ISSN: 2664-9640

Computer Big Data Technology in the Statistics of Active Monitoring Data for Power Supply Service Appeals

Xusheng Liu*, Yeteng An, Ying Wang, Shihui Xu

State Grid Customer Service Center Co. Ltd, Tianjin 300306, China

dingmaomaogirl@163.com

Abstract

In order to better dig deeper into the information contained in power supply service complaint work orders, the research is based on natural language processing technology to conduct in-depth text mining of power customer complaint work orders. We use LDA model and other word segmentation technology to analyse the accepted content in the complaint work order for word frequency statistics. The LDA model algorithm is used to calculate the keyword importance weight value to extract the keyword frequency with a large weight value as the final result of customer complaint text mining. And use word cloud analysis technology to visualize the analysis results. Research shows that the system can effectively perform intelligent clustering of customer appeal texts, and can realize statistics and analysis of active monitoring data of power supply service appeals.

Keywords

Computer, big data technology, power supply service, active monitoring data.

1. Introduction

As the electricity distribution market continues to open up, new distribution networks will allow external capital investment. Registration of new distribution and distribution companies has become a hot area for social capital. Electricity sales companies in various places have sprung up. They are very likely to become powerful competitors in the field of power sales for grid companies in the near future. The loss of customer resources will be Inevitably, how to reduce the loss of customer resources will be an urgent problem to be solved. The primary measure to ensure customer stability is to improve customer satisfaction, which means reducing customer complaints [1]. Therefore, natural language processing technology is used to conduct text mining and analysis on customer complaint work orders to understand the main problems of customer complaints, and to improve differentiation in a targeted manner. The current service strategy is an important measure to improve customer satisfaction and increase customer stickiness.

The article proposes a method for identifying customer demands of power supply services based on the LDA model. The recognition method includes the following processes: Text pre-processing: text segmentation and text quantification of the text in the work order, mainly to segment the long content of the text according to certain rules; LDA theme generation: random selection of samples from the full sample as training samples With test samples, performance parameters are obtained by training the training samples, and then the test samples are used to test and obtain performance parameters with a higher recognition rate [2]. The performance parameters are used to generate recognition text to identify customer requests contained in the work order; automatic text classification.

ISSN: 2664-9640

2. LDA model

Current probabilistic topic models are generally based on the same idea—a text is a random mixture of several topics. Different models will further make different statistical assumptions and obtain model parameters in different ways.

2.1. Model introduction

A text usually needs to discuss several topics, and the specific vocabulary in the text reflects the specific topic discussed. In statistical natural language processing, the method of modelling text topics is to treat the topics as the probability distribution of words, and the text is a random mixture of these topics [3]. Assuming there are T topics, the i word w_i in the given text can be expressed as follows:

$$P(w_i) = \sum_{j=1}^{T} P(w_i \mid z_i = j) P(z_i = j) \tag{1}$$

The probability of word w in text d is:

$$P(w \mid d) = \sum_{j=1}^{T} \varphi_w^{(z=j)} \bullet \psi_{z=j}^{(d)} \tag{2}$$

Although the specific values of α and χ will affect the extent to which the topic and vocabulary are used, the ways in which different topics are used have hardly changed, and the ways in which different words are used are basically the same. Therefore, a symmetric Dirichlet distribution can be assumed [5]. That is, all α take the same value, and all χ take the same value.

2.2. Gibb's sampling

For the LDA model of this article, only the vocabulary assignment of the topic, that is, the variable sampling, is needed. Denote the posterior probability as $P\left(z_i = j \mid z_{-i}, w_i\right)$, and the calculation formula is as follows:

$$P\left(z_i = j \mid z_{-i}, w_i\right) = \frac{\dfrac{n_{-i,j}^{(w_i)} + \chi}{n_{-i,j}^{(\cdot)} + W\chi} \bullet \dfrac{n_{-i,j}^{(d_i)} + \alpha}{n_{-i,\bullet}^{(d_i)} + T\alpha}}{\sum_{j=1}^{T} \dfrac{n_{-i,j}^{(w_i)} + \chi}{n_{-i,j}^{(\cdot)} + W\chi} \bullet \dfrac{n_{-i,j}^{(d_i)} + \alpha}{n_{-i,\bullet}^{(d_i)} + T\alpha}} \tag{3}$$

Abandon the vocabulary mark and use w to represent the unique word. For each single sample, the values of φ and ψ can be estimated as follows:

$$\hat{\varphi}_w^{(z=j)} = \frac{n_j^{(w)} + \chi}{n_j^{(\cdot)} + W\chi}, \quad \hat{\psi}_{z=j}^{(d)} = \frac{n_j^{(d)} + \alpha}{n_\bullet^{(d)} + T\alpha} \tag{4}$$

3. Key technology solutions

3.1. Refinement of complaint feature labels

Complaint feature label extraction is to first process the content of the historical complaint sample work order with DataVec, combine with Baidu word database for word segmentation, extract the complaint-specific vector label words, and then systematically mark these label words in the original sample work order content, and then Individual label words are reorganized and refined, and finally the complaint samples are segmented and stop words are removed, so as to obtain the complaint formatted sample data. The specific process is shown in Figure 1.

