Food Control 78 (2017) 33–42

Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

Early warning modeling and analysis based on analytic hierarchy process integrated extreme learning machine (AHP-ELM): Application to food safety

ZhiQiang Geng [a, b], ShanShan Zhao [a, b], GuangCan Tao [c], YongMing Han [a, b, *]

[a] College of Information Science & Technology, Beijing University of Chemical Technology, Beijing 100029, China
[b] Engineering Research Center of Intelligent PSE, Ministry of Education in China, Beijing 100029, China
[c] Guizhou Academy of Testing and Analysis, Guiyang, Guizhou 550002, China

ARTICLE INFO

Article history:
Received 27 October 2016
Received in revised form
25 January 2017
Accepted 19 February 2017
Available online 20 February 2017

Keywords:
Food safety
Extreme learning machine
Analytic hierarchy process
Artificial neural network
Early warning modeling

ABSTRACT

Since the actual food safety monitoring data have characteristics of high-dimension, complexity, discreteness and nonlinear properties, it is difficult to accurately predict the risk of actual food inspection process. Therefore, this paper proposes a predictive modeling approach based on analytic hierarchy process (AHP) integrated extreme learning machine (ELM) (AHP-ELM). The proposed approach utilizes the AHP model to obtain the effective process characteristic information (PCIs). Compared with the analytic hierarchy process (AHP) integrated traditional artificial neural network (ANN) approach, the AHP-ELM prediction model is effectively verified by executing a linear comparison between all PCIs and the effective PCIs through daily inspection data source from the supervision and inspection department repository of China quality supervision system. Finally, the PCIs and the prediction value are obtained to provide more reliable food information and identification of potentially emerging food safety issues. The proposed method is applied to the food safety early warning and monitoring system in China. The result shows that the proposed model is effective and feasible in processing the complex food inspection data. Meanwhile, it can help to improve the quality of food products, ensure food safety and reduce the risk of food safety.

1. Introduction

In recent years, a series of food safety incidents not only endanger people's health and life safety directly (Xue & Zhang, 2013), but also undermine the reputation of the food processing industry in China (Lu & Wu, 2014; Pei et al., 2011). The frequent occurrences of major food safety incidents expose the imperfection and the low efficiency of Chinese food safety supervision system (Liu, Xie, Zhang, Cao, & Pei, 2013). The risk analysis-based food safety supervision system is commonly considered as the most effective approach to ensure food safety. Developed countries and regions, like Europe, the United States, Japan and South Korea have widely and commonly established food safety regulations or warning systems (Mensah & Julien, 2011; Neltner et al., 2011). The successful experience of developed countries is very useful for China to perfect our food safety regulation systems. In order to ensure food safety and protect public health, the Standing Committee of the 12th National People's Congress approved the Food Safety Law of the People's Republic of China on October 1, 2015. In the second chapter: the food safety risk monitoring and evaluation, the 17th article provides that the country builds the food safety risk assessment system. The system analyzes the risks of biological, chemical and physical hazards in food, food additives and food-related products based on the food safety risk monitoring data, relevant information and scientific method. The establishment of an effective food safety supervision and early warning system of potential food safety issues can avoid or reduce losses caused by the food safety hazards.

Currently, the early warning methods of food safety are mainly divided into the qualitative analysis method and the quantitative data prediction method. The qualitative analysis method is mainly used to draw warning conclusions from policy and theory analysis. This method is very common in food control research, analysis of

* Corresponding author. College of Information Science & Technology, Beijing University of Chemical Technology, Beijing 100029, China.
E-mail address: hanym@mail.buct.edu.cn (Y. Han).

http://dx.doi.org/10.1016/j.foodcont.2017.02.045
0956-7135/

the participants' attitudes towards food safety risk, and analysis of the participants' awareness of food safety issues (Clayton, Smith, Neff, Pollack, & Ensminger, 2015; Manning & Soon, 2013). However, the data are mainly obtained from the questionnaire, rather than the laboratory testing, so they cannot be used to accurately build the early warning model (Chow & Mullan, 2010). The quantitative data prediction method is the early warning method based on data analysis. The researchers analyzed the laboratory testing data including the results of time series and regression analysis in the daily monitoring network to get the early warning results of the food safety (De Dominicis, Commissati, Gritti, Catellani, & Suman, 2015; Den Besten & Zwietering, 2012; Qi, Zhong, & Liu, 2014). However, these statistical analysis-based methods need to know the structure of systems or the clear relationship between variables, which are often unrealistic in practice. Many regression analysis models deviate from reality, because the structure of the model itself is not clear (Granato, Calado, & Jarvis, 2014; Pan, Sun, Cheng, & Pu, 2016). The existing research method can make warning only if the current statistics or variations exceeded the standard or has reached the critical point. This kind of method is not consistent with the nature of risk prediction. We need a system to warning before the risk develops to a critical point (Griffith, 2013).

Recently, as an advanced intelligent method, artificial neural network (ANN) has been used in process modeling and early warning. Besides, a variety of network models have been derived, such as back propagation neural network (BP) (Wang., Wang., & Chen, 2011; Ding, Su, & Yu, 2011), radial basis function neural network (RBF) (Wong., Seng., & Ang, 2011; Yu, Xie, Paszczynski, & Wilamowski, 2011) and so on. However, some problems are exposed in the traditional BP and RBF, such as the slow convergence usually requires thousands of iteration, the computational complexity increases rapidly when the network has many layers and nodes. In addition, the BP algorithm is very easy to fall into local optima so that the efficiency of the network training process cannot be guaranteed (Xie., Yu., & Wilamowski, 2011). In 2006, Huang et al. proposed a learning algorithm for a single hidden layer feed-forward neural network, called extreme learning machine (ELM) (Huang, Zhou, Ding, & Zhang, 2012, 2015). Under the premise of random selection of input layer weight, the algorithm calculates the output layer weight by using the Moore-Penrose generalized inverse matrix, and the generalization ability of the network is guaranteed. Compared with the BP and the RBF, the ELM has faster learning speed, less adjustable parameters and strong generalization ability (Wang, Cao, & Yuan, 2014; Huang, 2014; Pratama., Lu., & Zhang, 2015). Moreover, the ELM has been successfully used in speech recognition (Lu et al., 2016), face recognition (Peng, Wang, Long, & Lu, 2015), biological information processing (Li, Deng, He, & Du, 2010) and fault diagnosis (Yang, Wang, & Zhong, 2016).

However, since data produced by the inspection process have the characteristics of high-dimensional attributes, complexity, discreteness and nonlinear properties, the food safety inspection process model based on the ELM is inaccurate. To deal with the high dimensional data modeling problem, scholars explored many high dimensional data feature extraction methods, including principal component analysis (PCA) (Geng., Chen., & Han, 2016), input training neural network (Chen, Jiang, Li, Jia, & Ghamisi, 2016) and analytic hierarchy process (AHP). Among them, the AHP is a hierarchical and multi-criteria analysis method that combines qualitative analysis and quantitative analysis (Bernardon et al., 2011; Kou & Lin, 2014; Li, Anavatti, & Ray, 2014; Yigit, Isik, & Ince, 2014) to obtain the process characteristic information through data compression and data filtering. By combining the AHP and the ELM, this paper firstly extracted and preprocessed the raw data to fuse the classified reduction indicators. Then we got three kinds of early warning indicators as the output of the ELM: the risk of heavy metal pollution, the risk of chemical contaminant pollution and the risk of pathogenic bacteria pollution. Finally, we established a food safety warning model based on the AHP-ELM. The proposed method has been applied to the food safety early warning and monitoring system in one province in China. The result shows that the proposed model is effective and feasible in dealing with complex food inspection data. Meanwhile, it provides a new idea for the early warning of food safety and can help to improve the quality of food products, ensure food safety and reduce the risk of food safety.

