Multi-modal Knowledge-aware Hierarchical Attention Network for Explainable Medical Question Answering

Yingying Zhang[1,2], Shengsheng Qian[1], Quan Fang[1], Changsheng Xu[1,2,3]
[1]National Lab of Pattern Recognition, Institude of Automation, CAS, Beijing 100190, China
[2]University of Chinese Chinese Academy of Sciences
[3]Peng Cheng Laboratory, ShenZhen, China
zhangyingying2017@ia.ac.cn,{shengsheng.qian,qfang,csxu}@nlpr.ia.ac.cn

ABSTRACT

Online healthcare services can offer public ubiquitous access to the medical knowledge, especially with the emergence of medical question answering websites, where patients can get in touch with doctors without going to hospital. Explainability and accuracy are two main concerns for medical question answering. However, existing methods mainly focus on accuracy and cannot provide a good explanation for retrieved medical answers. This paper proposes a novel *Multi-Modal Knowledge-aware Hierarchical Attention Network* (MKHAN) to effectively exploit multi-modal knowledge graph (MKG) for explainable medical question answering. MKHAN can generate path representation by composing the structural, linguistics, and visual information of entities, and infer the underlying rationale of question-answer interactions by leveraging the sequential dependencies within a path from MKG. Furthermore, a novel hierarchical attention network is proposed to discriminate the salience of paths endowing our model with explainability. We build a large-scale multi-modal medical knowledge graph and two real-world medical question answering datasets, the experimental results demonstrate the superior performance on our approach compared with the state-of-the-art methods.

CCS CONCEPTS

• **Information systems → Question answering**.

KEYWORDS

multi-modal knowledge graph representation, medical question answering, interpretability

ACM Reference Format:
Yingying Zhang, Shengsheng Qian, Quan Fang, Changsheng Xu. 2019. Multi-modal Knowledge-aware Hierarchical Attention Network for Explainable Medical Question Answering. In *Proceedings of the 27th ACM International Conference on Multimedia (MM'19), Oct. 21–25, 2019, Nice, France.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3343031.3351033

MM '19, October 21–25, 2019, Nice, France

ACM ISBN 978-1-4503-6889-6/19/10…$15.00
https://doi.org/10.1145/3343031.3351033

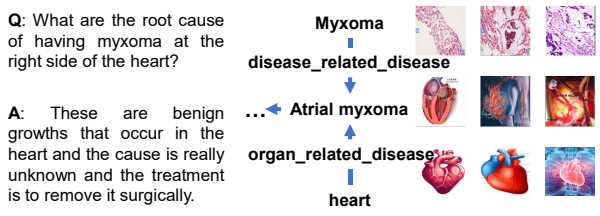


Figure 1: An illustrative example of a question-answer pair (left), and related entities with images (right).

1 INTRODUCTION

The escalating costs of healthcare and developing world-wide-web drive more consumers to spend time online to seek medical information. On medical question answering websites, the consultant can communicate with a professional doctor one-to-one, or look through other users' consulting histories. Figure 1 illustrates example. Appropriate explanations of medical answers may help health seekers adopt them, which can improve the trustworthiness and transparency of the medical question answering system. Therefore, explainability becomes critically important for medical question answering system to provide convincing results.

There are many powerful deep learning based question answering algorithms, which can be roughly categorized into two types: representation-based [32, 33] and interaction-based [24, 39]. However, they focus on improving the retrieval accuracy but seldom include reasonable explanations for retrieved answers to convince users. In the real scenario, doctors need to have profound domain knowledge to answer the medical questions. Prior efforts [4, 9, 18] demonstrate that the incorporation of knowledge graph (KG) can bridge the knowledge gap between experts and the public. For example, GRAM [4] and KAME [18] learned interpretable medical code representations consistent with the hierarchical ontology structures, and used these representations to predict diagnosis. KABLSTM [27] designed a context-guided attentive convolutional neural network to leverage knowledge embeddings to enrich the representations of questions and answers(QAs). Therefore, a great potential is expected to exploit KG for explainable medical question answering. However, those methods inject KG to enrich the representations of QAs, but ignore underlying rationales of question-answer pairs and multi-modal information [26]. In the medical question answering system, there is extra question-answer connectivity information derived from knowledge graph. As shown in Figure 1, as the question mentioned "myxoma" and "heart", the doctor might refer to

"atrial myxoma", then think of cause and treatment of this disease. Besides, there are multiple images that intuitively describe the appearances of this entity, which enrich the entities' representation with their hidden semantics. These relational paths tell the consultant why the doctor suggests "remove it surgically" with explicit reasons. This observation motivates us to extract the ontologies in the q/a and map them to entities in the KG find multi-hop relational paths between entities. Furthermore, high-quality representations of the extracted paths are expected to be learned for fully capturing the semantic meanings of entities and entity relations encoded in the multi-modal knowledge graph (MKG), which is essential to inject knowledge into the medical question answering system. Different paths reflect different diagnostic reasoning processes between medical ontologies, but some of them may not be bound with the question-answer context and doctor's reasoning logic. There are several entity mentions in the QAs, forming multiple entity pairs, each with multiple paths connecting them. "pain → tumor → swollen" and "pain → inflammation→ swollen" are two paths related to different diseases with the same symptoms. Therefore, how to model such complex connectivity between a question-answer pair to mimic the doctor's reasoning logic, is of critical importance to endow the medical question answer system with explainability.

In order to build an explainable medical question answering framework, we have to address the following challenging issues:

- *Challenge 1*: How to efficiently learn the representation of the paths connecting a question-answer pair based on the multi-modal medical knowledge graph? Injecting knowledge graph into question answering system can capture question-answer interactions. However, existing methods do not make full use of entity associations and multi-modal information.
- *Challenge 2*: How to discriminate the strengths of different paths to infer doctor's logic? Explainability is as important as accuracy for medical question answering. However, existing methods neglect the explanation for the retrieved answers.

To shed some light on these challenges, we propose a new solution, named *Multi-modal Knowledge-aware Hierarchical Attention Network* (MKHAN), which not only generates representations for paths by accounting for structural, linguistic and visual information of the entity but also performs reasoning based on paths to infer doctor's logic. To learn the rich path semantics (*Challenge 1*), we build a multi-modal knowledge graph and learn the representation of the entities within the path. Specifically, we retrieve the relevant images from the Internet for each entity, and apply a translation-based method to combine the structural, linguistic and visual information of each entity. To endow the explainability of the retrieved answers (*Challenge 2*), we apply a hierarchical attention network to discriminate the strengths of paths. Specifically, we use a self-attention mechanism to discriminate the paths related to each entity pair, then highlight the salient entity mentioned in question/answer guided by the question-answer context.

