Research Article
Study of Bayesian Network-based Vegetable Traceability Model for Quality Security

[1]Qingcong Zhao and [2]Qingling Zhao
[1]School of Information Management, Beijing University of Science and Technology Information, Beijing 100192, China
[2]Library of Beijing Union University, Beijing 100101, China

Abstract: In order to solve the problem of vegetable product's Quality Security, a Vegetable Traceability Model for Quality Security based on Bayesian network is proposed. The key factors that affect the quality and safety of vegetables in the process of vegetable production management are analyzed in the same time. Finally, the case study is given and the experimental results have shown efficiency and rationality of this modeling.

Keywords: Bayesian network, traceability mode, vegetable quality security

INTRODUCTION

Over the last few years, China has increasingly attached importance to the quality and safety of agricultural products (Cheng, 2012). By drawing lessons from the experience of developed countries in managing food safety, China established a vegetable quality and safety traceability system and strengthened the supervision of vegetable products (Qian *et al*., 2014). But this traceability system was not closely integrated with the production process management of vegetables, it was only a simple query of the plant records, packaging records, etc (Xing *et al*., 2013). It belongs to the supervision afterwards. Finding out critical factors about vegetable quality security from source of vegetable planting, combining vegetable production management flow to build prefect traceability model, this can take into account the prior supervision and after-tracking and is also a breakthrough point for the establishment of quality and safety traceability system.

Vegetable product quality and safety traceability system is a network structure, choosing an appropriate model can make the system more optimized. Drop described the food traceability models of Gozinto graphs-based and proposed a model to achieve (Van Dorp, 2003). There are many problems in the traceability system (Liu *et al*., 2014). The Bayesian network combines the knowledge of graph theory and probability theory, provides a compact graphical method for expression of complex probabilistic between random variables, there is a great advantage in solving NP problems (Campos *et al*., 2004). This study establishes a model of vegetable quality and safety based on Bayesian network.

MATERIALS AND METHODS

Materials: For establishing a traceability system for vegetable quality and safety, it is necessary to identify the hazards that may occur in the supply chain process (Zhao and Bai, 2009). From March to October 2014, at Shunxin agriculture Corporation Shandong branch, which is in Shandong Province, our research team analyzed the possible hazards of vegetables in the production process. The results of the analysis are shown in Table 1.

According to the Fig. 1, in order to ensure the quality and safety of vegetables, we must take attention to several issues in the process of the supply of vegetables. These problems are: whether there are hidden dangers in the source of vegetable seeds, planting environment, production process of agricultural inputs (fertilizers, pesticides), processing equipment and process and these hidden dangers will affect the entire supply chain, or only affect a specific batch of products; Whether the production process meets the requirements. These are also the most basic elements of quality safety traceability.

Analysis of factors affecting the quality and safety of vegetable supply chain: The production of vegetable products need to go through a long chain which from "the field" to "the table", including the supply of agricultural inputs, production, processing, circulation and sales of vegetable products. The whole production

Table 1: Hazard analysis of vegetable supply chain process

| Procedure	Hazard	Cause of hazard
Grow seedlings	Pathogenic bacteria, Pesticide residues, irradiation residues, Pollution of heavy metals and harmful chemicals	Seeds from infected areas, mixed species, Fertilizer, contaminated water, Pollution caused by the use of herbicides and pesticides
Selection of planting land	Pathogenic bacteria and parasites, Heavy metals and pesticide residues	The early crop residues in the soil, Water, environment pollution of soil
Field management after planting	Heavy metals and pesticide residues, Pathogenic bacteria, nematodes,	Farm implements used Contaminated soil, Fertilizer and contaminated water
Harvest/inspection	Pathogenic bacteria pollution, Pollution of pesticides and hazardous chemicals, Metal fragments	The spread of the collection personnel and equipment, Spraying an excess of pesticides and fertilizers, Metal fragments may come from the raw material recovery process
Processing/ classification	Pathogenic bacteria pollution	Processing personnel, equipment and communication
Tie/box	Pathogenic bacteria, Harmful chemical substances, Foreign body	The spread of processing personnel, equipment and packaging materials, Pollution of packaging materials and spray marking ink, Mud, sand, glass, etc.
Transportation/sales	Pathogenic bacteria	High temperature caused corruption
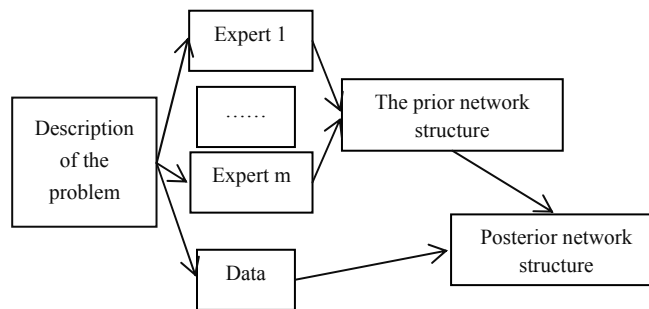
Fig. 1: Expert evaluation and historical data integration Bayesian networks structure modeling

