An CNN-LSTM Attention Approach to Understanding User Query Intent from Online Health Communities

Ruichu Cai[1], Binjun Zhu[2]
and Wenyin Liu[6]
School of Computer Science
Guangdong University of Technology
Guangzhou, China
Email: {cairuichu,binjzh}@gmail.com
liuwy@gdut.edu.cn

Lei Ji[3]
and Jun Yan[5]
Microsoft Research Asia
Beijing, China
Email: {leiji,junyan}@microsoft.com

Tianyong Hao[4]
School of Informatics
Guangdong University of Foreign Studies
Guangzhou, China
Email: haoty@126.com

***Abstract*—Understanding user query intent is a crucial task to Question-Answering area. With the development of online health services, online health communities generate huge amount of valuable medical Question-Answering data, where user intention can be mined. However, the queries posted by common users have many domain concepts and colloquial expressions, which make the understanding of user intents very difficult. In this paper, we try to find and predict user intent from the realistic medical text queries. A CNN-LSTM attention model is proposed to predict user intents, and an unsupervised clustering method is applied to mine user intent taxonomy. The CNN-LSTM attention model has a CNN encoders and a Bi-LSTM attention encoder. The two encoder can capture both of global semantic expression and local phrase-level information from an original medical text query, which helps the intent prediction. We also utilize extra knowledge like part-of-speech tags and named entity tags to enrich feature information. Based on the experiments on a health community query intent(HCQI) dataset, we compare our model with baseline models and experiment results demonstrate the effectiveness of our model.**

I. Introduction

Traditional medical area is being effected by information technology which provides more low-cost and preferable medical services. In recent years, online health communities are very popular and are in a rapid development. The online health communities provide users with a convenient way to seek medical information support, where users can post queries describing their health situation and need. Most health communities enable interactions between users and clinicians while newly generated queries are recommended to suitable clinicians in real time to speed up the response procedure. Consequently, large amount of Question-Answer records are generated and contains inestimable medical knowledge.This QA data can benefit many medical related tasks, such as knowledge extraction [1], Question-Answering system [2] and disease inference [3].

In this paper, we try to find and predict user query intent in online health communities for better understanding requirement of the users. This is a fundamental research and may benefit many relevant medical NLP task like medical dialogue, query analysis, question similarity. Firstly, we try to mine potential user intents and build a user intent taxonomy from the original QA data. Then, we attempt to predict the existing user's intents from a realistic medical text query. It is a challenge to complete this two work in a medical domain:

For intent taxonomy mining:

- Medical text query is complex and have rich information. Most of the users would describe their problem and situation as detailed as possible. Their queries have a lot of information and even are hard to be read artificially.
- An intent taxonomy should cover most of typical users intent. If a intent taxonomy can't cover most of the user intents, is is useless in practical application.
- Each intent type in the taxonomy should be reasonable and fit the domain. Each intent type in the taxonomy should not be too general or too specific. The user intent in online health communities could be different from the traditional medical question intent, which causes it's inappropriate to use the question taxonomy proposed by the medical experts.

For intent prediction:

- The questions with the same intent have diverse expressions. For example, two questions *"What is the best way to treat rhinitis?"* and *"How to get rid of rhinitis?"* ask the treatment of rhinitis but have different words completely and grammatical representations.
- The medical query containing many complex medical concepts. Some prolix text about symptoms description, treatment history, medicine history are common in the queries. Capturing real intent in this such query would be more difficult.
- The queries could be posted by common users with less medical knowledge. Unlike the answers by clinicians always has frequent and accurate expressions, this query could exist many mistakes and colloquial expression.

2375-9259/17 $31.00 © 2017 IEEE
DOI 10.1109/ICDMW.2017.62

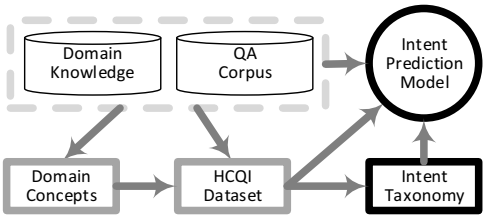

Fig. 1. Work Progress. 1) The concepts in medical domain is mined by a unsupervised clustering method. 2)The health community query intent(HCQI) dataset is labeled manually based on the domain concepts, and the queries are from the realistic QA corpus with randomly selection. 3)The intent taxonomy is constructed by a statistic for HCQI dataset. 4)All the data are used for constructing and training a user intent prediction model.

This paper propose a work progress to mine a user query intent taxonomy and predict user query intent, which is shown in Fig. 1. The medical Question-Answer corpus crawled from the online health communities and the domain knowledge are utilized to mine domain concepts and some potential user intents. Then we randomly select some queries from the QA corpus and label a user intent dataset. After summarization for the labeled dataset, 29 user intents are selected to construct a complete intent taxonomy. Finally, all the data (QA corpus, domain knowledge, labeled user intent dataset) are used to build an CNN-LSTM attention model for intent prediction.

