. (2000). Introduction to Graphical Modelling. New York, Springer-Verlag.
Fallman, D. (2007). "Why Research-oriented Design Isn't Design-oriented Research: On the Tensions between Design and Research in an Implicit Design Discipline." Journal on Knowledge, Technology and Policy **20**(3).
Fayyad, U. M., S. G. Djorgovski and N. Weir (1996). Automating the Analysis and Cataloging of Sky Surveys. Advances in Knowledge Discovery and Data Mining. U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy, AAAI/MIT Press**:** 471-493.
Fayyad, U. M., G. Piatesky-Shapiro and P. Smyth (1996). From Data Mining to Knowledge Discovery: An Overview. Advances in Knowledge Discovery and Data Mining. U. M. Fayyad, G. Piatesky-Shapiro, P. Smyth and R. Uthurusamy. Menlo Park (Ca), AAAI-Press**:** 1-37

Fayyad, U. M., G. Piatesky-Shapiro and P. Smyth (1996). From Data Mining to Knowledge Discovery: An Overview. Advances in Knowledge Discovery and Data Mining. U. M. Fayyad, G. Piatesky-Shapiro, P. Smyth and R. Uthurusamy. Menlo Park (Ca), AAAI-Press: 1-37.

Fayyad, U. M., G. Piatetsky-Shapiro and P. Smyth (1996). "The KDD process for extracting useful knowledge from volumes of data." Communications Of The ACM **39**(11): 27-34.

Fowler, A. (1999). "Feedback and feedforward as systemic frameworks for operations control." International Journal Of Operations & Production Management **19**(2): 182-204.

Freedman, D. A. (2004). "Graphical models for causation, and the identification problem." Evaluation Review **28**(4): 267-293.

Freitas, P. S. A. and A. J. L. Rodrigues (2006). "Model combination in neural-based forecasting." European Journal of Operational Research **173**(3): 801.

Gómez-Pérez, A. (1994). From knowledge Based Systems to Knowledge Sharing Technology: Evaluation and Assessment. KSL-94-73, Knowledge Systems Laboratory, Stanford University, CA.

Gestal, M., M. P. Gomez-Carracedo, J. M. Andrade, J. Dorado, E. Fernandez, D. Prada and A. Pazos (2004). "Classification of apple beverages using artificial neural networks with previous variable selection." Analytica Chimica Acta **524**(1-2): 225.

Glinz, M. (2005). Rethinking the Notion of Non-Functional Requirements. Proceedings of the Third World Congress for Software Quality 2005, Munich, Germany.

Goodwin, G. C., S. F. Graebe and M. E. Salgado (2001). Control System Design, Prentice Hall.

Gopalakrishnan, B. and M. L. McCoy (2008). "Adaptive process control for turning operation using expert systems." International Journal of Industrial and Systems Engineering **3**(6): 711 - 726.

Gorte, B. G. H. and A. Stein (1998). "Bayesian classification and class area estimation of satellite images using stratification." IEEE Transactions on Geoscience and Remote Sensing **36**(3): 803-812.

Goulermas, J. Y., D. Howard, C. J. Nester, R. K. Jones and L. Ren (2005). "Regression techniques for the prediction of lower limb kinematics." Journal Of Biomechanical Engineering-Transactions Of The Asme **127**(6): 1020-1024.

Gracia, J., M. d'Aquin and E. Mena (2009). Large scale integration of senses for the semantic web Proceedings of the 18th international conference on World wide web Madrid, Spain

Grijsen, J. G., X. C. Snoeker and C. J. M. Vermeulen (1992). An information system for flood early warning 3rd international conference on floods and flood management, Florence, Italy, Delft Hydraulics Publication

Gruber, T. R. (1995). "Toward principles for the design of ontologies used for knowledge sharing." International Journal Of Human-Computer Studies **43**(5-6): 907-928.

Haimila, S. (2001). "KM in practice: the helping hand of BI." KMWorld **10**(10).

