WFL Publisher
Science and Technology
Meri-Rastilantie 3 B, FI-00980
Helsinki, Finland
e-mail: info@world-food.net

Journal of Food, Agriculture & Environment Vol.11 (1): 234-243. 2013
www.world-food.net

Fuzzy Bayesian network research on knowledge reasoning model of food safety control in China

Jianming Sun [1]*, Zhihui Sun [2] and Xiaofei Chen [3]
[1] *School of Computer and Information Engineering, [2]School of Industry, [3]Vocational Technology, Harbin University of Commerce, Harbin, Heilongjiang 150028, China.*e-mail:sjm@hrbcu.edu.cn*

Received 30 September 2012, accepted 28 January 2013.

Abstract

In order to take pre-crisis diagnosis, the safety warning, and definition of responsibility problems on the control of current food safety, a fuzzy Bayesian network of food safety risk knowledge reasoning model was established according to the data characteristics in the field of food safety control. Following the data research of the traceability system in the Bureau of Quality Supervision in a certain city in China, food safety risk related to index was extracted, whose value was defined by statistical methods. The sample data was thus achieved and the reasoning and diagnosis model of food safety control knowledge was set up by fuzzy Bayesian network algorithm based on the genetic algorithm. The application results show that fuzzy Bayesian network algorithm based on the genetic algorithm increased the computational complexity and running time due to the fuzzy math treatment, but the use of fuzzy logic can directly reflect the fuzzy random question reasoning and diagnosis on the possibility of high risk in certain step of the food production process. Compared with the general Bayesian network, the fuzzy Bayesian network has a higher accuracy of reasoning.

Key words: Fuzzy Bayesian network, food safety control, reasoning model, diagnosis model, fuzzy logic.

Introduction

In recent years, China's food safety accidents occurred frequently, which means that problems exist in food safety control. Therefore, how to push forward food safety control of knowledge reasoning and diagnosis reasoning become a concern, so that dynamic testing and safety warning on the risk analysis and food safety risk can be taken scientifically and responsibility for risk can be confirmed scientifically after the accident. Because the food safety involves many complicated factors, fuzzy Bayesian network and its application in the food safety knowledge control become the key point in this paper. Von Braun [1] of IEPRI points out in research on food safety that, in addition to the basic access to safety, other elements such as health, sanitation and capacity for taking care of social vulnerable groups also have important influence on food safety. Food safety always means sufficient, safe, nutritious and cultural food safety [2]. Antle [3] proposes the effective food safety regulation theory, and constructs the meat enterprises theory and econometric cost function model with the Rosen's product model of competitive enterprise production quality differences and Gertler and Waldman's quality-adjusted cost function model, hoping to test hypothesis of "product safety does not affect production efficiency" [3]. After the survey data analysis by the manufacturer, it was found that this assumption does not hold. Due to the existence of the above problem, China's food safety control knowledge reasoning model is proposed based on fuzzy Bayesian network. Data fuzzy partition and fuzzy random problems can be solved through the definition of hybrid event and fuzzy probability. Fuzzy Bayesian network is defined by the definition of conditional fuzzy probability table to solve the problem of combing fuzzy mechanism and Bayesian network. Following the research data of the traceability system in the Bureau of Guangzhou Municipal Quality Supervision in China [4], food safety risk related index is extracted,whose value is defined by statistical methods.The sample data is thus achieved. Through the genetic algorithm, learning process of the reasoning error feedback structure and parameters is reasoned and the network structure and parameters are optimized, then fuzzy Bayesian network is established. Reasoning model of food safety control knowledge based on fuzzy Bayesian network for reasoning and diagnosis can solve the actual problem in the application [5-7]. This paper compares the similarities and differences in algorithm complexity, accuracy, precision, succinctness between fuzzy Bayesian network and Bayesian network based on genetic algorithm. Through experiment, fuzzy Bayesian network established by the proposed algorithm proves to have a higher reasoning accuracy rate than general Bayesian network.

Model

Mixed event and fuzzy probability are defined by fuzzy mathematics principle. We will present concept of condition fuzzy probability table for the first time to solve representation problems of fuzzy and random variables and definition of fuzzy Bayesian network. Gauss membership parameters are to be found by clustering. Structure learning and parameter learning are optimized by genetic algorithm. Feedback of the reasoning horizontal error and membership degree error are used to find the optimal network structure. At the same time through modifying the parameters of the membership function synchronically correct network parameters, in particular parameter α for definition of fuzzy probability can be optimized, and then the fuzzy Bayesian network can be established.

Data fuzzification: In this algorithm, membership function of the data fuzzification uses the Gauss membership function. The attribute A is divided into three fuzzy subsets, of which a is one of fuzzy subset. First its class center m is found through cluster, the sample standard deviation σ is calculated, then membership function of a, subset of the attribute A, is the following: $\mu_a(x_i) = \exp(-(x_i-m)^2/\sigma^2)$, x_i is the value in the attribute A of sample data i [8, 9].

Calculation of the conditional fuzzy probability: Because the data has been fuzzified, the original method to calculate the conditional probability table used for frequency estimation of probability must undertake corresponding adjustment, membership needs to be taken into consideration [10-12]. For example, as shown in Fig. 1, condition fuzzy probability table P(A/B) of network node A is calculated. When the data is fuzzified, the specific steps are as follows: (1) if attributes A and B have fuzzy subsets $\{a_1, a_2\}$ and $\{b_1, b_2\}$,respectively, and all the sample sets $\{x_1, x_2,..., x_n\}$, therefore, each sample attached to the four fuzzy subsets membership degree is achieved as followes:

$$\{\mu_{a1}(x_1),\mu_{a2}(x_1),\mu_{b1}(x_1),\mu_{b2}(x_1)\},\{\mu_{a1}(x_2),\mu_{a2}(x_2),\mu_{b1}(x_2),\mu_{b2}(x_2)\},$$
$$\{\mu_{a1}(x_n),\mu_{a2}(x_n),\mu_{b1}(x_n),\mu_{b2}(x_n)\}.$$

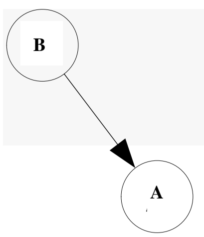

Figure 1. Diagram of fuzzy sets A and B.

