

PAPER • OPEN ACCESS

Semantic Query Expansion based on Entity Association in Medical Question Answering

To cite this article: Xue Cui *et al* 2020 *J. Phys.: Conf. Ser.* **1642** 012022

View the [article online](#) for updates and enhancements.

You may also like

- [Question answering system based on tourism knowledge graph](#)
Yuan Sui
- [Application Research of a Practical and New Intelligent Question Answering System](#)
Hua Zhou, Yaofei Tan, Luyang Wei et al.
- [Two layers LSTM with attention for multi-choice question answering in exams](#)
Yongbin Li



The Electrochemical Society
Advancing solid state & electrochemical science & technology

242nd ECS Meeting

Oct 9 – 13, 2022 • Atlanta, GA, US

Early hotel & registration pricing
ends September 12

Presenting more than 2,400
technical abstracts in 50 symposia

The meeting for industry & researchers in

BATTERIES
ENERGY TECHNOLOGY
SENSORS AND MORE!



Register now!



**ECS Plenary Lecture featuring
M. Stanley Whittingham,**
Binghamton University
Nobel Laureate –
2019 Nobel Prize in Chemistry



Semantic Query Expansion based on Entity Association in Medical Question Answering

Xue Cui¹, Pengjun Zhai^{1,*} and Yu Fang¹

¹Department of Computer Science and Technology, Tongji University, Shanghai, China

*Corresponding author: 1810369@tongji.edu.cn

Abstract. Query expansion technology is a common method to solve the semantic deviation problem between questions and answers in the question answering system. In medical question answering field, the current semantic-based query expansion method has the following two shortcomings. (1) In the stage of obtaining candidate extended terms, only the medical entity's concept in the query is used, but the medical entity's association between questions and answers is ignored (2) During the extended terms filtering stage, the interference of negative medical entities to the mutual information is ignored. Given the above-mentioned deficiencies, this paper proposes a semantic query expansion method based on entity association in medical question answering. This method first combines the query intent with the inference association between medical entities to obtain candidate extended terms from the medical knowledge graph. Then filter candidate extended terms using the selection method based on negative medical entity recognition and mutual information and thus get the final extended terms according to the co-occurrence association between medical entities. Experimental results show that our proposed method outperforms the query expansion method based on the thesaurus which commonly used in the current medical question and answering system.

1. Introduction

With the rapid development of the Internet, more and more people tend to seek medical help from online health communities. However, the rapidly increasing number of problems has brought a huge response burden to the doctors. In order to ease doctors' workload and meet the needs of users to get answers quickly, a large number of researchers have devoted themselves to the research in medical question and answering field. In medical question-answering systems, different expressions and different amounts of information between questions and answers cause term mismatches and semantic deviations, which are the key factors that affect the accuracy of the system. Therefore, researchers introduced query expansion technology to solve these problems, that is, to improve system's performance by expanding related terms to the original query.

In the current medical question answering field, query expansion methods mainly divided into keyword-based query expansion and semantic-based query expansion. However, the keyword-based query expansion method only selects keywords from the statistical level and ignores the semantic information of the query. Therefore, many unrelated medical entities may be extended to bring "noise" to the original query, thereby affecting the accuracy of answer selection. Semantic-based query expansion uses medical ontology libraries or medical semantic dictionaries to mine potential semantics in queries, but the current research on semantic-based query expansion is based on using



the concept of medical entities to select candidate extended terms, which ignore the important role of the medical entity's inference association between questions and answers in guiding the acquisition of candidate extended terms. In the extended terms filtering stage, some researchers used mutual information to filter candidate extended terms, but they ignored the interference of negative medical entities to mutual information. In view of this, this paper proposes a semantic query expansion method based on entity association in medical question answering. This method first combines query intent with the inference association between medical entities to obtain candidate extended terms from the medical knowledge graph, and then uses the extended terms selection strategy which based on negative medical entity recognition and mutual information to select final extended terms.

