

Construction of Customer Appeal Classification Model Based on Speech Recognition

Sheng Cao*, Yaling Zhang, Shengping Yan, Xiaoxuan Qi, and Yuling Li

Abstract

Aiming at the problems of poor customer satisfaction and poor accuracy of customer classification, this paper proposes a customer classification model based on speech recognition. First, this paper analyzes the temporal data characteristics of customer demand data, identifies the influencing factors of customer demand behavior, and determines the process of feature extraction of customer voice signals. Then, the emotional association rules of customer demands are designed, and the classification model of customer demands is constructed through cluster analysis. Next, the Euclidean distance method is used to preprocess customer behavior data. The fuzzy clustering characteristics of customer demands are obtained by the fuzzy clustering method. Finally, on the basis of naive Bayesian algorithm, a customer demand classification model based on speech recognition is completed. Experimental results show that the proposed method improves the accuracy of the customer demand classification to more than 80%, and improves customer satisfaction to more than 90%. It solves the problems of poor customer satisfaction and low customer classification accuracy of the existing classification methods, which have practical application value.

Keywords

Association Rules, Customer Demands, Euclidean Distance

1. Introduction

In the fierce market competition, customer service has become one of the important problems faced by enterprises in the market. Improving the ability of the customer service system to analyze and understand the demands of power customers is one of the important ways to improve the quality of customer service in the power industry. To efficiently and purposefully address the concentrated needs of power customers and improve customer satisfaction, scholars in relevant fields have conducted research and made some progress on how to classify customer demands [1,2]. Wang et al. [3] proposed a correlation analysis model between customer demands and marketing inspection business based on data correlation analysis. The model first intelligently classified customer needs to improve the efficiency and accuracy of demand screening. This paper established a correlation analysis model between customer demand and marketing audit based on a priori algorithm. Through data correlation analysis, it studied the correlation between customer demands and marketing audit, and designed a collaborative management model for customer demands and marketing audit business. Chen et al. [4] proposed a business fluctuation monitoring model

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manuscript received May 30, 2022; first revision September 22, 2022; second revision December 7, 2022; accepted January 29, 2023.

* **Corresponding Author:** Sheng Cao (cs7842339@163.com)

Marketing Service Center, State Grid Qinghai Electric Power Company, Xining, China (Cs7842339@163.com, zyling1006@163.com, 15003685939@163.com, qxxuan0701@163.com, ly1_feng@126.com)

based on customer demands. This method is based on the massive business data of big data platforms, and combined with customer behavior characteristics and business characteristics, divides all three-level business subclasses into business theme scenarios. For each scenario type, the SARIMA model is established according to time series to calculate the threshold warning of each time node, and the model results are refreshed and displayed in real-time through chart display tools and a data console. This method can improve the accuracy of customer classification, but customer satisfaction is poor. Peng et al. [5] developed a deep neural network-based method for power customer demand prediction. This method tapped the hidden customer demand and improved customer satisfaction, but the accuracy of customer classification was poor. Yue and Fang [6] proposed a classification method for telecommunications customer complaints based on filtering models and random forests. Through the filtering model, the high-dimensional mixed features of telecommunications customer attributes are obtained, and the information gain ratio of telecommunications customer complaints is calculated, which is beneficial for determining the priority of customer complaints. Through the decision tree method, irrelevant redundant features are eliminated, achieving a dimensionality reduction effect. A customer complaint classification model based on random forests is established. Niranjana et al. [7] built a multi-product closed-loop supply chain customer classification model.

In view of the above issues, this paper effectively constructs a customer demand classification model based on speech recognition. This model can analyze the periodic data characteristics of customer demand data and determine the influencing factors of customer demand behavior. By setting the Myer frequency cepstrum coefficient (MFCC) parameter of the customer service recording audio length, it can obtain the characteristics of the customer's voice signal. In addition, emotional association rules for customer needs have been established, and a customer needs classification model has been constructed using clustering analysis algorithms. Meanwhile, this paper also innovatively uses Euclidean distance method to preprocess customer behavior data. After obtaining the fuzzy clustering features of customer needs through fuzzy clustering methods, customer demands are classified based on naive Bayesian algorithm. The experimental results show that this method can effectively improve customer satisfaction and customer classification accuracy.

2. Feature Extraction of Voice Signals Requested by Customers

2.1 Period Data Characteristics of Customer Demand Data

Power supply service demand refers to the demand of power customers within the company's operating region to express their dissatisfaction with the damage caused to their rights and interests by the power supply enterprise within their responsibilities. The responsibilities of power supply enterprises include power supply services, operations, power outages, power supply quality, and power grid construction.

