RDKG: A Reinforcement Learning Framework for Disease Diagnosis on Knowledge Graph

Shipeng Guo[1,2], Kunpeng Liu[3], Pengfei Wang[1,2], Weiwei Dai[4], Yi Du[1,2],
Yuanchun Zhou[1,2], Wenjuan Cui[1,2,*]
[1]*Computer Network Information Center, Chinese Academy of Sciences, Beijing, China*
[2]*University of Chinese Academy of Sciences, Beijing, China*
[3]*Portland State University, Oregon, US*
[4]*Changsha Aier Eye Hospital, Changsha, China*
{spguo, pfwang,duyi,zyc,wenjuancui}@cnic.cn, kunpeng@pdx.edu, daiweiwei@aierchina.com

2023 IEEE International Conference on Data Mining (ICDM) | 979-8-3503-0788-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICDM58522.2023.00122

***Abstract*—Automatic disease diagnosis from symptoms has attracted much attention in medical practices. It can assist doctors and medical practitioners in narrowing down disease candidates, reducing testing costs, improving diagnosis efficiency, and more importantly, saving human lives. Existing research has made significant progress in diagnosing disease but was limited by the gap between interpretability and accuracy. To fill this gap, in this paper, we propose a method called Reinforced Disease Diagnosis on Knowlege Graph (RDKG). Specifically, we first construct a knowledge graph containing all information from electronic medical records. To capture informative embeddings, we propose an enhanced knowledge graph embedding method that can embed information outside the knowledge graph into entity embedding. Then we transform the automatic disease diagnosis task into a Markov decision process on the knowledge graph. After that, we design a reinforcement learning method with a soft reward mechanism and a pruning strategy to solve the Markov decision process. We accomplish automated disease diagnosis by finding a path from symptoms to disease. The experimental results show that our model can effectively utilize heterogeneous information in the knowledge graph to complete the automatic disease diagnosis. Besides, our model demonstrates supreme performance in both accuracy and interpretability.**

***Index Terms*—Electronic Medical Records, medical knowledge graph, reinforcement learning, automatic disease diagnosis.**

I. Introduction

Automatic disease diagnosis is a significant research area in healthcare management, particularly highlighted during the COVID-19 pandemic. Machine learning has been successfully applied to automate disease diagnosis [1]–[4], but current methods often only focus on predicting diseases, overlooking the interpretability of the diagnostic process, which limits the practical use of existing methods [5]. Thus, we pose a question: **Can we build an ML framework that efficiently predicts diseases while maintaining interpretability?**

In recent years, knowledge graphs have gained attention for efficient data storage [6] and effective information discovery [7], [8]. They contain verified domain knowledge, ensuring data reliability. Disease knowledge graphs offer valuable interconnected paths to understand symptom-disease relationships, enhancing interpretability. To address our research question, we introduce Reinforced Disease Diagnosis on Knowledge Graph (RDKG). RDKG identifies meaningful paths in disease knowledge graphs to provide comprehensive symptom-disease information. This uncovers connections between symptoms and diseases, enhancing disease diagnosis model interpretability. RDKG comprises two key modules: (i) constructing and representing disease knowledge graphs and (ii) mining interconnected paths using reinforcement learning.

Specifically, we first construct a medical knowledge graph (KG) from electronic medical records annotated by medical experts. We use symptoms, diseases, surgeries, and other related information as nodes in the KG, and the doctors' pre-diagnosis records as textual information for these records. In order to obtain high-quality KG embeddings, we employ an enhanced embedding approach that combines traditional KG embeddings with textual embeddings. Secondly, we employ reinforcement learning to uncover meaningful interconnected paths within the disease knowledge graph. These paths extend from symptoms all the way to the ultimate diseases, thus revealing the entire reasoning process. Given the constraints of limited data and the presence of a long-tail distribution, we introduce soft reward mechanisms and pruning mechanisms to address these challenges.

In summary, we propose a Reinforced Disease Diagnosis on Knowledge Graph (RDKG) method. This method offers the following contributions: (1) We present an enhanced method for effectively representing the knowledge graph, leveraging additional textual information. (2) We formulate disease diagnosis and reasoning as a path construction problem on the knowledge graph. Subsequently, we reformulate it as a Markov Decision Process and address it within a reinforcement learning framework, incorporating a soft reward mechanism and a pruning strategy. (3) We conduct extensive experiments to demonstrate the improved performance and reasoning capabilities of our proposed method.

II. Related Work

A. Automatic Disease Diagnosis

One approach to automatic disease diagnosis is treating it as a classification problem, using deep learning models like [9]. However, these models are black boxes and lack

* Corresponding author.