2. Preprocessing of food inspection data

2.1. Data source and preprocessing

In this paper, from 2010 to 2014, a total of 18657 lines of the daily inspection data of food safety have been obtained from the Analysis and Testing Institute of one province in China. A part of the inspection data are shown in Table 1.

Table 1 shows the characteristics of food safety inspection data:

(1) High dimensional attribute. Each food inspection information consists of different attributes, such as manufacturers information, origin information, product information, production date, inspection items, inspection results and so on. The attributes are independent of each other and represent different dimensions, so we need a method to incorporate all of these items into a model or a system.
(2) Complexity. The description of inspection items are very complicated and their values contain many descriptive language. For example, because the inspection process includes sensory inspection, the table shows color is milky white as a descriptive result. These descriptive information are unquantifiable, leading to the increasing of analysis difficulty.
(3) Discreteness. The entire table has about 50 kinds of inspection indicators. The dimension of one indicator is different from the other. And the national standard requirements for each indicator are also different. This means that these indicators have different discrete domains, and the inspection results are disordered in the discrete domain. In addition, some inspection results are boundary descriptions and cannot determine their specific values, for example, the number of *Escherichia coli* does not exceed 100.
(4) Nonlinear properties. Each inspection information includes numerical information and description information. The numerical information contains a lot of null values and the description information contains some discrete values. Therefore, in a complete inspection information, the distribution of information between different attributes is asymmetrical and unbalanced. In other words, the emergence frequencies of two kinds of attributes are very different.

We have the following preliminary treatments of the raw data (Sampers, Toyofuku, Luning, Uyttendaele, & Jacxsens, 2012):

(1) Remove useless information. Each sample includes 31 descriptive fields. For analysts, it is necessary to simplify the information to emphasize the safety features. We remove food sensory quality items that are not closely related to food safety, such as the taste, the color, the shape and so on.
(2) Remove extra symbols. For example, the inspection result of total arsenic in a sample is "<0.011". While the standard limit value of total arsenic is less than or equal to 0.25, so we remove the "<" symbol in the result and record the inspection value as "0.011".

Table 1
Raw data of food inspection between 2010 and 2014.

Bar code	Commodity name	Specifications	Date of production	Date of inspection	Inspection item name	Unit	Technical indicator	Inspection result	Individual evaluation	Inspection standard
6923644270940	Deluxe organic milk	250 ml per box	8/11/2013 00:00:00	1/3/2014 00:00:00	color	–	Milky white or slightly yellow.	Milky white	Qualified	GB 25190-2010
6923644270940	Deluxe organic milk	250 ml per box	8/11/2013 00:00:00	1/3/2014 00:00:00	Chromium	mg/kg	≤0.3	0.021	Qualified	GB/T 5009.123 −2003
6907992100272	Pure milk	900 g per tin	5/3/2013 00:00:00	5/3/2013 00:00:00	protein	g/ 100 g	≥2.9	3.19	Qualified	GB/T 5009.123 −2003
6917878020112	Nestle LACTOGEN	250 ml per box	4/2/2014 00:00:00	4/8/2014 00:00:00	iron	mg/ 100 g	5.00–10.01	6.34	Qualified	
6907992100289	Yili Pure milk	1 L	11/8/2013 00:00:00	11/8/2013 00:00:00	Melamine	mg/kg	≤2.5	Negative	Qualified	GB/T 5009.123 −2003
6907992507064	Low lactose milk	250 ml per pack	11/21/2013 00:00:00	11/21/2013 00:00:00	AFT M1	μg/kg	≤0.5	Negative	Qualified	GB/T 5009.123 −2003

(3) Use a minimum value instead of zero. For example, the national standard of China requires to prohibit the addition of melamine in dairy products and the actual inspection result of the sample was not detected. In order to ensure the generated information matrix is a positive definite matrix, we use a very small positive number 0.001 instead of zero.
(4) If the inspection result is positive and does not contain special characters, we add the value to the information matrix directly. For example, the inspection result of vitamin D is 9.70, so we add 9.70 to the information matrix.

A part of pre-processed inspection data are shown in Table 2:

2.2. Classification of inspection indicators

There are many kinds of food inspection indicators (Jacxsens et al., 2010), including pesticides, veterinary drug, biological toxins, food additives, non-food additives (Huang, Min, Duan, Wu, & Li, 2014), prohibited ingredients, persistent organic pollutants, processed products, pathogenic bacteria (Gentschew & Ferguson, 2012; National food safety standards of China, 2013) and so on. For existing data table, there are about 50 kinds of inspection indicators. In this paper, we selected 11 kinds of important inspection indicators and then categorized them (Van Boekel et al., 2010; Dong et al., 2015). This 11 kinds of indicators can directly determine whether the product is safe (National food safety standards of China, 2012; National food safety standards of China, 2010) Table 3.

3. The AHP-ELM method

3.1. Analytic hierarchy process (AHP)

The traditional AHP model constructs the judgment matrix by establishing a hierarchical structure model and combining with expertise. Firstly, the process is affected by people's subjective consciousness, thus it is prone to errors in judgment. The rough comparison and calculation process of the traditional AHP method does not meet the requirements of high accuracy in the actual inspection process. Secondly, the data of food inspection process are complex and the data items are various, so it is not realistic for people to build up a precise mathematical model. Using the data driven technology, we only need to capture the input and output signals without knowing the internal principle of the production process. It's obviously impossible to predict all items one by one as there are too many inspection indicators and results are complicated. Therefore, this paper uses the AHP algorithm based on correlation function differential driver to fuse different indicators in the inspection sample, and ultimately constructs the output of food safety inspection data.

Suppose the matrix Y is the initial sample data and its attributes:

Definition 1. Suppose the j parameters of the underside correlation function (Han, Geng, & Liu, 2014) of the sample is:

$$c_{ij}(y) = \begin{cases} 0 & y \notin \left[y_j(1), y_j(4)\right] \\ \dfrac{y_{ij} - y_j(1)}{y_j(2) - y_j(1)} & y \in \left[y_j(1), y_j(2)\right] \\ 1 & y \in \left[y_j(2), y_j(3)\right] \\ \dfrac{y_j(4) - y_{ij}}{y_j(4) - y_j(3)} & y \in \left[y_j(3), y_j(4)\right] \end{cases} \tag{1}$$

where $i = 1, 2 \ldots n$ and $j = 1, 2 \ldots m$.

The preprocessed information matrix is as follows:

$$C_{n\times m} = \begin{bmatrix} c_{11} & c_{12} & \ldots & c_{1m} \\ c_{21} & c_{22} & \ldots & c_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ c_{n1} & c_{n2} & \ldots & c_{nm} \end{bmatrix} \tag{2}$$

Center normalization:

$$\begin{cases} c'_{ij} = \left(c_{ij} - \overline{c_j}\right) \Big/ S_j \\ \overline{c_j} = \dfrac{1}{n}\sum_{i=1}^{n} c_{ij} \\ S_j = \sqrt{\dfrac{1}{n}\sum_{i=1}^{n}\left(c_{ij} - \overline{c_j}\right)} \end{cases} \tag{3}$$

where $i = 1, 2 \ldots n$, $j = 1, 2 \ldots m$, then move the negative to zero (the zero is positive, represented by zero plus c positive decimal ε) $d_{ij} = c'_{ij} - t_j + \varepsilon, i = 1, 2 \ldots n \; j = 1, 2 \ldots m$. where $t_j = \min(a'_{ij}) < 0, j = 1, 2 \ldots m$. And get the positive matrix $D^j_{n\times m}$:

$$\mathrm{D}^j_{n\times m} = \begin{bmatrix} d_{11} & d_{12} & \ldots & d_{1m} \\ d_{21} & d_{22} & \ldots & d_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ d_{n1} & d_{n2} & \ldots & d_{nm} \end{bmatrix} \tag{4}$$