The main contributions of this paper can be summarized as follows: (1) A semantic query expansion method based on entity association in medical question answering is proposed. This method selects the query keywords by combining medical dictionary and query intention. Then obtains the candidate medical entities with the help of the inference association between medical entities and query keywords. Finally, the final query extended terms are obtained by using a filtering strategy based on negative medical entity recognition and mutual information. (2) Experiments show that our method achieves good performance, especially on ACC@1 and ACC@3. For these two indicators, our method improves 3.89% and 3.60% respectively compared to the original query, and improves 3.23% and 3.42% compared to the query expansion method based on thesaurus.

The rest of the paper is organized as follows: Section II overviews the related work, section III describes our method in detail, section IV shows the experimental settings and analysis. Finally, we draw a conclusion and discuss future work in section V.

2. Related work

In the question answering field, the semantic deviation between questions and answers is a key factor affecting the performance of the question answering system. In order to solve this problem, researchers have conducted extensive research on query expansion techniques.

During the candidate extended terms acquisition stage, most semantic-based query expansion uses external knowledge base to obtain candidate extended terms. In the general field, Meng [1] used the knowledge graph to obtain the hypernym and hyponym of query terms as candidate extended terms. Dipasree[2] combined pseudo-relevant feedback with WordNet to obtain candidate extended terms. Zhang [3] used HowNet to measure the relevance between the extended terms and the concept of the query to select candidate extended terms. In the medical field, because the knowledge base in the general field contains insufficient professional medical terms, people use knowledge base in the professional field such as UMLS [4], MESH [5], semantic dictionary, and the ontology library built by themselves as extended terms' source. Hersh[6] found related terms of queries through the gene-related thesaurus to expand the original query. Palotti [7] used MetaMap [8] to map the original query terms to the UMLS concept and then selected some related concepts to expand the original query. Chen[9] used MeSH to map query terms. Medical entities appearing in the top N search results are used as candidate extended terms. Song[10] extracted MESH subject terms from the titles, abstracts, and documents of the top 10 documents as candidate extended terms based on the Google search results. Wang[11] used the Personalizing PageRank algorithm to select medical entities related to the query from the medical ontology graph as candidate extended terms. However, these methods only use medical entities' concepts to select extended terms but ignoring the guiding role of the medical entity's inference association between questions and answers in obtaining extended terms

In the extended term filtering stage, some studies use mutual information to filter candidate extended terms. In the general field, Xia [12] used ontology and mutual information to collaboratively filter extension terms. Wang[13] assigned comprehensive weights to extended terms through mutual information, and set a threshold based on the weights as the basis for selecting extended terms. Yan[14] combined the topic relevance and mutual information to filter candidate terms. Huang[15] used mutual information to assign semantic weights to the words on the associated semantic tree for screening. In the medical field, some researchers also filter extended terms based on the relevance

degree of the term pairs reflected in the mutual information. Shen[16] used the highest mutual information entity as the extended term. Tu[17] used terms whose normalized mutual information exceeded the artificial threshold as candidate extended terms, and selected the top n terms as the final extended terms. Ornuma[18] used seven words as the co-occurrence window to calculate mutual information between entities so as to select extended terms. However, none of their methods take into account the interference of negative medical entities on the mutual information value between medical entities.

3. Methods

For medical question answering systems based on question answering pairs, this paper proposes a semantic query expansion method based on entity association. The overall process of the method is divided into three stages. As shown in Figure 1, in the first stage, we classify the query by query intent classifier. In the second stage we infer medical entities that may exist in the answer according to query keywords and inference association between medical entities, then choose them as candidate extended terms. In the third stage, through the extended term selection strategy based on negative medical entity recognition and mutual information, the final extended terms are obtained by the co-occurrence association of the medical entities.

