CARPATHIAN JOURNAL OF FOOD SCIENCE AND TECHNOLOGY

journal homepage: http://chimie-biologie.ubm.ro/carpathian_journal/index.html

APPLICATION OF FOOD RISK ASSESSMENT AND EARLY WARNING METHOD BASED ON DATA MINING

Siwei Li*

School of Innovation and Entrepreneurship of Xuzhou Institute of Technology, Xuzhou, Jiangsu, 221008, China;
** lsw0635@163.com*

Article history:
Received:
15 February 2016
Accepted in revised form:
15 March 2016

Keywords:
Data mining;
food safety;
risk assessment;
early warning system

ABSTRACT

In recent years, continuously emerging food safety issues have attracted wide attention and become more and more obvious as trade liberalization, economic globalization and food trade develop rapidly. This study first gives an introduction on comprehensive assessment mathematical model, multidimensional data model, association rule mining as well as Apriori algorithm, then analyzes core concept of each algorithm; secondly, discusses process and core concept of risk assessment through analyzing food risk assessment and early warning system concretely, establishes multidimensional data model based on food safety risk indicators and observes data from different angles to make better decisions; finally, performs instance analysis according to risk assessment method.

1. Introduction

With the advent of poison rice, substandard milk powder and poison bean sprout, food safety issues have impacted people over and over again, resulting in more concerns on food safety.

China has conducted extensive researches on food safety risk assessment as attentions on food safety heat up gradually. Scholars from various countries have put forward some methods for evaluating food safety risks in recent years, including probability exposure assessment model (Changhua et al., 2012; Liwei and Hong, 2008; Jingxian et al., 2009) and gray relative analysis method (GRA) (Huixi and Rentian, 2008; Zeyi and Yaobo, 2000), and China mainly develops index assessment method (Liling et al., 2003; Rentian et al., 2008) and comprehensive assessment method. Establishing food safety risk assessment index is the foundation. Chen Yinyu et al (Yinyu et al., 2007; Beibei et al., 2008) take category of risk factors in food safety into an overall consideration, which involve features of relevant food, establishment of enterprise safety and health system as well as running conditions, national laws and regulations on production and supervision, quality of law-executor, etc., but fail to analyze risk factors under the category. Pan Chunhua et al (Chunhua et al., 2010; Jianjun and Shengpu, 2011) set some rules combining national safety standard, and then early warning information for food safety is produced after simple statistics on food detection data according to some rules. Therefore, food safety risk assessment and pre-warning are the preconditions of guaranteeing food security.

2. Materials and methods

2.1. Mathematical model for comprehensive assessment

Comprehensive assessment is a mathematical method for evaluating multiple abstract systems based on several indexes, considering transforming several indexes into an index that can reflect the comprehensive situation as the core concept (Shoukang, 2003). The assessment with sole index and clear process is called as “individual assessment”, however, place of comprehensive assessment different from individual assessment lies in assessment standard or index system varying complexity, but not the number of objects involving in the assessment (Shun and Shuxin, 2010). For example, evaluating a student’s performance belongs to an individual assessment, while an assessment on the teaching quality of a school is considered as a comprehensive evaluation.

Mathematical model for comprehensive evaluation is shown below:

(1)Confirmation of evaluation index set

Evaluation index set is on behalf of evaluation index system, generally speaking, it is expressed with a vector, and every component can present the state of the system in many respects. To analyze and evaluate running or development conditions of objects being evaluated comprehensively, it is usually expressed as $x=\left\{x^{(1)},x^{(2)},...,x^{(n)}\right\}$, of which, n state vectors is $x^{(i)}=\left\{x_{i1},x_{i2},...,x_{im}\right\}^T (i=1,2,...,n)$, and m refers to m evaluation indexes.

(2)Confirmation of evaluation set

Evaluation set, showing the level that evaluation index is likely to belong to, is usually expressed with $v=(v_1,v_2,...,v_k)$, of which, $v_i(i=1,2,...,k)$ means different evaluation levels may exist in k.

(3)Confirmation of evaluation index weight

In the practical application analysis, evaluation factors have various influences on evaluation goal, and corresponding weight vector is confirmed as $w=(w_1,w_2,...,w_n)^T$ according to practical effect of m evaluation indexes, besides, $\sum_{i=1}^{m} w_i = 1$ exists.