Han, J. and M. Kamber (2001). Data mining: concepts and techniques, Morgan Kaufmann Publishers.

Hansen, M. H. and B. Yu (2001). "Model selection and the principle of minimum description length." Journal Of The American Statistical Association **96**(454): 746-774.

Harrison, S. R. (1991). "Validation of agricultural expert systems." Agricultural Systems **35**(3): 265-285.

Haykin, S. (1999). Neural networks : a comprehensive foundation, London [etc.] : Prentice Hall.

Heckerman, D. and J. S. Breese (1996). "Causal independence for probability assessment and inference using Bayesian networks." Ieee Transactions On Systems Man And Cybernetics Part A-Systems And Humans **26**(6): 826-831.

Herschel, R. T. and N. E. Jones (2005). "Knowledge management and business intelligence: the importance of integration." Journal of Knowledge Management **9**(4): 45 - 55.

Hevner, A. R., S. T. March, J. Park and S. Ram (2004). "Design Science in Information Systems Research." Management Information Systems Quaterly **28**(1): 75-106

Hoehndorf, R., J. Bacher, M. Backhaus, S. E. G. Jr, F. Loebe, K. Prufer, A. Uciteli, J. Visagie, H. Herre and J. Kelso (2009). BOWiki: an ontology-based wiki for annotation of data and integration of knowledge in biology. Proceedings of the Bio-Ontologies Special Interest Group Workshop: Knowledge in Biology.

Hofstede, G. J., A. J. M. Beulens and L. Spaans-Dijkstra (2004). Transparency: Perceptions, Practices and Promises. The emerging World of Chains and Networks: Bridging Theory and Practice. T. Camps, P. Diederen, G. J. Hofstede and B. Vos, Reed Business Information.

Hofstede, G. J., H. Schepers, L. Spaans-Dijkstra, J. Trienekens and A. Beulens (2004). Hide or confide? : the dilemma of transparency. 's-Gravenhage, Reed Business Information.

Holsapple, C. W. and A. B. Whinston (2000). Decision Support Systems: A Knowledge-based Approach, West Publishing Company, Minneapolis/St Paul, .

Houten, S.-P. A. v. (2007). A suite for developing and using business games, TU Delft. **PhD Thesis**.

Hsu, Y. C., A. P. Chen and C. H. Wang (2008). A RFID-enabled traceability system for the supply chain of live fish. IEEE International Conference on Automation and Logistics, Qingdao, China.

Ince, H. and T. B. Trafalis (2006). "A hybrid model for exchange rate prediction." Decision Support Systems **42**(2): 1054.

ISO (2001). ISO 9126-1:2001, Software engineering - Product quality, International Organization for Standardization, Part 1: Quality model.

Jackson, M. C. (1993). Systems methodology for the management sciences, New York: Plenum Press.

Jansen, M. H. (1998). The environment as a resource: developing environmental information systems based on enterprise resource planning software, Technical University of Eindhoven. **PhD thesis**.

Jedermann, R. and W. Lang (2007). "Semi-passive RFID and beyond: steps towards automated quality tracing in the food chain." International Journal of Radio Frequency Identification Technology and Applications **1**(3): 247 - 259.

Jonker, J.-J., N. Piersma and R. Potharst (2006). "A decision support system for direct mailing decisions." Decision Support Systems **42**(2): 915.

Keen, P. G. W. and M. S. S. Morton (1978). Decision Support Systems: An Organizational Perspective, Addison-Wesley.

Kelepouris, T., K. Pramatari and G. Doukidis (2007). "RFID-enabled traceability in the food supply chain." Industrial Management & Data Systems **107** (2): 183 - 200.

Kim, Y. and W. N. Street (2004). "An intelligent system for customer targeting: a data mining approach." Decision Support Systems **37**(2): 215.

