

# Model of Combined Transport of Perishable Foodstuffs and Safety Inspection Based on Data Mining

Tongjuan Liu, Anqi Hu

Information School, Beijing Wuzi University, Beijing, China

Email: 3283846941@qq.com

**How to cite this paper:** Liu, T.J. and Hu, A.Q. (2017) Model of Combined Transport of Perishable Foodstuffs and Safety Inspection Based on Data Mining. *Food and Nutrition Sciences*, 8, 760-777.

<https://doi.org/10.4236/fns.2017.87054>

**Received:** June 11, 2017

**Accepted:** July 24, 2017

**Published:** July 27, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

There is still no effective means to analyze in depth and utilize domestic mass data about agricultural product quality safety tests in china now. The neural network algorithm, the classification regression tree algorithm, the Bayesian network algorithm were selected according to the principle of selecting combination model and were used to build models respectively and then combined, innovatively establishing a combination model which has relatively high precision, strong robustness and better explanatory to predict the results of perishable food transportation meta-morphism monitoring. The relative optimal prediction model of the perishable food transportation metamorphism monitoring system could be got. The relative perfect prediction model can guide the actual sampling work about food quality and safety by prognosticating the occurrence of unqualified food to select the typical and effective samples for test, thus improving the efficiency and effectiveness of sampling work effectively, so as to avoid deteriorated perishable food's approaching the market to ensure the quality and safety of perishable food transportation. A solid protective wall was built in the protection of general perishable food consumers' health.

## Keywords

Perishable Foodstuffs, Transport Monitoring, Dada Mining, Sample Detection, Neural Network, the Classification and Regression Tree Algorithm (CART), Bayesian Network

## 1. Introduction

Explore the intrinsic correlation and rules among the massive data, then form the model, and this process is what we call data mining techniques [1] [2]. The

application of data mining is very broad, such as: retail, manufacturing, finance, banking, insurance, communications and medical services, etc. [3], and database industries which have the value and demand of analysis can all use data mining tools for purposeful mining analysis. In foreign countries, data mining technology has been widely used in many areas where some of the typical applications are as follows: using data mining techniques to analyze DNA in biological research field, identify customers' purchase patterns, and predicate the frauds which often occurs in banks and insurance companies [4] [5], the project initiated by the United States Eber Smith and Douglas joint venture, and use the data mining technology to provide a set of cold chain monitoring and analysis technology for the health care industry [6] [7] [8]; International academic journals have also appeared data mining topics or special issue of artificial intelligence, database and information processing and other areas. On the field of data mining, compared with abroad, domestic research started later, At present, many domestic research institutes, universities and colleges have carried out research about basic theory, application and other related aspects of data mining, knowledge discovery [9]-[15]. Compared with the United States and the European Union, the domestic early warning work is very limited, and there are only two types of methods concerning the current domestic agricultural food risk early warning means: the typical case notification and the simple mathematical statistics. However, with regard to the analyses in-depth and applications of a large number of agricultural food test data, there is still no effective means so far [9] [10].

Through consulting literatures home and abroad which are relevant to data mining [1]-[15] and food safety testing technology [16] [17] [18] [19] [20], I found that the studies which apply data mining to food transport metamorphism monitoring and forecasting to guide the actual food safety sampling work are limited; moreover, the studies which establish a prediction model that has high precision, strong stability and perfect explanatory are scarce. Therefore, in order to meet people's requirements for the quality safety of perishable food in transit, it is essential to establish a model which has relatively high accuracy, strong robustness, and better interpretation to predict perishable food deterioration in transit. The following takes monitoring strawberry deterioration in transit for example, collecting data and innovatively combining the models to establish a model which has relatively high precision, strong robustness, better interpretation to predict perishable food deterioration in transit. By providing typical sample test data, the predictive model guides food testing to ensure the quality and safety of perishable food during transportation.

## 2. Overview of Combination Models [21] [22] [23]

### 2.1. The Basic Principles of the Combination Model

The combination of multiple classified model structures whose every single model separately belongs to one type (the prediction result of each single model severally contains useful information from different angles) is called the combination model (the current research focus of the international machine learning

community), and its principle is to use complementary advantages of different types of single models (They are not exclusive, but interrelated reciprocally among single models), the combination model after series connections can overcome the shortcomings of single models to solve a problem jointly. That is, the purpose of the combination model is to improve the robustness of the composite system by reducing the error rate of change of the combination model (That is, turn the change rate of the system prediction error into the combination of change rates of the prediction errors of the single models). Although the combination model has been widely used in many fields (e.g. speech recognition, face recognition, medical diagnosis), research and application in monitoring of perishable food during transportation are rare.

In this paper, the neural network model, Bayesian network model, classification and regression tree model are selected to be combined. It's on the basis of the 3 principals that select the one-way models which constitute the combination model [21], and the 3 principals are: the optimal number of models, the difference of models, the principle of applicability. It's up to the merits and demerits of every model, and also, it is determined by the purpose of establishing a prediction model with high precision, strong stability and good explanation. The three models are selected based on the principle that the number of models is optimal. The neural network, the classification and regression tree and the Bayesian network are selected according to the principle of model difference and applicability. Neural network model has high accuracy of classification, but its stability is relatively poor and it's not interpretable, furthermore, it is not robust enough in the face of changes of the sensor data; Bayesian network model has not only good interpretation but also strong stability, however, it has lower accuracy of classification than Neural network model; classification and regression tree (CART) has good interpretation, but its robustness is weak. According to the advantages and disadvantages of the three models, build a combination model which has higher accuracy of classification, stronger stability and better interpretation than the single models, via track monitoring, predict the metamorphism of perishable food.

## 2.2. The Structure of the Combination Model

Serial structure, parallel structure and hybrid structure are the three structures of the combination model [22].