(4)Confirmation of comprehensive evaluation function

After confirming evaluation index set and weight, comprehensive evaluation function is presented as $y=f(w,x)$, thus, function value of comprehensive evaluation index is figured out: $y_i = f(w,x^{(i)})(i=1,2,...,n)$, and n systems are classified based on evaluation set according to $y_i(i=1,2,...,n)$, to confirm evaluation level.

2.2. Multidimensional data model

Data warehouse (DW), the core issue of multidimensional data model, is on the basis of multidimensional data model, together with online analytical processing (OLAP) tool. In other words, multidimensional data model shows a multi-dimensional space, “dimension” stands for the object that user observes, and points in the space are metrics (Shengen and Shan, 2005). In the multidimensional data model, data are organized into a multidimensional model structure, and each dimension contains several abstract layers with hierarchical definition, which provides flexibility for user observing data from multiple perspectives (Ming and Xiaofeng, 2007). For example, user may be interested in time, commodity, region and sales volume of the object when analyzing food selling (Figure 1), in the following figure, little blocks can be assumed as sales volume or other application instances for storing data according to actual situations.

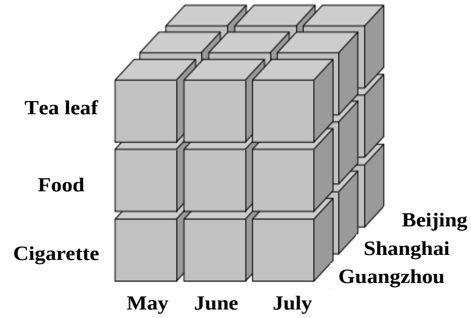

Figure 1. Data cube

2.3. Association rule mining

Association rule mining is a branch of data mining technology. Association rules, mainly applied in discovering the connection between different commodities in the transaction database, are able to reflect customer's purchasing behaviors, including input attribute and output attribute. Comparing traditional rules, difference of association rules is in that it has one or more output attributes, and output attribute of one rule can be taken as input attribute of another rule (Hongxia and Song, 2010; Jun et al., 2010; Haipeng, 2010).

2.4. Apriori algorithm

In 1994, Agrawal et al (Chunyuan, 2012) proposed that Apriori algorithm could be used in finding frequent item set in database, and the difference of which consisted in applying prior knowledge in data mining, compared with traditional algorithm. In addition, frequent item set was found to have 2 very important anti-monotone natures.

One is that sub-item set of frequent item set is surely frequent item set.

The other one is that superset of infrequent item set must be infrequent.

Based on above two natures, Apriori algorithm uses layer-by-layer iterative search algorithm together with K - item set to explore (k+1) - item set. L_k is the set of frequent K - item set and C_k refers to the set of candidate K - item set.

3. Results and discussions

3.1. Analysis on application of food risk evaluation and early warning system

3.1.1.Integrative construction of risk evaluation system

Food safety risk evaluation system is an important part of food safety risk assessment and early warning, and its reliability not only provides theoretical foundation for food risk warning, but also creates advantages for food safety supervision.

This study sets up a scientific risk evaluation system due to shortcomings in risk evaluation system and inconvenience for experts evaluating risk indexes. This system first analyzes and discusses food safety risk indexes that produce risks in detail and perfects risk evaluation index system; secondly, builds multidimensional data model on food safety risk evaluation indexes to formalize information knowledge so that experts can set level and weight; and finally, evaluates food risks using comprehensive evaluation method according to set level and weight. Process of food safety risk evaluation is shown in Figure 2.

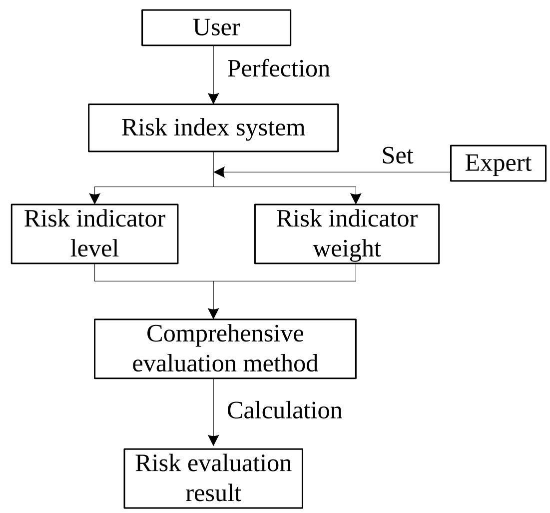

Figure 2. Process of food safety risk evaluation

3.1.2. Risk early warning

Though numerous scholars have done a large number of studies on food safety early warning, shortages exist in many aspects, for instance, early-warning object and source of the early warning information.