(2) Calculate

$$\mu(b_1) = \frac{\sum_{i=1}^{n}\mu_{b1}(x_i)}{n},\ \mu(b_2) = \frac{\sum_{i=1}^{n}\mu_{b2}(x_i)}{n}.$$

By definition (hybrid event and fuzzy probability) test different results in the same conditions of U. The result is shown by $u_1(1\le i \le m)$, which is one of the fuzzy subsets on the X domain, then $A_i = \{X = u_i\}$ is called hybrid event, and its probability of occurrence is defined as a fuzzy probability $P(A_i)$.

The definition expression is shown as follows:

$$P(A_i) = \frac{\mu(u_i)^{1/a}}{\sum_{i=1}^{m}\mu(u_i)^{1/a}} \quad 0 \prec a \prec 1$$

Calculate to get the fuzzy probability of b_1, b_2

$$P(b_1) = \frac{\mu(b_1)^{1/a}}{\mu(b_1)^{1/a}+\mu(b_2)^{1/a}} \text{ and } P(b_2) = \frac{\mu(b_2)^{1/a}}{\mu(b_1)^{1/a}+\mu(b_2)^{1/a}}$$

(3) From the expression $\mu_{a_1\cap b_1}(x_i) = \mu_{a_1}(x_i) \wedge \mu_{b_1}(x_i) = \min\{\mu_{a_1}(x_i), \mu_{b_1}(x_i)\}$ Calculate to get the membership set of

$\alpha_1 \cap b_1$, $\{\mu_{a_1\cap b_1}(x_1), \mu_{a_1\cap b_1}(x_2), \cdots, \mu_{a_1\cap b_1}(x_n)\}$ and $\mu(a_1 \cap b_1) = \frac{\sum_{i=1}^{n}\mu_{a_1\cap b_1}(x_i)}{n}$

Similarly, calculate to get $\mu(a_2 \cap b_1) = \frac{\sum_{i=1}^{n}\mu_{a_2\cap b_1}(x_i)}{n}$

(4) The fuzzy probability of α_1, b_1 at the same time

$$P(a_1 \cap b_1) = \frac{\mu(a_1\cap b_1)^{1/\alpha}}{\mu(a_1\cap b_1)^{1/\alpha}+\mu(a_2\cap b_1)^{1/\alpha}}$$

(5) Calculate under the condition of $B = b_1$, when $A = \alpha_1$ conditional fuzzy probability $P(A = a_1|B = b_1) = \frac{P(a_1\cap b_1)}{P(b_1)}$

Similarly, $P(A = a_2|B = b_1) = \frac{P(a_2\cap b_1)}{P(b_1)}$, $P(A = a_1|B = b_2) = \frac{P(a_1\cap b_2)}{P(b_2)}$,

$$P(A = a_2|B = b_2) = \frac{P(a_2\cap b_2)}{P(b_2)}$$

Definition of fuzzy Bayesian network: Because this algorithm aims at fuzzy Bayesian network with fuzzy random variables, in the data has been made the fuzzy below the premise that change [13-14], it is necessary to adapt the conditional probability tables (CPT),which express the relation between the original variables' strength to the actual situation. The specific approach is to put the membership function of the relevant attribute fuzzy set into the corresponding conditional probability table, and then the condition fuzzy probability table (CFPT) is designed.

As node structure is shown in Fig. 1, if attribute A and attribute B has three fuzzy subsets, respectively,$\{a_1, a_2, a_3\}$, $\{b_1, b_2, b_3\}$then, condition fuzzy probability table of the designing attribute A is shown in Table 1.

As is shown in Table 2, Gauss membership function is used in this algorithm during the data fuzzification, so the corresponding position of membership function in CFPT is marked by parameters of Gauss membership function, i.e. the class center m and the sample standard deviation σ.

Genetic operator of the network structure: In this paper, connection matrix is used to represent the fuzzy Bayesian network structure, and the elements in the matrix represent links between the nodes and their direction relationship [15]. Thus, design of the optimal network structure of genetic operator can realize through connection matrix. The following part will elaborate how to operate crossover and mutation of the connection matrix. Connection matrix crossover operator includes lines cross which link lines in connection matrix, and columns cross which link columns in connection matrix (Plates 1 and 2). Pick a random integer i ($0\le i \le m$) to determine cross lines or columns, randomly select to do line crossing or column crossing. Because fuzzy Bayesian network is a directed acyclic graph, the cross operation generates a new connection matrix to be performed acyclic test. Those who meet the acyclic test will be put into new species, while those who won't will generate random number I, and do line or column cross at random selection, and then do crossover operation again.

Line crossing: Exchange all elements on line I between two connection matrices.

Column crossing: Exchange all elements on column I between two connection matrices.

Connection matrix variation refers to network structure variation,

Table 1. CFPT.

$P(A \mid B)$	Membership function of the attribute A	Membership function of the attribute B	Fuzzy subset of attribute A	Fuzzy subset of attribute B
	$\mu_{a_1}(x)$	$\mu_{b_1}(x)$	a_1	b_1
	$\mu_{a_2}(x)$	$\mu_{b_1}(x)$	a_2	b_1
	$\mu_{a_3}(x)$	$\mu_{b_1}(x)$	a_3	b_1
	$\mu_{a_1}(x)$	$\mu_{b_2}(x)$	a_1	b_2
	$\mu_{a_2}(x)$	$\mu_{b_2}(x)$	a_2	b_2
	$\mu_{a_3}(x)$	$\mu_{b_2}(x)$	a_3	b_2
	$\mu_{a_1}(x)$	$\mu_{b_3}(x)$	a_1	b_3
	$\mu_{a_2}(x)$	$\mu_{b_3}(x)$	a_2	b_3
	$\mu_{a_3}(x)$	$\mu_{b_3}(x)$	a_3	b_3

Table 2. CFPT based on Gauss membership function.