The appeal volume starts to rise at 7 am every day, reaches the peak at about 10 am, then drops to noon to form a small trough, and then fluctuates slowly. It reaches a small peak at 4 pm and 8 pm and drops sharply after 8 pm, and the trough of the whole day is from 0 am to 6 am. In general, the daily distribution of complaints is consistent with the law of people's work, life and work and rest [8].

2.2 Regional Characteristics of Demand Data

As can be seen from Fig. 1 of regional distribution, complaints are concentrated in District A of the

city. District A is the largest area in the city, with active cultural activities, plenty of diplomatic and business buildings, and a large population density. Zone B, located in the center of the city, is well developed in cultural tourism. Zone E is the suburb with the largest demand in the region, ranking third. Excluding individual outer suburbs, covering a small area has less demand. The demand gap in other regions is not significant, and the demand is the lowest in recent years. Zone Q is a new economic development zone [9].

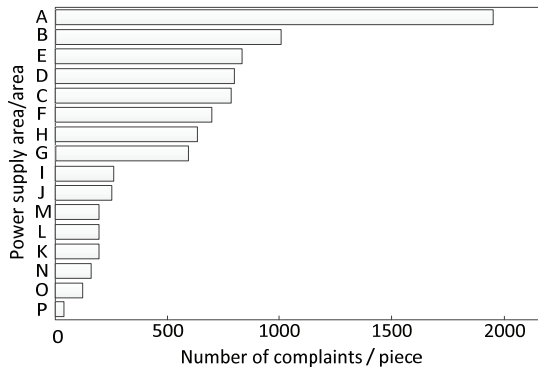


Fig. 1. Regional distribution of appeal quantity.

The demand business distribution data of each region is shown in Table 1. It shows that the complaint business in most regions focuses on the business service demand, followed by the metering device problems caused by the electricity tariff business and characteristic business in the power supply service demand and business demand. In the power supply service category, the enterprise employees have direct contact with customers, and the service complaints mostly involve the service personnel's business level and service attitude. It follows that the enterprise needs to strengthen its personnel's service awareness and business norms. From the perspective of customers, electricity tariff is a sensitive issue, which should be paid attention to in time.

Table 1. Percentage of regional distribution of appeal data (unit: %)

	Service demand	Business appeal	Power supply quality	Power grid construction	Stop power transmission
Area A Power Supply Company	31.84	45.46	7.65	2.60	12.45
Area B Power Supply Company	28.41	49.75	3.69	5.48	12.66
Area C Power Supply Company	27.99	53.18	4.07	6.49	8.27
Area D Power Supply	28.05	33.17	21.57	5.24	11.97
Area E Power Supply	25.30	35.37	17.87	7.43	14.03
Area F Power Supply	27.75	39.34	15.88	8.30	8.73
Area G Power Supply	28.40	29.58	22.12	7.85	12.04
Area H Power Supply	31.38	46.31	9.40	3.69	9.23
Area I Power Supply	30.97	29.56	19.50	10.22	9.75
Area J Power Supply	28.47	27.40	20.64	11.03	12.46
Area K Power Supply	33.07	26.38	12.20	14.57	13.78
Area L Power Supply	34.87	49.23	3.08	3.08	9.74
Area M Power Supply	30.77	18.97	18.46	19.49	12.31
Area N Power Supply	33.95	22.22	21.60	17.90	4.32
Area O Power Supply Company	33.06	37.90	11.29	8.87	8.87
Area P Power Supply Company	41.46	41.46	4.88	2.44	9.76

Since the research object of customer behavior is the variable behavior, it will change with the characteristics of customers and environmental changes, which can be expressed by public signs:

$$B = f(P, E) \quad (1)$$

Among them, B is customer behavior, P is personal factors, and E is environmental factors (e.g., cultural environment, social and family factors, etc.).

2.3 Feature Extraction of Speech Signal

The essence of feature extraction of the speech signal is to change the speech signal from an analog signal to a digital signal, use some characteristic parameters reflecting the characteristics of the speech signal to represent speech, and digitize the speech signal for processing by computer [10]. Whether feature parameters can distinguish speech signals and accurately extract feature parameters will directly affect the accuracy of speech recognition. There are many measurement standards for speech signal characteristics, the commonly used are linear predictive cepstral coefficient (LPCC) and MFCC. The use of LPCC has a small amount of computation and is easy to implement, but it has poor anti-noise ability. In addition, since LPCC is linear at all frequencies and completely different from human auditory features, the MFCC is adopted here to extract speech features [11]. MFCC is raised based on the auditory model of the human ear and has been widely used in recent years to better improve the recognition performance of the system [12]. The human cochlea is actually like a filter bank, which has a linear relationship with the frequency below 1,000 Hz and a logarithmic relationship with the frequency above 1,000 Hz. MFCC is precisely the correlation between the two to calculate frequency-related spectral features. Since the Mayer frequency is not linearly correlated with the frequency, the calculation accuracy of MFCC decreases when the frequency exceeds a certain range. Therefore, in practice, only MFCC at low frequency is usually calculated, and MFCC at a medium frequency and high frequency is discarded.

