IOP Conference Series: Materials Science and Engineering

Intelligent Assessment of 95598 Speech Transcription Text Quality Based on Topic Model

Intelligent Assessment of 95598 Speech Transcription Text Quality Based on Topic Model

Bochuan Song[1], Peng Wu[1,*], Qiang Zhang[1], Bo Chai[1], Yuanbo Gao[1], He Yang[1]

[1]Artificial Intelligence on Electric Power System State Grid Corporation Joint Laboratory (GEIRI), Global Energy Interconnection Research Institute, Beijing 102211, China

[*]Corresponding author Email: wup@geiri.sgcc.com.cn

Abstract. The quality of speech transcripts is of great significance to power system management and is an important basis for supporting subsequent data analysis. In this paper, based on the characteristics of speech transcription texts of customer service, this paper proposes an analysis method combining manual processing and latent Dirichlet allocation topic model, analyzing the transcribed texts. First, data preprocessing is performed on the State Grid's work order data, and then the text topic distribution calculation is performed by the LDA topic model, and the topic parameter is set to a total of 100 topics. Next, the unsupervised clustering of the documents is performed by the k-means method, and the similarity between the files is obtained. Finally, the quality of the data is analyzed by combining manual labeling and manual evaluation. For the first time, this paper marks and identifies the State Grid's work order analysis data, which is a pioneering work for natural language processing technology in the field of power grid.

1. Introduction

95598 is the main window for the customer service of the State Grid Corporation. At the current stage, the 95598 phone service serves the customers manually. State Grid customer service generates 100,000 to 150,000 pieces of work order data per day. The information passing of 95598 telephone service mainly relies on the work order text record, the work order information extraction is completed by the customer service full-time staff, and the main information extraction is inevitably different from person to person. At the same time, it is difficult to extract unstructured text information such as work orders. Therefore, it is necessary to study the topic recognition requirements and topic classification technology of customer service call based on customer voice texts, improve the topic recognition and classification accuracy in the context of customer service calls, and explore the automatic identification of new demands.

There have been some researches on 95598 customer service in China. Paper [1] establishes a model of 95598 work order text automatic classification, and by realizing the automatic classification of 95598 work order text it provides a method and data basis for analyzing the impact of 95598 phone service on customers' electricity demand. The paper [2] takes the work order monitoring data of a company's 95598 complaints in the first quarter as a sample, and analyzes the number of complaints and the processing time of the complaints. Paper [3] carries out text analysis based on the work order data of State Grid Corporation of China, in which a set of customer service satisfaction analysis framework based on text mining is proposed. Through this framework, the identification of unsatisfactory work orders and the identification of unsatisfactory reasons can be realized.

The two basic problems of text classification are: the feature representation of the text and the establishment of the classification model. The commonly used methods for feature representation of the text are the word bag model and the TF-IDF method, which are relatively shallow methods, and it is difficult to capture the semantic information in the text. This paper proposes a method using the LDA topic model for the current problem. The LDA topic model has been extensively studied as a mature analytical method. Paper [4] systematically expounds LDA topic model parameter estimation and Gibbs sampling algorithm, introduces common LDA improvement and extension models, and finally analyzes the application of LDA model in text mining field. Paper [5] constructs an LDA model suitable for dynamic topic content mining, and then studies the evolution of content topics over time from the perspective of topic similarity and intensity. Paper [6] proposes a text similarity calculation method based on LDA topic model. This method mines the relationship between different topics and words hidden in the text, and clusters the text similarity matrix to evaluate the clustering effect. Paper [7] proposes a social label recommendation method based on the LDA topic model. This method uses the LDA topic modeling technology to extend the traditional recommendation method based on the relationship between objects to the unified recommendation method based on the relationship between objects and the characteristics of the resource content.

The main contribution of this paper is to apply the LDA topic model to the text quality assessment problem of the grid customer service, laying a solid foundation for the development of text understanding tools in the future. The main content of the article is arranged as follows:

The first chapter introduces the background of the research; the second chapter mainly introduces the text quality assessment preprocessing; the third chapter introduces the LDA topic model and feature extraction; the fourth chapter introduces K-means and its use in this research; fifth chapter introduces the manual rules for quality assessment; the sixth chapter show the results, in chapter 7 the full text is summarized.