Use $D^j_{n\times m}$ to get the n-dimensional matrix:

$$COR = DD^T = \begin{bmatrix} o_{11} & o_{12} & \ldots & o_{1n} \\ o_{21} & o_{22} & \ldots & o_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ o_{n1} & o_{n2} & \ldots & o_{nn} \end{bmatrix} \tag{5}$$

Table 2
Food inspection pre-processed data.

id	coliform group	total numbers of colony	*Escherichia coli*	lead	copper	arsenic	acephate	hexachlorocyclohexane	fenitrothion	cypermethrin	dichlorodiphenyltrichloroethane
1	2	13	0.001	0.201	0.02	0.045	0.017	0.045	0.044	9.3	0.174
2	2	49	0.001	0.243	3.93	0.037	0.09	0.028	0.084	17.3	0.161
3	0.001	47	1	0.276	1.55	0.006	0.055	0.196	0.021	3.1	0.04
4	3	32	1	0.017	0.81	0.176	0.054	0.054	0.422	10.6	0.049
5	3	46	2	0.126	4.86	0.126	0.025	0.094	0.464	6.6	0.153
6	2	47	3	0.29	1.58	0.058	0.001	0.005	0.13	8.4	0.02
7	3	31	1	0.065	4.2	0.099	0.016	0.012	0.127	4.9	0.008
8	2	28	2	0.012	4.76	0.193	0.015	0.096	0.16	13.1	0.078
9	3	50	3	0.297	2.02	0.052	0.092	0.145	0.132	10.5	0.098
10	1	0.001	0.001	0.3	2.04	0.006	0.097	0.12	0.256	17	0.156
11	0.001	26	3	0.012	0.67	0.035	0.082	0.099	0.471	16.4	0.01
12	3	10	0.001	0.218	3.47	0.064	0.058	0.146	0.295	0.6	0.124
13	3	32	1	0.033	2.51	0.101	0.001	0.122	0.274	2.3	0.168
14	1	15	0.001	0.018	1.41	0.006	0.06	0.109	0.117	10.2	0.148
15	3	2	3	0.093	2.93	0.036	0.03	0.198	0.297	14.3	0.065
16	1	28	2	0.21	4.37	0.088	0.002	0.056	0.419	13.2	0.077
17	0.001	40	2	0.18	0.13	0.013	0.054	0.075	0.327	0.2	0.124
18	0.001	22	1	0.019	4.07	0.06	0.043	0.118	0.47	1.4	0.085
19	0.001	34	2	0.072	1.49	0.17	0.03	0.178	0.233	1	0.047
20	3	44	2	0.129	4.39	0.074	0.077	0.145	0.447	1.3	0.191

Table 3
Initial inspect indicators and its category.

ID	inspect indicators	category
1	coliform group	pathogenic bacteria
2	total numbers of colony	pathogenic bacteria
3	*Escherichia coli*	pathogenic bacteria
4	lead	heavy metal
5	copper	heavy metal
6	arsenic	heavy metal
7	acephate	chemical contaminant
8	hexachlorocyclohexane	chemical contaminant
9	fenitrothion	chemical contaminant
10	cypermethrin	chemical contaminant
11	dichlorodiphenyltrichloroethane	chemical contaminant

For the n symmetric matrix, the eigenvalue $W = (\omega_1, \omega_2, \ldots, \omega_n)^T$ can be obtained by the root mean square of product method (geometric mean method):

$$\omega_i = \frac{\left(\prod_{j=1}^{n} o_{ij}\right)^{1/n}}{\sum_{i=1}^{n} o_i} (i = 1, 2, \ldots, n) \tag{6}$$

Use W to fuse sample, the food inspection features fusion data are obtained:

$$Y^T W = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ y_{n1} & y_{n2} & \cdots & y_{nm} \end{bmatrix}^T \begin{bmatrix} \omega_1 \\ \omega_2 \\ \cdots \\ \omega_n \end{bmatrix} \tag{7}$$

The AHP can filter the redundant information in the high-dimensional input, extract the feature of components and effectively remove the noise.

3.2. Extreme learning machine

The ELM neural network is composed of input layer, hidden layer and output layer, and the neurons of the two adjacent layers are all connected. The weights between the input layer and the hidden layer can be randomly generated, and the weights between the hidden layer and the output layer are obtained by using the least square method. Because the parameters are not adjusted, so the ELM has an extremely fast training speed. The network structure is shown in Fig. 1:

Suppose there are p arbitrary training samples, which the set is represented as: $P = \{(x_p, y_p) | p = 1, 2, \ldots, P; x_p \in R^V; y_p \in R^W\}$

The ELM network model with L hidden layer nodes is defined as follows:

$$y_p = \Phi(x_p) = \sum_{l=1}^{L} \beta_l g(\omega_l \cdot x_p + b_l) \tag{8}$$

where $\omega_l = [\omega_{l1}, \omega_{l2}, \ldots, \omega_{lV}]^T$ are the weight vector between the *Vth* input layer node and the *lth* hidden layer node, *bl* is the threshold of the *lth* hidden neuron, $\beta_l = [\beta_{l1}, \beta_{l2}, \ldots, \beta_{lW}]^T$ are the weight connecting the *lth* hidden layer neurons to the *Wth* output layer neurons, $\omega_l \cdot x_p$ is the inner product between ω_l and x_p, $g(\cdot)$ is a linear activation function.

In the ELM theory, β_l, ω_l and b_l are existed for the *lth* hidden layer neurons. And a activation function has been defined, which makes the single hidden layer feed forward neural network can approximate the sample expectation of y_p with the error of zero. Therefore, the Eq. (8) can be rewritten as the following equation:

$$G\beta = y \tag{9}$$

$$G = \begin{bmatrix} g(\omega_1 \cdot x_1 + b_1) & \ldots & g(\omega_l \cdot x_1 + b_L) \\ \ldots & \ldots & \ldots \\ g(\omega_1 \cdot x_P + b_1) & \ldots & g(\omega_l \cdot x_P + b_L) \end{bmatrix}_{P \times L} \tag{10}$$

$$\beta = \left[\beta_1^T, \beta_2^T, \ldots, \beta_L^T\right]_{L \times W}^T \tag{11}$$

$$y = \left[y_1^T, y_2^T, \ldots, y_L^T\right]_{L \times W}^T \tag{12}$$

where G is the output matrix of hidden layer, β is the vector of weights between the hidden layer neurons and the output layer neurons, and y is the outputs vector of the ELM model. The output weights β can be calculated by the least square solution of minimal 2- norm of Eq. (9):

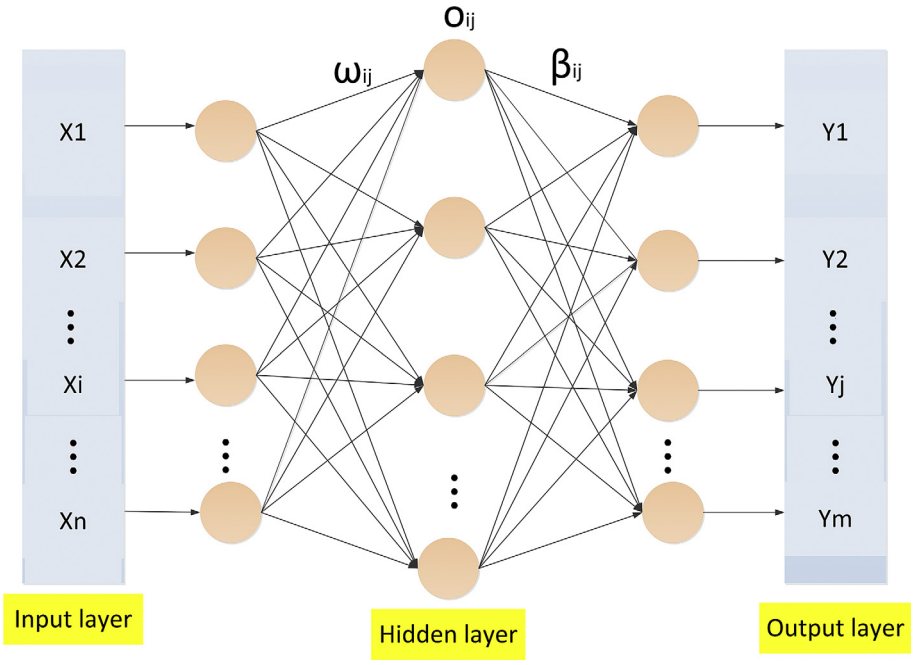

Fig. 1. Conventional ELM structure.