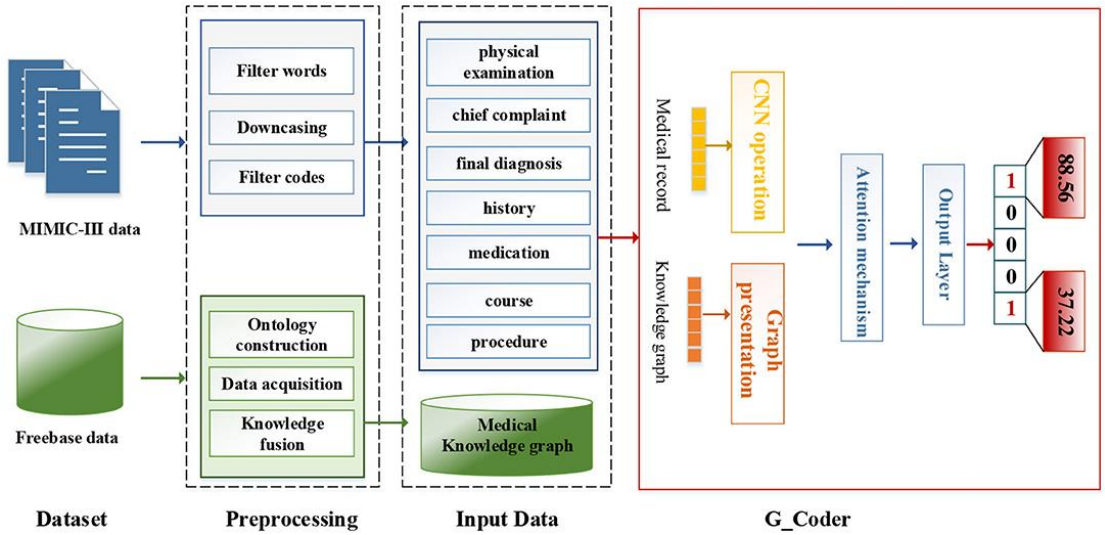

Figure 1. Complaint feature label extraction process

3.2. Identification of Suspected Complaints

Suspected complaint identification is based on text similarity judgment to identify and classify suspected complaints. Once the document segmentation is expressed by spatial vectors, the semantic similarity between texts can be measured by the geometric relationship between these two vectors in the space. After model training and evaluation, the similarity is set to 70%, and the accuracy of complaint recognition reaches about 91.5%. The specific implementation process is shown in Figure 2. Based on the above learned model results, all 95598 incoming calls and receiving orders are determined and identified one by one.

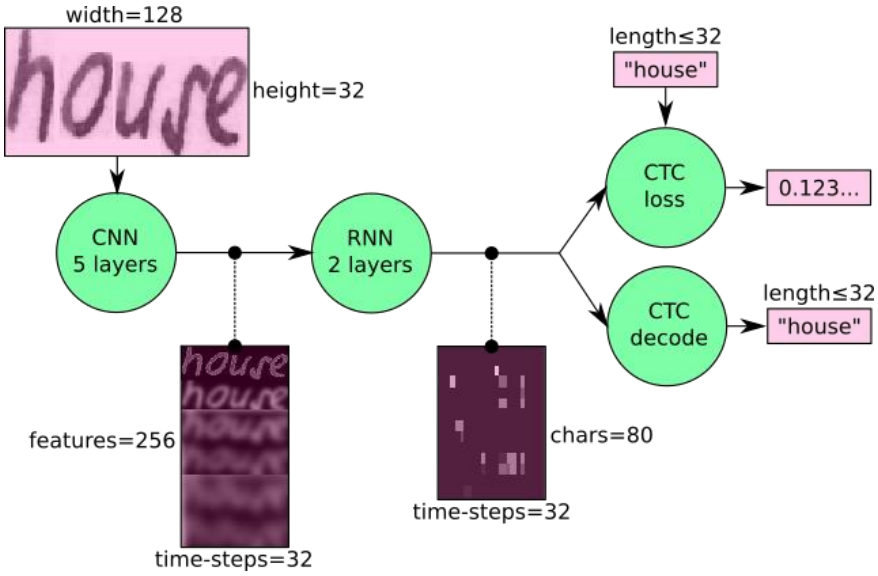

Figure 2. Model learning and training implementation process

4. Experimental detection

Based on natural language processing technology, conduct in-depth text mining on power customer complaint work orders, use word segmentation technology to analyse the content of complaint work orders, carry out feature selection and dimensionality reduction processing on the word segmentation results, and carry out word frequency statistics and use word cloud analysis The technical analysis results are visually displayed, the main problems of current power customer complaints are controlled, and differentiated service strategies are provided for different types of power customers to improve customer satisfaction and loyalty.