In summary, the contributions of this work are as follows:

- We are among the first to study on explainable medical question answering over multi-modal knowledge graph, which is challenging on both data acquisition and model design. We propose a novel *Multi-modal Knowledge-aware Hierarchical Attention Network* (MKHAN), which utilizes the domain knowledge to explain the retrieved answers. It interprets the proper answer by performing reasoning over multi-modal knowledge graph, and discriminating the strengths of paths with hierarchical attention network.
- We build a multi-modal medical knowledge graph which incorporates structural, linguistic and visual features to learn the representations of entities and relations.
- We build two large scale real-world Chinese medical question answering datasets to evaluate our model. Compared with the state-of-the-art methods, our approach not only improves the accuracy, but also holds the explanation capacity.

2 RELATED WORK

2.1 Medical-related Research

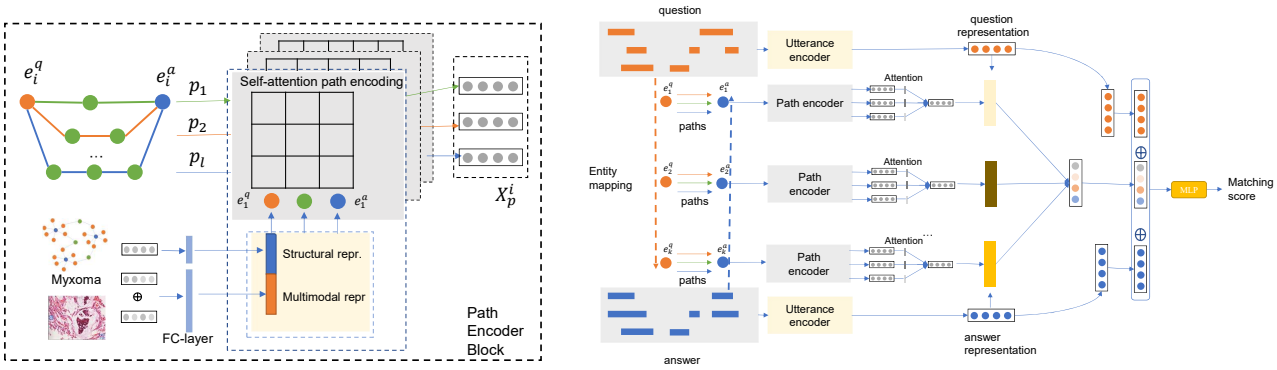
The existing approaches related to medical are diverse, including disease inference [10, 23], medical diagnosis [14, 16, 17, 30], and medical question answering [1, 8, 9]. DeepEHR [29] listed recent advances for electronic health record analysis, containing multi-outcome prediction [19], hospital re-admission prediction [21], EHR concept representation [3, 25], and suicide risk stratification [34].

However, most approaches were designed for well-organized structured data, e.g. EHR. In addition, some methods [8, 9] were based on TREC-CDS, where the topic description, summary and expected medical answer type were clearly labeled and accessible. Therefore, these approaches are not applicable to the online medical question-answering data that are unstructured multimedia content.

2.2 Knowledge Graph Representation

Translation-based methods, including TransE [2], TransH [36], TransD [12] and TransR [15] have achieved tremendous success on knowledge representation learning in the recent years. These approaches mapped the entities and relations into a low-dimensional vector space, and regarded relations as translations from head to tail entities. TransE [2] was under the assumption that the sum of the representations of head and relation is close to the tail's, denoted as $\boldsymbol{h} + \boldsymbol{r} \approx \boldsymbol{t}$, but it could not handle reflexive and one-to-many/many-to-many relations. To tackle this shortcoming, TransH [36] mapped head and tail to relation-specific hyperplane and then performed translation operation, while TransR [15] projected entities and relations into different vector spaces. Furthermore, TransD [12] used dynamic mapping matrices for learning the multiple representations of entities. There was also a growing research trending that incorporated multi-modal information into knowledge graph learning. IKRL [38] used image feature to enrich entity representation while [20] combined multi-model (visual and linguistic) with structural representations. Both of them [20, 38] were based on TransE and were not capable of reflexive relations.

However, the above methods either do not make full use of rich modality, or cannot cope with reflexive/one-to-many/many-to-many relations, such as "disease_related_disease". To this end, we propose an approach which can leverage rich information from multi-modal data, as well as handle the reflexive relations.



(a) Path representation based on multi-modal knowledge graph.

(b) Path Reasoning via Hierarchical attention network.

Figure 2: Schematic overview of our model architecture. (a) We extract paths connecting entities in questions and answers, then learn the representation of the path with self-attention. (b) We perform path reasoning via hierarchical attention network to discriminate the contribution of the paths and find explanation for the answers.

2.3 Interaction-based question answering

Retrieving and ranking answers from the vast corpus is an essential function for the medical question answering websites. Most existing approaches applied LSTM/CNN to learn the representation of QAs, then considered word-level interactions [6, 28, 40] or social communication [7] to yield better matching results. SMatrix [28] utilized a similarity matrix considering lexical as well as sequential information to model the complicated matching relations between QAs. K-NRM [39] used a translation matrix to model word-level similarities, a kernel-pooling layer to extract multi-level match features, and a learning-to-rank layer that produced the score.

However, these interaction-based methods are not explainable. In this paper, we come up with an approach that uses paths connecting the question-answer pairs to increase the interpretability for question-answer matching results. To the best of our knowledge, it is the first work that introduces the knowledge graph to provide explanation for retrieved medical answers.

3 METHOD

3.1 Basic Notations

We denote the question answering corpus as $\mathcal{D}$, while m/n is the length of the question q and answer a, and V is the vocabulary. A medical MKG is defined as $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{T})$, where $\mathcal{E}$ for entities and $\mathcal{R}$ for relations, and $\mathcal{T} = \{(h, r, t)|h, t \in \mathcal{E}, r \in \mathcal{R}\}$ is the KG triples.

In addition, to model the explainable interactions within the question-answer pair, we extract entities from question/answer as $\mathcal{S}^q$ and $\mathcal{S}^a$, and formally define the path connecting a question-answer pair as a sequence of entities within $\mathcal{G}$: $p = [e_1 \xrightarrow{r_1} e_2 \xrightarrow{r_2} \cdots \xrightarrow{r_{L-1}} e_L]$, where $e_1 \in \mathcal{S}_q$ and $e_L \in \mathcal{S}_a$, $e_1 \neq e_L$, r_i is the relation and L is the length of the path. We refer to paths that connect the ith entity pair as $\{p_1^i, p_2^i, \cdots, p_l^i\}$.

With the above notations, the inputs of the proposed MKHAN are the MKG $\mathcal{G}$, a question-answer pair, and paths connecting them. For each question-answer pair, we aim to predict the matching score, as well as explain the matching result.

3.2 Overall Framework

Our purpose is to calculate the matching score for a given question-answer pair. To this end, we present an approach called MKHAN. Figure 2 shows the workflow of MKHAN, which consists of two modules: path representation based on multi-modal knowledge graph and path reasoning via hierarchical attention network.