We complete all the intent understanding work step from building the medical intent taxonomy to training a neural network prediction model. Overall, this paper makes the following contributions:

- An unsupervised query clustering method is used to mine domain concepts and potential user query intents in a limited domain.
- A user intent taxonomy is constructed for intent prediction and a health community query intent(HCQI) dataset is labeled for training and validation a intent prediction model.
- A CNN-LSTM attention model is proposed for predicting user intent in original medical text query. Compared with other typical model to predict user intent, this model can understand the point for user intent in a query and have better performance in the user query intent prediction experiment.

II. Related Work

Mining medical question taxonomy is meaningful in medical informatics domain. [4] used 1101 questions from family doctors to create a hierarchical medical question taxonomy. [5] proposed a question taxonomy for pregnant woman, and also indicated that the question taxonomy in [4] is complex and contains potentially overlapping categories. For a fixed question taxonomy, it is hard to fit all kinds application scenarios.

To predict user intent from online health communities, [6] summarized 5 categories of user intent from the online health community. [7] proposed a user intent taxonomy which has 23 kinds of user intents and trained a neural network based neural network heterogeneous intent model. To build a suitable intent taxonomy, it's better to generate a new intent taxonomy instead of using a fix and universal taxonomy.

Intent understanding is also a popular research area for natural language processing in recent years. [8] utilized CNN for text classification and get excellent result compared with traditional method. [9] utilized the RNN to train utterance classifier model and got the superior performance. [10] compared the RNN and LSTM in utterance classification task and find the length of utterance would influence the performance, and LSTM is better at handle the long sentence. While the traditional RNN only considers unilateral relations, bidirectional RNN structure [11] can solve the bilateral relationship of given point. [12] used a bidirectional LSTM and conditional random field algorithm for named entity recognition.

The attention mechanism was first applied to the NLP field in machine translation task [13]. The encoder-decoder framework used the attention mechanism to associate the expression of each word in the source language with the word that is predicting. [14]proposed a hierarchical attention networks for document classification and the attention mechanism was used to assign different weights to different words and different sentences. The attention mechanism can also help the NLP tasks including answering selection, sentence understanding.

III. Mining User Intent Taxonomy

To mine a user intent taxonomy from the online health communities with the realistic QA corpus, we proposed an unsupervised clustering method. Because of the difficulty of extracting intent information from the long and complex medical query, we cluster the short sentences segmented by the completed text queries. After the density-based clustering, each of the generated clusters has short sentences with similar meaning. With an innovative intent definition, we summarize existing medical concepts and potential user intents (concept pairs) with the generated clusters. After a practical labeling for a user intent dataset, we select appropriate user intents to build our final user intent taxonomy.

A. Intention Definition

We refer the definition of user intent in [6]. In medical domain, there are 5 main concepts: disease, symptom, medicine, surgery, and examination. For user intent in online health community, we can also enumerate some common query concepts, including reason, treatment, instruction, recover etc. The concepts are the basic elements to construct user intent.

For a complete medical query, we can get two kinds of information: **declared information** and **expected information**. For example, the query, *"What determines when type 1 diabetes develops?"* , have a declared information (type 1 diabetes) and an expected action (query for the reason). Thus we transform the detail information to certain concept pair $< disease, reason >$. The concept pair is used for representing for a user intent in a query. We assume D and E are the

TABLE I
CATEGORIES OF NAMED ENTITY (THE CATEGORY FOLLOWED * WOULD BE RECOGNIZED BY DICTIONARIES)

Pattern Word	Example	Pattern Word	Example
[n_disease]	heart disease	[n_location]*	Beijing
[n_medicine]	aspirin	[n_food]*	banana
[n_surgery]	Appendectomy	[n_appellation]*	father
[n_body]	head	[n_jobTitle]*	Director
[n_examination]	urine test	[n_hospital]*	Union Hospital
[n_symptom]	headache; fever	[n_department]*	Ophthalmology

declared concept set and the expected concept set, d and e are the concept of declared and expected information, respectively, a user intent in query is $t =< d, e > \ d \in D, e \in E$. For one text query q, there could be one or more intents. It can be represented as $I(q) = \{t_i | t_i \in T\}$, where $T = [t_1, t_2, , t_n]$ is a intent taxonomy constructed by n intents.

B. Short Sentences Clustering

To extract the concepts from original text queries, we try to cluster the queries and gather the queries with similar meanings. In detail, firstly, the particular entity words and some special words would be transferred to the pattern words to reduce the diversity. Then each complete text query would be segmented to short sentences and each sentence would be transformed into the feature including semantic vectors, word importance in the medical domain and the POS tags. Finally, we utilize the DBSCAN to cluster the short sentences and filter the noise sentences.

