

Knowledge Acquisition Approach Based on Rough Set in Online Aided Decision System for Food Processing Quality and Safety

Liu Peng, Liu Wen, Li Qiang, Yang Li, Duan Min and Dai Yue

China National Institute of Standardization

*liupeng@cnis.gov.cn, liuwen@cnis.gov.cn, liqiang@cnis.gov.cn,
yangli@cnis.gov.cn, duanmin@cnis.gov.cn, daiyue@cnis.gov.cn*

Abstract

For the problem that the use effect of food processing information system is poor due to absence of knowledge acquisition measures and knowledge self-updating function, a knowledge acquisition approach based on rough set is put forward. First, the approach establishes a set of predicted samples for the relationship between food processing parameters and product quality; then uses the discretization of continuous attribute, attribute reduction and rule extraction algorithm of rough set to acquire automatically predicted knowledge from a large number of predicted sample sets, and then saves the predicted knowledge in the knowledge base of expert system; finally, realizes the extraction of knowledge of food processing process based on the inference engine, which greatly enhances the effectiveness and practicality of the acquired knowledge in online aided decision system of the food processing quality and safety.

Keywords: *Food processing; Expert system; Knowledge Acquisition; Rough set theory*

1. Introduction

An important aspect of food safety is the quality safety in food processing process. The food production and processing process of modern food industry often involves multiple links and steps, and each link is related to food quality and safety, so it is of important significance to conduct online monitoring and support decision. Monitoring parameters in food production and processing process include necessary test data, processing craft parameters, personnel operating records, processing environment records and other data, which are various and bring significant data processing volume. In order to respond to the current quality and safety situation in the food production and processing process in time, it is necessary to use online aided decision system of the food processing quality and safety to assist operators to respond promptly and give advice of decision.

At present, there are few studies for online aided decision system of the food processing quality and safety at home and abroad, but there are some reports involved in other industries. Wang Ling, et. al. save a variety of complex decision rules in the knowledge base, and then gradually call other rules for prediction according to a preliminary judgment result of gas monitoring data, and the speed of this sequential reasoning prediction approach is slower[8]. On the basis of the study of granular computing theory, Zhang Qinghua, (2011) proposed an incremental knowledge acquisition approach based on granular computing. Namely by means of establishing an original granular knowledge tree of the decision information system, it will search for matched knowledge granule in the original granular knowledge tree for newly added data, update the granular knowledge tree based on decision values, and fast and efficiently handle the dynamic information system[6]. For features of a wide variety of dairy products and complex production craft, combined with the popular neural network technology, Wang Hui (2012) used a normalization approach to process sample data, and made a dairy simulation experiment for the quality of dairy product and completed it in

the system [5]. On the basis of decision tree theory, Zhang Jing (2011) proposed a knowledge acquisition approach based on the decision tree. This approach makes full use of the advantages that the decision tree can integrate the knowledge representation with acquisition, so that the knowledge representation and knowledge acquisition can be conducted at the same time, which overcomes the shortcoming that the knowledge representation and knowledge acquisition is separated in traditional artificial intelligence system [7]. Wei Yongfu (2012) analyzed the study situation of CAPP (Computer Aided Process Planning) system and some commonly used craft decision approaches, proposed the overall framework of craft planning system and craft & processing approach decision based on craft knowledge base and inference engine, and verified the feasibility of the decision approach with examples [9]. Above aided decision systems do not take approaches of knowledge acquisition, and are not equipped with self-learning function in knowledge base, which results in poor aided decision effect. In this paper, the rough set approach is introduced to online aided decision system of the food processing quality and safety to achieve the knowledge acquisition function of quality and safety knowledge base in food production and processing progress, and effectively improve the intelligent level of the online aided decision system of the food processing quality and safety.