$P(A \mid B)$	Membership function of the attribute A	Membershipfunction of the attribute B	Fuzzy subset of attribute A	Fuzzy subset of attribute B
	m_{a_1},σ_{a_1}	m_{b_1},σ_{b_1}	a_1	b_1
	m_{a_2},σ_{a_2}	m_{b_1},σ_{b_1}	a_2	b_1
	m_{a_3},σ_{a_3}	m_{b_1},σ_{b_1}	a_3	b_1
	m_{a_1},σ_{a_1}	m_{b_2},σ_{b_2}	a_1	b_2
	m_{a_2},σ_{a_2}	m_{b_2},σ_{b_2}	a_2	b_2
	m_{a_3},σ_{a_3}	m_{b_2},σ_{b_2}	a_3	b_2
	m_{a_1},σ_{a_1}	m_{b_3},σ_{b_3}	a_1	b_3
	m_{a_2},σ_{a_2}	m_{b_3},σ_{b_3}	a_2	b_3
	m_{a_3},σ_{a_3}	m_{b_3},σ_{b_3}	a_3	b_3

including side increase, decrease, and the reverse operation. The specific operating embodies in the transformation of the connection matrix elements from 1 to 0, and 0 to 1.

Assuming that the number of lines and columns of the connection matrix are n, the operations of mutation operation of the connection matrix are as follows:

(1) Randomly select two random integers i and J, and $1 \le i \le n$, $1 \le j \le n$;

(2) if $C_{i,j} = 0$, and $C_{j,i} = 0$, then make $C_{i,j} = 1$ if $C_{i,j} = 1$, then make $C_{i,j} = 0$ an acyclic test is needed after transformation to ensure that there is no loop. If it does not meet the requirement of acyclic test, then repeat step 1 and 2.

After the above operation, the network structure variation may come to one of three situations:

(1) a directed edge of the network structure is "increased":

$$C_{i,j} = C_{j,i} = 0 \Rightarrow C_{i,j} = 1$$

(2) a directed edge of the network structure is "deleted":

$$C_{i,j} = 0 \Rightarrow C_{i,j} = 1$$

(3) a directed edge of the network structure is "reversed":

$$C_{i,j} = 0,\ C_{j,i} = 1 \Rightarrow C_{i,j} = 1,\ C_{j,i} = 0$$

$$\begin{pmatrix} x_{11}, x_{12}, \cdots, x_{1m} \\ x_{21}, x_{22}, \cdots, x_{2m} \\ \cdots\cdots\cdots\cdots \\ x_{i1}, x_{i2}, \cdots, x_{im} \\ \cdots\cdots\cdots\cdots \\ x_{m1}, x_{m2}, \cdots, x_{mm} \end{pmatrix} \times \begin{pmatrix} y_{11}, y_{12}, \cdots, y_{1m} \\ y_{21}, y_{22}, \cdots, y_{2m} \\ \cdots\cdots\cdots\cdots \\ y_{i1}, y_{i2}, \cdots, y_{im} \\ \cdots\cdots\cdots\cdots \\ y_{m1}, y_{m2}, \cdots, y_{mm} \end{pmatrix} =$$

$$\begin{pmatrix} x_{11}, x_{12}, \cdots, x_{1m} \\ x_{21}, x_{22}, \cdots, x_{2m} \\ \cdots\cdots\cdots\cdots \\ y_{i1}, y_{i2}, \cdots, y_{im} \\ \cdots\cdots\cdots\cdots \\ y_{m1}, y_{m2}, \cdots, y_{mm} \end{pmatrix} \times \begin{pmatrix} y_{11}, y_{12}, \cdots, y_{1m} \\ y_{21}, y_{22}, \cdots, y_{2m} \\ \cdots\cdots\cdots\cdots \\ x_{i1}, x_{i2}, \cdots, x_{im} \\ \cdots\cdots\cdots\cdots \\ x_{m1}, x_{m2}, \cdots, x_{mm} \end{pmatrix}$$

Plate 1.

$$\begin{pmatrix} x_{11}, x_{12}, \cdots, x_{1i}, \cdots, x_{1m} \\ x_{21}, x_{22}, \cdots, x_{2i}, \cdots, x_{2m} \\ \cdots\cdots\cdots\cdots \\ x_{m1}, x_{m2}, \cdots, x_{mi}, \cdots, x_{mm} \end{pmatrix} \times \begin{pmatrix} y_{11}, y_{12}, \cdots, y_{1i}, \cdots, y_{1m} \\ y_{21}, y_{22}, \cdots, y_{2i}, \cdots, y_{2m} \\ \cdots\cdots\cdots\cdots \\ y_{m1}, y_{m2}, \cdots, y_{mi}, \cdots, y_{mm} \end{pmatrix} =$$

$$\begin{pmatrix} x_{11}, x_{12}, \cdots, y_{1i}, \cdots, y_{1m} \\ x_{21}, x_{22}, \cdots, y_{2i}, \cdots, y_{2m} \\ \cdots\cdots\cdots\cdots \\ x_{m1}, x_{m2}, \cdots, y_{mi}, \cdots, y_{mm} \end{pmatrix} \times \begin{pmatrix} y_{11}, y_{12}, \cdots, x_{1i}, \cdots, x_{1m} \\ y_{21}, y_{22}, \cdots, x_{2i}, \cdots, x_{2m} \\ \cdots\cdots\cdots\cdots \\ y_{m1}, y_{m2}, \cdots, x_{mi}, \cdots, x_{mm} \end{pmatrix}$$

Plate 2.

Adaptive value function: In order to improve the applied value of the model, this paper intended to increase forward reasoning errors, and backward reasoning errors with the corresponding membership degree error besides the classification errors. Assuming that reasoning level error is shown by $f_1(A)$, reasoning membership degree error is shown by $f_2(B)$, they can be defined by the following expression:

$$f_1(B) = f_c(B) + f_c(FR) + f_c(BR),\ f_2(B) = f_\mu(B) + f_\mu(FR) + f_\mu(BR)$$

where $f_c(B)$ and $f_\mu(B)$ are classification error and the membership degree error, $f_c(FR)$, $f_\mu(FR)$ are forward reasoning error and the membership degree error, $f_c(BR)$, $f_\mu(BR)$ are backward reasoning error and the membership degree error.

