# Research on the Classification Modeling for the Natural Language Texts with Subjectivity Characteristic

Chen Xiao Yu[1], Song Ying[3]
Computer School
Beijing Information Science & Technology University
Beijing, China

Gao Feng[2], Zhang Xiao Min[4]
Academy of Agricultural Planning and Engineering
Ministry of Agriculture and Rural Affairs
Beijing, China

***Abstract*—The methods of natural language text classification have the characteristic of diversification, and the text characteristics are the basis of the method effectiveness; this paper takes the car service complaint data as an example to study the classification modeling for the texts with subjectivity characteristic. The effective handling of car service complaints is important for improving user experience and maintaining brand reputation; manual classification commonly has the disadvantages of experience dependence, prone to error, heavy workload and so on; corresponding automatic classification modeling research is of great practical significance. The core links of the research method in this study include word segmentation, text vectorization, feature selection and dimensionality reduction based on correlation, classification modeling based on progressive method and random forest, and model reliability analysis; the research results show that the car service complaint texts could be effectively classified based on the method in this study, which could provide a reference for related further research and application.**

***Keywords*—Car service complaint; text classification; machine learning; natural language texts***

## I. INTRODUCTION

Car service complaints widely exist in the use of cars of different types, brands, and use times. Effective classification of service complaint texts is an important part of the basis for the efficient and reasonable handling of corresponding complaints. The classification of service complaints could be done by users or receiving staff, if this work is handed over to the users, on one hand, it might increase the user's irritability and dissatisfaction, and on the other hand, the users possibly make mistakes because they might don't know much about the professional field; if the complaint receiving staff conducts manual classification, there are also problems such as heavy workload, experience dependence, and error-prone; therefore, it is of great practical value to carry out the automatic classification modeling and realize automatic classification of service complaint.

Machine learning methods are widely used in data processing such as classification and regression, and they are also quite widely used in automatic classification of natural language texts [1-4]. The text classification modeling methods based on machine learning have the characteristics of diversification and different applicability. In recent years, related researchers have carried out a lot of studies on different types of texts, which provides a well foundation for follow-up text classification modeling research. In general, it has the necessary theoretical basis and important practical significance to carry out automatic classification modeling for car service complaint texts so as to realize automatic classification of the complaints, through reasonably using machine learning methods based on the characteristics of the complaint texts.

Chinese text classification modeling based on machine learning commonly involves data preprocessing, text vectorization, classification modeling, and model reliability analysis, among which, text vectorization and classification modeling are commonly the core links. In related research, the methods used in text vectorization mainly involve three categories, including word frequency-based methods such as TF-IDF and Bag-of-words [5-7], distributed static word vector-based methods such as Word2vec [8-11], distributed dynamic word vector-based methods such as BERT [12-13]. Different types of methods commonly have different characteristics and applicable scenarios; the methods based on word frequency is relatively simple in principle and convenient to implement, but text vectorization based on this kind of methods would lose context information; the distributed static word vector methods and the distributed dynamic word vector methods both understand and represent words based on context, and could effectively retain context information in the process of text vectorization; the main difference between the static word vector methods and the dynamic word vector methods is the polysemy distinction of a word, the static word vector methods couldn't distinguish the different interpretations of a word in different contexts, while the dynamic word vector methods have the ability to distinguish. In text classification modeling, after the texts are converted to vectors, it is commonly necessary to furtherly process classification information so as to obtain text classification prediction results; in related researches, the basic methods used in classification information processing mainly involve two categories, including classical machine learning methods [14], such as Naive Bayes and random forests, and various neural network methods [15-19]. The neural network methods used in text classification have the characteristic of diversification, including basic neural network methods such as CNN, RNN,

Funding Project: Promote the Classification Development of College-Construction of Professional Degree Sites of Electronic Information (Intelligent Computing) (5112211039).

LSTM, BILSTM, and various improved neural network methods suitable for different application scenarios. At the same time, the neural network method has the characteristic of flexible superposition, different kinds of neural networks could be stacked up for use. Applying the methods based on different types of neural networks stacking to the classification information processing, the effect of multi-layer classification information processing could be achieved, which commonly could improve the quality of classification information extraction and the classification accuracy of the model. In addition, ensemble learning methods are gradually being used in natural language text classification research in recent years [20]. Ensemble learning methods could comprehensively use a variety of basic methods, which could firstly build multiple individual models based on different kinds of basic methods, and then perform horizontal integration of basic models through certain strategies, obtaining prediction results through two-stage process processing. In general, the natural language text classification modeling based on machine learning has the characteristics of multi-link research process and diversified research methods; different types of basic methods commonly have different characteristics and applicability differences; in text classification modeling, technical routes and technical methods could be designed and selected based on actual text characteristics.