1) Sentence Pattern Transferring: We transfer certain entity words to the corresponding pattern word, which can represent a category of entity. The entity categories are shown in Table. I. Moreover, we also replace the common interrogatives and some particular verb to the pattern words.

2) Word Weight: The word frequency gap between two corpus can be used for calculating the word weight. Assume there are a corpus A (such as the general news corpus) and a sub corpus A' (such as the medical news corpus). If the word frequency in the sub-domain corpus A' is higher than the word frequency in corpus A, the word should be more important in the sub domain. Instead, the word should be less important. For example, the common stop words like "the" and "a" would have similar word frequency in the two corpus. But for some domain-specific words like "treatment", "hospital" and "medicine", their frequency in medical domain corpus could be dozens of times in general corpus.

We define corpus A, sub-domain corpus A' and word space W. W_A and $W_{A'}$ are the word spaces for A and A', where $W_A \subset W$, $W_{A'} \subset W$. $P_A(w)$ and $P_{A'}(w)$ are the word frequency in corpus A and A'.

The weight of word w in sub-domain corpus A' is defined as:

$$H(w, A, A') = \frac{P_{A'}(w) - P_A(w)}{P_{A'}(w)} P_A(w) \ , \ w \in W_A$$

$$H(w, A, A') = \frac{P_{A'}(w) - \min P_A}{P_{A'}(w)} \min P_A \ , \ w \notin W_A$$

We collected a general domain corpus and a medical domain corpus as A and A' to calculate word frequency. The weight $H_m(w)$ of word w in medical domain is:

$$H_m(w) = H(w, A, A')$$

3) Short Sentence Representation: We used a hybrid knowledge representation method to extract features from the short sentences. The Word2Vec [15] can extract deep semantic features for words from large-scale unannotated corpus and represent the word in a dense vector in a semantics space. For a short sentence, we select the 4 kinds of words (*noun, verb, adjective and interrogative*) according to the POS tag to represent the difficult components for a sentence. For example, we can select all the nouns in a sentence to represent the topic and select all the interrogative to represent the query type. The 4 vectors are calculated with the word2vec vector and the word weight proposed above. The weighted average of 4 component vectors would be the feature vector for the sentence.

We define POS Tag category space $P = \{p_1, p_2, ...p_m\}$. For the word w, $POS(w) \in P$ is its POS tag and $Vec(w) \in \mathbb{R}^d$ is a word vector to represent its semantics where d is the dimension of word vector. The sentence s with k words have the word sequence $L(s) = [w_1, w_2, ...w_k], w \in W$. We select the words with difficult POS Tags to build corresponding part-of-speech vector. Assume we select a POS tag set $P_t \in P$, the corresponding vector is:

$$T(s, P_t) = \sum_{i=1}^{l} \frac{H_m(w_i)}{\sum_{j=1}^{l} H_m(w_j)} Vec(w_i)$$

$$w_i, w_j \in \{POS(e) \in P_t | e \in L(s)\}$$

We utilize the word weight calculated by the word frequency diversity of two corpus to enhance the influence of important words in medical domain. We select four kinds of part-of-speech as noun, verb, adjective and interrogative to build the vectors, whose POS tags are P_{norn}, P_{verb}, P_{adj} and P_{inter}. These four categories of words are used to determine the essential meaning of a sentence. Afterward, we sum the four vectors with parameter k_t to weight influence of each composition to the final feature representation of the sentence. We add the constraint for k_t to adjust the contribution of the POS vectors.

$$F(s) = k_t T(s, P_t) \ \ t \in \{norn, verb, adj, inter\}, \sum k_t = 1$$

4) DBSCAN Clustering: DBSCAN [16] is a density-based unsupervised clustering algorithm with efficient anti-noise capability. We utilize DBSCAN to cluster the short query sentences with the feature have proposed above. Though the clustering, the diverse sentence have less similar sentence would be filtered. Thus each cluster has the short sentences with similar semantics and expression. For the clustering

result partly shown in Table. II, we obtain some conformal expressions containing the collective user intents.

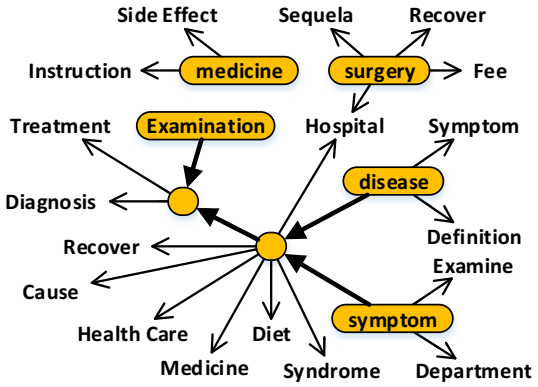

Fig. 2. Dynamical User Intent Taxonomy. The word with a rounded rectangle is the **concept of declared information** and the word without rectangle is the concept of **expected information**. The filled circle in graph is the relay node, which can connect other nodes. In the graph, any complete directed path connects the two kinds of concepts is a user intent.