$$\widehat{\beta} = G^{+}y \tag{13}$$

where G^{+} is the Moore-Penrose generalized inverse of G. Therefore, if known the set of learning samples, the number of hidden layer neurons and the activation function of hidden layer, the ELM algorithm steps are:

(1) Randomly generate the vector of weights ω between the input layer neurons and the hidden layer neurons and the threshold b of the hidden neuron;
(2) Calculate the output matrix G of the hidden layer;
(3) Calculate the vector of weights $\widehat{\beta}$ between the hidden layer neurons and the output layer neurons by Eq. (13).

3.3. The framework of early warning model of food safety

Suppose there are K training samples with high dimensional input, which the set is represented as:$S = \{(x_{ij})|i = 1,2,\ldots,K; j = 1,2,\ldots,J\}$, the input of each training sample contains J attributes. For the training samples S, the modeling process based on AHP-ELM is as follows:

Step 1: Construct the output. Using the method of inspection indicators classification in Section 2.2 to divide J input attributes of the set S, record the classification group number M, the attributes number of each group and the specific input attributes. The J group attributes are used as the input vectors of the AHP model. After feature extraction and fusion, a new set of samples $S' = \{(y_{it})|i = 1,2,\ldots,K; t = 1,2,\ldots,M\}$ are obtained, where M group attributes are the output vectors of the AHP model.

Step 2:The establishment of ELM neural network. The training set S as the input of ELM neural network, the sample set S' obtained by step(1) as the expected output of the ELM neural network, so the network has J neurons in the input layer and M neurons in the output layer. The contribution of each feature is different due to the different description and the non-uniform dimension. In order to reduce the difficulty of training, each variable should provide the same importance and would vary in not too wide range. So we normalize the data set before training by using the following conversion formula:

$$\begin{cases} x_{ij} = \dfrac{x_j^{\max} - x_{ij}}{x_j^{\max} - x_j^{\min}} \\ y_{it} = \dfrac{y_t^{\max} - y_{it}}{y_t^{\max} - y_t^{\min}} \end{cases} \tag{14}$$

If $x_j^{\max} = x_j^{\min}$, $x_{ij} = -1$, where in Eq. (14) $x_j^{\max} = \max\{x_{1j}, x_{2j}, \ldots, x_{Kj}\}$, $x_j^{\min} = \min\{x_{1j}, x_{2j}, \ldots, x_{Kj}\}$, $i = 1,2,\ldots,K$, $j = 1,2,\ldots,J$. If $y_t^{\max} = y_t^{\min}$, $y_{it} = -1$, where in Eq. (14) $y_t^{\max} = \max\{y_{1t}, y_{2t}, \ldots, y_{Kt}\}$, $y_t^{\min} = \min\{y_{1t}, y_{2t}, \ldots, y_{Kt}\}$, $i = 1,2,\ldots,K$, $t = 1,2,\ldots,M$.

Step 3:Training process. Using the ELM neural network to train the normalized training sample set $(S, S') = \{(x_{ij}, y_{it})|i = 1,2,\ldots,K; j = 1,2,\ldots,J; t = 1,2,\ldots,M\}$, then record the vector of weights ω between the input layer neurons and the hidden layer neurons, the threshold b of the hidden neuron and the vector of weights β between the hidden layer neurons and the output layer neurons.

Step 4:Recall process. The AHP-ELM model has been obtained by step(3) training. We use the set S to calculate the recall output $D = \{(y'_{it})|i = 1,2,\ldots,K; t = 1,2,\ldots,M\}$ of the model, then use Eq. (15) to anti-normalize the recall output.

$$y'_{it} = y_t^{\max} - y'_{it}\left(y_t'^{\max} - y_t'^{\min}\right) \tag{15}$$

where $y_t'^{\max} = \max\{y'_{1t}, y'_{2t}, \ldots, y'_{Kt}\}$, $y_t'^{\min} = \min\{y'_{1t}, y'_{2t}, \ldots, y'_{Kt}\}$, $i = 1,2,\ldots,K$, $t = 1,2,\ldots,M$. We compare the recall output and the expected output, then calculate the recall relative error and the recall standard deviation.

Step 5:Generalization process. Using a set of data sample $(\widehat{S}, \widehat{S}') = \{(\widehat{x}_{ij}, \widehat{y}_{it})|i = 1,2,\ldots,K; j = 1,2,\ldots,J; t = 1,2,\ldots,M\}$ which is different from the training samples as a generalization sample of