ISSN: 2664-9640

4.1. Implementation of text segmentation

Text segmentation refers to the use of a computer to automatically segment the words of the text. Through the Jieba package in the big data software Python, the hidden Markov model is used to realize the word segmentation of the content of customer complaints.

4.2. Keyword frequency extraction

Through the above-mentioned feature selection and dimensionality reduction of the word segmentation results, the filtering of irrelevant words is realized, and keywords related to the electric power business are left. Combining the actual power business, the existing keywords are further screened, the keyword importance weight value is calculated through the TFIDF (term frequency-inverse document frequency) algorithm, and the keyword frequency with a large weight value is extracted as the final result of customer complaint text mining.

4.3. Visual display

Using Python software to use word cloud analysis to realize the text mining results of complaint tickets is shown in Figure 3 below.

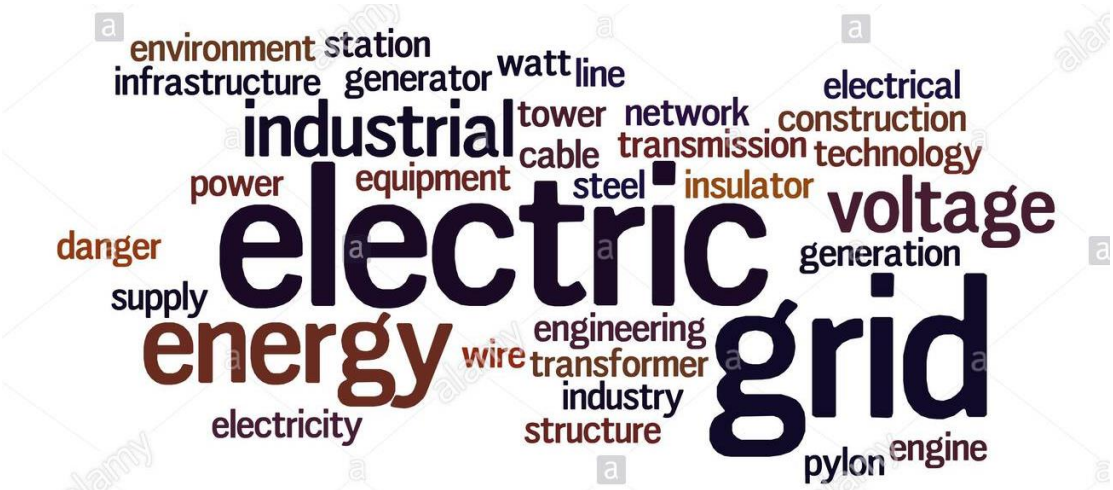

Figure 3. Complaint text word cloud

It can be seen from the figure that in customer complaints, the words "business hall", "power outage", and "fault" appear frequently, indicating that customers have major opinions on business halls, power outages, failures, etc. You can start from these aspects. Measures such as improving the service level of the business hall, reducing power outages or outage information notifications, strengthening troubleshooting to reduce failures, etc., so as to improve customer satisfaction and improve customer complaints.

5. Conclusion

This paper uses text mining technology based on natural language processing, combined with power business requirements, and special research on hot business work orders, to break the blind spots of customers' power demand, improve the management level of users' electricity demand, and realize hot complaint business work orders the reason is digging. The application of special topics will improve the work efficiency of the customer service department, provide decision support for the realization of active and accurate customer service, and enhance customer service capabilities.

References

[1] Kabir, A. I., Ahmed, K., & Karim, R. Word Cloud and Sentiment Analysis of Amazon Earphones Reviews with R Programming Language. Informatica Economica, 24(4) (2020) 55-71.

[2] Wang, J., Zhang, G., Wang, W., Zhang, K., & Sheng, Y. Cloud-based intelligent self-diagnosis and department recommendation service using Chinese medical BERT. Journal of Cloud Computing, 10(1) (2021) 1-12.
[3] Tan, S. P. Classifying Multilingual Text in Quality Assurance. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(3) (2021) 1987-1992.
[4] Bortolini, R., & Forcada, N. Analysis of building maintenance requests using a text mining approach: Building services evaluation. Building Research & Information, 48(2) (2020) 207-217.
[5] Diprose, W. K., Buist, N., Hua, N., Thurier, Q., Shand, G., & Robinson, R. Physician understanding, explainability, and trust in a hypothetical machine learning risk calculator. Journal of the American Medical Informatics Association, 27(4) (2020) 592-600.
[6] Sindhani, M., Parameswar, N., Dhir, S., & Ongsakul, V. Twitter analysis of founders of top 25 Indian startups. Journal for Global Business Advancement, 12(1) (2019) 117-144.