As the serial number increases, the precision is also increased (The final result of the combination model of the serial structure will outperform the prediction results of the single base classifiers) and the structure is simple. The accuracy of this paper is the key to the prediction model of monitoring of perishable food metamorphism in transit. Therefore, the serial structure is selected to combine the neural network model, the classification and regression tree model, the Bayesian network model. Because the learning result of the previous classifier of the series structure feeds the next base classifier, and so on, until the last base classifier learning. Therefore, the neural network model that has high accuracy

of prediction, relatively volatile stability and poor interpretation is put in the first place, then the classification and regression tree model that has good interpretation but weak robustness is put in the second place, finally the Bayesian network model that has strong robustness, good interpretation but lower accuracy than neural network is put in the third place, thus, a prediction model with higher accuracy, better interpretation and stronger robustness than the single models is come into being to predict the metamorphism of perishable food through track monitoring. From the result of the example in this paper, we can conclude that the combination model which has been verified and assessed does achieve the intended purpose.

The complexity of the parallel structure and the hybrid structure not only makes it more difficult to build the model, but also reduce the interpretation of the model. Besides, on the improvement of the precision, both of them are not as good as the serial structure. Therefore, the parallel structure and the hybrid structure are not suitable for the establishment of the forecasting model of the metamorphism of perishable food through track monitoring.

### **3. Experimental Design of Data Mining for Metamorphic Monitoring of Perishable Food Transportation**

In recent years, food safety accidents continue to cause people to pay special attention to food safety, and pay more attention to low-temperature preservation of perishable food. Temperature sensitive foods are not only required to control temperature in production, storage, and marketing, but also in temperature control during transportation and distribution. If the transport and distribution process can't be effective in controlling the temperature caused by food quality problems, not only cause economic losses, and may threaten human life. Therefore, in order to meet people's requirements for food quality and safety, we can monitor the temperature and humidity and mechanical damage of perishable food in the process of transportation. In this paper, the data of the strawberry transportation and metamorphism monitoring are taken as an example to establish and evaluate the data. The excavation experiment design [1]-[15] will be described in detail in Sections 3.1, 3.2, 3.3 and 3.4.

#### **3.1. Task Description**

In this paper, the main task of data mining is to take strawberry transport metamorphism monitoring as an example, the strawberry transport metamorphism monitoring collected time, temperature, humidity and mechanical damage and other data as the sample data, using the collected affect deterioration information attributes as input, prediction results as output. Based on the model of combination of complementary, For the neural network algorithm classification and regression tree algorithm, Bayesian network algorithm model were modeled respectively by using of data mining software clementine12.0 in the training set and test set, then the three specific model series combination, and the combination model improved the prediction accuracy of the model, so as to obtain the

best prediction model of perishable food transport deterioration monitoring system.

### 3.2. Data Preparation

In this paper, three parameters of temperature, humidity and mechanical damage of the metamorphic factors of a strawberry transportation were used as modeling independent variables, and metamorphism was used as the modeling target variable. At the same time according to the model to establish the need to set the type of independent variable as gather, the target set to mark.

There are three steps for data preparation: the first step is data selection, the second step is data preprocessing, and the third step is data transformation.

1) First of all, the data selection, from 6:00 a.m. to 24:00 PM, measured every fifteen minutes, to use 1000 strawberries transport monitoring process of temperature, humidity and mechanical damage of 73 groups of data as experimental data, whether or not the number of spoiled strawberries as the target data, to simplify the data mining initial data was to achieve the data selection Final goal.

2) Secondly, the data preprocessing, the quality of data preprocessing directly affects the effect of data mining, so data preprocessing is an indispensable step in data mining. The content of preprocessing usually refers to the elimination of duplicate records and the completion of data type transformations (for example, the continuous data is discretized to facilitate symbolic induction, and the discrete type is transformed into a continuous value type to facilitate neural network induction) and so on. Because of the continuity of the field of decision tree is more difficult to predict, Bayesian algorithms are required to deal with the discrete field, consistent with requirements of each model data type combination model combination, so the pretreatment of the continuous field discretization. In this paper, we need to discretize the parameters, mainly for the temperature and humidity division of different intervals, different intervals corresponding to different discrete values. For example, the temperature from  $-0.7^{\circ}\text{C}$  to  $26^{\circ}\text{C}$  was divided into four stages respectively by 1 (low temperature:  $-0.9^{\circ}\text{C}$  -  $0.9^{\circ}\text{C}$ ), 2 (medium-low temperature:  $1^{\circ}\text{C}$  -  $10^{\circ}\text{C}$ ), 3 (medium-high temperature:  $11^{\circ}\text{C}$  -  $20^{\circ}\text{C}$ ), 4 (high temperature:  $21^{\circ}\text{C}$  -  $30^{\circ}\text{C}$ ) to express; humidity from 50% RH - 95% RH is divided into six stages with 1 (90% RH - 95% RH), 2 (76% RH - 89% RH), 3 (70% RH - 75% RH), 4 (65% RH - 69% RH), 5 (56% RH - 64% RH), 6 (50% RH - 55% RH), strawberry metamorphic was indicated by 1, strawberry was not deteriorated indicated by 0; mechanical damage was calculated according to the number of strawberries with mechanical damage.

In order to escape the incomplete parameter in the pretreatment, select the appropriate parameters according to the object of study. In this paper, metamorphic strawberry transport monitoring as an example, the fundamental reason is caused by the deterioration of the strawberry pathogen bacteria, microbial pathogens by temperature, humidity, illumination and mechanical damage caused by the change of generation, or from the field (infected leaves and soil) directly bring in fruits and vegetables. Storage in almost all the major influence

factors of fruits and vegetables are related to temperature, such as respiration, transpiration, ripening, ethylene production, growth, so the temperature is one of the most important key factors is the primary environmental conditions must be considered. But the water content is one of the basic characteristics of perishable foods, the perishable food in free water, bound water and hydrate, object water availability is called "water activity". Due to microbial and fungal reproduction will cause the deterioration of perishable food, so we must determine the microbial water activity factor therefore humidity must also be considered the perishable food deterioration factor. But mechanical damage is caused by the injury caused by pathogenic bacteria and perishable food epidermal microorganisms are also the main factors to be considered. Although the photosynthesis light play a key role in the process of cultivation of strawberry, but the transport storage is not necessary. So this thesis chooses the factors of temperature, humidity and mechanical damage as a parameter perishable food transport monitoring. Therefore, there would be no parameter incomplete, if the parameter redundancy preprocessing can transform the data dimensionality reduction to solve.