To date, food safety early warning is mainly targeted to test item or spot check, and early warning on food is relatively less. After food inspection, supervision department can obtain food risk level; if the risk level is high, early-earning information is sent and relevant departments make corresponding responses.

From another perspective, the food safety early warning is aimed at analyzing food safety monitoring data statistically, then, experts formulate corresponding rules for early warning, but the loss of original data and noise lead to deviation and error in the source of the early warning information.

3.2. Analysis on multidimensional data model of food safety risk index

3.2.1 Establishment of multidimensional data model

When experts evaluate food safety risks, level and weight should be set for risk evaluation index, and setting one-time cannot be suitable for all situations will happen in the future.

According to analysis above, plenty of elements are required to be considered when setting level for food safety risk evaluation indexes. Thus, traditional two-dimensional data analysis can no longer realize expected results, and multidimensional data modeling is demanded for risk evaluation index.

3.2.2. Display of multidimensional data cube

Multidimensional data cube with multidimensional data model is able to show the relationship between various dimensions more intuitively and make a convenience for setting risk index. On the foundation of diverse factors and dataset considered by experts in setting risk index, multidimensional data cube is analyzed from different angles.

Specialists focus on time, food, enterprise, nation and unqualified food when setting evaluation level for credit in enterprise. Here, because four-dimensional data cube is fairly complicated to be established, time dimension is confirmed first to analyze the influence of changes of other dimensions on setting risk evaluation index. Three-dimensional data cube made up of nation, enterprise plus food is displayed in Figure 3.

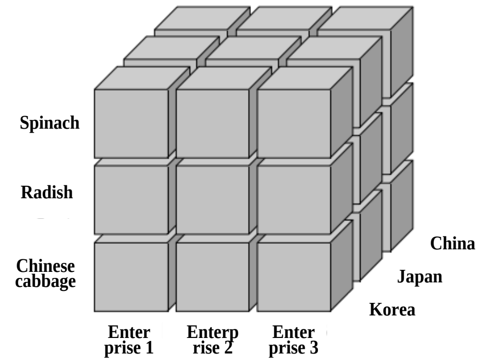

Figure 3. Credit in enterprise three-dimensional data cube

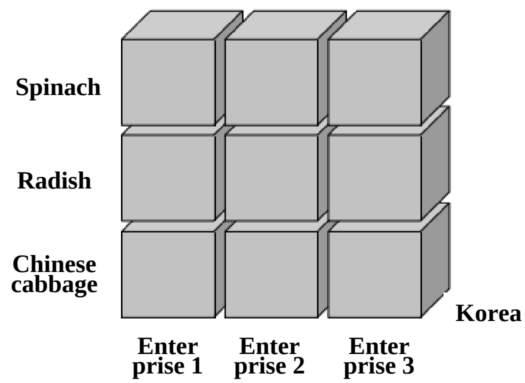

Figure 4. Volume analysis of credit in enterprise data cube

Data stored in small blocks in Figure 3 can be used for expressing unqualified food, credit in enterprise is related with time, food, enterprise and nation, and those correlation dimensions decide the unqualified food value. From a point of view of nation dimension, rolling up multidimensional data cube can acquire the tangent plane (Figure 4).

Specialists pay much attention to time, food, nation, test item and standard residual quantity when setting evaluation level for foreign technology. Time dimension is confirmed to analyze the influence of changes of other dimensions on setting risk evaluation index, and data cube consisting of nation, test item plus food dimensions is shown in Figure 5.

