

A New Sales Forecasting Method for Industrial Supply Chain

Mahmoud Zadeh

University of Newcastle, Newcastle Australia
mahmoud.zadeh@newcastle.edu.au

Abstract

With the continuous innovation of computer technology, it can solve the problems of low accuracy, non-intelligence, and inability to process complex samples in the sales forecasting methods of industrial supply chains. This paper proposes a sales forecasting method for the industrial supply chain based on the Gaussian mixture model. By analyzing the characteristics of the original sales data of the industrial supply chain, the eigenvalue correlation ranking vector is generated. Then predict the parameters such as the number of clusters in the Gaussian mixture model. By comparing the accuracy of the prediction results, the recall rate and the F-value, the eigenvalues, and the number of clusters that can achieve better prediction results are determined. This paper compares the Gaussian mixture model with the artificial neural network model and the convolutional neural network model on the original sales data set of the same industrial supply chain. The experimental results show that, compared with the artificial neural network model and the convolutional neural network model, the method has better performance in all three indicators, and can better predict sales transactions.

Keywords: Sales forecast, Gaussian mixture model, Artificial neural network, Industrial supply chain

1. Introduction

With the continuous innovation of computer technology, product sales-related data in the industrial supply chain can be collected and stored, and it is growing at an unprecedented rate. Real-time analysis of big data promotes the continuous in-depth research and discussion of more effective analysis tools and methods in academia and industry. The data structure and artificial intelligence analysis system for big data have attracted more and more attention from researchers. Among them, the machine learning theory extended by artificial intelligence is a method of finding rules from samples and predicting the output of future samples or unobservable data. A theoretical method that transforms unstructured data or semi-structured data into structured data and performs modeling analysis, which has become a current research hotspot.

Most of the transaction forecasting methods in the existing industrial supply chain sales system use qualitative methods to forecast sales. Most of these methods are theoretically and weakly maneuverable, and the forecast results often fail to meet the psychological expectations of customers and enterprises. Take the mechanical and electrical product sales system as an example. Its sales data often have features such as large data volume, large feature values, and irregular data. It is impossible to know which feature values have greater reference values for the prediction results. In this regard, traditional methods use qualitative sales forecast analysis,

Article history:

Received (April 5, 2021), Review Result (June 1, 2021), Accepted (August 30, 2021)

which not only consumes human, financial, and material resources, but also has difficulty in taking into account factors such as the market and different user groups, which leads to large errors in forecast results.

Therefore, this paper proposes an industrial supply chain sales forecasting method based on Gaussian Mixture Model (GMM). By processing historical business data in the sales system, analyzing customer consumption habits and consumption capabilities, on this basis, GMM is used to predict and train the sales data set to evaluate the performance of the model under different feature values and cluster numbers. To generate a more mature model, and finally, realize the sales forecast.

2. Theoretical research

Moon et al [1] pointed out that sales forecasting is a key link in enterprise supply chain management. Through sales forecasting, enterprises can formulate scientific and reasonable raw material procurement plans, production plans, staffing plans, inventory plans, and marketing plans. Therefore, the sales forecasting decision support system has important research significance for the business decision-making of enterprises. Sales data is dynamic, non-linear, and irregular time series data, which is affected by various factors such as seasonal weather, emergencies, dealers' sales capabilities, and the number of lower-level dealers. In the early days, sales forecasts were made artificially by sales staff based on product market and sales experience. This method has high labor costs and poor field portability. Later, with the rapid development of big data and data processing technology, more and more scholars at home and abroad have carried out more in-depth research on sales forecasting methods in various industries.

Among them, the more widely used methods mainly include the neural network forecasting method and clustering forecasting method. Thomassey et al [2], Silva et al [3], and Loureiro [4] proposed a neural network-based sales forecast model, which provides a basis for companies to formulate sales strategies. Kindelan et al [5] and Hainaut et al [6] improved the sales forecasting method based on neural networks. Among them: Kindelan et al [5] optimizes the BP neural network based on a genetic algorithm, and uses the real sales data of a convenience store for a period to experiment. The results show that this method can better solve the BP neural network's shortcomings and prediction results that are easy to fall into local optimality. The problem of low accuracy has improved the accuracy of sales forecast to a certain extent; Hainaut et al [6] proposed an algorithm for optimizing the weights and thresholds of BP neural network by mixing adaptive harmony algorithm and genetic algorithm. Experiments show that the algorithm solves the problems of neural networks easily falling into local optimal values and slow convergence speed to a certain extent.

