Hindawi
Mobile Information Systems
Volume 2022, Article ID 7229389, 12 pages
https://doi.org/10.1155/2022/7229389

*Research Article*

# Research on Dual-Dimensional Entity Association-Based Question and Answering Technology for Smart Medicine

**Pengjun Zhai, Yu Fang, and Xue Cui**

*Department of Computer Science and Technology, Tongji University, Shanghai 200082, China*

Correspondence should be addressed to Yu Fang; fangyu@tongji.edu.cn

Received 14 June 2022; Accepted 8 August 2022; Published 9 September 2022

Academic Editor: Le Sun

With the development of the Internet of Things, intelligent medical devices and intelligent consultation platforms have been rapidly popularized, providing great convenience for medical treatment to patients and consultation to doctors. In the face of large-scale medical electronic information data, how to automatically and accurately learn professional knowledge and realize application is very important. The existing intelligent medical question answering models typically use query expansion to improve the accuracy of model matching answers but ignore the corresponding entity association between questions and answers, and the method of randomly generating negative samples cannot train the model to capture more semantic information. To solve these problems, a question answering method based on dual-dimensional entity association for intelligent medicine is proposed. This method learns semantics from the dual-dimension of question and answer respectively. In the question dimension, query extension words with strong relevance to query intention are obtained through entity association in the medical knowledge graph. In the answer dimension, answer sentences are segmented and sampled by employing a variety of similarity distances to generate negative samples in different ranges, provide different levels of correlation information between entities for model training, and then integrate the trained model to improve the accuracy and robustness of the question answering model. The experimental results show that the question answering model proposed in this paper has a good improvement in accuracy.

## 1. Introduction

As an applied branch of the automatic question answering (QA) model, the medical question answering model has been a highline of research and application with the improvement of NLP (natural language processing) technology. The QA model consists of two stages [1]: question processing and question-answer matching. The former analyses and classifies the questions, extracts keywords, and reconstructs questions, while the latter searches and matches the answers based on semantic and syntactic analysis. At present, query expansion [2] is widely used to narrow the deviation between questions and answers to understand questions more accurately. Meanwhile, many research studies match questions and answers by rules [3], clustering [4], similarity [5], and neural network [6, 7] for training the model to make the obtained answers closer to the golden answer.

However, query expansion methods based on keywords or semantics only start with the surface information such as statistics, medical dictionaries, and mutual information to mine candidate extension words, ignoring the key role of the relation between medical entities in question-answer corpus and negative medical entity recognition in obtaining extension words. In addition, the training methods of the QA model based on a single similarity or neural network collecting negative samples only focus on the relation between entities in a sample at a specific hierarchy, which lacks diversity and stability. Table 1 gives two examples of our QA data, and there is a relation of disease causes symptoms between symptom entities such as “vomiting,” “abdominal pain,” and the disease entity of “gastrointestinal dysbacteriosis” in question 1 (Q1). However, the symptom entities of “protruding navel” and “reducible tumor” are ignored because of the query deviation, and the negative entities such as “no fever” also interfere with the QA model because they

Table 1: QA examples from the Chinese pediatric corpus.

| |
|---|
| Example#1<br>Q1: Baby has a protruding navel, reducible tumor, no fever, occasionally vomiting, and abdominal pain. How is this going on.<br>A1: Baby abdominal distension, acid reflux, abdominal pain sometimes accompanied by vomiting, fever, consider caused by gastrointestinal dysbacteriosis, and imperfect Digestive system is also an essential factor. |
| Example#2<br>Q2: Small blisters appeared on the child's hands and feet, a few days later, had a fever, and have been a loss of appetite. How is this going on.<br>A2: According to the description of symptoms, it is supposed to herpes zoster is caused by virus infection. There will be blister shaped bulges in the area of nerve distribution, accompanied by fever and loss of appetite. It is recommended that you go to the Hospital for examination and treatment.<br>A2': According to the description of symptoms, it is supposed to hand-foot-mouth disease. The disease often occurs in the mouth, hands and feet, with blister shaped bulge, fever, loss of appetite and other symptoms. It is recommended that you go to the hospital for antiviral treatment. |

are not recognized. In addition, the answer 2 (A2) and A2' of Q2 have high lexical similarity because they contain such entities as "blister shaped bulge," "fever," and "loss of appetite."

An elegant framework is proposed in this paper to capture the potential entity association in Chinese medical QA corpus with the help of the medical knowledge graph, and adopt the idea of model integration [8] to train model focusing on multiple similarities at the same time by combining multiple base learners. Specifically, the main contributions are as follows:

(i) We harvest the Chinese medical QA sentences from the XunYiWenYao (http://3g.xywy.com/) and 39 Health (http://www.39.net/). The query intention classifier is trained on the pre-processed data by using the semi-supervised self-training method for realizing the automatic annotation of the query intentions in the original questions.

(ii) The proposed medical integrated QA approach focusing on dual-dimensional entity association, which focuses on the entity association from the dimensions of question and answer, reduces the noise of introducing extension words, and captures the entity relevance at different hierarchy through integrated learning so that to the model can match the golden answer more accurately.

(iii) The results on the QA dataset show that the effectiveness of the proposed model DDEA (a dual-dimensional entity association-based question and answering technology for smart medicine) in obtaining query extension words and matching questions and answers, and can more effectively capture the potential semantic information under different query intentions.

## 2. Related Work

*2.1. Question Answering Based on Query Expansion.* Since the content between the question sentence and answer sentence is different, the semantic deviation is caused, which significantly affects the accuracy of the QA model. Therefore, the query expansion method is introduced into the QA model, which makes up the semantic gap between questions and answers by adding words related to the answers to the original query. In the field of medical, external medical knowledge resources such as MeSH [9], UMLS [10], and several medical ontology databases [11] are employed as the source of extension words. However, the query expansion only based on synonyms is incapable of accurately capturing the semantic information in the corpus. Yang et al. [12] developed a QA system, which trained the classifier based on the Euclidean distance and word to select the appropriate words from the medical expert vocabulary to enhance the query of users. Although the performance of the QA model improved by ontology databases, it is difficult to construct and introduce in a specific system because of the large scale. Wang et al. [13] extended queries by finding the most relevant semantic association with potential feature vectors and association terms with triples. Li et al. [14] based on the weight of medical terms propose a query reconstruction method that weighted medical terms with self-information and then combined the weighted medical terms and the original query in proportion. Shen et al. [15] reflected the degree of association between words by utilizing the mutual information and took the entity concept with the highest mutual information value as the query extension word. To promote the performance of the QA model, Aicha et al. [16] extended queries with medical entities and semantic relations based on the external resource of OWL. Nasir et al. [17] presented a method of joint knowledge and related feedback, which analysed the diversity of query words and finds synonyms through different methods. Hu et al. [18] employed domain concepts to extend queries by fusing the domain knowledge graph and extracted context features to obtain context-aware of queries. However, the method of query expansion only adopts the concept of a domain entity and ignores the critical role of the relation between entities.