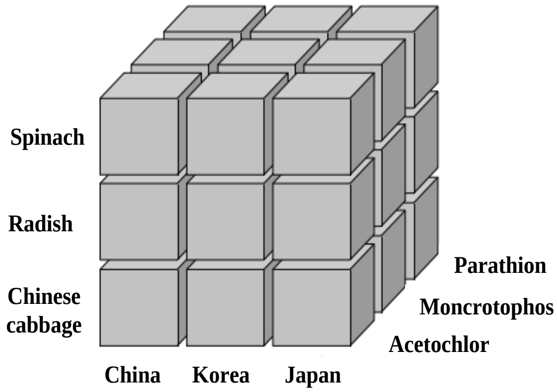

Figure 5. International technology multidimensional data cube

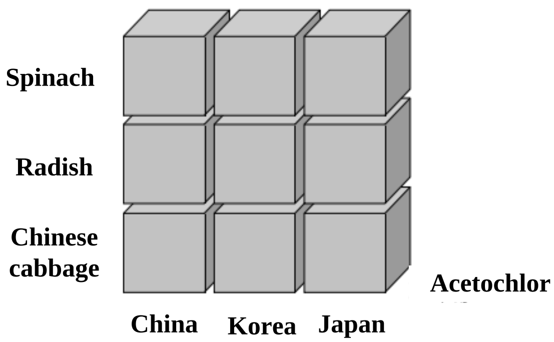

Figure 6. Volume analysis of international technology multidimensional data cube

Data stored in small blocks in Figure 5 are digital measured value, presenting standard residual quantity. Standard residual quantity is correlated with test item, food name, nation and set time in data model, and those correlation dimensions decide the standard residual quantity of test item. From nation dimension, rolling up multidimensional data cube can obtain the tangent plane (Figure 6). Relevant statistics data of standard acetochlor residues of each food in various countries in a specific period can be seen from it, and some food's standard acetochlor residues in each country are observed as well.

3.3. Mining process and result analysis

Factors that can affect food safety risk are analyzed using Apriori algorithm in association rules mining. Apriori algorithm is conducted for Boolean data while food category, risk evaluation result and time are multivalued data, so the obtained data should be converted. Food category mapping is S_{xy}, of which, x refers to category number and y expresses subclass number. Time information quarter mapping is $J_1 - J_4$, month mapping is $M_1 - M_{12}$, and enterprise risk level is set as Q_L, Q_M, Q_S, high risk, medium risk and low risk respectively. The official sets above risks as G_L, G_M, G_S, trade risk as T_L, T_M, T_S, and assumes risk evaluation result as P_L, P_M, P_S. Above data are encoded as shown in Table 1.

Table 1. Coded data

S_{01}, S_{015}, Q_S, G_S, T_S, J_3, M_9, P_S
S_{02}, S_{021}, Q_M, G_M, T_L, J_2, M_4, P_M
S_{02}, S_{022}, Q_M, G_S, T_L, J_2, M_4, P_M
S_{01}, S_{015}, Q_L, G_M, T_S, J_3, M_8, P_M
S_{02}, S_{021}, Q_M, G_M, T_M, J_2, M_5, P_M
S_{02}, S_{021}, Q_M, G_L, T_M, J_2, M_6, P_M
S_{02}, S_{022}, Q_S, G_M, T_M, J_1, M_1, P_S
S_{01}, S_{015}, Q_L, G_L, T_M, J_2, M_5, P_L

Encoded data in Table 1 are iterated applying Apriori algorithm, and minimum confidence coefficient and minimum support are set as 75% and 30% respectively. Data mining experiment is performed using Apriori algorithm, and data are the results of food inspection and reported record after food safety risk assessment in inspection and quarantine database in Qingdao from 2014 to 2015. Minimum confidence coefficient and minimum support are set as 80% and 20% respectively, to carry out association rules mining. After obtaining 1, 200 rules, association rules are restrained and screened from prior information because of too much data. As to risk evaluation level of predicted and inspected food, association rules are restrained as follows: consequent must have and only risk assessment results, and antecedent has to contain time types, together with some other attributes, the more the better. In this way, the risk evaluation results of inspected food will be more accurate.

After deleting and constraining the association rules, the number of association rules decreases by 70%, and key association rules are found out:

(1) Food category = plant food, enterprise risk = high risk, official risk = high risk, quarter = the second quarter ——> risk assessment level = high risk;

(2) Food subclass = Chinese cabbage, enterprise risk = high risk, official risk = medium risk, trade risk = low risk, quarter = the third quarter ——> risk assessment level = medium risk;

(3) Food subclass = spinach, enterprise risk = medium risk, month = August ——> risk assessment level = medium risk.