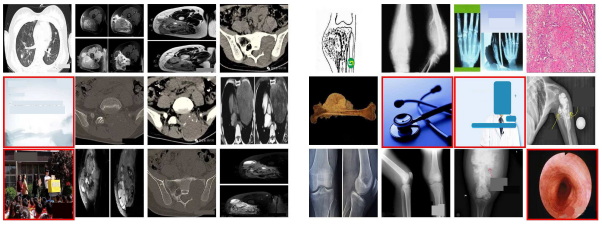
- *Path Representation based on Multi-Modal Knowledge Graph.* In this part, we build and learn the representation for medical MKG by integrating structural, linguistic, and visual features for entities and relations with a translation based method. Then we generate the path representation based on entity.
- *Path Reasoning via Hierarchical Attention Network.* This portion is designed to model the complex connectivity between a QA pair, then predict the score and give the explanation.

3.3 Path Representation based on Multi-modal Knowledge Graph

To learn the representations of the paths, we first build a medical MKG and learn the representation of the entities, then generate the path representations based on the entities.

3.3.1 Visual Modality Representation. There exists a medical knowledge graph that consists of medical ontologies and their relations, however, the visual information is not included. To enrich the knowledge graph with visual modality, we retrieve the top returned images from Google Image Search Engine [1] for each entity. We remove noisy images which are irrelevant to the corresponding entity, example images are in the Figure 3. We filter out the noisy image by its noisy score [22], which is calculated by adding up the distances of this image with all others retrieved by the same entity. We utilize the Euclidean metric as the distance measurement and the visual feature is a 2048-dim vector which is calculated using pretrained ResNet50. The image is filtered if its noise score is greater than a threshold, and the remained images set for ith entity denoted as $\mathcal{I}_i$.

[1] https://images.google.com/



(a) (b)

Figure 3: Results of noisy image filtering. The images with red border are filtered as noise.

Since most entities have more than one image, it is important to aggregate representation for each corresponding entity from multiple image instances. Simply summing up or averaging may lose the detailed information, thus we use the noisy score to determine the weight of each image. Hence, we define the aggregated image-based representation for the ith entity $\boldsymbol{e}_i^I$ as follows:

$$\boldsymbol{e}_i^I = \sum_{k=1}^{n_i} \frac{exp(-N_{ik}) \cdot \boldsymbol{e}_{ik}^I}{\sum_{k=1}^{n_i} exp(-N_{ik})}, \tag{1}$$

where N_{ik} and $\boldsymbol{e}_{ik}^I$ are the noisy score and visual feature of the kth image in the $\mathcal{I}_i$, n_i is the size of $\mathcal{I}_i$.

3.3.2 Multi-modal Knowledge Graph Representation Learning. Following the idea in TransH [36], we model a relation as a hyperplane together with a translation operation on it. We choose it rather than the TransE [2], because we need to deal with reflexive relations.

Given the structural feature $\boldsymbol{e}^S \in \mathbb{R}^d$, linguistic feature $\boldsymbol{e}^T \in \mathbb{R}^d$ and visual feature $\boldsymbol{e}^I \in \mathbb{R}^d$of an entity, where d is the dimension of the features, we first map them into a common space. We denote the mapped structural feature of head and tail as $\boldsymbol{h}^S$ and $\boldsymbol{t}^S$. For the multi-modal feature, we first combine the linguistic and visual features by concatenating them, then feed them into a fully-connected layer. We define the final multi-modal representation of head and tail as $\boldsymbol{h}^M$ and $\boldsymbol{t}^M$. When mapping to the hyperplane, the projection is defined as follows by restricting the hyperplane vector $\|\boldsymbol{w}_r\|_2 = 1$:

$$\begin{aligned} \boldsymbol{h}_\perp^S &= \boldsymbol{h}^S - \boldsymbol{w}_r^\top \boldsymbol{h}^S \boldsymbol{w}_r, \quad \boldsymbol{t}_\perp^S = \boldsymbol{t}^S - \boldsymbol{w}_r^\top \boldsymbol{t}^S \boldsymbol{w}_r, \\ \boldsymbol{h}_\perp^M &= \boldsymbol{h}^M - \boldsymbol{w}_r^\top \boldsymbol{h}^M \boldsymbol{w}_r, \quad \boldsymbol{t}_\perp^M = \boldsymbol{t}^M - \boldsymbol{w}_r^\top \boldsymbol{t}^M \boldsymbol{w}_r, \end{aligned} \tag{2}$$

then we define the projection of combined representation $\boldsymbol{h}_\perp^C$ and $\boldsymbol{t}_\perp^C$ by adding up the structural and multi-modal projected representation as follows:

$$\boldsymbol{h}_\perp^C = \boldsymbol{h}_\perp^S + \boldsymbol{h}_\perp^M, \quad \boldsymbol{t}_\perp^C = \boldsymbol{t}_\perp^S + \boldsymbol{t}_\perp^M \tag{3}$$

Next, we define the triple energies in consideration of both structural and multi-modal representation. Firstly, we can extend structural energy proposed in [36] by replacing the structural representation with multi-modal and combined representation, which enforces the translation operation of the structural/multi-modal/combined representation between the head and the tail:

$$\begin{aligned} E_S = \left\| \boldsymbol{h}_\perp^S + \boldsymbol{d}_r - \boldsymbol{t}_\perp^S \right\|, \quad E_M = \left\| \boldsymbol{h}_\perp^M + \boldsymbol{d}_r - \boldsymbol{t}_\perp^M \right\|, \\ E_C = \left\| \boldsymbol{h}_\perp^C + \boldsymbol{d}_r - \boldsymbol{t}_\perp^C \right\|. \end{aligned} \tag{4}$$

Then, we define the Structural-Multimodal energies [38] to ensure that the structural and the multi-modal representation are learned in the same space:

$$E_{MS} = \left\| \boldsymbol{h}_\perp^S + \boldsymbol{d}_r - \boldsymbol{t}_\perp^M \right\|, \quad E_{SM} = \left\| \boldsymbol{h}_\perp^M + \boldsymbol{d}_r - \boldsymbol{t}_\perp^S \right\|. \tag{5}$$

Considering all above, the final energy is defined as follows:

$$E(h, r, t) = E_S + E_M + E_C + E_{SM} + E_{MS} \tag{6}$$

Objective Function: Following the thought of [20], we employ head-centric and tail-centric loss jointly, so we define negative sample list for the head-centric and tail-centric view as follows:

$$\begin{aligned} \mathcal{T}'_{tail} &= \{(h, r, t') | h, t' \in \mathcal{E} \wedge (h, r, t') \notin \mathcal{T}\}, \\ \mathcal{T}'_{head} &= \{(h', r, t) | h', t \in \mathcal{E} \wedge (h', r, t) \notin \mathcal{T}\}. \end{aligned} \tag{7}$$