979-8-3503-0788-7/23/$31.00 ©2023 IEEE
DOI 10.1109/ICDM58522.2023.00122

explanation. An alternative approach [10] employs convolutional neural networks with SHAP explanations, but some question its rationale. NEEDED [4] modeling the automatic disease diagnosis as classification tasks and solving it via a Transformer architecture. Another method, Diaformer [3], uses transformers but requires many inquiry turns. An efficient way is to convert multi-step reasoning into a Markov decision process and apply reinforcement learning (RL) [11]. However, this framework uses simplified data and may face the curse of dimensionality in complex real-world dialogues.

B. Knowledge Reasoning on Knowledge Graph

An automatic diagnosis model based on a medical knowledge graph utilizes heterogeneous information within the graph to uncover potential associations and provide more evidence for accurate diagnosis. This approach treats diagnosis as a knowledge graph path reasoning task, where the goal is to find the best path from symptoms to disease, ensuring interpretability by showing each decision step and its basis. Knowledge reasoning involves using existing process knowledge to infer new knowledge. Current knowledge graph reasoning methods encompass logical rules, distributed representations, and neural networks [12]–[14]. Neural network-based knowledge graph reasoning, particularly using reinforcement learning, has gained prominence [15]–[17]. Reinforcement learning has made significant progress [18], [19]. Deeppath [20] introduced reinforcement learning into knowledge graph reasoning, introducing a controllable multi-hop reasoning approach and framing path learning as reinforcement learning. Methods like [21] efficiently search graphs for answers based on input questions using reinforcement learning, avoiding precomputed paths. Our approach employs reinforcement learning for knowledge graph reasoning, incorporating soft rewards and a pruning strategy. The soft reward mechanism encourages wider exploration of results, while the pruning strategy enhances method efficiency by reducing the action space size.

III. METHOD

A. Enhanced KG Embedding

Enhanced Embedding The EMR encompasses a significant amount of unstructured textual data, which presents challenges in extraction but holds crucial clinical insights. Our enhanced embedding method aims to incorporate this information into knowledge embedding. To achieve this, we convert the unstructured textual data into embeddings, denoted as v_{t_i}, and then compute their average to obtain the fused embedding, v'.

$$v' = \frac{1}{n} \sum v_{t_i} \tag{1}$$

The triple (h, r, t) of the knowledge graph is expanded to (h, v', r, t), incorporating the fused embedding. Additionally, the scoring function of TransE is modified as follows:

$$f_r = \begin{cases} -\|(h + v') + r - t\|_2^2 & \text{if } h \in M \\ -\|h + r - (t + v')\|_2^2 & \text{if } t \in M \end{cases} \tag{2}$$

and the formula for calculating the entity projection in TransH is transformed as follows:

$$\begin{cases} h_\perp = (h + v') - w_r^\top (h + v')\, w_r & \text{if } h \in M \\ t_\perp = (t + v') - w_r^\top (t + v') w_r & \text{if } t \in M \end{cases} \tag{3}$$

the scoring function of TransR converts to:

$$f_r = \begin{cases} -\|M_r(h + v') + r - M_r t\|_2^2 & \text{if } h \notin M \\ -\|M_r h + r - M_r(t + v')\|_2^2 & \text{if } t \in M \end{cases} \tag{4}$$

where M represents the symptom set. We design two methods to obtain textual information embedding:

1) Textual information is processed into one-hot vectors and fed into a Neural Network layer to obtain vector embeddings.

2) Obtain textual information embeddings using pre-trained models. We used the Word2Vec model to obtain the textual information embedding and used a projection matrix to align its dimensions with the knowledge graph embedding. The reason we didn't opt for recent popular large-scale pre-trained language models like BERT [22] is that our dataset contains relatively short textual information with a small vocabulary, and there's virtually no word ambiguity or polysemy. In such cases, Word2Vec suffices for the task. Another factor to consider is that models like BERT have a plethora of parameters, which leads to lower efficiency and significantly longer training times for knowledge graph embedding learning.

B. Reinforced Disease Diagnosis Modeling

This section introduces reinforcement learning modeling. For simplicity, we denote the embeddings of the corresponding entities with lowercase letters below.

Markov Decision Process (MDP) We employ reinforcement learning methods to implement the Markov Decision Process. Reinforcement learning consists of state, action, reward, and transition [23], [24]. The definitions of these terms are as follows: ***State.*** The state at time t is denoted as $s_t = (m, e_t, h_t)$, where $m \in \mathbf{M}$ represents the starting point of the current path, e_t is the entity selected by the agent at time t, and h_t denotes the historical information recorded at time t. The window size of selected history information is denoted as k. The historical information h_t can be derived from $e_{t-k}, r_{t-k+1}, e_{t-k+1}, r_{t-k+2}, ..., r_t$. Initially, h_t is empty, and the initial state is represented as $s_0 = (m, m, \varnothing)$.

