Approved Rough Set Knowledge Acquisition Approach in Intelligent Decision System

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

China National Institute of Standardization, China.
liupeng@cnis.gov.cn, liuwen@cnis.gov.cn, yangli@cnis.gov.cn,
liqiang@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

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 [1, 5]. 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[2, 3, 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 [4, 5].

2 Realization of Knowledge Acquisition Approach Based on Rough Sets

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 [6] 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:

In the decision information system $D^S=<U, C \cup D, V, I>$ family, for arbitrary $x \in U$, the certainty $Cer(r)$ of decision rule $r : \bar{d}(U[x]_C) \rightarrow \bar{d}(U[x]_D)$ corresponding to the object $x$ is defined as [1, 2, 3]:

$$Cer(r) = \left| \frac{\left| U[x]_C \cap I[x]_D \right|}{\left| U[x]_C \right|} \right|$$  \hspace{1cm} (1)

In the decision information system $D^S=<U, C \cup D, V, I>$, for arbitrary $x \in U$, the coverage $Cov(r)$ of decision rule $r : \bar{d}(U[x]_C) \rightarrow \bar{d}(U[x]_D)$ corresponding to the object $x$ is defined as:

$$Cov(r) = \left| \frac{\left| U[x]_C \cap I[x]_D \right|}{\left| U[x]_D \right|} \right|$$  \hspace{1cm} (2)

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:
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 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.

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.

<table>
<thead>
<tr>
<th>Rule No.</th>
<th>Rule 1</th>
<th>Rule 2</th>
<th>Rule 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Certainty</td>
<td>0.977</td>
<td>1.000</td>
<td>0.992</td>
</tr>
<tr>
<td>Coverage</td>
<td>0.963</td>
<td>0.876</td>
<td>0.978</td>
</tr>
<tr>
<td>Upgrading Degree</td>
<td>0.00718</td>
<td>0.00724</td>
<td>0.00719</td>
</tr>
</tbody>
</table>

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 1. Comparing the results of each evaluation parameter rules

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Number of Instance</th>
<th>Number of Condition Attribute</th>
<th>Number of Sample Identification</th>
<th>Number of Rule</th>
<th>Rule Matching Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test data</td>
<td>140</td>
<td>3</td>
<td>136</td>
<td>3</td>
<td>97.14%</td>
</tr>
</tbody>
</table>
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

5. Wei Yongfu, Research and Development of Process Decision System Based on Knowledge Base Inference and Genetic Algorithm, Guangxi University (2012)