A Dynamic Source Tracing Method for Food Supply Chain Quality and Safety Based on Big Data

Jun Song,1 Hong Huo,1 Teng Li,1 and Lingyun Chu2

1School of Management, Harbin University of Commerce, Harbin 150028, China
2School of Foreign Languages, Harbin University of Commerce, Harbin 150028, China

Correspondence should be addressed to Hong Huo; 102184@hrbcu.edu.cn

Received 28 February 2022; Revised 6 May 2022; Accepted 11 May 2022; Published 23 May 2022

Abstract

The data of food quality tracing information have several features, such as wide coverage range, many circulation links, complex data sources, low authenticity, and difficult information sharing. The continuous development of big data technology provides infinite possibilities for the construction of food quality source tracing systems. Currently, there are many studies on the application of food quality source tracing systems; however, most of them are in the field of food quality databases, and few have concerned about its application in the field of big data. Therefore, to fill in this research gap, this paper aimed to study a dynamic source tracing method for food supply chain quality and safety based on big data. At first, this paper summarized the variables of food supply chain quality and safety, constructed a Petri net model and a Bayesian network model for food quality prediction and source tracing, and realized the prediction of food quality features. Then, this paper applied two data analysis and processing methods—the density-based clustering algorithm and the cosine similarity algorithm—to preliminarily process the collected quality tracing information of each link in the food supply chain and analyzed the influencing factors of food quality. Finally, experimental results proved the effectiveness of the constructed model. Relying on the real-timeliness and authenticity of big data, this paper guarantees the credibility of the traceable information in the tracking process and improves the accuracy through real-time stream processing of the updated data, providing unlimited possibilities for the comprehensive tracking of food sources.

1. Introduction

On average, about 62.3 food quality and safety incidents occur in China per day, the overall situation is very serious, and more than 3/4 of the incidents occurring in the production and processing links are caused by human factors [1–10]. In view of the frequent food quality and safety problems and for the purpose of improving food quality and safety, it is necessary to build a food quality and safety source tracing regulatory system to track the information of all circulation links in the food industry, including production, processing, transportation, and sales [11–15]. Since the collection of food quality source tracing information requires the cooperation of all participants in the food supply chain, it has the characteristics of a wide coverage range, many circulation links, complex data sources, low authenticity, and difficult information sharing [15–22]. Today's smart energy manufacturing relies on big data, cloud computing, and the Internet of things. By combining big data with food, it is possible to grasp the basic laws of food production and improve the resource utilization efficiency and production efficiency of food production. Through data analysis, processing, data mining, and decision-making, the combination between big data and food would lower the trial-and-error cost in the development of the modern food industry, making the final decision more accurate. In this way, it is possible to reduce the costs and enhance the profits. The continuous development of big data technology provides infinite possibilities for the construction of food quality source tracing systems.

As intelligent technologies such as the Internet of Things (IoT) and big data have emerged and been widely applied, many researchers have done jobs to combine these emerging technologies with food quality source tracing systems. For example, Hao et al. [23] proposed a traceability scheme combining reverse search and recursive algorithm based on...
the theory of blockchain technology and designed and implemented a food safety tracing system. Then, they used experimental results to prove that the proposed scheme is more effective than the traditional source tracing models. In terms of epidemic prevention and control, food safety monitoring, data analysis, and food safety traceability are even more important. Studies on food safety source tracing systems based on big data, artificial intelligence (AI), and IoT provide ideas and methods for solving the problems of low reliability and difficult data storage during the application of traditional source tracing systems. For example, Zheng et al. [24] took rice as an example and proposed a food safety source tracing system based on RFID + QR code technology and IoT big data storage technology; by analyzing system requirements, the proposed system uses RFID and big data storage functions to obtain information in the food production process and then realizes the whole-process tracking of food production information through the designed dynamic query platform and mobile terminals. Balamurugan et al. [25] proposed an efficient food traceability management technology that makes use of the IoT. Then, they deduced a solution for data transmission and used the improved cycle data and the improved Petri network model to perform food source tracing. Singh and Jenamani [26] proposed a Casandra-based NoSQL database suitable for storing traceability data. Then, in order to test the performance of the database, they designed a test platform in the laboratory and simulated the operation of the supply chain of the Indian public distribution system to generate data. Food is one of the most important elements in international trade. Bordel et al. [27] proposed a food traceability system for international food trade based on the blockchain network and RFID tags, the system adopts a REST interface, NoSQL database, and JavaScript code. It realizes a distributed solution that collects, tags, and stores reliable information about food flows.