Klein, M. R. and L. B. Methlie (1995). Knowledge based decision support systems: with applications in business. Chichester, John Wiley & Sons

Knorr, E., R. Ng and V. Tucakov. (2000). "Distance-Based Outlier: algorithms and applications." VLDB Journal **8**(3-4): 237-253.

Knublauch, H., R. W. Fergerson, N. F. Noy and M. A. Musen (2004). The Protégé OWL Plugin: An Open Development Environment for Semantic Web Applications. Third International Semantic Web Conference, Hiroshima, Japan.

Kodratoff, Y. and V. Moustakis (1994). "Managing Machine Learning Application Development and Organizational Implementation." Tutorial notes, ECAI.

Kohavi, R. and G. H. John (1997). "Wrappers for feature subset selection." Artificial Intelligence **97**(1-2): 273-324.

Kumar, V. (1998). "An early warning system for agricultural drought in an arid region using limited data." Journal of Arid Environment **40**: 199-209.

Kumar, V. (1998). "An early warning system for agricultural drought in an arid region using limited data." Journal of Arid Environments **40**: 199-209.

Kurgan, L. A. and P. Musilek (2006). "A survey of knowledge discovery and data mining process models." Knowledge Engineering Review **21**(1): 1-24.

Larose, D. T. (2005). Discovering knowledge in data: an introduction to data mining. New Jersey, John Wiley & Sons.

Lattin, J. M., P. E. Green and J. D. Carroll (2003). Analyzing Multivariate Data Duxbury Press.

Lee, P. G. (2000). "Process control and artificial intelligence software for aquaculture." Aquacultural Engineering **23**(1-3): 13-36.

Leeuw, A. C. J. d. (2000). Bedrijfskundig management - Primair proces, strategie en organisatie. Assen, Van Gorcum & Comp. B.V.

Leeuw, A. C. J. d. (2002). Organisaties: management, analyse, ontwerp en verandering, Assen [etc.] : Van Gorcum.

Levinson, D. R. (2009). Traceability in the Food Supply Chain, Office of Inspector General.

Li, Y., M. R. Kramer, A. J. M. Beulens and J. G. A. J. v. d. Vorst (2006). Applying Data Mining for Early Warning in Food Supply Networks. 18th Belgium-Netherlands Conference on Artificial Intelligence, Oct. 2006, Namur, Belgium.

Li, Y., M. R. Kramer, A. J. M. Beulens and J. G. A. J. v. d. Vorst (2006). Using data mining to improve operations management in food supply networks. International agri-food chains and networks: management and organization. J. Bijman, S. W. F. Omta, J. H. Trienekens, J. H. M. Wijnands and E. F. M. Wubben, Wageningen Academic Publishers**:** 163-177.

Li, Y., M. R. Kramer, A. J. M. Beulens and J. G. A. J. v. d. Vorst (2006a). Using data mining to improve operations management in food supply networks. International agri-food chains and networks: management and organization. J. Bijman, S. W. F. Omta, J. H. Trienekens, J. H. M. Wijnands and E. F. M. Wubben, Wageningen Academic Publishers**:** 163-177.

Li, Y., M. R. Kramer, A. J. M. Beulens and J. G. A. J. v. d. Vorst (2006b). Applying Data Mining for Early Warning in Food Supply Networks. 18th Belgium-Netherlands Conference on Artificial Intelligence, Namur, Belgium.

Liao, S.-H., Y.-N. Chen and Y.-Y. Tseng (2009). "Mining demand chain knowledge of life insurance market for new product development " Expert Systems with Applications **36**(5): 9422-9437.

Lin, C. and P.-J. Hsieh (2004). "A fuzzy decision support system for strategic portfolio management." Decision Support Systems **38**(3): 383.

Liu, H. and H. Motoda (1998). Feature Selection for Knowledge Discovery and Data Mining, Kluwer Academic Publishers.