3) Finally, the data transformation. In order to make the transportation monitoring data dimension reduction, we need identify the really useful characteristics from transportation monitoring the initial characteristics, and reduce the number and characteristics of variables to be considered in data mining, filtering time and sequence number, only temperature and humidity as well as mechanical injury three parameters, and the discrete data corresponding to three parameter.

### 3.3. Data Segmentation

In the model development process, through the unused data to verify the model was a standard model of development, this criterion can validate the model robustness, namely: if not only in the modeling data set model runs well, and in a similar data set on the model of running the same good, explains the model was not only suitable for modeling data set, but for all the data sets are applicable. Before the model was built, The last part required work was data segmentation, That was the final sample split into training data for training the model set, and the test data for the model test in two parts. The data splitting node were set up in the model, then the data are randomly divided into two parts by the data splitting node , Some were used to build the model, and the other was used to evaluate the accuracy of the model. The use of data segmentation could greatly avoid the emergence of over-fitting, can guarantee the effectiveness and reliability of the model. In this paper, the sample data was divided into 70% proportion the training data and 30% proportion the test data.

### 3.4. Data Mining Model Building

In this paper, the data mining tool Clementine12.0 software successfully used

70% of the training data and 30% of the test data to divide the sample data, and model the Bayesian network. Bayesian network algorithms for different parameters, the classification regression tree and the neural network algorithm for different parameters were modeled respectively, the best prediction model of three kinds of models were selected to establish the series based on the three principles of model combination. The establishment of specific modeling and model evaluations would be covered in detail in Chapters 4 and 5.

## 4. Modeling Evaluation Standard

Data mining software Clementin 12.0 provides a combination of three universal evaluation methods [12] [13]: The first is to use the analysis node to examine the coverage, hit rate, the correct rate of prediction (numerical assessment), etc. The second is to use the assessment chart to assess the quality of the graphics (graphical assessment); the third is to use the verification data (data segmentation) test model is good or bad. Section 3.3 introduces the third assessment method, Section 4.1, Section 4.2 details the first two evaluation methods. In this paper, six universal indicators of numerical evaluation and graphical evaluation are used as the basis for evaluating the merits of the model.

### 4.1. Model Numerical Evaluation Standard

The numerical evaluation index includes three: the overall accuracy rate, precision rate (hit rate) and recall rate (coverage). The overall accuracy rate can be obtained directly from the correct rate shown in the partition table in the analysis report. The recall rate is based on the percentage of the coincidence matrix in the analysis report. In general, in predicting the final effect of the model, we must first ensure the overall accuracy of the model forecast, and then ensure that the model predicted hit rate, and then on this basis, try to improve the coverage of the model.

The overall accuracy rate: the correct rate of the entire sample.  $A$  represents a level (overall classification) Probability.  $P$  represents the positive and negative numbers of the correct forecasts. With  $T$  that the total number of samples, the overall correct rate is calculated as:

$$A = P \setminus T \quad (1)$$

Hits: Indicates the percentage of perishable food that is predisposed to perishable food by the model, which is an indicator of the accuracy of the model. The formula is:

$$V = p \setminus t \quad (2)$$

(2) where  $V$  is the hit rate,  $p$  indicates the correct number of positive forecasts, and  $t$  indicates the total number of positive forecasts.

Check the rate (coverage): that is accurately predicted by the model of food deterioration of the sample concentration of the actual percentage of deterioration. It uses an indicator to describe the accuracy. The formula is:

$$S = p \setminus a \quad (3)$$



(3) where  $S$  denotes the coverage rate,  $p$  indicates the positive number of instances to be correctly predicted,  $a$  represents the actual number of positive instances.

## 4.2. Model Evaluation Chart Evaluation Standard

The higher left side of the cumulative chart represents a better model, rather than the cumulative extension of the left side of the graph, and the lower part of the right side shows a better model. Profit charts, response charts, performance charts, profit charts, and ROI charts are the most common assessment of universality. This paper selects the income table, response graph, and performance chart as the evaluation criteria based on the selected three models.

Gains charts: A good or bad judgment of a model's revenue graphs is to see if the lines in the table are steeply rising to 100% and gradually become gentle. If yes, the model is a good model. If not, then the model is not a good model. If a line in the model's revenue graph is raised from the left end to the right end with a diagonal diagonal shape to a higher position, the model does not provide any information. With  $R$  that the gain,  $m$  quantile number of matching,  $M$  said the total number of matching, the gain is calculated as:

$$R = (m/M) \times 100\% \quad (4)$$

Response charts: The lines in the response graph just start from 100% or more, and the model is a good model. The entire image of the curve has been around the response rate around the description model without providing any information. The response graph is calculated as:

$$N = (s/r) \times 100\% \quad (5)$$

(5) where  $N$  is the response,  $s$  is the number of successes in the quantile, and  $r$  is the number of records in the quantile.

Lift charts: The performance model is a good model that is characterized by a high degree of stability from the beginning of the function graph just above 1.0 and then to the right. The entire image of the curve has been around 1.0 indicating that the model did not provide any information. It is calculated as:

$$E = (s \setminus r) \setminus (T \setminus C) \quad (6)$$

(6) where  $E$  is the function,  $s$  is the number of successes in the quantile,  $r$  is the number of records in the quantile,  $T$  is the total number of successes, and  $C$  is the total number of records.

