A Telecom-Domain Online Customer Service Assistant Based on Question Answering with Word Embedding and Intent Classification

Jui-Yang Wang[1], **Ming-Feng Kuo**[1], **Jen-Chieh Han**[1], **Chao-Chuang Shih**[1],
Chun-Hsun Chen[2], **Po-Ching Lee**[2], and **Richard Tzong-Han Tsai**[1,*]

[1]Department of Computer Science and Information Engineering,
National Central University, Taiwan, R.O.C.
[2]Telecommunication Laboratories,
Chunghwa Telecom Co., Ltd., Taiwan, R.O.C.
[*]corresponding author
thtsai@csie.ncu.edu.tw

Abstract

In the paper, we propose an information retrieval based (IR-based) Question Answering (QA) system to assist online customer service staffs respond users in the telecom domain. When user asks a question, the system retrieves a set of relevant answers and ranks them. Moreover, our system uses a novel reranker to enhance the ranking result of information retrieval. It employs the word2vec model to represent the sentences as vectors. It also uses a sub-category feature, predicted by the k-nearest neighbor algorithm. Finally, the system returns the top five candidate answers, making online staffs find answers much more efficiently.

1 Introduction

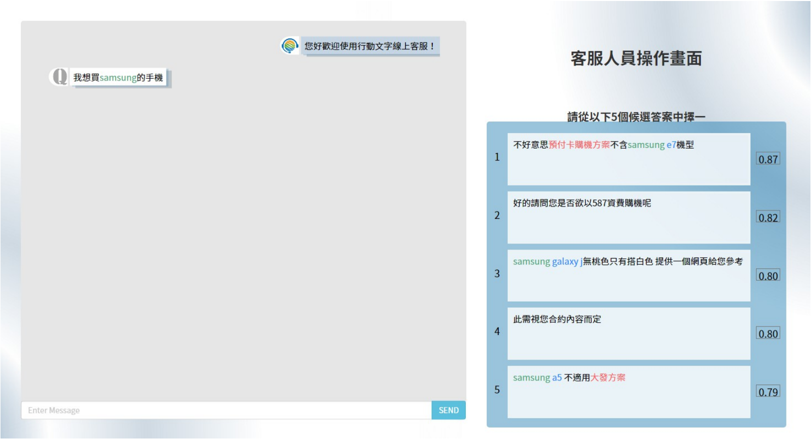
Online customer services have been used for decades. Providing the services with satisfactory quality requires a large number of well-trained online customer service staffs, making the cost not affordable for most companies. For online customer service staffs, the most time consuming tasks is to find appropriate answers from the log database of previous conversations or the frequently-asked-question database. Academic researchers and industrial engineers have made considerable efforts on developing automatic question-answering (QA) systems to support online customer services.

Currently, the most popular approach to find the answer given the queried question is information-retrieval-based. In this approach, the queried question is treated as a set of keywords. Then, these keywords are sent to an information retrieval engine, such as Apache Lucene[1], to search similar questions in the log or FAQ databases. All in-database questions are ranked according to their keyword-based similarities to the queried question. The main incapability of keyword-based similarity is that it considers only the appearance of surface keywords but overlooks the semantics. Thanks to the emergence of distributed representations of words (Mikolov et al., 2013), words are transformed to vectors that capture precise semantic word relationships. Therefore, the similarity between two questions could be measured in terms of their semantic meanings.

Currently, several studies have proved the effectiveness of incorporating word embedding features in answer ranking models (Zhou et al., 2015; Zhou and Huang, 2017; Belinkov et al., 2015; Tran et al., 2015). In our system, we derive our reranking model based on two SemEval-2016 studies (AlessandroMoschitti et al., 2016; Mihaylov and Nakov, 2016). We train a word embedding model to transform all questions into distributed vectors (Le and Mikolov, 2014; Dai et al., 2015). Due to every in-database question contains intention labels assigned by Chunghwa telecom customer service staffs, we train a multi-class classifier to identify the input questions intention. Finally, we design a reranking model using the word embedding features and the intention feature. Our system provides a web-based interface that present the user-staff conversation on the left pane and the candidate answers on the right pane, as shown in Figure 1.

[1] https://lucene.apache.org/

The Companion Volume of the IJCNLP 2017 Proceedings: System Demonstrations, pages 17–20,
Taipei, Taiwan, November 27 – December 1, 2017. ©2017 AFNLP



您好歡迎使用行動文字線上客服！

我想買samsung的手機

Enter Message

SEND

客服人員操作畫面

請從以下5個候選答案中擇一

1	不好意思預付卡購機方案不含samsung e7機型	0.87
2	好的請問您是否欲以587資費購機呢	0.82
3	samsung galaxy j無桃色只有搭白色 提供一個網頁給您參考	0.80
4	此需視您合約內容而定	0.80
5	samsung a5 不適用大發方案	0.79

Figure 1: The user interface of our system. The dialogue part is on the left side, and the top 5 answers are beside it. Name entities in the answers are distinguished by different colors. The user in the figure types ”I wants to buy a Samsung smartphone.”. The responses are ”I’m sorry. Samsung edge7 is not included in the prepaid card plan.”, ”OK. Are you going to buy cell phone the 587 plans?”, ”There is no Samsung galaxy j in Pink Gold. There is only white ones left. This is a web site that you can reference.”, ”It depends on your cell phone plan.” and ”Samsung a5 isn’t included in the Dafa plan.”.

