WORKING PAPER
MANSHOLT GRADUATE SCHOOL

Applying Data Mining for Early Warning in Food Supply Networks

Yuan Li, Mark R. Kramer, Adrie J.M. Beulens and Jack G.A.J. van der Vorst

DISCUSSION PAPER No. 29
2006

Mansholt Graduate School

Hollandseweg 1, 6706 KN Wageningen,
The Netherlands
Phone: +31 317 48 41 26
Fax: +31 317 48 47 63
Internet: http://www.mansholt.wur.nl/
e-mail: office.mansholt@wur.nl

Working Papers are interim reports on work of Mansholt Graduate School (MGS) and have received only limited reviews[1]. Each paper is refereed by one member of the Editorial Board and one member outside the board. Views or opinions expressed in them do not necessarily represent those of the Mansholt Graduate School.

The Mansholt Graduate School's researchers are based in three departments: 'Social Sciences', 'Environmental Sciences' and 'Agrotechnology and Food sciences' and two institutes: 'LEI, Agricultural Economics Research Institute' and 'Alterra, Research Institute for the Green World'. In total Mansholt Graduate School comprises about 250 researchers.

Mansholt Graduate School is specialised in social scientific analyses of the rural areas and the agri- and food chains. The Graduate School is known for its disciplinary and interdisciplinary work on theoretical and empirical issues concerning the transformation of agriculture, rural areas and chains towards multifunctionality and sustainability.

Comments on the Working Papers are welcome and should be addressed directly to the author(s).

[1] Working papers may have been submitted to other journals and have entered a journal's review process. Should the journal decide to publish the article the paper no longer will have the status of a Mansholt Working Paper and will be withdrawn from the Mansholt Graduate School's website. From then on a link will be made to the journal in question referring to the published work and its proper citation.

Applying Data Mining for Early Warning in Food Supply Networks

Yuan Li

PhD Candidate
Information Technology Group
Wageningen University
Hollandseweg 1
6706 KN Wageningen, The Netherlands
Phone: +31 317 483533, Fax: +31 317 4 85646
email: Yuan.Li@wur.nl

Mark R. Kramer

Assistant Professor
Information Technology Group
Wageningen University
Hollandseweg 1
6706 KN Wageningen, The Netherlands
Phone: +31 317 4 84346, Fax: +31 317 4 85646
email: Mark.Kramer@wur.nl

Adrie J.M. Beulens

Professor
Information Technology Group
Wageningen University
Hollandseweg 1
6706 KN Wageningen, The Netherlands
Phone: +31 317 4 84460, Fax: +31 317 4 85646
email: Adrie.Beulens@wur.nl

Jack G.A.J. van der Vorst

Professor
Operations Research and Logistics Group
Wageningen University
Hollandseweg 1
6706 KN Wageningen, The Netherlands
Phone: +31 317 482387, Fax: +31 317 4 85646
email: Jack.vanderVorst@wur.nl

Applying Data Mining for Early Warning in Food Supply Networks

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 paper investigates the requirements on data mining posed by early warning in food supply networks, and maps those requirements to available data mining methods. Results of a preliminary case study show that data mining is a promising approach as part of early warning systems in food supply networks.

Keywords: data mining, requirements, early warning, food supply networks

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 (Berry and Linoff 1997). One of the reasons is that historically food supply networks were less automated than other businesses. However, in recent years, the 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.

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 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 food quality problems.

2 Food Supply Networks

According to Santoso 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 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.

Food supply networks have specific characteristics compared to other supply networks (Van der Vorst, Beulens and van Beek 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 very 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.

2.1 Information Systems in Food Supply Networks

In food supply networks, the use of information systems has increased over the last decades. One reason is that the need for transparency in food supply networks has increased (Hofstede, Beulens and Spaans-Dijkstra 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 (Barendsz 1998), and HACCP (Horchner et al. 2006) 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.

3 Data Mining

According to Simoudis (1996), 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. Nowadays the term data mining is more and more regarded as equivalent to KDD (Knowledge Discovery in Data) (Han and Kamber 2001). So, in this paper we will treat data mining as another term for KDD.

There are two approaches for data mining: verification driven, with the aim to confirm a hypothesis, or discovery driven, which is the automatic discovery of information by the use of appropriate tools. The underlying principle of the conventional scientific cycle “observation-hypothesis-experiment” fits well with the processes of Data Mining. The discovery driven mining works well for the observation-hypothesis step, whereas the verification driven mining works well for the hypothesis-experiment step (Crawford and Crawford 1996).