process can be divided into three parts, that is, production, processing, circulation. Vegetable products are contaminated in any link, which will lead to unsafe vegetable products, may cause harm to human health.

In this study, views of scholars and experts have been synthesized. The factors that affect the quality and safety of vegetables are divided into three aspects: planting environment, agricultural investment and management. The planting environment plays a decisive role in the safety of vegetable production. Industrial "three wastes" (waste water, waste gas, solid waste) unreasonable emission and pesticide indiscriminately using may cause the atmosphere, water, soil pollution and ultimately cause the insecurity of vegetables. Agricultural inputs include a variety of agricultural production materials, such as chemical fertilizers, pesticides, seeds, *etc*. At present, using unqualified products in the wrong way, especially pesticides, is the main factor to influence the safety of vegetable products. In the process of vegetable supply chain, production, processing, circulation all have the problem of management, management methods and staff awareness of the safety of vegetable products. The methods of management and the awareness of the staff both have an effect on the safety of vegetable products. For example: in the production process, because of interests, Farmers' illegal use of banned pesticides and harmful additives, stored in the circulation in the wrong way, lead to the problem of the quality of vegetables.

According to the above description, Combined with the possible dangers of the supply chain of vegetables and the causes of the damage that listed in Table 1, the main factors that affect the quality and safety of vegetables are: air pollution, soil pollution, water pollution, plant selection, pest management, fertilization management, weed control, process management, transportation management, sales management and so on. These factors will directly or indirectly affect the quality and safety of vegetables, these factors themselves are closely related to each other and this relationship is dynamic: the change of a factor may result in the change of the other or a few factors, which ultimately affect the quality and safety of vegetables. Using Bayesian network to describe vegetable traceability model for quality and security, which can reflect the relationship between the factors and the change, we can deduce the trend of other related factors and its effect on the quality and safety of vegetables. When the changes of certain factors can be confirmed; According to the problems of vegetable product, we can find out the most likely problems and the causes of the problems by the posterior probability and as soon as possible to correct and remedy, so that the loss and impact as far as possible to minimize.

Bayesian network: Bayesian network, also known as belief networks, indicates that the variable dependency between the probability of a directed acyclic graph,

where each node represents a random variable, each probability dependence edges represent variables, each node corresponds to a conditional probability distribution table, indicating the probability of the number of dependent relationship between the variable with the parent node (Khakzad *et al.*, 2011). The conditional probability table can be obtained by the expert knowledge directly or through historical case data statistics.

Definition 1 Bayesian structure: about a set of variables U = {α, β, ……}, the Bayesian network is composed of two parts:

- A network structure S, independent assertion of variable condition in U, S is a directed acyclic graph; the nodes in the graph correspond to the variables in U and the conditional dependency relationship between the variables is represented by the edge.
- The P probability distribution conditions associated with each variable (Hu, 2006). S and P define a U joint probability distribution.

According to the definition of the structure of the Bayesian network, the steps for building a Bayesian network:

Step 1: Identified as model variables and their interpretation.
Step 2: The establishment of a conditional independence assertions directed acyclic graph.

The step 2 is the most critical step. The Bayesian network structure contains all the variables of the joint probability distribution and performance in the form of conditional probability as function (1):

$$P(U) = P(X_1, X_2, ... Xn) = \prod_{i=1}^{n} P(X_i \mid Pa_i) \qquad (1)$$

That Pa_i indicates the parent node set of X_i.

Otherwise, in the case of a given parent node, X and the other nodes are independent as function (2):

$$\prod_{i=1}^{n} P(X_i \mid Pa_i) = \prod_{i=1}^{n} P(X_i \mid X_1, X_2, ... X_{i-1}) \qquad (2)$$

For each variable X, there is a subset of {X_i,…, X_{i-1}} make Xi and {Xi,…,Xi-1}\\$\prod_i$ is Conditional independence, that is $P(X_i|X_1, X_2,…, X_{i-1}) = P(X_i|\prod_i)$, for any X.