Lobell, D. B., J. I. Ortiz-Monasterio, G. P. Asner, R. L. Naylor and W. P. Falcon (2005). "Combining field surveys, remote sensing, and regression trees to understand yield variations in an irrigated wheat landscape." Agronomy Journal **97**(1): 241-249.

Lohman, C., L. Fortuin and M. Wouters (2004). "Designing a performance measurement system: A case study." European Journal Of Operational Research **156**(2): 267-286.

Mannino, M. V. and Z. Walter (2006). "A framework for data warehouse refresh policies." Decision Support Systems **42**(1): 121.

March, S. T. and G. Smith (1995). "Design and Natural Science Research on Information Technology." Decision Support Systems: 251-266.

McCall, J. A., P. K. Richards and G. F. Walters (1977). Factors in Software Quality. Volume I. Concepts and Definitions of Software Quality. Springfield, Va., National Techical Information Service.

Mendoza, J. E., A. L. Medaglia and N. Velasco (2009). "An evolutionary-based decision support system for vehicle routing: The case of a public utility " Decision Support Systems **46**(3): 730-742

Meulen, B. V. D. (2008). European Food Law Handbook, Wageningen Academic Publishers.
Meulman, J. J., A. J. V. d. Kooij and W. J. Heiser (2004). Principal Components Analysis With Nonlinear Optimal Scaling Transformations for Ordinal and Nominal Data. The Sage handbook of quantitative methodology for the social sciences. D. Kaplan. Calif. : Sage, Thousand Oaks**:** 49-70.
Michie, D., D. J. Spiegelhalter and C. C. Taylor (1994). Machine Learning, Neural and Statistical Classification. New York, McGraw-Hill.
Millan, F. R., V. Roa and M. S. Tapia (2001). "Mathematical modeling of food moisture adsorption isoterms using artificial neural networks." Interciencia **26**(5): 190-+.
Mintzberg, H. (1973). The Nature of Managerial Work, New York [etc.] : Harper & Row.
Mizoguchi, R. and M. Ikeda (1996). Towards Ontology Engineering. Technical Report AI-TR-96-1, I.S.I.R., Osaka Univ.
Mockler, R. J. (1972). The management control process, Appleton-Century-Crofts.
Moe, T. (1998). "Perspectives on traceability in food manufacture." Trends in Food Science & Technology **9**: 211-214.
Moreno, L., R. M. Aguilar, J. D. Pineiro, J. I. Estevez, J. F. Sigut and C. Gonzalez (2001). "Using KADS methodology in a simulation assisted knowledge based system: application to hospital management." Expert Systems with Applications **20**(3): 235-249.
NACMCF (1997). Hazard Analysis and Critical Control Point - Principle and Application Guidelines, National Advisory Committee on Microbiological Criteria for Foods.
Naylor, J. B., J. Griffiths and M. M. Naim (2001). "Knowledge-based system for estimating steel plant performance." International Journal Of Operations & Production Management **21**(7): 1000-1019.
Neches, R., R. Fikes, T. Finin, T. Gruber, R. Patil, T. Senator and W. R. Swartout (1991). "Enabling Technology For Knowledge Sharing." AI Magazine **12**(3).
Nemati, H. R., D. M. Steiger, L. S. Iyer and R. T. Herschel (2002). "Knowledge warehouse: an architectural integration of knowledge management, decision support, artificial intelligence and data warehousing." Decision Support Systems **33**(2): 143.
Ni, H. and S. Gunasekaran (1998). "Food quality prediction with neural networks." Food Technology **5**: 60-65.
Nielsen, J. (1993). Usability Engineering. San Francisco, Morgan Kaufmann.
Okeefe, R. M. and A. D. Preece (1996). "The development, validation and implementation of knowledge-based systems." European Journal Of Operational Research **92**(3): 458-473.
Ondersteijn, C. J. M., J. H. M. Wijnands, R. B. M. Huirne and O. van Kooten (2006). Quantifying the Agri-Food Supply Chain, Springer.
Ozden Gur Ali, F. and W. A. Wallace (1997). "Bridging the gap between business objectives and parameters of data mining algorithms." Decision Support Systems **21**(1): 3.
Pérez, A. M. G. and M. G. Martínez (2007). From Supply Chain Network Orientation to Supply Chain Network Management. Mediterranean Conference of Agro-Food Social Scientists, Barcelona, Spain.
Papakiriakopoulos, D., K. Pramatari and G. Doukidis (2009). "A decision support system for detecting products missing from the shelf based on heuristic rules " Decision Support Systems **46**(3): 685-694
Parmenter, D., Ed. (2007). Key performance indicators: developing, implementing, and using winning KPIs, Wiley
Patel, D., E. R. Davies and I. Hannah (1995). "Towards a breakthrough in the detection of contaminants in food products." Sensor Review **15**(2): 27-28.
Pearl, J. (2000). Causality: Models, Reasoning, and Inference Cambridge University Press
Petersen, B., S. Knura-Deszczka, E. Ponsgen-Schmidt and S. Gymnich (2002). "Computerised food safety monitoring in animal production." Livestock Production Science **76**: 207-213.