2. Text quality assessment preprocessing

In order to complete the text quality analysis of the work order record, we preprocessed the text first. This includes converting data from excel to plain text documents, document encoding conversions, de-stopping words, and word segmentation. The second step uses the LDA topic model to calculate the text topic distribution and extract the semantic features of the text. In the third step, unsupervised clustering of the work order record documents based on extracted features is performed, and the similarity between the documents is obtained. The fourth step is to combine the manual annotation and manual rules to analyze the quality of the data. The general analysis process is shown in Figure 1.

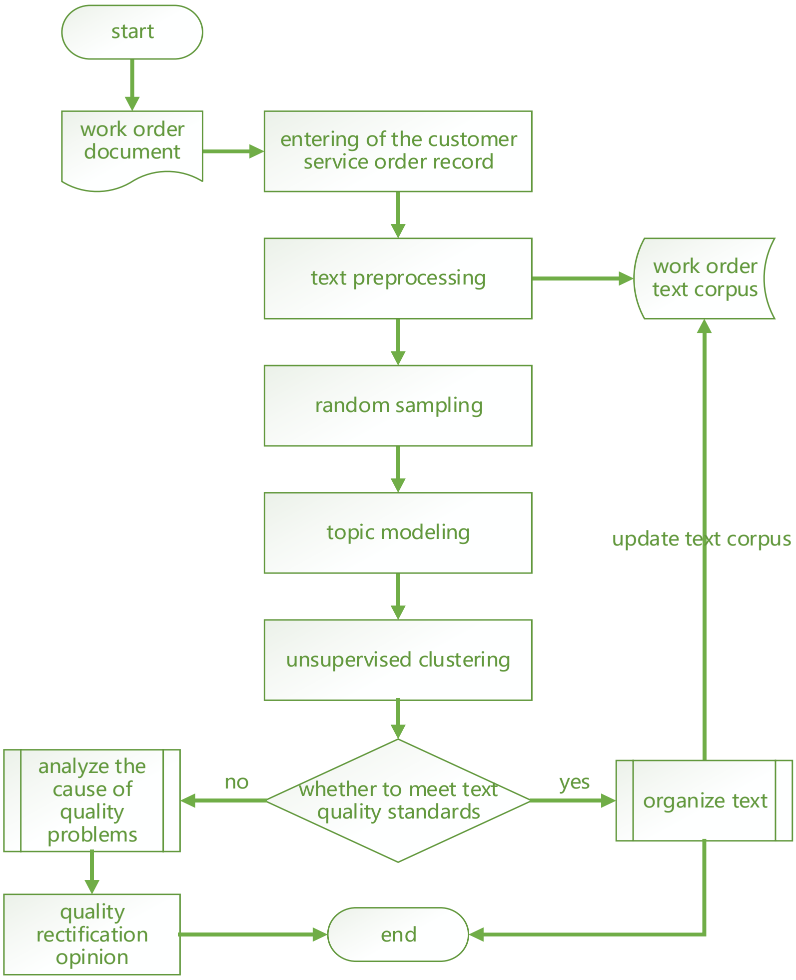

Figure 1. flow chart of text quality analysis of work order records

The data analysis object is the state grid customer service work record data. The analysis targets 27 provinces and all cities in the business scope of the State Grid. The State Grid customer service generated 100,000 to 150,000 pieces of work order data per day, and we randomly selected 4,000 work order service records for data quality analysis. The work order records are classified. The main categories mainly include two categories, namely, the first-level hotspot classification and the second-level hotspot classification. See Table 1 for details. The first-level hotspot classification includes: electricity price, electricity short message, password of service, fast charging network, business hall, electricity fee paying, sensitive event, important customers and major service event. The secondary hotspots are classified in Table 1, appearing in the form of fields, such as the field of errors in reading of electricity bill.

Table 1. Information of data classification

Table Name	Numbering	Field Name	Data Type
electricity price	1	errors in reading of electricity bill	NVARCHAR2
	2	incomplete troubleshooting	NVARCHAR2
	3	long troubleshooting time	NVARCHAR2
	4	frequent power outages	NVARCHAR2