Therefore, we minimize a margin based on ranking loss between the energies of the gold and negative triples:

$$\begin{aligned} \mathcal{L}_{\text{head}} &= \sum_{(h,r,t)\in\mathcal{T},(h,r,t')\in\mathcal{T}'_{\text{tail}}} \max\left(\gamma + E(h, r, t) - E\left(h, r, t'\right), 0\right), \\ \mathcal{L}_{\text{tail}} &= \sum_{(h,r,t)\in\mathcal{T},(h',r,t)\in\mathcal{T}'_{\text{head}}} \max\left(\gamma + E(t, -r, h) - E\left(t, -r, h'\right), 0\right), \end{aligned} \tag{8}$$

where γ is the margin parameter. The final loss is the sum of head-view loss and tail-view loss.

$$\mathcal{L}_{kg} = \mathcal{L}_{\text{head}} + \mathcal{L}_{\text{tail}} \tag{9}$$

3.3.3 Self-Attention based Representation. Based on the multi-modal knowledge graph representation learning, we can define $\boldsymbol{E} \in \mathbb{R}^{|\mathcal{E}|\times 2d}$ by concatenating mapped structural and multi-modal representation of each entity. Thus each path can be represented as $\boldsymbol{X}^p \in \mathbb{R}^{L\times 2d}$ by concatenating node embeddings.

We employ self-attention technique to extract information from paths. Following [35], we define attention operation with the residual connection as follows:

$$Attention(X) = \text{softmax}(\frac{XW^Q(XW^K)^\top}{\sqrt{d}_k,})XW^V + X, \tag{10}$$

where W^Q, W^K, W^V are different transform matrices of the input X. Then we can combine the sequence information of the path as:

$$\boldsymbol{x_p} = \sum Attention(\boldsymbol{X}^p) \in \mathbb{R}^{2d}. \tag{11}$$

3.4 Path Reasoning via Hierarchical Attention Network

3.4.1 Utterance Encoder. All words in the q,a will be passed into an embedding layer, which looks up embedding for each word indices in the word embedding matrix $\boldsymbol{W} \in \mathbb{R}^{|V|\times d}$, where d is the dimension of embeddings. We denote $R(q) \in \mathbb{R}^{m\times d}$ and $R(a) \in \mathbb{R}^{n\times d}$ as the representation matrix as q and a. Similarly, we can generate the representation for q,a considering sequence information as:

$$\boldsymbol{x_q} = \sum Attention(R(q)), \quad \boldsymbol{x_a} = \sum Attention(R(a)), \tag{12}$$

3.4.2 Path-based Context Embedding. The paths imitate the doctor's reasoning process when answering the question, and can be regarded as contexts and interactions between a question-answer pair. Since there might exist several paths, we propose a novel hierarchical attention mechanism to discriminate the contributions of different paths.

Firstly, given l paths connecting the ith entity pair, we first compute their path representation $X_p^i \in \mathbb{R}^{l \times 2d}$ according to Section 3.3.3. Then we compute their self-attention based encoding and sum up the l paths' representation. $\boldsymbol{x}_{p^i}$ is the combined representation of the paths connecting the ith entity pair, and we define the combined paths' set as $\mathcal{S}_{q \to a}$:

$$\boldsymbol{x}_{\boldsymbol{p}^i} = \sum Attention(X_p^i) \tag{13}$$

Secondly, we calculate the attention between paths and question-answer pair. We learn the attention weights over paths conditioned on the involved question-answer pair, and adopt a two-layer feed-forward architecture. For the ith entity pair's path p^i, the attention can be calculated by:

$$\alpha^{(1)}_{q,p^i,a} = f(W^{(1)}[\boldsymbol{x}_q \oplus \boldsymbol{x}_{p^i} \oplus \boldsymbol{x}_a] + b^{(1)}), \tag{14}$$

$$\alpha^{(2)}_{q,p^i,a} = f(W^{(2)}\alpha^{(1)}_{q,p^i,a} + b^{(2)}), \tag{15}$$

where $f(\cdot)$ is ReLU activation function, $W^{(1)}/W^{(2)}$ and $b^{(1)}/b^{(2)}$ are the wights and biases, "$\oplus$" denotes the concatenation operation, then the attention weights are normalized by a softmax function,

$$\alpha_{q,p^i,a} = \frac{\exp\left(\alpha^{(2)}_{q,p^i,a}\right)}{\sum_{p^i \in \mathcal{S}_{q \to a}} \exp\left(\alpha^{(2)}_{q,p^i,a}\right)}. \tag{16}$$

Finally, we aggregate all these interactions by a weighted sum:

$$\boldsymbol{c}_{q \to a} = \sum_{p^i \in \mathcal{S}_{q \to a}} \alpha_{q,p^i,a} \cdot \boldsymbol{x}_{p^i}, \tag{17}$$

where $\boldsymbol{c}_{q \to a}$ combines path representations based on their contributions to reveal the reasoning process. Therefore, our model can infer the rationales underlying the question-answer pair to explain the retrieved answer.

3.4.3 Matching Schema. Until now, given a question-answer pair, we have the embedding of question q, answer a, and the explainable paths connecting them. We combine these embedding vectors into a jointed representation of the question-answer pair as below:

$$\boldsymbol{x}_{q,a} = \boldsymbol{x}_q \oplus \boldsymbol{c}_{q \to a} \oplus \boldsymbol{x}_a, \tag{18}$$

$\boldsymbol{x}_{q,a}$ encodes the information from three aspects: the involved question, answer, and the corresponding paths based context. We feed $\boldsymbol{x}_{q,a}$ into an MLP layer in order to predict the matching score:

$$score(q, a) = MLP(\boldsymbol{x}_{q,a}). \tag{19}$$

Objective Function: We design a pair-wise learning to rank loss function to learn the parameters:

$$\mathcal{L}_{rank} = \sum_{(q,a^+,a^-) \in \mathcal{D}} max(1 - score(q, a^+) + score(q, a^-), 0), \tag{20}$$

where a^+ is the relevant answer and a^- is the irrelevant with respect to q. Both the relevant and irrelevant answers are sampled from the same domain (clinics) with the question.

Table 1: Statistics of the Medical MKG.

MKG	#Type	#Rel	#Ent	#Triples	#Images
Count	6	17	59882	599223	741935

3.5 Optimization

We update the parameters of knowledge graph learning and question-answer matching iteratively. Specifically, given a question-answer pair, we first find the entities according to their utterances, and then update knowledge graph learning's parameters with the relative entities. After that we extract the paths connecting the question-answer pair, and update parameters in question answering matching network. The updating procedure is shown in Algorithm 1.

Algorithm 1 Training Algorithm.