Action. The action space $\mathbf{A}_t$ of $s_t = (m_t, e_t, h_t)$ is defined as the set of all edges and nodes connected to e_t that have not appeared before. It can be represented as follows:

$$\mathbf{A}_t = \{(r, e) | (e_t, r, e) \in KG, e \notin \{e_0, ..., e_{t-1}\}\} \tag{5}$$

In medical KG, some common symptom entities are connected to a multitude of different nodes, which can lead to a potentially large number of candidates for certain states. To improve the decision-making process, we propose a pruning strategy that partitions the action set $\tilde{\mathbf{A}}_t(s)$ and restricts its maximum size to α, while giving priority to the most promising results. The pruning strategy is expressed as follows:

$$\tilde{\mathbf{A}}_t = \{(r, e) \mid \mathrm{rank}(f_p((r, e) \mid u) \leq \alpha, (r, e) \in \mathbf{A}_t\} \tag{6}$$

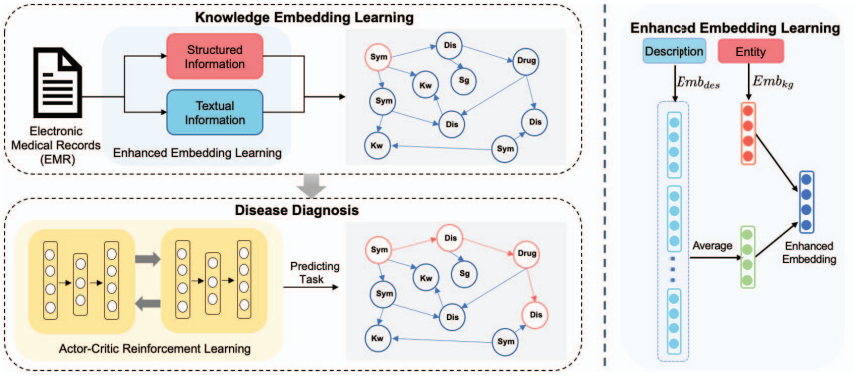

Fig. 1. Framework overview. Structural and textual information is extracted from EMR, and an enhanced embedding method (details are shown on the right of the figure) is employed to incorporate textual information into knowledge embedding. Disease diagnosis is based on actor-critic reinforcement learning, returning the path from symptom to disease. Sym, Dis, Kw, and Sg in the figure represent symptom, disease, keyword, and surgery, respectively.

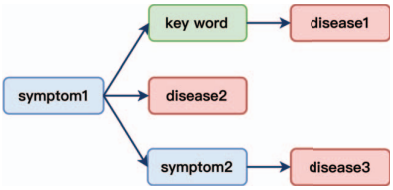

Fig. 2. With a binary reward mechanism, it is possible to define rewards for symptom1 and disease2, but it is not possible to define a reward for symptom1 and disease1 or disease3.

For a symptom entity m and other type entity e, we assume that there is a unique shortest path $p_k = (m, r_0, e_1, ..., r_k - 1, e_k)$, the scoring function for action pruning is defined as:

$$f_p = \begin{cases} \left\langle m + \sum_{i=1}^{k} r_i, e_k \right\rangle & \text{if } e_k \notin D \\ \left\langle m' + \sum_{i=1}^{k} r_i, e_k \right\rangle & \text{if } e_k \in D \end{cases} \tag{7}$$

where $\langle \cdot, \cdot \rangle$ is dot product. Please note that before calculating the scores for the pruning strategy, it is necessary to ensure that the current path does not contain the target node, and then proceed with the calculation. Otherwise, the node should be skipped. When computing with disease entities is involved, we use enhanced embedding of symptoms.

Reward. In reinforcement learning, binary reward mechanisms are commonly used. Here, the agent receives a fixed reward upon achieving a goal, while no reward or a penalty is given otherwise. This simplicity works well in tasks like games [25] but is suited for scenarios with fixed objectives.

However, when constructing knowledge graphs from real Electronic Medical Records (EMR) data, certain common diseases have ample recorded instances, with many connected nodes in the graph. A binary reward mechanism can facilitate transitions from symptoms to diseases in such cases. Yet, less frequent symptoms may not be present in limited EMR data, causing a lack of direct connections between symptom nodes and target disease nodes in the knowledge graph, as shown in Figure 2. In such situations, a binary reward mechanism will fail to produce accurate results.

Therefore, we design a soft reward mechanism that ensures the model can compute a corresponding reward for any disease appearing in the path, thereby mitigating the issue of long-tail distribution of the data. In the knowledge graph embedding stage, we aim to make the embeddings of connected entities as close as possible using techniques like TransE. Therefore, we can compute the reward using the following approach:

$$\mathbf{R}_t = \begin{cases} \max\left(0, f_r\left(m', e_t\right)\right), & \text{if } e_t \in D \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

where m' is enhanced embedding of m and $f_r = \langle m', e_t \rangle$.