The specific relationship between Mayer frequency and the actual frequency of audio is as follows:

$$f_{mel}(f) = 2595 \cdot \log\left(1 + \frac{f}{700\text{Hz}}\right) \quad (2)$$

The extraction of MFCC parameters includes the following steps: after the audio signal is divided into frames and windowed, the time domain signal is transformed into the power spectrum of the signal through the fast Fourier transform (FFT). Then a set of triangular window filters (usually 24) with linear distribution on the Mayer frequency scale are used to process the power spectrum of the signal.

Numerous experiments show that to improve the recognition performance of the algorithm, difference parameters can be added to the speech features. In this program, the first- and second-order different parameters are used to more accurately represent the spectrum characteristics of speech. The flow chart of feature extraction of the speech signal is shown in Fig. 2.

In this paper, 256 is taken as the number of points in each frame of the speech sampling sequence, and the order of the Mel filter is 24. To facilitate the synthesis of samples, now 20 MFCC parameters are extracted uniformly regardless of the length of the collected audio.

After the feature signal is extracted, "OK" and "Cancel" speech signals are added with marks "1" and "2", respectively to show the difference. Then all the features extracted from the training samples are

stored in data in the form of an array. In the mat file, there are 21 columns in the array. The first column is identification, and the last 20 columns are speech characteristic signals.

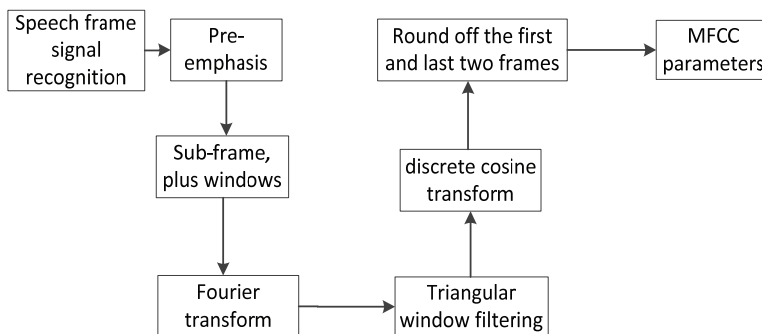


Fig. 2. Flow chart of speech signal feature extraction.

3. Construction of Customer Demand Classification Model

3.1 Design Association Rules for Customer Demand Classification

Simple correlation is used to describe the degree of linear correlation between two random variables. In this paper, Pearson simple correlation coefficient is selected, which is the most used method at present. Pearson simple correlation coefficient is used to express the closeness and direction of linear correlation between two variables ρ . And R represents the overall Pearson correlation system and the sample Pearson correlation coefficient, respectively.

3.2 Build a Customer Demand Classification Model through Cluster Analysis

The basic idea of clustering analysis is to use similarity to measure the proximity between things, thereby achieving classification. The essence of fuzzy cluster analysis is to construct a fuzzy matrix according to the attributes of the research object itself and determine its classification relationship according to a certain membership degree.

Since there may be differences in each indicator unit and order of magnitude, a direct comparison cannot be carried out in many cases. If the original data is directly used for analysis, it is likely to highlight the weight of some indicators with large orders of magnitude in the classification and ignore the characteristic indicators with small orders of magnitude. As a result, changing a unit will completely overturn the clustering results and get different results. Therefore, the original data should generally undergo dimensional processing before clustering analysis, so that the values of each characteristic indicator are unified within a specific range and comparable. There are many data standardization methods for samples. The mean normalization method, standard deviation normalization method, range normalization method, and maximum normalization method are common data processing methods.

3.3 Classification of Customer Demands based on Naive Bayesian Algorithm

This paper uses the naive Bayes algorithm to classify customer demands, so this paper focuses on the

naive Bayes algorithm. As is well-known, the Bayesian theorem describes the relationship between two conditional probabilities. From it, two Bayesian classification algorithms have been developed: the Bayesian belief network and naive Bayes.

If the data population is X , the sample obtained from it is x , and H is a descriptive condition.

A priori probability: the probability that y occurs in the data space X , which is called the a priori probability of x , expressed of $P(x)$.