*2.2. Question Answering Based on Model Integration.* Computing the similarity between sentences or texts is greatly helpful in the QA model [19]. Scott et al. [20] proposed an enhanced lexical semantic resource model to raise the effectiveness of the QA model. Zhang et al. [6] employed the end-to-end multi-scale CNN based on the word vector to model question and answer respectively to find the correct answer using the similarity of question and answer. Based on

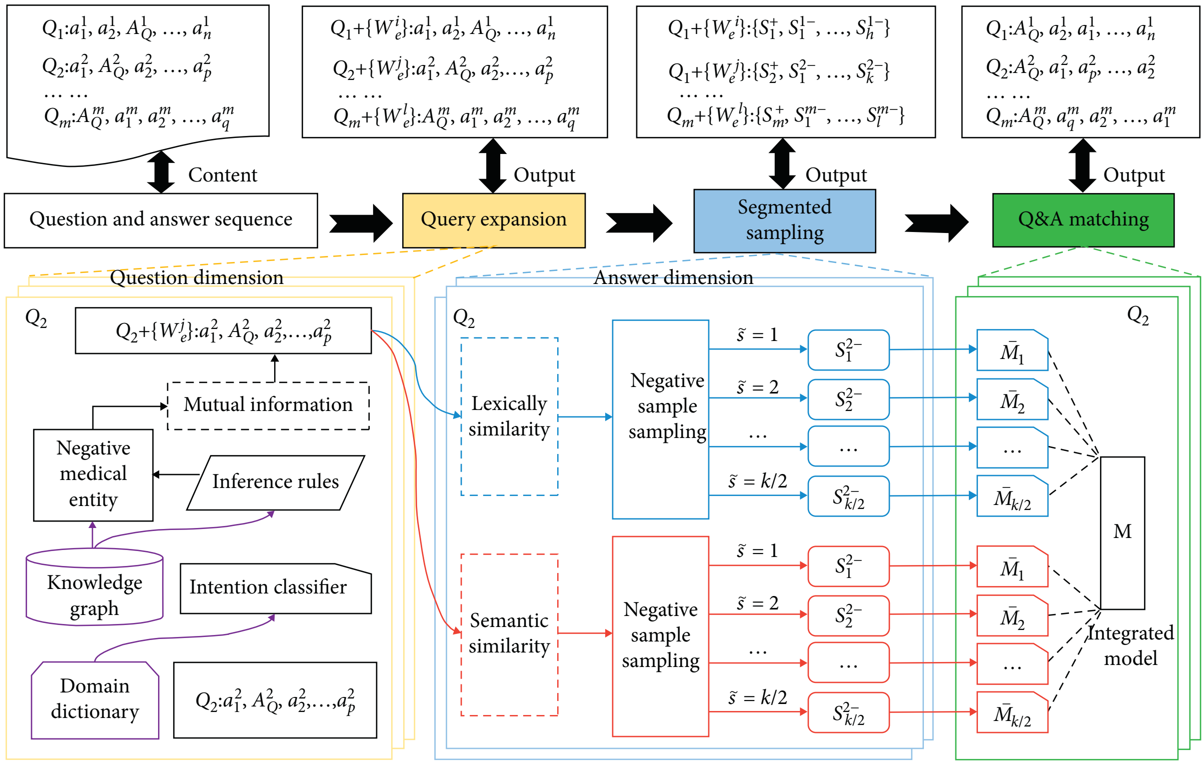

$Q_1{:}a_1^1, a_2^1, A_Q^1, \ldots, a_n^1$
$Q_2{:}a_1^2, A_Q^2, a_2^2, \ldots, a_p^2$
… …
$Q_m{:}A_Q^m, a_1^m, a_2^m, \ldots, a_q^m$
Content
Question and answer sequence
$Q_1+\{W_e^i\}{:}a_1^1, a_2^1, A_Q^1, \ldots, a_n^1$
$Q_2+\{W_e^j\}{:}a_1^2, A_Q^2, a_2^2, \ldots, a_p^2$
… …
$Q_m+\{W_e^l\}{:}A_Q^m, a_1^m, a_2^m, \ldots, a_q^m$
Output
Query expansion
$Q_1+\{W_e^i\}{:}\{S_1^+, S_1^{1-}, \ldots, S_h^{1-}\}$
$Q_1+\{W_e^j\}{:}\{S_2^+, S_1^{2-}, \ldots, S_k^{2-}\}$
… …
$Q_m+\{W_e^l\}{:}\{S_m^+, S_1^{m-}, \ldots, S_l^{m-}\}$
Output
Segmented sampling
$Q_1{:}A_Q^1, a_2^1, a_1^1, \ldots, a_n^1$
$Q_2{:}A_Q^2, a_1^2, a_p^2, \ldots, a_2^2$
… …
$Q_m{:}A_Q^m, a_q^m, a_2^m, \ldots, a_1^m$
Output
Q&A matching
Question dimension
$Q_2$
$Q_2+\{W_e^j\}{:}a_1^2, A_Q^2, a_2^2, \ldots, a_p^2$
Mutual information
Negative medical entity
Inference rules
Knowledge graph
Intention classifier
Domain dictionary
$Q_2{:}a_1^2, A_Q^2, a_2^2, \ldots, a_p^2$
Answer dimension
$Q_2$
Lexically similarity
Negative sample sampling
$\tilde{s} = 1$
$\tilde{s} = 2$
…
$\tilde{s} = k/2$
$S_1^{2-}$
$S_2^{2-}$
…
$S_{k/2}^{2-}$
Semantic similarity
Negative sample sampling
$\tilde{s} = 1$
$\tilde{s} = 2$
…
$\tilde{s} = k/2$
$S_1^{2-}$
$S_2^{2-}$
…
$S_{k/2}^{2-}$
$Q_2$
$\bar{M}_1$
$\bar{M}_2$
…
$\bar{M}_{k/2}$
$\bar{M}_1$
$\bar{M}_2$
…
$\bar{M}_{k/2}$
M
Integrated model

Figure 1: The framework of the medical integrated QA model focusing on the dual dimensional entity relation. Take the question and answer sample of question 2 as an example, $Q_2$ and $A_Q^2$ are the question 2 and its positive answer, and $a_p^2$ is the negative answer sample. For the given question 2, the query extension words are first obtained based on the knowledge graph and mutual information in the question dimension, and then segment sampling is realized for the expanded answer from the two perspectives of lexical similarity and semantic similarity in the answer dimension. After that, train the corresponding base model and integrate them to complete the matching of questions and answers.

distributed representation technology, Word2Vec, Si et al. [21] matched the correct answer by obtaining the cosine similarity between the sentence pairs. However, the performance of these models based on single models is unstable, which affects QA models' accuracy.

The method of integrating multiple excellent base models began to appear and be applied to the QA model to increase the effectiveness of the model. Hou et al. [22] selected the first N optimal classifiers for integration by training samples used the cross validation. Wang et al. [23] trained the models with the same hyperparameter and structure for several rounds and integrate them to select the answers for questions. Chen et al. [24] trained several stochastic artificial neural networks and integrated the models to predict the answers with high accuracy. Liu et al. [25] trained heterogeneous base models by the bagging method and then integrated them. Bandyopadhyay et al. [26] integrated BioBERT models pre-trained on the different corpus to achieve better results. Yang et al. [27] presented an adaptive decision fusion method, which adaptively combines classifiers with different levels of features to cultivate an answer selection model with robustness and effectiveness. However, all negative samples above with integration strategies come from random sampling, which cannot explicitly control the diversity of the base models. Therefore, Mehri et al. [28] put forward the idea of Multi-Granularity Training (MGT). The concept of MGT is to obtain the training samples set by calculating the semantic similarity between the positive and negative sample sets, and then sorting and segmenting the negative samples, thereby the MGT explicitly controlling the granularity of the semantic representation learned by the base model. However, the single dimension of semantic similarity distance does not guarantee that the model can fully employ the diversity of negative samples, resulting in insufficient diversity of the base models.