Input: Question answering corpus $\mathcal{D}$, MKG $\mathcal{G}$.
Output: Matching function $\mathcal{F}(q, a|\Theta, \mathcal{D}, \mathcal{G})$.

```
Initialize all parameters;
for number of training iteration do
    Sample minibatch of positive and negative question answer pairs from corpus D;
    Extract the entities in the questions and answers as S(h);
    Sample true and false triples from G for each entities in the S(h);
    Update parameters according to Equation 9;
    Find the paths between entities in the question-answer pair;
    Update parameters according to Equation 20.
end for
```

4 EXPERIMENTS

4.1 Data Preparation

4.1.1 Knowledge Graph. We use the open medical knowledge graph Symptom-in-Chinese[1] in the experiments, which is collected and built from eight medical question answering consult websites, three Chinese encyclopedias and several electronic medical health records. We choose six most related types of entity and seventeen relations between them. We then retrieve several images for each entity and build a multi-modal medical knowledge graph. The statistics of the medical multi-modal knowledge graph is listed in Table 1.

4.1.2 Dataset. We build two real-world datasets from two popular medical question answering websites: Dingxiang Doctor [2] and Chunyu Doctor [3], where qualified doctors give some helpful advices to the consultants. Table 2 shows the statistics of the two datasets, where the Avg Q-Len and Avg A-Len denote the average words in the QAs, respectively. We preprocess QAs by removing the punctuation and cutting them into words with Chinese text segmentation tool Jieba[4].

[1] http://openkg.cn/dataset/symptom-in-chinese
[2] https //ask.dxy.com/
[3] https //www.chunyuyisheng.com/pc/qalist/
[4] https://github.com/fxsjy/jieba

Table 2: Statistics of Medical Question Answering Datasets.

Dataset	QA pairs	Avg. Q-Len	Avg A-Len	Clinics
Chunyu	245085	32	75	16
Dingxiang	273003	115	284	25

4.1.3 Path Extraction. In practice, it is infeasible and labor intensive to find all connected paths over the knowledge graph. The number of paths grows exponentially with regards to the length of the path, as suggested in [5, 31]. Longer paths not only bring more edges between two entities but also introduce more noisy entities. Moreover, as pointed out by [5], performance dropped significantly when the number of hops increased from 2 to 3. Therefore, we extract all qualified paths by deep-first search algorithm, each with length up to 3.

4.2 Comparing Methods

The experiments on the two Chinese medical question answering datasets employ the following comparing methods: (1)*BOW*, *Doc2Vec* [13], which are simple but efficient representations used in NLP. We use the cosine similarity as the matching score. (2)*SMatrix* [28], which introduces the similarity matrix based architecture to model the complicated relations between questions and answers. (3)*K-NRM* [39], which is a state-of-the-art deep semantic matching approach. It uses a translation matrix to model word-level similarities, and employs a kernel-pooling technique to extract multi-level soft match features. Finally, those features are fed to a learning-to-rank layer to predict the ranking score. (4)*KABLSTM* [27], which incorporates knowledge graph into question answering by leveraging knowledge embeddings to enrich the representation of questions and answers.

In addition, we use a variant of MKHAN as a baseline: *MKHAN-NM*: MKHAN-NM is a variant of MKHAN without using the multi-modal information of the entities, thus the comparison between other baselines and MKHAN-NM can reflect the influences of incorporating the explainable paths. The comparison between MKHAN-NM and MKHAN can reflect how multi-modal information benefits the matching qualities.

4.3 Evaluation Metrics and Parameter Setting

Precision and normalized discounted cumulative gain (nDCG [11]) are used as the evaluation metrics. Precision is the average number of times that the correct answer has the highest matching score. The nDCG score is a standard metric for evaluating rankings. It penalizes each rating based on its position in the results and then normalizes the gain with the ideal discounted cumulative gain. Since Precision and nDCG require a global scan of all answers for each question, which is time-consuming, we follow [28, 37] and use a variant of Precision and nDCG based on sampling. For each question in the test data, we keep one ground truth and sample F fake answers from the same clinics as candidate answers. Then we evaluate the Precision and nDCG on each candidate set and take the average score as the final results.

As for parameter setting, we set the dimension of words, images, and entities embedding to 150 for all methods. For MKHAN, the length of the path is up to 3. We test the negative samples F for 5 and 19, separately. For each dataset, we use $p\%$ of the data for training and the remaining $(1-p)\%$ for evaluation. $p\%$ varies from 30% to 70%. We evaluate the model every 400 batches and stop training when the valid precision does not increase for 5 rounds.

4.4 Experimental Results and Analysis

4.4.1 Quantitative Results. We list the experimental results in Table 3–4, From these results, we have the following observations:

(1) The performance of approach BOW is stable under various training sizes, while other approaches' performances improve when adding more training data. This shows the representation-based methods make full use of training data.
(2) SMatrix and K-NRM achieve better results than BOW and Doc2Vec on both datasets, showing that the interactions of the QA pairs can provide valuable information to improve the qualities of matching with a well designed neural model.
(3) KABLSTM beats SMatrix and K-NRM on Chunyu dataset and achieves high precision on Dingxiang dataset, verifying the incorporation of KG is useful in the medical domain.
(4) MKHAN consistently achieves the best performance, demonstrating that our proposed multi-modal knowledge-aware hierarchical attention network is effective in question-answer matching. Moreover, it beats the variant MKHAN-NM in most cases, verifying the incorporation of multi-modal information is beneficial for medical question answering.

4.4.2 Attention and Interpretability. We analyze the attention within the paths of the entity pair and among the entities of a question-answer pair. Figure 4 demonstrates an example of a question-answer pair, paths, and attention within them. The question-answer pair is that a consultant has noticed some abnormal examination results and suspected that she/he had the tumor, and the doctor gave her/him advices. The words with the yellow or blue background can be mapped into entities in the medical knowledge graph. Blocks in the center show the path from entities of the question to the answer, with the attention weights on the path or the block (bold). The paths here can be viewed as potential reasoning process of the doctor according to the description of the consultant. The attention between the entity pairs and question-answer pair can be viewed as a selection for entity pairs conditioned on the question-answer context. As shown in Figure 4, The entity pair "(tumor markers, color ultrasound)" gains the highest attention score (0.3225) among all entity pairs, which is identical with expectation, since it is most related to the "tumor" topic of the question-answer pair. The paths connecting one entity pair can be viewed as the possible reasoning process, and attention weights can be explained as a possibility. For example, the path "runny nose → toxic pneumon → chest" gains the highest attention for "(runny noise, chest)", which reveals the rationale underlying the question-answer pair, and can explain why the doctor suggested checking for chest.

