

A Reinforcement Learning-based Temporal Knowledge Graph Reasoning Model for Chronic Gastritis Diagnosis Prediction

Zhou Sun, Xiaolong Qu, Yuhang Wang, Haiyu Liu, Lei Yao, Dongmei Li, Guanli Song, Runshun Zhang, Xiaoping Zhang

Submitted to: JMIR Medical Informatics
on: September 07, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 4

Supplementary Files..... 18

 Figures 19

 Figure 1..... 20

 Figure 2..... 21

 Figure 3..... 22

 Figure 4..... 23

 TOC/Feature image for homepages 24

 TOC/Feature image for homepage 0..... 25

A Reinforcement Learning-based Temporal Knowledge Graph Reasoning Model for Chronic Gastritis Diagnosis Prediction

Zhou Sun¹; Xiaolong Qu¹; Yuhang Wang¹; Haiyu Liu²; Lei Yao³; Dongmei Li¹; Guanli Song²; Runshun Zhang²; Xiaoping Zhang²

¹ School of Information Science and Technology Beijing Forestry University, Beijing, China Beijing, China CN

² National Data Center of Tradition Chinese Medicine of China Academy of Chinese Medical Sciences, Beijing, China Beijing, China CN

³ University of Wisconsin-Milwaukee, Milwaukee, WI, USA WI, USA US

Corresponding Author:

Xiaoping Zhang

National Data Center of Tradition Chinese Medicine of China Academy of Chinese Medical Sciences, Beijing, China
National Data Center of Tradition Chinese Medicine of China Academy of Chinese Medical Sciences, Beijing, China
Beijing, China
CN

Abstract

Background: The clinical progression of chronic gastritis entails intricate temporal dependencies, making it difficult for conventional methods to capture both the disease's dynamic trajectory and the underlying relationships among medical events.

Objective: To predict chronic gastritis diagnosis dynamically, we proposed RL4TKGR, a Reinforcement Learning-based Temporal Knowledge Graph Reasoning model for chronic gastritis diagnosis prediction.

Methods: Specifically, RL4TKGR incorporated an innovative reinforcement learning policy network that employed a dual-path encoding module to separately model historical and non-historical diagnostic information. Furthermore, RL4TKGR integrated a dual-channel reward function with a dynamic weight allocation mechanism, which adaptively balanced the two information sources, thereby addressing the strategic bias problem and enabling interpretable reasoning.

Results: Experiments conducted on the self-constructed Chronic Gastritis Temporal Knowledge Graphs (CG-TKG) demonstrated that RL4TKGR delivered state-of-the-art performance across MRR and Hits@1/3/10 metrics, outperforming existing baseline models. Ablation experiments and case analyses further validated the efficacy of each module design and the model's practical utility in predicting disease subtypes and therapeutic medications.

Conclusions: This paper advanced the predictive accuracy of chronic gastritis while establishing a novel and effective paradigm for interpretable reasoning over complex temporal medical data.

(JMIR Preprints 07/09/2025:83544)

DOI: <https://doi.org/10.2196/preprints.83544>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in a JMIR journal, I will be able to make the full manuscript visible to all users.

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in a JMIR journal, I will be able to make the full manuscript visible to all users.

Original Manuscript

A Reinforcement Learning-based Temporal Knowledge Graph Reasoning Model for Chronic Gastritis Diagnosis Prediction

Zhou Sun¹, Xiaolong Qu¹, Yuhang Wang¹, Haiyu Liu², Lei Yao³, Dongmei Li^{1,*}, Guanli Song⁴, Runshun Zhang⁴, Xiaoping Zhang^{2,*}

¹School of Information Science and Technology Beijing Forestry University, Beijing, China

²National Data Center of Tradition Chinese Medicine of China Academy of Chinese Medical Sciences, Beijing, China

³University of Wisconsin-Milwaukee, Milwaukee, WI, USA

⁴Guang'anmen Hospital China Academy of Chinese Medical Sciences, Beijing, China

*Corresponding Authors

Abstract

Background: The clinical progression of chronic gastritis entails intricate temporal dependencies, making it difficult for conventional methods to capture both the disease's dynamic trajectory and the underlying relationships among medical events.