The basic characteristics of car service complaint texts include quite rich emotional expressions and professional vocabularies. After data preprocessing and basic analysis, based on the characteristics of the service complaint texts, this study uses the Jieba word segmentation tool for word segmentation, word frequency-based method for text vectorization, and correlation method for feature selection; the classification model is trained based on progressive training method and random forest, and the feature dimension is adjusted through the modeling feedback; model reliability is assessed based on the effect of data amount on the modeling and the effect of text length on the probability distribution of classification predictions. It is expected that this study could provide effective reference for subsequent related research and application.

## II. TECHNICAL ROUTE

The technical route of this study includes data sorting, data characteristic analysis, splitting words, feature extraction, classification modeling, model reliability analysis, conclusion and outlook. In feature extraction, the method based on word frequency and correlation analysis is used; in classification modeling, random forest and progressive method are used to construct classification model.

The technical route is shown as Fig. 1.

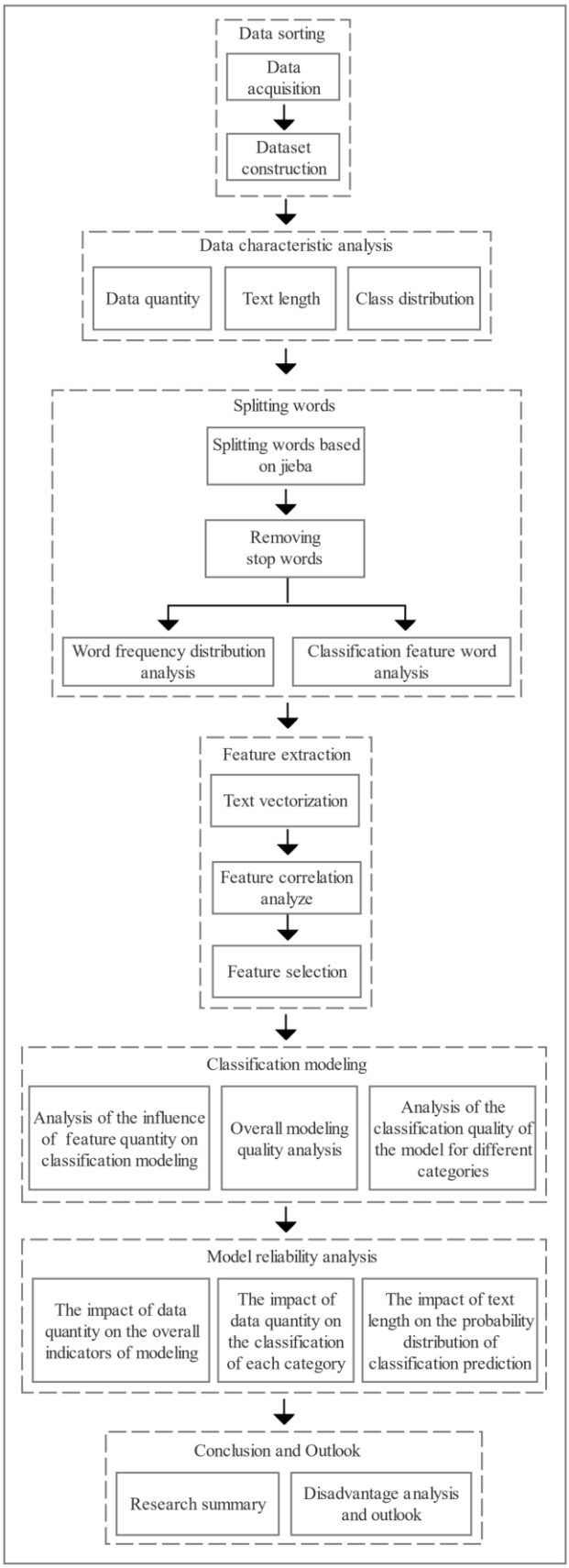


Fig. 1. Technical Route Figure.

## III. DATA

The research data of this study is from the Beijing Car Quality Net Information Technology Limited Company. The dataset used includes seven classes of car service complaint text data, which includes service attitude, personnel technology, service charge, not keeping promise, sale fraud, accessory dispute, service process is not perfect. The total data amount of the dataset is 2100, and the data amount of each class is 300.

The characteristics of data amount and text length of the dataset are shown in Table I.

The data distribution of the dataset in terms of car type, purchase and use time, and brand is shown in Fig. 2. The coverage of the dataset in terms of car type, purchase and use time, and brand is relatively comprehensive, and the overall distribution uniformity is quite well.