## 3. Methods

The QA model in this paper is applied to the Chinese medical automatic question answering system. We are crawling the medical question-answer pairs as the corpus and then training the query intention classifier by utilizing a semi-supervised and self-training method to tag the intention of the questions. In addition, we develop a dual-dimensional entity association-based question and answering technology for smart medicine, DDEA. Figure 1 shows the framework that gives a question and its answers as an example, which extends the question by the relation between medical entities and mutual information in the question dimension and then completes the negative sample segmentation sampling based on the multi-similarity

Table 2: Examples of interrogative words in questions.

| Question types | Interrogative words |
|---|---|
| Diagnosis | What diseases, what is it, what is going on… |
| Treatment | How to heal/treat, what should I do, how can I… |
| Symptom | Symptom, show that, display… |
| Pathogeny | Why, how is it that, what causes… |
| Diagnosis-treatment | What disease and how to treat, how did this and what to do… |

Table 3: Statistics of the domain dictionary.

| Word types | Disease | Symptom | Drug | Test |
|---|---|---|---|---|
| #Entity | 87948 | 49864 | 128036 | 9870 |

#Entity represents the number of entities.

Table 4: Statistics of feature words.

| | |
|---|---|
| Trigger feature words | Deny, no, not exist, exclude |
| Stop feature words | But, however, yet |

distance in the answer dimension to training the base models that focus on the relevance between entities of different levels. Finally, the base models are integrated to achieve more accurate automatic QA matching for different query intentions.

*3.1. Corpus Annotation.*

Data cleaning: delete invalid data such as unreadable expression and without answers, and 7987 pairs of eligible question-answer pairs are obtained by deleting individual questions other than the classify of disease diagnosis, symptom, treatment, and pathogeny in the corpus to ensure the balance of the corpus. Next, unify disease names for the corpus and about 20% of the questions, i.e., 1408 questions, are selected to tag the intent category as an initial training set for the classifier.

Intention classification: questions with the same intention often contain similar interrogative words even though they are different in various questions. The type of question is divided into five categories: disease diagnosis, treatment, symptom, pathogeny, and diagnosis-treatment. The latter is a compound type of question, which has both disease diagnosis interrogative words and treatment interrogative words. The examples of interrogative words are given in Table 2.

Automatic annotation: the query intention has a positive impact on extracting the key medical entities in questions, making the QA model accurately infer the medical entity types that may appear in the answers to avoid noises for the original query caused by introducing irrelevant extension words. Considering the sample imbalance in the dataset, the SVM [29] is selected, which is insensitive to the sample imbalance, as the initial classifier, and train the classifier by utilizing the interrogative features and TF-IDF features of questions. Finally, two physicians are invited to check and correct the annotation results of the classifier.

*3.2. Obtaining Query Expansion Words.* Query expansion is a vital step in the stage of question processing. Its accuracy directly affects the QA model performance. There, the candidate extension words more relevant with question are obtained by focusing on the relation between entities for questions under different query intentions, and the interference of negative medical entities on the value of mutual information between entities is eliminated through the negative medical entity recognition to obtain more accurate extension words.

Due to lack of a complete medical knowledge base in Chinese, we first obtain a domain dictionary of disease, symptom, drug, and test by integrating 39 Health, Sogou thesaurus (https://pinyin.sogou.com/dict/cate/index/132), ICD-9-CM (https://www.cdc.gov/nchs/icd/icd9cm.htm), and ICD-10 (https://www.cdc.gov/nchs/icd/icd10.htm). The dictionary information is shown in Table 3. Meanwhile, we extract the triples with labeled departments from an open Chinese medical general knowledge graph (https://github.com/liuhuanyong/QASystemOnMedicalKG) and develop a Chinese medical knowledge graph by utilizing the medical data crawled from 39 Health. Then, combined with the medical domain dictionary, the initial query keywords are selected for the questions with intention labels. Here, the symptom entities are extracted as the initial query keywords for the question of disease diagnosis and diagnosis-treatment, the disease entities are extracted as the initial query keywords for the question of treatment and pathogeny. The query keywords $W_k$ are obtained by removing the negative medical entities through trigger feature words and stop feature words based on the context algorithm [30] from the initial query keywords. Table 4 shows the trigger feature word and stop feature word included in the corpus.

The possible types of medical entities in the answer corresponding to the question are predicted based on the query keywords $W_k$ by using the inference rule $R$ and the types of query intentions, and the inference rule is as follows:

$$(\mathrm{QbelongsTo}\,I) \longrightarrow \left(\mathrm{QhasEntity}\,T_Q\right) \longrightarrow (A\,\mathrm{hasEntity}\,T_A), \tag{1}$$

where $Q$ and $A$ are question and answer, respectively; $I$ and $T_Q$ are the query intention and the type of query keyword $W_k$ in question, respectively; and $T_A$ is the type of medical entity in answer. Then, if there are symptom entities in the questions of disease diagnosis, it may exist in the answer, i.e., $T_A \approx T_Q$.

Based on the medical knowledge graph KG in the Chinese language, the questions with various intentions are classified and expanded by adopting reasoning rule $R$ combining the query keyword $W_k$ and the type of query intention $T$ in the question. The query keyword $W_k$ is normalized based on the entity in the KG to avoid the negative impact of the deviation between colloquial words in corpus and the same concept in KG. Then, combined with the medical entity type $T_A$ in answers obtained by reasoning

```
Input: question $Q$, question type $T$, interface rule $R$, Chinese Pediatric knowledge graph KG, threshold $\theta$
Output: expansion words list $w_e[0,\ldots,l]$
$w_e[0,\ldots,j] \leftarrow$ getKeywordsByType (checkNegative ($Q$, type))
for $i$ in 1, 2, ..., $j$ do
    Normalize $w_k[i]$ with KG
        Infer Medical Entity type $T_A$ by R
        Search Medical Entity by Type $T_A$ with KG: $M_t$
end for
if $T$ equal diagnose then
    Merge all response entity: $M = M_1 \cap M_2 \cdots \cap M_t$
else
    $M = M_1 \cap M_2 \cdots \cap M_t$
end if
Get Candidate query extension words: $w'_e[0,\ldots,h] \leftarrow M$
for $n$ in 0, 1, ..., $h$ do
    Calculate the degree of association $M(Q)$
        $NM(Q) \leftarrow$ Normalize $M(Q)$
    if$NM(Q) > 0$then
        $W_e + = W'_e[n]$
    else
        $W_e == W'_e[n]$
    end if
    return$W_e[0,\ldots,l]$
end for
```

Algorithm 1: Query expansion based on knowledge graph.

rule $R$, the medical entity $M_t$ is extracted from the entities of disease, symptom, drug, and pathogeny in KG. For the question of disease diagnosis, the intersection of disease medical entities corresponding to symptoms is regarded as candidate query extension words $W'_e$ because multiple diseases may cause several symptoms. In addition, the union of the corresponding drug and symptom medical entities is considered as candidate query extension words of treatment and symptom questions, respectively. The candidate query extension words of the diagnosis-treatment question are the union of the disease and drug entities. It is difficult to generalize with a few extension words because the pathogeny usually contains many sentences for the question of pathogeny, and this type of question will not be dealt with temporarily to avoid the noise.