## 5. The Building of Network Model Based on Perishable Food Transportation Monitoring

### 5.1. Evaluation of Neural Network Model Based on Strawberry Transport Deterioration Monitoring

The mathematical model of a classification algorithm that simulates the behavior of human or animal neural network behavior and distributed parallel informa-



tion processing is called the artificial neural network (Artificial Neural Networks, ANN) [24] [25] [26] [27]. It can prestore and categorize data by establishing different types of neural networks. Compared to other classification algorithms, Not only does it have a high level of accuracy but also the ability to handle parallel distribution is very strong. Moreover, the accuracy is high and the ability to handle parallel distribution is strong, too, and it has strong robustness and fault tolerance for noisy data. During neural network training, the initial value is more sensitive and should be tested several times. Because the convergence is very different from the initial value, we used the comparison method to obtain the value. At the time of training, multiple initio weights are generated at random, and the best one is chosen as the initial weight [25].

1) Training data the numerical evaluation of the neural network model of different parameters was evaluated by 70% proportion

The Numerical evaluation of the neural network of the two parameters is calculated by the calculation formula of the overall accuracy, percentage and coverage. The overall accuracy and hit ratio and coverage values are shown below:

From **Table 1**, for neural network we can know that the overall accuracy of Quick parameters and the overall accuracy of the RBFN parameters calculated on the test set were both 90.00%. However, the hit ratio of the Quick parameter was higher than the hit ratio of the RBFN parameter, so it is better to predict the neural network model of the Quick parameter.

2) The Quick neural network and the RBFN neural network evaluation chart are shown in the figure below:

On the test set, the analysis of **Figure 1** and **Figure 2** shows that the results of Quick neural network response and lift graph obtained were similar to the results of RBFN neural network modeling.

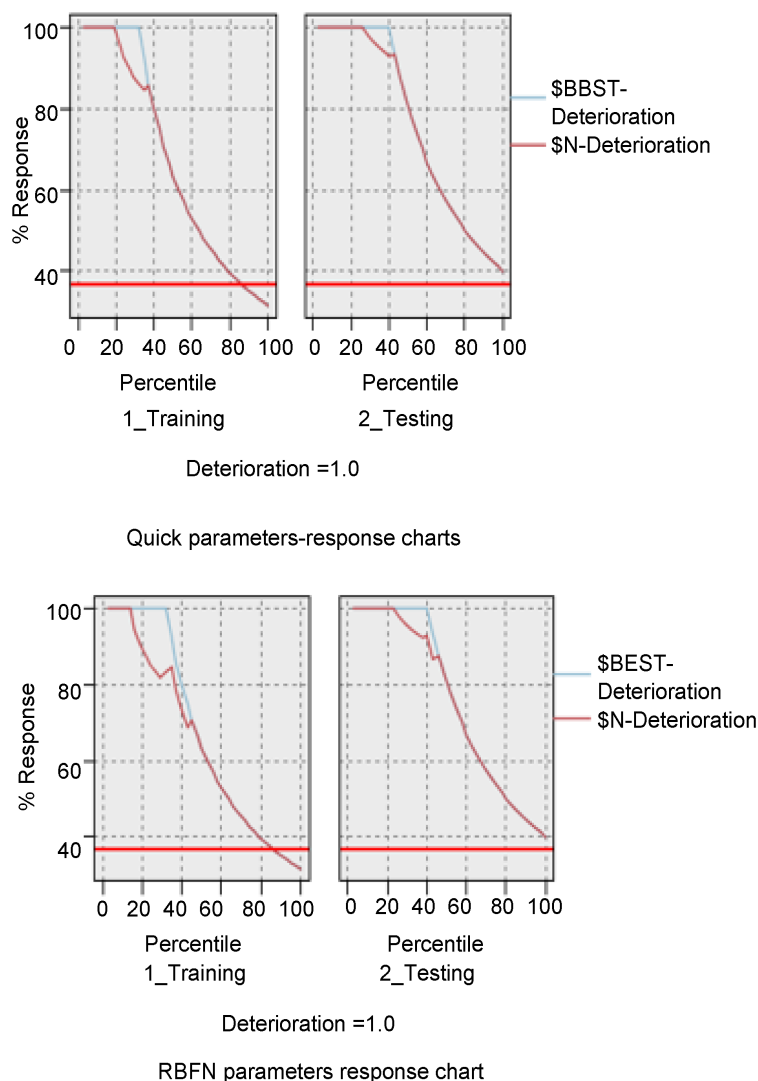
By the results of the Numerical evaluation of two parameters of Bias network 1) and Graphical evaluation of two parameters for Bias network 2) were analyzed and compared, The Quick neural network achieved better predictions on the test set than RBFN neural network modeling.

## 5.2. A regression Tree Model Was Evaluated Based on the Analysis of the Metamorphic Classification of Strawberry Transport

Classification regression tree technique [27] [28] is a data set classification decision tree technique. It can automatically detect the underlying structure, important patterns and relationships of highly complex data, and the underlying

**Table 1.** Two kinds of neural network parameters comparison numerical evaluation table.

	The neural network of 70% Quick parameter	The neural network of 70% RBFN parameter
The overall accuracy	90.00%	90.00%
hit ratio	90.00%	83.30%
coverage	90.00%	100.00%