Given an unknown data samples of X, classification method will predict that X belongs to class of the highest posterior probability. According to Bayes theorem, posterior probability is given as

$$P(c|X) = \frac{P(X|c) \times P(c)}{P(X)}.$$

For all C, $P(X)$ is constant, so when the maximum of $P(X|c) \times P(c)$ will meet. Assuming that the classification result of the sample I in the data sets is $\overline{c_i}$, while the actual classification result is c_i, whose corresponding membership degree of fuzzy subsets are $\mu_{\overline{ci}}$ and μ_{ci}. Assuming that the sample number is n, then classification level error of this sample learning and the membership degree error respectively are:

$$f_c(B)=\sqrt{\sum_{i=1}^{n}(\bar{c}_i-c_i)^2},\ f_\mu(B)=\sqrt{\sum_{i=1}^{n}\left(\mu_{\bar{c}_i}-\mu_{c_i}\right)^2}$$

Forward reasoning is top-down reasoning. The basic idea is as follows: to find the available knowledge from the existing information or facts, to select knowledge through conflict resolution, and to change the solution state step by step until the issue is resolved. In this paper, in order to establish a more objective, practical, and adaptive fuzzy Bayesian network, the following forward reasoning strategy will be taken:

1. Select non-leaf node A in the network (generally from the root node of the network). Assuming that node A has two (or more) child nodes B and C, and the node B has only one parent node A, while node C also has parent node C and D.
2. The level and degree of membership node B is reasoned through the actual value of node A in sample data with the CFPT.
3. The level and degree of membership node C is reasoned through the actual value of node A and D (or more) in sample data with the CFPT. So far forward inference starting from node A is completed.
4. If there is a non-leaf node without forward reasoning, then repeat the above operation until all non-leaf nodes are reasoned. If the operation overlaps that of step 3, then step is not taken.

Backward reasoning refers to the bottom-up reasoning, its basic idea is as follows: firstly, let us take the hypothesis of a target, then find the knowledge set in the knowledge base whose conclusion leads to the target, then check every condition in the knowledge set. If a certain condition item in the knowledge set can be meet the user's session, or can match the current database, then the conclusion of that knowledge can join the current database, and that goal is proved; otherwise, the condition item of the knowledge as a new subgoal, recursive implement the process until all subgoals of the "and" relation or one subgoal of the "or" relation appears in the database, the target is solved; or until the subgoal cannot be further decomposed and the database can not realize the matching, the target hypothesis is false. In this paper, in order to establish a more reflective, practical, and adaptive fuzzy Bayesian network, following backward reasoning strategy will be taken:

1. Select non-root node A in the network (generally from the leaves of the network);
2. If A has N parent nodes of $B_1, B_2,..., B_n$, according to the actual value of sample A, with the network CFPT, reason the level and membership degree of nodes of $B_1, B_2,..., B_n$.
3. If B_1 has other child nodes such as A', the level of B_1 is not needed in backward reasoning of A'.
4. If a non-root node still exists, then repeat the above steps until all the nodes of leaf nodes are reasoned to new level.

Selection of operator: The basic principle of genetic algorithm is Darwin's principle of natural selection. Selection is the driving force of genetic algorithm [16,17]. High selective pressure would terminate the search precociously, while low selective pressure slow the search. So at initial stage of the algorithm, the lower selective pressure is adopted, which is conducive to the expansion of the searching space; while in the termination phase higher pressure is appropriate, which helps to find the best solution. In this article, the offspring individuals do not directly replace father generation, instead, the best individual is chosen from all the parent and offspring individuals as new individual. Specifically, the ranking selection method ($m + l$) is used. Specific operations are as follows: assuming that parent population size is m, its offspring's population size is l. All of the parent individuals and all the offspring individuals together sort from good to bad by target priorities, after sorting the first m optimal ones as a new generation of population [18, 19]. The above method has the advantages of improving the performance of genetic algorithm by increasing the crossover rate and mutation rate method. If we choose in the expanding sampling space composed of all the parents and offsprings. We do not worry about too much random variables caused by high crossover rate and mutation rate.

Flow of algorithm: First, the data should be classified, then the classified data should be fuzzified, and initialize the data of population. If the number of individuals have already produced with specified initial population size,then finish initialization,else connect the data matrix crossover and mutation,and calculate the cfpt.Generate intermediate population among 1, 2 and 3 species among populations,and merge sort choice among population generate new population, then use fuzzy reasoning calculation objective function. Finally returns the number of individual judgment generated in compliance with specified initial population size. Flow chart of fuzzy Bayesian network algorithm based on genetic algorithm is given in Fig. 2.

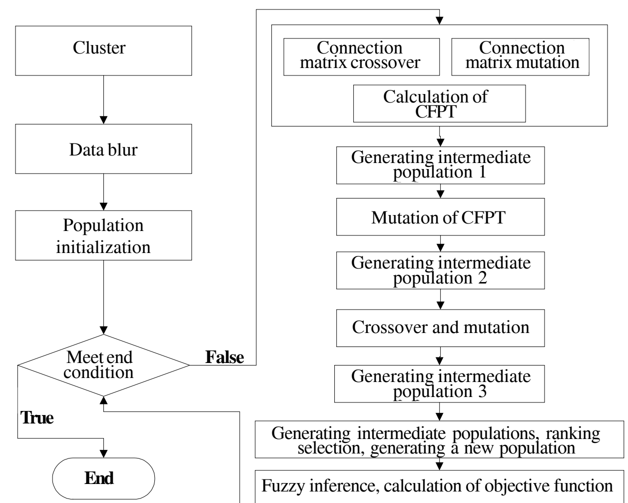

Figure 2. Flow chart of fuzzy Bayesian network algorithm based on genetic algorithm.

Reasoning Model of Food Safety Knowledge Control Based on Fuzzy Bayesian Network

In this paper, following the data research of the traceability system in the Bureau of Quality Supervision in Guangzhou, food safety risk related index is extracted ,whose value is defined by statistical methods.The sample data is thus achieved and the reasoning and diagnosis model of food safety control knowledge is set up by fuzzy Bayesian network algorithm based on the genetic algorithm.