The review of the above literature shows that many scholars at home and abroad have studied food quality tracking methods. However, there are problems with the consistency and universality of the data storage for the tracking system. Besides, few scholars have studied food quality tracing in the field of big data. Therefore, this paper attempts to research a dynamic source tracing method for food supply chain quality and safety based on big data. The purpose is to explore the key technologies of the product quality tracing system in the food supply chain, establish a quality tracing system suitable for various types of foods, and improve consumer satisfaction and the core competitiveness of enterprises.

The main contents of this paper (1) summarize the variables of food quality and safety source tracing; (2) introduce the food quality source tracing mode; (3) model the food production process based on Petri net; (4) build a Bayesian network model for food quality prediction and source tracing to realize the prediction of food quality features; (5) give the framework of the food quality source tracing system; (6) adopt two data analysis and processing methods, the density-based clustering algorithm and the cosine similarity algorithm, to preliminarily process the collected quality tracing information of each link in the food supply chain; (7) give the technical route map of food quality source tracing, and analyze the influencing factors of food quality; (8) use experimental results to prove the effectiveness of the constructed model. Relying on the real-timeliness and authenticity of big data, this paper guarantees the credibility of the traceable information in the tracking process and improves the accuracy through real-time stream processing of the updated data, providing unlimited possibilities for the comprehensive tracking of food sources.

2. Construction of the Food Quality Source Tracing Model

Food production has many processing procedures, and there are many reasons for the causing quality and safety problems; therefore, the quality tracing information generated during the production and circulation process has many types and is transmitted and accumulated between the various supply chain links.

In the developed world, food safety tracking systems have been established to supervise the whole process of food production, from the planting and breeding of animals and plants (sources of food production) to the detailed records of production, processing, and sales. The purpose is to monitor and manage all the links for the food from the source to consumers, such as to ensure the safety and quality of the food on the tables. Any food in the market must carry a label of the required information. Besides, the production and processing records of the food must be traceable. Overall, countries around the world are very concerned with the safe development of agricultural products.

Figure 1 shows a diagram of the food quality source tracing mode. According to the figure, when food of a certain batch output from the supply chain has quality problems and its quality needs to be analyzed, the data of actual production workshop information, food quality feature information, and production process information collected from the production process of this food batch, are analyzed to find the reasons that cause the quality problems of this food batch.

During the multistage food processing, the three key links that are most closely related to quality are: raw material supply, production, and circulation. Common problems include chemical contamination, shelf-life problem, excessive harmful ingredients, and poor production sanitation conditions, etc. Figure 2 gives a summary of food quality and safety source tracing variables.

Starting from the perspectives of food production process and food quality source tracing, at first, this paper constructed a model of the food production process based on Petri net. Figure 3 shows the flow of food quality source tracing. A Petri net was used to describe the relationship between quality tracing activities and activity links in the above figure, and it can be expressed as: $M=(O, D, G)$, $O=\{o_1, o_2, o_3, o_4, o_5, o_6, o_7, o_8, o_9, o_{10}, o_{11}, o_{12}\}$, $D=\{d_1, d_2, d_3, d_4, d_5, d_6, d_7\}$, and $G=\{(o_0, d_0), (d_0, o_1), (d_0, o_2), (d_0, o_3), (o_1, o_4), (o_3, d_4), (o_10, d_0), (d_1, o_2), (d_2, o_4), (d_2, o_5), (d_2, o_6), (d_3, o_6), (o_5, d_2), (o_7, d_2), (o_8, d_7), (d_7, o_{12}), (o_2, d_7), (o_4, d_7), (o_6, d_4), (o_8, d_3), (o_{11}, d_7), (d_3, o_5), (d_4, o_7), (o_4, d_6)\}$. 