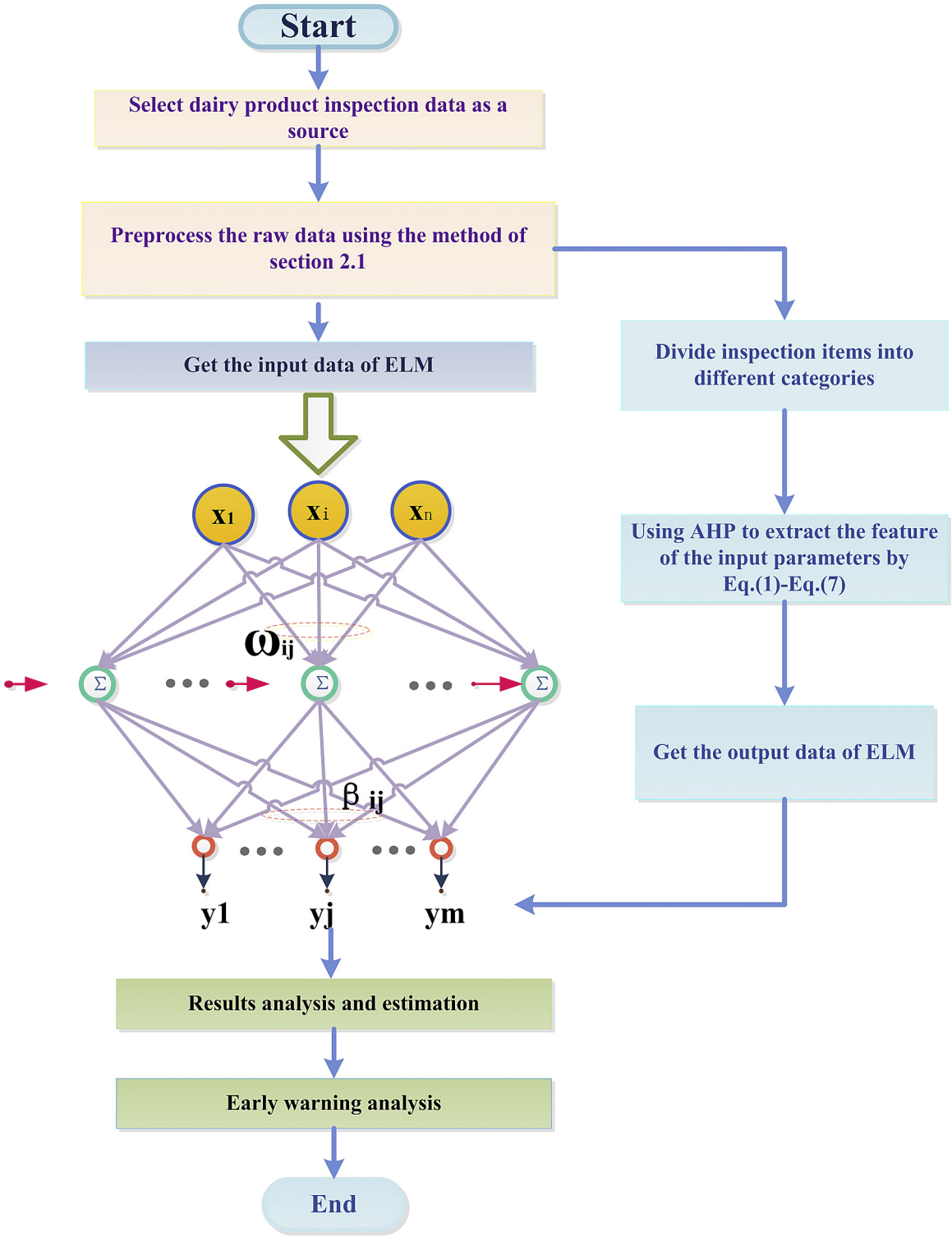

Fig. 2. The framework of the AHP–ELM algorithm.

the AHP-ELM network model. We compare the generalization output and the expected output $\widehat{S}'$, then calculate the generalization relative error and the generalization standard deviation.

The framework of early warning model of food safety is shown in Fig. 2.

4. Early warning analysis of the food safety

We select 150 inspection data of the dairy products in 2012 to establish model from the Analysis and Testing Institute of one province in China. Firstly, we preprocess these raw data, and then select 11 inspection indicators: coliform group, total numbers of colony, *Escherichia coli*, lead, copper, arsenic, acephate,

Table 4
Results of the AHP fusion.

id	Pathogenic bacteria	Heavy metal	Chemical contaminants
1	0.9267	0.8614	1.8216
2	1.6467	1.7705	2.9061
3	1.2733	1.2356	1.9236
4	1.9733	1.0603	2.4848
5	2.5867	2.0144	2.8285
6	2.6067	1.5584	0.7871
7	1.9533	1.5276	0.7285
8	1.8933	1.9433	2.0106
9	3.0000	1.6423	2.9610
10	0.3333	1.4198	3.7891
11	1.5200	0.2873	3.2027
12	1.2000	1.7282	2.6069
13	1.9733	1.0791	2.1619
14	0.6333	0.3110	2.6616
15	2.0400	1.0388	2.9851
16	1.5600	2.0094	2.2373
17	1.4667	0.6462	2.2478
18	0.7733	1.1401	2.5283
19	1.3467	1.3603	1.9686
20	2.5467	1.6601	3.5147

hexachlorocyclohexane, fenitrothion, cypermethrin and dichlorodiphenyltrichloroethane. Above-mentioned indicators which can directly affect the safety of dairy products are selected as the input items of neural network. Then, we use the AHP model to extract the features of the input parameters and obtain pathogenic bacteria, heavy metals and chemical contaminants as the output items of neural network. A part of AHP fusion results are shown in Table 4.

In this paper, the 11 inspection indicators can be selected as the input variables of the neural network. And we set the AHP fusion results as the output variables of the network which refer to pathogenic bacteria, heavy metal and chemical contaminant. We randomly select 135 groups data as the training data, and the remaining 15 groups data as the generalization data. At the same time, we assign the initial hidden layer node as 20 by using the gradual increasing method, in order to shorten the training time (Gao, 1998; Han, Geng, & Zhu, 2016). So, the nodes number of input layer, hidden layer and output layer of ELM network are 11, 20 and 3.

In order to validate the effectiveness and robustness of the AHP-ELM model, we build the BP and RBF network model to compare and verify the neural network proposed in this paper. In the single-layer BP network, we set the learning factor as 0.1, the momentum factor as 0.9, the iteration number as 1000, the excitation function as sigmoid function, the nodes number of input layer, hidden layer and output layer are 11, 20, 3. In the RBF network, we set learning factor as 0.1, the iteration number as 1000, the nodes number of input layer, hidden layer and output layer are 11, 20, 3. In addition, the average relative generalization error (ARGE), root mean square error (RMSE) and generalization standard deviation(Std) are obtained by Eq. (16), Eq. (17) and Eq. (18), respectively.

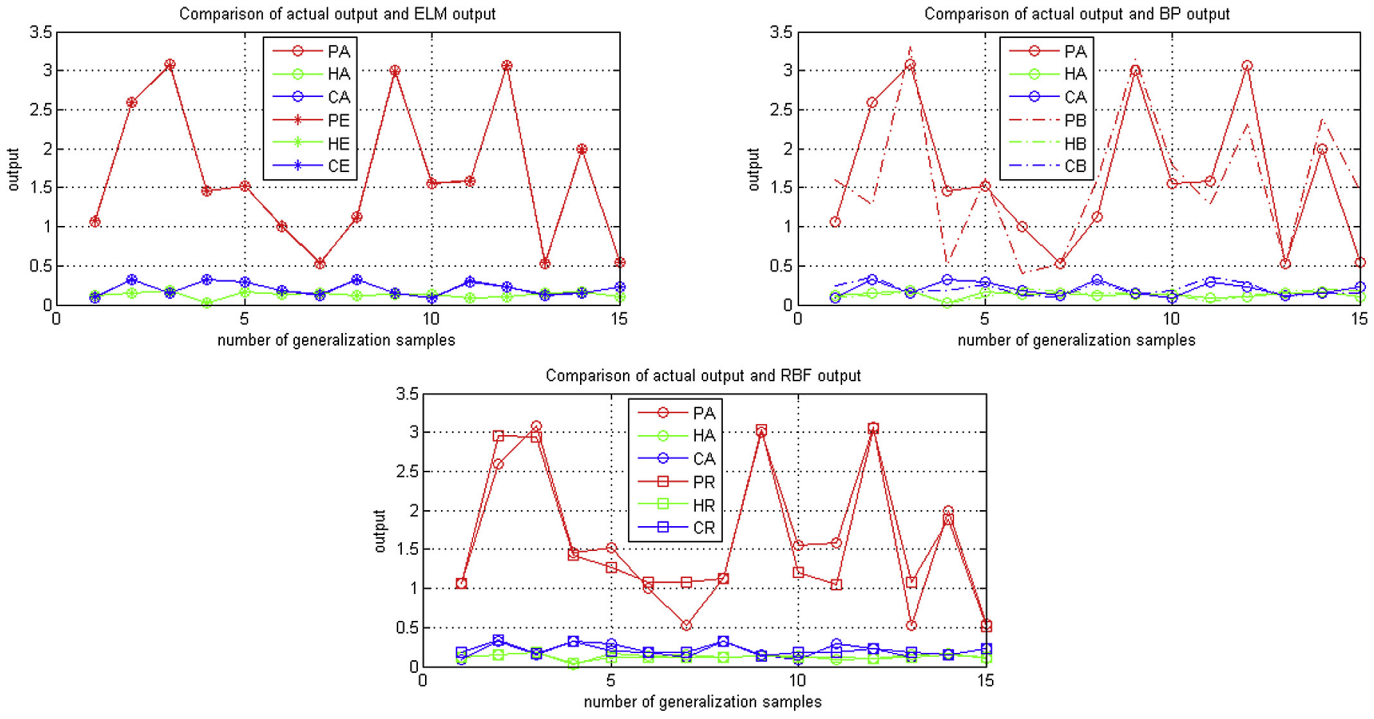

Fig. 3. Comparison of different generalization results and actual results.

