Guiding Users of Early Warning and Proactive Control Systems in Food Supply Chain Networks

Yuan Li[1], Mark R. Kramer[1], Adrie J.M. Beulens[1], Jack G.A.J. van der Vorst[2]
[1] Information Technology Group, Wageningen University
[2] Operations Research and Logistics Group, Wageningen University
Hollandseweg 1, 6706 KN Wageningen, The Netherlands
Yuan.Li@wur.nl

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

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

2. Context

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 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 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.

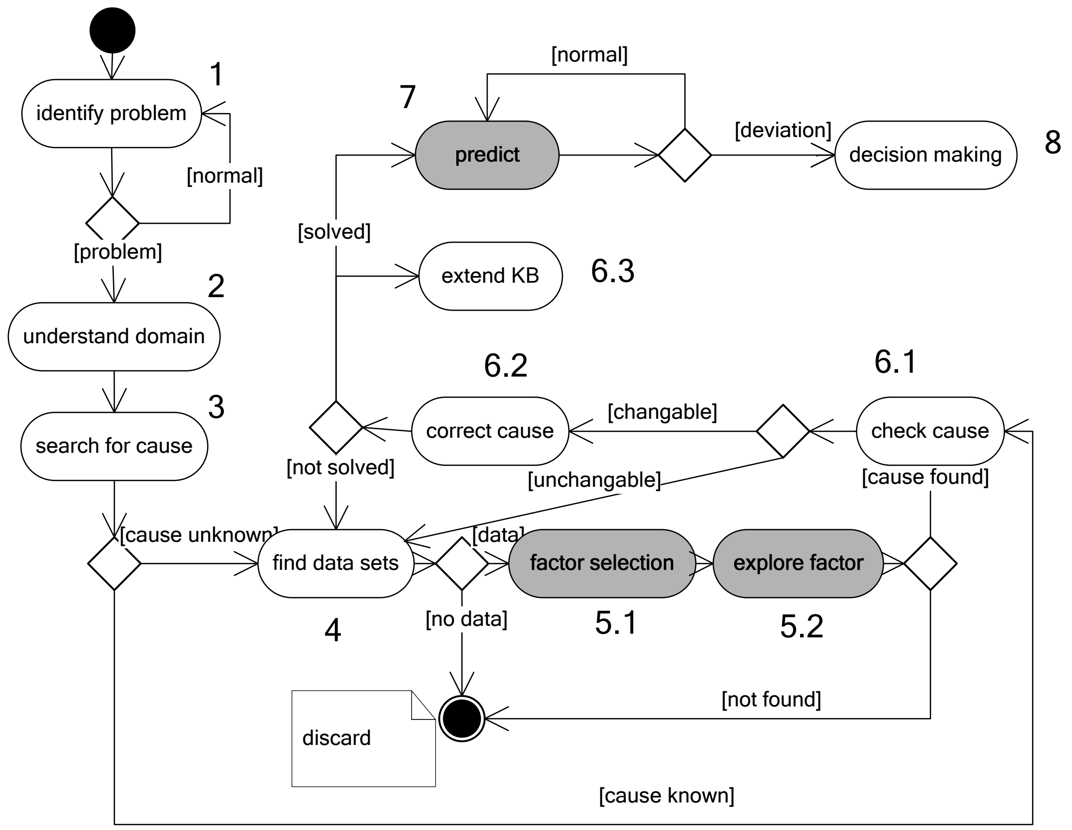

Figure 1: processes for early warning and proactive control in FSCN; processes marked in grey need support on DM

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.

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 2 as arrow-linked white boxes. They are generalized from literature in Knowledge Discovery in Database (KDD) process models (Fayyad et al. 1996; Kurgan et al. 2006). The grey boxes in Figure 2 indicate where we support the process.