Application of data mining in food supply networks is cheap and flexible when domain knowledge is scarce (Verdenius and Hunter 2001). For many food quality problems, detailed domain knowledge is scarcer than exemplary data. We can only observe the inputs and outputs of processes, without extensive knowledge about internal functions. Consequently collecting data in those cases is less costly than acquiring the required process knowledge. Of course, a basic understanding is needed. Flexibility appears from the ease of adapting models. In food supply networks, food products are subject to change once there is innovation in production processes or development in raw materials. Models constructed with data mining methods can easily adapt to such evolving behavior.

Even when domain knowledge is available, data mining is still helpful for validation of models derived from domain knowledge. When a derived model performs considerably worse than data mining methods, its validity can be questioned.

Existing applications of data mining for food quality problems in food supply networks mainly consider specific tasks at individual stages. For example, neural networks have been applied for classification of apple beverages (Gestal et al. 2004), evaluation of meat quality (Whittaker et al. 1991), and prediction of milk shelf-life (Vallejo-Cordoba, Arteaga and Nakai 1995). However, in our research we are aiming at multiple stages of the network. As discussed above, various influential factors for food quality originate from multiple stages of food supply networks. Thus, in order to construct early warning systems for food quality problems, it is necessary to include multiple stages of food supply networks.

4 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 and Vermeulen 1992) or earthquakes (Wu et al. 1999), and help to prevent or reduce their impact. In food technology, Costello 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 2). The main distinguishing attribute of our approach is the aim for adaptivity. Other early warning systems are designed only predicting for specific predefined problems. Due to uncertainty in food supply networks, we do not have predefined knowledge about all kinds of problems that could occur.

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. So it will be helpful not only for us to construct early warning systems, but also for other stakeholders to deal with similar problems.

5 Requirements

5.1 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. In order to realize early warning in food supply networks, we defined the following essential functional requirements. These requirements are derived from existing early warning systems discussed above and from 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.
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.
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.

4. *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.
5. *Different representation forms*: Many different kinds of relations between factors may exist. So early warning systems require the possibility to show different representation forms.
6. *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).

The whole process is iterative. 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.

5.2 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 et al. (1996) and Freitas (1997). 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 and Nakai 1995; Ni and Gunasekaran 1998; Rousu 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 1997). 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 and Hannah 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 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 (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 1998).

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, PiatetskyShapiro and Smyth 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. Baker et al. (2002) describe how they applied Bayesian networks to assess the risk of botulism.

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 1 and Table 2. The requirements of describing different kinds of relations and novel relation incorporation will be dealt with in following subsections.

In Table 1, we summarize the use of data mining for different functional requirements as reported in literature. In Table 2, we compare some commonly used data mining methods against the data mining functions mentioned above. We will use such tables for technique selection.

Different Representation Forms. 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 (Marcelis 2001). In other instances the relation takes the form of a conditional probability distribution (van Boekel, Stein and van Bruggen 2004). For some cases we do not have enough detailed knowledge and we can only give a model as a black box (Geeraerd et al. 1998). Table 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 3 also provides an overview of capabilities of different methods for incorporating new knowledge into models constructed with these methods.

5.3 Other Remarks towards 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, ordinal, 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 (Kalousis and Hilario 2001), and landmarking (Pfahringer, Bensusan and Giraud-Carrier 2000). 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.

6 Conclusion

This paper 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 could relieve those experts from many time consuming tasks, and also 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.

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 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.

References

Adomavicius, G. and A. Tuzhilin (2001) Expert-driven validation of rule-based user models in personalization applications. *Data Mining Knowledge Discovery* 5(1/2): 33-58

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

Balderasy, M.-A., et al. (2005) Discovering hidden association rules. *International Workshop on Data Mining Methods for Anomaly Detection*, Chicago, Illinois, USA

Barendsz, A.W. (1998) Food safety and total quality management. *Food Control* 9(2-3): 163-170

Barker, G.C., N.L.C. Talbot and M.W. Peck (2002) Risk assessment for Clostridium botulinum: a network approach. *International Biodeterioration & Biodegradation* 50: 167 - 175

Berry, M. and G. Linoff (1997) *Data mining techniques for marketing*. New York, Wiley

Costello, B., et al. (2003) A prototype sensor system for the early detection of microbially linked spoilage in stored wheat grain. *Measurement Science & Technology* 14(4): 397-409

Crawford, J. and F. Crawford (1996) Data Mining in a Scientific Environment. *AUUG 96 & Asia Pacific World Wide Web 2nd Joint Conference*, Australia

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

Freedman, D. A. (2004) Graphical models for causation, and the identification problem. *Evaluation Review* 28(4): 267-293

Freitas, A. A. (1997) Generic, Set-Oriented Primitives to Support Data-Parallel Knowledge Discovery in Relational Database Systems. UK, University of Essex. Ph.D. Thesis.