Table 5
Comparison of food inspection data set.

Item	BP	RBF	RBF	ELM	ELM
Hidden nodes number	20	20	30	10	20
Running time/ms	4526	3275	4665	386	468
Average relative generalization error/%	11.746	17.936	7.879	4.962	0.528
Root mean square error	0.2239	0.2861	0.1024	0.0719	0.0071
Generalization standard deviation	0.1270	0.2495	0.1265	0.0475	0.0041

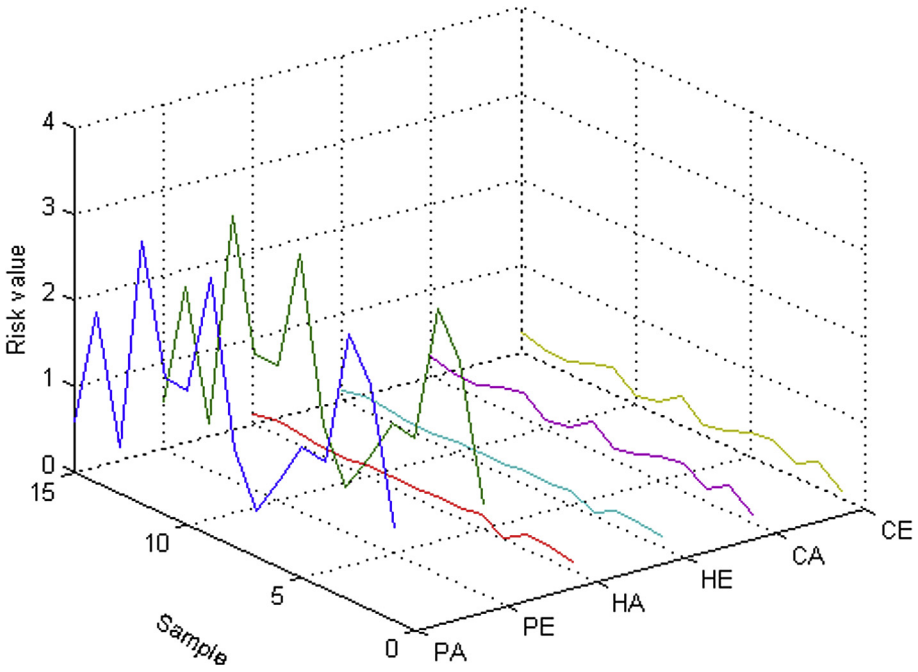

Fig. 4. Generalization curve of AHP-ELM.

Table 6
Comparison of actual value and predictive value.

ID	PA	PE	HA	HE	CA	CE
1	1.0715	1.0699	0.1193	0.1193	0.0938	0.0938
2	2.5996	2.5997	0.1512	0.1515	0.3155	0.3141
3	3.0799	3.0824	0.1794	0.1787	0.1575	0.1568
4	1.4650	1.4733	0.0274	0.0274	0.3245	0.3226
5	1.5195	1.5186	0.1690	0.1704	0.2865	0.2869
6	1.0050	0.9963	0.1303	0.1281	0.1800	0.1824
7	0.5348	0.5309	0.1442	0.1437	0.1259	0.1268
8	1.1209	1.1168	0.1245	0.1237	0.3264	0.3262
9	2.9966	2.9807	0.1306	0.1295	0.1445	0.1431
10	1.5553	1.5662	0.1309	0.1328	0.0889	0.0890
11	1.5844	1.5791	0.0932	0.0914	0.2988	0.2994
12	3.0612	3.0700	0.0954	0.0965	0.2328	0.2338
13	0.5348	0.5309	0.1442	0.1437	0.1259	0.1268
14	1.9880	1.9972	0.1624	0.1632	0.1465	0.1458
15	0.5467	0.5468	0.1102	0.1110	0.2241	0.2238

$$\begin{cases} RGE_{ji} = Abs\left(\dfrac{Pout_{ji} - Eout_{ji}}{Eout_{ji}}\right) \\ ARGE = \dfrac{\sum_{j=1}^{k}\sum_{i=1}^{n} RGE_{ji}}{k \cdot n} \times 100\% \end{cases} \tag{16}$$

$$\begin{cases} MSE_{ji} = \left(Pout_{ji} - Eout_{ji}\right)^2 \\ RMSE = \dfrac{\sum_{j=1}^{k}\sqrt{\dfrac{\sum_{i=1}^{n} MSE_{ji}}{n}}}{k} \end{cases} \tag{17}$$

$$Std = \frac{\sum_{j=1}^{k}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Pout_{ji} - \overline{Pout_j}\right)^2}}{k} \tag{18}$$

where $Pout_{ji}$ is the predictive output and $Eout_{ji}$ is the expected output, $i = 1, 2,\ldots,n$; $j = 1,2,\ldots,k$.

The results of pathogenic bacteria actual value(PA), heavy metal actual value(HA), chemical contaminants actual value(CA), pathogenic bacteria predictive value of ELM(PE), heavy metal predictive value of ELM(HE), chemical contaminants predictive value of ELM(CE), pathogenic bacteria predictive value of BP(PB), heavy metal predictive value of BP(HB), chemical contaminants predictive value of BP(CB), pathogenic bacteria predictive value of RBF(PR), heavy metal predictive value of RBF(HR), chemical contaminants predictive value of RBF(CR) are shown in Fig. 3, and the prediction accuracy and running time of each model are shown in Table 5.

As we can see from Fig. 3, the generalization results of ELM are more close to the actual value than BP and RBF. It can be seen from Table 5 that both ARGE and RMSE of the AHP-ELM model are better than that of AHP-BP and AHP-RBF. When the number of hidden layer nodes is 20, the ARGE of ELM is 0.528%, however, the ARGE of BP and RBF are 11.746% and 17.936% respectively. Secondly, the standard deviation of the ELM is smaller than the other two networks. When the number of hidden layer nodes is 20, the standard deviation of ELM is 0.0041, however, the standard deviation of BP and RBF are 0.1270 and 0.2495 respectively. Thirdly, in terms of the time required for the entire training and generalization process, ELM network is significantly faster, it is about 11 times faster than the BP network, and about 8 times faster than the RBF network. Therefore, when we use the ELM to make early warning analysis of food safety, we can predict more quickly, more stable and with less error.

We analyze 15 inspection data of the dairy products in 2012. Fig. 4 shows the comparison of the actual risks of these dairy products and the risks predicted by AHP-ELM model. Table 6 are the risks data processed by AHP-ELM model and expected results.