IOP Conf. Series: Materials Science and Engineering **563** (2019) 042001 doi:10.1088/1757-899X/563/4/042001

electricity short message	1	not receiving a power message	NVARCHAR2
	2	electricity message sending error	NVARCHAR2
password of service	1	Pocket Power password query	NVARCHAR2
	2	website password query	NVARCHAR2
	3	electricity password query	NVARCHAR2
fast charging network	1	electric vehicle charging consulting	NVARCHAR2
	2	reservation of charging	NVARCHAR2
business hall	1	business hours	NVARCHAR2
	2	business hall locating	NVARCHAR2
	3	business hall equipment	NVARCHAR2
	4	business staff attitude	NVARCHAR2
	5	business staff service specification	NVARCHAR2
electricity fee paying	1	Alipay payment	NVARCHAR2
	2	bank withholding	NVARCHAR2
	3	electric ebao	NVARCHAR2
	4	Pocket Power	NVARCHAR2
sensitive event	1	obvious fault of power supply equipment	NVARCHAR2
	2	casualty	NVARCHAR2
	3	forced changes of customer purchase methods	NVARCHAR2
	4	longtime abnormal voltage	NVARCHAR2
	5	unfulfilled compensation	NVARCHAR2
important customers	1	social customers	NVARCHAR2
	2	economic customers	NVARCHAR2
	3	high-risk customers	NVARCHAR2
	4	key account contact persons and company leaders at all levels	NVARCHAR2
	5	system leaders	NVARCHAR2
	6	media VIP	NVARCHAR2
	7	love card account	NVARCHAR2

major service event	1	transferred or supervised complaints	NVARCHAR2
	2	media-involved complaints	NVARCHAR2
	3	complaints that will be reported to the media	NVARCHAR2
	4	charge complaints	NVARCHAR2
	5	three special events	NVARCHAR2
	6	report on corruption and bribery	NVARCHAR2
	7	power outages in important places	NVARCHAR2

3. Feature extraction based on LDA topic model

This paper uses the LDA topic model to complete the semantic feature extraction of speech transcribed text. The LDA topic model is a document topic generation model, which is essentially a three-layer Bayesian probability model with three layers of words, topics and documents. The document generation process based on the LDA topic model can be expressed as follows: 1) for each document, extract a topic in the topic distribution; 2) extract a word from the word distribution corresponding to the extracted topic; 3) repeat the above process until a document is generated. The model structure of the LDA topic model is shown in Figure 2.

In Fig. 2, $\vec{\alpha}$ is the parameter of the Dirichlet prior distribution of the topic distribution of each document, $\vec{\beta}$ is the parameter of the Dirichlet prior distribution of the word distribution of each topic, $\vec{\vartheta}_m$ Is the topic distribution of the document m, $\vec{\varphi}_k$ is the word distribution of the topic k, $z_{k,n}$ is the topic of the nth word of the document m, and $w_{m,n}$, the exact word, is the nth word in the document m, K stands for K topics, M stands for M documents, and N_m stands for document m with N_m words.

For document m, its corresponding generation probability can be calculated according to formula (1)

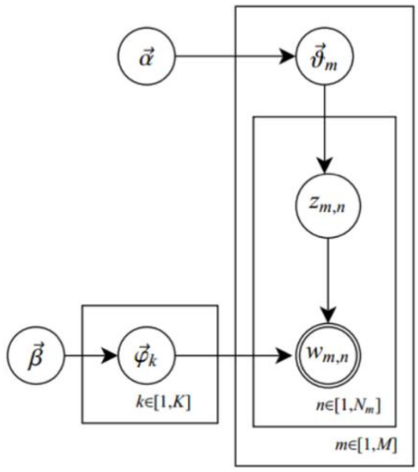

Figure 2. LDA topic model

$$P(\boldsymbol{W},\boldsymbol{Z},\boldsymbol{\vartheta},\boldsymbol{\varphi};\alpha,\beta)= \prod_{k=1}^{K}P(\varphi_k;\beta)\prod_{m=1}^{M}P(\vartheta_m;\alpha)\prod_{n=1}^{N}P(z_{m,n}\mid\vartheta_m)P(w_{m,n}\mid\varphi_{z_{m,n}}) \quad (1)$$

In the formula (1), when the topic of the nth word of the mth document is $z_{m,n}$, the corresponding word distribution is $\varphi_{z_{m,n}}$ The meaning of other parameters is as described above. It can be seen that the LDA topic model is a three-layer probability model. Probability calculation is a traversal summation over all probability spaces.