Considering that there are some infrequent medical entities in the answers, which will bring noises to obtain the extension words, this paper filters candidate query extension words $W'_e$ through negative medical entity recognition and mutual information. Here, mutual information refers to the correlation between two words. It is obtained by calculating the frequency of the two words appearing in the common window,

$$I(w_1, w_2) = \log \frac{c(w_1, w_2) \times N}{c(w_1) \times c(w_2)}, \quad (2)$$

where $c(w_1, w_2)$ is the number of occurrences $w_1$ and $w_2$ appearing in the co-occurrence window at the same time, $c(w_1)$ and $c(w_2)$ are the number of occurrences $w_1$ and $w_2$ appearing in the corpus, respectively, and $N$ is the quantity of entities in the corpus.

To avoid the interference of the word frequency of negative medical entities to calculate the mutual information of medical entities, the context algorithm is employed to recognize the negative medical entities, and the associated word frequency is mapped to 0. Supposing that medical entities in a question are independent of each other, the degree of association between the extension word $W'_e$ and the question $Q$ can be obtained by calculating the sum of mutual information of each word pair, and the normalized result $NM(Q)$ is compared with the hyperparameter of extension threshold to select the final expansion words $W_e$,

$$M(Q) = \sum_{W_k \in Q} \sum_{W'_e \in A} I(W_k, W'_e), \mathrm{NM}(Q) = \frac{(M_{\max} - M(Q))}{(M_{\max} - M_{\min})}, \quad (3)$$

where $M_{\max}$ and $M_{\min}$ are the maximum and minimum of $M(Q)$, respectively. Algorithm 1 illustrates the specific process.

*3.3. Question and Answer Matching.* The randomly generated negative samples may have low similarity and poor correlation with the positive samples, which will harm the model training and reduce the effectiveness of the model. Inspired by the integration method of the Multi-Granularity Training (MGT) model [28], we develop an idea of negative sample generation based on Multi-Similarity Segmented Sampling (MSSS) in the answer dimension, that is, the initial negative samples are segmented and sampled at two levels of semantic similarity and lexical similarity to construct training sample sets focusing on different levels of relevance between entities, which is used to train distinct base models, and then these base models are integrated to match the questions and answers. The negative sample generation strategy of MSSS enables the base models to learn multi-type statement representation such as semantic and lexical, and multi-granularity statement representation from subtle to abstract.

The TF-IDF algorithm [31] is frequently employed to calculate text similarity based on lexical due to it can capture the importance of words in the text through the statistical method. The domain words are often more discriminative and important than common words for the medical question answering corpus in the Chinese language. Thus, a TF-IDF algorithm of domain weighting is utilized to calculate the lexical similarity between positive and negative samples:

$$\begin{aligned} T_n(W_n) &= tf_{W_n} * i\,df_{W_n}, \\ T_n\left(W_f\right) &= \alpha * tf_{W_f} * i\,df_{W_f}, \end{aligned} \tag{4}$$

where $T_n(W_n)$ and $T_n(W_f)$ are the TFIDF of the common words $W_n$ and domain words $W_f$, respectively, $tf$ is the term frequency and $i\,df$ is the inverse document frequency, respectively, and $\alpha$ is the weight.

In addition, the tree structure Chinese Medical Subject Headings (CMeSH, http://cmesh.imicams.ac.cn/index.action?action=index), which can clearly show the semantic relation between medical vocabulary, is adopted to calculate the similarity between positive and negative samples at the semantic level. Specifically, based on CMeSH and referring to the calculation method of semantic similarity proposed by Jiang et al. [32], we first gain the semantic similarity $\mathrm{Sim}\,(W_{f1}, W_{f2})$ between domain words and then the semantic similarity $\mathrm{Sim}\,(S_+, S_-)$ between positive and negative samples is calculated:

$$\begin{aligned} \mathrm{Sim}\left(W_{f1}, W_{f2}\right) &= \frac{1}{\left(1 + Di\,st\left(W_{f1}, W_{f2}\right)\right)}, \\ \mathrm{Sim}\,(S_+, S_-) &= \frac{2\sum_{W_{f+}\in S_+}\sum_{W_{f-}\in S_-}\mathrm{Max}\left(\mathrm{Sim}\left(W_{f+}, W_{f-}\right)\right)}{\left|S_+\right| + \left|S_-\right|}, \end{aligned} \tag{5}$$

where $\mathrm{Dist}\,(W_{f1}, W_{f2})$ is the semantic distance between domain words $W_{f1}$ and $W_{f2}$, and $|S_+|$ and $|S_-|$ are the number of domain words in the sentence of positive and negative samples.

By sorting $T_n(W_n)$, $T_n(W_f)$, and $\mathrm{Sim}\,(W_{f1}, W_{f2})$, two kinds of negative sample sequences of similarity are obtained and then sample them by segments $\widetilde{s}$ to gain different negative sample sets $S_{Q_e}^{\widetilde{s}-}$. under different segments for extended questions $Q_e$,

$$S_{Q_e}^{\widetilde{s}-} = \left\{a \in A \mid d\left(A_{Q_e}, a\right) \in \left(\mathrm{Max}d\left(\mathrm{Sim}_{Q_e}, \widetilde{s} - 1\right), \mathrm{Max}d\left(\mathrm{Sim}_{Q_e}, \widetilde{s}\right)\right)\right\}. \tag{6}$$

Here, $A$ denotes the answer, $A_{Q_e}$ is the correct answer to question $Q_e$, $d\,(A_{Q_e}, a)$ is the cosine similarity between positive and negative answers, $\mathrm{Max}d\,(\mathrm{Sim}_{Q_e}, \widetilde{s})$ is the maximum semantic similarity value in segment $\widetilde{s}$, and $Max\ d\,(T_{Q_e}, \widetilde{s})$ is the maximum lexical similarity value in segment $\widetilde{s}$. Here, for any $i \in [1, k-1]$, the $i$th negative sample $N_{Q,i}^{l}$ in segment $\widetilde{s}$ of question $Q_e$ satisfies $N_{Q,i}^{l} \sim \mathrm{Uniform}\,(S_{Q_e}^{\widetilde{s}-})$. Figure 2 illustrates that the negative sample sets are generated by segmented sampling under the given similarity, which takes the number of segments $\widetilde{s} = 3$ as an example, where $A_Q$ is the positive sample and $a_1 \sim a_9$ are the negative samples, and there are three negative sample sets in Figure 2 such as $a_1 \sim a_3$ in the first segment, $a_4 \sim a_6$ in the second segment, and $a_8 \sim a_9$ in the third segment. As this example, multiple training sample sets can be generated. In this paper, we train a base model $\overline{M}_i$ on each training sample set to make the various training sample sets fully utilized by QA, and the base models are integrated through a combination method of weighted average. The prediction probability $P\,(A_Q|\mathrm{Q})$ of the final integrated model is as follows:

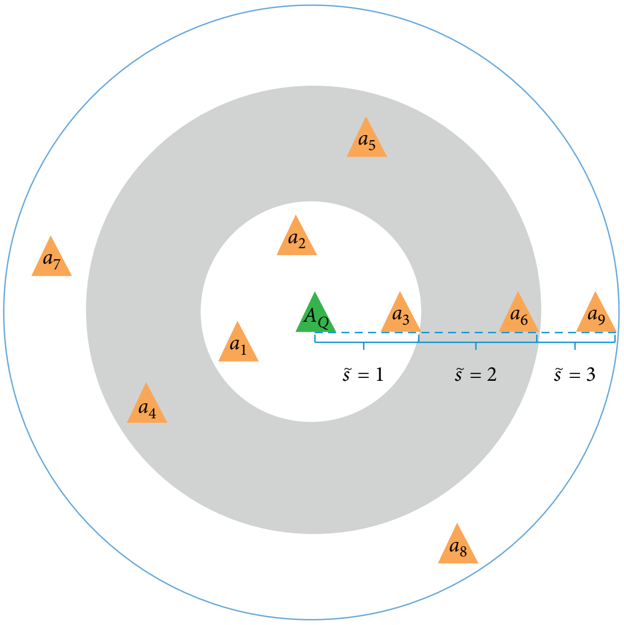


FIGURE 2: Diagram of segmented sampling.

$$\begin{aligned} P\left(A_Q|\mathrm{Q}\right) &= \sum\nolimits_{i=1}^{G} \mu_i p_i\left(A_Q|\mathrm{Q}\right), \\ \mu_i &= \frac{\widehat{p}_i\left(A_Q|\mathrm{Q}\right)}{\sum\nolimits_{i=1}^{G} \widehat{p}_i\left(A_Q|\mathrm{Q}\right)}, \end{aligned} \tag{7}$$

where $G$ is the total number of the base models, $\mu_i$ is the weight of the $i$th base model, and $p_i\,(A_Q|\mathrm{Q})$ is the predicted result of the $i$th base model. Here, the weight $\mu_i$ depends on the accuracy $\widehat{p}_i\,(A_Q|\mathrm{Q})$ of the base model on the validation set, and the weight ratio of the base model with high accuracy in the integrated model is relatively large.

We minimize formula (8) to train the model:

$$\mathscr{L}\,(\Theta) = -\left(\lambda_1 NM\,(\mathrm{Q}) + \lambda_2 P\left(A_Q|\mathrm{Q}\right)\right), \tag{8}$$

where $\lambda_1$ and $\lambda_2$ are the weights and $\Theta$ is the set of all the parameters of the DDEA. The question answer matching algorithm is shown in Algorithm 2.

```
Input: Question $Q$, segment number $\bar{s}$, Algorithm 1
Output: Integration model $IM$
    $Q_e$←Algorithm 1
      Get positive answer $A_{O_o}$ and source negative sample $S\_$
    for $W$ in $S\_$ do
        Get TFIDT$T_n(W)$ for common and domain words respectively
        for$\widetilde{W}$ in $A_{O_o}$ do
            Calculate Sim$(S_+, S_-)$
        end for
    end for
    $S_1^-$ = keywordSimilarity $(A_{O_o}, S_-)$←getNegativeSampleBy$T_n(W)$
      $S_2^-$ = sematicSimilarity $(A_{O_o}, S_-)$←getNegativeSampleBySim $(S_+, S_-)$
    for $i$ in $1, 2, \ldots, \tilde{s}$do
        Get NegativeSamples $S_{O_o}^{i-}$ for $S_1^-$ and $S_2^-$
          Triple.append (construct $(Q_e, A_{O_o}, S_{O_o}^{i-})$)
    end for
    for $j$ in $1, 2, \ldots, 2\tilde{s}$ do
      Get base model ←train $(Triple_i)$
        Measure base models $p$ and get weights $\mu$
        $IM = sign(\sum \mu p)$
      return $IM$
    end for
```

Algorithm 2: Question answer matching based on MSSS.

Table 5: Statistics of CMQA for the question answering model in Chinese pediatric.

| Dataset | #Question[1] | #Answer[2] | Ave. #CharQ[3] | Ave. #CharA[4] | Positive : negative[5] |
|---|---|---|---|---|---|
| Train | 6287 | 6287 | 97 | 171 | 1 : 50 |
| Valid | 850 | 850 | 94 | 173 | 1 : 100 |
| Test | 850 | 850 | 95 | 169 | 1 : 100 |
| Total | 7987 | 7987 | 95 | 170 | — |

[1,2]#Question and #Answer represents the quantity of questions and answers, respectively. [3,4]Ave. #CharQ and Ave. #CharA represent the average of questions and answers, respectively. [5]Positive: negative is the ratio of positive and negative samples.

## 4. Experiments

*4.1. Dataset.* We are crawling the question-answer pairs in Chinese medical from XunYiWenYao and 39 Health and annotating them by the automatic annotation method of query intention mentioned in Section 3.1. Therefore, the Chinese medical question answering (CMQA) dataset is developed, which is composed of triples $(q_i, a_i+, a_i-)$, where $q_i$, $a_i+$, and $a_i-$ are the question, positive answer, and negative answer, respectively. The negative samples in the triples are generated by the method based on MSSS in Section 3.3. The training set, validation set, and test set of the experimental data are basically divided according to the proportion of 10%, 10%, and 80% of the data volume and also combined with the length of characters in question sentences and answer sentences. To reduce the experimental error caused by the data, we try to keep the length of characters in question sentences and answer sentences consistent. Therefore, 6287, 850, and 850 question-answer pairs in the training set, validation set, and test set of the experimental data are finally selected. Table 5 shows the CMQA.

Table 6: Statistics of the query intention category for questions.

| Question type | Diagnosis | Treatment | Symptom | Pathogeny | Dia-Tre |
|---|---|---|---|---|---|
| Number | 2815 | 1578 | 2574 | 939 | 81 |

Dia-Tre represents diagnosis-treatment.