Although scholars at home and abroad have been committed to solving the local optimal value and convergence problems of neural networks and have made certain progress, they still cannot complete sales forecasts with high performance when faced with huge sales data. Therefore, more and more scholars have begun to do further research on sales forecasting from the perspective of clustering methods. Iotti et al [7] proposed a product reclassification prediction model based on sales data. The model uses the K-means clustering method to cluster products based on the common performance of sales characteristics. On this basis, time series forecasting methods and hidden Markov forecasting methods are used to predict product sales qualitatively and quantitatively respectively. Taking the actual sales data of a certain category of an e-commerce website as the experimental data, the results show that the prediction model has realized the sales prediction function of e-commerce products to a certain extent.

Kurematsu et al [8] from the perspective of the managers of apparel enterprises, proposed a sales forecasting model combining cluster analysis and CAST decision tree algorithm to further explore the factors affecting apparel sales, and predict sales on this basis. Empirical analysis shows that this method is effective in predicting turnover to a certain extent. K-means clustering algorithm [9] is not only used in sales forecasting but also similar recommendation system problems. Li et al [10] proposed an e-commerce recommendation method based on the clustering of user-item preference values. This method is used to predict the target user's interest in unrated products. The open-source data set MovieLens is used as the experimental data. It shows that this method can better solve the problem of the performance degradation of the recommendation system due to the increase of userspace.

With the in-depth study of clustering algorithms by scholars at home and abroad, GMM, as an algorithm that uses probability to describe the classification of data points, has gradually been applied to text clustering, prediction problems, recommendation systems, etc. [11]. Steiner et al [12] clustered the music feature texts with high similarity based on GMM to realize the personalized service of music recommendation. Yao et al [13] proposed a GMM-based text clustering method to achieve the normalization of the brand entities of e-commerce products. An et al [14] predicts user interests in social networks based on the GMM method and uses real experimental data sets for experimental verification. The results show that GMM has good prediction accuracy. Piao et al [15] improved the collaborative filtering algorithm used in traditional recommendation systems based on GMM. And use the public data set MovieLens to verify the algorithm. The results show that the algorithm has stronger prediction accuracy. Yang et al [16] proposed a knowledge recommendation system based on GMM to predict users' ratings of knowledge items, and then to achieve personalized knowledge recommendations for users. Experiments were carried out with the open-source data sets fisheries and MovieLens, and the results showed that the prediction accuracy of this method was better. high.

At present, there are few studies on the application of GMM to sales price forecasting, but many related studies have shown that GMM has good performance in dealing with forecasting problems. Therefore, it is aimed at electromechanical products with large data volume, complex data attribute values, and a large amount of information. For sales data, this article chooses GMM as the processing method and uses artificial neural networks and convolutional neural networks as a comparison.

3. Research method design

3.1. User data acquisition and processing

This article uses the real sales data of a mechanical and electrical product sales company as the data source. After desensitization, 418,282 pieces of original data are retained as the standard data set for the experiment.

The data features included in the original data set of this article are id (a transaction record number), RFQ (enterprise and customer agreement record number), ACCT (customer number), Coverage, SKU, SKU_Category, EB_Flag, RFQ_TYPE (agreement between enterprise and customer) Product type), List_Price (the price of the product), RFQ_Price (the price of the product negotiated by the company and the customer), RFQ_Qty (the number of the product negotiated by the company and the customer), Order_Qty (the number of products the customer finally ordered). This article first calculates the mean, maximum, minimum, median, etc. of each feature of the data set, and understands the numerical distribution of different features of

the data. Then calculate the correlation between the various features of the data, compare the importance of the features, and provide a basis for the follow-up work.

Data cleaning rules mainly include filling default values, smoothing noisy data, and identifying outliers. Default values, noise data, and outliers are common in real data sets. These abnormal data can be generated for various reasons such as sales system failures, data recording errors, and abnormal database data updates. Therefore, you first need to visualize the data to check whether there is noisy data, and then further check whether there are default values or outliers in the obtained data.