Posterior probability: when the given data sample is x , the probability that condition h holds is called the posterior probability of H under condition x , expressed by $P(H|x)$. A posteriori probability synthesizes a variety of information such as data as a whole, samples and a priori.

A priori probability can be obtained from experience, history and analysis, and a posteriori probability is usually not easy to obtain directly. The Bayesian theorem provides a method to calculate $P(H|x)$ from $P(H)$, $P(x)$ and $P(x|H)$, such as Eq. (3). The classification problem is to obtain its category through a given text sample, that is, to determine the posterior probability. Therefore, the Bayesian algorithm has been widely used in the field of classification because of its idea of inferring the unknown from the known.

$$P(H|x) = \frac{P(x|H)P(H)}{P(x)} \quad (3)$$

As a statistical classification method based on Bayesian theorem, the basic idea of naive Bayesian algorithm is to obtain a posterior probability of all attributes using the prior probability of training samples according to Bayesian formula. The attribute with the highest posterior probability is found and recognized as the category of the object. In terms of classification of customer demands, it is implemented according to the above constraints.

4. Test and Analysis

The experimental sample comes from 390G voice sample data of customer service complaints of a power supply company in a certain area and is recognized and classified on a microcomputer in the simulation experiment. The simulation experimental environment is the Win10 operating system with Intel i5-9400F processor and NVIDIA GeForce GTX1070 graphics card. The experiment uses Python programming language to train and test the model on the naive Bayesian algorithm learning framework. Microsoft SQL Server 200 is selected as the tool to build the data warehouse. The data mining module developed by Microsoft Visual Basic 6.0 is used to process the data, and the processing results are saved in the SQL server database. The design model, traditional method 1 [2], and traditional method 2 [3] are used to classify customer needs respectively. The accuracy results of statistical classification are shown in Fig. 3.

According to Fig. 3, the classification accuracy of this model is higher than 80% and significantly higher than the other methods. It proved that the classification performance of this method is superior.

The classification results of the design of the model, traditional method 1, and traditional method 2 are applied to the power system, and the user satisfaction results of the three statistical methods are shown in Fig. 4.

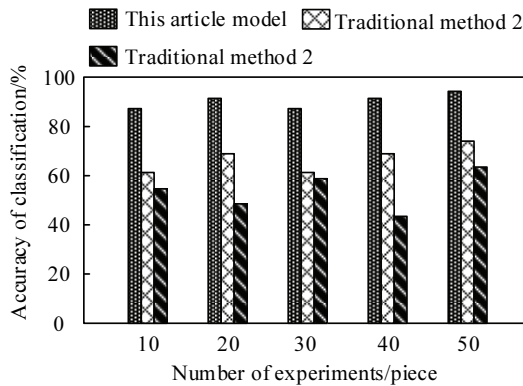


Fig. 3. Results of classification test accuracy.

According to Fig. 4, the user satisfaction obtained after the application of the model is higher than 90%, which is significantly better than the other two methods, indicating that the practical application effect of this method is good.

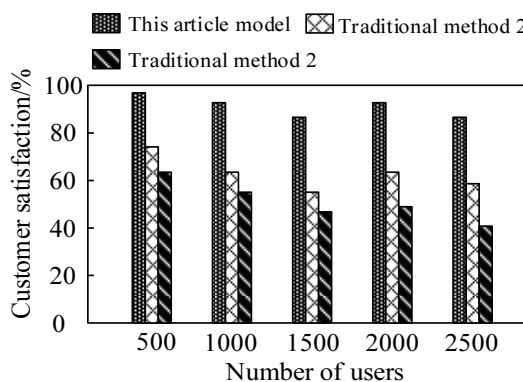


Fig. 4. Satisfaction results.

5. Conclusion

This paper constructs a customer appeal classification model based on speech recognition. The purpose of this study is to analyze the temporal data characteristics of customer demand data, determine the influencing factors of customer demand behavior, and determine the feature extraction process of customer voice signals. Firstly, the model extracts the audio length MFCC parameters to obtain the characteristics of the voice signal required by the customer. Secondly, this paper designs emotional association rules for customer needs, and constructs a customer demands classification model through clustering analysis. Customer demand classification grounded on naive Bayesian algorithm and customer demand recognition model based on speech recognition have been completed. The experimental results demonstrate that the proposed method improves the accuracy of customer demand classification and customer satisfaction, and has practical application value. However, due to the limited conditions, this method mainly focuses on classification accuracy, but the classification efficiency is not significantly

improved. In view of this, future studies can improve classification efficiency based on ensuring the classification accuracy.

