Design of Intelligent Customer Service Report System Based on Automatic Speech Recognition and Text Classification

*Yunlong Zou**, *Xiangyu Liu*, *Hongyan Xu*, *Yingzhe Hou* and *Jialiang Qi*

Northeast Normal University, Changchun, China

Abstract. In combination with features such as intensive labor and speech in the customer service report field, this paper discusses the design of a customer service report system based on artificial intelligence automatic speech recognition technology and big data text classification technology. The proposed system realizes functions like a flat IVR menu, quick transcription and input of work orders, dynamic tracking of failure hotspots, automatic classification and accumulation of the knowledge base, speech emotion detection and real-time supervision of service quality, and it can improve the user experience and reduce the labor strengths of customer service staff. The automatically accumulated knowledge base can further assist with feedback to resolve the difficult problem that the emerging intelligent network Q&A and intelligent robots rely on a manually summarized knowledge base.

1 Introduction

As an important interface for direct user experience feedback and customer interaction, the customer service report system is widely used by different service agencies, and it is significant for the improvement of service quality and user satisfaction, summarization of user demands, and the creation of service improvement points and value-added points [1].

Customer service report reception is a labor-intensive industry, in which a lot of simple and repeated efforts are required. With the development of services, customer service staff have been required to provide an increasing amount of services. Concerning a variety of problems, such as increased labor strength, higher pressure on emotion control [2], higher agency training costs, and the more complicated report system, the traditional service methods and management modes shall be transformed. Moreover, as a main portal for customer feedback, the customer service report is also a speech-intensive industry, especially for a telephone-based customer service system using a call center as the representative, and the massive data cannot be processed effectively by using traditional methods of data collection, arrangement, and analysis, such as the manual input of work orders and the manual sorting of the knowledge base. There are problems such as low data collection efficiency, partial information extraction, lagging fault tracking, difficulty in monitoring service quality, and slow update of the knowledge base. For the labor-intensive

* Corresponding author e-mail: zouyl@nenu.edu.cn

and speech-intensive features, it is suitable to use technologies like automatic speech recognition (ASR) technology in the artificial intelligence field and text classification technology in the big data field to perform optimization and to realize reasonable format transformation of speech data and reprocessing, thereby greatly improving data utilization; bringing advantages such as quick input of work orders, dynamic tracking of failure hotspots, efficient supervision of service quality, and automatic classification of knowledge base; and laying the foundation for additional functions like automatic recommendations for answers, multi-language translation, failure trend statistics, and service value-added point analysis.

2 Significance and application of automatic speech recognition and text classification

Automatic speech recognition can convert the speech of a person to text and is one of the key technologies for human-machine interaction in the field of artificial intelligence (AI) [3]. This technology aims to convert speech signals to computer-readable input, such as keystrokes, binary code, or a character sequence, by using computer-based recognition and understanding. The ASR technologies mainly include feature extraction technology, pattern matching criterion, model training technologies, and so on. Today, ASR technologies are relatively mature and platforms like iFlytech, Baidu, Alibaba, Tencent, Microsoft, IBM, and Qualcomm provide commercial interfaces. The declared accuracy rate of these interfaces is over 95%. Currently, ASR technology is widely used in a variety of fields, such as speech input method, speech message transcription, caption generation, meeting records, and telephone quality inspection. In combination with ASR technology and other technologies, such as speech emotion recognition, speech synthesis TTS, semantics understanding, and machine translation, more complicated applications can be built.

Text classification indicates that a document set $D = \{d_1, d_2, \dots, d_n\}$ and a class set (label set) $C = \{c_1, c_2, \dots, c_n\}$ are defined, a classification function f is obtained by using a learning method of algorithm, and each document d_i in the document set D is mapped to one or more classes in the class set C . Text automatic classification started at the end of the 1950s, and models include the Boolean model, the probability statistics model, and the vector space model. Various classification algorithms have been proposed based on these three models, in which the Naive Bayes (NB) classification algorithm is simple and of high-performance [4]. At present, the NB classification algorithm is extensively used in information filtering, mail classification, search engines, prediction of query intention, analysis of academic hotspots, analysis of social hotspots, and high-frequency vocabulary for foreign language examination.