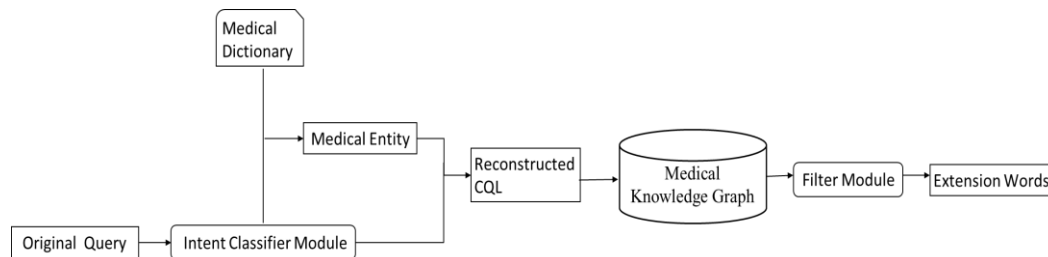


Figure 1. Semantic query expansion based on entity association in medical question answering.

3.1. Query intent classification

The determination of the query intent can help us to reason out the categories of medical entities that may exist in the answers and extract the query keywords targeted, so as to avoid introducing unrelated categories extended terms, which bring a lot of noise to the original query. We train an intent classifier to determine the intent of the original query.

Since the dataset we used does not have intent category labels, we use the self-training method to train a semi-supervised query intent classifier. In view of the fact that most questions in our dataset are disease diagnosis category, disease symptoms category and disease treatments category. And disease reason category is rare, in order to avoid the bad impact of sample imbalance on the classifier, we select SVM which is not sensitive to sample imbalance as our classifier. The target function is shown as follows:

$$\min_{\frac{1}{2}} \|\omega\|^2 \quad s.t. \quad y_i(\omega^T x_i + b) \geq 1, i = 1, \dots, n \quad (1)$$

Questions in the dataset can be divided into five categories according to the query intent: disease diagnosis category, disease treatment category, disease symptom category, diagnosis plus treatment category, and disease reason category. We find that questions with different intention categories have different interrogative words. These interrogative words can distinguish the query categories well. For example, for the disease treatment category, the interrogative words include how, how to do, how to treat. For the disease reason category, the interrogative word is why. Among them, the diagnosis plus treatment category is a compound category query, and it has both diagnosis category's interrogative feature words and treatment category's interrogative feature words. Finally, we use the interrogative word feature and the tf-idf feature to train the classifier.

3.2. Acquisition of candidate extended terms

After obtaining the intent category of the query through the intent classifier, we integrate a medical entity dictionary from 39 health website, ICD-10, ICD-9-CM and Sogou thesaurus. Then we map the terms in the query to this medical entity dictionary to extract the candidate query key medical entities. Among them, for disease diagnosis query, we extract symptom entities as initial key medical entities, and for disease treatment query, disease symptoms query, and diagnosis plus treatment query, we extract disease entities, disease entities, and symptom entities as candidate query key medical entities respectively. Then we draw on the Context algorithm proposed by HenkHarkema[19], and combine the trigger item and termination item to remove the denied medical entity from the candidate query key entities so as to obtain the query keywords. The trigger item includes the trigger feature words, and the termination item includes the stop feature words, a period and a semicolon. The trigger words and stop feature words in the dataset are shown in Table 1.

Table 1. Trigger feature words and stop feature words.

Trigger feature words	Deny, No, Not exist, Exclude
Stop feature words	But, However, Yet

After obtaining the query keywords, we combine them and the query intent to infer the categories of medical entities that may exist in the answer based on the inference rules shown in (2).

$$[rule: (Q \text{ belongsTo } C), (Q \text{ hasEntity } M) \rightarrow (A \text{ hasEntity } N)] \quad (2)$$

In the above formula, Q represents the query, A represents the answer, C represents the query intent, M represents the category of query key medical entity, and N represents the corresponding medical entity's category in the answer. The specific inference rules are as follows: For the disease diagnosis query, if there is a symptom entity in the query, we can infer the corresponding disease entities which cause this symptom by querying the knowledge graph. Similarly, for disease symptom query and disease treatment query, we can infer the symptom entity and the drug entity which may exist in the answer according to the disease entity in the query.