In this experiment, each category of the work order record naturally forms the topic of a document. When a document is divided into categories, such as "website password query", the semantic information it contains should be more than just a semantic message of the website password. It is possible that these semantic information also includes similar topics, such as "electricity password query" and "Pocket Power password query". One of the semantic information expressed by the vector extracted by the LDA topic model is the distribution of pre-set topics on this document. From a generation perspective, it is also a speculation of the model's distribution of topics in this document. For example, document M is extracted into three topics of A, B, and C, and the corresponding distribution is (0.3, 0.4, 0.3). The meaning expressed can be understood as that document M may contain 0.3 topic A, 0.4 topic B, and 0.3 topic C; It can also be understood as that the probability that the document M is divided into the topic A is 0.3, the probability of being divided into the topic B is 0.4, the probability of being divided into the topic C is 0.3. This experiment sets the number of topics of the LDA topic model to 100. The preprocessed work order record is used as input to the LDA topic model. Through the LDA topic model's document-topic and topic-word generation process, the semantic features in the document can be better extracted.

4. Unsupervised clustering based on K-means

In the process of data analysis and data mining of texts, we usually need to measure the differences and similarities between texts, and group similar texts into one category. Currently commonly used clustering algorithms are: K-means algorithm, DSCAN algorithm and so on.

Suppose there are two text objects, all of which contain N-dimensional features. $D_x = \{w_1^x, w_2^x, \dots, w_n^x\}$ and $D_y = \{w_1^y, w_2^y, \dots, w_n^y\}$, then we calculate the similarity of D_x and D_y. Commonly used similarity calculation methods include Euclidean distance, Manhattan distance, Minkowski distance, cosine similarity, Jaccard Similarity and so on. This work uses cosine similarity to calculate the similarity of documents.

The unsupervised clustering method of K-means is used in the quality analysis of work order records in this paper. The K-means algorithm is a hard clustering algorithm and is a representative of a typical prototype-based objective function clustering method. It uses distance from the data point to the prototype as the objective function of the optimization. The function is used to find the extremum point to get the adjustment rules of the iterative operation. The K-means algorithm takes the Euclidean distance as the similarity measure, which is to find the optimal classification of a certain initial cluster center vector V, so that the evaluation index J is the smallest. The algorithm uses the error squared criterion function as the clustering criterion function.

Its algorithmic idea is relatively simple and easy to implement. The algorithm is as follows:

```
    select K points as the initial centroid
    repeat
        assign each point to the nearest centroid to
form K clusters
        recalculate the centroid of each cluster
    until
    new centroid and original centroid changes less
than the threshold or the number of iterations
reaches the maximum number of iterations
```

In the experiment, we extracted the features extracted from the LDA topic model as input to the K-means model. In the LDA topic model, the number of topics is set to 100, so the dimension of each data point of the K-means model is also 100. The K-means clustering results will serve as the basis for the final text quality assessment.

5. Manual standard of quality assessment

In order to give the text quality assessment the ability to quantify and expand, this paper has developed a document readability quality evaluation standard. The standard was developed by randomly extracting 100 documents and conducting manual review and summary. The final readability evaluation criteria are shown in Table 2.

Table 2. Work order record readability evaluation criteria

Rating level	Rating description
10 points	The documentation is completely correct and readable.
9 points	There may be minor errors, but it does not affect reading, and each sentence can be fully understood.
8 points	There are serious errors in a sentence that need to be understood in combination with the context.
7 points	There are serious errors that prevent a sentence from being understood.
6 points	In addition to the above problems, errors lead to the lack of key information, such as phone, account number and address.
5 points	Through the entire document, you can understand the topic information of the article, such as the first-level and second-level hotspot classification.
4 points	We cannot understand the topic of the article, such as the first-level and second-level hotspot classification.
3 points	In addition to the above problems, most of the content is difficult to understand.
2 points	Only one or two long sentences can be understood.
1 point	The entire document is difficult to understand

As seen from Table 2, the key factor in the development of the evaluation criteria is whether the document can express the central topic, that is, the first-level and second-level hotspot classification described above. This is determined by the core tasks of this paper, that is, the criteria for evaluation are to better classify documents to first-level and second-level hotspot classification.

6. Data quality analysis result

6.1 Quality assessment on data readability

Data readability quality is assessed by using a combination of manual evaluation and topic model.

The overall readability of the work order data is good, The number of documents reaching 5 points (Through the entire document, you can understand the topic information of the article, such as the first-level and second-level hotspot classification) or above takes up 83% of the sample from the result of manual sampling. So the distribution of words under the topic has a low rate of garbled and transcribed text errors. We select documents with more than 5 points to calculate the similarity of the document with random topic distribution. The result is shown in Figure 3.