Pfahringer, B., H. Bensusan and C. Giraud-Carrier (2000). Meta-learning by landmarking various learning algorithms. Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000). San Francisco, California, Morgan Kaufmann**:** 743--750.

Pinto, H. S. and J. P. Martins (2004). "Ontologies: How can they be built?" Knowledge And Information Systems **6**(4): 441-464.

Poppe, C. and U. Kjærnes (2003). TRUST IN FOOD IN EUROPE - A Comparative Analysis. Professional Report No. 5. Oslo, NATIONAL INSTITUTE FOR CONSUMER RESEARCH.

Power, D. J. (1997). "What is a DSS?" The On-Line Executive Journal for Data-Intensive Decision Support **1**(3).

Power, D. J. and R. Sharda (2007). "Model-driven decision support systems: Concepts and research directions." Decision Support Systems **43**(3): 1044-1061.

Prasad, R., K. R. Ranjan and A. K. Sinha (2006). "AMRAPALIKA: An expert system for the diagnosis of pests, diseases, and disorders in Indian mango." Knowledge-Based Systems **19**(1): 9.

Quinlan, J. R. (1992). C4.5: Programs for Machine Learning San Francisco,CA, Morgan Kaufmann

Rajan, J. and V. Saravanan (2008). A Framework of an Automated Data Mining System Using Autonomous Intelligent Agents. International Conference on Computer Science and Information Technology.

Rao, L., L. Owen and D. Goldsman (1998). Development and application of a validation framework for traffic simulation models. Proceedings of the 1998 Winter Simulation Conference.

Recio, B., J. Ibanez, F. Rubio and J. A. Criado (2005). "A decision support system for analysing the impact of water restriction policies." Decision Support Systems **39**(3): 385.

Recio, B., F. Rubio and J. A. Criado (2003). "A decision support system for farm planning using AgriSupport II." Decision Support Systems **36**(2): 189.

Roos, T., H. Wettig, P. Grunwald, P. Myllymaki and H. Tirri (2005). "On discriminative Bayesian network classifiers and logistic regression." Machine Learning **59**(3): 267-296.

Rose, N., J. P. Mariani, P. Drouin, J. Y. Toux, V. Rose and P. Colin (2003). "A decision-support system for Salmonella in broiler-chicken flocks." Preventive Veterinary Medicine **59**(1-2): 27-42

Roth, A. V., A. A. Tsay, M. E. Pullman and J. V. Gray (2008). "Unraveling The Food Supply Chain: Strategic Insights From China And The 2007 Recalls." The Journal of Supply Chain Management **44**: 22-39.

Rousu, J., L. Flander, M. Suutarinen, K. Autio, P. Kontkanen and A. Rantanen (2003). "Novel computational tools in bakery process data analysis: a comparative study." Journal Of Food Engineering **57**(1): 45-56.