In this study, we extract pediatric data from the public medical knowledge graph as the external knowledge base, and use the diseases entities, symptoms entities, drugs entities, and causes entities in the knowledge graph as the source of extended terms. The query keywords obtained in the previous stage is combined with the medical entity category that may exist in the answer to form a CQL query, then we use this CQL to extract candidate extended terms from the knowledge graph. The overall algorithm's pseudocode is shown in Figure 2.

In Figure 2, after filtering out the query keywords $A[0...j]$, the classification expansion strategy is adopted for different intent categories queries. Because the expression of the same concept in spoken language and the knowledge graph is often biased, that is, the description of medical entities in the knowledge graph is more professional, the normalized operation of the key query entities $normalized(A[i]), DB$ is necessary. For the disease diagnosis category queries, the method obtains the disease entities that corresponding to the query keywords from the knowledge graph. Since the same symptoms may be caused by multiple diseases, the intersection of the disease entities obtained from each symptom entity in the query is taken as the candidate extended terms $B[0... J]$. For the queries that ask about symptoms and treatment of disease, the typical symptoms entities or drug entities corresponding to the query keywords are extracted respectively from knowledge graph as candidate extended terms $B[0... J]$. For the diagnosis plus treatment query, the disease entity is first obtained according to the same method of the disease diagnosis query, and then the commonly used drug entity is obtained according to the method of the disease treatment query. Finally, both the disease entity and drug entity are output as candidate extended terms. As for the disease reason query, because the disease reason often be composed of multiple sentences, thus it can not summarized by a few single

terms. In order to avoid bringing unnecessary noise, this category of queries are not dealt with temporarily.

```

Input: question Q, question category t, medical knowledge graph
DB
Output: expansion words list B[0...j]
1  A[0...j] <- getKeywordsByCategory (checkNon(Q,t))
2  for i in 1,2,...,j do
3    Normalize A[i] with DB
4    Search Medical Entity by category t with DB: Mt
5  end
6  if t equal diagnose then
7    Merge all response entity: M=M1 ∩ M2... ∩ Mt
8  else
9    M=M1 ∪ M2... ∪ Mt
10 end if
11 B[0...j]<-M
12 return B[0...j]

```

Figure 2. Expansion algorithm based on knowledge graph.

3.3. Extended terms filtering strategy

Due to the information deviation between the knowledge graph and the corpus, in order to avoid bringing the noise to the original query by introducing some medical entities that do not appear frequently in the corresponding answers, this paper combines negative medical entity recognition algorithm and mutual information to filter the candidate extended terms.

Mutual information can reflect the correlation degree between two terms according to the terms' co-occurrence. The higher the mutual information value, the stronger the correlation between the two terms. For two terms w_1 , w_2 , the mutual information between them can be calculated as (3):

$$I(w_1, w_2) = \log \frac{\frac{c(w_1, w_2)}{N}}{\frac{c(w_1)}{N} \times \frac{c(w_2)}{N}} = \log \frac{c(w_1, w_2) \times N}{c(w_1) \times c(w_2)} \quad (3)$$