Modeling indicator extraction: Food safety control is a system engineering, whose performance is affected by both the internal factors in the food safety systems such as food production conditions, level of science and technology of food production, impact of food hygiene policy, etc. and by external factors such as natural disasters, environmental pollution and other factors. So various possible factors must be taken into consideration when the performance evaluation indicator system for the food safety control is constructed. Theoretically, all the influential factors should be included on the food safety risk assessment, but considering the possibility of monitoring implementation, the construction of the food safety risk evaluation index system can only be designed in main factors which cause food insecurity. Two aspects are included mainly: one is the degree of microbial pollution in food; another is the content of poisonous and harmful substances in food. In order to effectively control food hygiene and safety, manufacturers may use the temperature as a control index. According to one professor in the University of Florida, the temperature control is the most effective tool to guarantee food quality and safety. In fact, microbes, like all the living things, need the right environment for growth, reproduction. If the required conditions, such as temperature and humidity, for the growth of bacteria can be known and controlled, bacterial reproduction can be effectively prevented, and even the number of the bacteria will be reduced. Temperature monitoring is not only suitable for the production process, it is also important in the storage and transportation. In fact, in the HACCP system, most related temperature control procedures focus on control of food safety. Key points in this key chain are set in the following parts: raw material procurement, production and processing, packaging, storage, distribution, sales, food and information feedback from consumers. Key value is restricted in each key point. The monitoring, recording, contrasting, adjusting, treatment, and prevention of that value can prevent the unsafe factors on people's health.

Design traceability system form: The posts that need to fill out the form are as follows: system management, inspection, material management, production management, product warehouse management, product inspection, health check (Fig. 3). Food safety risk related indicators are extracted from the fields in the 29 forms with the above analysis. Specific meaning of each index is as follows: food safety risk: it refers to the evaluation on food safety risk, if the food safety index IFS<<1, it means that the inspected food has no safety risk, and the index value is 1; otherwise, if IFS<1, the inspected food safety status is acceptable, and the index value of 2; otherwise, if IFS>l, the inspected food safety status is not acceptable, the risk is high, and the index value is 3.

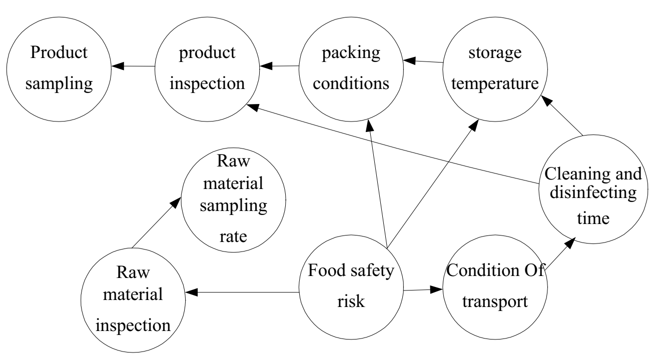

Figure 3. Network structure diagram of reasoning model of food safety control knowledge based on fuzzy Bayesian network.

Raw material inspection: It refers to the inspection of raw materials and auxiliary materials including freshness and microbial content. The risk grade is evaluated by the inspector with professional knowledge. Value set is {1, 2, 3}, each represents {low level, middle level, high level}.

Raw materials sampling rate: It refers to the proportion of the extracted raw materials samples of the total number of all materials samples. It is mainly used to describe whether the inspector's operation is standard, whether raw materials inspection is representative. The general value is 5%.

Cleaning and disinfecting time: The cleaning purpose is to maintain the environmental sanitation, improve product quality, maintain equipment performance, and reduce energy consumption. Disinfection aims to remove pathogens, improve quality, and prolong shelf period. Cleaning and disinfection equipment, disinfection methods, disinfection temperature are different in food production, therefore, the disinfection time is used as index in this paper to investigate whether cleaning and disinfection condition meets the requirements .

Storage temperature: Storage is designed to ensure the demand of food production or continuous market supply. Due to improper storage operation, the food in the storage process may spoil. The main influencing factors including temperature, light, sealing and so on. Microbes, like all the living things, need the right environment for growth, reproduction. If the required conditions, such as temperature and humidity, for the growth of bacteria can be known and controlled, bacterial reproduction can be effectively prevented, and even the number of the bacteria will be reduced. So here, the storage temperature is used as indicators for assessment of storage condition.

Product inspection: It is mainly to test whether a product complies with the corresponding product standards, whether it is defective, whether harmful substances are contained, and whether the production process is appropriate. The risk grade is evaluated by the inspector with professional knowledge. Value set is {1, 2, 3}, each represents {low level, middle level, high level}.

Product sampling rate: It refers to the proportion of the extracted product of the total number of all products. It is mainly used to describe whether the inspector's operation is standard, whether product inspection is representative. The general value is 5%.

Packaging conditions: Packaging materials, selection of technology and improper use will produce great problems on food safety and quality. Some residual impurities such as adhesives, paint, and ink remain in the packaging materials. Improper treatment of such impurities can contaminate food, ranging from heavy smell in the product to some toxic substances permeate into food. In the food packaging operation process, if the environment is not up to the standard of aseptic degree or packaging sterilization is not complete, bacteria will multiply, and food will spoil. The risk grade is evaluated by the inspector with professional knowledge. Value set is {1, 2, 3}, each represents {low level, middle level, high level}.

Traffic condition: It mainly refers to whether the environment temperature, light, and humidity in the transport processes meet the requirements. The risk grade is evaluated by the inspector with professional knowledge. Value set is {1, 2, 3}, each represents {low level, middle level, high level}.

Table 3. Part of modeling data.

Sample No.	Condition of transport	Cleaning and disinfecting time (m)	Storage Temperature (°C)	Packing condition	Product inspection	Product sampling rate	Raw material inspection	Raw materials sampling rate	Food safety risk
1	1	25	5	3	3	1%	2	5%	3
2	2	30	3	3	1	10%	3	5%	1
3	1	50	-1	3	1	6%	1	9%	2
4	2	45	6	3	3	4%	2	10%	2
5	2	50	0	2	3	7%	3	8%	2
6	3	30	4	1	3	5%	1	7%	1
7	1	30	-1	3	2	2%	2	9%	3
8	2	45	3	1	1	5%	2	9%	3
9	1	30	7	2	2	3%	3	8%	1
10	3	50	4	3	1	4%	1	3%	2
11	3	50	3	1	1	1%	1	2%	2
12	2	60	6	1	2	2%	1	5%	1
13	2	60	1	2	2	4%	2	8%	1
14	3	30	2	3	3	1%	2	9%	3
15	2	40	4	1	3	6%	2	1%	1
16	1	30	2	2	2	2%	3	1%	1
17	2	30	0	3	1	1%	1	3%	2
18	3	55	5	3	3	2%	1	2%	3
19	2	50	7	3	3	5%	2	7%	2
20	3	60	5	2	3	10%	1	4%	3
…	…	…	…	…	…	…	…	…	…

Table 4. Algorithm of parameter settings.