Rowe, G. and G. Wright (1999). "The Delphi technique as a forecasting tool: issues and analysis." International Journal of Forecasting **15**(4): 353-375.

Santoso, T., S. Ahmed, M. Goetschalckx and A. Shapiro (2005). "A stochastic programming approach for supply chain network design under uncertainty." European Journal of Operational Research **167**(1): 96-115.

SAS Institute (2004). Getting Started with SAS Enterprise Miner 4.3. Cary, North Carolina, SAS Institute Inc.

Scholten, H., A. Kassahun, J. C. Refsgaard, T. Kargas, C. Gavardinas and A. J. M. Beulens (2007). "A methodology to support multidisciplinary model-based water management." Environmental Modelling & Software **22**(5): 743-759.

Schreiber, G., H. Akkermans, A. Anjewierden, R. d. Hoog, N. Shadbolt, W. V. d. Velde and B. Wielinga (2000). Knowledge Engineering and Management: The CommonKADS Methodology, MIT Press.

Shaalan, K., M. Rizk, Y. Abdelhamid and R. Bahgat (2004). "An expert system for the best weight distribution on ferryboats." Expert Systems with Applications **26**(3): 397-411.
Shang, J., P. R. Tadikamalla, L. J. Kirsch and L. Brown (2008). "A decision support system for managing inventory at GlaxoSmithKline." Decision Support Systems **46**(1): 1-13.
Shaw, M. J., C. Subramaniam, G. W. Tan and M. E. Welge (2001). "Knowledge management and data mining for marketing." Decision Support Systems **31**(1): 127-137.
Shim, J. P., M. Warkentin, J. F. Courtney, D. J. Power, R. Sharda and C. Carlsson (2002). "Past, present, and future of decision support technology." Decision Support Systems **33**(2): 111.
Sigut, J., J. Pineiro, E. Gonzalez and J. Torres (2007). "An expert system for supervised classifier design: Application to Alzheimer diagnosis." Expert Systems with Applications **32**(3): 927-938.
Silverstein, C., S. Brin, R. Motwani and J. Ullman (2000). "Scalable techniques for mining causal structures." Data Mining And Knowledge Discovery **4**(2-3): 163-192.
Simon, H. A., G. B. Dantzig, R. Hogarth, C. R. Plott, H. Raiffa, T. C. Schelling, K. A. Shepsle, R. Thaler, A. Tversky and S. Winter (1987). "Decision-Making And Problem-Solving." Interfaces **17**(5): 11-31.
Simoudis, E. (1996). "Reality check for data mining." IEEE Expert: 26-33.
Slack, N., S. Chambers, C. Harland, A. Harrison and R. Johnston (1998). Operation managements. London, Pitman Publishing.
Smyth, P. and R. M. Goodman (1992). "An information theoretic approach to rule induction from databases." IEEE Trans. Knowledge and Data Engineering **4**(4): 301-316.
Soeria-Atmadja, D., A. Zorzet, M. G. Gustafsson and U. Hammerling (2004). "Statistical Evaluation of Local Alignment Features Predicting Allergenicity Using Supervised Classification Algorithms." International Archives of Allergy and Immunology **133**: 101-112.
Spirtes, P., C. Glymour and R. Scheines (2000). Causation, Prediction and Search, MIT Press.
Sprague, R. H. (1980). "A Framework for the Development of Decision Support Systems." MIS Quarterly **4**(4): 1-26.
Sprague, R. H. and E. D. Carlson (1982). Building effective decision support systems, Englewood Cliffs : Prentice-Hall.
SPSS (2003). Clementine 8.0 User's Guide, Integral Solutions Limited.
Trienekens, J. H. and A. J. M. Beulens (2001). The implications of EU food safety legislation and consumer demands on supply chain information systems. 11th Annual World Food and Agribusiness Forum, Sydney, Australia, International Food and Agribusiness Management Association.
Turban, E. (1995). Decision support and expert systems: management support systems. Englewood Cliffs, Prentice Hall.
Turban, E., D. Leidner, E. McLean and J. Wetherbe (2008). Information Technology for Management: Transforming Organizations in the Digital Economy John Wiley&Sons, Inc.
Uschold, M. and M. Gruninger (1996). "Ontologies: Principles, methods and applications." Knowledge Engineering Review **11**(2): 93-136.
Vaikousi, H., C. G. Biliaderis and K. P. Koutsoumanis (2008). "Development of a Microbial Time/Temperature Indicator Prototype for Monitoring the Microbiological Quality of Chilled Foods." Applied and Environmental Microbiology **74**(10): 3242-3250.
Vallejo-Cordoba, B., G. E. Arteaga and S. Nakai (1995). "Predicting milk shelf-life based on artificial neural networks and head-space gas chromatographic data." Journal of Food Science **60**: 885-888.
Van Aken, J. E. (2004). "Management Research Based on the Paradigm of the Design Sciences: The Quest for Field-Tested and Grounded Technological Rules." Journal of Management Studies **41**(2): 219-246.