2. Online Aided Decision System of the Food Processing Quality and Safety Based On Knowledge Acquisition of Rough Set

The online aided decision system of the food processing quality and safety based on knowledge acquisition of rough set is composed of the subsystem of knowledge acquisition, data in knowledge base, subsystem of inference engine, subsystem of real-time data acquisition, etc. Its workflow is shown in Figure 1. First, it determines a set of collected samples and a set of decision knowledge and the establishes a decision information table, and then uses the discretization of continuous attribute, feature attribute reduction, rule acquisition and other approaches to acquire knowledge of the online aided decision system of the food processing quality and safety, and save it in the knowledge base. For the newly collected data, the system will input the online monitoring data of food processing quality and safety to the inference engine, and get new knowledge by a certain reasoning mechanism. Knowledge that passes the verification will be saved by a certain principle and learned by an incremental mode. With a continuous accumulation of online data collection of food processing quality and safety, the knowledge in the base is increased and updated continuously, and aided decision precision of the online aided decision system of the food processing quality and safety will be improved gradually.

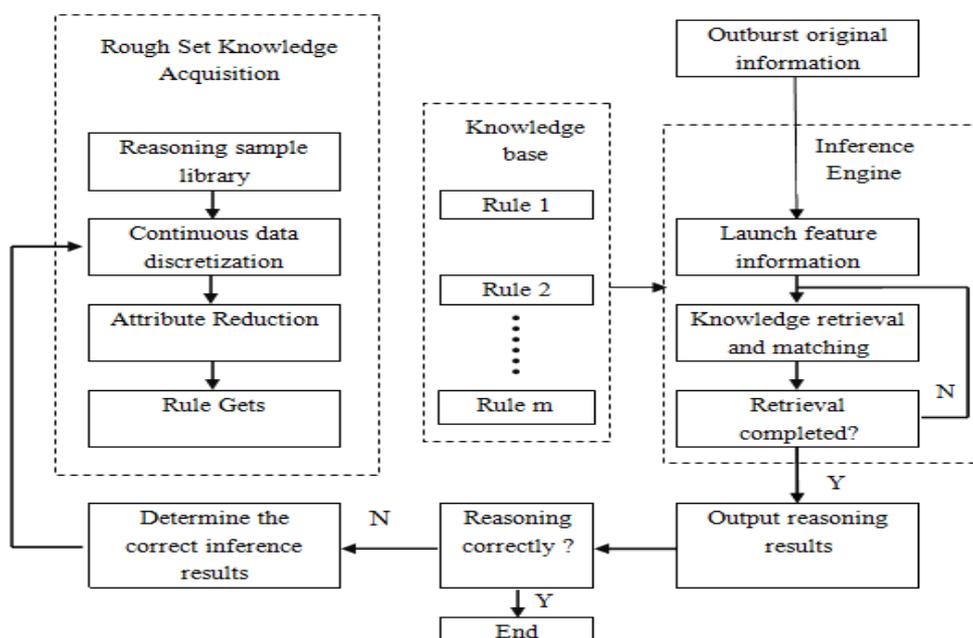


Figure 1.

3. Realization of Knowledge Acquisition Approach Based on Rough Sets

3.1. Knowledge Acquisition and Evaluation

The attribute reduction, to some extent, eliminates the redundant or unimportant information in the system, but it cannot directly get satisfactory decision rules from the decision system after being reduced. In order to get simpler decision rules, the decision system must be further processed, i.e. decision rule acquisition.

Currently, various decision rule acquisition algorithms have been put forward, such as average value reduction algorithm, value reduction algorithm based on decision matrix, and default rule acquisition algorithm based on projection, etc. The algorithm in reference [9] is used in this paper, that is, the mining algorithm extraction rule based on heuristic rule.

But not all extracted rules are practically valued and needed to be evaluated, those of which comply with evaluation requirements can be saved in the knowledge base for decision making. Evaluation indicators of common rules in the rough set theory have certainty and coverage, which are defined as follows:

Set the decision information system $DS = \langle U, C \cup D, V, f \rangle$, for arbitrary $x \in U$, the decision rule r_x corresponding to the object x , i.e., $r_x : des([x]_C) \rightarrow des([x]_D)$.

Wherein, $des([x]_C)$ is called the antecedent of rules (condition part) and $des([x]_D)$ is called the consequent of rules (decision part). $des([x]_C)$ shows the description of x under the condition attribute set C . $des([x]_D)$ shows the description of x under the decision attribute set D .