Algorithm	Populati on size	Maximum iterations	The number of crossover probability	Mutation probability
FBN	8	70	0.5	0.5

Table 5. CFPT of Attribute 0.

$P(x_0)$	CFPT of Attribute 0	Attribute 0
0.037645	1,1.12	1
0.290156	2,1.89	2
0.672199	3,1.15	3

Table 6. CFPT of Attribute 1.

$P(x_1\|\pi_1)$	Membership function parameters of attribute 1	Membership function parameters of attribute 0	Node 1	Node 0
0.457775	1,1.121	1,1.12	1	1
0.254721	2,1.888	1,1.12	2	1
0.287505	3,1.151	1,1.12	3	1
⋮	⋮	⋮	⋮	⋮

Modeling data: There are 100 samples used for modeling data, part of the modeling data are shown in Table 3. The 100 samples are randomly divided into 10 parts, 8 randomly selected as a model, 2 as model test.

The model construction: C# programming language and Microsoft Visual Studio 2005 programming are used in this paper to realize fuzzy Bayesian network algorithm based on genetic algorithm. The related tests are taken in the paper. parameter settings are shown in Table 4, the results of modeling are shown in Tables 3 -11.

Comparison between FBN and BN based on genetic algorithm: Aimed at data of 100 samples shown in Table 3, models are constructed by FBN and BN based on genetic algorithm. The parameter settings of the two kinds of algorithm are shown in Table 12, the results of modeling are shown in Fig. 4 and Table 13. The model structure comparison map under the two algorithms is shown in Fig. 4; all contrast index comparison results are shown in Table 13.

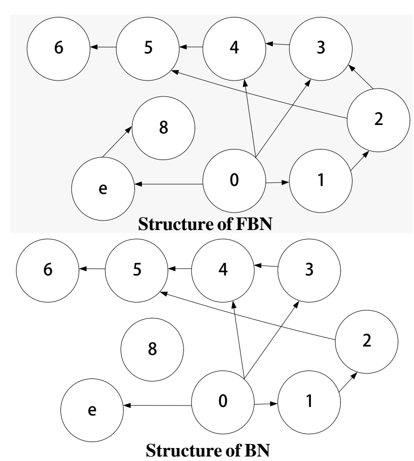

Figure 4. Structural comparison table between FBN and BN.

From a structural comparison (Fig. 4), FBN has one more node than BN, it also has 2 directed edges, and these directed edges or nodes are fuzzy random variables or have relationships among

Table 7. CFPT of Attribute 2.

$P(x_2\|\pi_2)$	Membership function parameters of attribute 2	Membership function parameters of attribute 1	Node 2	Node 1
0.492204	15.5,1.536	1,1.12	1	1
0.17647	41.91,1.161	1,1.12	2	1
0.331326	58.98,1.531	1,1.12	3	1
⋮	⋮	⋮	⋮	⋮

Table 8. CFPT of Attribute 3.

$P(x_3\|\pi_3)$	Membership function parameters of attribute 3	Membership function parameters of attribute 2	Membership function parameters of attribute 0	Node 3	Node 2	Node 0
0.315081	-2.77,1.063	15.5,1.536	1,1.12	1	1	1
0.053979	0.22,1.505	15.5,1.536	1,1.12	2	1	1
0.63094	9.31,1.387	15.5,1.536	1,1.12	3	1	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 9. CFPT of Attribute 4.

$P(x_4\|\pi_4)$	Membership function parameters of attribute 4	Membership function parameters of attribute 3	Membership function parameters of attribute 0	Node 4	Node 3	Node 0
0.408398	1,1.241	-2.77,1.063	1,1.12	1	1	1
0.31364	2,1.496	-2.77,1.063	1,1.12	2	1	1
0.277962	3,1.918	-2.77,1.063	1,1.12	3	1	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 10. CFPT of Attribute 5.

$P(x_5\|\pi_5)$	Membership function parameters of attribute 5	Membership function parameters of attribute 4	Membership function parameters of attribute 2	Node 5	Node 4	Node 2
0.022482	1,1.99	1,1.241	15.5,1.536	1	1	1
0.30251	2,1.787	1,1.241	15.5,1.536	2	1	1
0.675008	3,1.746	1,1.241	15.5,1.536	3	1	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 11a. CFPT of Attribute 6.

$P(x_6\|\pi_6)$	Membership function parameters of attribute 6	Membership function parameters of attribute 5	Node 6	Node 5
0.042862	4.00%,1.928	1,1.99	1	1
0.266325	6.93%,1.395	1,1.99	2	1
0.690813	8.36%,1.265	1,1.99	3	1
⋮	⋮	⋮	⋮	⋮

Table 11b. CFPT of Attribute 7.

$P(x_7\|\pi_7)$	Membership function parameters of attribute7	Membership function parameters of attribute 0	Node 7	Node 0
0.184063	1,1.305	1,1.12	1	1
0.239771	2,1.868	1,1.12	2	1
0.576166	3,1.703	1,1.12	3	1
⋮	⋮	⋮	⋮	⋮

Table 11c. CFPT of Attribute 8.

$P(x_8\|\pi_8)$	Membership function parameters of attribute8	Membership function parameters of attribute7	Node 8	Node 7
0.176855	3.74%,1.17	1,1.12	1	1
0.164612	6.14%,1.689	1,1.12	2	1
0.658533	8.08%,1.555	1,1.12	3	1
⋮	⋮	⋮	⋮	⋮

nodes. Compared with Fn, FBN has a certain degree of improvement in the following aspects: maximum likelihood estimation, reasoning error, reasoning correct rate, especially the reasoning error improvement rate reached 40%. On the other hand, we can also

Table 12. Algorithm of parameter settings.

Algorithm	Population size	Maximum iterations	Crossover probability	Mutation probability
FBN	8	70	0.5	0.5
BN	8	50	0.5	0.5

find that, at the time of the operation, FBN costs more than 1.48 min than BN, i.e. more than 47.74%.on the aspect of the convergence algebra, FBN is 8 generations more than BN, an increase of 80%. In minimum description length, FBN is 18.79% more than BN, that is to say, FBN is complex than BN.