Van der Vorst, J. G. A. J. (2000). Effective Food Supply Chains - Generating, Modelling and Evaluating Supply Chain Scenarios Wageningen University. **Ph.D. thesis**.

Van der Vorst, J. G. A. J. (2004). Supply Chain Management: Theory and Practices. The emerging World of Chains and Networks: Bridging Theory and Practice. T. Camps, P. Diederen, G. J. Hofstede and B. Vos, Reed Business Information.

Van der Vorst, J. G. A. J. (2005). Performance measurement in agrifood supply chain networks: an overview. Quantifying Supply Chain Performance. C. Ondersteijn. Dordrecht, Kluwer.

Van der Vorst, J. G. A. J., A. J. M. Beulens and P. van Beek (2005). Innovations in logistics and ICT in food supply chain networks. Innovation in Agri-food Systems - Product Quality and Consumer Acceptance. W. M. F. Jongen and M. T. G. Meulenberg, Wageningen Academic Publishers, The Netherlands 245-292.

Van Gestel, T., B. Baesens, P. Van Dijcke, J. Garcia, J. A. K. Suykens and J. Vanthienen (2006). "A process model to develop an internal rating system: Sovereign credit ratings." Decision Support Systems **42**(2): 1131.

Verdenius, F. (2004). Methodological Aspects of Designing Induction-based Applications, Wageningen University. **PhD thesis**.

Verdenius, F. (2005). Methodological Aspects of Designing Induction-based Applications, Universiteit van Amsterdam. **PhD thesis:** 160.

Verschuren, P. and R. Hartog (2005). "Evaluation in Design-Oriented Research." Quality & Quantity **39**: 733-762

Weiss, S. M. and N. Indurkhya (1998). Predictive data mining: a practical guide. San Francisco, Morgan Kaufmann Publishers.

Wettschereck, D., D. W. Aha and T. Mohri (1997). "A Review and Empirical Evaluation of Feature Weighting Methods for a Class of Lazy Learning Algorithms." Artificial Intelligence Review **11**(1-5): 273 - 314

Wilson, B. W. (1993). Systems: concepts, methodologies and applications, John Wiley & Sons.

Windrum, P., Fagiolo, Giorgio and Moneta, Alessio (2007). "Empirical Validation of Agent-Based Models: Alternatives and Prospects " Journal of Artificial Societies and Social Simulation **10**(2): 8.

Winterhalder, L. and R. Kleinschmidt (2000). "Operative production planning and control as a precondition for supply chain management." Stahl Und Eisen **120**(3): 57-60.

Wirth, R., C. Shearer, U. Grimmer, T. P. Reinartz, J. Schlosser, C. Breitner, R. Engels and G. Lindner (1997). Towards Process-Oriented Tool Support for Knowledge Discovery in Databases. Principles of Data Mining and Knowledge Discovery**:** 243-253.