TABLE I. DATA DESCRIPTION

| No. | Class | Data amount | Average length of text | Maximum length of text | Minimum length of text | Text length standard deviation |
|---|---|---|---|---|---|---|
| 1 | Service attitude | 300 | 18.5033 | 24 | 15 | 1.8352 |
| 2 | Personnel technology | 300 | 19.1167 | 26 | 15 | 1.9806 |
| 3 | Service charge | 300 | 18.6133 | 24 | 14 | 1.9156 |
| 4 | Not keeping promise | 300 | 18.5500 | 24 | 13 | 1.8670 |
| 5 | Sale fraud | 300 | 18.7100 | 29 | 14 | 2.2065 |
| 6 | Accessory dispute | 300 | 18.5267 | 30 | 15 | 1.9618 |
| 7 | Service process is not perfect | 300 | 18.7033 | 26 | 14 | 2.2193 |
| 8 | All | 2100 | 18.6748 | 30 | 13 | 2.0098 |

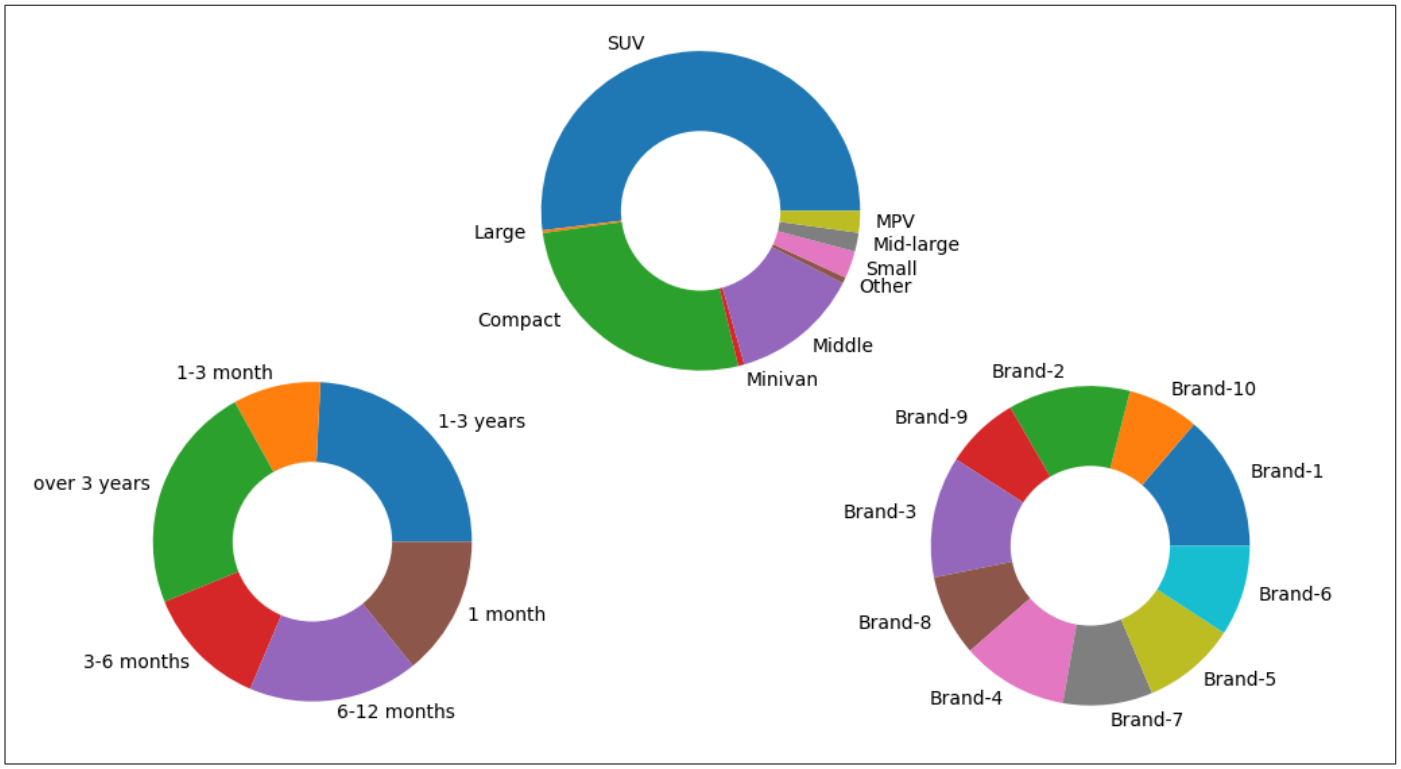


Fig. 2. Data Distribution Characteristics.

## IV. Word Segmentation

Chinese text classification commonly needs to separate the texts into words, in this study, Jieba tool is used for word segmentation, which is widely used in Chinese text segmentation. After word segmentation, the data process of removing stop words is conducted to remove the feature words with low classification relevance. After removing stop words, the word segmentation result is shown in Table II. In the different classes, the highest word amount is 2865 and the lowest amount is 2611; the highest unique word amount is 743 and the lowest amount is 516; the highest value of repetition rate is 0.8129 and the lowest value is 0.7154.