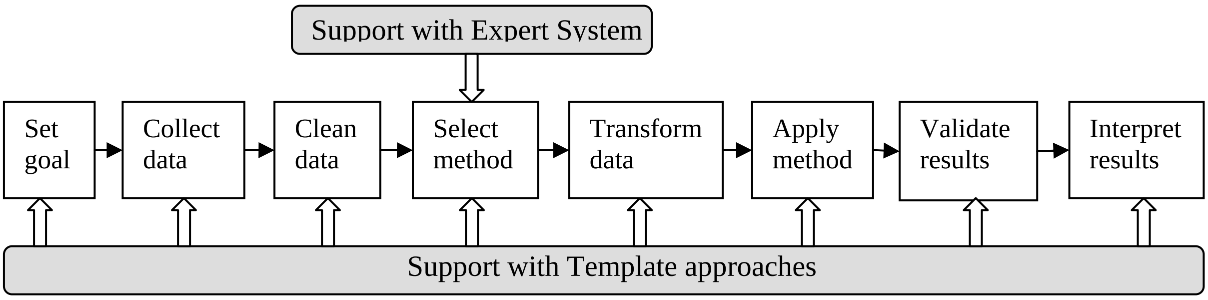

Figure 2: Relation between expert system, template approaches, and KDD process

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 et al. 2007). Figure 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 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 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 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 et al. 2001), landmarking (Pfahringer 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 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 paper 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.

3. Designing Template approaches and Expert System

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 et al. 1996; Kurgan et al. 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 3. This generic process needs to be validated in further case studies.

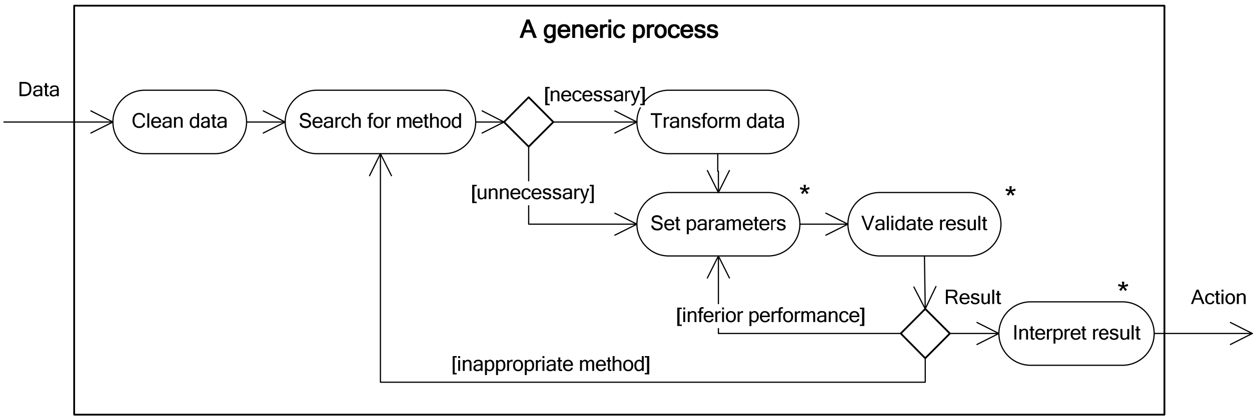

Figure 3: a generic process for applying DM methods in EW&PC systems; the steps marked with a star are method-specific

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 3. This implies that, for each supported method, there should be comprehensive and detailed information on how to set parameters, validate and interpret results.

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 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 1.

Table 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 (1992)
4	Nearest Neighbours	Aha (1992)
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 2, current research on categorizations of DM method properties is not suitable for EW&PC. In our research (Li 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. As mentioned in section 2.2, 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 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 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.

