BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **14**, Special Issue

Sofia • 2014 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2014-0049

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

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

China National Institute of Standardization
Emails: 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: *In connection with the problem that the food processing information system is poor due to the absence of knowledge acquisition and a knowledge self-updating function, a knowledge acquisition approach, based on a Support Vector Machine (SVM) is proposed. First, the approach establishes a set of predicted samples for the relationship between the food processing parameters and product quality; then it uses discretization of the continuous attributes, attributes reduction and a rule extraction algorithm of SVM to automatically acquire predicted knowledge from a large number of predicted sample sets. After that it saves the predicted knowledge in the knowledge base of an expert system; finally, the method realizes extraction of the knowledge about the food processing process based on the inference engine, which greatly enhances the efficiency and applicability of the acquired knowledge in an online aided decision system of food processing quality and safety.*

Keywords: *Food processing, expert system, knowledge acquisition, SVM theory.*

1. Introduction

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

At present there are few studies of online aided decision systems of food processing quality and safety at home and abroad, but there are some reports included in other industries. Wang Ling [6] 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. On the basis of the study of granular computing theory, Zhang et al. [4] 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 a matched knowledge granule in the original granular knowledge tree for newly added data, update the granular knowledge tree based on decision values, and quickly and efficiently handle the dynamic information system. For features of a wide variety of dairy products and complex production craft, combined with a popular neural network technology, Wang Hui [3] used a normalization approach to process the sample data and made a dairy simulation experiment for the quality of dairy products and completed it in the system. On the basis of decision tree theory, Zhang Jing [5] proposed a knowledge acquisition approach based on the decision tree. This approach makes full use of the advantages that the decision tree can integrate from 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 are separated in traditional artificial intelligence systems. Wei Yongfu [7] analyzed the study situation of CAPP (Computer Aided Process Planning) system and some commonly used craft decision approaches, proposed the overall framework of a craft planning system and craft and processing approach decision, based on the craft knowledge base and inference engine and verified the feasibility of the decision approach by examples. The above aided decision systems do not take the approaches of knowledge acquisition and are not equipped with a self-learning function in the knowledge base, which results in a poor aided decision effect. In this paper a Support Vector Machine (SVM) approach is introduced to an 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, and efficiently improve the intelligent level of the online aided decision system of the food processing quality and safety.

2. An online aided decision system of the food processing quality and safety, based on knowledge acquisition of SVM

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

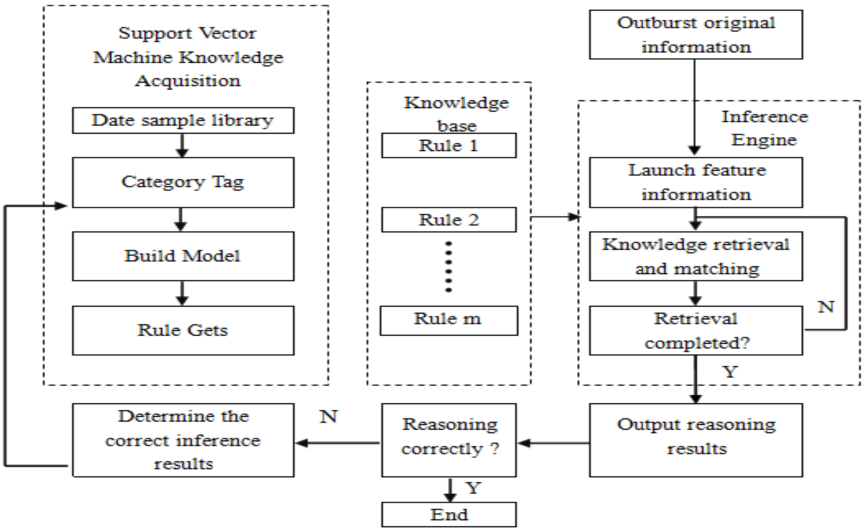

Fig. 1

3. Realization of the knowledge acquisition approach based on SVM

Based on the principle of structural risk minimization of the statistic learning theory, SVM [1] combines the maximum interface classifier theory with the methodology, based on a kernel. SVM shows excellent generalization ability, has many unique advantages in solving finite samples, nonlinearity, circumpolar latitude pattern recognition and other aspects. SVM also replaces the empirical risk principle by a structural risk minimization principle, applies the kernel function processing to nonlinear problems which possesses high training speed, seeks the best compromise between the model complexity and learning ability, ensuring better precision and generalization performance of SVM.