In this paper, the data is normalized, that is, the data is processed into floating-point data from 0 to 1, which represents the weight ratio of the feature in all features. The specific conversion process is as follows: the id contained in the data file is used to mark each record, and can be processed as a label as a sequential natural number gradually increasing from 1; RFQ, ACCT, coverage, SKU, sku_category, EB_flag, RFQ_TYPE can be processed as in this The frequency in all records of the feature; list_price, rfq_price, rfq_qty, order_qty can be normalized according to the formula $(x-x_{\min})/(x_{\max}-x_{\min})$.

Based on the above analysis and processing of the data set, to prevent over-fitting, this paper uses cross-validation during the experiment to form the required training set and test set.

3.2. Model construction

This article uses GMM to describe the probability of different characteristics affecting the success of a transaction. The likelihood function is expressed as the sum of the Gaussian models and the weights, as shown in equation (1). deal price:

$$p(x) = \sum_{k=1}^K \pi_k N(x | \mu_k, \sigma_k) \quad (1)$$

Where: $N(x | \mu_k, \sigma_k)$ is the k-th model (component) in the Gaussian mixture model; μ_k and σ_k are the k-th Gaussian model mean and variance; π_k is the mixing coefficient, that is, the weight. π_k needs to satisfy:

$$\sum_{k=1}^K \pi_k = 1 \quad 0 \leq \pi_k \leq 1 \quad (2)$$

In the process of using this likelihood function in this paper, the input is any row of the user \times (whether the feature is traded) matrix, and the output is the probability matrix of the user belonging to each cluster.

3.3. Training the model

This article uses the constructed model to train the data and then obtains the value of each parameter in the likelihood function. The training process includes two parts: the determination of the optimal component quantity and the parameter estimation.

In the parameter estimation problem, it is necessary to find an optimal balance between the model complexity and the model's ability to describe the data set. For GMM, the core of the problem lies in determining the optimal number of components. This paper compares the performance indicators of the model under different combinations of the number of clusters and the number of features to determine the optimal number of clusters. During the experiment, cross-checks are performed ten times to prevent overfitting.

The parameter estimation process in the likelihood function of the Gaussian mixture model uses the EM algorithm [17]. The basic idea is: randomly initialize a set of parameters $\Phi(0)$, update the expected $E(Y)$ of Y according to the posterior probability $P(Y | X; \Phi(0))$, and then use $E(Y)$ instead of Y to find the new model parameter $\Phi(1)$, so iterate until Φ stabilizes.

First of all, this paper establishes the maximum likelihood function based on formula (1):

$$L_{\theta} = \sum_{i=1}^N \log \left\{ \sum_{k=1}^K \alpha_k N(x_i | \mu_k, \sigma_k) \right\} \quad (3)$$

In the likelihood function of the Gaussian mixture model, the probability of a single point is very small, and the data will be smaller after the continuous multiplication operation, which will easily cause the floating-point number to underflow, which will affect the subsequent calculations. Therefore, the EM algorithm performs a logarithmic operation on the likelihood function when calculating, and realizes the likelihood function shown in formula (3).

The EM algorithm determines the parameters including two parts: E step (Expectation) and M step (Maximization).

(1) Step E, that is, calculate the probability of each data (ie each record) produced by every single Gaussian model. For each data x_i , the probability that it is generated by the k th component is:

$$\gamma(i, k) = \frac{\pi_k N(x_i | \mu_k, \sigma_k)}{\sum_{j=1}^K \pi_j N(x_i | \mu_j, \sigma_j)} \quad (4)$$

Since μ_k and σ_k in equation (4) are also values that need to be estimated, it is necessary to assign initial values to μ_k and σ_k when using the iterative method, and in the process of iteration, the value obtained in the previous iteration is used as the calculated value.

(2) Step M, that is, estimate the parameters in Step E, and then obtain the maximum likelihood function. Through the E step, the probability of each data x_i generated by the k th component can be obtained, and then it needs to be specific to each component to solve the parameters. Since each component is a standard Gaussian distribution, it is easy to find the parameter values corresponding to the maximum likelihood function:

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^N \gamma(i, k) x_i \quad (5)$$

$$\sigma_k^2 = \frac{1}{N_{ki}} \sum_{i=1}^N \gamma(i, k) (x_i - \mu_k)(x_i - \mu_k)^T \quad (6)$$

$$N_k = \sum_{i=1}^N \gamma(i, k) \quad (7)$$

$$\pi_k = \frac{N_k}{N} \quad (8)$$

Repeat steps E and M until the values converge, and substitute the final parameters into the objective function to complete the model.