*4.2. Details.* In this study, a few comparative experiments are implemented to highlight the positive influence of entity relation in the knowledge graph for obtaining the query extension words and validate the improvement of the integration model based on multi-similarity negative sample generation on the effect of questions-answers matching. The ACC@1, ACC@3, MRR, MAP [33], and

Table 7: Statistics of CMQA for the question answering model in Chinese pediatric.

| Models | ACC@1 | ACC@3 | MRR | MAP | NDCG |
|---|---|---|---|---|---|
| *Original query* | | | | | |
| BIGRU | 32.04 | 54.89 | 47.30 | 47.30 | 58.78 |
| Stack-CNN | 35.69 | 59.48 | 50.91 | 50.91 | 61.63 |
| Multi-CNN | 43.70 | 64.78 | 57.64 | 57.64 | 66.98 |
| Multi-stack-CNN | 44.64 | 66.90 | 58.27 | 58.27 | 67.43 |
| BIGRU-CNN | 44.76 | 65.72 | 58.49 | 58.49 | 67.75 |
| Mean value | 40.17 | 62.35 | 54.52 | 54.52 | 64.51 |
| *Query expansion method based on thesaurus (QE-T)* | | | | | |
| BIGRU | 34.63 | 54.42 | 48.54 | 48.54 | 59.64 |
| Stack-CNN | 36.28 | 60.90 | 51.86 | 51.86 | 62.48 |
| Multi-CNN | 43.93 | 65.49 | 57.33 | 57.33 | 66.70 |
| Multi-stack-CNN | 44.29 ↓ | 65.25 ↓ | 58.12 ↓ | 58.12 ↓ | 67.35 ↓ |
| BIGRU-CNN | 44.99 | 66.67 | 58.58 | 58.58 | 67.76 |
| Mean value | 40.82 | 62.55 | 54.89 | 54.89 | 64.79 |
| *Our query expansion method (QE-KG)* | | | | | |
| BIGRU | 38.16 | 59.25 | 52.52 | 52.52 | 62.93 |
| Stack-CNN | 40.16 | 63.02 | 54.82 | 54.82 | 64.78 |
| Multi-CNN | 46.53 | 68.43 | 60.00 | 60.00 | 68.95 |
| Multi-stack-CNN | 47.47 | 69.26 | 61.04 | 61.04 | 69.71 |
| BIGRU-CNN | **47.94** | **69.85** | **61.62** | **61.62** | **70.16** |
| Mean value | 44.05 | 65.96 | 58.00 | 58.00 | 67.31 |

Table 8: Results of different models by integrate learning.

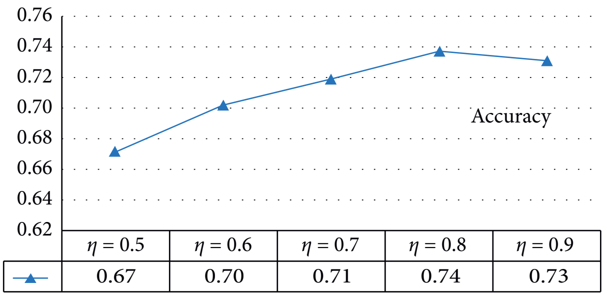
| Integrate models | Base model type | ACC@1 | ACC@3 | MRR | NDCG |
|---|---|---|---|---|---|
| Single model | — | 47.94 | 69.85 | 61.62 | 70.16 |
| Negative random sampling | Random initialization (6) | 49.12 | 71.50 | 63.01 | 71.43 |
| | Checkpoint (6) | 50.77 | 72.20 | 63.98 | 72.11 |
| | Training set (6) | 54.18 | 76.21 | 67.42 | 74.94 |
| | Mean value | 51.36 | 73.30 | 64.80 | 72.83 |
| MGT | Lexical similarity (6) | 55.63 | 77.47 | 68.30 | 75.37 |
| | Semantic similarity (6) | 56.39 | 78.33 | 68.91 | 76.20 |
| | Mean value | 56.01 | 77.90 | 68.61 | 75.79 |
| DDEA (our) | Lexical similarity (3) + semantic similarity (3) | **57.58** | **79.65** | **69.57** | **76.78** |

NDCG [34] are selected to evaluate the performance of models.

The experiments employ a scikit-learn machine learning module based on the Pytorch framework (https://pytorch.org/) and utilize the word vector published by Chinese-word-vectors [35]. The Adam Optimizer is selected, and the learning rate is initially set to 0.001, the dropout is set to 0.3, and the weight of domain words in the TF-IDF algorithm and the number of segments for segmented sampling are set to 0.6 and 3, respectively.

*4.3. Results.* From Table 6, the statistics of query intention classification for questions in the CMQA that are automatically tagged with query intention and the results of this annotation are checked by doctors.

To evaluate the significance of the proposed query expansion method based on the entity relation in the knowledge graph and the mutual information of negative medical entities, the QA models based on the hybrid neural network designed by Zhang [36] are selected as the QA encoder, such as stack-CNN, multi-CNN, BIGRU, BIGRU-CNN, and multi-stack-CNN, where the last one is the combination of the first two frameworks. Table 7 illustrates the answer select performance of models based on the original query, thesaurus expansion (QE-T), and our query expansion (QE-KG), and extending the question can improve the precision. However, the results of multi-stack-CNN under QE-T are lower than those of the original query, which indicates that extending the question only by matching the thesaurus of keywords may introduce noise. The ACC@1 of QE-KG is 3.88% higher than that of the original query on average. Comparing to QE-T, it increased by 3.23%. These results show that the method of query expansion based on the entity relation can improve the accuracy of extension words by utilizing the potential knowledge information and reduce the noise introduced by extension words by recognizing negative medical entities. In addition, the MAP and NDCG of QE-KG are improved, which respectively shows that the accuracy and the correlation between the returned answers of QE-KG are higher.

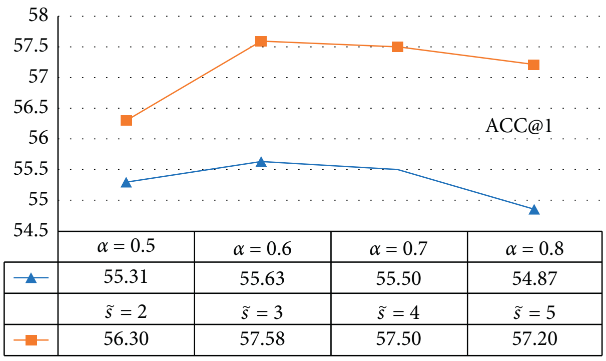


| | $\eta$ = 0.5 | $\eta$ = 0.6 | $\eta$ = 0.7 | $\eta$ = 0.8 | $\eta$ = 0.9 |
|---|---|---|---|---|---|
| ▲ | 0.67 | 0.70 | 0.71 | 0.74 | 0.73 |

Figure 3: Accuracy of the intention classifier under different confidence levels.

To prove the validity of the integrated learning method of base models trained by negative samples generated through segmented sampling based on multi-similarity in the accuracy and robustness of the QA method, this paper takes the BIGRU-CNN with the best performance as the base model structure, and implements and compares the integrated models based on a single model, negative random sampling, MGT [28], and entity relation of our DDEA. Table 8 shows the results of different models by integrate learning. Here, we design the experiments of the integrated model based on the negative random sampling from three aspects: parameter random initialization, different checkpoints, and training set with random sampling. The number of base models is set to 6, and the specific reasons are explained in Section 4.4. There are two negative samples generated methods by segmented sampling based on the lexical similarity of domain word weight and the semantic similarity of CMeSH for the integrated model based on MGT, and three base models are selected from the perspective of lexical similarity and semantic similarity for DDEA to ensure that the number of base models is consistent with that of the contrast models. From Table 8, the single model is inferior to other integrated models, and compared with the integrated models based on negative random sampling, the MGT based on similarity sampling shows more than 4.65% improvement on the ACC@1 on average, and that of DDEA outperforms the MGT-based methods by up to 1.57%, which show that the accuracy of the QA model is improved by learning more comprehensive language representation from multiple levels such as lexical and semantic through DDEA. In addition, the result of DDEA on NDCG is 0.99% higher than that of MGT, which shows that DDEA is more stable.