C. Intent Taxonomy

The clustering result is utilized to mine the information concepts and build an intent taxonomy. Firstly, a group of similar sentence pattern in a cluster can be used to summarize a information concepts. For example, for a short sentence pattern, "what is the reason" ("是什么原因" in Chinese), we can know the expected information is the cause of something. Thus we use the word "cause" to represent this category of expected information. For some complement sentence patterns like "how to treat the [n-disease]" ("如何治疗 [n-disease]" in Chinese), we can use the "disease" and "treatment" stand for the concepts of declared information and expected information. The concepts and the intents composed by a concept pair can build the primary intent taxonomy. Secondly, we randomly select the queries from the corpus which we used for clustering. As the same time, we label intents with the information concepts and construct the primary intent taxonomy. The new concepts pair could be concluded when we label the queries data, which create the new intent label and can construct a complete intent taxonomy. Finally, we build an intent taxonomy with a specific corpus by our method. All the concepts are placed on the graph. All the concept pair of an intent are connected with directed arrow, which is shown in Fig. 2.

IV. CNN-LSTM Attention Model

We propose a CNN-LSTM attention model to predict the user intent from queries in online health communities and the model structure is shown in Fig. 3. The model transform the input query to a feature matrix firstly, and utilize a CNN encoder and Bi-LSTM Attention Encoder to represent the query. Finally, the two feature vectors are concatenated for predicting the user intent.

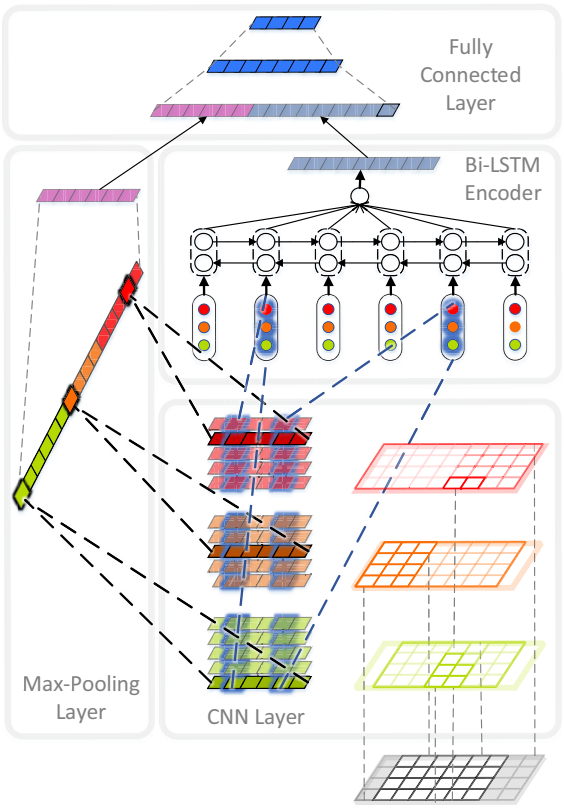

Fig. 3. CNN-LSTM Attention Model. This model use different size of filters extract the feature maps from sentence. Then a Max-Pooling layer and a Bi-LSTM attention encoder are connected in parallel. The features from two layers are concatenated and the fully connected layers are applied to reduce the feature dimension. The shadow blocks in input feature matrix are the padding for keeping the feature maps have the same length with input matrix.

A. Input Feature Matrix

We transfer the question text to the feature matrix as the input for our joint model. The query text would be segmented to words and each word could get one semantic vector with the Word2Vec [15] model and two binary vectors with POS and NER tags. Therefore the three kinds of feature can be used to present the word vector and all the word vectors can construct the input feature matrix.

A medical QA corpus is used for training the Word2Vec model. We also use tokenizer [17] for the word segmentation and POS tagging and the model [18] trained by our NER corpus for NER tagging. Besides, the existing entity lexicon is also used to tag the entities like the word of location and food. Because the Chinese interrogatives with same meaning could have lots of different forms, we also recognize the common interrogative tags in Chinese as the special POS tags. All the NER categories tag are shown in Table. I.

B. CNN encoder

1) CNN layer: We implement a convolution layer with one dimensional convolution operation [8]. Filter with a certain length l would slide over the sequence and extract the phrasal level feature. Each filter would generate one feature sequence as feature map, and the feature in the feature map is corresponding for the slide window.
We use **3** group of filters with the length from are **2** to **4**. The number of each group has **200** filters. We utilize the multiple filters to extract more feature maps, and different length of filters can extract different phrase-level representation. Finally, we concatenate all the feature maps and structure a CNN feature matrix. Relu is applied as activation function.