2 Discrete Dynamics in Nature and Society
It is assumed that $X=(x_{ij})$ represents the incidence matrix, and it satisfies $x_{ij}=q(d_i, e_j)-q(e_j, d_i)$. From the existing descriptions of Petri net, it can be obtained from the following equations:

$$x_{ij} = \begin{bmatrix} X_1 & X_2 \\ X_3 & X_4 \\ X_5 & X_6 \end{bmatrix},$$

$$X_1 = \begin{bmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{bmatrix},$$

$$X_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$X_3 = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$X_4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix},$$

$$X_5 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$X_6 = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{1}$$

Assuming $t=(x_1, x_2, \ldots, x_m)$ represents the $D$ vector of $M$, $e=(y_1, y_2, \ldots, y_m)$ represents the $E$ vector of $M$, and $x=(t_1, t_2, \ldots, t_m)$ represents an arbitrary transition sequence, then, the state equation of $M$ is $N=N_0+X_1D$, $N$, which is the successor of $x$ and $N=N_0|x>N$; assuming $N$ is an identity of $M$, $[N_0]$ is called the reachable identity set of $M$, then, there are:

$$N \in [N_0+d \in D \land N^t \in [N_0+N^t|d>N^u] \implies N^u \in [N_0]. \tag{2}$$

The Petri net can be expressed rigorously in math and described visually with graphs. With rich system description means and system behavior analysis technologies, the Petri net provides a solid conceptual foundation for computer science.

For the food quality source tracing model based on Petri net, its food quality evaluation index system has the characteristics of bounded and accessible, therefore, aiming at the accessibility of specific food source tracing information, this paper verified the effectiveness of the constructed model.
For the purpose of predicting the features of food quality, this paper constructed a Bayesian network model for food quality prediction and source tracing and defined the network node variables and states. Assuming $M = [N, O]$ represents the constructed Bayesian network, wherein $N = <U, R>$ represents the structure of the Bayesian network, $O = [o_1, o_2, \ldots, o_n]$ represents the probability of association between node variables, $U = [u_1, u_2, \ldots, u_m]$ represents the set of node variables in the network structure, and $R = [r_1, r_2, \ldots, r_m]$ represents the set of directed edges of association attributes. Then, the strength of association attributes represented by the directed edges between the node variables can be described by the conditional probability, which can be calculated by the following formula:

$$
O(X|Y) = \frac{O(X \cap Y)}{O(Y)} = \frac{O(Y|X)O(X)}{O(Y)}
$$

Assuming $\gamma$ represents the observed quantity, $\xi$ represents the unknown parameter, $o(\gamma, \xi)$ represents the joint distribution of $\gamma$ and $\xi$, $\rho(\gamma)$ and $\rho(\xi)$ represent edge density, and $\pi(a)\chi(\xi)$ represents the prior probability distribution of $\xi$, then, based on the multivariate Bayes theorem, we obtain the following formula:

$$
o(\gamma|\xi) = \frac{\chi(\xi)\rho(\gamma|\xi)}{\rho(\gamma)} = \frac{\chi(\xi)\rho(\gamma|\xi)}{\int \chi(\xi)(y|\xi)\rho(\xi)}
$$

Formula (5) gives the formula for solving the multivariate dependent joint conditional probability distribution under normal conditions:

$$
O(a_1, a_2, a_3, \ldots, a_m) = O(a_1)O(a_2|a_1)O(a_3|a_1, a_2) \ldots O(a_m|a_1, a_2, \ldots, a_{m-1}).
$$
After simplification, there is:

$$O(a_1, a_2, \ldots, a_j) = \prod_{i=1}^{m} O(a_j | \text{Parents}(a_j)).$$  \hspace{1cm} (6)

After training the constructed network model and completing the calculation of conditional probability, the tabular expression form of the model could be obtained.

3. The Processing of Food Quality Source Tracing Information Based on Big Data

Figure 4 shows the framework of food quality source tracing system, which is consisted of base layer, data layer, business layer, and application layer. The base layer includes the information of the actual production workshops, the information of food quality features, and the information of production process. The data layer supports data integration management, including source tracing data file storage, dataset of food quality tracing information, and data operation log. The business layer is mainly responsible for data analysis and processing, i.e., identifying abnormal data by density-based clustering algorithm, classifying data by cosine similarity algorithm, getting the source tracing results by analyzing the influencing factors of food quality, and taking the value corresponding to the maximum probability change rate.