Geeraerd, A. H., et al. (1998) Application of artificial neural networks as a non-linear modular modeling technique to describe bacterial growth in chilled food products. *International Journal of Food Microbiology* 44(1-2): 49

Gestal, M., et al. (2004) Classification of apple beverages using artificial neural networks with previous variable selection. *Analytica Chimica Acta* 524(1-2): 225

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., et al. (2005) Regression techniques for the prediction of lower limb kinematics. *Journal Of Biomechanical Engineering-Transactions Of The Asme* 127(6): 1020-1024

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

Han, J. and M. Kamber (2001) *Data mining: concepts and techniques*, Morgan Kaufmann Publishers

Heckerman, D. (1996). Bayesian networks for knowledge discovery. *Advances in Knowledge Discovery and Data Mining*. U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy, AAAI/MIT Press**:** 273-306

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

Horchner, P. M., et al. (2006) HACCP-based approach to the derivation of an on-farm food safety program for the Australian red meat industry. *Food Control* 17(7): 497-510

Kalousis, A. and M. Hilario (2001) Model Selection via Meta-learning: a Comparative Study. *International Journal of Artificial Intelligence Tools* 10(4)

Knorr, E., R. Ng and V. Tucakov. (2000) Distance-Based Outlier: algorithms and applications. *VLDB Journal* 8(3-4): 237-253

Kodratoff, Y. and V. Moustakis (1994) Managing Machine Learning Application Development and Organizational Implementation. *Tutorial notes, ECAI*

Kohavi, R. and G. H. John (1998). The Wrapper Approach. *Feature Selection for Knowledge Discovery and Data Mining*. H. Liu and H. Motoda, Kluwer Academic Publishers**:** 33-50

Li, Y., et al. (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

Liu, H. and H. Motoda (1998) *Feature Selection for Knowledge Discovery and Data Mining*, Kluwer Academic Publishers

Lobell, D. B., et al. (2005) Combining field surveys, remote sensing, and regression trees to understand yield variations in an irrigated wheat landscape. *Agronomy Journal* 97(1): 241-249

Marcelis, L. F. M. (2001). Yield and quality prediction of vegetables: the case of cucumber. *Food process modelling*. L. M. M. Tijskens, M. L. A. T. M. Hertog and B. M. Nicolai, Woodhead Publishing Limited**:** 183-200

Milde, H., et al. (1999) Qualitative Model-Based Decision Tree Generation for Diagnosis in Real World Industrial Application. *KI - Kunstliche Intelligenz* 13(3): 30-35

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-+

Ni, H. and S. Gunasekaran (1998) Food quality prediction with neural networks. *Food Technology* 5: 60-65

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

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

Pitz, D. W. and J. W. Shavlik (1995) Dynamically adding symbolically meaningful nodes to knowledge-based neural networks. *Knowledge-Based Systems* 8(6): 301

Roos, T., et al. (2005) On discriminative Bayesian network classifiers and logistic regression. *Machine Learning* 59(3): 267-296

Rousu, J., et al. (2003) Novel computational tools in bakery process data analysis: a comparative study. *Journal Of Food Engineering* 57(1): 45-56

Santoso, T., et al. (2005) A stochastic programming approach for supply chain network design under uncertainty. *European Journal of Operational Research* 167(1): 96-115

Silverstein, C., et al. (2000) Scalable techniques for mining causal structures. *Data Mining And Knowledge Discovery* 4(2-3): 163-192

Simoudis, E. (1996) Reality Check for Data Mining. *IEEE EXPERT* 11(5): 26-33

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., et al. (2004) Statistical Evaluation of Local Alignment Features Predicting Allergenicity Using Supervised Classification Algorithms. *International Archives of Allergy and Immunology* 133: 101-112

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 Boekel, M. A. J. S., A. Stein and A. H. C. van Bruggen (2004) *Bayesian statistics and quality modelling in the agro-food produciton chain*. Dordrecht, Kluwer Academic

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

Verdenius, F. (2005) Methodological Aspects of Designing Induction-based Applications, Universiteit van Amsterdam. PhD thesis**:** 160.