Transition. Given state $s_t = (m, e_t, h_t)$ and action $a_t = (r_{t+1}, e_{t+1})$, the formula for transitioning to the next state s_{t+1} is:

$$\mathbf{P}[s_{t+1} = (m, e_{t+1}, h_{t+1}) | s_t, a_t] = 1 \tag{9}$$

please be noted that the initial state s_0 is random.

Policy Learning We build an actor-critic network to solve the KG problem. We design a two-layer fully connected (FC) neural network to learn a representation $\boldsymbol{x}_t$ from $\boldsymbol{s}_t$, a policy network to learn policy $\boldsymbol{\pi(A_t)}$ and a value network to learning value v_t according to $\boldsymbol{x}_t$. The structures of the networks are defined as follows:

$$\begin{gathered} \boldsymbol{x}_t = \sigma\left(\sigma\left(\boldsymbol{s}_t \boldsymbol{W}_1 + \boldsymbol{b}_1\right) \boldsymbol{W}_2 + \boldsymbol{b}_2\right) \\ \boldsymbol{\pi(A_t)} = \operatorname{softmax}\left(\left(\boldsymbol{x}_t \boldsymbol{W}_p + \boldsymbol{b}_p\right) \odot A_t\right) \\ v_t = \boldsymbol{x}_t \boldsymbol{W}_v + \boldsymbol{b}_v \end{gathered} \tag{10}$$

Here, σ is a nonlinear activation function, for which we use the rectified linear unit (ReLU) in this study, $\odot$ indicates the Hadamard product, and $\{\boldsymbol{W_1}, \boldsymbol{W_2}, \boldsymbol{W_p}, \boldsymbol{W_v}, \boldsymbol{b_1}, \boldsymbol{b_2}, \boldsymbol{b_p}, \boldsymbol{b_v}\}$ are the parameters and bias in the neural network.

C. Disease Inference

The complete inference process can be interpreted as follows: given a set of symptoms $(m_0, ..., m_n)$ and textual information $(t_0, ..., t_j)$, we use an enhanced embedding method to embed the textual information into multiple symptom embeddings, resulting in enhanced symptom embeddings $(m'_0, m'_1, ..., m'_n)$. These enhanced symptom embeddings are then fed into the policy network, generating different states $(s_0, ..., s_n)$. Path inference is performed, obtaining multiple sets of paths $(p_0, ..., p_n)$ starting from each symptom. Next,

we compute the fused symptom embedding $m_f = \frac{1}{n}\sum 0^n m'i$. We select the path with the highest score or likelihood as the returned result.

IV. Experimental Results

A. Experimental Setup

Dataset Description We conduct validation experiments on two datasets, Aier and CNschma. **The Aier dataset** is derived from actual electronic medical records (EMR) provided by Changsha Aier Eye Hospital. The EMR includes initial inquiries conducted by doctors, such as the patient's medical history, symptoms, and post-onset experiences. These inquiries are recorded as unstructured textual information, which was referred to as "textual information" mentioned earlier. **The CNschma dataset** is obtained from the public knowledge graph website OpenKG[1]. We constructed the input and output for the disease diagnosis model based on random sampling and introduced noise into the input set to test the model's robustness. The quantities of entities and relationships in these two datasets are displayed in Table II and Table I.

Acknowledgment: Aier is a private dataset with full authorization from Aier Eye Hospital Group Company Limited, and has been desensitized for privacy preservation. CNschema is publically available.

TABLE I
RELATIONSHIP DESCRIPTION AND QUANTITATIVE STATISTICS.

Relation	Description	Aier	CNschema
disease_symptom	disease ⟷ symptom	484	6186
disease_surgery	disease ⟷ surgery	682	1926
disease_drug	disease ⟷ drug	974	4978
related_symptom	symptom ⟷ symptom	478	2489
related_disease	disease ⟷ disease	470	155
mentions	symptom ⟷ keyword	17511	13127
described_as	disease ⟷ keyword	13156	7467

TABLE II
ENTITY EXAMPLE AND ITS QUANTITY STATISTICS.