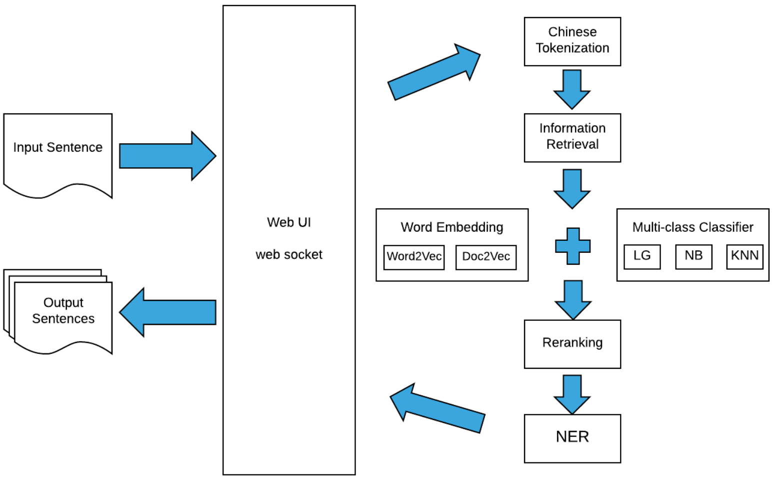


Figure 2: The architecture of our system

2 Architecture

In this section, we introduce the architecture of our system, as shown in Figure 2. Firstly, it preprocesses the input question and then retrieves top 100 QA pairs from the Chunghwa Telecom (CHT) customer-staff conversation log database. Then, our system employs a word embedding model to transform questions into vectors.

There are two models, Word2Vec[2] and Doc2Vec[3], to implement the word embedding. The word embedding make great result in finding the appropriate responses. To optimize the result, we also do multi-class classification to obtain the category features of each pair. We compare different classifiers, logistic regression (LG), Naïve Bayes (NB) and k-nearest-neighbor (KNN). The system concatenates the results of Word2Vec, Doc2Vec and multi-class classification to be the feature of reranking. Finally, our system calculates the cosine similarity between the feature of the input query and the feature of QA pairs and choose the top 5 candidates. Additionally, the system recognizes the proper nouns in the output sentences and return the relevant information with the answers of top 5 candidates to the user. We will next detail each step in the following sections.

[2] https://radimrehurek.com/gensim/models/word2vec.html

[3] https://radimrehurek.com/gensim/models/doc2vec.html

2.1 Chinese Tokenization

Our system mainly used by the people who speak in Chinese, and Chinese words in the sentence don't have the labels to show the boundary between them. Thus, the system has to do the Chinese tokenization[4]. To solve the sequence labeling problem, people often utilize Hidden Markov Model (HMM) and Conditional Random Filed (CRF) because of their high accuracy. Our system should immediately realize the input messages. It also needs to update the new words and retraining the tokenizer regularly. As a result, we adopt HMM algorithm. It is fast, and it has high accuracy. It basically takes time to count the N-gram Frequency instead of modifying the weight of features, so its training time is far less than CRF.

2.2 Information Retrieval - Okapi BM25

For the information retrieval, we adopt the Okapi Best Matching 25 (BM25) algorithm. BM 25 is arguably one of the most important and widely used information retrieval function. It is effective and it has strong retrieval capacity. However, it lacks for variety, and it's hard to unite the properties of multiply entities.

2.3 Word2Vec

After the information retrieval, we take the top 100 entities to be the candidates. Each word in a candidate will be transformed into a fixed-length vector using Word2Vec model. Word2Vec is a high efficiency model using a real number vector to represent a word. It utilize the concept of deep learning, transforming the text into a k dimension vector through training. The similarity of the word vectors can represent the similarity of the words' semantic. Because sentences consist of words, we calculate the arithmetic mean of all word vectors in the same QA pair to represent the pair.

2.4 Doc2Vec

We have tried another word embedding model, Doc2Vec, to map the sentences to unique vectors as well. Doc2Vec's concept is alike to Word2Vec except the additional paragraph vector. The paragraph vector represents the missing information from the current context and can act as a memory of the topic of the paragraph. Word2Vec and Doc2Vec can effectively make the inappropriate candidates' rank drop and left the good answers in the top 5 order. Besides being a feature to rerank the 100 pairs, the result of Doc2Vec is also the input feature of multi-class classification which is discussed in next section.