The reason of the above result is mainly as follows: during the process of the discretization, the only interval ownership of the attribute value makes some characteristics to be filtered, and these characteristics are preserved in fuzzy system. Therefore, after the fuzzy mechanism is introduced, correct rates of the reasoning will rise, and reasoning error will be reduced, but due to the introduction of membership, either this or that relation under "discrete division" among the nodes are not so absolute, but allows intermediate transition, thereby increasing the network node number and the number of directed edges. On the other hand, FBN increased the calculation of the membership degree, the forward and backward reasoning expenses, so the time cost of FBN is greater than BN. FBN based on genetic algorithm needs to be improved in the rate of algorithm.

Table 13. Comparison between FBN and BN based on genetic algorithm.

Evaluation index \ Algorithm	BN	FBN	The value of the improvements	Improvement rate
MLE	-251.35	-206.22	45.13	17.96%
MDL	328.90	390.71	-61.81	-18.79%
ICE	5	3	2	40.00%
ICR	82.41%	88.69%	6.28%	7.62%
Time	3.1(m)	4.58(m)	-1.48	-47.74%
Convergence algebra	10	18	-8	-80%
Node	{1, 2, 3, 4, 5, 6, 7}	{1, 2, 3, 4, 5, 6, 7, 8}		
The number of edges	9	11		

Table 14. Forward reasoning result of attribute 1 (transport condition).

Sample number	The original level rate	Each level of reasoning of fuzzy probability			Reasoning level	Reasoning of the fuzzy probability	Match or not
		Low level	Middle level	High level			
1	1	0.902	0.039	0.059	1	0.902	Yes
2	2	0.024	0.89	0.086	2	0.89	Yes
3	1	0.89	0.059	0.051	1	0.89	Yes
4	2	0.068	0.909	0.023	2	0.909	Yes
5	2	0.113	0.771	0.116	2	0.771	Yes
6	3	0.187	0.082	0.731	3	0.731	Yes
7	1	0.89	0.081	0.03	1	0.89	Yes
8	2	0.248	0.35	0.402	3	0.402	No
9	1	0.941	0.054	0.006	1	0.941	Yes
10	3	0.012	0.113	0.875	3	0.875	Yes
11	3	0.34	0.399	0.262	2	0.399	No
12	2	0.045	0.887	0.068	2	0.887	Yes
13	2	0.081	0.866	0.053	2	0.866	Yes
14	3	0.041	0.445	0.514	3	0.514	Yes
15	2	0.103	0.849	0.048	2	0.849	Yes
16	1	0.928	0.043	0.029	1	0.928	Yes
17	2	0.069	0.857	0.074	2	0.857	Yes
18	3	0.23	0.16	0.611	3	0.611	Yes
19	2	0.386	0.556	0.059	2	0.556	Yes
20	3	0.044	0.015	0.941	3	0.941	Yes
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

The result test: As there are many nodes in FBN, to avoid repeating, we take forward reasoning of attribute 1 (transport condition) as an example, as shown in Table 14. From the table, forward reasoning results of attribute 1 are quite ideal, the reasoning of majority of the samples are basically correct. Of 100 samples 89 can be recognized, so the recognition rate is up to 89%, of course, reasoning accuracy still needs further improvement.

With the above model, we can use that model for reasoning. The examples are as follows: (1) Under the known circumstance of part factors affecting the food safety risk, the level of the fuzzy probability and severity that food safety risks may occur is reasoned. For example, the known material sampling rates is 2%, from the CFPT of attribute 8, we know that Gauss membership function parameters of the raw materials sampling rate shown in Table 15, substitute 2% in the Gauss membership function, and apply calculation formulas of fuzzy probability on mixed event, we can get membership under the various levels of raw materials sampling rate and fuzzy probability of each level, as shown in Table 16, α, the parameter value of fuzzy probability formulas on mixed event is 0.92. Compared the membership size, we can know it is most likely that the raw materials sampling rate attached to low levels, and its probability of occurrence is 0.902. Combined with the CFPT of attribute 0, attribute 7, and attribute 8, according to the Zadeh fuzzy reasoning method, membership of the food safety risk is calculated, and its value is 0.385. The calculation process is as follows:

Table 15. Parameters of Gauss membership function of attribute 8 (raw materials sampling rate).

	Low level (1)	Middle level (2)	High level (3)
mean	3.74%	6.14%	8.08%
standard deviation	0.0317	0.02689	0.02555

Table 16. The reasoning of the potential risk of attribute 8 (the raw materials sampling rate).

	Low level (1)	Middle level (2)	High level (3)
Membership	0.7402509	0.093906736	0.00347983
Fuzzy probability	0.9017464	0.095593828	0.00265977

$$P(x_0=3|x_8=1)=\frac{P(x_0=3,x_8=1)}{P(x_8=1)}=$$

$$\frac{\sum_{\gamma=1}^{3}P(x_0=3|x_7=\gamma)*P(x_7=\gamma|x_0=3)*P(x_0=3)}{P(x_8=1)}=0.2$$

The rule of the results is expressed as: R_1: Raw material sampling rate has a low level, degree of membership is 0.74, → food safety risk degree is high (fuzzy probability is 0.2, membership is 0.385).

The analysis of results: Raw material inspection rate is 2% ,which belongs to low level compared with the general level of 5%.It means that the inspector's operation is not standardized, thus fuzzy probability of food safety accident is 0.2, indicating the fuzzy probability of the accident is smaller. The reason is as follows:

Although the inspector's operation is not very standardized, which may make the unqualified raw material enter the production process, but the subsequent cleaning and disinfection can exclude the danger. Meanwhile, if product inspection is standardized, potential risk of raw materials can still be found in inspection, thus excluding the risk, so that the fuzzy probability of the food safety accidents will be greatly reduced.

(2) Similarly we can perform diagnostic reasoning: when the food safety risk occurs, influence of certain link can be diagnosed. For example, membership of the food safety risk under the high level is 0.67, then use the above model for reasoning to get the following rules.

The rule of the results is expressed as R_2: Food safety risk degree is high, 0.67, → transport condition has a low level (fuzzy probability is 0.438, membership degree is 0.67).