We also select 15 inspection data of the dairy products from Analysis and Testing Institute of one province in 2014 to make early warning analysis of food safety. After the method in Section 2.1 preprocessed, then we use the AHP-ELM model to predict the 2014 dairy products safety risks. Compared with the 2012 dairy products safety risks in Table 4, the results are shown in Fig. 5:

ELM neural network forecasts the dairy products safety risks of

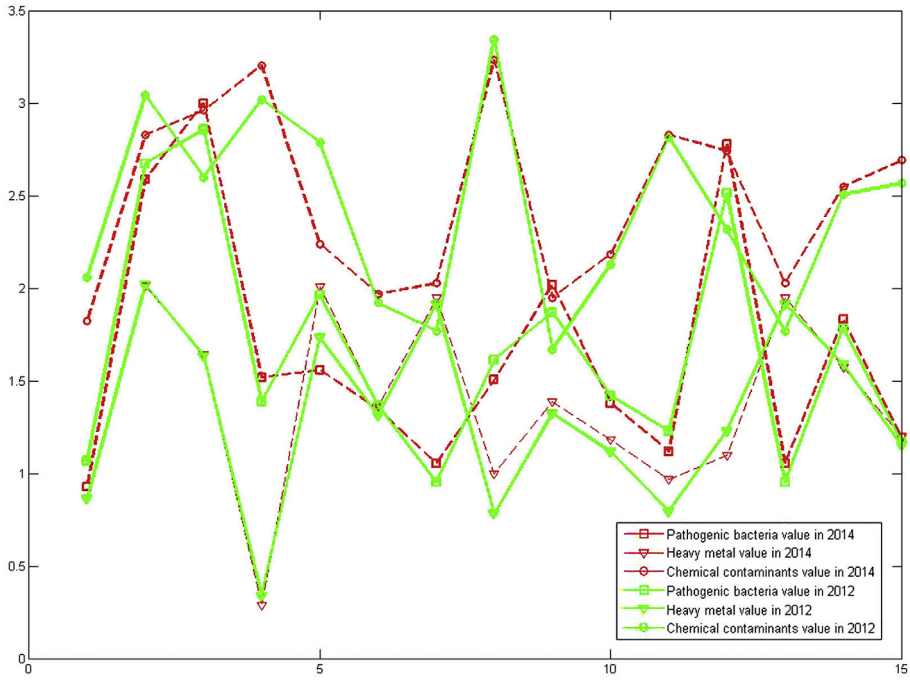

Fig. 5. Risk value comparison of dairy products between 2012 and 2014.

one province in 2014, including pathogenic bacteria, heavy metals and chemical contaminants. For example, it can be seen from Fig. 5 that the pathogenic bacteria risk values of the fifth sample in 2012 and 2014 were 1.94 and 1.57, respectively, which indicate that the pathogenic bacteria risk of this product in 2014 was reduced. The heavy metal risk values of the fifth sample in 2012 and 2014 were 1.74 and 2.05, respectively, which indicate that the heavy metal risk of this product in 2014 was increased. The chemical contaminants risk values of the fifth sample in 2012 and 2014 were 2.76 and 2.26, respectively, which indicate that the chemical contaminants risk of this product in 2014 was reduced. Therefore, we should pay attention to the heavy metal pollution of this product.

For all dairy inspection samples of one province in 2014, the safety risks were caused by pathogenic bacteria, heavy metals, chemical pollutants, and their average values were 0.7969, 0.6842, 1.5698. Compared with the dairy products safety risks of this province in 2012, the risks increased in 2014, so we analyzed that there are hidden dangers of food safety issues. Among the three dimensions of pathogenic bacteria, heavy metals and chemical contaminants, the highest degree of risk are the chemical contaminants, so the possibility of chemical pollution of dairy products in the future is also the highest, it should cause the attention of the relevant departments.

5. Discussion

First, the food safety modeling and application approach based on the AHP-ELM was proposed. Compared with the traditional quantitative data analysis approach, the robustness and effectiveness of AHP-ELM model were validated through the dairy product inspection data source from the Analysis and Testing Institute of one province.

Second, the AHP method could help to extract data features and reduce data dimensions, which can make the quantitative analysis becomes simple. And by using the ELM network based on the fused data, the input-output model of the food safety risks could be established. The model would serve as an operation guide for the food safety risks discovery and early warning, it meets the needs of the national food safety policy for early warning. And it could also be used for predicting the potential risk value related to the input inspection data. This proposed modeling method could solve problems like, when using the AHP to evaluate the food safety risks, the data samples have too many input indicators that cannot directly analysis, and when using the ELM to predict the food safety risks, the potential risk of food can be predicted quickly without using AHP to fuse a lot of inspection data by category.

Third, the current data pre-processing method is mostly based on expert opinion, therefore, we will deal with more data for more objective, such as, using some data pre-processing methods like the principal component analysis(PCA).

Finally, the proposed modeling method is applied to analyze the inspection data and predict the risks of food safety effectively, and the parameters of model are adjusted by training samples. Therefore, we will improve our model so that parameters can be adjusted automatically, such as designing a self-organizing AHP-ANN model, which has more practical value and is more suitable for practical application.

6. Conclusions

This paper proposes an AHP-ELM approach that could help to build a multi-input-multi-output optimization and prediction model of food safety inspection data. This proposed algorithm analyzes the food safety risks based on the inspection data that have characteristics of high-dimension, complexity, discreteness and nonlinear properties. In order to reduce the difficulty of processing and forecasting data, we use the AHP method to filter out redundant information and extract the characteristic variables of data as the output of the ELM. The AHP-ELM approach solves the lag problem of the traditional quantitative data analysis. Moreover, in terms of accuracy and training time, through the inspection data modeling of dairy products in 2012, we verify that the robustness and effectiveness of the AHP-ELM model are better than the AHP-BP model and AHP-RBF model. Finally, the AHP-ELM approach is applied in predicting the safety risks of dairy products in 2014. The results show that the proposed method could obtain the risks of

pathogenic bacteria, heavy metal and chemical contaminants. Meanwhile, the food inspection departments could focus on items with higher risk values by comparing the risk values of previous years. And the food inspection departments could utilize these results to effectively improve the early warning efficiency of food safety.

For further research, we will study the principal component analysis (PCA) method to optimize the data pre-processing process. We need to reduce the data dimension and maximum retain the useful information of the original variables. Moreover, the particle swarm optimization (PSO) will be used to analyze and predict the food safety risks to compare with current studies. Furthermore, we will use some self-organizing neural networks for the food safety modeling.

Acknowledgments

This research was partly funded by National Natural Science Foundation of China (61673046, 61603025), the Natural Science Foundation of Beijing, China (4162045) and the Technology Support Plan (2015BAK36B04).

References

Bernardon, D. P., Sperandio, M., Garcia, V. J., Canha, L. N., Abaide, A. R., & Daza, E. F. (2011). AHP decision-making algorithm to allocate remotely controlled switches in distribution networks. *IEEE Transactions on Power Delivery, 26*(3), 1884–1892.

Chen, Y., Jiang, H., Li, C., Jia, X., & Ghamisi, P. (2016). Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing, 54*(10), 6232–6251.

Chow, S., & Mullan, B. (2010). Predicting food hygiene: An investigation of social factors and past behaviour in an extended model of the Health Action Process Approach. *Appetite, 54*(1), 126–133.

Clayton, M. L., Smith, K. C., Neff, R. A., Pollack, K. M., & Ensminger, M. E. (2015). Listening to food workers: Factors that impact proper health and hygiene practice in food service. *International Journal of Occupational and Environmental Health, 21*(4), 314–327.

De Dominicis, E., Commissati, I., Gritti, E., Catellani, D., & Suman, M. (2015). Quantitative targeted and retrospective data analysis of relevant pesticides, antibiotics and mycotoxins in bakery products by liquid chromatography-single-stage Orbitrap mass spectrometry. *Food Additives and Contaminants Part A-chemistry Analysis Control Exposure & Risk Assessment, 32*(10), 16–17.

Den Besten, H. M., & Zwietering, M. H. (2012). Meta-analysis for quantitative microbiological risk assessments and benchmarking data. *Trends in Food Science and Technology, 25*(1), 34–39.

Ding, S., Su, C., & Yu, J. (2011). An optimizing BP neural network algorithm based on genetic algorithm. *Artificial Intelligence Review, 36*(2), 153–162.

Dong, Q. L., Barker, G. C., Gorris, L. G., Tian, M. S., Song, X. Y., & Malakar, P. K. (2015). Status and future of quantitative microbiological risk assessment in China. *Trends in Food Science and Technology, 42*(1), 70–80.

Gao, D. Q. (1998). On structures of supervised linear basis function feedforward three-layered neural networks. *Chinese Journal of Computers, 21*(1), 80–85.