2) Max-Pooling layer: We apply a max-over-time pooling operation [19] with the feature maps built by the CNN filters. For a feature map $c = [c_1, c_2, , c_n]$ $c \in \mathbb{R}$, $\hat{c} = max\{c_i\}$ is the maximum feature, which is selected to represent intensity of this particular filter in a query. This operation try to capture the most important feature from a feature map with the highest value. Finally, each filter would be represented as a feature and the feature dimension would also be reduced.

C. Bi-LSTM Attention Encoder

1) LSTM: Long short-term memory (LSTM) [20] is a recurrent neural network (RNN) architecture. The excellent memory of LSTM can also handle the long-term dependencies of concepts in the sentence. LSTM can extract the interaction information of crucial concepts in the sentence and the overall emotional expression. We use the ordered phrase-level information as the input for LSTM to extract the interaction of concept and semantic representation in the question text, which is the crucial method to understanding the intent in the question sentence and representing the corresponding feature. A LSTM model process a vector sequence input $X = [x_1, x_2, ..., x_n]$ from begin to end and calculates a hidden state for each time step as:

$$h_i = LSTM(h_{i-1}, x_i)$$

2) Bidirectional LSTM: One shortcoming of conventional RNNs is that they are only able to make use of previous context. Bidirectional LSTM(Bi-LSTM) [11] have one LSTM layer in each direction (forward and backward), and then concatenate the output at each time step to represent that sentence position. This wrapping for RNNs model can capture both previous and posterior content for corresponding position. A Bi-LSTM model have two LSTMs ($\overrightarrow{LSTM}$,$\overleftarrow{LSTM}$) reading same input sequence with different direction. We concatenate the two hidden state $[\overrightarrow{h_i},\overleftarrow{h_i}]$ as the final hidden state output h_i of input x_i.

$$\overrightarrow{h_i} = \overrightarrow{LSTM}(h_{i-1}, x_i)\,; \overleftarrow{h_i} = \overleftarrow{LSTM}(h_{i+1}, x_i)$$

$$h_i = [\overrightarrow{h_i}; \overleftarrow{h_i}]$$

3) Attention Mechanism: [21] proposed utilize a attention mechanism for a translation task. In a medical query, each word may contribute unequally to the representation of the user intent. Question sentence may be more important than other description sentence, and some key entities words can decide the type of user intent. Therefore, we utilize attention mechanism to extract such words that are important to the intent of the query and aggregate the representation of those informative words to form a feature vector.
The CNN and Bi-LSTM layer map the input query q to a sequence of hidden state $H = [h_1, h_2, ..., h_n]$. The output dimension of single LSTM is **300**. BiLSTM concatenated the outputs of two LSTM model and the dimension of output hidden state is **600**. To encode all the hidden states H, we define:

$$e_i = b^\top \tanh(Wh_i)$$

$$\alpha_i = \frac{\exp e_i}{\sum_j^n \exp(e_j)}$$

$$v = \sum_i^n \alpha_i h_i$$

where v is encode state calculated by the weighted sum of H, and α_i is the weight of h_i. Parameters $W \in \mathbb{R}^{d\times d}$,$b \in \mathbb{R}^d$ are used for transforming h_i to a scalar. In this way, the query could be encoded into a vector v, and each attention scalar α_i can demonstrate the attention degree for the i-th word in query q.

D. Joint Layer

The Bi-LSTM attention encoder can capture the main semantic information and the question type information for representing the question intent. The CNN encoder can capture the word level information, including the word concept and named entity type. Therefore, we concatenate both of the feature vectors to represent the ultimate feature of a query. Two fully connected layers with sigmoid and softmax activation function are applied to reduce the dimension. The last layer have the same dimension with the size of intent category. Thus the model output is the probability distribution of each intent category.

E. CNN-LSTM Attention Model Structure

The CNN-LSTM attention model is shown in Fig. 3. In our model, a text query would be transform to a input matrix firstly with the method in Sec. IV-A, which can be processed by neural network. We apply a CNN Encoder to extract feature from the input matrix firstly. Each filter of CNN convolute a feature map from the input matrix and each feature in the feature map have particular information related to corresponding convolutional area. For each feature map from different filter, we select the most influential feature to represent it with Max-Pooling layer, and get the first feature vector. For all feature maps, the features with same index in each feature map are be concatenate, which construct a vector sequence with the same length of query. A Bi-LSTM

Attention encoder is used to encode this vector sequence, and build the second feature vector. The attention encode layer can distinguish the crucial features from hidden state output of a Bi-LSTM layer. Finally, the two feature are concatenated and fed into a multi-layer perceptron. The last layer in the MLP have same dimension with our existing number of user intent. And the output after the softmax activation function can predict the probability distribution of user intent.

V. Experiment

A. Data and Preprocessing

1) Medical QA Corpus: xywy.com is one of the most popular online medical health community in China, where a large number of user post their health questions and the clinician provide professional diagnosis and advice. We crawled question-answers data from the website and got about **20 million** valid QA pairs.