This paper needs to pay attention to the relevance degree between each candidate medical entity and the key medical entities selected from the original query, so we use a pair of question and answer sentences as the co-occurrence window of the entity pair and calculate their mutual information. As shown in (3), $c(w_1, w_2)$ represents the number of times w_1 appears in the query of the co-occurrence window and w_2 appears in the answer of the same window simultaneously. $c(w_1)$, $c(w_2)$ represents the number of times medical entity w_1 , w_2 appears in the corpus respectively, N represents the number of all medical entities in the corpus. The co-occurrence of the denied medical entity and normal medical entities in the co-occurrence window has no practical meaning. We use a pair of question and answer as an example, the question is "My baby has eaten a lot these days, and always pulls some loose stools, but he doesn't show any discomfort. We have checked that he is not dysentery, what is going on?". The answer is "I looked at your child's condition. If it is not dysentery, it is most likely to be a dysbacteriosis". We can see that since dysentery is a denied medical entity in the answer, it shouldn't make contribution to the relevancy between itself and loose stools. So we need to recognize it during the getting mutual information matrix stage, otherwise, we may select it instead of dysbacteriosis incorrectly as our extended term due to it may have higher mutual information with the query. Therefore in the getting mutual information matrix stage, we identify the negative medical entities in the questions and answers by the negative medical entity recognition algorithm we

mentioned before. And ignore the co-occurrence entity frequency associated with them, so as to avoid interfering with the relevance degree of the overall medical entities in the corpus. The process of identifying negative medical entities is the same as the step of filtering negative medical entities from the candidate query key medical entities in the previous section.

In order to better evaluate the correlation degree $M(Q)$ between the extended medical entity w and the original query Q , we calculate the mutual information value between the extended entity and the overall query sentence, and use this as the filtering basis for the extended entity. Assuming that each key medical entity q_i in the query Q is independent, we use the sum of each group $I(w, q_i)$ as the correlation between the extended entity w and the overall query. The calculation formula is shown in (4).

$$M(Q) = \sum_{q_i \in Q} I(q_i, w) \quad (4)$$

In order to facilitate setting the filtering threshold, this paper normalized the obtained mutual information values. The formula is shown in (5), where M_{max} and M_{min} represent the maximum and minimum values of $M(Q)$ respectively.

$$NM(Q) = (M_{max} - M(Q)) / (M_{max} - M_{min}) \quad (5)$$

Finally, the candidate extended terms whose normalized mutual information value $NM(Q)$ with the original query is greater than the filtering threshold are selected as the final extended terms.

4. Experiment

4.1. Experimental Setting

4.1.1 Dataset. Due to lack of a standard Chinese medical questions and answers dataset, we have collected 7987 pairs of pediatric questions and answers from XunYiWenYao website and 39 health website as our dataset. We use triples $(q_i, ai+, ai-)$ to train our question answering model, which includes the question q_i , the correct answer $ai+$ and the wrong answer $ai-$ which randomly selected from the candidate answer pool. We generate 50 tuples of $(q_i, ai+, ai-)$ for each question in train set., and 100 tuples for each question in dev set and test set. The detailed statistical information of the dataset is shown in Table 2.

4.1.2 Preprocessing. In the pre-processing stage, we first modify the typos and remove the special characters from the dataset. Then unify the names of aliases with different names for the same disease in the corpus. Finally, we select 20% of the questions to mark the intention category as the initial training set of the semi-supervised classifier.

In the comparative experiment, we choose the query expansion method based on thesaurus which is the most common query expansion method in the medical QA field as a comparison. Due to lack of an open Chinese medical thesaurus, we collect a medical synonym dictionary by referring to Zhu's method[20]

Table 2. Statistics of the dataset

Dataset	Question Number	Answer Number	Characters Per Question	Characters Per Answer
Train	6287	6287	95	170
Dev	850	850	95	170
Test	850	850	95	170
Total	7987	7987	95	170

4.1.3 Classifier. We use SVM as the initial classifier for semi-supervised classification, and the training

set is the questions with marked intent categories in the dataset. The features used in the training include tfidf feature and interrogative words features. Besides, we set the class_weight parameter to balanced so as to alleviate the sample imbalance problem in the dataset.

4.1.4 Question answering model for answer selection. In this paper, the original question, the extended question based on the thesaurus and the extended question based on our method are used in the answer selection task in the question and answering system in order to evaluate the effectiveness of the method. In the answer selection part of our experiment, we use pre-trained word vectors published by Chinese-Word-Vectors[21]. For the question and answering models, we choose stack-CNN, multi-CNN, multi-stack-CNN, BIGRU and BIGRU-CNN these five models in the hybrid neural network framework designed by Zhang [22].