The distribution differences of the global high frequency words in different classes could reflect the applicability of the data processing method based on word frequency in the classification modeling for the corresponding text data to a large extent. The analyze result for the word frequency distribution of the global high frequency words of the dataset in the different classed is shown in Fig. 3. Overall, the higher the global word frequency value, the higher the difference of the word frequency distribution of the feature words in the different classes; the distribution of the global high frequency feature words of the dataset in different classes shows a high degree of discrimination as a whole, which shows that the data processing method based on word frequency is suitable for the classification modeling scene targeted in this study. The class high frequency feature words are shown in Table III.

TABLE II. Word Segmentation Results

| No. | Class | Number of characters | Number of separated words | Number of unique words | Repetition rate |
|---|---|---|---|---|---|
| 1 | Service attitude | 5551 | 2611 | 743 | 0.7154 |
| 2 | Personnel technology | 5735 | 2788 | 719 | 0.7421 |
| 3 | Service charge | 5584 | 2682 | 618 | 0.7696 |
| 4 | Not keeping promise | 5565 | 2758 | 516 | 0.8129 |
| 5 | Sale fraud | 5613 | 2865 | 558 | 0.8052 |
| 6 | Accessory dispute | 5558 | 2802 | 620 | 0.7787 |
| 7 | Service process is not perfect | 5611 | 2731 | 544 | 0.8008 |
| 8 | All | 39217 | 19237 | 2092 | 0.8913 |

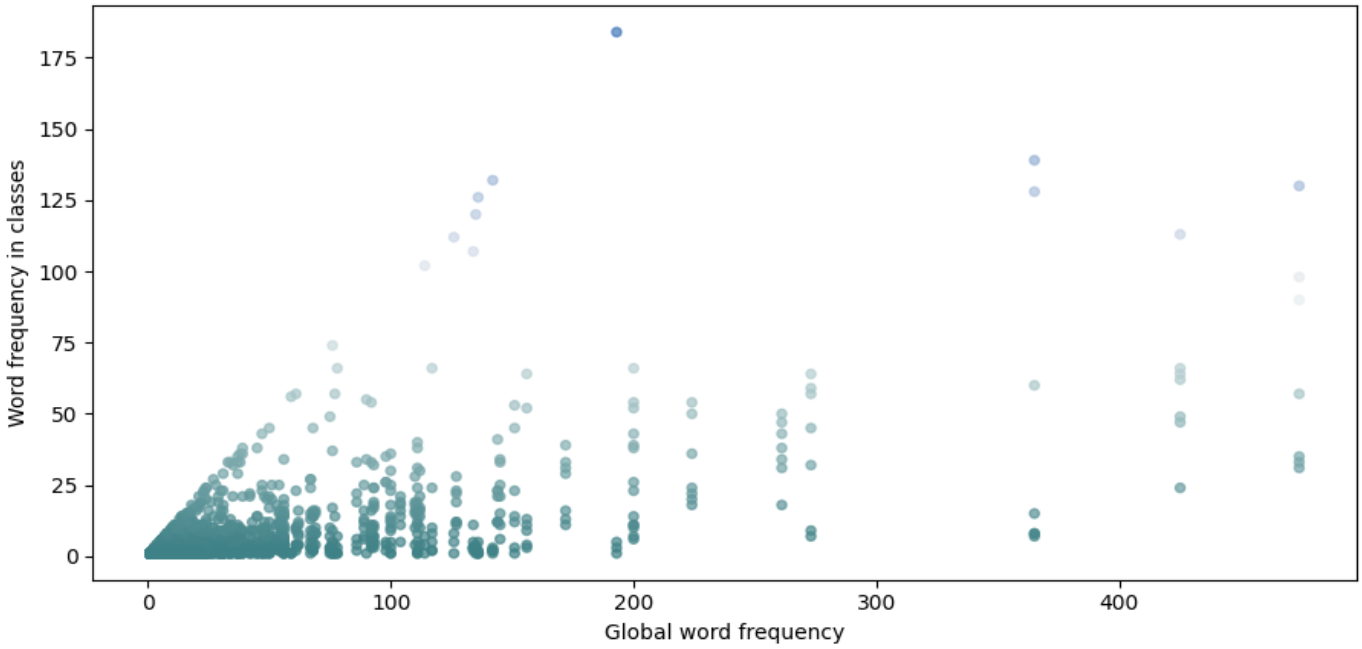


Fig. 3. Word Frequency Distribution.