2) Medical Domain Knowledge: There are **272,816** words in total of 12 classes of named entity shown in Table. I, like disease (8,356), medicine (60,647), surgery(11,252), symptom(14,508), department(200). In addition, we collected named entities and frequently-used nouns from public lexicon and various professional medical websites.

3) Word2Vec: We use the Word2Vec [15] to train the word vectors with all the question and answer text from the data we have crawled. (4.34 GB). We train the word2vec model with the Skip-gram architecture and setting the windows size as 7. The dimension of word vector is 100.

B. Short Sentence Cluster

We used **132,298** complement text queries for the sentence segmentation and clustering. The result have **744** clusters and **35,979** short sentences. Each cluster have more than 5 short sentences. This representative pattern of each cluster can display the main semantics in this cluster. We can find more the information concept category with this intuitionistic cluster result. The clustering result partly shown in Table. II, which include the domain concepts and the potential user intents we summarized.

C. HCQI Dataset and Intent Taxonomy

We randomly selected the data in 2015 and choice the queries have the length range from 6 to 150. After manual annotation by three annotators with the medical information concepts which partly shown in Table. II, we build the health community query intention (**HCQI**) dataset. which have valid **11,964** queries with **15000** labels. As some categories of intent have less labels, we assign the intent category which have less than 100 labels into a similar or high-level category. Finally, **29** intent categories in Table. III are selected to represent the intent categories in our corpus. The dataset is split in training(60%), validation(20%) and testing(20%).

TABLE III
THE STATISTIC FOR 11964 INTENT LABELS OF TEXT QUERIES.

Intent Label	Count	Intent Label	Count
<disease,medicine>	823	<symptom,medicine>	410
<disease,treatment>	663	<symptom, examination>	367
<disease,hospital>	542	<symptom,recover>	295
<disease,symptom>	540	<symptom,department>	248
<disease,cause>	485	<symptom,diet>	212
<disease,diet>	479	<symptom,hospital>	192
<disease,surgery>	439	<symptom,syndrome>	138
<disease,recover>	218	<symptom,care>	135
<disease, care>	189	<examination,treatment>	534
<disease,syndrome>	153	<examination,instruction>	168
<disease,diagnosis>	129	<examination,diagnosis>	533
<disease,definition>	100	<examination,treatment>	147
<symptom,cause>	1782	<medicine,instruction>	357
<symptom,treatment>	1118	<medicine,side_effect>	201
<symptom,diagnosis>	631		

D. CNN-LSTM attention model evaluation

1) Baseline Models:

- *Classic Classification Models*: We extract the n-gram features of segmented text queries and use classic classification models like Logistic Regression (LR) and Support Vector Machine (SVM) for our baseline.
- *CNN Models*: [22] proposed a widely used CNN model for text classification: CNN-static, CNN-non-static, CNN-rand.
- *Bi-LSTMs model*: We also use the stacked Bi-LSTM layers for our baseline models. The output of the last time step in the last Bi-LSTM layer would be concatenated and fed to the 2 fully connection layers, which is utilized to reduce the features dimension. The softmax activation function would be applied to the last layer. 1Bi-LSTM, 2Bi-LSTM and 3Bi-LSTM are this LSTM model with 1-3 stacked Bi-LSTM layers.
- *LSTM model*: We apply a attention layer in Sec. IV-C3 after the last Bi-LSTM layer in the Bi-LSTMs model. And the models with 1-3 stacked Bi-LSTM layers are named as 1Bi-LSTM-AT,2Bi-LSTM-AT and 3Bi-LSTM-AT.

To be fair, we use the input feature matrix in Sec. IV-A as the input for the CNN and LSTM models. All the neural network based model shared the same input.

2) Evaluation Metrics: We apply the ranking based evaluation metric to our result. The Average Precision(AP) and Ranking Loss(RL) are two commonly used metrics for multi-labeled task by the ranking of labels. [23]. We also proposed two metrics, Top k Precision and Recall, to evaluate the top k result in our prediction result.

Assume a dataset $G = [g_1, g_1, ...g_n]$ with n labeled data. For a data g, we define $L(g)$ is the true label set, $L'(g,k)$ is the top K predicted label set. We define the function $P(G_i,k) = \begin{cases} 0, L(g) \cap L'(g,k) = \emptyset \\ 1, L(g) \cap L'(g,k) \neq \emptyset \end{cases}$ to show if the top k

TABLE II
TYPICAL SENTENCE AND SUMMARIZED CONCEPT PAIR FROM THE CLUSTERING RESULT. d AND e ARE THE DECLARED CONCEPT AND EXPECTED CONCEPT