4.1.5 Metrics. We use the five evaluation indicators ACC@1, ACC@3, MRR, MAP, and NDCG to evaluate the query expansion method's effectiveness. Since the question answering system often returns one or a few high-scoring answers as the result, we hope that the correct answers will appear in the top few answers, so we focus on ACC@1 and ACC@3 indicators especially. Besides, we use the MRR, MAP, and NDCG indicators to make a more comprehensive comparison between different methods.

4.2. Results and Analysis

After the dataset is classified by the semi-supervised intent classifier, we rechecked the classification labels of the dataset. The statistical results are shown in Table 3.

Table 3. Statistics of question category.

Disease diagnosis	Disease treatment	Disease symptom	Diagnosis plus treatment	Disease reason
2815	1578	2574	939	81

Figure 3 shows the accuracy of query intent classifier. We can see that the choice of confidence threshold has a great impact on the accuracy of the classifier. We finally choose 0.8 which has the best classification performance as the confidence threshold for classifier training. The final accuracy rate reaches 73.7%.

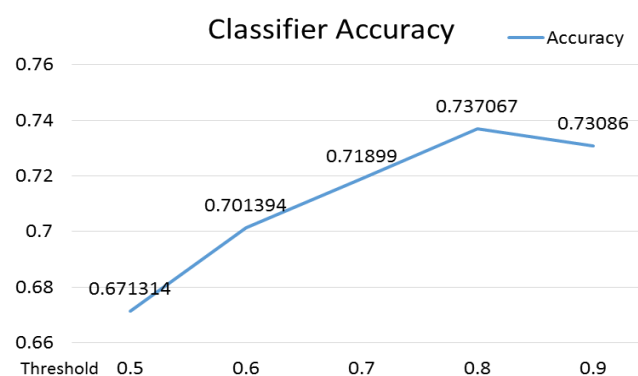


Figure 3. Results on Query Intent Classifier Accuracy

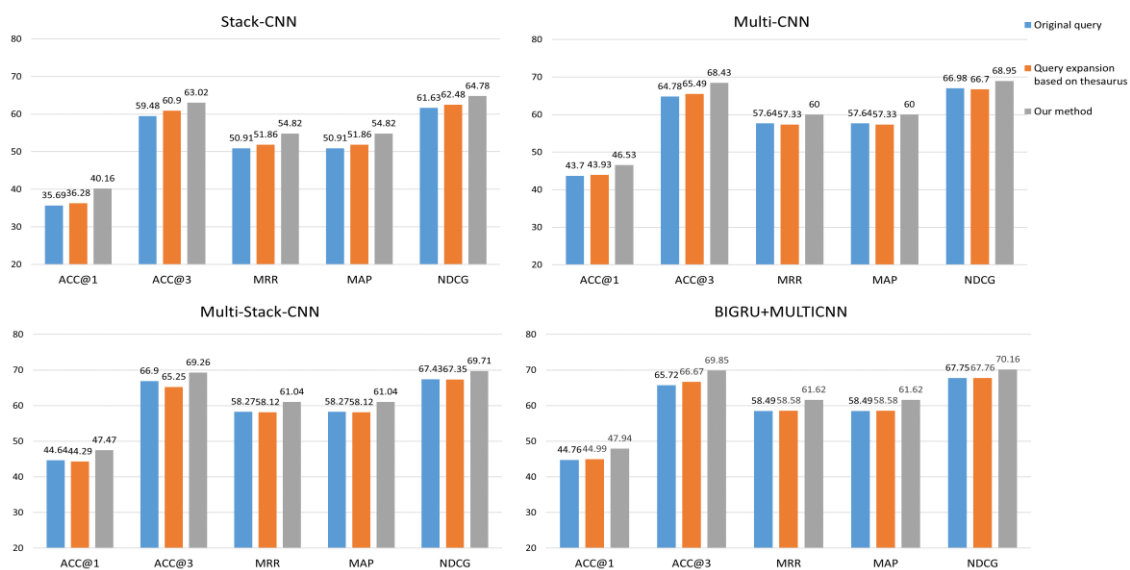
Table 4 shows three comparative experiments results obtained on the BIGRU model using the original query, the extended query based on the thesaurus and the extended query using our method. We use the evaluation tool [13] provided by TREC to evaluate the experimental effects of the three methods.