Entity	Example	Aier	CNschema
Symptom	Decreased vision	148	1741
Disease	Retinal Detachment	150	801
Surgery	Scleral Buckling Surgery	364	140
Medicine	Artificial Tears	544	1516
keyword	Obstruction, 4 days,left eye	2255	9744

Baseline Methods We compare our model with various prediction approaches: SVM and XGBoost are classical machine learning methods. Both of these methods treat the disease diagnosis task as a classification task, where they predict possible diseases based on multiple input symptoms. Then, we compare with the latest several automatic disease diagnosis models. Unlike our approach of using enhanced embedding to incorporate the inquiry results (textual information) into the embedding of KG, GAMP uses a Generative Adversarial Network (GAN) as the foundation of its Reinforcement Learning (RL) model [26]. It utilizes the generator of the GAN as a policy network to generate symptom inquiries, thereby optimizing the diagnostic process. Diaformer [3] is a method that formalizes automatic diagnosis as a sequence generation task. It introduces a symptom attention generation architecture to directly learn the underlying relationship between symptom inquiry selection and disease diagnosis.BED [27] introduces a medical diagnosis model based on the Bayesian method, employing the Quick Medical Reference (QMR) belief network and Bayesian inference to facilitate medical inquiry and disease inference.

Implementation Details The model training is conducted in two steps. Firstly, we train KG embedding using three KG Embedding algorithms: TransE, TransH, and TransR. The embedding size is set to 100. The maximum path length (K) for automatic disease diagnosis is set to 2 as the default value. The batch size is set to 64, and the number of negative samples is set to 10. Once the KG embedding training is completed, we freeze the embeddings of all entities and relations and proceed to train the reinforcement learning module. In the reinforcement learning part, the history information vector h_t is obtained by the concatenation of the embeddings of e_{t-1} and r_t, and the state vector $s = (m_t, e_t, e_{t-1}, r_t)$ is of size 400. The maximum size of the pruned action space is set to 200. The discount factor γ is 0.99. For the policy/value network, $\mathbf{W}_1 \in \mathbb{R}^{400\times512}, \mathbf{W}_2 \in \mathbb{R}^{512\times256}, \mathbf{W}_{\mathrm{p}} \in \mathbb{R}^{256\times80}$ and $\mathbf{W}_{\mathrm{v}} \in \mathbb{R}^{256\times1}$. Both steps' learning rate is set to 0.001. In the path inference stage, we set the sample size at each step as $K_1 = 25, K_2 = 5, K_3 = 1$.

B. Experimental Results

Overall Performance This experiment aims to answer: *Can RDKG effectively predict the disease in the automatic disease diagnosis task?* Table III shows the comparison in terms of precision, recall, F1-macro, Hits@{1, 3, 5}. We observe that RDKG outperforms other baseline methods on both of these datasets. On the Aier dataset and CNschema dataset, RDKG achieves a significant improvement in precision, recall, and Hits@1. Furthermore, our experimental results consistently demonstrate that the proposed RDKG model outperforms other baseline methods, including knowledge graph embedding methods such as TransE, TransH, and TransR, in disease diagnosis tasks. Despite minor fluctuations, RDKG consistently exhibits superior performance in disease diagnosis tasks. These findings highlight the effectiveness and robustness of RDKG in handling such tasks.

Ablation Study This experiment aims to answer: *How do embedding strategy and restriction strategy impact the performance of RDKG?* Table IV shows the results of different embedding and restriction strategies on the Aier dataset. Generally, RDKG benefits from advanced KG embedding (Emb_{kg}) methods (TransR, TransH $>$ TransE) and restriction strategies (With Restriction $>$ No Restriction). The restriction strategy is that there cannot be duplicate nodes on the disease diagnosis path, and the second node cannot be a disease type. In addition, we also evaluated the RDKG with different textual information embedding (Emb_{txt}) methods: Base (without textual information embedding function), well-trained Word2Vec (W2V) [28], and Neural Network (NN). We can observe that the combinations of different KG embedding, textual information embedding, and restriction strategies (TransH, NN,

[1]https://github.com/nuolade/disease-kb

TABLE III

OVERALL PERFORMANCE COMPARISON OF DIFFERENT METHODS. THE BEST RESULTS ARE HIGHLIGHTED IN **BOLD**. AND THE SECOND-BEST RESULTS ARE HIGHLIGHTED IN UNDERLINE. (HIGHER VALUES INDICATE BETTER PERFORMANCE.)

Model	Aier						CNschema					
	Precision	Recall	F1-macro	HITS@1	HITS@3	HITS@5	Precision	Recall	F1-macro	HITS@1	HITS@3	HITS@5
SVM	0.059	0.033	0.037	0.199	0.316	0.381	0.304	0.306	0.287	0.365	0.384	0.402
XGBoost	0.092	0.044	0.053	0.245	0.368	0.342	0.342	0.276	0.282	0.368	0.437	0.467
GAMP	0.035	0.092	0.090	0.203	0.353	0.386	0.316	0.371	0.322	0.382	0.425	0.449
BED	0.094	0.130	0.107	0.181	0.264	0.313	0.410	0.377	0.377	0.406	0.410	0.417
Diaformer	0.055	0.098	0.0664	0.222	0.351	0.406	0.662	0.673	0.683	0.745	0.829	0.859
RDKG(TransE)	0.133	0.150	0.138	0.302	0.431	0.505	0.712	0.713	0.698	0.765	0.843	0.876
RDKG(TransH)	**0.163**	0.176	**0.162**	0.333	0.458	0.527	0.709	0.711	0.702	0.767	0.848	0.882
RDKG(TransR)	0.146	**0.181**	0.157	**0.352**	**0.480**	**0.533**	**0.733**	**0.735**	**0.720**	**0.784**	**0.851**	**0.892**

TABLE IV

COMPARISON OF DIFFERENT COMPONENTS OF RDKG ON AIER DATASET.