From the above equation, the variable collection ($\prod_1$, …, $\prod_n$) and (Pa_1, …, Pa_n) corresponding, that is, the process of determining the parent node by looking

for the conditions between the variables independent relationship is completed to determine the conditional independence relations will determine each node's parent.

Step 3: Assigning the conditional probability distribution $P(X_i|Pa_i)$ for each node.

If there are errors in the order of the variables, the correct network structure may not be possible. In order to get the correct result, at most to consider n! Different order, very complicated. Taking into account the variable between causality easily found; causality and conditional independence assertions corresponding. Therefore, we use another method to build Bayesian networks: The network structure of the Bayesian network can be obtained from the reason that the direct result is connected to the arc.

Bayesian network structure modeling method of knowledge and data fusion: After the structure is determined, the next step is to determine the conditional probability distribution $P(X_i|Pa_i)$, this can be obtained from historical data, also can be determined by the experience of the experts. For complex problems, the relationship between variables and variables is difficult to be judged and the conditional probability distribution is difficult to be determined. Therefore, many researchers have studied a method of learning Bayesian networks from data and to shape the model instead of a human expert. From the data, Learning Bayesian networks from data can save the labor of experts and get more objective results, the drawback of this method is the slow convergence of the algorithm and the learning process requires a large amount of sample data support, if the data is less, the result is not necessarily accurate.

The combination of data and expert knowledge, using the advantages of both and make up for their shortcomings, is a more feasible method of modeling. Modeling methods are shown in Fig. 1.

The specific method as follows: A Bayesian network containing n variables, each of the variables are determined by the causal relationship between the experts, the total is n (n-1) /2. In order to avoid the subjectivity of expert knowledge, we can use the theory of evidence theory to integrate the views of many experts. Not all of the variables can directly determine the causal relationship, sometime is not clear. So only by virtue of expert opinion cannot get the final results. We adopt the method of combining knowledge with sample data: the expert's opinion has been used to determine the causal relationship between the majorities of variables, so that a large part of the network structure has been determined and then use historical data to determine the rest of the network structure.

Table 2: The relationship between the cause and effect of vegetables' quality and safety

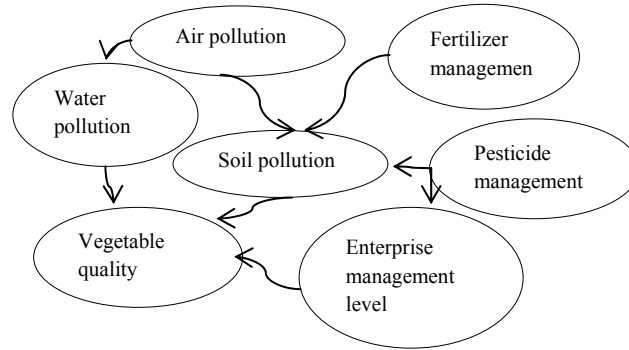| Factors/Factors	AP	SP	WP	PM	FM	EPL	VQ
AP		→	→				→
SP	←		←	←	←		→
WP	←	→					→
PM		→					→
FM		→					→
EML							↔
VQ						↔	

Fig. 2: The initial model of the Bayesian network for the traceability of the quality and safety of vegetables

RESULTS AND DISCUSSION

Vegetable traceability model for quality security based on bayesian network: According to the main factors that affect the quality and safety of vegetables, we can preliminarily determine the system variables. In order to make the model have the practical operability, we must reduce the system variables, in order to simplify the model. Combined with expert advice, we put the factors of management category, such as processing, sales, transportation management and so on for the enterprise management level to express, the management and control of weeds and insect pests are unified as pesticide management. The system variables in the final model are: Air Pollution (AP), Soil Pollution (SP), Water Pollution (WP), Pesticide Management (PM), Fertilizer Management (FM), Enterprise Management Level (EML) and vegetables' Quality (VQ). According to the expert opinion we can get the preliminary table of the relationship between the cause and effect (Table 2). The arrows to the right in this table representative row is the parent node of the column and the arrows to the left in this table representative column is the parent node of the row, bidirectional arrow indicates that the relationship cannot be determined, the blank indicates that there is no relationship between the two.