Verdenius, F. and L. Hunter (2001). The power and pitfalls of inductive modelling. *Food process modelling*. L. M. M. Tijskens, M. L. A. T. M. Hertog and B. M. Nicolai, Woodhead Publishing Limited**:** 105-136

Whittaker, A. D., et al. (1991) Ultrasonic signal classification for beef quality grading through neural networks. *Automated Agriculture for the 21st Century Symposium*

Wu, Y. M., et al. (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

Table 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 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 and Weir 1996)	Valid (Patel, Davies and Hannah 1995)	Valid (Agarwal 2005)	Valid (Balderasy et al. 2005)	Valid (Knorr, Ng and Tucakov. 2000)
Classification	Valid (Verdenius and Hunter 2001)	Valid (Gestal et al. 2004)	Valid (Gorte and Stein 1998)		Valid (Soeria-Atmadja et al. 2004)
Regression	Valid (Lobell et al. 2005)	Valid (Millan, Roa and Tapia 2001)	Valid (Roos et al. 2005)		Valid (Goulermas et al. 2005)
Dependence model			Valid (Heckerman 1996)	Valid (Smyth and Goodman 1992)	
Causal model			Valid (Pearl 2000)		

Table 3. Representation forms of DM methods and extensibility of corresponding models

Data mining method	**Representation form** (Fayyad, Piatesky-Shapiro and Smyth 1996)	**Novel knowledge incorporation**
Decision trees	Decision trees	Easy (Milde et al. 1999)
Association rules	Rules	Easy (Adomavicius and Tuzhilin 2001)
Neural networks	Linear or Nonlinear model	Difficult (Pitz and Shavlik 1995)
Nearest neighbors	Example-base methods	Difficult
Bayesian networks	Probabilistic graphical dependency model	Easy (van Boekel, Stein and van Bruggen 2004)

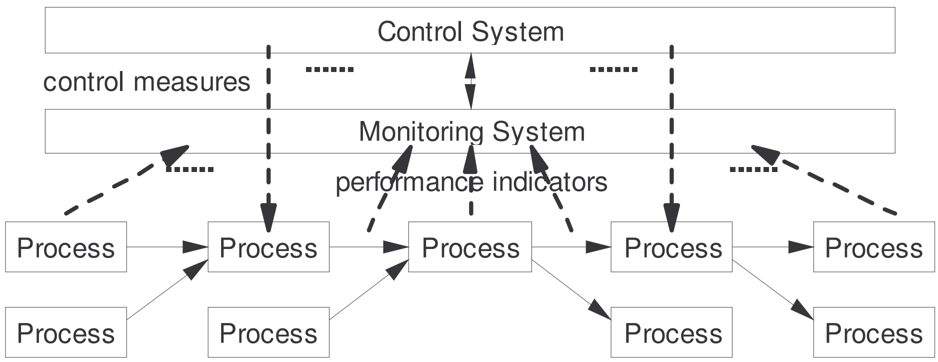

Fig. 1. An example supply network

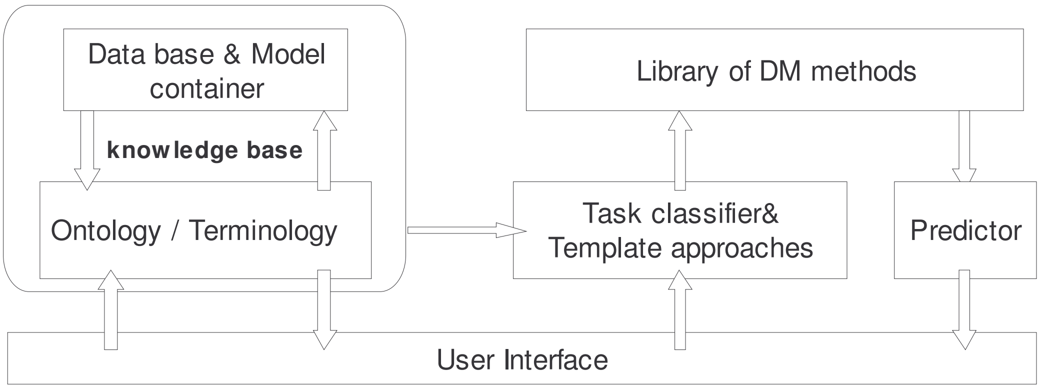

Fig. 2. Framework for early warning system in food supply networks