In the decision information system $DS = \langle U, C \cup D, V, f \rangle$ family, for arbitrary $x \in U$, the certainty $Cer(r_x)$ of decision rule $r_x : des([x]_C) \rightarrow des([x]_D)$ corresponding to the object x is defined as [1, 2, 3]:

$$Cer(r_x) = \frac{|[x]_C \cap [x]_D|}{|[x]_C|} \quad (1)$$

The certainty $Cer(r_x)$ reflects the credibility of decision rules, which is the accuracy of condition category allocated to the decision category in the measuring rule r_x .

In the decision information system $DS = \langle U, C \cup D, V, f \rangle$, for arbitrary $x \in U$, the coverage $Cov(r_x)$ of decision rule $r_x : des([x]_C) \rightarrow des([x]_D)$ corresponding to the object x is defined as:

$$Cov(r_x) = \frac{|[x]_C \cap [x]_D|}{|[x]_D|} \quad (2)$$

The coverage $Cov(r_x)$ is used to evaluate the quality of decision rules, reflecting indicators of two aspects: precision of the condition category in decision rules allocation to decision category and coverage of the condition category to the decision category.

The above 2 measuring methods have their own characteristics; however, when the test sample is analyzed by making use of the knowledge base during practical application, there are usually mutual match in multiple certainty or possibility classification rules. The two existing standards show certain limitation.

Based on above situation, this paper, based on the classification capacity of rules, defines the classification rule evaluation function:

$$assessfunc = \frac{|[X]_C \cap [X]_D|}{|[X]_C| |[X]_D|} \quad (3)$$

The parameter is called reflecting the upgrading of condition category to decision category in the decision rule.

The upgrading assessfunc is used to evaluate the quality of decision rules; meanwhile, the equal weight reflects coverage of evaluating decision rules to the decision category.

3.2. Knowledge Inference Engine

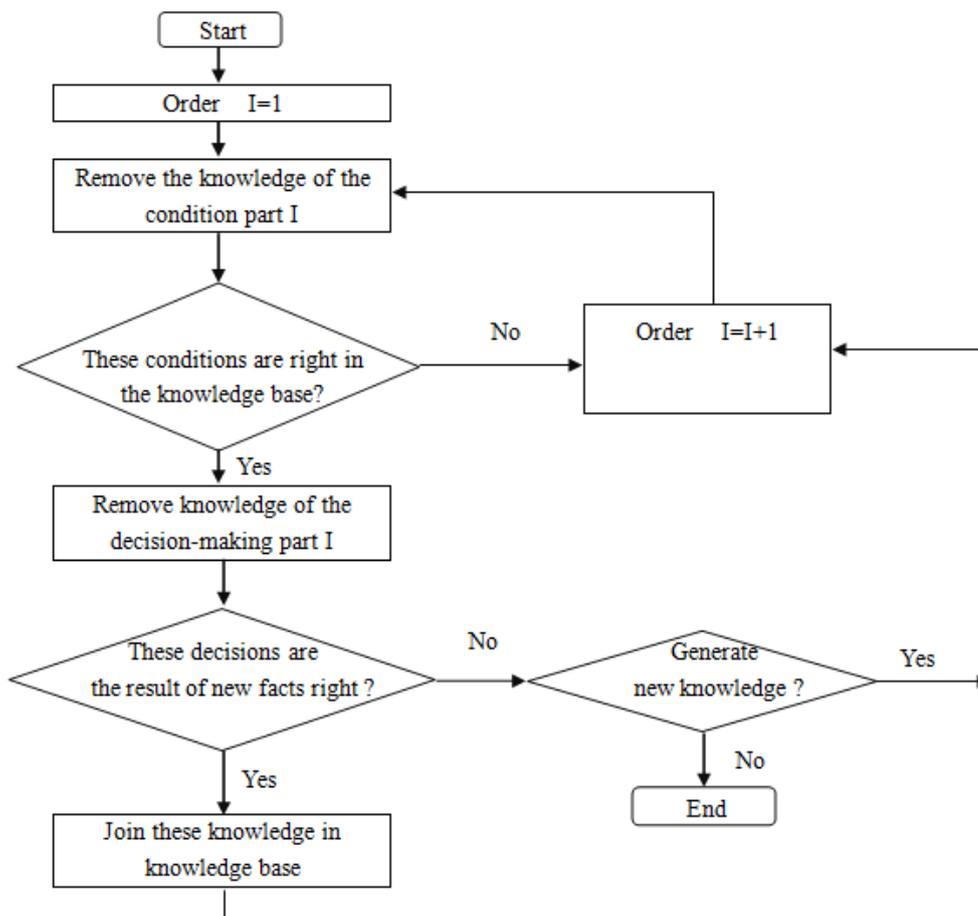


Figure 2.