Emb_{kg}	Emb_{txt}	No Restriction						With Restriction					
		Precision	Recall	F1-macro	HITS@1	HITS@3	HITS@5	Precision	Recall	F1-macro	HITS@1	HITS@3	HITS@5
TransE	Base	0.112	0.132	0.111	0.252	0.391	0.475	0.115	0.148	0.122	0.263	0.400	0.483
	W2V	0.095	0.127	0.103	0.275	0.419	0.491	0.090	0.115	0.097	0.288	0.411	0.481
	NN	0.104	0.138	0.111	0.291	0.425	0.499	0.133	0.150	0.138	0.302	0.431	0.505
TransH	Base	0.097	0.113	0.094	0.263	0.430	0.501	0.160	0.176	0.158	0.305	0.446	0.501
	W2V	0.142	0.168	0.150	0.324	0.449	0.504	0.122	0.151	0.130	0.314	0.425	0.472
	NN	0.097	0.115	0.092	0.282	0.467	0.518	**0.163**	0.176	**0.162**	0.333	0.458	0.527
TransR	Base	0.113	0.143	0.117	0.277	0.416	0.507	0.135	0.163	0.138	0.291	0.411	0.525
	W2V	0.087	0.121	0.096	0.294	0.447	0.483	0.105	0.135	0.112	0.302	0.466	0.491
	NN	0.097	0.137	0.119	0.325	0.475	0.513	0.146	**0.181**	0.157	**0.352**	**0.480**	**0.533**

TABLE V

DIAGNOSIS EFFECTIVENESS OF OUR MODEL UNDER DIFFERENT SIZES OF PRUNED ACTION SPACES ON AIER DATASET.

Emb_{kg}	*Pruning Size*	H@1	H@2	H@3	H@4	H@5	H@6	H@7	H@8	H@9	H@10	Precision	Recall	F1-macro
TransE	400	0.222	0.389	0.417	0.500	0.556	0.556	0.569	0.569	0.569	0.583	0.088	0.110	0.084
	300	0.236	0.347	0.403	0.458	0.500	0.556	0.583	0.597	0.597	0.597	0.086	0.111	0.081
	200	0.250	0.347	0.389	0.444	0.486	0.500	0.542	0.556	0.556	0.556	0.089	0.118	0.081
	100	0.264	0.347	0.403	0.444	0.486	0.528	0.528	0.556	0.569	0.569	0.099	0.122	0.098
TransH	400	0.253	0.411	0.433	0.486	0.486	0.499	0.510	0.551	0.578	0.602	0.100	0.120	0.096
	300	0.253	0.415	0.443	0.477	0.511	0.516	0.556	0.584	0.597	0.611	0.104	0.125	0.099
	200	0.266	0.422	0.463	0.472	0.528	0.528	0.556	0.597	0.597	0.611	0.125	0.146	0.121
	100	0.269	0.431	0.463	0.486	0.542	0.556	0.569	0.625	0.625	0.653	0.132	0.140	0.138
TransR	400	0.295	0.397	0.443	0.476	0.520	0.533	0.542	0.581	0.604	0.604	0.144	0.167	0.140
	300	0.296	0.409	0.443	0.486	0.528	0.542	0.542	0.579	0.594	0.624	0.145	0.167	0.141
	200	0.302	0.423	0.444	0.499	0.533	0.556	0.611	0.639	0.639	0.653	0.148	0.167	0.147
	100	**0.306**	**0.444**	**0.472**	**0.500**	**0.542**	**0.583**	**0.611**	**0.639**	**0.653**	**0.667**	**0.148**	**0.167**	**0.147**

With Restriction) and (TransR, NN, With Restriction) achieve better performances on the Aier dataset. In detail, for the KG embedding strategy, TransH and TransR outperform TransE due to their advancement. And for the textual information embedding strategy, NN outperforms W2V because W2V is a well-trained model, which doesn't work well with each medical dataset. Moreover, generating duplicate nodes and extrapolating the correct result prematurely may have bad impacts on disease diagnosis modeling, thus the results with the restriction strategy can outperform the results without the restriction strategy.