*4.4. Discussion.* In this paper, the QA model DDEA starts from the relation between entities to mine the extension words that have higher relevance with the key medical entities of the questions, and then trains base models based on negative samples generated by the segmented sampling of multiple similarities to promote the precision of the model. The classification of query intention can help the model specifically extract the key medical entities in questions. The threshold of confidence has a significant influence on the query intention classifier accuracy. From Figure 3, the classifier accuracy is up to the highest when it is set to 0.8.



| | α = 0.5 | α = 0.6 | α = 0.7 | α = 0.8 |
|---|---|---|---|---|
| ▲ | 55.31 | 55.63 | 55.50 | 54.87 |
| | $\tilde{s}$ = 2 | $\tilde{s}$ = 3 | $\tilde{s}$ = 4 | $\tilde{s}$ = 5 |
| ■ | 56.30 | 57.58 | 57.50 | 57.20 |

Figure 4: ACC@1 according to the value of different parameters.

According to the characteristics of the Chinese medical QA dataset, the words are divided into common words and domain words in the aspect of lexical similarity and the importance of domain words is emphasized by weight, which will directly affect the performance of the results of lexical similarity. Therefore, we evaluate the integrated model of BIGRU-CNN trained by six negative samples generated by segmented sampling based on the lexical similarity with different weights, and the ACC@1 is highest when the weight is 0.6, which is shown in Figure 4. In addition, the number of base models and the learning granularity of the base model are directly determined by the number of segments in the process of negative sample generation. Figure 4 explains the ACC@1 results for various numbers of segments, and the optimal number of segments is three. In addition, it shows a drastical decline when the number of segments goes down, making it hard for the model to take advantage of multi-granularity learning. The degree of discrimination between the base models is reduced when the number of segments is four or five, which leads to the reduction of ACC@1 and the calculation time is longer. Therefore, the number of segments is set to three, i.e., the three base models for each similarity level, a total of six base models.

Table 9 provides several examples of QA models for a disease diagnosis question to show the highest score answers returned by these models, including original query, query expansion based on thesaurus (QE-T), query expansion based on knowledge graph (QE-KG), integrated model based on semantic similarity with QE-KG (QE-KG + MGT), and our model DDEA for illustrating the advantages of DDEA more intuitively. BIGRU-CNN with the best performance is selected as the model structure. The answers in Table 9 show that the query expansion model QE-KG has a higher matching degree between the extension words and correct answers. In contrast, the query expansion model QE-T causes query deviation due to the introduction of medical entities with low relevance with the correct answers. Through the comparison of answers, QE-KG outperforms the QE-T because QE-KG can usefully shorten the gap of semantic among questions and answers by matching potential medical entities in answers for the original query to

Table 9: Summary of different works pertaining to face and speech fusion.

| Models | Question | Returned answers | Correct answers |
|---|---|---|---|
| Original query | Baby is born more than 40 days: has protruding navel, reducible tumor, occasionally vomiting, and abdominal pain, how is this going on. | Baby abdominal distension, acid reflux, abdominal pain sometimes accompanied by vomiting, fever, consider caused by gastrointestinal dysbacteriosis, and imperfect digestive system is also an important factor. | If the child's navel is raised, consider that it is related to mild umbilical hernia. You can observe whether the child's abdomen has a mass raised from the body surface. If the mass cannot disappear after lying on his back, and the child has vomiting, abdominal distension, and other symptoms, you should see a doctor immediately. |
| QE-T | Baby is born more than 40 days: has protruding navel, reducible tumor, occasionally vomiting, and abdominal pain, how is this going on. *cysts, nausea* | Brain cyst is a swelling formed by cerebrospinal fluid surrounded in the pia mater. If not treated in time, with the increase of the cyst, it can cause headache, dizziness, nausea and vomiting, and other symptoms. | |
| QE-KG | Baby is born more than 40 days: has protruding navel, reducible tumor, occasionally vomiting, and abdominal pain, how is this going on. *umbilical hernia, infantile umbilical hernia* | Umbilical hernia includes congenital type, adult type, and infant type. Infantile umbilical hernia is characterized by raised walnut size tumor in the umbilicus. In crying and after exercise, the tumor becomes larger and feels more solid. | |
| QE-KG + MGT | | Hernias include femoral hernia, indirect inguinal hernia and umbilical hernia. When the baby is crying or defecating, infantile umbilical hernia can press the raised abdominal contents back to the umbilical part. If the bulge progressive increase, it can be pressed by the abdominal belt. | |
| DDEA | | If the child's navel is raised, consider that it is related to mild umbilical hernia. You can observe whether the child's abdomen has a mass raised from the body surface. If the mass cannot disappear after lying on his back, the child has vomiting, abdominal distension and other symptoms, you should see a doctor immediately | |

obtain the correct disease entity of "umbilical hernia," but there are still differences between the returned answers and the correct answers. The integrated model DDEA successfully captures the key information, either the correct diseases or "protruding navel, vomiting" in the question, and DDEA is based on the negative samples generated by segmented sampling with multi-similarity, which makes the model learn the language representation in multi-granularity from subtle to abstract but also ensure it to learn the language representation in multi-level that increases the precision of the model prediction.

## 5. Conclusions

In this paper, we propose a new model termed DDEA, which effectively avoids weak relevance between the query extension words and the query intentions through the relation of entities from different medical scenarios in the question dimension. Meanwhile, the interference to mutual information value from the negative medical entities is reduced by recognizing these entities in the stage of extension word screening, thus improving the accuracy of query expansion. In the answer dimension, DDEA adopts a negative sample generation strategy of segmentation sampling based on the multi-level similarity to integrate the base models trained by these sample sets so that the model focuses on the relevance of entities at different levels, which improves the accuracy and stability of the QA model. The results show that DDEA outperforms the MGT integration method, and DDEA is able to capture multi-granularity knowledge information.

In the future, we will employ a more complex integration method for training a more accurate and intelligent QA model, and make the model automatically generate answers in the case of low matching for questions and answers by the deep learning methods.

## Data Availability

The data have not been made available due to privacy concerns. If necessary, the authors will select some experimental data as samples, which can be obtained from the corresponding author upon request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

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

## References

[1] B. Ojokoh and E. Adebisi, "A review of question answering systems," *Journal of Web Engineering*, vol. 17, no. 8, pp. 717–758, 2019.

[2] H. K. Azad and A. Deepak, "Query expansion techniques for informat-ion retrieval: a survey," *Information Processing & Management*, vol. 56, no. 5, pp. 1698–1735, 2019.

[3] B. Cairns, R. D. Nielsen, J. J. Masanz et al., "The MiPACQ Clinical Question Answering System," *American Medical Informatics Association*, vol. 2011, pp. 171–180, Article ID 3243235, 2011.

[4] A. Mishra and S. K. Jain, "A survey on question answering systems with classification," *Journal of King Saud University-Computer and Information Sciences*, vol. 28, no. 3, pp. 345–361, 2016.