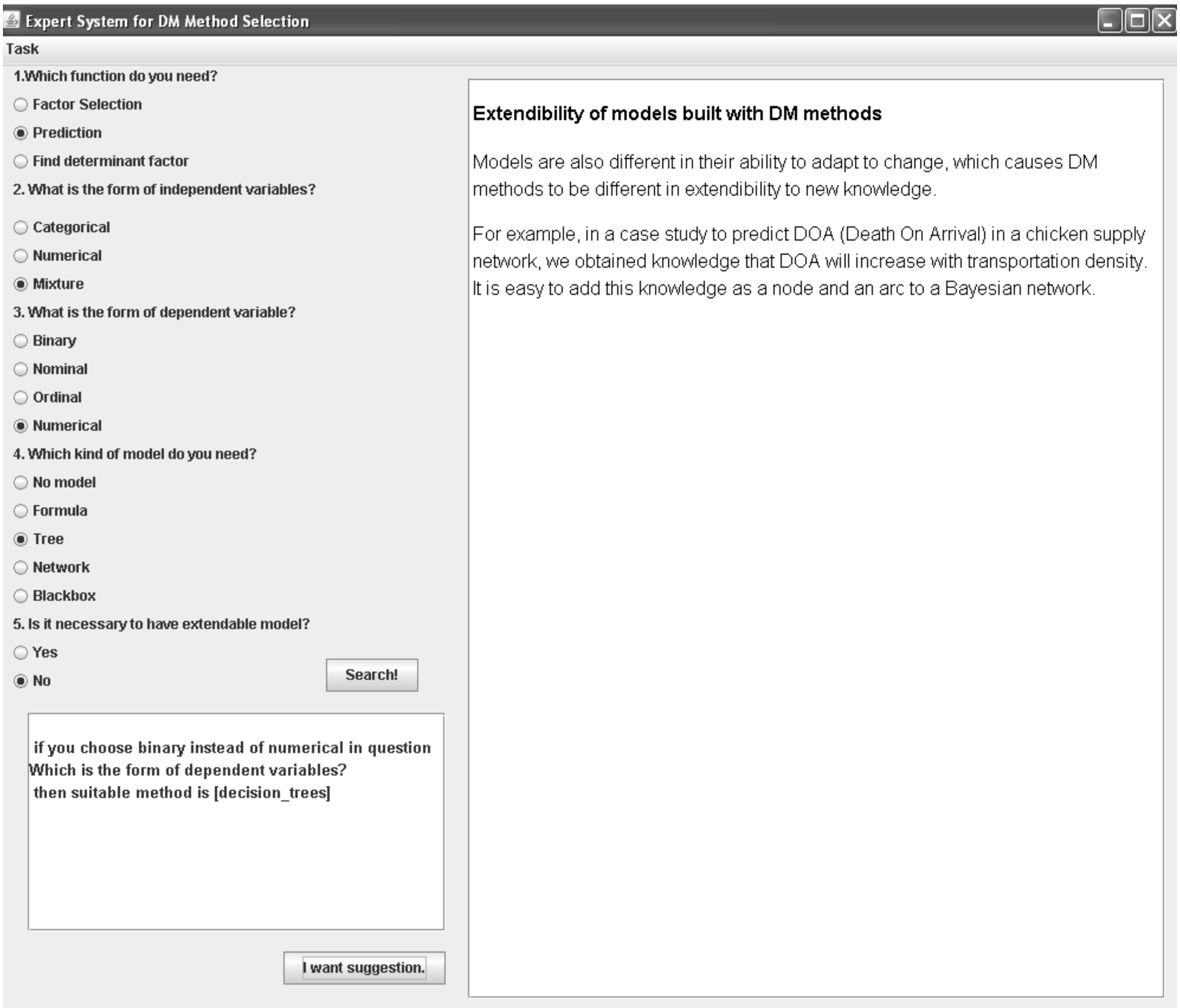

Figure 4: prototype User Interface of Expert System for method selection

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 gives an impression of how the system operates.

The generic process for using DM methods in EW&PC systems (as in Figure 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.