Because the quality tracing information generated during production and circulation process has various types and is transmitted and accumulated between various supply chain links, the quality description variables have multiple correlations. In order to analyze the influencing factors of food quality, it is necessary to preliminarily process the collected quality tracing information of each link in the food supply chain, this paper mainly adopted two data processing methods: the density-based clustering algorithm and the cosine similarity algorithm.

3.1. Data Processing. In the density-based clustering algorithm, all variables were extracted from the dataset of food quality source tracking information and it satisfies $a_i \in A$, then, for the set of variables in set $A$ with a distance from $a_i$ less than the value of $\sigma$, its neighborhood can be defined as follows:

$$R_\sigma(a_i) = \{ a_j \in A: \text{dist}(a_i, a_j) \leq \sigma \}.$$ \hspace{1cm} (7)

When the number of other variables in the neighborhood of $a_i$ is greater than the manually set minimum value of variables in the neighborhood, then $a_i$ is called a core variable. If there are core variables in the neighborhood of noncore variable $a_i$ then, $a_i$ is called an abounding variable, and there are as follows:

$$a_i \in A; \exists a_j \in A; a_j \in R_\sigma(a_i) \cap A.$$ \hspace{1cm} (8)

If core variables $a_i$ and $a_j$ belong to the same neighborhood, then, it is said that $a_j$ is directly density-reachable from $a_i$. If core variable $a_i$ belongs to set $A$ which contains $n$ subdatasets of food quality tracing information, where $\{b_1, b_2, \ldots, b_m\}$ represents the subdatasets, and it satisfies $a_i = b_n$, $a_i = b_n$; subsets $b_1, b_2, b_3, \ldots, b_m$ are all core variables, and $b_{i+1}$ is directly density-reachable from $b_i$, then, it can be inferred that $a_i$ is directly density-reachable from $a_j$ by the following variable:

$$a_i, a_j \in A; a_i \in A_2; a_j \in R_\sigma(a_i).$$ \hspace{1cm} (9)

If there is variable $a_i$, and it satisfies that variables $a_i$ and $a_j$ are both directly density-reachable from $a_i$, then, their called that $a_i$ and $a_j$ are density-connected.

The cosine similarity algorithm is to calculate the cosine value of the angle between two vectors, which is represented by $\beta$. The larger the angle, the lower the similarity between the two vectors. For $\beta$, the calculation formula of the cosine function is as follows:

$$\cos(\beta) = \frac{x^2 + y^2 - z^2}{2xy}.$$ \hspace{1cm} (10)

The length of vectors $x$ and $y$ in the Cartesian coordinate system can be calculated by the following formula:

$$x = \sqrt{a_1^2 + b_1^2},$$ \hspace{1cm} (11)

$$y = \sqrt{a_2^2 + b_2^2}.$$

Assuming vector $z$ represents the distance between $x$ and $y$, then, the calculation formula of the length is as follows:

$$z = \sqrt{(a_1 - a_2)^2 + (b_1 - b_2)^2}.$$ \hspace{1cm} (12)

The following formula shows the cosine function in the Cartesian coordinate system:

$$\cos(\beta) = \frac{a_1 \times a_2 + b_1 \times b_2}{\sqrt{a_1^2 + b_1^2} \times \sqrt{a_2^2 + b_2^2}}.$$ \hspace{1cm} (13)

Assuming $X_i$ represents each component of vector $X, Y_i$ represents each component of vector $Y$, then, the similarity of the multidimensional space can be calculated by the following formula:

$$SIM = \cos \beta = \frac{X \cdot Y}{\|X\| \|Y\|} = \frac{\sum_{i=1}^{m} X_i \times Y_i}{\sqrt{\sum_{i=1}^{m} (X_i)^2} \times \sqrt{\sum_{i=1}^{m} (Y_i)^2}}.$$ \hspace{1cm} (14)