Diagnostic Path Length Analysis This experiment aims to answer: *How does the length of the diagnostic path impact the performance of RDKG?* Disease diagnosis in the real world usually needs to go through multiple steps. Correspondingly, the analysis of the length of automatic diagnosis methods is also meaningful.To explore the effect of diagnostic path length L on RDKG performance, we conduct experiments on the Aier dataset without restriction strategy. Figure 3 shows the experimental results in the evaluation metrics HIT@{1,2,3,4,5}. We can observe that as the value of L decreases, the model performance is better in the metrics HIT@{1,2,3}. And longer diagnostic paths may capture more useful information, so they may perform better in HITS@{4,5}. Our model can dynamically adjust the diagnostic path length without retraining and can meet the needs of different application scenarios.

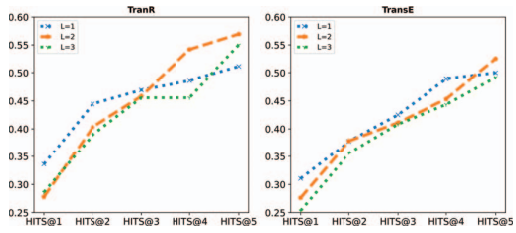

Fig. 3. Evaluation results under different path length L on Aier dataset.

Pruning Strategy Analysis In this experiment, we evaluate how the performance of our model varies with different sizes of pruned action spaces. Recall that, conditioned on the starting symptom, the action space is pruned according to the scoring function defined in section 4.3, where actions with higher scores are more likely to be retained. Through the evaluation of the two datasets, we find that the performance of our model is slightly influenced by the size of the pruned

action space. As shown in table V, the performance of the model will initially increase with the decrease of the pruning size, and then gradually smooth down. When the pruning size is set smaller, more nodes need to be pruned, which leads to a significant increase in training and testing time. For example, when $pruning\ size = 100$, it takes about 1.5 times as long as $pruning\ size = 200$.

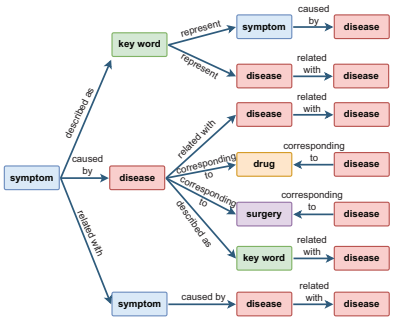

Fig. 4. All 3-hop path patterns found in the results.

V. Conclusions

In this paper, we transform the task of disease diagnosis into a Markov Decision Process on a medical knowledge graph and employ reinforcement learning algorithms to address this challenge. To enhance the representation of the knowledge graph, we introduce a novel enhanced embedding method. Experimental results demonstrate that our method outperforms all baseline models in terms of accuracy and interoperability.

VI. Acknowledgements

This work was supported by the National Key Research and Development Plan of China (No. 2022YFF0712200), Beijing Natural Science Foundation under Grant No. 4212030, Science and Technology Service Network Initiative, Chinese Academy of Sciences (No. KFJ-STS-QYZD-2021-11-001) and Youth Innovation Promotion Association CAS.

References

[1] J. Lu, J. Hu, G. Zhao, F. Mei, and C. Zhang, "An in-field automatic wheat disease diagnosis system," *Computers and electronics in agriculture*, vol. 142, pp. 369–379, 2017.

[2] N. Papandrianos and E. Papageorgiou, "Automatic diagnosis of coronary artery disease in spect myocardial perfusion imaging employing deep learning," *Applied Sciences*, vol. 11, no. 14, p. 6362, 2021.

[3] J. Chen, D. Li, Q. Chen, W. Zhou, and X. Liu, "Diaformer: Automatic diagnosis via symptoms sequence generation," vol. 36, no. 4, pp. 4432–4440, 2022.

[4] X. Ye, M. Xiao, Z. Ning, W. Dai, W. Cui, Y. Du, and Y. Zhou, "Needed: Introducing hierarchical transformer to eye diseases diagnosis," in *Proceedings of the 2023 SIAM International Conference on Data Mining (SDM)*. SIAM, 2023, pp. 667–675.

[5] J. Amann, A. Blasimme, E. Vayena, D. Frey, V. I. Madai, and P. Consortium, "Explainability for artificial intelligence in healthcare: a multidisciplinary perspective," *BMC medical informatics and decision making*, vol. 20, pp. 1–9, 2020.

[6] Y. Zhou, W. Wang, Z. Qiao, M. Xiao, and Y. Du, "A survey on the construction methods and applications of sci-tech big data knowledge graph," *Sci. Sin. Inf*, vol. 50, no. 7, p. 957, 2020.