Objective: To predict chronic gastritis diagnosis dynamically, we proposed RL4TKGR, a **R**einforcement **L**earning-based **T**emporal **K**nowledge **G**raph **R**easoning model for chronic gastritis diagnosis prediction.

Methods: Specifically, RL4TKGR incorporated an innovative reinforcement learning policy network that employed a dual-path encoding module to separately model historical and non-historical diagnostic information. Furthermore, RL4TKGR integrated a dual-channel reward function with a dynamic weight allocation mechanism, which adaptively balanced the two information sources, thereby addressing the strategic bias problem and enabling interpretable reasoning.

Results: Experiments conducted on the self-constructed Chronic Gastritis Temporal Knowledge Graphs (CG-TKG) demonstrated that RL4TKGR delivered state-of-the-art performance across MRR and Hits@1/3/10 metrics, outperforming existing baseline models. Ablation experiments and case analyses further validated the efficacy of each module design and the model's practical utility in predicting disease subtypes and therapeutic medications.

Conclusions: This paper advanced the predictive accuracy of chronic gastritis while establishing a novel and effective paradigm for interpretable reasoning over complex temporal medical data.

Keywords: Temporal Knowledge Graph Reasoning; Reinforcement Learning; Chronic Gastritis; Diagnosis Prediction; Traditional Chinese Medicine

1 Introduction

Chronic gastritis is a chronic inflammation of the gastric mucosa caused by multiple pathogenic factors. Its spectrum of precancerous lesions—including gastric atrophy, intestinal metaplasia, and dysplasia—has been shown to confer independent carcinogenic risks[1]. Consequently, elucidating the evolutionary trajectory of chronic gastritis and accurately predicting its progression are of critical importance for the early detection and intervention of gastric cancer, underscoring its significant clinical value.

In recent years, artificial intelligence technologies have been extensively explored for the diagnosis and treatment of chronic gastritis. In prediction studies based on static features, through the analysis of gastroscopic images, some models have demonstrated high accuracy and efficiency in the diagnosis of gastric diseases such as chronic atrophic gastritis, significantly improving the accuracy of real-time diagnosis[2-4]. Meanwhile, the study by Tao *et al.*[5] not only realized the diagnosis of gastric mucosal atrophy but also performed risk stratification based on its extent, providing decision support for subsequent monitoring intervals. However, static prediction models struggled to effectively capture the dynamic evolution trends of diseases and failed to consider the temporal characteristics inherent in their clinical progression adequately. Therefore, current research has gradually shifted from static, point-in-time diagnosis toward dynamic risk prediction and disease course evolution analysis based on temporal medical data. Existing studies have leveraged electronic medical record data and applied temporal analysis techniques to predict the progression risk from atrophic gastritis to gastric cancer, while also quantifying relevant disease trajectories and high-risk factors[6-8]. Nevertheless, temporal medical data primarily record sequences of indicator values changing over time, which inherently lack the capacity to organize information into structured tuple forms. This limitation constrains the ability to model deeper associations among medical events.

Temporal Knowledge Graphs (TKGs) extend the static knowledge graph framework by introducing the temporal dimension[9], and they represent the evolutionary processes of dynamic events using quaternary tuples (s, p, o, t) . In medical scenarios, Knowledge Graphs can provide doctors with reliable auxiliary diagnostic tools and have good explainability[10]. Moreover, TKGs can model the dynamic evolution of relationships between entities, demonstrating key value in disease progression prediction and drug interaction tracking[11]. Static pathological features are limited in capturing the dynamic trends of disease progression and overlook the clinical evolution of the disease course, whereas TKGs can embed timestamps into entity states to capture the progressive nature of chronic gastritis. Temporal data

primarily capture variations in indicator values over time but lack the ability to organize information into structured tuples, which limits their capacity to model the deeper associations among medical events. In contrast, TKGs enable the representation of the clinical evolution of chronic gastritis, thereby improving both the accuracy and interpretability of prediction. Nevertheless, studies that leverage TKGs to predict the dynamic progression of chronic gastritis remain scarce.