3.2. Obtaining the Source Tracing Results. Through the description of food quality source tracing, the source tracing process can be regarded as a diagnosis and reasoning process based on the constructed multiple Bayesian networks. The food quality source tracing problem can be described as assuming $A = \{a_1, a_2, \ldots, a_n, a_m\}$ represents the reasons that cause the food quality and safety problem, $W = \{w_1, w_2, \ldots, w_i, \ldots, w_m\}$ represents the types of food quality and safety problem, $wl$ represents the type of quality problem of a certain food batch, and $o(wl)$ represents the
probability that the quality problem is the $KL$-th type problem, then, $o(w_l)$ needs to satisfy:

$$
\sum_{i=1}^{m} o(w_i) = 1.
$$

(15)

The source tracing of food quality is actually to determine which type of quality and safety problem within the range of causes $[a_1, a_2, \ldots, a_0, \ldots, a_m]$ when $w_l$ occurs.

Figure 5 shows the technical route of food quality source tracing. The analysis of the influencing factors of food quality includes the following steps:

**Step 1. Analyze food quality information**

Separately, quantify each quality description variable, determine the main quality evaluation indicators of the food production or distribution process, and count the quality evaluation parameters $w_l$ of each link in the production or distribution process. Assuming $W$ represents the total quality evaluation parameter of food quality source tracing and $\theta_l$ represents the coefficient of influence of the quality information of each link on the total quality information, then, there is:

$$
W = \sum_{l=1}^{M} w_l \theta_l, \quad (l = 1, 2, 3, \ldots M).
$$

(16)

According to the above formula, the total quality evaluation indicator of food quality source tracing is equal to the weighted sum of the quality indexes of each production and circulation link.

**Step 2. Perform hierarchical index analysis on each production and circulation link**

Assuming $E_{w_l} = (e_{1l}, e_{2l}, e_{3l}, e_{4l}, \ldots, e_{il})$ represents the set of food quality information flow of link $l$, $l=(1,2,3,\ldots,m)$ and $\beta_l$ represents the weight of each quality information flow $e_{il}$ to the quality evaluation index $w_l$ of link $l$, then there is:

$$
w_l = \sum_{i=1}^{m} e_{il} \beta_i, \quad (i = 1, 2, 3 \ldots m).
$$

(17)

According to above formula, the food quality evaluation index of the independent production or circulation link is equal to the weighted sum of each quality index of this link.
Step 3. Analyze the direct influence relationship of adjacent production or circulation links

Assuming $\mu_{i,i+1}$ represents the correlation coefficient of the influence of a previous link on the latter link, if two adjacent links are correlated, then, there is $0 < \mu_{i,i+1} \leq 1$, if two adjacent links are not correlated, then, there is $\mu_{i,i+1} = 0$. The following formula gives the expression of the food quality evaluation index when two adjacent links have an influence relationship:

$$w_{i+1} = \mu_{i,i+1} w_i + \sum_{i=1}^{m} e w_{i+1} \beta_{i+1}.$$ (18)

According to above formula, when there's an influence relationship between two adjacent production and circulation links, the food quality evaluation index is equal to the quantitative value of the influence of the link and its coupling link plus the weighted sum of each quality index of the link. According to the data features of quality description variables, discretize the continuous variable $a_i$ to generate several state variables $ZT^a_i, ZT^c_i$, etc. The data involved in the process of food production and distribution have two types of variables: the discrete type, and the continuous type. The following formula gives the conversion formulas for continuous variables and discrete variables:

$$a_i \in (\bar{a} \pm \gamma), a_i = ZT^a_i,$$

$$a_i \notin (\bar{a} \pm \gamma), a_i = ZT^c_i.$$ (19)

Food quality source tracing is to find out the cause according to the features of the quality and safety problem when a batch of food has a quality and safety problem. Under the condition that the state variable probabilities related to the features of the quality and safety problem were known, this paper sorted the state variable probabilities, obtained the reason of the food quality and safety problem corresponding to the maximum probability change rate after the constructed network was updated, and then took the reason as the cause of the current food quality and safety problem.