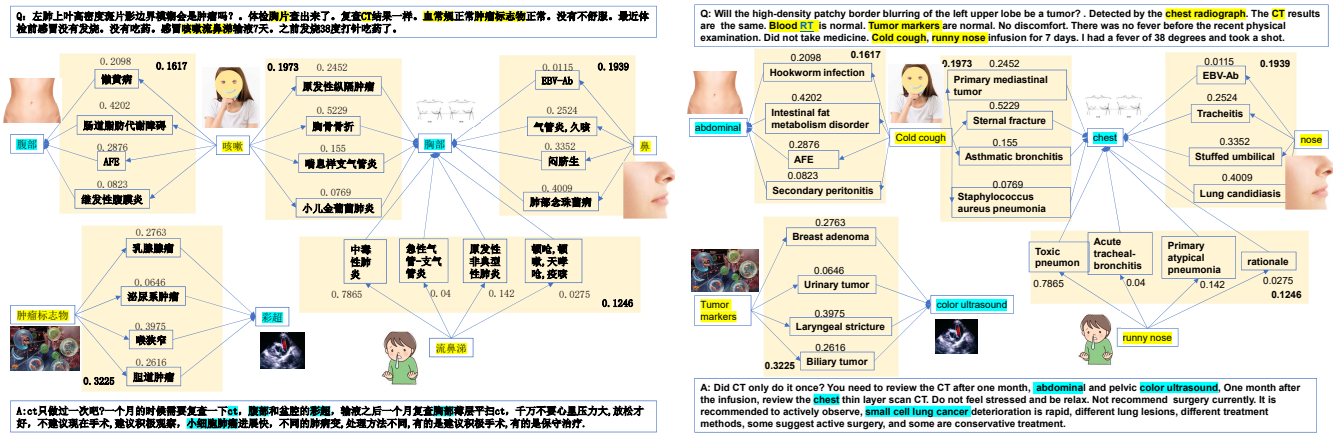
4.4.3 Case Study. We retrieve the top 6 answers to the question "What kind of medicine does lung cancer patient take to protect liver". Figure 5 shows the ranking results by K-NRM and MKHAN. The first row is the gold answer. As shown in Figure 5, MKHAN gives the gold answer higher rank comparing with K-NRM. It shows

Table 3: nDCG and Precision on Chunyu dataset.

	nDCG (F=5)			Precision (F=5)			nDCG (F=19)			Precision (F=19)		
p%	30%	50%	70%	30%	50%	70%	30%	50%	70%	30%	50%	70%
BOW	0.6950	0.6864	0.6812	0.4146	0.4001	0.3940	0.5175	0.5160	0.5173	0.2434	0.2401	0.2432
Doc2Vec	0.7170	0.7187	0.7281	0.4271	0.4290	0.4463	0.4996	0.5060	0.5157	0.1920	0.2053	0.2141
SMatrix	0.7756	0.7959	0.7984	0.5370	0.5695	0.5758	0.5904	0.6201	0.6321	0.3175	0.3541	0.3683
K-NRM	0.7325	0.7795	0.7294	0.4606	0.5432	0.4594	0.5302	0.5879	0.5898	0.2512	0.3205	0.3279
KABLSTM	0.8464	0.8443	0.8511	0.6653	0.6624	0.6744	0.7068	0.6974	0.7038	0.4710	0.4601	0.4688
MKHAN-NM	0.8612	0.8715	0.8770	0.6885	0.7076	0.7190	0.7097	0.7257	0.7222	0.4595	0.4826	0.4756
MKHAN	**0.8665**	**0.8756**	**0.8798**	**0.7021**	**0.7169**	**0.7251**	**0.7143**	**0.7303**	**0.7419**	**0.4656**	**0.4894**	**0.5086**

Table 4: nDCG and Precision on Dingxiang dataset.

	nDCG (F=5)			Precision (F=5)			nDCG (F=19)			Precision (F=19)		
p%	30%	50%	70%	30%	50%	70%	30%	50%	70%	30%	50%	70%
BOW	0.7945	0.7951	0.7961	0.5826	0.5831	0.5850	0.6500	0.6477	0.6502	0.4120	0.4085	0.4108
Doc2Vec	0.8602	0.8707	0.8928	0.6975	0.7183	0.7650	0.6241	0.6423	0.6950	0.3534	0.3841	0.4601
SMatrix	0.8436	0.8814	0.8883	0.6645	0.7382	0.7506	0.7279	0.7591	0.7905	0.5120	0.5500	0.5986
K-NRM	0.9239	0.9298	0.9270	0.8294	0.8420	0.8363	0.8302	0.8542	0.8599	0.6798	0.7181	0.7282
KABLSTM	0.9232	0.9232	0.9233	0.8198	0.8240	0.8222	0.8207	0.8315	0.8282	0.6451	0.6630	0.6592
MKHAN-NM	0.9395	0.9429	**0.9466**	0.8542	0.8606	**0.8688**	0.8516	0.8753	0.8613	0.6866	0.7300	0.7013
MKHAN	**0.9448**	**0.9484**	0.9400	**0.8642**	**0.8749**	0.8543	**0.8610**	**0.8796**	**0.8763**	**0.7039**	**0.7316**	**0.7374**



(a) Original Chinese question-answer pair.

(b) Question-answer pair translated to English.

Figure 4: Example of question-answer pair and paths. The question proposed by the consultant is on the top, and bottom is the doctor's answer. The words with the yellow or blue background can be mapped into entities of the medical knowledge graph. In the center, there are the paths connecting question-answer pair with attention weight.

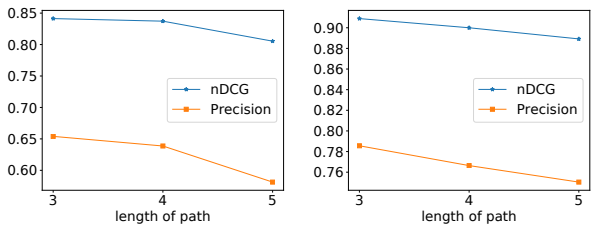
that it is useful to introduce multi-modal knowledge graph for medical question answering.

4.4.4 Parameter Analysis. To find out how the length of paths influences the result of matching, we test the length of 3, 4 and 5 in the experiment. We take 30% of ordinary data for experiment and use 70% for training and the last for validation and test. The results are shown in Figure 6. We can easily find out that the longer path does not improve the evaluation results. It can be attributed to the noise introduced by longer path.

肺癌病人护肝吃什么药?　　What kind of medicine does lung cancer patient take to protect liver?