It can be observed from above example that rules acquired from association rule mining reveal connotative correlations between risk evaluation indexes in risk evaluation data. The generation of the rule has a practical significance; on the one hand, these rules obtained from a large number of historical evaluation data mining are in line with the actual meaning; on the other hand, enterprise risk will impact food export to some extent.

4. Conclusions

Realizing food safety evaluation and early warning system construction with computer technology is of great help for food safety supervision, but it comes true with difficulty due to relatively complicated information of food safety issue. Thus, this study puts emphasis on presenting the application of evaluation method, in the meantime, proposes corresponding points to process and core concept of early warning system. However, accuracy of risk early warning remains to be further improved as a certain differences are bound to exist in the collected data and actual data.

5. References

Beibei, X., Wenjun, L., Hui, L. et al. (2008). Risk Identification and Rank Confirmation in Import and Export Commodity Risk Evaluation. Inspection and Quarantine Science, 2, 56-59.

Changhua, L., Yinong, H., Yeping, T. et al. (2012). Risk Assessment and Risk Warning on Quality and Safety of Animal and Animal Products. Journal of Food Safety and Quality, 3(1), 45-52.

Chunyuan, T. (2012). Food Safety Risk Assessment and Early Warning System Based on Data Mining. Qingdao Technological University, Shandong.

Huixi, H., Rentian, J. (2008). The Improvement of National Food Security Early Warning System. Science and Technology of Food Industry, 3, 252-256.

Hongxia, W., Song, L. (2010). Research and Application about Association-rules Algorithm in Drought Forecast System. Control & Automation, 26(4), 264-269.

Haipeng, X. (2010). Forecasting Method of Stock Tend Based on Association Rules. Computer & Digital Engineering, 38(3), 150-1531.

Liling, H., Zhijian, S., Dongdong, Z., et al. (2003). Goods Entry and Exit Inspection and Quarantine Risk Management. Inspection and Quarantine Science, 13(6), 1-6.

Rentian, J., Suhua, Z., Mingzhuo, W. (2008). Construction of Food Safety Early Warning System Framework . Chinese Fisheries Economics, 26(5), 61-65.

Liwei, W., Hong, L. (2008). The Risk Assessment and Application of Chemical Contaminants in Food. Shanghai Journal of Preventive Medicine, 20(1), 26-28.

Jingxian, P., Jianjun, X., Yuanfeng, W. et al. (2009). The Safety Assessment System of Food Quality in the Domestics and Overseas. Science and Technology of Food Industry, (11), 265-267, 270.

Zeyi, Z., Yaobo, F. (2000). Fuzzing Mathematics Method in the Comprehensive Assessment of Food Contamination.

Chinese Journal of Environmental Science, 21(3), 22-26.

Yinyu, C., Xiaoli, L., Chun, C., et al. (2007). Application of Risk Management in Inspection and Supervision of Imported Food. Food Scienc, (8), 589-592.

Chunhua, P., Tonglin, Z., Mingwu, Z. et al. (2010). Research on Food Safety Information Early Warning System. Transactions of the Chinese Society of Agricultural Engineering, 26(S1), 329-333.

Jianjun, X., Shengpu, G. (2011). Study on the Construction of Data Analysis System to Food Safety Warning. Journal of Chinese Institute of Food Science and Technology, 11(4), 169-172.

Shoukang, Q. (2003). Comprehensive Evaluation Principle and Application. Publishing House of Electronics Industry, Beijing.

Shun, L., Shuxin, D. (2010). Fuzzy Comprehensive Evaluation Based on Data Envelopment Analysis. Fuzzy Systems and Mathematics, 24(2), 93-98.

Jun, X., Fang, G., Bojun, D., et al. (2010). Intelligent Recommendation of Urban Power Load Forecasting Models Based on Association Rules. Journal of Tianjin University (Science and Technology), 43(12)¸1079-1085.

Shengen, L., Shan, W. (2005). Multidimensional Data Model ER (H). Chinese Journal of Computers, 28(12), 2059-2067.

Ming, F., Xiaofeng, M. (2007). Concept and Calculation of Data Mining (Second Edition). China Machine Press.