As the control center of the expert system, the inference engine aims to solve the issues raised by current users when running the knowledge base and other mechanisms. The algorithm for knowledge inference engine updating the knowledge base is as follows:

(1) Calculate the level of similarity $s(k)$ between the feature information of samples under test and various rules of knowledge base:

$$s(k) = \left[\sum_{k=1}^p (x_{ik} - r_{jk})^2 \right]^{-\frac{1}{2}} \quad (4)$$

Wherein,

$$x_{ik} - r_{jk} = \begin{cases} 1, & \text{if } x_{ik} \neq r_{jk} \\ 0, & \text{else} \end{cases} \quad (5)$$

Wherein: p indicates the number of classification rules of each pieces of knowledge; x_{ik} indicates the k_{th} attribute value of Data i . r_{jk} indicates the k_{th} attribute value of Rule j .

(2) Select the classification rule with maximum $s(k)$ value to make decisions, and if there is only one classification rule with maximum $m(k)$ value, decide as per the rule, otherwise go to (3).

(3) If there are two classification rules and above with the same and maximum $s(k)$ value, for the classification rules of the same decision category and calculate the rule assessment function $assessfunc$, select the decision attribute values of $assessfunc$ and maximum value as the new decision attributes; if there are several modes that meet the $assessfunc$ and maximum value, select the mode with the maximum individual number as the new knowledge.

3. Experimental Verification

Take the fermented milk data of a dairy plant in Hebei Province as an example, so as to verify the application effects of the knowledge acquisition approach based on rough sets on the online aided decision system of the food processing quality and safety.

The raw materials greatly affect the processing quality of dairy products. In the dairy product formula, the principal raw material falls on the fresh milk, wherein the protein, fat content and acidity are key indicators to determine the raw materials. Besides, in the actual production process, for some critical process parameters, the samples can be delivered to the laboratory for detection, or the monitoring instrument will directly read the real-time values of these parameters. Thus, combined with actual production, select the acidity, total bacteria and sterilization temperature in the numerous process parameters as the process input parameters affecting the product quality. These parameters are not only related to the important parts of dairy products production, but also there are not simple linear relations among various parameters. At the same time, there is more than one quality parameter that affects the final dairy products, the protein content, however, acts as the key parameter and is the final assurance of product quality.

There are the meanings of the parameters, A1: milk protein content; A2: milk fat content; A3: acidity; A4: total number of bacteria; A5: sterilization temperature; R: fermented milk protein content.

Take the 584 sets of raw data generated in the production process of dairy products enterprises as the sample data, of which first 414 sets of data will be used as training samples, while remaining 170 sets of data as detection samples. Identify the test samples with the extracted rules. Firstly, discretize the continuous attributes of training samples, and see Table 1 for discrete intervals.

Table 1. Attribute Discretization Results

Condition	Discrete Intervals		
	1	2	3
Attribute			
A1	(2.87, 2.90)	(2.90, 2.91)	(2.91, 2.94)
A2	(3.00, 3.02)	(3.02, 3.04)	(3.04, 3.05)
A3	(68, 70)	(70, 71)	(71, 72)
A4	(115, 139)	(139, 174)	(174, 202)
A5	(95.0, 95.2)	(95.2, 95.4)	(95.4, 95.6)
R	(2.71, 2.73)	(2.73, 2.76)	(2.76, 2.79)

Carry out the attribute reduction for the discrete data, and the core attribute set after the reduction of sample data shall be (A1, A2, A4). A3 and A5 are reduced.

Then calculate the processing rule set of dairy products production by heuristic rules in reference [9], and see Table 2 for the results.