TABLE III. CLASS HIGH FREQUENCY FEATURE WORDS

| **No.** | Service attitude | Personnel technology | Service charge | Not keeping promise | Sale fraud | Accessory dispute | Service process is not perfect |
|---|---|---|---|---|---|---|---|
| 1 | 4S shop | Repair | Dealer | Fulfill | Publicity | Accessory | Upgrade |
| 2 | Dealer | 4S shop | Return | Promise | Inconsistent | Repair | Machine system |
| 3 | Repair | Dealer | 4S shop | Dealer | Sell | Replace | Car |
| 4 | Factory | Maintenance | Car purchase | Not receive | Dealer | Factory | Car machine |
| 5 | * | Accident | Charge | Maintenance | Factory | 4S shop | * |
| 6 | Not | * | Deposit | Car purchase | Car | Dealer | * |
| 7 | Vehicle | * | * | Not yet | * | Original factory | * |
| 8 | After sale | Cause | Charge | * | Function | Accident | Factory |
| 9 | * | Vehicle | Not yet | Car | * | * | Not yet |
| 10 | Bad | * | Deposit | Subsidy | * | Not | Update |
| 11 | Not yet | Damage | * | 4S shop | * | Cause | 4S shop |
| 12 | Attitude | Change | * | Delivering car | Sale | * | Version |
| 13 | When | Bad | Sale | Deposit | Suspected | Vehicle | Solve |
| 14 | Solve problem | * | * | * | * | * | * |
| 15 | Accident | Specification | When | Replace | 4S shop | * | Provide |
| 16 | Delay | * | Maintenance | Order car | * | Accessory | * |
| 17 | * | Engine | Purchase | When | Light | After sale | Dealer |
| 18 | * | Not yet | * | Factory | * | Quality | Navigation |
| 19 | Fault | Personnel | Pay | Presentation | Parking | Damage | * |
| 20 | * | Record | * | * | New car | * | * |

## V. FEATURE EXTRACTION AND CLASSIFICATION MODELING

Based on basic analysis for the data characteristics, this study adopts the bag-of-words method based on word frequency for text vectorization, and operates feature word correlation analysis for the feature selection and word vector dimensionality reduction.

Based on correlation analysis, feature words with a degree of correlation higher than 0.95 are partly removed to reduce the dimension of the word vectors and improve the efficiency of model training, model classification. Through the feature selection, 257 features with high relevance are removed, and 1689 features are reserved. Fig. 4 shows the correlation matrix of part global high frequency features in the dataset.

In this study, the random forest combined with progressive model training method is used for the classification model training. In the model training, the features involved are gradually increased for obtaining the optimal amount of the modeling features by comparing the results of multiple rounds of training, in order to avoid too little features used do not contain enough classification information for supporting high quality classification modeling, or the model prediction accuracy, model classification efficiency are negatively affected due to too much features used. The research results show that in this study, the optimal modeling effect is obtained when using 1689 features.

Table IV shows the comparison of the progressive multi-round training results from the aspects of feature selection ratio, feature amount, overall accuracy, overall precision, overall recall, overall f1-score, highest f1-score in the classes, and lowest f1-score in the classes.

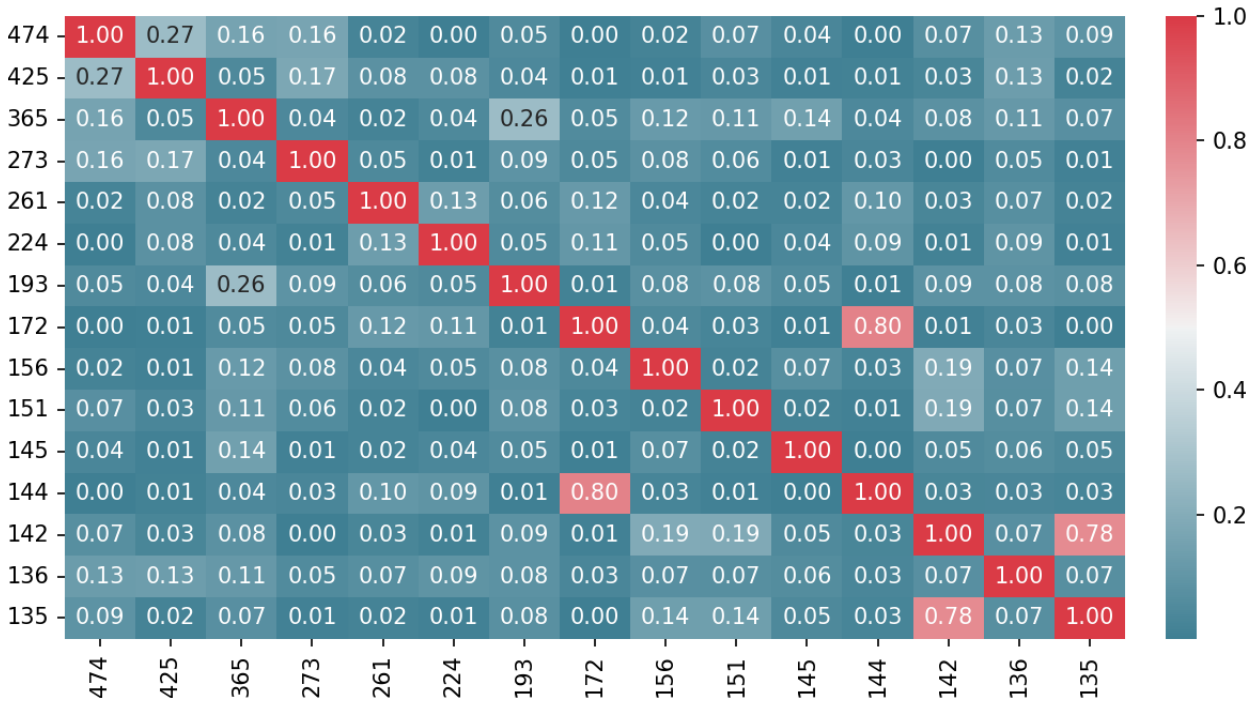


Fig. 4. Correlation Matrix Heatmap.