K-NRM	MKHAN	Chinese	English
2	1	保肝药很多,吃吉非替尼,建议定期复查肝功，如果肝功能损害严重,应考虑停药，如果肝功没有异常,不是一定需要吃保肝药的。	Many liver-protecting drugs, gefitinib, it is recommended to review liver function regularly, if liver function damage is serious, should consider stopping the drug, if the liver function is not abnormal, it is not necessary to eat liver-protecting drugs.
5	3	血栓就是这样,吃着抗凝药,要溶开很慢的。化疗还需综合评估吧，血栓就怕脱落，血栓化疗有风险。	The thrombus is like this. If you eat anticoagulant, it should be dissolved slowly. Chemotherapy needs to be comprehensively evaluated. Thrombosis is dangerous。Thrombosis chemotherapy is risky.
4	5	纵隔型淋巴瘤容易出现头颈部浮肿，如果不是肿瘤性的浮肿,三天内就会消肿。一般摔倒后不会引起脖子浮肿,往往与肿瘤性压迫有关,所以有可能只有抗肿瘤治疗后才能消肿，淋巴瘤的主要治疗是化疗药结合局部放疗。	Mediastinal lymphoma is prone to head and neck edema, if it is not a tumor edema, it will be swollen within three days. Generally, it will not cause neck edema after falling, and it is often related to tumor compression, so it may only be able to reduce swelling after anti-tumor treatment. The main treatment of lymphoma is chemotherapy combined with local radiotherapy.
3	6	还是过去查一下。囊肿本身是空心的,有时候里面的液体物质张力太大,摸上去也会觉得硬，检查一下放心，如果有问题就得考虑手术。	You shall go to check it. The cyst itself is hollow. Sometimes the liquid material inside is too tight, and it will feel hard when touched. If there is a problem, you must consider the operation.
6	2	肠梗阻不适合化疗,先把身体状况调理好了,积极纠正一般状况，争取化疗吧,手术不太适合了,肺上也有了转移灶，还是胃肠间质瘤。可以试试伊马替尼(格列卫）。	Intestinal obstruction is not suitable for chemotherapy. First adjust the physical condition, actively correct the general condition, surgery is not suitable, because of lung metastasis, or gastrointestinal stromal tumor. You can try imatinib (Gleevec), to control the tumor and save lives.
1	4	十二直肠癌预后较差,术后也需配合放化疗等综合治疗，如需化疗肯定要住院治疗。	It is also necessary to cooperate with radiotherapy and chemotherapy after surgery. If chemotherapy is required, hospitalization is necessary.

Figure 5: Top-6 retrieval results over Chunyu dataset by the K-NRM and the proposed MKHAN.



(a) Result on Chunyu dataset.　　(b) Result on Dingxiang dataset.

Figure 6: The nDCG and Precision with different lengths of paths.

4.4.5 Results on Knowledge Graph Embedding Side. Although we show that MKHAN can benefit retrieval precision as well as interpretability, it is still essential to verify whether our multi-modal knowledge graph learning mechanism can improve the knowledge graph representation. However, since the knowledge graph is partially trained when learned together with question answering module, it is hard to evaluate the influence of multi-modal content to the knowledge representation. Therefore, we compare our knowledge module individually with other knowledge representation methods. We compare our approach with TransH [36], which only uses structural information, and approach from [20] (Base), which is a TransE-based multi-modal knowledge graph learning method. Table 5 shows the results on triple classification with 70% data for training. The test data consists of 50% real triples and 50% fake triples. All three methods are trained with head and tail view loss. We evaluate the seventeen types of relations individually, and calculate average accuracy, standard deviation and maximum accuracy among all relations. We can see that our knowledge representation approach outperforms all other models, verifying that our multi-modal knowledge graph learning is effective.

Table 5: Triple Classification Accuracy.

Method	Avg±std	Max
TransH	0.7418 ± 0.11	0.9155
Base	0.6449 ± 0.07	0.7915
Ours	**0.7685±0.13**	**0.9459**

5 CONCLUSIONS

In this paper, we propose an end-to-end framework with multi-modal medical knowledge graph to automatically model explainable interactions between the medical question-answer pairs. We build a multi-modal knowledge graph and learn the representations of the entities within the path to learn the rich path semantics, and apply a hierarchical attention network to endow the explainability for the retrieved answers. Its favorable performance verifies the effectiveness of our method compared with others. In the future, we will consider to integrate social information in our framework such as doctors' profiles, answer histories, to improve the reliability of the retrieved answers.

ACKNOWLEDGMENTS

This work was supported in part by the National Key Research and Development Program of China (No. 2017YFB1002804), the National Natural Science Foundation of China under Grants 61432019, 61702509, 61802405, 61832002, 61572503, 61872424 and 61720106006, the Key Research Program of Frontier Sciences, CAS, Grant NO. QYZDJ-SSW-JSC039, the Beijing Municipal Science & Technology Commission (No. Z181100008918012) and the K.C.Wong Education Foundation.

REFERENCES

[1] Asma Ben Abacha and Pierre Zweigenbaum. 2015. MEANS: A medical question-answering system combining NLP techniques and semantic Web technologies. *Inf. Process. Manage.* 51 (2015), 570–594.

[2] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Modeling Multi-relational Data. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'13)*. Curran Associates Inc., USA, 2787–2795.

[3] Edward Choi, Mohammad Taha Bahadori, Elizabeth Searles, Catherine Coffey, Michael Thompson, James Bost, Javier Tejedor-Sojo, and Jimeng Sun. 2016. Multi-layer Representation Learning for Medical Concepts. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. ACM, New York, NY, USA, 1495–1504.

[4] Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F. Stewart, and Jimeng Sun. 2017. GRAM: Graph-based Attention Model for Healthcare Representation Learning. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17)*. ACM, New York, NY, USA, 787–795.

[5] Wanyun Cui, Yanghua Xiao, Haixun Wang, Yangqiu Song, Seung won Hwang, and Wei Yang Wang. 2017. KBQA: Learning Question Answering over QA Corpora and Knowledge Bases. *Proceedings of the VLDB Endowment* 10 (2017), 565–576.

[6] Jun da Hu, Shengsheng Qian, Quan Fang, and Changsheng Xu. 2018. Attentive Interactive Convolutional Matching for Community Question Answering in Social Multimedia. In *ACM International Conference on Multimedia*.

[7] Hanyin Fang, Fei Wu, Zhou Zhao, Xinyu Duan, Yueting Zhuang, and Martin Ester. 2016. Community-based Question Answering via Heterogeneous Social Network Learning. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI'16)*. AAAI Press, 122–128.

[8] Travis R. Goodwin and Sanda M. Harabagiu. 2016. Medical Question Answering for Clinical Decision Support. (2016), 297–306.

[9] Travis R. Goodwin and Sanda M. Harabagiu. 2017. Knowledge Representations and Inference Techniques for Medical Question Answering. *ACM Transactions on Intelligent Systems and Technology* 9 (2017), 14:1–14:26.

[10] Donglin Guo, Min Li, Ying Yu, Yaohang Li, Guihua Duan, Fang-Xiang Wu, and Jianxin Wang. 2018. Disease Inference with Symptom Extraction and Bidirectional Recurrent Neural Network. *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (2018), 864–868.

[11] Kalervo Järvelin and Jaana Kekäläinen. 2000. IR Evaluation Methods for Retrieving Highly Relevant Documents. (2000), 41–48.

[12] Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Knowledge Graph Embedding via Dynamic Mapping Matrix. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)"*. Association for Computational Linguistics, Beijing, China, 687–696.