Cnt	Typical Sentence (English)	Typical Sentence (Chinese)	Concept Pair(<d,e>)
1811	how to treat	如何治疗	<*,treatment>
1170	can I find you for treatment	能否找您看病	<*,in_treatment>
408	what should be noted	应该注意什么	<*,care>
398	what medicine to eat	吃什么药	<*,medicine>
394	what medicine should I use	如何用药	<*,instruction>
388	feel [n_symptom] recently	最近[n-symptom]	<symptom,*>
286	what kinds of tests do I need	需要做什么检查	<*,examination>
267	how to use the medicine	如何用药	<medicine, instrution>
227	what is this disease	是什么病	<*,disease>
174	what about the cost	费用大约多少钱	<*,fee>
144	minor side effect	副作用小	<*,side_effect>
122	how to treat [n_disease]	如何治疗 [n_disease]	<disease, treatment>
64	is there any danger	有危险吗	<*,risk>
62	want to know what the disease is	想知道是什么病	<*,cause>
59	which department should I register with	挂哪个科	<*,department>
57	whether surgery is needed	能否做手术	<*,surgery>

labels can hit any true label. We also define the function $P(g,k) = \begin{cases} 0, L(g) \subseteq L'(g,k) \\ 1, L(g) \nsubseteq L'(g,k) \end{cases}$ to show if the top k labels can cover all the true labels. Thus we define the precision and recall for top k predicted labels as: $Precision(G,k) = \frac{1}{n}\sum_{i=1}^{n} P(G_i,k)$ and $Recall(G,k) = \frac{1}{n}\sum_{i=1}^{n} R(G_i,k)$.

3) Evaluation Result: As shown in Fig. 4, the four groups of baseline models are used to compare with our CNN-LSTM attention model. In this experiment, most of the neural network based models have better capability to predict user query intent than classic classification model, and our CNN-LSTM attention model got the best performance. The SVM models with n-gram feature obtained low performance than LR models. Furthermore, the LR model with bi-gram feature got the excellent result closing to the result of 2Bi-LSTM and CNN-Dynamic, which are the best performance in Bi-LSTM and CNN models. It indicates that the traditional text classification models can also handle the task because user intent have strong relation to some key words. However, after the attention mechanism was applied to the Bi-LSTMs models, the Bi-LSTM attention models got the best performance in the all of the basic model groups. It proves that the attention mechanism can really help the user intent understanding and concentrate on encoder the task relevant information dynamically. Our CNN-LSTM attention model, which contains a CNN encoder and a Bi-LSTM attention encoder, can extract the comprehensive information to improve our prediction, including both the phrase-level information for the key works and the overall information for semantic representation. Our model get the best performance in all the models result. Its top-1 precision and recall have 1.6 and 2.0 promotion compared with the result of Bi-LSTM with attention mechanism(Bi-LSTM-AT).

E. Case Study

The CNN-LSTM attention model is used to predict unlabeled queries from online health communities. Some typical queries and predicting result are demonstrated in this sector to verify if the model could understanding the realistic medical queries.

"I have the diabetes, could I use the dimethylbiguanide?" / "糖尿病可以吃二甲双胍吗?"

<disease,medicine>	0.856	<disease,diet>	0.143
<disease,treatment>	0.001	<symptom,diet>	0.000

This short query described the user have diabetes, which is a common diseases. And what the user expected information is to judge if he/she can take the dimethylbiguanide(a kind of antidiabetes drug). Our model can predict the correct intent even the query contains some professional entity names.

F. Discussion

Our CNN-LSTM attention model extracts knowledge from text and makes use of professional knowledge for intent mining task, which gets high accurate prediction result on the raw text queries. Our model could distinguish the different of concepts for entities in sentences which are important for the prediction. Moreover, both the local phrase information and the interaction of concepts in the sentences can be detected by our model, which can be used to predict the intent in the queries.

VI. CONCLUSION

This paper proposed a novel query intent prediction model by using CNN-LSTM attention for user intent understanding. It contains the steps including mining the user intent taxonomy and labeling the dataset for training and validating the prediction model. In the experiment, we found the neural