TABLE IV. CLASSIFICATION MODEL TRAINING

| No. | Feature selection ratio | Feature amount | Accuracy | Precision | Recall | F1-score | Highest f1-score in the classes | Lowest f1-score in the classes |
|---|---|---|---|---|---|---|---|---|
| 1 | 10% | 169 | 0.8000 | 0.8041 | 0.8000 | 0.8008 | 0.8966 | 0.7119 |
| 2 | 20% | 338 | 0.8000 | 0.8041 | 0.8000 | 0.8008 | 0.8966 | 0.7119 |
| 3 | 30% | 507 | 0.8000 | 0.8041 | 0.8000 | 0.8008 | 0.8966 | 0.7119 |
| 4 | 40% | 676 | 0.8000 | 0.8041 | 0.8000 | 0.8008 | 0.8966 | 0.7119 |
| 5 | 50% | 845 | 0.8000 | 0.8041 | 0.8000 | 0.8008 | 0.8966 | 0.7119 |
| 6 | 55% | 929 | 0.8333 | 0.8423 | 0.8333 | 0.8358 | 0.8966 | 0.7188 |
| 7 | 60% | 1013 | 0.8333 | 0.8418 | 0.8333 | 0.8352 | 0.8814 | 0.7213 |
| 8 | 65% | 1098 | 0.8333 | 0.8482 | 0.8333 | 0.8368 | 0.9123 | 0.7188 |
| 9 | 70% | 1182 | 0.8333 | 0.8404 | 0.8333 | 0.8346 | 0.8852 | 0.7213 |
| 10 | 75% | 1267 | 0.8286 | 0.8336 | 0.8286 | 0.8298 | 0.8929 | 0.7097 |
| 11 | 80% | 1351 | 0.8381 | 0.8438 | 0.8381 | 0.8396 | 0.9180 | 0.7302 |
| 12 | 85% | 1436 | 0.8238 | 0.8274 | 0.8238 | 0.8249 | 0.8667 | 0.7619 |
| 13 | 90% | 1520 | 0.8286 | 0.8326 | 0.8286 | 0.8296 | 0.8772 | 0.7213 |
| 14 | 95% | 1605 | 0.8333 | 0.8436 | 0.8333 | 0.8359 | 0.9153 | 0.7419 |
| 15 | 100% | 1689 | 0.8476 | 0.8550 | 0.8476 | 0.8495 | 0.9123 | 0.7619 |

## VI. MODEL RELIABILITY ANALYSIS

Model reliability analysis is of great significance to the classification modeling, since the modeling is commonly based on limited data, and the modeling indexes mainly reflect the overall quality of the model. The applicability of the model to new data and the reliability of the model classification under specific condition are commonly difficult to evaluate based on the overall modeling result indexes.

This study evaluates the reliability of the model from two aspects: the effect of data amount on the modeling and the effect of text length on the model prediction probability distribution. In term of the effect of data amount, Table V shows the comparison of the multi-round model training result parameters under the condition of increasing data amount. In term of the effect of text length, the first 8 and last 8 text data in the global text length rank are selected for model classification prediction probability distribution analysis. The model classification prediction probability distribution is shown in Table VI.