In clinical practice, patients' historical diagnosis information plays an essential role in predicting diagnoses for subsequent patients with the same disease. The patterns of disease representation similarity in patient symptoms [12] provide a theoretical foundation for temporal diagnosis and treatment modeling. Reinforcement learning excels at deriving optimal strategies through interaction with the environment, making it well-suited to simulate the clinical decision-making process in which physicians formulate future decisions based on patients' historical medical records and current conditions. Meanwhile, the medical decision-making process can be regarded as a sequential decision-making problem of finding the optimal path in massive information, which aligns closely with the modeling paradigm of reinforcement learning. Therefore, we constructed a Temporal Knowledge Graph for chronic gastritis (CG-TKG) based on real-world dynamic diagnostic and treatment data. By modeling and analyzing patients' historical medical records and incorporating the characteristics of clinical data, we proposed a Reinforcement Learning-based Temporal Knowledge Graph Reasoning model that leveraged historical diagnostic information to overcome the limitations of traditional static prediction and enabled the prediction of disease subtypes and treatment drugs for chronic gastritis.

To summarize, the contributions of this paper are as follows:

(1) We proposed a Reinforcement Learning-based Temporal Knowledge Graph Reasoning model for chronic gastritis diagnosis prediction (RL4TKGR). We designed a reinforcement learning policy network that exploited associations within historical diagnostic data to predict chronic gastritis diagnosis, whereby RL4TKGR captured both features from patients' historical medical records and latent features of non-historical diagnoses, thus enabling interpretable reasoning on the CG-TKG.

(2) We proposed a dual-channel aggregated reward function and a dual-path encoded policy network based on historical diagnosis. By constructing reward functions for historical and non-historical information, the model guided the learning of the corresponding paths within the policy network. Furthermore, the event balance factor β was used to dynamically and interpretably adjust the weight allocation in the dual-channel aggregation reward function and the action scoring module, effectively addressing the strategy bias issue caused by static weight allocation.

(3) In the self-constructed dataset CG-TKG, RL4TKGR outperformed the suboptimal baseline by an average of 8.86%, 3.04%, 5.09%, and 5.52% in terms of MRR, Hits@1/3/10 metrics respectively. It can efficiently perform chronic gastritis diagnosis prediction and improved the accuracy of clinical temporal data analysis.

2 Related Work

2.1 Temporal Knowledge Graph Reasoning

A TKG is a special type of knowledge graph that adds a temporal dimension to the traditional triple representation (entity, relation, entity), forming a new quadruple: (entity, relation, entity, time). This representation is capable of capturing the temporal evolution of relationships among entities, which is of great significance for understanding and predicting their dynamic interactions [13].

The TTransE model [14] was an extension of the TransE model [15] by adding a temporal dimension. Its scoring function considered temporal information, enabling the model to handle quadruples in temporal knowledge graphs. The TA-DistMult model [16] was a temporal version of the DistMult model [17], which integrated temporal information by learning a sequence encoder. This model used a Recurrent Neural Network (RNN) to learn time-aware representations of relation types. Similarly, the TA-TranE model [18] was a temporal variant of the TransE model, incorporating temporal information into the standard embedding framework for link prediction. This model leveraged an RNN to encode sequences of temporal tokens, thereby learning time-aware representations.

The DE-Simple model [19] equipped static models with a diachronic entity embedding for temporal knowledge graph completion. The xERTE model [20] proposed an explainable reasoning framework for future link prediction in temporal knowledge graphs, achieved through a temporal relation attention mechanism and a novel reverse representation update scheme. TiTer [21] was the first to use path-based reinforcement learning for temporal knowledge graph reasoning and can handle unseen timestamps and entities. In addition, the RTTI model [22] further proposed a new method for representing time intervals on this basis, using the median of two timestamps and embedding changes to represent time intervals.

Existing works have adopted various Graph Neural Network methods to model the evolution patterns of temporal knowledge graphs. RE-GCN [23] designed a recurrent evolutionary network to capture both structural dependencies and temporal patterns. To address the problem of length diversity in evolution patterns, Cen [24] combined relational graph neural networks and length-aware convolutional networks to learn evolution features of different historical sequence lengths. RE-Net [25] modeled the temporal conditional probability distribution of event sequences through a recursive event encoder, while employing a neighborhood aggregator to handle concurrent events and enable global structural reasoning. CeNet [26] further considered the influence of potential unobserved factors, leveraging contrastive learning to

model dependencies between historical and non-historical events, thereby improving predictive performance.