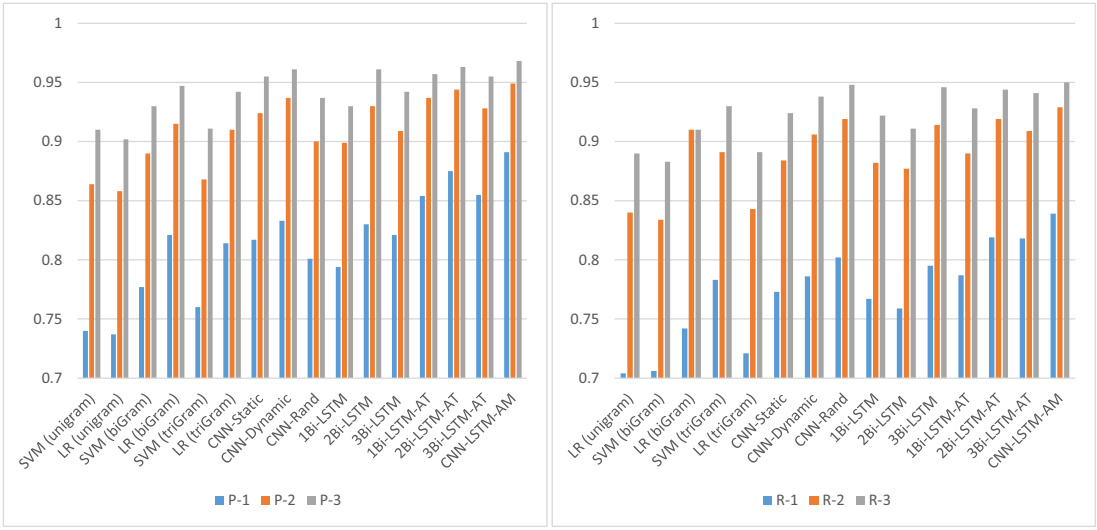

Fig. 4. Top-k Precision and Recall for user query intent prediction

network have great performance in our intent mining task. The CNN encoder effectively extracted local word combination features and the Bi-LSTM encoder used the orderly features to extract global intent features. Our CNN-LSTM attention model extracted knowledge from text and made use of professional knowledge to intent mining task, which got high accurate predict result for the original text queries.

REFERENCES

[1] Y. Li, C. Liu, N. Du, W. Fan, Q. Li, J. Gao, C. Zhang, and H. Wu, "Extracting medical knowledge from crowdsourced question answering website," *IEEE Transactions on Big Data*, 2016.

[2] Y. Yin, Y. Zhang, X. Liu, Y. Zhang, C. Xing, and H. Chen, "Healthqa: A chinese qa summary system for smart health," in *International Conference on Smart Health*. Springer, 2014, pp. 51–62.

[3] L. Nie, M. Wang, L. Zhang, S. Yan, B. Zhang, and T.-S. Chua, "Disease inference from health-related questions via sparse deep learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 8, pp. 2107–2119, 2015.

[4] J. W. Ely, J. A. Osheroff, P. N. Gorman, M. H. Ebell, M. L. Chambliss, E. A. Pifer, and P. Z. Stavri, "A taxonomy of generic clinical questions: classification study," *Bmj*, vol. 321, no. 7258, pp. 429–432, 2000.

[5] J. A. Shenson, E. Ingram, N. Colon, and G. P. Jackson, "Application of a consumer health information needs taxonomy to questions in maternal-fetal care," in *AMIA Annual Symposium Proceedings*, vol. 2015. American Medical Informatics Association, 2015, p. 1148.

[6] T. Zhang, J. H. Cho, and C. Zhai, "Understanding user intents in online health forums," in *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*. ACM, 2014, pp. 220–229.

[7] C. Zhang, W. Fan, N. Du, and P. S. Yu, "Mining user intentions from medical queries: A neural network based heterogeneous jointly modeling approach," in *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2016, pp. 1373–1384.

[8] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.

[9] X. Zhang and H. Wang, "A joint model of intent determination and slot filling for spoken language understanding," in *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI-2016)*. IJCAI, 2016.

[10] S. V. Ravuri and A. Stolcke, "Recurrent neural network and lstm models for lexical utterance classification." in *INTERSPEECH*, 2015, pp. 135–139.

[11] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997. [Online]. Available: https://arxiv.org/pdf/1603.01360

[12] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, "Neural architectures for named entity recognition," *arXiv preprint arXiv:1603.01360*, 2016.

[13] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[14] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proceedings of NAACL-HLT*, 2016, pp. 1480–1489. [Online]. Available: http://www.aclweb.org/anthology/N16-1174

[15] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in neural information processing systems*, 2013, pp. 3111–3119.

[16] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *Kdd*, vol. 96, no. 34, 1996, pp. 226–231.

[17] H.-P. Zhang, H.-K. Yu, D.-Y. Xiong, and Q. Liu, "Hhmm-based chinese lexical analyzer ictclas," in *Proceedings of the second SIGHAN workshop on Chinese language processing-Volume 17*. Association for Computational Linguistics, 2003, pp. 184–187.

[18] J. R. Finkel, T. Grenager, and C. Manning, "Incorporating non-local information into information extraction systems by gibbs sampling," in *Proceedings of the 43rd annual meeting on association for computational linguistics*. Association for Computational Linguistics, 2005, pp. 363–370.

[19] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," *Artificial Neural Networks–ICANN 2010*, pp. 92–101, 2010.

[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[21] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[22] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," *arXiv preprint arXiv:1404.2188*, 2014.

[23] M.-L. Zhang and Z.-H. Zhou, "A review on multi-label learning algorithms," *IEEE transactions on knowledge and data engineering*, vol. 26, no. 8, pp. 1819–1837, 2014.