**Figure 1.** Comparison of response graphs of two parameters in neural network.

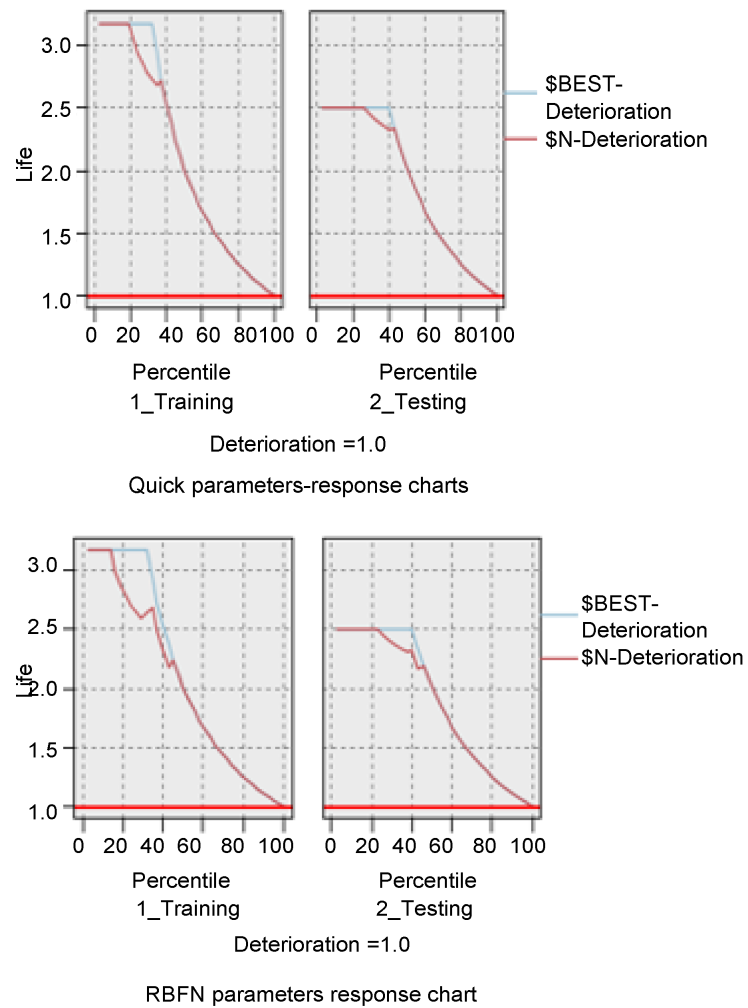
structures, important patterns and relationships of the highly complex data that are automatically detected can construct accurate and reliable forecasting models.

The classification regression tree model is evaluated based on the calculation formula of the overall accuracy, percentage and coverage. The overall accuracy and hit ratio and coverage values of the Classification regression tree model are shown below.

It is shown in **Table 2** that the overall accuracy of Classification regression tree was 85.00%, Hit ratio was 88.89%, and the coverage rate was 80.00%.

### 5.3. Strawberry Transportation Deteriorated Monitoring Bayesian Network Model Assessment

The advantage of Bayesian network [29] [30] is that it provides a natural and intuitive way to solve the problem of uncertainty. The Bayesian network (One of



**Figure 2.** Comparison of lift graphs of two parameters in neural network.

**Table 2.** 70% the proportion of the training data segmentation classification and regression tree numerical evaluation.

70% training data segmentation and classification and regression tree CART	
The overall accuracy	85.00%
Hit ratio	88.89%
Coverage	80.00%

the most effective theoretical models of reasoning and uncertainty) represents a non-loop diagram of the dependencies between variables. It is made up of nodes and the arc of the connecting node. Because each node corresponds to a random variable, each node also corresponds to a conditional probability table. Qualitative representations of dependencies between nodes and the dependent relationships of nodes can be quantified by conditional probability tables. Bayesian network theory is based on a solid foundation; the knowledge structure is expressed in the way of natural; the reasoning ability is strong, in simple and intuitive way to explain the actual problem.

### 1) Numerical evaluation of two parameters of Bias network

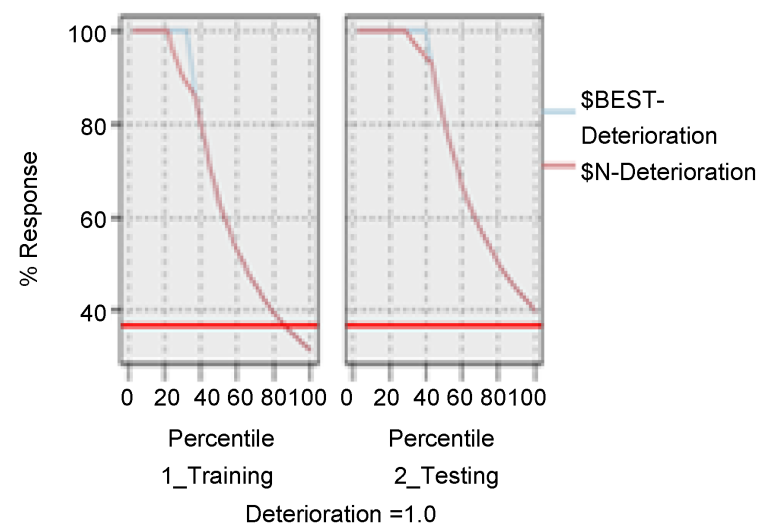
Based on the calculation formula of the overall accuracy, Hit ratio and coverage, TAN Bayes and Markov's Blanket Bayes on the test set model was evaluated in **Table 3**.

The data comparison analysis of **Table 3** shows the conclusion that Markov's Blanket Bayesian network was more accurate than the overall accuracy of the TAN Bayesian network, with better prediction results.