As a new research hotspot in machine learning field now, SVM has been successfully applied in many fields, including pattern recognition, function approximation, data mining, nonlinear system control, medical diagnosis and other aspects, and has been expanded to comprehensive assessment, economic forecast and so on.

Based on the SVM method of knowledge acquisition, the concrete six steps are as follows in the next

Algorithm

Step 1. Select an appropriate kernel function $K(x_i, x)$ and kernel parameters and regularization C.

Step 2. Construct and solve the optimization problem:

$$\max W(\alpha, \alpha^*) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j) +$$

$$+\sum_{i=1}^{n}[\alpha_i(y_i - \varepsilon) - \alpha_i^*(y_i + \varepsilon)],$$

subject to $\sum_{i=1}^{n}(\alpha_i - \alpha_j^*) = 0$, $0 \le \alpha_i$, $\alpha_i^* \le C$, $i = 1, 2, \ldots, n$.

Get the optimal solution α_i, α_i^*, $i = 1, 2, \ldots, n$.

Step 3. A nonlinear regression function to get access to knowledge

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(x_i, x) + b.$$

Step 4. The typical vector quantity must be determined by k-HBFCM. Then one axis of the ellipsoid has to be made by a typical vector quantity and support vector, which is the farthest from the typical vector quantity. The other axis and vertex must be located with the help of geometric analysis to create a relative ellipsoid area.

Step 5. Partition testing must be made on the ellipsoid area. If the results show that there is no overlapping between the different classifications, this area can be converted as a rule.

Step 6. Otherwise, a new ellipsoid must be generated. This processing has to be repeated until the testing shows that there is no overlapping between the different classifications in the new ellipsoid area or it has reached limited iterations. The whole ellipsoid area by partitioning must be converted as a relative rule.

When applying SVM to an On-line Assistant Decision Making System for Food Processing Quality Safety, there are lots of problems to be solved, among which the two hardest ones are:

Firstly, it costs too much when SVM does training to a (super) large-scale sample set and requires high time complexity. SMO algorithm, Light-SVM and the furthest neighbor do some improvements in this. But such problems, like relatively high space complexity and unguaranteed precision and low efficiency of a cutting non-support vector still exist.

Secondly, SVM has a smaller generalization ability when two training samples mix together, because it is necessary to guarantee the relatively great generalization ability of classifiers, seek for a larger interval and have a smaller cost in wrong classification. SVM cares for the boundary of two categories only. The points mixed in another category always cause over-fitting, which reduces the generalization ability.

3. Modified SVM arithmetic

In order to solve the problems above mentioned, we need such arithmetic: on one hand, it can delete or cut down the non-support vector training samples efficiently; on the other hand, when two types of training sample sets badly mix up and overlap with each other, it can solve the aliasing exactly to make SVM show good generalization.

The basic theory basis is that the support vector in SVM will not be in the correctly set partition section beyond the interval of two types of sample sets. In order to conveniently introduce the arithmetic, we first introduce some concepts [8]:

In the class $K= \{t_1, t_2,\ldots, t_n\}$, containing a point with a number of N, t_i is a sample of class K and each sample t_i is an m–dimensional vector, and generic centre of mass

$$C=\frac{1}{N}\sum_{i=1}^{N}t_i. \tag{1}$$

This concept is used to express the mean value of the vector of each sample in the same class – generic radius

$$R=\sqrt{\frac{\sum_{i=1}^{N}(t_i-C)^2}{N}}. \tag{2}$$

This concept shows the mean value of the Euclidean distance between each sample in the same class and the generic centre of mass.

Distance of generic centre of mass $M = |C_1 - C_2|$ refers to the Euclidean distance between two centers of mass.

Generic Centripetal Degree: Each sample t_i will calculate samples with a number of L nearest to it. We may as well suppose $D_1, D_2, ..., D_L$ are the distance of samples we research and use $1/D_i$ to stand for the influence factor of the point i regarding the classification adscription of the researched samples. It should be defined in some different cases.