Geng, Z., Chen, J., & Han, Y. (2016). Energy efficiency prediction based on PCA-FRBF model: A case study of ethylene industries. In *IEEE transactions on systems, man, and Cybernetics: Systems*. PP(99), 1–11.

Gentschew, L., & Ferguson, L. R. (2012). Role of nutrition and microbiota in susceptibility to inflammatory bowel diseases. *Molecular Nutrition & Food Research, 56*(4), 524–535.

Granato, D., Calado, V., & Jarvis, B. (2014). Observations on the use of statistical methods in food science and technology. *Food Research International, 55*, 137–149.

Griffith, C. J. (2013). Food safety: Where from and where to? *British Food Journal, 108*(1), 6–15.

Han, Y., Geng, Z., & Liu, Q. (2014). Energy efficiency evaluation based on data envelopment analysis integrated analytic hierarchy process in ethylene production. *Chinese Journal of Chemical Engineering, 22*(11), 1279–1284.

Han, Y., Geng, Z., & Zhu, Q. (2016). Energy optimization and prediction of complex petrochemical industries using an improved artificial neural network approach integrating data envelopment analysis. *Energy Conversion and Management, 124*, 73–83.

Huang, G., Zhou, H., Ding, X., & Zhang, R. F. (2012). Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, 42*(2), 513–529.

Huang, Y., Min, S., Duan, J., Wu, L., & Li, Q. (2014). Identification of additive components in powdered milk by NIR imaging methods. *Food Chemistry, 145*, 278–283.

Huang, G., Bai, Z., Kasun, L. L., & Vong, C. M. (2015). Local receptive fields based extreme learning machine. *IEEE Computational Intelligence Magazine, 10*(2), 18–29.

Huang, G. (2014). An insight into extreme learning Machines: Random neurons, random features and kernels. *Cognitive Computation, 6*(3), 376–390.

Jacxsens, L., Uyttendaele, M., Devlieghere, F., Rovira, J., Gomez, S. O., & Luning, P. A. (2010). Food safety performance indicators to benchmark food safety output of food safety management systems. *International Journal of Food Microbiology, 141*(Suppl 1), S180–S187.

Kou, G., & Lin, C. (2014). A cosine maximization method for the priority vector derivation in AHP. *European Journal of Operational Research, 235*(1), 225–232.

Li, K., Deng, J., He, H., & Du, D. (2010). Compact extreme learning machines for biological systems. *International Journal of Computational Biology and Drug Design, 3*(2), 112–132.

Li, C., Anavatti, S. G., & Ray, T. (2014). Analytical hierarchy process using fuzzy inference technique for real-time route guidance system. *IEEE Transactions on Intelligent Transportation Systems, 15*(1), 84–93.

Liu, S., Xie, Z., Zhang, W., Cao, X., & Pei, X. (2013). Risk assessment in Chinese food safety. *Food Control, 30*(1), 162–167.

Lu, F., & Wu, X. (2014). China food safety hits the "gutter". *Food Control, 41*, 134–138.

Lu, Z., Quo, D., Garakani, A. B., Liu, K., May, A., Bellet, A., et al. (2016). A comparison between deep neural nets and kernel acoustic models for speech recognition. *International Conference on Acoustics, Speech, and Signal Processing, 161*(2), 214–217.

Manning, L., & Soon, J. M. (2013). Mechanisms for assessing food safety risk. *British Food Journal, 115*(3), 460–484.

Mensah, L. D., & Julien, D. (2011). Implementation of food safety management systems in the UK. *Food Control, 22*(8), 1216–1225.

National food safety standards of China: Raw milk (GB/19301—2010).(2010). http://www.nhfpc.gov.cn/sps/s7891/201004/7c44639d854445f1963df8d61533ddee.shtml.

National food safety standards of China: the Limitation of Contaminants in food (GB/2762—2012).(2012). http://www.nhfpc.gov.cn/zhuzhan/zcjd/201304/e7b0c0e77f754c5aa2d3708acef0415e.shtml.

National food safety standards of China: the Limitation of Pathogenic bacteria in food (GB/29921—2013).(2013). http://www.nhfpc.gov.cn/zhuzhan/zcjd/201403/db2dc26b9938424cb8ecc52463a90792.shtml.

Neltner, T. G., Kulkarni, N. R., Alger, H. M., Maffini, M. V., Bongard, E. D., Fortin, N. D., et al. (2011). Navigating the U.S. Food additive regulatory program. *Comprehensive Reviews in Food Science and Food Safety, 10*(6), 342–368.

Pan, T., Sun, D., Cheng, J., & Pu, H. (2016). Regression algorithms in hyperspectral data analysis for meat quality detection and evaluation. *Comprehensive Reviews in Food Science and Food Safety, 15*(3), 529–541.

Pei, X., Tandon, A., Alldrick, A. J., Giorgi, L., Huang, W., & Yang, R. (2011). The China melamine milk scandal and its implications for food safety regulation. *Food Policy, 36*(3), 412–420.

Peng, Y., Wang, S., Long, X., & Lu, B. (2015). Discriminative graph regularized extreme learning machine and its application to face recognition. *Neurocomputing, 149*, 340–353.

Pratama, M., Lu, J., & Zhang, G. (2015). A novel meta-cognitive extreme learning machine to learning from data streams. In *Systems, man, and cybernetics (SMC), 2015 IEEE international conference on, Kowloon* (pp. 2792–2797).

Qi, X., Zhong, L., & Liu, L. (2014). A framework for a regional integrated food security early warning system: A case study of the dongting lake area in China. *Agriculture and Human Values, 32*(2), 315–329.

Sampers, I., Toyofuku, H., Luning, P. A., Uyttendaele, M., & Jacxsens, L. (2012). Semi-quantitative study to evaluate the performance of a HACCP-based food safety management system in Japanese milk processing plants. *Food Control, 23*(1), 227–233.

Van Boekel, M. A., Fogliano, V., Pellegrini, N., Stanton, C., Scholz, G., Lalljie, S. P., et al. (2010). A review on the beneficial aspects of food processing. *Molecular Nutrition & Food Research, 54*(9), 1215–1247.

Wang, Y., Cao, F., & Yuan, Y. (2014). A study on effectiveness of extreme learning machine. *Neurocomputing, 74*(16), 2483–2490.

Wang, J., Wang, C. H., & Chen, C. L. P. (2011). On the BP training algorithm of Fuzzy Neural Networks (FNNs) via its equivalent fully connected neural networks (FFNNs). Systems, Man, and Cybernetics (SMC). In *2011 IEEE international conference on, Anchorage, AK* (pp. 1376–1381).

Wong, Y. W., Seng, K. P., & Ang, L. M. (2011). Radial basis function neural network with incremental learning for face recognition. IEEE transactions on systems, man, and cybernetics. *Part B (Cybernetics), 41*(4), 940–949.

Xie, T., Yu, H., & Wilamowski, B. (2011). Comparison between traditional neural networks and radial basis function networks. In *2011 IEEE international symposium on industrial Electronics, Gdansk* (pp. 1194–1199).

Xue, J., & Zhang, W. (2013). Understanding China's food safety problem: An analysis of 2387 incidents of acute foodborne illness. *Food Control, 30*(1), 311–317.

Yang, Z., Wang, X., & Zhong, J. (2016). Representational learning for fault diagnosis of wind turbine equipment: A multi-layered extreme learning machines approach. *Energies, 9*(6), 379.

Yigit, T., Isik, A. H., & Ince, M. (2014). Web-based learning object selection software using analytical hierarchy process. *IET Software, 8*(4), 174–183.

Yu, H., Xie, T., Paszczynski, S., & Wilamowski, B. M. (2011). Advantages of radial basis function networks for dynamic system design. *IEEE Transactions on Industrial Electronics, 58*(12), 5438–5450.