2.2 Application of Temporal Knowledge Graph Reasoning in Disease Prediction

Knowledge graph reasoning technology in the medical field is evolving from static modeling to dynamic temporal analysis[27]. Carvalho *et al.*[28] proposed an ontology-based electronic medical record temporal graph construction framework, which realized dynamic tracking of patient status through semantic annotation and timestamp mapping, establishing a foundation for the advancement of knowledge graph reasoning in the medical domain.

Considering the temporal characteristics of diabetes progression, Geng *et al.*[29] proposed an incremental LSTM model, which embedded entity relations through TransR and integrated graph topological structures, achieving precise modeling of complication evolution on clinical knowledge graphs. Chaturvedi *et al.*[30] leveraged a pre-trained Transformer to extract fine-grained temporal relations from clinical texts, constructing a patient-centered multi-source temporal graph and providing a scalable risk prediction framework for chronically monitored diseases such as type II diabetes. Considering the task characteristics of long-term health risk prediction, Postiglione *et al.*[31] proposed the MedTKG framework, which integrated the dynamic temporal information from Electronic Health Records (EHR) with the static hierarchy of medical ontologies, and validated the effectiveness of ontology-enhanced temporal reasoning on the MIMIC-III dataset. All of above efforts have focused on chronic disease evolution and complication prediction. For acute disease deterioration prediction, Song *et al.*[11], accounted for the rapid worsening of disease trajectories by integrating temporal information into a GRU network while preserving graph structural properties with TransR, thereby developing a diagnostic support system capable of real-time prediction of acute diseases and their complications.

2.3 Application of Reinforcement Learning in Disease Prediction

As one of the three major paradigms of machine learning, reinforcement learning centers on the dynamic interaction between an agent and its environment, optimizing decision-making by maximizing target objectives through reward signals. Unlike supervised learning, reinforcement learning models the decision-making trajectory via a Markov decision process[32].

The core advantage of reinforcement learning in disease prediction lies in its dynamic decision optimization capability. Compared with static prediction models, reinforcement learning can simulate the multi-stage process of disease development and formulate personalized intervention strategies according to the real-time status of patients. In the context of Alzheimer's disease prediction, Chaudhari *et al.*[33] proposed the CAdam-RL-DCNN model, which innovatively integrated the improved Coyote optimization algorithm with the Adam optimizer, combined with a reinforcement learning decision mechanism, to solve the problems of feature extraction efficiency and class imbalance. Zhang *et al.*[34] constructed the Weighted Dueling Double DQN model, integrated clinical expert rules to guide action selection, and ensured decision reliability through Doubly Robust Off-Policy Evaluation, providing a new paradigm for dynamic treatment in ICUs.

Despite these advances, most existing TKG reasoning models in the medical domain relied on static weight fusion or fixed embedding schemes, limiting their ability to dynamically adapt to individualized diagnostic and treatment trajectories or to differentiate historical and non-historical clinical events, thus leading to policy bias. Likewise, current reinforcement learning-based disease prediction models often depended on static rules or weight fusion, which constrained their capacity to dynamically capture the relevance of historical events. To address these challenges, we proposed a reinforcement learning-based temporal knowledge graph reasoning model based on historical diagnosis information to effectively overcome the bias problem of static fusion and significantly improve the accuracy and interpretability of chronic gastritis prediction.

3 Methodology

The framework of RL4TKGR is shown in Fig. 1, with the policy network as its core component. The policy network consists of four modules: the diagnosis embedding module, the dual-path encoding module, the action scoring module, and the dynamic weight allocator module. The dual-channel reward function guides the learning of the policy network, comprising both a historical reward function and a non-historical reward function, which are adaptively integrated through a dynamic weight allocation module.