In this paper, the state probability of quality authenticity of node variables in the output layer of the network was studied as the feature of the quality and safety problem changed from high = 100% to low = 100%, and the following formula gives the calculation of probability change:

$$\Delta o = \frac{\Delta o_{INC} + \Delta o_{DEC}}{2}.$$ (20)

4. Analysis of Experimental Results

Figure 6 and 7, respectively, show the execution time distribution of the map phase and the reduce phase of big data analysis. According to Figure 6, with the increase of the scale of quality tracing information data generated in the production and circulation process, the execution time of the map phase increased accordingly, but the increment was limited, which is mainly because the source tracing information was not stored in this phase. By comparing Figures 6 and 7, we can find that with the increase of the scale of quality tracing information data generated in the production and circulation process, the execution time of the reduce phase increased with it, and the generation time of source tracing results was relatively long, which is because the source tracing information had been stored in this phase; that is, the source tracing information was written into the HDFS file. Figure 8 shows the percentage of source tracing results generation time in the entire execution time. According to the figure, with the increase of the scale of a quality tracing information data generated in the production and circulation process, the percentage of the generation time of source tracing results in the entire execution time was controlled between 28% and 51%. Therefore, according to the analysis of the above experimental results, the food quality source tracing method adopted in modeling had little impact on the performance of the bottom execution framework of big data analysis.

Figure 9 shows the comparison results of the distribution of source tracing results generation time of different models, and the reference model is a fine-grained source tracing model, the word count model. According to Figure 9, with the increase of the scale of quality tracing information data, the generation time of source tracing results of the two models grew continuously, but there's a big difference in the growth rate. Based on these results, it can be seen that the source tracing efficiency of the proposed model was better than that of the reference model, and its performance was even better as the input data volume increased.

According to Figure 10, when the scale of quality tracing information data generated in the production and circulation process is fixed, with the increase in the number of production and circulation links, the source tracing results generation time of the reference model and the proposed model grew continuously, but at different rates. According to the figure, the generation time of source tracing result of the proposed model grew approximately linearly. This is because the proposed model’s retrieval scales of the source tracing information of each link were basically the same, so the source tracing results generation time grew linearly with the number of production and circulation links. However, for the reference model, with the increase of the number of production and circulation links, the source tracing results generation time grew rapidly. The model is a many-to-one model, and its retrieval scales multiplied, resulting in the source tracing results generation time grew faster and the performance was lower.

In this paper, the quality tracing information data were first processed into Netica software format data, then, a Bayesian network model was constructed, and the structure and parameter learning was carried out. After that, the probability of the proposed model was reasoned to evaluate the effectiveness of the constructed network in food quality and safety source tracing. After completing the analysis as the feature of quality and safety problems changed from high = 100% to low = 100%, the calculation results of the state probability of quality authenticity of node variables in the output layer of the network after the automatic update are shown in Table 1. Through observation, it was found that
when the food safety and quality feature changed from high = 100% to low = 100%, the state probabilities corresponding to the four variables of raw material supply information, production and processing process information, circulation process information, and quality inspection and storage information all changed; wherein the change of the probability of production and processing process information was the largest. Therefore, this paper judged that the reason for this type of food quality change was due to the processing environment and the processing technique.

Next, for the four variables of raw material supply information, production and processing process information, circulation process information, and quality inspection and storage information, the probability changes of the corresponding quality evaluation criteria were calculated, and the results are shown in Table 2. The specific variables include raw material batch/type 1, supplier/buyer information 2, raw material quality/key component content 3, production and processing environment information 4, production and processing equipment information 5, production and processing personnel information 6, production and processing link information 7, distribution time/person and other

### Figure 6: Execution time of the map phase of big data analysis.

### Figure 7: Execution time of the reduced phase of big data analysis.

### Figure 8: Percentage of source tracing results generation time in the entire execution.

### Figure 9: Distribution of source tracing results generation time.
information 8, delivery time information 9, distribution temperature and humidity environment 10, quality inspection result information 11, storage location/time information 12, IM/EX-warehouse information 13, and other information 14.