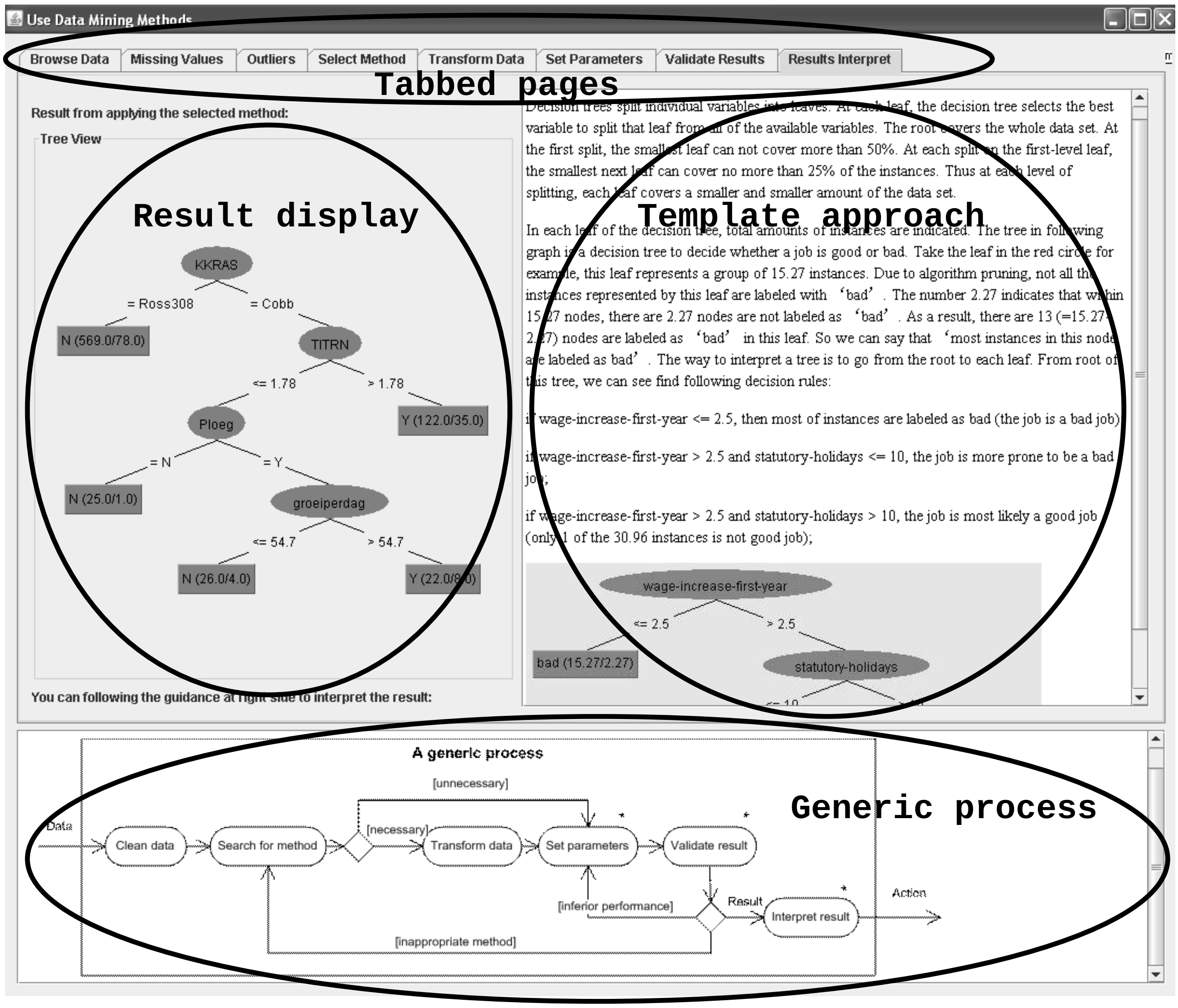

Figure 5: user interface of the prototype system

4. Select Method

Managers can use the Expert System for method selection (as shown in Figure 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, 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. 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 paper is on supporting DM method application for factor selection, casual factor exploration, and prediction. However, as indicated in Figure 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.

Reference

Aha, D. W. (1992). Tolerating noisy, irrelevant, and novel attributes in instance-based learning algorithms. International Journal Of Man-Machine Studies 36: 267-287.

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

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.

Cooper, G. F. and E. Herskovits (1992). A Bayesian Method For The Induction Of Probabilistic Networks From Data. Machine Learning 9(4): 309-347.

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.

Edwards, D. (2000). Introduction to Graphical Modelling. New York, Springer-Verlag.

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.

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.

Haykin, S. (1999). Neural networks : a comprehensive foundation. London [etc.] : Prentice Hall.

Kurgan, L. A. and P. Musilek (2006). A survey of knowledge discovery and data mining process models. Knowledge Engineering Review 21(1): 1-24.

Lattin, J. M., P. E. Green and J. D. Carroll (2003). Analyzing Multivariate Data Duxbury Press.

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, Oct. 2006, Namur, Belgium.

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.

Mintzberg, H. (1973). The Nature of Managerial Work. New York [etc.] : Harper & Row.

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.

Quinlan, J. R. (1992). C4.5: Programs for Machine Learning San Francisco,CA, Morgan Kaufmann

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.

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

Spirtes, P., C. Glymour and R. Scheines (2000). Causation, Prediction and Search. MIT Press.

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