In this paper, the model trained by the EM algorithm is used to predict the data in the test set and test the accuracy of the model's prediction performance.

4. Experimental analysis

4.1. Data preprocessing

The performance evaluation indicators used in this article include accuracy, recall, and F-value. The data set used is the real sales record data of a Canadian mechanical and electrical product sales company. The data features are not clear and cannot be directly applied to the model for training. Further analysis is required. Therefore, this article first analyzes the data itself and calculates the mean-variance under different eigenvalues. The results are shown in [Table 1]. Then on this basis, the original data set is cleaned to eliminate noise data. Then further analyze the data characteristics and analyze the correlation between the data characteristics. The results are shown in [Table 2]. Finally, the preprocessing of the data set is realized. This paper analyzes the results in [Table 1], and it can be seen that: (1) The data type and value upper and lower bounds of each feature are quite different. To facilitate the subsequent use of the model for training and prediction, the data preprocessing process needs to be normalized. (2) The data set contains noisy data. For example, if the minimum value in Order_Qty is a negative number, outlier detection and replacement are required in the later data preprocessing process. (3) Combining the results in [Table 1] and [Table 2], part of the noise data is displayed visually, so the data needs to be further cleaned, detected, and removed in the future. (4) The data label of the data set is Order_Qty. Analysis of the results shows that the number of samples with 0 trading volume is much greater than the number of samples with non-zero trading volume. When dividing the training set and the test set in the later experiment, you need to consider the transaction samples and non-transaction samples.

Table 1. Analysis of experimental data set

Data characteristics	Quantity	Mean	Minimum	Lower quartile	Median	Upper quartile	Max
RFQ	4.182820e+05	1.204412e+10	1.204209e+10	1.204303e+10	1.204397e+10	1.204525e+10	1.204642e+10
ACCT	4.182820e+05	1.108886e+10	1.102465e+10	1.108173e+10	1.109369e+10	1.110756e+10	1.111017e+10
Coverage	4.182820e+05	1.13012	1.00000	1.00000	1.00000	1.00000	2.00000
SKU	4.182820e+05	1.006754e+07	1.000000e+07	1.003717e+07	1.006880e+07	1.010248e+07	1.012984e+07
SKU_Category	4.182820e+05	17721.6556	9211.0000	10369.0000	12026.0000	24297.0000	36062.0000
RFQ_TYPE	4.182820e+05	1.53980	0.00000	1.00000	1.00000	1.00000	9.00000
List_Price	4.182820e+05	294.98828	0.00000	17.25080	47.57100	170.25570	302136.4500
RFQ_Price	4.182820e+05	187.27420	0.05050	8.87790	25.72470	97.88920	241709.1600
RFQ_Qty	4.182820e+05	27.92749	1.00000	1.00000	1.00000	5.00000	244900.0000
Order_Qty	4.182820e+05	5.75621	-49.00000	0.00000	0.00000	0.00000	4259.00000

Table 2. Correlation analysis among various characteristics

Feature name	Correlation value
RFQ	0.58032
ACCT	1.00000
Coverage	0.00000
SKU	0.00000
SKU_Category	0.52708
EB-Flag	-0.13636
RFQ_TYPE	1.00000
List_Price	0.89733
RFQ_Price	0.92787
RFQ_Qty	0.23979
Order_Qty	-0.02594

This paper uses the decision tree regression model to calculate the feature correlation, that is, remove the features one by one, use the remaining features to predict the removed features, and achieve the calculation of the correlation between the features by scoring the prediction results. The order of the degree of relevance is: features with a negative value > features with a value between 0 and 1 > features with a value of 1. Among them, the correlation of features between 0 and 1 will increase as the value approaches 0.

4.2. Experimental results and performance evaluation

This paper evaluates the performance of the method through accuracy, recall, and F-value, and further determines the number of features of the data set and the number of clusters to be used in this paper by calculating and comparing the values of the three indicators under different cluster numbers and feature number combinations.