2.5 Intention Prediction

The intention of the input question is effective for searching similar question to it. For example, the question ”How to renew my plan for acquiring an iphone 7” is similar to ”I want to buy an iphone 7, could you suggest me a suitable plan?” Word-based similarity measures are hard to detect their similarity because they have few words in common. However, incorporating intention information could mitigate this problem. Fortunately, every question in the CHT customer-staff conversation log database has intention labels. Therefore, we could train a multi-class classification model to predict the intention label. We have tried three modes: logistic regression, Naïve Bayes and k-nearest-neighbor. The results will be presented in Section 3.1

2.6 NER

The contents of the answers usually contain many proper nouns such as plans, devices, Store location, etc. To respond to the user, we need the information of them. Consequently, it is important to distinguish which word is a proper noun. We use CRF Model to address the problem and train several models for the name of special offers, product name, location. The accuracy is 91%, 71.4%, 85.86%, respectively. We design the features like the suffix, brand name, product id, etc. The training data is the CHT dialogue corpus and some data collected from sogi[5].

3 Experiment and Results

3.1 Intention Prediction

In this section, we compare the different classifiers, logistic regression, Naïve Bayes and KNN. The training data is CHT dialogue corpus. It consists of 113425 sentences classified in 10 categories by the customer service staff in CHT. However, it contains a lot of noisy data which impact the performances. We manually select the most common 120 sentences in each category to be the test data, and there are 1200 sentences in total. Table 1 shows the comparison of different mod-

[4] https://github.com/fxsjy/jieba

[5] https://www.sogi.com.tw/

els. Using Doc2vec vector as feature, KNN significantly outperform other classifiers.

Type of Classifier	**Pre.**	**Rec.**	**F-1**
Logistic Regression	0.61	0.62	0.6
Gaussian Naïve Bayes	0.61	0.5	0.48
Bernoulli Naïve Bayes	0.52	0.45	0.44
K-Nearest Neighbor	**0.72**	**0.62**	**0.63**

Table 1: Comparison of different classifiers on the test data.

3.2 Reranker

To validate our approach, our system compares with the baseline, BM25, implemented in Solr. The test data is the most common query in each category. It retrieves 100 QA pairs per query. We evaluate the result of reranking using Mean Average Precision (MAP) of top 5 candidates and all candidates respectively. Table 2 and Table 3 show the performance of the experiments. Different word embedding models have different advantages. The average of Word2Vec vectors contains semantic information. Doc2vec vectors extract syntactic information from sentences. Moreover, multi-class classifiers aid to identify the categories of sentences.

Method	**MAP**
BM25	0.9197
Word2Vec	0.9671
Word2Vec + Doc2Vec	0.9692
Word2Vec + Doc2Vec + KNN	1

Table 2: MAP of top 5 QA pairs in different method.

Method	**MAP**
BM25	0.7034
Word2Vec	0.7774
Word2Vec + Doc2Vec	0.8198
Word2Vec + Doc2Vec + KNN	0.8279

Table 3: MAP of all candidate QA pairs in different method.

4 Conclusion

We present a novel system to economizes on manpower and material resources in the the online customer service. It outperforms the famous information retrieval algorithm through word representation and multi-class classification, and it achieves excellent performance.

Although there are many NN-based technique in our system, some components still need to be improved. For example, attention LSTM model shows that it's effective to know the focus word in the sentence. Therefore, we could use it to deal with intention prediction. In the reranker component, the answer part in a QA pair has not been used to match the input question in our module. The answer part plays an important role that we can use neural network to learn. These directions are worthy of our in-depth study in the future.

References

PreslavNakov LluısMarquez AlessandroMoschitti, WalidMagdy HamdyMubarak AbedAlhakim-Freihat, James Glass, and Bilal Randeree. 2016. Semeval-2016 task 3: Community question answering. *Proceedings of SemEval*, pages 525–545.

Yonatan Belinkov, Mitra Mohtarami, Scott Cyphers, and James R Glass. 2015. Vectorslu: A continuous word vector approach to answer selection in community question answering systems. In *SemEval@ NAACL-HLT*, pages 282–287.

Andrew M Dai, Christopher Olah, and Quoc V Le. 2015. Document embedding with paragraph vectors. *arXiv preprint arXiv:1507.07998*.

Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 1188–1196.

Todor Mihaylov and Preslav Nakov. 2016. Semanticz at semeval-2016 task 3: Ranking relevant answers in community question answering using semantic similarity based on fine-tuned word embeddings. In *SemEval@ NAACL-HLT*, pages 879–886.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Quan Hung Tran, Vu Tran, Tu Vu, Minh Nguyen, and Son Bao Pham. 2015. Jaist: Combining multiple features for answer selection in community question answering. In *SemEval@ NAACL-HLT*, pages 215–219.

Guangyou Zhou, Tingting He, Jun Zhao, and Po Hu. 2015. Learning continuous word embedding with metadata for question retrieval in community question answering. In *ACL (1)*, pages 250–259.

Guangyou Zhou and Jimmy Xiangji Huang. 2017. Modeling and learning distributed word representation with metadata for question retrieval. *IEEE Transactions on Knowledge and Data Engineering*, 29(6):1226–1239.