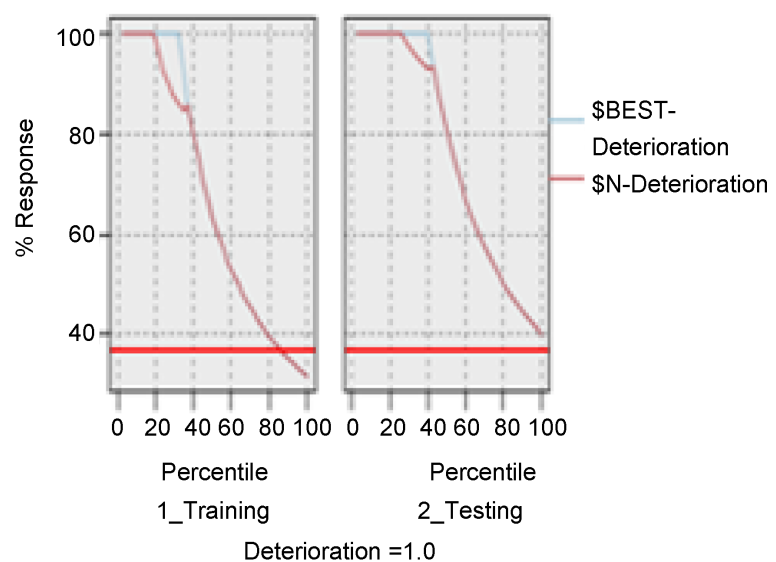
### 2) Graphical evaluation of two parameters for Bias network

The evaluation diagram was shown in the figure below:

On the test set, the analysis of **Figure 3** and **Figure 4** shows that the results of TAN Bayesian network response and lift graph obtained were similar to the results of the Markov Blanket Bayesian network.

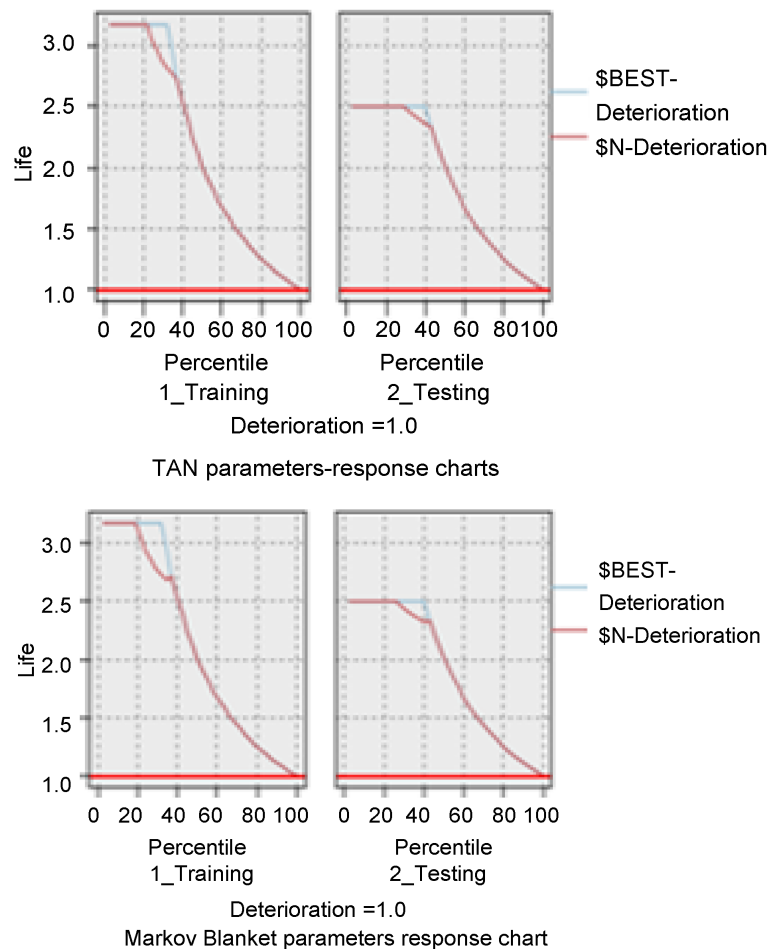


TAN parameters-response charts



Markov Blanket parameters response chart

**Figure 3.** Comparison of response graphs of two parameters in Bias network.



**Figure 4.** Comparison of lift graphs of two parameters in Bias network.

**Table 3.** 70% the proportion of the training data segmentation Bayesian network two parameter evaluation table.

	TAN Bayesian network	Markov Bayesian network Blanket
The overall accuracy	84.00%	86.00%
Hit ratio	90.48%	87.59%
Coverage	82.61%	91.30%

By the results of the Numerical evaluation of two parameters of Bias network 1) and Graphical evaluation of two parameters for Bias network were analyzed and compared, the prediction of Markov Blanket Bayesian network modeling was much better on the test set than Markov Blanket Bayesian network modeling.

**5.4. Based on the Analysis of the Tandem Optimal Combination Model Evaluation of Strawberry Transport Metamorphic Monitoring, the Comparison Was Made**

From the Section 5.1 neural network Numerical evaluation and Assessment evaluation map of results and in Section 5.2 of the classification and regression

tree CART Numerical evaluation and Assessment evaluation map of results and in section 5.3 of the Bayesian network numerical evaluation and Assessment evaluation map of results to get the optimal combination of serial, Based on the 70% proportional training data and 30% proportional testing data were used to segment sample data, respectively established Quick neural network model, classification and regression tree CART model and Markov Blanket Bayesian network model combined in series in order to improve the accuracy of the models.

#### 1) Numerical evaluation

Based on the numerical evaluation calculation formula, the results of the overall accuracy and hit ratio and the coverage of the combination optimization model were calculated (**Table 4**).

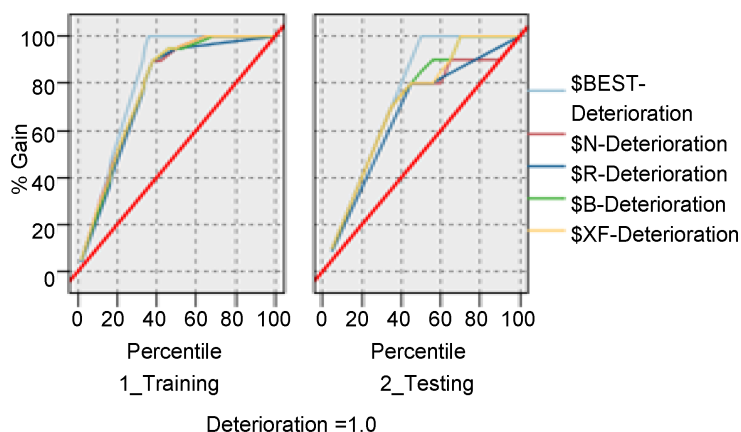
**Table 4.** Comparison of numerical evaluation before and after model combination.