According to the table, when the food safety quality changed from high \( \geq 100\% \) to low \( \leq 100\% \), the probability change of the raw material quality/key component content was the largest, so this paper judged that this food quality change was caused by the raw material quality/key component content variable in the raw material supply information. Therefore, the link in which the food quality and safety problems occurred could be preliminarily located, and the problem-tracing personnel can verify the accuracy of the judgements made by the model by checking the value of raw material quality/key component content in the raw material supply information.

The probability change of the delivery time information in the table is 0.2, which means that the delivery time information has little impact on food quality features. Because in the food production and distribution process, the delivery time information is checked repeatedly before distribution and the food quality has a relatively long shelf life, a short-time delivery delay will not cause food quality defects, which is consistent with the actual production situations of enterprises.

5. Conclusion

This paper researched a dynamic source tracing method of food supply chain quality and safety based on big data. At first, it summarized the variables of food quality and safety source tracing, constructed a Petri net model and a Bayesian network model for food quality prediction and source tracing, and realized the prediction of food quality features. Then, this paper applied two data analysis and processing methods—the density-based clustering algorithm and the cosine similarity algorithm—to preliminarily process the collected quality tracing information of each link in the food supply chain and analyzed the influencing factors of food quality. The experimental results showed the distribution of the execution time of the map phase and reduce phase in big...
data analysis and the percentage of source tracing results generation time in the entire execution time, which verified that the food quality source tracing method adopted by the proposed model has little impact on the performance of the bottom execution framework of big data analysis. Moreover, this paper also gave the distribution of source tracing results generation time and the distribution of tracing time under different numbers of production and circulation links, and the results proved that the proposed model was more efficient in generating source tracing results and its performance was higher. At last, the paper also gave the probability changes of the variables of food quality and safety source tracing, the calculation results listed in the table were in line with the actual production situations of enterprises, which verified the accuracy of the judgment made by the model.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the Project of Philosophy and Social Science Research in Heilongjiang Province (Grant no. 19JE268).