TABLE V. THE EFFECT OF DATA AMOUNT ON MODEL TRAINING

| No. | Data use ratio | Data amount | Accuracy | Precision | Recall | F1-score | Highest f1-score in the classes | Lowest f1-score in the classes |
|---|---|---|---|---|---|---|---|---|
| 1 | 10% | 210 | 0.7143 | 0.7381 | 0.7143 | 0.7197 | 1.0000 | 0.3333 |
| 2 | 20% | 420 | 0.7143 | 0.7381 | 0.7143 | 0.7197 | 1.0000 | 0.3333 |
| 3 | 30% | 630 | 0.7143 | 0.7381 | 0.7143 | 0.7197 | 1.0000 | 0.3333 |
| 4 | 40% | 840 | 0.7143 | 0.7381 | 0.7143 | 0.7197 | 1.0000 | 0.3333 |
| 5 | 50% | 1050 | 0.7143 | 0.7381 | 0.7143 | 0.7197 | 1.0000 | 0.3333 |
| 6 | 55% | 1155 | 0.7845 | 0.7953 | 0.7815 | 0.7829 | 0.9714 | 0.5625 |
| 7 | 60% | 1260 | 0.7381 | 0.7475 | 0.7381 | 0.7398 | 0.9143 | 0.6061 |
| 8 | 65% | 1365 | 0.8175 | 0.8357 | 0.8169 | 0.8199 | 0.9189 | 0.6977 |
| 9 | 70% | 1470 | 0.8571 | 0.8678 | 0.8571 | 0.8580 | 0.9767 | 0.7755 |
| 10 | 75% | 1575 | 0.8165 | 0.8243 | 0.8156 | 0.8171 | 0.8980 | 0.6667 |
| 11 | 80% | 1680 | 0.7917 | 0.7928 | 0.7917 | 0.7908 | 0.8571 | 0.6939 |
| 12 | 85% | 1785 | 0.8715 | 0.8777 | 0.8716 | 0.8714 | 0.9630 | 0.8077 |
| 13 | 90% | 1890 | 0.8466 | 0.8510 | 0.8466 | 0.8466 | 0.9310 | 0.7241 |
| 14 | 95% | 1995 | 0.8100 | 0.8220 | 0.8103 | 0.8126 | 0.8772 | 0.6667 |
| 15 | 100% | 2100 | 0.8476 | 0.8550 | 0.8476 | 0.8495 | 0.9123 | 0.7619 |

TABLE VI. THE PREDICTION PROBABILITY DISTRIBUTION FOR THE FIRST 8 AND LAST 8 TEXTS IN THE GLOBAL TEXT LENGTH RANK

| No. | Service attitude | Personnel technology | Service charge | Not keeping promise | Sale fraud | Accessory dispute | Service process is not perfect |
|---|---|---|---|---|---|---|---|
| F-1 | 0.0700 | 0.1300 | 0.0300 | 0.0000 | 0.0000 | 0.7200 | 0.0500 |
| F-2 | 0.1000 | 0.3700 | 0.0500 | 0.1000 | 0.1450 | 0.1000 | 0.1350 |
| F-3 | 0.0500 | 0.9100 | 0.0000 | 0.0000 | 0.0200 | 0.0000 | 0.0200 |
| F-4 | 0.1000 | 0.6800 | 0.0000 | 0.0000 | 0.1000 | 0.0800 | 0.0400 |
| F-5 | 0.0100 | 0.0500 | 0.1700 | 0.0000 | 0.0700 | 0.0100 | 0.6900 |
| F-6 | 0.0100 | 0.0500 | 0.1700 | 0.0000 | 0.0700 | 0.0100 | 0.6900 |
| F-7 | 0.0700 | 0.8000 | 0.0800 | 0.0100 | 0.0000 | 0.0100 | 0.0300 |
| F-8 | 0.7500 | 0.0800 | 0.0100 | 0.0200 | 0.0000 | 0.0400 | 0.1000 |
| L-1 | 0.0100 | 0.0400 | 0.0000 | 0.9300 | 0.0000 | 0.0100 | 0.0100 |
| L-2 | 0.0300 | 0.0600 | 0.7100 | 0.1300 | 0.0500 | 0.0100 | 0.0100 |
| L-3 | 0.0000 | 0.0200 | 0.9600 | 0.0000 | 0.0100 | 0.0100 | 0.0000 |
| L-4 | 0.0000 | 0.0200 | 0.9600 | 0.0000 | 0.0100 | 0.0100 | 0.0000 |
| L-5 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.9900 | 0.0100 | 0.0000 |
| L-6 | 0.0000 | 0.0100 | 0.0000 | 0.0000 | 0.0300 | 0.0100 | 0.9500 |
| L-7 | 0.0300 | 0.0400 | 0.0000 | 0.0000 | 0.0200 | 0.0300 | 0.8800 |
| L-8 | 0.9700 | 0.0000 | 0.0000 | 0.0000 | 0.0100 | 0.0200 | 0.0000 |

## VII. CONCLUSION AND OUTLOOK

This study focus on the issue of car service complaint classification modeling, the dataset used involves 7 classed of service complaint texts, and the data amount of every class is all 300; the core links of the research process include word segmentation, text vectorization, the feature selection and dimensionality reduction based on correlation analysis, the classification modeling based on random forest and progressive method, and the model reliability analysis based on data amount and text length; the results show that based on the method of this study, the effective classification for the car service complaint texts could be realized. In this study, when the feature amount reaches 1689, the optimal modeling effect is obtained, the values of overall accuracy, overall precision, overall recall, overall f1-score, the highest f1-score in the classes, the lowest f1-score in the classes respectively reach 0.8476, 0.8550, 0.8476, 0.8495, 0.9123 and 0.7619; in the model reliability analysis, when the data use ratio reaches 85%, the modeling effect is generally stable; the analysis results of

the effect of text length on the model classification prediction probability distribution show that the distribution is overall high discriminative.