	Combination optimization model	Quick neural network model	Classification and regression tree model	Markov Bayesian network Blanket
The overall accuracy	95.00%	90.00%	85.00%	85.00%
Hit ratio	88.89%	90.00%	88.89%	87.59%
Coverage	100.0%	90.00%	80.00%	91.30%

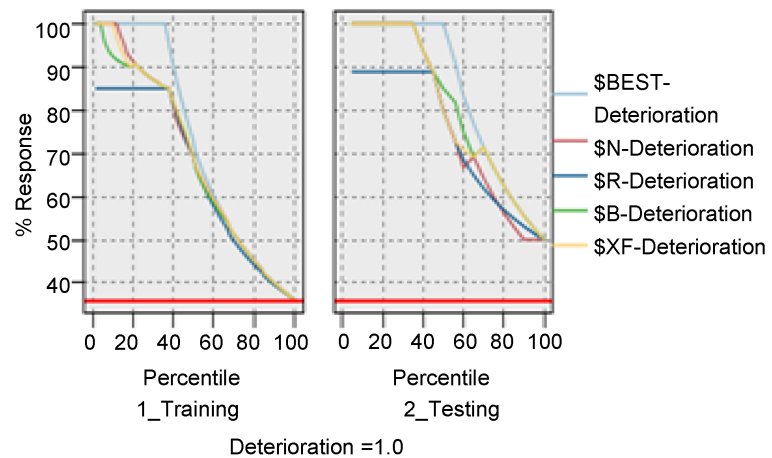
We can know from the table above that on the test set for the combined model results for the overall accuracy of 95.00%, 88.89% hit ratio, 100.00% coverage. Therefore, the prediction results of the combined model are better than that of the single model.

#### 2) Assessment evaluation map

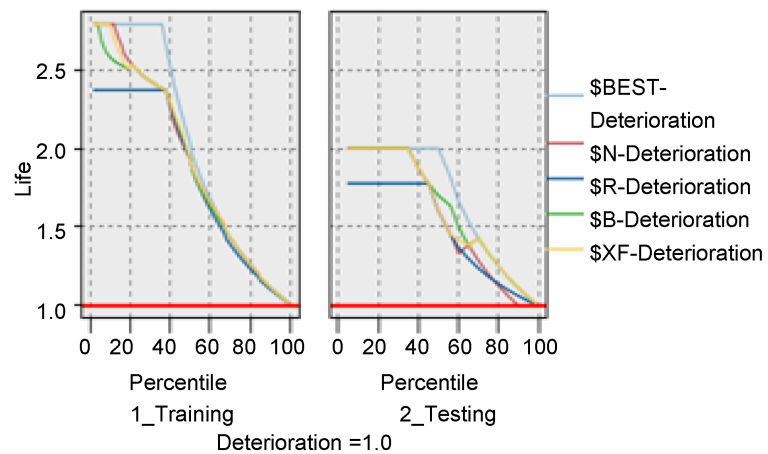
From **Figure 5**, **Figure 6** and **Figure 7**, we can know that ginger yellow line represents the gain, response, and efficacy combination model evaluation line than neural network, Bayes network and classification and regression tree represents a single model evaluation line closer to the light blue best line. Therefore, prediction model of the serial combination is better than the single prediction model.



**Figure 5.** Comparison of Gain graph of three model combination optimization.



**Figure 6.** Comparison of Response graph of three model combination optimization.



**Figure 7.** Comparison of lift graph of three model combination optimization.

By the results of the Numerical evaluation of two parameters of Bias network 1) and Graphical evaluation of two parameters for Bias network were analyzed and compared, the prediction of combination forecasting model was much better on the test set than a single model prediction results. And the accuracy of model prediction was enhanced by using the advantages of individual models.

## 6. Conclusion

The data mining system is widely developed and applied in business, economy, finance and management, but for perishable food transportation metamorphic monitoring field that does not see more, it is used to establish perishable food metamorphism prediction model for the transportation destination perishable food fast detection to extract the effective and typical samples for guiding the rarer. There are two points of innovation in this article: First, the data mining technique is applied to the prediction model in perishable monitoring of perishable food; Second, use the model of combination of the three principles, classifi-



cation, regression tree and select neural network Bayesian network model of three complementary combinations to create a relatively high precision, strong stability, good explanatory model to predict perishable food transportation monitoring metamorphism. Through the combination forecast model of transportation destination perishable food fast detection to extract the effective work instruction, the typical test sample effectively improves the efficiency and effect of sampling inspection work, to avoid bad perishable food into the market and ensure the transportation safety of perishable food quality.

This paper, by using data mining technology for perishable food transportation metamorphic monitoring makes a preliminary exploration, and obtained some achievements, but there are still some limitations in theory and operation practice, and further studies are needed. In this paper, only 1000 strawberries were selected for the study, and the number of samples required for the data mining was still to be increased.

## Acknowledgements

This work was supported by funding project for Youth Talent Cultivation Plan of Beijing City University Under the grant number (CIT&TCD201504051), and Beijing outstanding talent training project (2014000020124G093) and Beijing Intelligent Logistics System Collaborative Innovation Centre.