[13] Quoc Le and Tomas Mikolov. 2014. Distributed Representations of Sentences and Documents. In *Proceedings of the 31st International Conference on Machine Learning (Proceedings of Machine Learning Research)*, Eric P. Xing and Tony Jebara (Eds.), Vol. 32. PMLR, Bejing, China, 1188–1196.

[14] Yaliang Li, Nan Du, Chaochun Liu, Yusheng Xie, Wei Fan, Qi Li, Jing Gao, and Huan Sun. 2017. Reliable Medical Diagnosis from Crowdsourcing: Discover Trustworthy Answers from Non-Experts. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining (WSDM '17)*. ACM, New York, NY, USA, 253–261.

[15] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning Entity and Relation Embeddings for Knowledge Graph Completion. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI'15)*. AAAI Press, 2181–2187.

[16] Siqi Liu, Sidong Liu, Tom Weidong Cai, Hangyu Che, Sonia Pujol, Ron Kikinis, David Dagan Feng, and Michael J. Fulham. 2015. Multimodal Neuroimaging Feature Learning for Multiclass Diagnosis of Alzheimer's Disease. *IEEE Transactions on Biomedical Engineering* 62 (2015), 1132–1140.

[17] Fenglong Ma, Yaqing Wang, Houping Xiao, Ye Yuan, Radha Chitta, Jing Zhou, and Jing Gao. 2018. A General Framework for Diagnosis Prediction via Incorporating Medical Code Descriptions. *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (2018), 1070–1075.

[18] Fenglong Ma, Quanzeng You, Houping Xiao, Radha Chitta, Jing Zhou, and Jing Gao. 2018. KAME: Knowledge-based Attention Model for Diagnosis Prediction in Healthcare. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM '18)*. ACM, New York, NY, USA, 743–752.

[19] Riccardo Miotto, Li Li, Brian A. Kidd, and Joel Dudley. 2016. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. In *Scientific reports*.

[20] Hatem Mousselly Sergieh, Teresa Botschen, Iryna Gurevych, and Stefan Roth. 2018. A Multimodal Translation-Based Approach for Knowledge Graph Representation Learning. In *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics*. Association for Computational Linguistics, New Orleans, Louisiana, 225–234.

[21] Phuoc Nguyen, Truyen Tran, Nilmini Wickramasinghe, and Svetha Venkatesh. 2017. Deepr: A Convolutional Net for Medical Records. *IEEE Journal of Biomedical and Health Informatics* 21 (2017), 22–30.

[22] Fudong Nian, Bing-Kun Bao, Teng Li, and Changsheng Xu. 2017. Multi-Modal Knowledge Representation Learning via Webly-Supervised Relationships Mining. In *Proceedings of the 25th ACM International Conference on Multimedia (MM '17)*. ACM, New York, NY, USA, 411–419.

[23] L. Nie, M. Wang, L. Zhang, S. Yan, B. Zhang, and T. Chua. 2015. Disease Inference from Health-Related Questions via Sparse Deep Learning. *IEEE Transactions on Knowledge and Data Engineering* 27, 8 (Aug 2015), 2107–2119.

[24] Liang Pang, Yanyan Lan, Jiafeng Guo, Jun Xu, Shengxian Wan, and Xueqi Cheng. 2016. Text Matching As Image Recognition. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI'16)*. AAAI Press, 2793–2799.

[25] Trang Pham, Truyen Tran, Dinh Phung, and Svetha Venkatesh. 2016. DeepCare: A Deep Dynamic Memory Model forźPredictive Medicine. In *Proceedings, Part II, of the 20th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining - Volume 9652 (PAKDD 2016)*. Springer-Verlag, Berlin, Heidelberg, 30–41.

[26] S. Qian, T. Zhang, C. Xu, and J. Shao. 2016. Multi-Modal Event Topic Model for Social Event Analysis. *IEEE Transactions on Multimedia* 18, 2 (Feb 2016), 233–246. https://doi.org/10.1109/TMM.2015.2510329

[27] Ying Shen, Yang Deng, Min Yang, Yaliang Li, Nan Du, Wei Fan, and Kai Lei. 2018. Knowledge-aware Attentive Neural Network for Ranking Question Answer Pairs. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18)*. ACM, New York, NY, USA, 901–904.

[28] Yikang Shen, Wenge Rong, Zhiwei Sun, Yuanxin Ouyang, and Zhang Xiong. 2015. Question/Answer Matching for CQA System via Combining Lexical and Sequential Information. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI'15)*. AAAI Press, 275–281.

[29] Benjamin Shickel, Patrick James Tighe, Azra Bihorac, and Parisa Rashidi. 2018. Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. *IEEE Journal of Biomedical and Health Informatics* 22 (2018), 1589–1604.

[30] Heung-Il Suk, Seong-Whan Lee, and Dinggang Shen. 2014. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage* 101 (2014), 569–582.

[31] Zhu Sun, Jie Yang, Jie Zhang, Alessandro Bozzon, Long-Kai Huang, and Chi Xu. 2018. Recurrent Knowledge Graph Embedding for Effective Recommendation. In *Proceedings of the 12th ACM Conference on Recommender Systems (RecSys '18)*. ACM, New York, NY, USA, 297–305.

[32] Yi Tay, Minh C. Phan, Luu Anh Tuan, and Siu Cheung Hui. 2017. Learning to Rank Question Answer Pairs with Holographic Dual LSTM Architecture. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17)*. ACM, New York, NY, USA, 695–704.

[33] Yi Tay, Luu Anh Tuan, and Siu Cheung Hui. 2018. Cross Temporal Recurrent Networks for Ranking Question Answer Pairs. In *Proceedings of the 32rd AAAI Conference on Artificial Intelligence*.

[34] Truyen Tran, Tu Dinh Nguyen, Dinh Phung, and Svetha Venkatesh. 2015. Learning Vector Representation of Medical Objects via EMR-driven Nonnegative Restricted Boltzmann Machines (eNRBM). *J. of Biomedical Informatics* 54, C (April 2015), 96–105.

[35] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. In *NIPS*.

[36] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge Graph Embedding by Translating on Hyperplanes. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI'14)*. AAAI Press, 1112–1119.

[37] Yu Ping Wu, Wei Chung Wu, Can Xu, and Zhoujun Li. 2018. Proceedings of the AAAI Conference on Artificial Intelligence. In *AAAI*.

[38] Ruobing Xie, Zhiyuan Liu, Huanbo Luan, and Maosong Sun. 2017. Image-embodied Knowledge Representation Learning. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*. 3140–3146.

[39] Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17)*. ACM, New York, NY, USA, 55–64.

[40] Xiaodong Zhang, Sujian Li, Lei Sha, and Houfeng Wang. 2017. Attentive Interactive Neural Networks for Answer Selection in Community Question Answering. In *Proceedings of the 31th AAAI Conference on Artificial Intelligence*.