[7] Z. Qiao, P. Wang, P. Wang, Z. Ning, Y. Fu, Y. Du, Y. Zhou, J. Huang, X.-S. Hua, and H. Xiong, "A dual-channel semi-supervised learning framework on graphs via knowledge transfer and meta-learning," *ACM Transactions on the Web*, 2023.

[8] H. Dong, Z. Ning, P. Wang, Z. Qiao, P. Wang, Y. Zhou, and Y. Fu, "Adaptive path-memory network for temporal knowledge graph reasoning," *arXiv preprint arXiv:2304.12604*, 2023.

[9] J. Liu, Z. Zhang, and N. Razavian, "Deep ehr: Chronic disease prediction using medical notes," in *Machine Learning for Healthcare Conference*. PMLR, 2018, pp. 440–464.

[10] D. Zhang, S. Yang, X. Yuan, and P. Zhang, "Interpretable deep learning for automatic diagnosis of 12lead electrocardiogram," *Iscience*, vol. 24, no. 4, p. 102373, 2021.

[11] Z. Wei, Q. Liu, B. Peng, H. Tou, T. Chen, X.-J. Huang, K.-F. Wong, and X. Dai, "Task-oriented dialogue system for automatic diagnosis," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2018, pp. 201–207.

[12] X. Chen, S. Jia, and Y. Xiang, "A review: Knowledge reasoning over knowledge graph," *Expert Systems with Applications*, vol. 141, p. 112948, 2020.

[13] R. Lu, Z. Cai, and S. Zhao, "A survey of knowledge reasoning based on kg," in *IOP Conference Series: Materials Science and Engineering*, vol. 569, no. 5. IOP Publishing, 2019, p. 052058.

[14] F. Yang, Z. Yang, and W. W. Cohen, "Differentiable learning of logical rules for knowledge base reasoning," *Advances in neural information processing systems*, vol. 30, 2017.

[15] X. Wang, K. Liu, D. Wang, L. Wu, Y. Fu, and X. Xie, "Multi-level recommendation reasoning over knowledge graphs with reinforcement learning," in *Proceedings of the ACM Web Conference 2022*, 2022, pp. 2098–2108.

[16] P. Wang, K. Liu, L. Jiang, X. Li, and Y. Fu, "Incremental mobile user profiling: Reinforcement learning with spatial knowledge graph for modeling event streams," in *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 2020, pp. 853–861.

[17] Y. Li, "Deep reinforcement learning: An overview," *arXiv preprint arXiv:1701.07274*, 2017.

[18] M. Xiao, D. Wang, M. Wu, Z. Qiao, P. Wang, K. Liu, Y. Zhou, and Y. Fu, "Traceable automatic feature transformation via cascading actor-critic agents," in *Proceedings of the 2023 SIAM International Conference on Data Mining (SDM)*. SIAM, 2023, pp. 775–783.

[19] M. Xiao, D. Wang, M. Wu, K. Liu, H. Xiong, Y. Zhou, and Y. Fu, "Traceable group-wise self-optimizing feature transformation learning: A dual optimization perspective," *arXiv preprint arXiv:2306.16893*, 2023.

[20] W. Xiong, T. Hoang, and W. Y. Wang, "Deeppath: A reinforcement learning method for knowledge graph reasoning," *arXiv preprint arXiv:1707.06690*, 2017.

[21] R. Das, S. Dhuliawala, M. Zaheer, L. Vilnis, I. Durugkar, A. Krishnamurthy, A. Smola, and A. McCallum, "Go for a walk and arrive at the answer: Reasoning over paths in knowledge bases using reinforcement learning," *arXiv preprint arXiv:1711.05851*, 2017.

[22] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[23] K. Liu, P. Wang, D. Wang, W. Du, D. O. Wu, and Y. Fu, "Efficient reinforced feature selection via early stopping traverse strategy," in *2021 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2021, pp. 399–408.

[24] P. Wang, D. Wang, K. Liu, D. Wang, Y. Zhou, L. Sun, and Y. Fu, "Hierarchical reinforced urban planning: Jointly steering region and block configurations," in *Proceedings of the 2023 SIAM International Conference on Data Mining (SDM)*. SIAM, 2023, pp. 343–351.

[25] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.

[26] Y. Xia, J. Zhou, Z. Shi, C. Lu, and H. Huang, "Generative adversarial regularized mutual information policy gradient framework for automatic diagnosis," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, 2020, pp. 1062–1069.

[27] H. Guan and C. Baral, "A bayesian approach for medical inquiry and disease inference in automated differential diagnosis," *arXiv preprint arXiv:2110.08393*, 2021.

[28] Y. Weng, "Chineseword2vecmedicine," https://github.com/WENGSYX/Chinese-Word2vec-Medicine, 2021.