[5] S. Wang, J. Jiao, and X. Zhang, "A semantic similarity-based subgraph matching method for improving question answering over RDF," *Companion Proceedings of the Web Conference*, vol. 2020, pp. 63-64, Article ID 3382698, 2020.

[6] S. Zhang, X. Zhang, H. Wang, J. Cheng, P. Li, and Z. Ding, "Chinese medical question answer matching using end-to-end character-level multi-scale cnns," *Applied Sciences*, vol. 7, no. 8, p. 767, 2017.

[7] S. Xu, F. Liu, Z. Huang, Y. Peng, and D. sheng Li, "A bert-based semantic matching ranker for open-domain question answering," in *Proceedings of the 4th International Conference on Natural Language Processing and Information Retrieval*, pp. 31–36, New York, NY, USA, December 2020.

[8] Z. Sun, Q. Song, X. Zhu, H. Sun, B. Xu, and Y. Zhou, "A novel ensemble method for classifying imbalanced data," *Pattern Recognition*, vol. 48, no. 5, pp. 1623–1637, 2015.

[9] T. B. Wright, D. M. Ball, and W. Hersh, "Query expansion using mesh terms for dataset retrieval: OHSU at the bioCADDIE 2016 dataset retrieval challenge," *Database*, vol. 2017, Article ID 5737054, 2017.

[10] K. Lu and X. Mu, "Query expansion using UMLS tools for health information retrieval," *Proceedings of the American Society for Information Science and Technology*, vol. 46, no. 1, pp. 1–16, 2009.

[11] C. Yun-zhi, L. Huijuan, L. Shapiro, R. S. Travillian, and L. Lanjuan, "An approach to semantic query expansion system based on Hepatitis ontology," *Journal of Biological Research-Thessaloniki*, vol. 23, no. S1, p. 11, 2016.

[12] M.-C. Yang, N. Duan, M. Zhou, and H.-C. Rim, "Joint relational embeddings for knowledge-based question answering," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 645–650, October 2014.

[13] H. Wang, Q. Zhang, and J. Yuan, "Semantically enhanced medical information retrieval system: a tensor factorization based approach," *IEEE Access*, vol. 5, pp. 7584–7593, 2017.

[14] L. Diao, H. Yan, F. Li, S. Song, G. Lei, and F. Wang, "The research of query expansion based on medical terms reweighting in medical information retrieval," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, p. 105, 2018.

[15] W. Shenwei, J.-Y. Nie, and X. Liu, "An Investigation of the Effectiveness of Concept-Based Approach in Medical Information Retrieval," in *Proceedings of the ShARe/CLEF eHealth Evaluation Lab*, May 2014.

[16] A. Ghoulam, F. Barigou, G. Belalem, and F. Meziane, "Query expansion using medical information extraction for improving information retrieval in French medical domain," *International Journal of Intelligent Information Technologies*, vol. 14, no. 3, pp. 1–17, 2018.

[17] J. A. Nasir, I. Varlamis, and S. Ishfaq, "A knowledge-based semantic framework for query expansion," *Information Processing & Management*, vol. 56, no. 5, pp. 1605–1617, 2019.

[18] Y. Hu, C. He, Z. Tan, C. Zhang, and B. Ge, "Fusion of domain knowledge and text features for query expansion in citation recommendation," in *Proceedings of the International Conference on Knowledge Science, Engineering and Management*, pp. 105–113, Springer, Cham, July 2020.

[19] G. Qiu, J. Bu, C. Chen, P. Huang, and K. Cai, *Syntactic Impact on Sentence Similarity Measure in Archive-Based QA system, Pacific-Asia Conference On Knowledge Discovery And Data Mining*, pp. 769–776, Springer, Berlin, Heidelberg, 2007.

[20] W. tau Yih, M.-W. Chang, C. Meek, and A. Pastusiak, *Question Answering Using Enhanced Lexical Semantic Models*, ACL-Association for Computational Linguistics, Redmond, 2013.

[21] S. Si, W. Zheng, L. Zhou, and M. Zhang, "Sentence similarity computation in question answering robot Journal of Physics: conference Series," *Journal of Physics: Conference Series*, vol. 1237, no. 2, Article ID 022093, 2019.

[22] Y. Hou, C. Tan, X. Wang, Y. Zhang, J. Xu, and Q. Chen, "Hitsz-icrc: exploiting classification approach for answer selection in community question answering," in *Proceedings of the 9th International Workshop on Semantic Evaluation*, pp. 196–202, June 2015.

[23] W. Wang, N. Yang, F. Wei, B. Chang, and M. Zhou, "Gated Selfmatching Networks for reading comprehension and question answering," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pp. 189–198, March 2017.

[24] C.-H. Chen, C.-L. Wu, C.-C. Lo, and F. J. Hwang, "An augmented reality question answering system based on ensemble neural networks," *IEEE Access*, vol. 5, pp. 17425–17435, 2017.

[25] H. Liu, Y. Du, and Z. Wu, "Aem: attentional ensemble model for personalized classifier weight learning," *Pattern Recognition*, vol. 96, Article ID 106976, 2019.

[26] D. Bandyopadhyay, B. Gain, T. Saikh, and A. Ekbal, *Iitp at Mediqa 2019: Systems Report for Natural Language Inference, Question Entailment and Question Answering*, BioNLP@ACL, 2019.

[27] M. Yang, L. Chen, Z. Lyu, J. Liu, Y. Shen, and Q. Wu, "Hierarchical fusion of commonsense knowledge and classifier decisions for answer selection in community question answering," *Neural Networks*, vol. 132, pp. 53–65, 2020.

[28] S. Mehri and M. Eskénazi, "Multi-granularity representations of dialog," 2019, https://arxiv.org/abs/1908.09890#:~:text=Neural%20models%20of%20dialog%20rely,at%20several%20levels%20of%20granularity.

[29] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[30] H. Harkema, J. N. Dowling, T. Thornblade, and W. W. Chapman, "Context: an algorithm for determining negation, experiencer, and temporal status from clinical reports," *Journal of Biomedical Informatics*, vol. 42, no. 5, pp. 839–851, 2009.

[31] G. Salton and M. McGill, “Introduction to modern information retrieval,” *Communications of the ACM*, 1983.

[32] J. J. Jiang and D. Conrath, “Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy,” 1997, https://arxiv.org/abs/cmp-lg/9709008.

[33] Y. Shi, A. Karatzoglou, L. Baltrunas, M. Larson, A. Hanjalic, and N. Oliver, “Tfmap: optimizing map for top-n context-aware recommendation,” in *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 155–164, New York, NY, USA, August 2012.

[34] K. Järvelin and J. Kekäläinen, “Ir evaluation methods for retrieving highly relevant documents,” vol. 51, no. 2, pp. 243–250, ACM SIGIR Forum, New York, NY, USA, 2017.

[35] S. Li, Z. Zhao, R. Hu, W. Li, T. Liu, and X. Du, “Analogical Reasoning on Chinese Morphological and Semantic Relations,” 2018, https://arxiv.org/abs/1805.06504.

[36] Y. Zhang, W. Lu, W. Ou et al., “Chinese medical question answer selection via hybrid models based on cnn and gru,” *Multimedia Tools and Applications*, vol. 79, no. 21-22, pp. 14751–14776, 2019.