Table 4. Experimental effect on BIGRU model.

Dataset	ACC@1	ACC@3	MRR	MAP	NDCG
Original query(%)	32.04	54.89	47.30	47.30	58.78
Query expansion method based on thesaurus (%)	34.63	54.42	48.54	48.54	59.64
Our Method (%)	38.16	59.25	52.52	52.52	62.93

In order to better show the superiority and universality of our method, this paper conducted comparative experiments on stack-CNN, multi-CNN, multi-stack-CNN, and BIGRU-CNN models. Figure 4 shows the comparative experimental results.

From Figure 4, we can see that the query expansion method based on the thesaurus has limited effectiveness, which indicates that only matching synonyms of keywords cannot effectively reduce the semantic difference between questions and answers for the medical question and answer corpus. However compared with the original query and the query expansion method based on the thesaurus, our method obtains the best results on ACC@1, ACC@3, MRR, MAP, and NDCG indicators. For the ACC@1 indicator, our method increase by 3.89% and 3.60% on average respectively, for the ACC@3 indicator, our method increase by 3.23% and 3.42% on average respectively, which indicates that our method can get the higher accuracy within the top answers. For the MAP indicator, our method increases by 3.48% and 3.11% on average respectively, indicating that our method improves the average accuracy of the question-answering system. For the NDCG indicator, the average value of this article is 2.79% and 2.52% higher on average respectively, indicating that the results of our method are more relevant to the query. In the experiment, we select multiple question-answering models to verify our method's effect. The experimental results show that our method all achieves an effective improvement on these five models. Among them, the BIGRU model achieves the best improvement effect and it has improved by 6.12% on ACC@1 and 5.22% on MRR and MAP, while the BIGRU_CNN model has obtained the best performance of these five models.

**Figure 4.** Experimental results on the other four models.

In view of the above experimental results, our method has improved the accuracy and relevance of the answer selection task compared to the original query and the query expansion method based on

thesaurus. The query expansion method based on the thesaurus introduces many medical entities that don't appear in the answers, therefore may cause the query shift phenomenon. Our method can infer the medical entities that may exist in the answer by combining the query intent and the inference association between medical entities, thereby can expand original query more precisely and shortening the semantic difference between the questions and answers, thus provides an effective pre-processing idea for the answer selection task in the medical field.

5. Conclusion

In this study, we propose a semantic query expansion method based on entity association in medical question answering. This method first combines the query intent with the inference association between medical entities to get candidate extended terms through the medical knowledge graph. Then uses the extended terms selection strategy which based on negative medical entity recognition and mutual information to select the most similar terms according to the co-occurrence association as final extended terms. The experimental results show that the method we proposed achieves good performance on ACC@1, ACC@3, MRR, MAP, and NDCG. And it is superior to the query expansion method based on the thesaurus which commonly used in the medical question and answering system. However, there is still some room for improvement in our method. During extended terms filtering stage, we set the expansion threshold manually, but it is hard to determine the optimal expansion threshold. So the next step of the research will be carried out for the selection of the best expansion threshold.

Acknowledgments

This research was funded by the National Key Research and Development Program of China (No. 2019YFB2101600).