[Table 3] shows the values of the three indicators under different feature numbers. Through analysis, it is found that when the feature value is 7, the accuracy, recall, and F-value have the largest values, that is, the method performance is the best. In particular, the three index values under different eigenvalues are obtained based on the maximum value of the number of clusters from 2 to 10.

Table 3. Accuracy, recall, and F value under different characteristic values

Eigenvalues	Accuracy	Recall rate	F value
7	0.71716	0.96972	0.82453
8	0.54870	0.98381	0.70449
9	0.65864	0.96374	0.78250
10	0.68422	0.98001	0.80583
11	0.59486	0.74945	0.66327
12	0.72457	0.88680	0.79752

After determining the eigenvalues of the data set, this paper conducts experiments on the number of different clusters to find the optimal number of clusters. Combined with the data results shown in [Figure 1], this paper determines that the optimal number of clusters is 7.

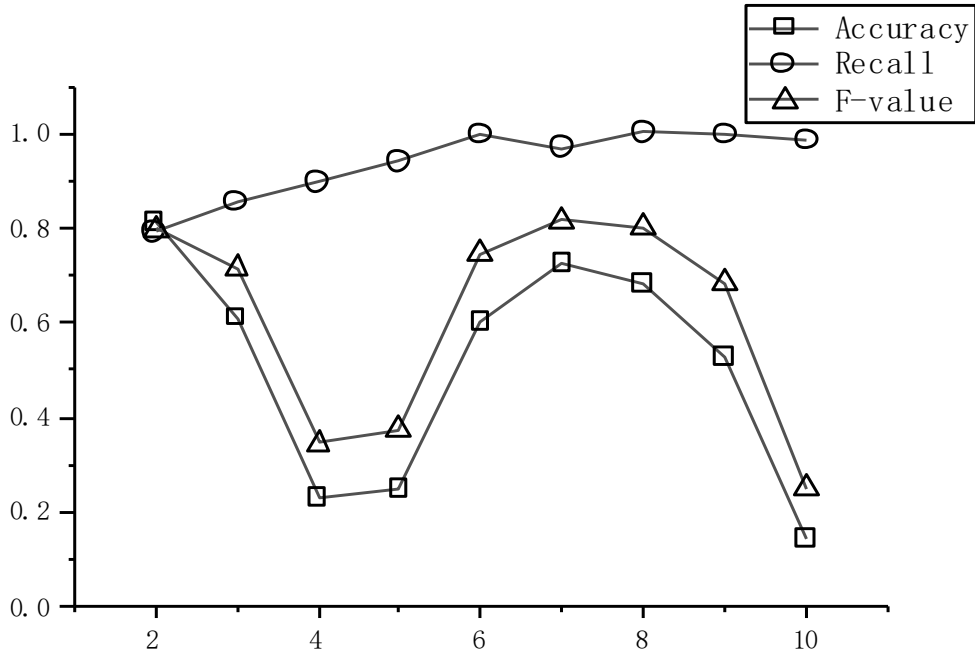


Figure 1. Accuracy, Recall, and F-value graphs under different clusters

For the data set in this paper, when the feature value is 7 and the number of clusters of GMM is set to 7, the accuracy of the prediction result is 0.71716, the recall rate is 0.96972, F value is 0.82453.

To further verify that the Gaussian mixture model can effectively predict sales transactions, this article uses artificial neural networks that are good at processing data information and more advanced convolutional neural networks to predict sales on the same data set, as shown in [Figure 2] and [Figure 3]. ROC curve. Based on this calculation, the prediction result of the artificial neural network model has been obtained: the accuracy rate is 0.37203, the recall rate is 0.39642, F value is 0.38383; The prediction result of the convolutional neural network model: the accuracy rate is 0.78022, the recall rate is 0.15127, the F value is 0.25341. F value is the result of comprehensive accuracy rate and recall rate. It is used to solve the contradiction between the two. The higher the F value, the better the model performance. Comparing the experimental results of the three models, this paper finds that the F value of the Gaussian mixture model is higher than the artificial neural network model and the convolutional neural network model. Therefore, the industrial supply chain sales forecasting method based on the Gaussian mixture model can predict sales transactions well.