- If the samples with a number of L and the sample we are going to research are in the same class, then the generic centripetal degree should be

(3)
$$E_i = \frac{\sum_{i=1}^{L} \frac{1}{D_i}}{L}.$$

- If the samples with a number of L and the sample we are going to research are not in the same class, then the generic centripetal degree should be

(4)
$$E_i = -\frac{\sum_{i=1}^{L} \frac{1}{D_i}}{L}.$$

- If among the samples with a number of L, samples with a number of m is in the same class with the sample we are going to research (suppose the distances are $D_1, D_2,..., D_m$) and the remaining samples from L to m are not in the same class with the sample we are going to research (suppose the distances are $D_{m+1}, D_{m+2},..., D_L$), then it indicates that there are some confusions and we should judge the order of severity by calculating generic centripetal degree, the generic centripetal degree can be calculated in the following formula:

(5)
$$E_i = \frac{\sum_{i=1}^{m} \frac{1}{D_i}}{m} - \frac{\sum_{i=m+1}^{L} \frac{1}{D_i}}{-(L-m)}.$$

This concept is used to the confusion degree between a sample point and samples around it.

Suppose that there are two classes of training sets. We use "+" to represent positive class and use "–" to stand for negative class. Then the arithmetic steps are as follows:

(1) Calculate the generic centre of mass of points in the two classes that is to say C+ and C– ; calculate the generic radius$R$+ and R–; calculate the distance of generic centre of mass L.

(2) Make a comparison: In order to be clearly understood, we may as well suppose $R+ > R-$.

If $M < \min(R+, R-)$:

– delete the samples in the positive class so that the distance between them and C+ is beyond R+;

– delete the samples in the negative class so that the distance between them and C– is less than R–.

If $\min (R+, R-) \leqq M \leqq \max(R+, R-)$:

– delete the samples in the negative class so that the distance between them and C– is less than R–;

– delete the samples in the negative class so that the distance between them and C – is beyond L.

If $M > \max(R+, R-)$:

– delete the samples in the positive class so that the distance between them and C– is beyond L;

– delete the samples in the positive class so that the distance between them and C+ is beyond L.

(3) After the following steps, the remaining samples are very likely to be the sample vectors of SV. Certainly, we also need to select the samples to cut down the confusion degree and calculate the generic centripetal degree – E_i of each remaining sample t_i. For the generic centripetal degree – E_i, we can suppose that the threshold value is ε ($\varepsilon > 0$):

if $E_i > \varepsilon$ or $-\varepsilon < E_i < \varepsilon$, we do not need to do anything;

if $E_i < -\varepsilon$, we delete the sample t_i.

(4) Next, we should classify and study the training set being cut down by SVM.

According to the above arithmetic description, we can note that in Steps 1 and 2, we select the samples which are most likely to become SV samples as initial samples. The order of complexity of the arithmetic is $O(n)$ and the farthest neighbor arithmetic is $O(n_2)$ (n refers to the number of samples), from which we can see that it is more efficient. In Step 3, we did not solve the confusion problem through NN or KNN, like the categorizer NN-SVM-KNN or SVM-KNN; instead, we introduced the concept Generic Centripetal Degree, and in this way we get the result, not considering the number of classes of samples around the researched sample, but also the distance between the researched sample and the samples around it. It is obvious that the result is more accurate. Thus this arithmetic meets the requirements of the initial period of the design.