The analysis of results: If the food safety risk belongs to a high level, the possibility of inferior transport conditions is 0.438, and the membership is 0.67, which tells the transport condition is serious. It is necessary to remind the food transportation logistics personnel to improve the conditions of transport, such as light, temperature, and humidity, especially temperature to meet the requirements.

Conclusions

In view of the knowledge reasoning with fuzzy random uncertainty problems, in-depth study on construction of fuzzy Bayesian network is carried out in this paper to get the following conclusions:
(1) A mixed event and fuzzy probability representation is defined, concept of CFTP is presented, and the parameter α of the fuzzy probability is optimized for the analysis. Through the value of the parameter, influence degree of randomness and fuzziness can be measured. The Bayesian network can apply in widen fields from the traditional, discrete, and random fields to continuous, fuzzy and fuzzy random mixed ones.
(2) "Genetic algorithm based on fuzzy Bayesian network algorithm" is put forward, many concepts are defined such as the encoding, the crossover and mutation operator of the network structure, mutation operator of CFPT, crossover and mutation operator of the fuzzy probability parameters, the reasoning membership error and the level error, etc. With the feedback of the two kinds of error on learning of network structure and parameter, learning of fuzzy Bayesian network structure, learning of parameter, and feedback of reasoning error are achieved in one optimal learning, establishes the structure, parameters can vary with the actual problems and adaptive change fuzzy Bayesian network inference. The introduction of "membership", "level error" and "double error" make the reasoning results conform to the actual situation. Experiments show that the algorithm based on fuzzy Bayesian network has a high rate of reasoning accuracy than the general bayesian network.
(3) In view of the knowledge reasoning of the current food safety control, the application of fuzzy Bayesian network in the knowledge reasoning of food safety control is put forward. Fuzzy Bayesian network has advantages in treatment of the mixed uncertainty reasoning with intuitiveness and easy apprehension, and treatment of fuzzy knowledge with powerful fuzzy logic. With the "fuzzy Bayesian network algorithm based on genetic algorithm", knowledge reasoning model of food safety risk is established, so that the structure and parameters of the model can adapt to the practical problems and the generalization ability of the model and its application value can be improved.
(4) Experiments show that knowledge reasoning model of the food safety control based on fuzzy Bayesian network can make forward, backward and mixed reasoning. It can also give fuzziness and the randomness metrics of the reasoning. Besides, it can realize potential risk analysis, scientific definition of liability afterwards, and safety pre-warning during the food production process so that food safety risks are minimized.

Acknowledgements

Humanities and Social Science Foundation of Ministry of Education of China of 2009, NO09YJC790063. Natural Science Foundation of Heilongjiang Province of China of 2009, NO A200905. University Young Academic Backbone Supported Program of Department of Education of Heilongjiang Province of 2011, NO1251G027.

References

[1] Von Braun, J., Bouis, H., Kumar, S. and Pandya-Lorch, R. 1992. Improving Food Security of the Poor: Concept, Policy and Programs. International Food Policy Research Institute,Washington DC, pp. 25-29.
[2] Tang, Z. P. 2002. The World Agriculture Research.China Agriculture Press,Beijing, pp. 78-79.
[3] Antle, J. M. 1995. Choice and Efficiency in Food Safety Policy. The AEI Press, Washington DC, pp. 56-58.
[4] Liu, W. J. 2006. Research on Chinese food safety control. Graduate School of Northwest Agriculture and Forestry University, Northwest Agriculture and Forestry University, pp. 16-29.
[5] Běhounek, L. 2008. On the difference between traditional and deductive fuzzy logic. Fuzzy Sets and Systems **159**(10):1153-1164.
[6] Pavlaka, O. and Talaová, J. 2010. Fuzzy vectors as a tool for modeling uncertain multidimensional quantities. Fuzzy Sets and Systems **161**(11):1585-1603.
[7] Liang, X. and Pedrycz, W. 2009. Logic-based fuzzy networks: A study in system modeling with triangular norms and uninorms. Fuzzy Sets and Systems **160**(24):3475-3502.
[8] Stefanini, L. A. 2010. Generalization of Hukuhara difference and division for interval and fuzzy arithmetic. Fuzzy Sets and Systems **161**(11):1564-1584.
[9] Tepnika, M., Bodenhofer, U., Daňková, M. and Novák, V. 2010. Continuity issues of the implicational interpretation of fuzzy rules. Fuzzy Sets and Systems **161**(14):1959-1972.
[10] Kroupa, T. 2008. Filters in fuzzy class theory. Fuzzy Sets and Systems **159**(14):1773-1787.
[11] Chrysafis, K. A., Papadopoulos, B. K. and Papaschinopoulos, G. 2008. On the fuzzy difference equations of finance. Fuzzy Sets and Systems **159**(24):3259-3270.
[12] Lo, C.H.,Wong, Y.K. and Rad, A.B. 2011.Bond graph based Bayesian network for fault diagnosis, Applied Soft Computing **11**(1):1208-1212.
[13] Pernkopf, F., Van Pham, T. and Bilmes, J. A. 2009. Broad phonetic classification using discriminative Bayesian networks. Speech Communication **51**(2):151-166.
[14] Ziegler, V. 2008. Approximation algorithms for restricted Bayesian network structures. Information Processing Letters **108**(2):60-63.
[15] Goubanova, O. and King, S. 2008. Bayesian networks for phone duration prediction. Speech Communication **50**(4):301-311.
[16] Butz, C. J., Hua, S., Chen, J. and Yao, H. 2009. A simple graphical approach for understanding probabilistic inference in Bayesian

networks. Information Sciences **179**(6):699-716.
[17]Silander, T., Roos, T. and Myllymäki, P. 2010. Learning locally minimax optimal Bayesian networks. International Journal of Approximate Reasoning 51(5):544-557.
[18]de Melo, A. C. V. and Sanchez, A. J. 2008. Software maintenance project delays prediction using Bayesian networks. Expert Systems with Applications **34**(2):908-919.
[18]Armañanzas, R., Inza, I. and Larrañaga, P. 2008. Detecting reliable gene interactions by a hierarchy of Bayesian network classifiers. Computer Methods and Programs in Biomedicine **91**(2):110-121.