Table 2. Rules Obtained

Rule No.	Classification Rule Set	
Rule 1	$A1 = 3 \text{ AND } A2 = 2 \text{ AND } A4 = 1 \rightarrow R = 3$	1
Rule 2	$A1 = 1 \text{ AND } A4 = 3 \rightarrow R = 1$	
Rule 3	$A1 = 2 \text{ AND } A2 = 3 \text{ AND } A4 = 2 \rightarrow R = 2$	3

Assess the acquired rules, and carry out calculation with the traditional certainty, coverage and upgrading degree defined in this paper. It shows that the rules with the better upgrading degree comprehensively reflects the certainty and coverage, and see Table 3 for the specific results.

Table 3. Comparing the Results of each Evaluation Parameter Rules

Rule No.	Certainty Degree	Coverage	Upgrading Degree
Rule 1	0.977	0.963	0.00718
Rule 2	1.000	0.876	0.00724
Rule 3	0.992	0.978	0.00719

Separately identify and verify the extracted classification rule set of protein quality for the final dairy products, and the experimental results include the number of sample identification, number of classification rules and matching rate of rules. The matching rate of rules is defined as the proportion of test samples that match with rule conditions in rule set in total samples. See Table 4 for the experimental results.

Table 4. The Experimental Results

Data Set	Number of Instance	Number of Condition Attribute	Number of Sample Identification	Number of Rule	Rule Matching Rate
Test data	140	3	136	3	97.14%

Table 4 shows that when verifying the test data sets with the classification rule sets, the identification rate of sample classification and matching rate of rules are very high; the acquired rules are able to accurately identify the protein quality law in final dairy products that possess similar features of law with the training samples, indicating that the rough set approach possesses the stronger knowledge induction ability.

4. Conclusions

The online aided decision system for food processing quality and safety that adopts the knowledge acquisition approach based on the rough sets, gain key technologies by using the rough set knowledge so as to determine the key processing factors affecting the protein quality of dairy products. The acquired related knowledge, after the cooperative research of the expert system and related personnel, shall be added to the knowledge base as the new knowledge, so as to effectively solve the intelligent level and knowledge

acquisition bottleneck problems of expert system. These research achievements of this paper are originated from the project of national science and technology support program *Research and Demonstration of Online Safety and Quality Monitoring and Control Technique during Food Processing* (2012BAD29B04).

References

- [1] M. Crosbie and G. Spafford, "Applying Genetic Programming to Intrusion Detection", The AAAI Fall Symposium on Genetic Programming, (2003), pp. 1-8.
- [2] G. Shafer, "A Mathematical Theory of Evidence", Princeton: Princeton University Press, (1976), pp. 133-185.
- [3] M. Jevsnik, V. Hlebee and P. Raspor, "Consumer's awareness of food safety from shopping to eating", Food Control, vol. 19, no. 8, (2008), pp. 737-745.
- [4] S. M. Goni, S. Oddone and J. A. Segura, "Prediction of foods freezing and thawing times: Artificial neural networks and genetic algorithm approach", Journal of Food Engineering, vol. 84, no. 1, (2008).
- [5] W. Hui, "Dairy Products Quality Prediction Research Based on Neural Network", Agricultural University of Hebei, (2012).
- [6] Z. Qinghua, X. Yuke and Z. Yulan, "Incremental Knowledge Acquisition Approach Based on Granular Computing", Journal of Electronics & Information Technology, vol. 02, (2011), pp. 435-44.
- [7] Z. Jing, "Knowledge Acquisition Approach Research Based on Decision Tree", Manufacturing Automation, vol. 08, (2011), pp. 154-156.
- [8] W. Ling, "Knowledge Acquisition Approach Based on Rough Set in Mine Gas Prediction Expert System", Industry and Mine Automation, vol. 03, (2013), pp. 49-52.
- [9] W. Yongfu, "Research and Development of Process Decision System Based on Knowledge Base Inference and Genetic Algorithm", Guangxi University, (2012).
- [10] S.-M. Chen, J.-S. Ke and J.-F. Chang, "Knowledge representation using fuzzy Petrinets", Knowledge and Data Engineering IEEE Transactions, vol. 2, (1990), pp. 311-319.