In summary, the car service complaint data could be effectively classified based on the method of this study, which could provide reference for the classification modeling of the natural language texts with subjectivity characteristic; at the same time, this study still belongs to theoretical research, and has not been applied to practice.

## REFERENCES

[1] Zhu Fang Peng, Wang Xiao Feng, Text classification for ship industry news [J], Journal of Electronic Measurement and Instrumentation, 2020, 34 (01): 149-155.

[2] Zhao Ming, Du Hui Fang, Dong Cui Cui, Chen Chang Song, Diet health text classification based on word2vec and LSTM [J], Transactions of the Chinese Society for Agricultural Machinery, 2017, 48 (10): 202-208.

[3] Bao Xiang, Liu Gui Feng, Yang Guo Li, Patent text classification method based on multi-instance Learning [J], Information Studies: Theory & Application, 2018, 41 (11): 144-148.

[4] Wen Chao Dong, Zeng Cheng, Ren Jun Wei, Zhang Yan, Patent text classification based on ALBERT and bidirectional gated recurrent unit [J], Journal of Computer Applications, 2021, 41 (02): 407-412.

[5] Hu Jing, Liu Wei, Ma Kai, Text categorization of hypertension medical records based on machine learning [J]. Science Technology and Engineering, 2019, 19 (33): 296-301.

[6] Yu Hang, Li Hong Lian, Lü Xue Qiang, Text classification of NPC report contents [J], Computer Engineering and Design, 2021, 42 (06): 1772-1778.

[7] Wang Xiang Xiang, Fang Hui, Chen Chong Cheng, Classification technique of cultural tourism text based on naive Bayes [J]. Journal of Fuzhou University (Natural Science Edition), 2018, 46 (05): 644-649.

[8] Zhou Qing Hua, Li Xiao Li, Research on short text classification method of railway signal equipment fault based on MCNN [J]. Journal of Railway Science and Engineering, 2019, 16 (11): 2859-2865.

[9] Feng Shuai, Xu Tong Yu, Zhou Yun Cheng, Zhao Dong Xue, Jin Ning, et al. Rice knowledge text classification based on deep convolution neural network [J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52 (03): 257-264.

[10] Niu Zhen Dong, Shi Peng Fei, Zhu Yi Fan, Zhang Si Fan, Research on classification of commodity ultra-short text based on deep random forest [J]. Transactions of Beijing Institute of Technology, 2021, 41 (12): 1277-1285.

[11] Zhang Yu, Liu Kai Feng, Zhang Quan Xin, Wang Yan Ge, Gao Kai Long, A combined-convolutional neural network for Chinese news text classification [J]. Acta Electronica Sinica, 2021, 49 (06): 1059-1067.

[12] Li Ke Yue, Chen Yi, Niu Shao Zhang, Social E-commerce text classification algorithm based on BERT [J], Computer Science, 2021, 48 (02): 87-92.

[13] Tian Yuan, Yuan Ye, Liu Hai Bin, Man Zhi Bo, Mao Cun Li, BERT pre-trained language model for defective text classification of power grid equipment [J]. Journal of Nanjing University of Science and Technology, 2020, 44 (04): 446-453.

[14] Zhao Yan, Li Xiao Hui, Zhou Yun Cheng, Zhang Yue. A study on agricultural text classification method based on naive bayesian [J]. Water Saving Irrigation,2018(02):98-102.

[15] Chen Ping, Kuang Yao, Hu Jing Yi, Wang Xiang yang, Cai Jing. Text categorization method with enhanced domain features in power audit field [J]. Journal of Computer Applications, 2020, 40 (S1): 109-112.

[16] Liu Zi Quan, Wang Hui Fang, Cao Jing, Qiu Jian, A classification model of power equipment defect texts based on convolutional neural network [J]. Power System Technology, 2018, 42 (02): 644-651.

[17] Ge Xiao Wei, Li Kai Xia, Chen Ming, Text classification of nursing adverse events based on CNN-SVM [J]. Computer Engineering & Science, 2020, 42 (01): 161-166.

[18] Wang Meng Xuan, Zhang Sheng, Wang Yue, Lei Ting, Du Wen, Research and application of improved CRNN model in classification of alarm texts [J]. Journal of Applied Sciences, 2020, 38 (03): 388-400.

[19] Wang Si Di, Hu Guang Wei, Yang Si Yu, Shi Yun, Automatic transferring government website e-mails based on text classification [J]. Data Analysis and Knowledge Discovery, 2020, 4 (06): 51-59.

[20] Zhang Bo, Sun Yi, Li Meng Ying, Zheng Fu Qi, Zhang Yi Jia, et al. Medical text classification based on transfer learning and deep learning [J]. Journal of Shanxi University (Natural Science Edition), 2020, 43 (04): 947-954.