In addition, the text features extracted by the LDA topic model are used for text similarity comparison. The similarity between texts under different scores is shown in Table 3.

It can be seen from Table 3 that the similarity between the low score documents is low, and the similarity between the high score documents is high. First, due to the difference in the topic distribution of documents with different score, the similarity between the high score document and the low score document has a proportional relationship, that is, the higher the score is, the higher the similarity is. Secondly, by manually checking high-quality documents and comparing the characteristics of the topic distribution, it is found that the information contained in the topic is richer, and the similarity with other same score document topics is higher. At the same time, compared with the low-quality documents, because of transcoding errors and the low information entropy, the similarity with the same low score document is low, and the readability is poor.

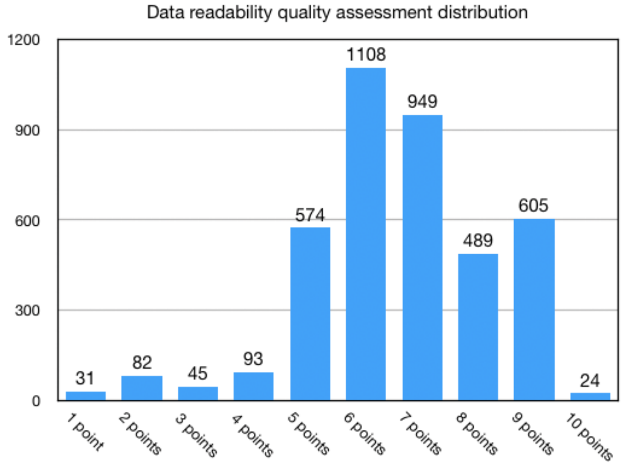

Figure 3. Data readability quality assessment distribution

Table 3. Text quality similarity comparison

Similarity comparison	5 points	6 points	7 points	8 points	9 points	10 points
5 points	0.7723	0.7768	0.7392	0.7567	0.7352	0.7233
6 points	0.7768	0.8569	0.8105	0.8732	0.8582	0.8598
7 points	0.7392	0.8105	0.7851	0.8254	0.8139	0.8013
8 points	0.7567	0.8732	0.8254	0.9467	0.9298	0.9341
9 points	0.7352	0.8582	0.8139	0.9298	0.9342	0.9383
10 points	0.7233	0.8598	0.8013	0.9341	0.9383	0.9432

Therefore, in the automatic evaluation of readability, a part of the document is sampled, and the vector center corresponding to the document with the score of 8, 9, and 10 points is calculated. For the new document to be evaluated, the K-means algorithm is first used for cluster analysis. If the new document is classified into the vector center corresponding to the 8, 9, 10-point document, and then the similarity between the vector and the vector center is calculated to be greater than a threshold (for example, 0.9), it can be judged as high-quality text.

6.2 Type assessment of major service event

The keyword search was used to perform the following keyword searches for documents and different types of major service event. The specific keywords are shown in Table 4, and the corresponding search results are shown in Figure 4. It can be seen from Figure 4 that significant service events have considerable sparsity and most cannot be discriminated based on keywords. In the 4,000 sampled documents, only 6 documents were classified as major service event types based on keywords for a given major service event, concentrating in the category of faults due to natural disasters, accounting for only 1.5‰. The proportion of major service events in the total number of services basically meets the service quality requirements of State Grid customer service.

Table 4. Example of major service event keyword list

event type of major service	Keyword
transferred or supervised complaints	none

Media-involved complaints	Weibo, public account, television, media, newspapers, press
complaints that will be reported to the media	ibid.
charge complaints	indiscriminate charging
three special events	none
report on corruption and bribery	bribery, arbitrary charges, corruption, collecting money, taking money, accusation
power outages in important places	railway station, airport, railway, plane, station
Collective complaints of more than 10 people	many, many people, many of us
Faults due to natural disasters	blizzard, typhoon, fire, hail, flash floods, floods, mudslides, earthquakes
Casualty	injury