5. Knowledge inference engine

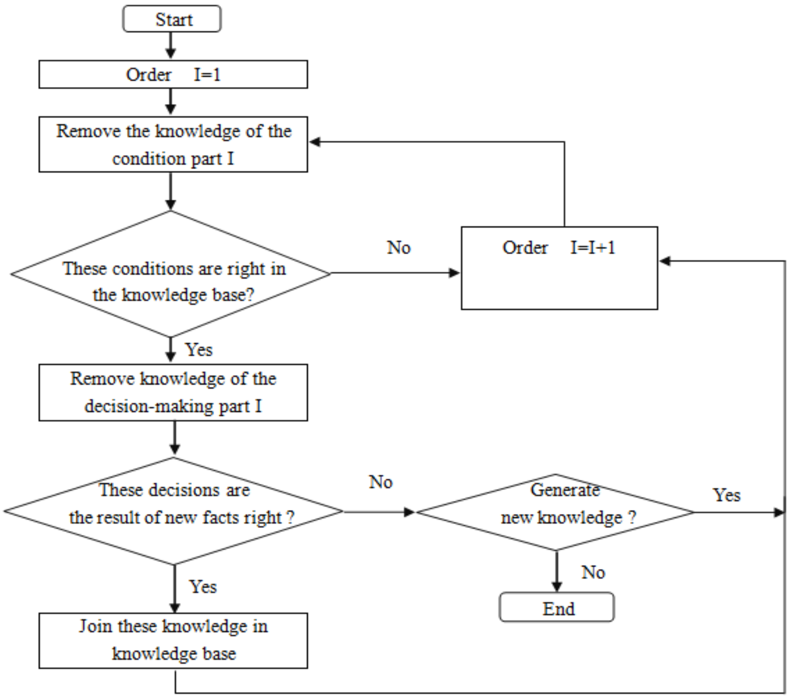
Start
Order I=1
Remove the knowledge of the condition part I
These conditions are right in the knowledge base?
No
Order I=I+1
Yes
Remove knowledge of the decision-making part I
These decisions are the result of new facts right ?
No
Generate new knowledge ?
Yes
No
End
Yes
Join these knowledge in knowledge base

Fig. 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 of the knowledge inference engine, updating the knowledge base is as follows:

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

$$s(k) = \left[\sum_{k=1}^{p} (x_{ik} - r_{jk})^2 \right]^{\frac{1}{2}}, \tag{6}$$

where

$$x_{ik} - r_{jk} = \begin{cases} 1 & \text{if } x_{ik} \neq r_{jk}, \\ 0 & \text{else.} \end{cases} \tag{7}$$

In the above formula, p indicates the number of classification rules of each piece of knowledge; x_{ik} indicates the kth attribute value of the data i; r_{jk} indicates the k-th attribute value of rule j.

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

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

6. Experimental verification

We conducted the experiments on PC (Intel dual core 2.8 GHz, 3G RAM). We have used the fermented milk data of a dairy plant in Hebei Province as an example, in order to verify the application effects of the knowledge acquisition approach based on SVM 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 the actual production, select the acidity, total bacteria and sterilization temperature in the numerous process parameters as 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 the various parameters. At the same time, there is more than one quality parameter that affects the final dairy products, the protein content, however, it acts as a key parameter and is the final assurance of product quality.

We have taken 584 sets of raw data generated in the production process of dairy products enterprises as sample data, of which the first 414 sets of data will be used as training samples, while the remaining 170 sets of data – as detection samples. The test samples are identified by the extracted rules. Firstly, discretize the continuous attributes of the training samples, and see Table 2 for the discrete intervals.

The meanings of the parameters are: A_1 – milk protein content; A_2 – milk fat content; A_3 – acidity; A_4 – total number of bacteria; A_5 – sterilization temperature; R – fermented milk protein content.

584 groups of raw data from the dairy enterprises during production will be used as sample data, among which, the former 414 groups of data will be used as training samples, while the rest 170 groups will be used as testing samples. The testing data will be identified by the extracted rules.

Firstly, the experts in dairy industry will be invited to discretize the R property of the training samples with the help of their expertise. It is classified into 3 intervals, as shown in Table 1.

Table 1

Condition	Discretization interval		
Property	X	Y	Z
R	(2.71, 2.74)	(2.74, 2.76)	(2.76, 2.79)

Secondly, the rule sets of the dairy products processing can be obtained by the Support Vector Machine (SVM) as noted in the paper. See Table 2 for the results obtained.

Table 2

Rule No	Classification rule set					
	A_1	A_2	A_3	A_4	A_5	R
Rule 1	(2.91, 2.94)	(3.02, 3.04)	–	(115, 139)	–	(2.76, 2.79)
Rule 2	(2.87, 2.90)	–	–	(174, 202)	–	(2.71, 2.73)
Rule 3	(2.90, 2.91)	(3.04, 3.05)	–	(139, 174)	–	(2.73, 2.76)

An improved algorithm is adopted for the test, with results shown in the Table 3.