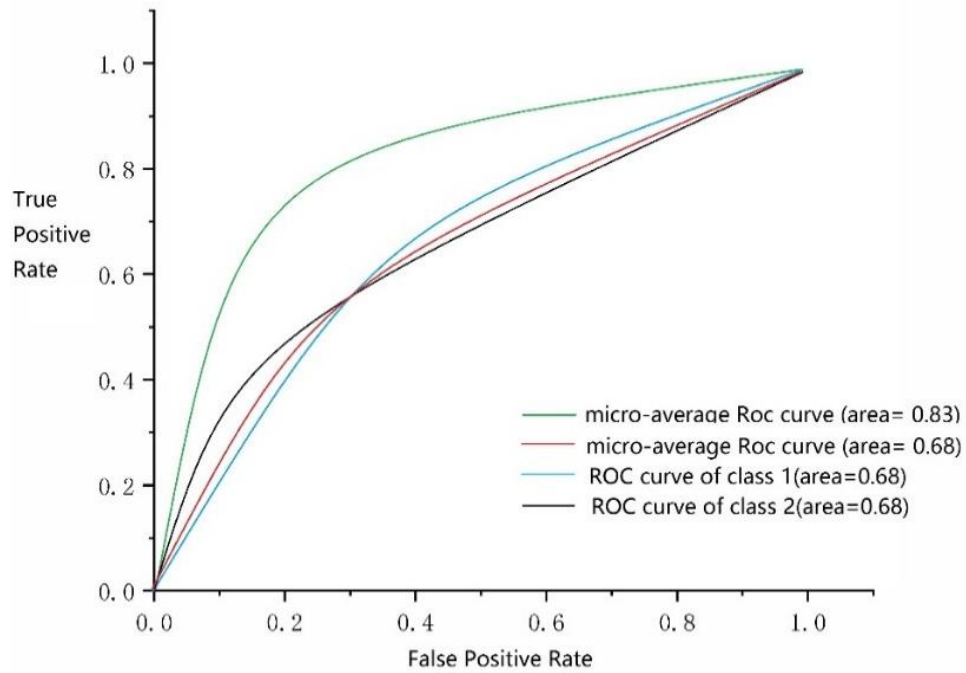


Figure 2. ROC curve of artificial neural network

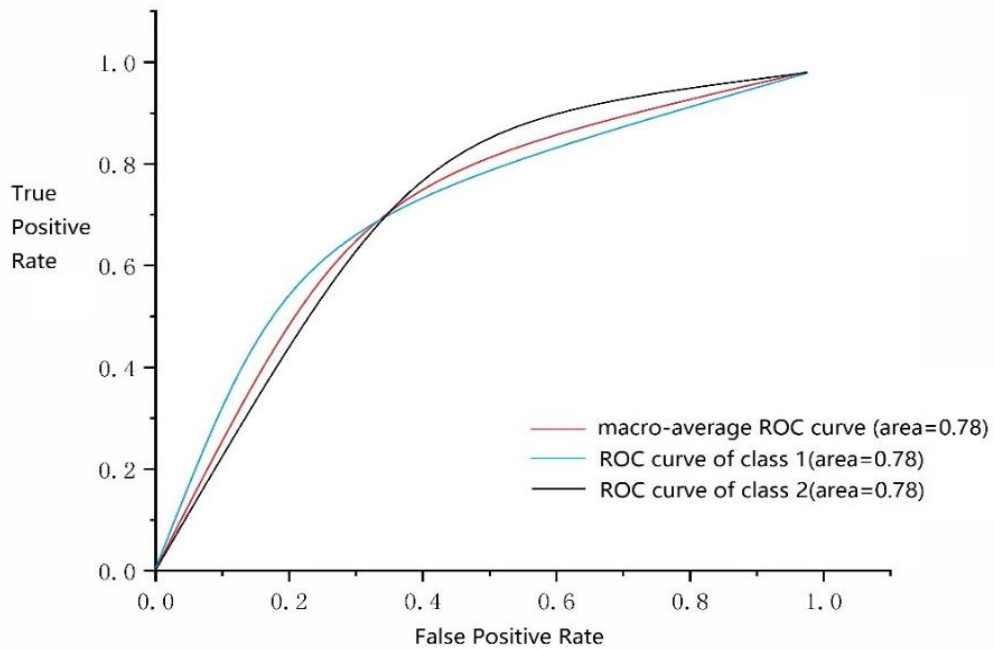


Figure 3. ROC curve of convolutional neural network

5. Conclusion

This paper proposes a sales forecasting method for the industrial supply chain based on the Gaussian mixture model. First, analyze the characteristics of the original sales data of the industrial supply chain, and generate the eigenvalue correlation ranking vector. Then, predict the parameters such as the number of clusters of the Gaussian mixture model. Then, perform Gaussian mixture clustering based on different cluster numbers and feature number combinations, and determine the feature values and clusters that can achieve better prediction results by comparing the accuracy of the prediction results, the recall rate, and the F value. This paper compares the Gaussian mixture model with the artificial neural network model and the convolutional neural network model on the original sales data set of the same industrial supply chain. The results show that the industrial supply chain sales forecasting method based on the Gaussian mixture model can predict sales transactions well.

References

- [1] M. A. Moon and J. T. Mentzer Jr., "Customer demand planning at lucent technologies: A case study in continuous improvement through sales forecast auditing," *Industrial Marketing Management*, vol.29, no.1, pp.19-26, (2000)
- [2] S. Thomassey, M. Happiette, and J. M. Castelain, "An automatic textile sales forecast using fuzzy treatment of explanatory variables," *Journal of Textile and Apparel, Technology and Management*, vol.2, no.4, (2011)
- [3] J Silva, J. V. Villa, and S. Cabrera, "Sale forecast for basic commodities based on artificial neural networks prediction," (2020)
- [4] A. Loureiro, V. L. Miguéis, and S. Da, "Exploring the use of deep neural networks for sales forecasting in fashion retail," *Decision Support Systems*, vol.114, pp.81-93, Oct. (2018)
- [5] A. A. Kindelan, L. M. Mata, and J. Godoy, "Analysis and projection of Pfizer's stock returns, in the period 2018-2020, through differential neural networks," *The Anáhuac Journal*, vol.19, (2019)
- [6] D. Hainaut, "A neural network analyzer for mortality forecast," *Astin Bulletin*, vol.48, no.2, pp.481-508, (2018)