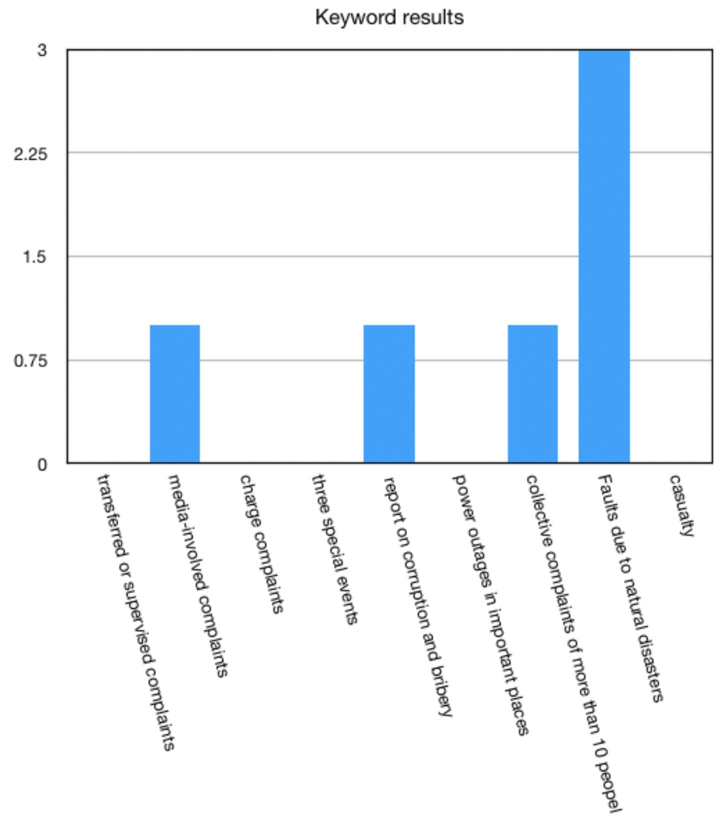

Figure 4. Key service query for keyword query statistics

7. Conclusion

The main problem with the quality of the data from the above analysis is the problem of data sparsity and readability of text quality. The summary of assessment is as follows:

1) There are problems with the quality results of the voice transcription text of the work order, mainly including

a) A small number of documents are basically difficult to understand, accounting for about 6.3% of the sampled data;

b) Most of the documents are at the level of understandable topic and customer appeal, accounting for 78% of the sampled data;

c) A few documents are very readable, accounting for 15.7% of the sampled data.

2) After the data quality assessment of the first-level hotspot classification and the second-level hotspot classification which is the customer's main concern, it can be concluded that there is a large data sparsity. Through the results of LDA clustering and manual labeling, it can be seen that the distribution of documents in the first- level hotspot classification and the second-level hotspot classification is quite uneven, and more documents cannot be covered by the current label, and are classified into other categories.

3) After data quality assessment for major service events that are of primary concern to customer service, it can be concluded that data sparsity is more serious. Through keyword filtering, you can see that there are very few documents related to major service events in the entire document, and based on keyword filtering, some major service events are not appearing in the document.

For the work order record data with poor quality, we recommend to deal with as follows:

1) For sparse data, it may be necessary for the data side to provide more data samples for verification or to manually create some data samples.

2) For poor quality data, it can be manually corrected according to the proportion.

This paper analyzes the problem of text classification of work orders of grid companies and evaluates the current work order data. Our work is fundamental to the application of natural language processing techniques to process work order data.

Acknowledgment

The work was supported by State Grid Corporation of China Headquarters Project (SGGR0000JSJS1701223) which is 'Application Research of New Generation Artificial Intelligence Technology in Grid Customer Service'.

References

[1] Zou Y F, He W, Zhao H Y, Cheng Y M and Yang H 2016 Application of text mining technology in electric power work order data analysis *Modern Electronics Technique* **39(17)** pp 149-152

[2] Wang Y Z, Guo Q H, Liu Y F, Liu Z and Liu J 2017 Analysis and Suggestions on 95598 Complaint Work Order Data *JiangXi Electric Power* **41(10)** pp 2-4

[3] Zhou Y Satisfaction Research of Grid customer service 2017 *Tianjin University of Finance and Economics*

[4] Zu X and Xie F 2015 Review of the Latent Dirichlet Allocation Topic Model 2015 *Journal of Hefei Normal University* **33(06)** pp 55-58, 61

[5] Hu J M and Chen G 2014 Mining and Evolution of Content Topics Based on Dynamic LDA Library and Information Service **58(2)** pp 138-142

[6] Wang Z Z, He M and Du Y P 2013 Text Similarity Computing Based on Topic Model LDA Computer Science **40(12)** pp 229-232

[7] Zhang L 2016 Research on Tagging Recommendation Method Based on LDA Topic Model Modern Information **36(02)** pp 53-56