## References

- [1] Xue, W. and Chen, H.G. (2014) Data Mining Method and Application of SPSS Modeler. Electronics Industry Press, Beijing, 1-327.
- [2] Xiong, P. (2011) Data Mining Algorithms and Clementine Practices. Tsinghua University Press, Beijing, 1-200.
- [3] He, J. (2016) The Development Trend of Data Mining Technology. *Western Leather*, No. 24, 21.
- [4] Wang, G., Hao, J.X., Ma, J., *et al.* (2011) A Comparative Assessment of Ensemble Learning for Credit Scoring. *Expert Systems with Applications*, **38**, 223-230.  
<https://doi.org/10.1016/j.eswa.2010.06.048>
- [5] Hsieh, N.C. and Hung, L.P. (2010) A Data Driven Ensemble Classifier for Credit Scoring Analysis. *Expert Systems with Applications*, **37**, 534-545.  
<https://doi.org/10.1016/j.eswa.2009.05.059>
- [6] Magpiesensing.com. "Sensing Cold Chain Analytics and Monitoring."  
<http://technical.ly/baltimore/wp-content/uploads/sites/3/2012/10/ColdChainAnalyticsMagpieSensing-Whitepaper-Final.pdf>
- [7] (2013) Centers for Disease Control and Prevention, Vaccine Storage and Handling.  
[www.cdc.gov/vaccines/pubs/pinkbook/vac-storage.html#storage](http://www.cdc.gov/vaccines/pubs/pinkbook/vac-storage.html#storage)
- [8] Zaleski, A. (2013) Magpie Analytic System Tracks Cold-Chain Products to Keep Vaccines, Reagents Fresh.  
<http://technical.ly/baltimore/2012/10/04/magpie-analytics-system-tracks-cold-chain-products-to-keep-vaccines-reagents-fresh/>
- [9] Zhu, J., Zhang, P. and Gao, Z.G. (2014) Application of Large Data Mining Technology to Early Warning of Agricultural Food Safety Risk of Imported and Exported Foods. *Agricultural Economy*, No. 3, 99-102.

- [10] Gao, X.F. and Yang, D.J. (2014) A Comparative Study of Food Safety Risk Warning at Home and Abroad. *Chinese Health Engineering*, **13**, 254-256.
- [11] Xue, S.J. and Chen, F. (2011) Research on Data Mining Technology in Telecom Customer Churn Warning System. *Microcomputer and Application*, **30**, 49-52.
- [12] Du, S.J. (2013) Evaluation of Data Mining Model Based on Clementine. *Software and Application of Computer*, **8**, 263-265.
- [13] Wu, Z.Y. (2006) Application of Data Mining in Customer Churn. University of Electronic Science and Technology of China, Chengdu, 63-64.
- [14] Ren, X.D. and Chen, J.Y. (2016) Research on Current Situation and Development Trend of Data Mining. *Information Communication*, **2**, 171-172.
- [15] Huo, K. (2015) Application of Data Mining Technology in Bank in the Era of Big Data. *Merchant*, **26**, 191-192.
- [16] Yang, W.Y., Shun, Y.M., Zhou, Y. and Wu, M.F. (2012) Application of Rapid Detection Technology in the Safety Control of Fruits and Vegetables. *Journal of Anhui Agricultural Sciences*, **40**, 147-149.
- [17] Li, X.Y. (2014) Research and Discussion on Rapid Detection Technology of Food Safety. *Food Research and Development*, **9**, 120.
- [18] Gao, G. (2015) How to Improve the Overall Effect of Food Safety Testing. *Changchun Shuangyang District Food and Drug Administration*, No. 27, 119.
- [19] Ma, X.N. and Du, M.H. (2015) Development and Prospect of Rapid Detection Instruments in the Field of Food Safety. *Journal of Food Safety and Quality Control*, No. 5, 1828-1833.
- [20] Yang, C.G., Wang, H.W. and Peng, X.T., etc. (2015) Advances in Rapid Detection of Food Pathogenic Microorganisms. *Journal of Food Safety and Quality Control*, No. 1, 43.
- [21] Wang, Y. (2013) Prediction of Heat Supply Network Based on Data Mining Optimized Combination Model. Shandong Petroleum University, Dongying, 56-59.
- [22] Xiang, H. (2011) Research and Application of Personal Credit Scoring Combination Model. Hunan University, Changsha, 84-89.
- [23] He, Y. (2014) Research on Stock Price Forecasting Based on Data Mining Combination Model. Southwestern University of Finance and Economics, Chongqing, 19-47.
- [24] Zhang, D.B., Xu, J.P., Xu, J.J., *et al.* (2010) Food Safety Early Warning Model Based on Monitoring Data and BP Neural Network. *Journal of Agricultural Engineering*, **1**, 221-226.
- [25] Feng, S.H. and Zh, P.Q.F. (2006) A New Method for Determining Initial Weights of Neural Networks. *Industrial Instrumentation and Automation*, No. 1, 65-68.
- [26] Morton, H. (2011) Municipal Credit Rating Modelling by Neural Networks. *Decision Support Systems*, **51**, 108-118.
- [27] Xu, Y.Y. (2013) Analysis and Discussion of Retail Merchandise Promotion Data Based on Classification. *Regression Tree and Neural Network Model*, **15**, 22-24.
- [28] Wang, W., Xue, F.C., Sh, D.W. and Wang, X.J. (2016) Study and Application of Summer Drought Prediction Model Based on CART Algorithm. *Meteorological Science*, **15**, 22-24.
- [29] Liu, Y.C., Cheng, S.W., Feng, J.D., Xu, X.J., Zhu, T.L. and Fang, Y.S. (2012) Traceability and Early Warning of Vegetable Quality and Safety Based on Bayesian Network. *Meteorological Science*, No. 20, 188-190.
- [30] Masegosa, A.R. and Moral, S. (2013) New Skeleton-Based Approaches for Bayesian

Structure Learning of Bayesian Net-Works. *Applied Soft Computing*, **13**, 1110-1120.



Scientific Research Publishing

**Submit or recommend next manuscript to SCIRP and we will provide best service for you:**

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact [fns@scirp.org](mailto:fns@scirp.org)