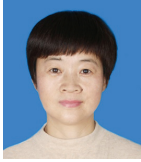
References

- [1] F. Mumali, "Artificial neural network-based decision support systems in manufacturing processes: a systematic literature review," *Computers & Industrial Engineering*, vol. 165, article no. 107964, 2022. <https://doi.org/10.1016/j.cie.2022.107964>
- [2] H. Zhu, L. Qi, and H. Q. Gu, "Research on intelligent classification of operator customer complaints based on artificial neural network algorithm," *Telecom Engineering Technics and Standardization*, vol. 34, no. 3, pp. 31-35, 2021.
- [3] X. Wang, R. H. Wang, and L. Li, "Customer demand and marketing audit business association analysis model based on data association analysis," *Electronic Paradise*, vol. 16, no. 28, pp. 89-98, 2020.
- [4] X. Chen, S. Wang, and J. Liu, "Big data application of business fluctuation monitoring model based on customer demand," *China New Technologies and Products*, vol. 2021, no. 17, pp. 34-36, 2021.
- [5] L. Peng, J. Zhu, and Y. F. Zou, "Prediction of power customer demands based on deep neural network," *Computer and Modernization*, vol. 2020, no. 5, pp. 22-28, 2020.
- [6] D. Y. Yue, and S. Fang, "Classification method of telecom customer complaints based on Filter model and random forest," *Computer Applications and Software*, vol. 37, no. 7, pp. 221-226, 2020.
- [7] T. Niranjana, P. Parthiban, and P. Goel, "Designing an integrated multichannel multiproduct closed loop sustainable supply chain network meeting customer requirements," *International Journal of Enterprise Network Management*, vol. 12, no. 2, pp. 113-130, 2021.
- [8] J. Richardson, "How data management drives omnichannel supply-chain planning," 2019 [Internet]. Available: <https://www.supplychainbrain.com/articles/29338-data-management-drives-omnichannel-supply-chain-planning-to-meet-evolving-customer-demand>.
- [9] Y. C. Liao and K. H. Tsai, "Innovation intensity, creativity enhancement, and eco-innovation strategy: the roles of customer demand and environmental regulation," *Business Strategy and the Environment*, vol. 28, no. 2, pp. 316-326, 2019.
- [10] T. Breckle, M. Manns, and J. Kiefer, "Assembly system design using interval-based customer demand," *Journal of Manufacturing Systems*, vol. 60, pp. 239-251, 2021.
- [11] G. He and Z. Zhang, "Design of camera control system based on speech recognition," *Electronic Design Engineering*, vol. 29, no. 12, pp. 9-13, 2021.
- [12] T. Jiang, Q. X. Li, and Q. M. Li, "Fine grained text sentiment classification method based on improved capsule network," *Computer Simulation*, vol. 38, no. 10, pp. 466-470, 2021.



Sheng Cao <https://orcid.org/0000-0001-8789-7890>

She graduated from North China Electric Power University in 2010, majoring in electrical engineering and automation. Now she is working in the Marketing Service Center of State Grid Qinghai Electric Power Company, serving as the director of the customer service department, and researching the direction of marketing. She has published five academic articles and participated in five scientific research projects.



Yaling Zhang <https://orcid.org/0000-0002-3835-4782>

She graduated from North China Electric Power University in 2010, majoring in electrical engineering and automation. Now she is working in the Marketing Service Center of State Grid Qinghai Electric Power Company, serving as the director of customer service department, and researching marketing direction. She has published five academic articles and participated in five scientific research projects.



Shengping Yan <https://orcid.org/0000-0002-7202-2634>

She graduated from Zhengzhou University in 2016, majoring in electrical engineering and automation. Now she is working in the Marketing Service Center of State Grid Qinghai Electric Power Company, responsible for the customer service department, and researching the direction of high-quality services. She has published five academic articles and participated in five scientific research projects.



Xiaoxuan Qi <https://orcid.org/0000-0001-7168-6925>

She graduated from Chongqing University in 2017, majoring in electrical engineering and automation. Now she is working in the State Grid Qinghai Electric Power Company Marketing Service Center, responsible for the customer service department, and researching the direction of customer service and power marketing. She has published five academic articles and participated in five scientific research projects.



Yuling Li <https://orcid.org/0000-0002-3537-5959>

She graduated from Tianjin University of Science and Technology in 2017, majoring in computer technology. She received a master's degree from Qinghai University in 2020, majoring in computer technology. Now she is working in the Marketing Service Center of State Grid Qinghai Electric Power Company, responsible for the customer service department, and researching service quality analysis direction. She has published one academic article and participated in one scientific research project.