- [7] M. Iotti and G. Bonazzi, "Profitability and financial sustainability analysis in Italian aquaculture firms by application of economic and financial margins," *American Journal of Agricultural and Biological Science*, vol.10, no.1, pp.18-34, (2015)
- [8] M. Kurematsu and H. Fujita, "A framework for integrating a decision tree learning algorithm and cluster analysis," *IEEE International Conference on Intelligent Software Methodologies*, IEEE, (2013)
- [9] K. Wagstaff, "Constrained K-means clustering with background knowledge," *Proceedings of ICML-2001*, (2001)
- [10] P. Li and C. H. Zhang, "Theory of the GMM kernel," *Proceedings of the 26th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee*, pp.1053-1062, (2017)
- [11] A. Hotho, S. Staab, and G. Stumme, "Explaining text clustering results using semantic structures," *7th European Conference on Principles and Practice of Knowledge Discovery*, (2003)
- [12] T. Steiner, R. Verborgh, and J. Gabarro, "Clustering media items stemming from multiple social networks," *Computer Journal*, vol.58, no.9, (2015)
- [13] L. Yao, Z. Li, and T. Z. Nie, "Research and implementation of a brand unity method across e-commerce platform," *2018 4th International Conference on Big Data Computing and Communications*, pp.236-242, (2018)
- [14] D. Y. An, X. H. Zheng, and C. M. Rong, "Gaussian mixture model-based interest prediction in social networks," *2015 IEEE 7th International Conference on Cloud Computing Technology and Science*, IEEE, pp.196-201, (2015)
- [15] C. H. Piao, Z. Jing, and L. J. Zheng, "Research on entropy-based collaborative filtering algorithm and personalized recommendation in e-commerce," *Service-Oriented Computing and Applications*, vol.3, no.2, pp.147-157, (2009)

- [16] N. Yang, G. X. Wang, and J. Hao, "A GMM-based user model for knowledge recommendation," 2017 3rd IEEE International Conference on Cybernetics (CYBCONF), IEEE, pp.1-5, **(2017)**
- [17] T. Higuchi, N. Itou, and S. Araki, "Online MVDR beamformer based on complex gaussian mixture model with spatial prior for Moise robust ASR," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol.99, pp.1-1, **(2017)**

This page is empty by intention.