Table 3

Classification	Numbers of removal samples in algorithm steps (1)-(2)	Percentage of numbers of removal samples in algorithm steps (1)-(2) to the total number
1	84	45.20
2	66	45.80%
3	28	33.33%

It can be concluded from above data that the impossible vector quantity removed during Step 1-2 may account for 42.99% of the total number. And it significantly reduces the complexity of algorithm, from the farthest $O(n_2)$ up to $O(n)$.

Besides, the mixing points removed during algorithm Step 3 account for about 40% of the remaining samples removed during the algorithm steps 1-2, which shows that it is very efficient to remove the mixing points between different classifications of samples by the algorithm Step 3. Moreover, it also shows that the influence of the parameters values during Step 3 on the mixing points, that is analyzed, is as follows (see Tables 4 and 5):

Table 4. Algorithm Step 3 with fixed $\varepsilon = -5$ to change the value of L

Results	L=2	L=4	L=5	L=7	L=9
Number of the removed mixing points	64	76	88	82	56
Percentage of the removed mixing points to the remaining total sample	27.12%	32.20%	37.29%	34.75%	23.73%

Table 5. Algorithm Step 3 with fixed $\varepsilon = -5$ to change the value of ε

Results	$\varepsilon = 0$	$\varepsilon = -1$	$\varepsilon = -2$	$\varepsilon = -3$	$\varepsilon = -4$	$\varepsilon = -10$
Number of the removed mixing points	46	67	80	69	56	44
Percentage of the removed mixing points to the remaining total sample	19.49%	28.39%	33.90%	29.24%	23.73%	18.64%

After increasing the value of L, it has been increased the removal rate of the mixing point, while the value of L is increased up to 5; the removal rate of the mixing point is basically kept changed. No matter how much the value of L is increased, the removal rate of the mixing point remained unchanged.

The removal rate of the mixing point decreases as the absolute value of ε increases, which shows that the increasing the absolute value of ε, the restriction for removal becomes more rigid, so that the mixing point of removal decreases.

After trimming the procedure (L = 5, ε = –2) with compiled data, SVM experiment will be made. The kernel function is Gauss kernel is

$$\exp\left(-\frac{1}{2\sigma^2}(x_j - x)^2\right),$$

where $\sigma = 0.5$, penalty parameter of SVM $C = 100$.

The results are shown in the Table 6.

Table 6

Algorithm	Samples of the training set	Accuracy of classification	Accuracy difference of classification	Times of speed improvement
By our algorithm	156	79.64%	Increased by 2.20%	1.59 times
By a traditional SVM	414	77.24%		

7. Conclusions

The online aided decision system for food processing quality and safety that adopts the knowledge acquisition approach based on SVM, helps key technologies by using the SVM knowledge, 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 the related personnel, will be added to the knowledge base as new knowledge, in order to efficiently solve the intelligent level and knowledge acquisition bottleneck problems of the expert systems. The research achievements of the paper have originated from the Project of the national science and technology support program Research and Demonstration of Online Safety and Quality Monitoring and Control Technique during Food Processing (2012BAD29B04).

References

1. Vapnik, V. N. An Overview of Statistical Learning Theory. – IEEE Trans. on NN, 1999.
2. Nello, C., S. T. John. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press, 2000.
3. Wang, Hui. Dairy Products Quality Prediction Research Based on Neural Network. Agricultural University of Hebei, 2012.
4. Zhang, Qinghua, Xing Yuke, Zhou Yulan. Incremental Knowledge Acquisition Approach Based on Granular Computing. – Journal of Electronics & Information Technology, Vol. **2**, 2011, 435-441.
5. Zhang, Jing. Knowledge Acquisition Approach Research Based on Decision Tree. – Manufacturing Automation, Vol. **8**, 2011, 154-156.
6. Wang, Ling. Knowledge Acquisition Approach Based on Rough Set in Mine Gas Prediction Expert System. – Industry and Mine Automation, Vol. **3**, 2013, 49-52.
7. Wei, Yongfu. Research and Development of Process Decision System Based on Knowledge Base Inference and Genetic Algorithm. Guangxi University, 2012.
8. Liu, Peng, Meng Haitao, Chen Xiaorong. A New Method to Advance Efficiency and Generalization of Support Vector Machine. – Journal of Guizhou University, Vol. **1**, 2007, 50-53.