3 Analysis of problems of the existing customer service report system

Customer service report systems at home and abroad can be mainly divided into two types. The first type is the telephone-based customer service report system (e.g., a call center). The telephone-based customer service report system is dominant in the customer service system due to its advantages, such as extensive adaptability, convenient communication, and precise answers. The telephone-based customer service report system is essential for agencies and services facing large user groups. At present, the telephone-based customer service mode mainly includes self-service classification, key menu interactive voice response (IVR) [5] and manual answering, and the service quality is mainly supervised with manual sampling and simple customer assessment. The major problems can be described as

follows: First, with the expansion of services, the IVR menu levels become deeper and a customer needs to spend much time and multiple interactions while searching for the required service points, which affects the user experience. Second, it is difficult to extract effective information from the massive speech data, which results in a variety of problems, such as the low-efficiency manual input of work orders, lagged tracking of failure hotspots, and the difficult timely summarization of service improvements. Third, the service quality supervision lags. It is time-consuming for a supervisor to spot-check records and auxiliary modes. Fourth, as the core of the customer service report system, the knowledge base needs to be manually summarized in advance. Thus, the knowledge base is easy to ossify and it is difficult to dynamically update and automatically accumulate; furthermore, the continuous expansion of the knowledge base also increases training costs of the customer service staff.

The second type of customer service report system includes the emerging network of online customer service, intelligent network answering system, entity intelligent robot, and so on. These types of intelligent answer systems play a better role in supplementing and replacing human customer service, but there are still some defects. First, the knowledge base cannot be automatically updated and supplemented, and it relies on manual summarization and is perpetuated and lags; moreover, answer entries are simple and rough, and they have low accuracy, thereby affecting user experience and restricting the development of the customer service system to some extent.

4 Design of an intelligent customer service report system based on ASR and text classification technology

4.1 Design of the system architecture

The system is divided into the interaction layer, speech processing layer, data processing layer, and knowledge integration layer based on the existing call center system or new system. The ASR interface (e.g., open SDK of iFlytech) of the AI platform is called when the speech or recorded speech is connected at the interaction layer [6] to convert the speech data to text and label retelling words of customer service staff. Persistent data is realized by using the data platform. Analysis and processing, such as high-frequency vocabulary statistics, label extraction, classification, and clustering, are realized by calling text classification algorithms. The knowledge integration layer deposits the data processing results, and the results are automatically induced and accumulated as the knowledge base, which is used for the interactive layer. The knowledge base may also be extended and used in the emerging intelligent network Q&A system, which resolves the problem that the knowledge base depends on manual summarization. The general architecture is shown in Fig. 1.

Fig. 1. System architecture

4.2 Design of the function modules

To combine scenario features and satisfy user requirements, the system shall include the function modules as shown in Fig. 2.

Fig. 2. System function module diagram

4.3 Design of the service flow

4.3.1 First, customers call and enter customer service via the traditional IVR menu or intelligently predicted flat IVR labels.

The system calls the ASR interface of the third-party AI platform to convert speech to text, and the text is displayed on the screen and allows manual editing, and it is saved as work orders. Meanwhile, the system detects the service quality, including sensitive words and speech emotion detection [7]. The emotion detection may be divided into levels of “negative, placid, unfriendly, and excited”. If a severe abnormality occurs, then a supervisor is recommended in real-time for intervention. The recorded speech is stored according to the design of the original call center.

4.3.2 Second, the customer service staff retells keywords, such as “network failure” and “fail to log into the system”, of the problem during the call.

The speech is automatically converted to the manually defined classification label. Labels, work orders, and the whole paragraph of the dialog texts are submitted to the database to quickly generate work orders and confirm the basis for immediate classification and persistent data. This replaces a lot of the manual input workload and improves efficiency.

4.3.3 Third, texts stored in the database are processed and analyzed using big data technology.

Various text classification algorithms, such as word frequency statistics, Naive Bayes, support vector machine, K nearest algorithm, and Rocchio algorithm are used to extract the high-frequency vocabulary and classification labels, and to realize text classification and clustering. Finally, data is integrated into kernel achievements and automatically induced and accumulated as the knowledge base. Meanwhile, the hot words are tracked and counted to form the failure hotspot warning and statistical charts as the basis for supervision and service improvement.

4.3.4 Fourth, to further realize multi-language translation, a machine translation interface based on the automatically generated knowledge base is used.

Then, to realize real-time recommendations and answers, the related entries of the knowledge base are determined according to the ASR contents obtained when a customer service staff answers a call. Finally, the recommendations and answers are used as a reference for the customer service staff, which reduces the user’s waiting time and the training costs of the customer service staff.

4.4 Key technologies

The difficulties during function implementation lie in connecting the ASR capability and classifying text based on big data. For ASR technology, the mature iFlytech open platform can be considered as an example. At present, the AI capabilities can be quickly connected via WebAPI streaming interfaces or SDK interfaces, and, based on the latest engine of the deep full sequence convolution neural network DFCNN, the ASR accuracy rate reaches 98%, speech input speed is 180 words/min, and the recognition response time is less than 200 ms. In addition, ASR hardware modules are provided, including a microphone array and offline recognition module. Some scholars have conducted beneficial research on speech emotion recognition.