References

- [1] Meng Mingming, Zhang Kun, Luan Bing, et al. A semantic query expansion method for knowledge graph question and answer[J]. *Computer Engineering*, 2019 (9): 44. (Chinese)
- [2] Pal D, Mitra M, Datta K. Improving query expansion using WordNet[J]. *Journal of the Association for Information Science and Technology*, 2014, **65(12)**: 2469-2478.
- [3] Zhang Zhenmei, Liu Ming, Bi Li, et al. Research on query expansion method based on HowNet[J]. *Computer Applications and Software*, 2018, **35(3)**: 27-31.(Chinese)
- [4] UMLS <http://www.nlm.nih.gov/research/umls>
- [5] MeSH:<http://www.nlm.nih.gov/mesh>
- [6] Hersh W R, Bhupatiraju R T, Price S. Phrases, Boosting, and Query Expansion Using External Knowledge Resources for Genomic Information Retrieval[C]//TREC. 2003: 503-509.
- [7] J. Palotti and A. Hanbury, "TUW@ TREC clinical decision support track 2015," Vienna University of Technology, Vienna Austria, Tech. Rep., 2015.
- [8] MetaMap:<http://metamap.nlm.nih.gov>
- [9] Chen Su, Yang Yan, Hu Qinmin, et al. Query expansion based on attention mechanism in the medical field[J]. *Journal of Computer Systems*, 2019, **28(8)**: 197-203.
- [10] Song Y, He Y, Hu Q, et al. ECNU at 2015 CDS track: two re-ranking methods in medical information retrieval[C]//Proceedings of the 2015 Text Retrieval Conference. 2015.
- [11] Wang J Z, Zhang Y, Dong L, et al. G-Bean: an ontology-graph based web tool for biomedical literature retrieval[J]. *BMC bioinformatics*, 2014, **15(S12)**: S1.
- [12] Xia Lei. Research on query semantic expansion model based on ontology and mutual information[D]. *Southwest University*, 2008.(Chinese)
- [13] Wang Shuili. Research on semantic query expansion technology based on mutual information[D]. *Henan University of Science and Technology*, 2011.(Chinese)
- [14] Yan Zequan. Research on query expansion technology based on topic model[D]. *Harbin Institute of Technology*, 2014.(Chinese)

- [15] Huang G, Wang S, Zhang X. Query expansion based on associated semantic space[J]. *Journal of Computers*, 2011, **6(2)**: 172-177.
- [16] Shen W, Nie J Y, Liu X, et al. An investigation of the effectiveness of concept-based approach in medical information retrieval GRIUM@ CLEF2014eHealthTask 3[J]. *Proceedings of the ShARe/CLEF eHealth Evaluation Lab*, 2014.
- [17] Tu Wei, Gan Lixin, Huang Lehui, et al. An extended information retrieval model based on improved mutual information[J]. *Computer Engineering and Science*, 2013, **35(3)**: 150-154.(Chinese)
- [18] Thesprasith O, Jaruskulchai C. Task 2a: Team KU-CS: Query Coherence Analysis for PRF and Genomics Expansion[C]//CLEF (Working Notes). 2015.
- [19] Harkema H, Dowling J N, Thornblade T, et al. ConText: an algorithm for determining negation, experienter, and temporal status from clinical reports[J]. *Journal of biomedical informatics*, 2009, **42(5)**: 839-851.
- [20] Zhu Kangling. The Influence of Synonym Acquisition on the Retrieval and Accuracy Rate of Medical Science Novelty Search[J]. *Chinese Journal of Medical Library and Information Science*, 2012, **21(3)**: 78-80.(Chinese)
- [21] Li S, Zhao Z, Hu R, Li W, Liu T, Du X (2018) Analogical reasoning on chinese morphological and semantic relations. In: *Proceedings of the 56th annual meeting of the association for computational linguistics (Volume 2: Short Papers)*, pp 138–143 Association for Computational Linguistics
- [22] Zhang Y, Lu W, Ou W, et al. Chinese medical question answer selection via hybrid models based on CNN and GRU[J]. *Multimedia Tools and Applications*, 2019: 1-26.