References

[1] Y. He, H. Huang, D. Li, C. Shi, and S. J. Wu, "Quality and operations management in food supply chains: a literature review," Journal of Food Quality, vol. 2018, Article ID 7279491, 14 pages, 2018.
[2] X. Zhang and W. Zhu, "The study on food quality supervision using collision game model," International Journal of Advances in Computing Technology, vol. 4, no. 4, pp. 283–293, 2012.
[3] X. Yu, X. Guo, Y. Feng, and Q. A Yao, "A RFID-based collaborative process management approach for real-time food quality management," INTERNATIONAL JOURNAL ON Advances in Information Sciences and Service Sciences, vol. 4, no. 10, pp. 34–43, 2012.
[4] W.-D. Huang, S. Deb, Y.-S. Seo, S. Rao, M. Chiao, and J. C. Chiao, "A passive radio-frequency pH-sensing tag for wireless food-quality monitoring," IEEE Sensors Journal, vol. 12, no. 3, pp. 487–495, 2012.
[5] X. Ran and X. Feng, "The design and evaluation of food cold-chain emergency logistics quality management index system," in Proceedings of the ICLEM 2012: Logistig for Sustained Economic Development—Technology and Management for Efficiency, pp. 872–878, Chengdu, China, October 2012.
[6] B. Ling, J. Tang, F. Kong, E. J. Mitcham, and S. Wang, "Kinetics of food quality changes during thermal processing: a review," Food and Bioprocess Technology, vol. 8, no. 2, pp. 343–358, 2015.
[7] H.-W. Xiao, Z. Pan, A. Martynenko, C.-L. Law, and P. K. Nema, "Innovative and emerging drying technologies for enhancing food quality," Journal of Food Quality, vol. 2018, Article ID 6943078, 2 pages, 2018.
[8] Y.-h. Chen, S.-j. Huang, A. K. Mishra, and X. H. Wang, "Effects of input capacity constraints on food quality and regulation mechanism design for food safety management," Ecological Modelling, vol. 385, pp. 89–95, 2018.
[9] D. Fan, H.-M. D. Wang, S. Janaswamy, and G. A. Cavender, "Green physical processing technologies for the improvement of food quality," Journal of Food Quality, vol. 2018, Article ID 6050424, 2018.
[10] S. Wang, X. Liu, M. Yang, Y. Zhang, K. Xiang, and R. Tang, "Review of time temperature indicators as quality monitors in food packaging," Packaging Technology and Science, vol. 28, no. 10, pp. 839–867, 2015.
[11] H. Luo, Y. Li, D. Liu, Y. Ouyang, and F. Yuan, "Design of a distributed food traceability platform and its application in food traceability at Guangdong province," MATEC Web of Conferences, vol. 128, Article ID 04014, 2017.
[12] M. Ingrassia, S. Bacarella, P. Columba, L. Altamore, and S. Chironi, "Traceability and labelling of food products from the consumer perspective," Chemical Engineering Transactions, vol. 58, pp. 865–870, 2017.
[13] R. J. Molyneux, "Traceability of food samples: provenance, authentication, and curation," Journal of Agricultural and Food Chemistry, vol. 65, no. 41, pp. 8977–8978, 2017.
[14] N. A. binti Azram and R. binti Atan, "SPI-based traceability model for food document tracing," in Proceedings of the 2015 9th Malaysian Software Engineering Conference (MySEC), pp. 212–216, Kuala Lumpur, Malaysia, December 2015.
[15] P. Liu, W. Liu, Q. Li, M. Duan, Y. Wang, and Y. Dai, "A research on tracing code of culture of food safety traceability based on RFID and improved EPC," in Proceedings of the 2016 International Conference on Logistics, Informatics and Service Sciences (LISS), pp. 1–6, Sydney, Australia, July 2016.
[16] Y. G. Kim and E. Woo, "Consumer acceptance of a quick response (QR) code for the food traceability system: application of an extended technology acceptance model (TAM)," Food Research International, vol. 85, pp. 266–272, 2016.
[17] P. Martins-Lopes, S. Gomes, L. Pereira, and H. Guedes-Pinto, "Molecular markers for food traceability," Food Technology and Biotechnology, vol. 51, no. 2, pp. 198–207, 2013.
[18] I. Šenk, G. Ostojić, L. Tarjan, S. Stankovski, and M. Lazarević, "Food product traceability by using automated identification technologies," IFIP Advances in Information and Communication Technology, vol. 394, pp. 155–163, 2013.
[19] E. Rajo-Iglesias, I. Cuinas, R. Newman, M. Trebar, L. Catarinucci, and A. A. Melcon, "Wireless corner: RFID-based traceability along the food-production chain," IEEE Antennas and Propagation Magazine, vol. 2, no. 5, pp. 196–207, 2014.
[20] J. Ren, "RFID enable food supply chain traceability and safety," in Proceedings of the 2015 International Conference on Logistics, Informatics and Service Sciences (LISS), pp. 1–5, Barcelona, Spain, July 2015.
[21] S. Charlebois, B. Sterling, S. Haratifar, and S. K. Naing, "Comparison of global food traceability regulations and requirements," Comprehensive Reviews in Food Science and Food Safety, vol. 13, no. 5, pp. 1104–1123, 2014.
[22] F. Dabbene, P. Gay, and C. Tortia, "Traceability issues in food supply chain management: a review," Biosystems Engineering, vol. 120, pp. 65–80, 2014.
[23] M. Hao, H. Tao, W. Huang, C. Zhang, and B. Yang, "Food safety traceability technology based on block chain," in Proceedings of the International Conference on Big Data.
[24] M. Zheng, S. Zhang, Y. Zhang, and B. Hu, "Construct food safety traceability system for people's health under the internet of things and big data," IEEE Access, vol. 9, pp. 70571–70583, 2021.

[25] S. Balamurugan, A. Ayyasamy, and K. S. Joseph, “Enhanced petri nets for traceability of food management using internet of things,” Peer-to-Peer Networking and Applications, vol. 14, no. 1, pp. 30–43, 2021.

[26] S. K. Singh and M. Jenamani, “Cassandra-based data repository design for food supply chain traceability,” VINE Journal of Information and Knowledge Management Systems, vol. 51, no. 2, pp. 193–217, 2020.

[27] B. Bordel, P. Lebigot, R. Alcarria, and T. Robles, "Digital food product traceability: using blockchain in the international commerce," Advances in Intelligent Systems and Computing, vol. 850, pp. 224–231, 2018.