Wolfert, S. (2002). Sustainable Agriculture: How to make it work?, Wageningen University. **PhD Thesis**.

Wolff, J. G. (2006). "Medical diagnosis as pattern recognition in a framework of information compression by multiple alignment, unification and search." Decision Support Systems **42**(2): 608.

Wu, Y. M., J. K. Chung, T. C. Shin, N. C. Hsiao, Y. B. Tsai, W. H. K. Lee and T. L. Teng (1999). "Development of an integrated earthquake early warning system in Taiwan - Case for the Hualien area earthquakes." Terrestrial Atmospheric And Oceanic Sciences **10**(4): 719-736.

Yao, J. T. (2003). Sensitivity analysis for data mining. Proceedings of the 22nd International Conference of the North American Fuzzy Information Processing Society *NAFIPS'2003*, Chicago, Illinois.

Yin, R. K. (1994). Case Study Research; design and methods, Thousand Oaks, California: Sage Publications.

Zachary, W. (1986). "A Cognitively-Based Functional Taxonomy of Decision Support Techniques." Human-Computer Interaction **2**(1): 25-63.

Ziarko, W., R. Golan and D. Edwards (1993). An Application of Datalogic/R Knowledge Discovery Tool to Identify Strong Predictive Rules in Stock Market Data. AAAI Workshop on Knowledge Discovery in Database. Washington, D.C.

About the author

Yuan Li was born on May 1981 in Xuzhou, Jiangsu Province, China. He holds a Bachelor degree in Software Engineering from Xidian University, China, and a Master degree in Artificial Intelligence from Katholieke Universiteit Leuven which he finished 'cum laude' in 2004, majoring in Engineering and Computer Science. Thereafter, he was appointed as a PhD researcher at the Information Technology Group of Wageningen University. This PhD research aims at applying various Data Mining methods to designing systems for preventing food quality and logistics problems.

Annex to statement
Name Yuan Li
PhD student, Mansholt Graduate School of Social Sciences (MG3S)
Completed Training and Supervision Plan

MANSHOLT GRADUATE SCHOOL of Social Sciences

Description	**Institute / Department**	**Year**	**ECTS***
Courses:			
MG3S Introduction Course	Mansholt Graduate School of Social Sciences (MG3S)	2005	1
Research Methodology: Designing and Conducting a PhD Research Project	MG3S	2005	2
Techniques for Writing and Presenting a Scientific Paper	Wageningen Graduate Schools (WGS)	2005	1.2
PhD Competence Assessment	WGS	2005	0.3
Project & Time Management	WGS	2005	1.2
Quantitative Research Methods	MG3S	2005	4
Food Risk Analysis	MG3S	2006	3
Multi-Agents Systems for Natural Resources Management	MG3S	2006	2
Bayesian Methods in Theory and Practice	MG3S	2006	4.2
Uncertainty Modelling and Analysis	SENSE	2006	4
Knowledge Management and Modelling	Vrije Universiteit Amsterdam	2006	6
Ontology Engineering	Vrije Universiteit Amsterdam	2005	3
Presentations at conferences and workshops:			**3**
Mansholt Multidisciplinary seminar: MG3S PhD day		2007	
7th International Conference on Management in AgriFood Chains and Networks, Ede, The Netherlands			
18th Belgium-Netherlands Conference on Artificial Intelligence, Namur, Belgium		2006	
8th International Conference on Management in AgriFood Chains and Networks, Ede, The Netherlands		2008	
2nd International Conference on Knowledge Generation, Communication and Management, Orlando, USA		2008	
Total (minimum 30 ECTS)			**34.9**

*One ECTS on average is equivalent to 28 hours of course work

The research described in this thesis was financially supported by the Nutrace project (Senter project TSIT3014)

Financial support from Wageningen University for printing this thesis is gratefully acknowledged