Text classification technology is one of the most typical application scenarios in NLP, and a lot of implementation methods of the text classification technology have accumulated. A big data platform may be used to run various text classification algorithms, for example, the MapReduce concurrent big data text classification method and the Spark Mllib including general learning algorithms like classification, regression, and dimension reduction. Alternatively, the text classification interface of a third-party platform, such as the Baidu EasyDL platform, is called for text classification. It does not need to know the details of the algorithm, and only a small amount of data can train a high-precision model. Alternatively, third-party free open-source text classification kits may be used for training, assessment, and classification of the user-defined text classification corpus, for example, Tsinghua University Chinese Text Classification (THUCTC) from Tsinghua University.

5 Advantages of an intelligent customer service report system based on ASR and text classification technology

- A flexible connection mode facilitates a combination with the traditional IVR customer service system and keeps industrial normalization and standardization.
- An agency with a call center or CRM can use program interfaces, software middleware, and hardware modules provided by the AI open platform to quickly acquire AI and big data technology capabilities.
- It facilitates the implementation of the flat IVR menu and the improvement of user experience.
- A user can speak freely, and the required service functions can be directly obtained through the human-machine voice interaction, which reduces the complexity of the IVR menu due to service expansion.
- Work orders can be automatically transcribed and entered, which greatly reduces the manual entry workload of customer service staff and improves work efficiency.
- Hotspots can be dynamically tracked and the knowledge base can be automatically accumulated by categories.
- The automatic accumulation of the knowledge base can effectively resolve massive special problems for minorities in manual reception, speed up experience sharing and knowledge transfer among customer service staff, and reduce the pressure of customer service staff and their service training.
- As automatically accumulated kernel resources, the knowledge base can feedback to resolve the problem that the emerging intelligent customer service knowledge base, such as network intelligent Q&A and intelligent robot, relies on manual summarization, and popularization of the intelligent customer service can be promoted.
- To realize multi-language translation of the knowledge base and improve adaptability and utilization of the knowledge base, AI machine translation technology and human-machine collaboration can be introduced based on the automatically accumulated knowledge base.
- The related entries of the knowledge base can be intelligently recommended according to the ASR contents obtained when the customer staff answers a call.
- The service quality inspection is improved to supervise service normalization in real-time.
- Illegal contents can be automatically sifted with quality inspection capabilities, such as speech emotion analysis technologies, keywords, speaking speed, and mute. The user intentions are determined and the reasons for incoming calls are matched to provide customers with anti-harassment detection.
- The strengths of user interfaces can be exploited to mine and analyze valuable information from massive user feedback data, thereby observing a customer's requirement and obtaining potential market intelligence.

6 Conclusion

With the rapid development of science and technologies, AI technologies and big data capabilities can be connected via a clicking "one-button". Currently, many open AI platforms become mature and provide diversified and convenient AI interfaces, including ASR, text recognition, image recognition, face recognition, natural speech processing, AR/VR. The big data platforms also support multiple existing text classification models and algorithms or use third-party platforms and tools with preset models. These conditions greatly reduce the thresholds for the customer service report system and other systems to introduce AI and big data technologies and promote the reform of management and service

modes in different industries. However, there is the problem that most of the AI technologies driven by big data are under the control of IT giants and some fees need to be paid for capability sharing. Thus, the economic cost must be considered. Moreover, the data will be transferred to the third-party cloud platform in call of the AI interfaces, and it may lead to hidden data security hazards [8]. These problems can be resolved by a variety of measures, such as agency negotiation, agreement signing, and relevant laws.

The author would like to thank Lili Zhang, Sijia Guo and Lin Yan for wonderful discussions. The author also would thank to the anonymous referee for the careful reading of the manuscript and many suggestions for its improvement.

References

1. J. Fen, Electronics World, **000(011)**, 186 (2019)
2. J. Fang, F. Wang, China New Telecommunications, **14**, 111 (2017)
3. W. Hongfang, C-Enterprise Management, **6**, 57 (2017)
4. H. Ming, S. Jian-jun, Ch. Ying, Information Science, **34(07)**, 147 (2016)
5. L. Sun, B. Tang, J. Tian, Z. Li, Electronic Science & Technology, **04(05)**, 73 (2017)
6. X. Liu, H. Xu, Y. Hou, J. Qi, L. Yan, Information Technology and Informatization, **03**, 33 (2019)
7. D. M Nazarov, A. D. Nazarov, D. B Kovtun, CBI 2020, **2**, 106 (2020)
8. Zhang Jie, Peng Qian. Research and Application of Call Center Service Quality Process Management Based on Speech Emotion Recognition Technology. Telecommunications Science, 2018, 34(S2): 215-218.
9. Y. Li, F. Wang, China Information Security, **07**, 41 (2020)