Whether the level of enterprise management can directly affect the quality and safety of vegetables, experts cannot determine their relationship, we can deduce it with combined the relevant statistical data. Taking Shunxin agriculture enterprise (Branch company of Shandong Province) as an example, according to the standard of processing management, transportation management, safety management,

Table 3: Conditional probability table of vegetable quality nodes (part)

Vegetable quality/ Enterprise management level	Advanced	Qualified	Unqualified
Qualified	96.4%	91.5%	64.6%
Unqualified	3.60%	9.50%	35.4%

enterprise management level is divided into advanced, qualified, unqualified three grades, the local agricultural product quality and safety inspection center for nearly three years, the level of enterprise management will directly affect the quality and safety of vegetables, the initial model of Bayesian network is shown in Fig. 2.

According to historical statistics, the conditional probability distribution of each node in the network can be determined, Take the enterprise management level and the quality of the vegetables as an example, the conditional probability table is shown in Table 3.

If the enterprise management level is not qualified, this condition can be used as evidence for a Bayesian network model and then get the posterior probability of unqualified vegetable quality will be as high as 35.4%, we can take some corresponding preventive measures, such as increasing the number of sampling, to prevent the inflow of substandard vegetables market.

If the conditional probability table of each node is given, we can use the Bayesian network simulation software GeNIe2. 0 to build the model and use its own reasoning function to avoid manual calculation process. In the model, if some nodes have evidence, such as the problem of the quality of vegetables, but also know that the pollution detection does not exist, it can be obtained from other nodes of the posterior probability, which makes it possible to examine the most likely part of the problem.

Establishment of food safety traceability system has become an effective way to solve the problem of food security, there are many researchers also conducted a study, but currently established traceability system mostly belongs to the post responsibility (Liu *et al.*, 2014; Qian *et al.*, 2014), rather than the prior prevention. This study build the vegetable traceability model for quality security based on Bayesian network, it realizes the aim of both "afterward" and "forward" monitor to ensure vegetable product's quality security. Complex problem involves many variables, the relationship between variables is difficult to judge and it is very difficult for the researchers to establish the Bayesian net based on the data (Khakzad *et al.*, 2011). In this study, we combine the data and expert knowledge, use the advantages of both, reduce the number of variables while ensure the validity of the model.

CONCLUSION

There are many factors that affect the quality and safety of vegetables and these factors are closely related with each other. In this study, we use Bayesian network to describe the quality and safety of vegetables, it can reflect the relationship between the factors and the change. The establishment of Bayesian networks is more difficult, especially for complex problems. The relationship between the variables and variables is difficult to judge. Combining data and expert knowledge to make use of the advantages of both to make up for their deficiencies will be a good way of modeling.

ACKNOWLEDGMENT

REFERENCES

Campos, L.M.D., J.M. Fernández-Luna and J.F. Huete, 2004. Bayesian networks and information retrieval: An introduction to the special issue. Inform. Process. Manag., 40(5): 727-733.

Cheng, Y., 2012. China's food quality and safety traceability system construction. Mod. Bus. Trade Ind., 24(20): 33-34. (In Chinese)

Hu, X., 2006. Bayesian network modeling technology and its application in decision-making. Hefei University of Technology, Hefei, pp: 45-62. (In Chinese)

Khakzad, N., F. Khan and P. Amyotte, 2011. Safety analysis in process facilities: Comparison of fault tree and Bayesian network approaches. Reliab. Eng. Syst. Safe., 96(8): 925-932.

Liu, M., Y.D. Dong, S. Zhang *et al.*, 2014. Design of quality and safety traceability system of agricultural products based on ASP.NET. Anhui Agric. Sci. Bull., 20(9): 141-143. (In Chinese)

Qian, J.P., X.X. Liu, X.T. Yang, B. Xing and Z.T. Ji, 2014. Construction of index system for traceability granularity evaluation of traceability system. T. Chinese Soc. Agric. Eng., 30(1): 98-104. (In Chinese)

Van Dorp, C.A., 2003. A traceability application based on gozinto graphs [C]. Proceeding of EFITA 2003 Conference. IOS Press, Debrecen, Hungary, pp: 280-285.

Xing, B., J.P. Qian, X.M. Wu and X.T. Yang, 2013. Design and implementation of multiple sources traceability system for fruit and vegetable product. J. Food Safety Qual., 4(6): 1705-1714. (In Chinese)

Zhao, Q.C. and R.P. Bai, 2009. Build vegetable traceability model for quality security based on Bayesian network. J. Agric. Mech. Res., 31(9): 1-5. (In Chinese)