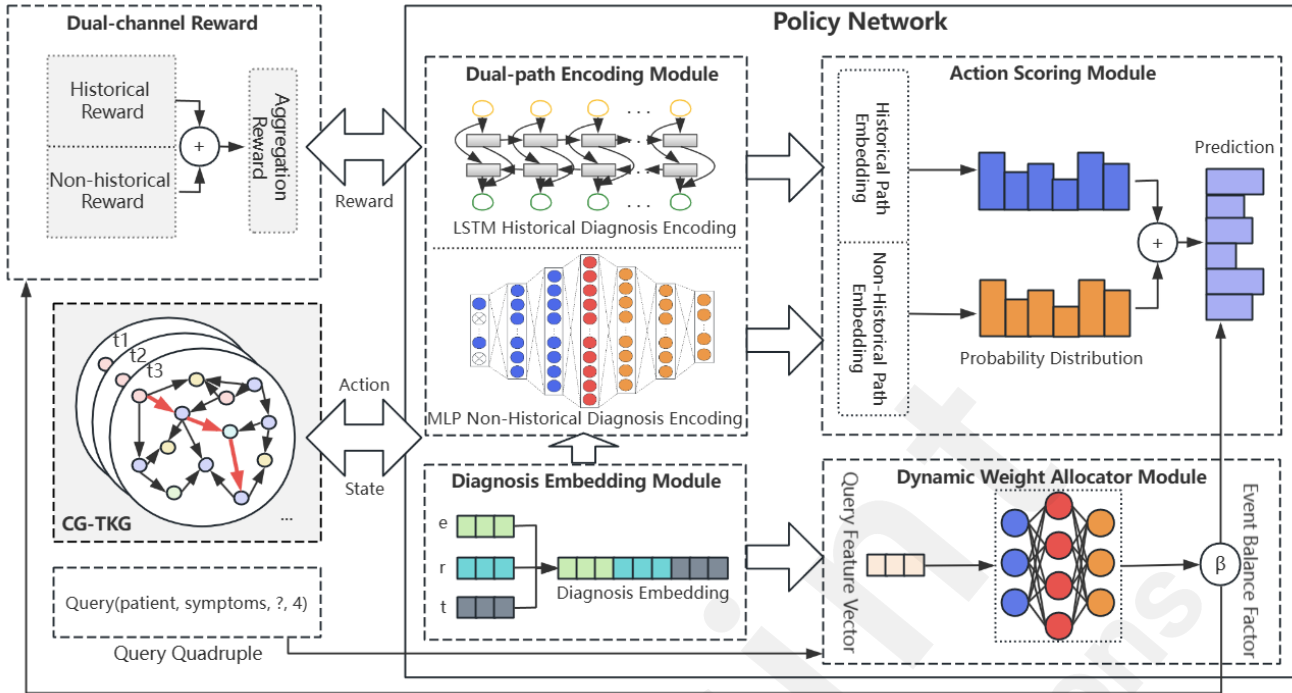


Fig. 1: Framework of the proposed RL4TKGR

3.1 Formal Definitions

3.1.1 Temporal Knowledge Graph

A TKG is defined as $G_{(1,T)} = (G_1, G_2, \dots, G_T)$, where $G_t = (E_t, R_t, F_t)$ denotes a snapshot at timestamp t . Here, E_t represents the entity set, R_t the relation set, and F_t the fact set at t . Each fact is expressed as $F_t = (e_s, r, e_o, t)$, where $e_s, e_o \in E_t$ indicate subject and object entities, and $r \in R_t$ denotes their relation. In the CG-TKG, t denotes the visit number, E_t denotes the set of medical entities at the t -th visit, R_t denotes the set of medical relations at the t -th visit, and F_t denotes the set of medical facts at the t -th visit.

3.1.2 Temporal Knowledge Graph Reasoning Task

The task of the TKG reasoning is defined as performing link prediction in knowledge graphs at future timestamps along the temporal evolution. Given a query q , the missing object entity in the query is predicted using the TKG composed of the set of known facts. In the task of the CG-TKG reasoning, the query q indicates that the missing medical entity is predicted based on the set of known medical facts in the CG-TKG, with a primary focus on predicting the types of chronic gastritis and corresponding therapeutic drugs. For example, the query $(patient\ ID, diagnosis, ?, 4)$ indicates predicting the disease diagnosis of a patient (identified by patient ID) at the fourth clinical visit, based on the set of medical facts from the first three visits.

3.2 Reinforcement Learning Framework

3.2.1 States

Define $\mathcal{S}$ as the state space, where each state can be represented as a 5-tuple $s_n = (e_n, t_n, e_q, t_q, r_q) \in \mathcal{S}$. Here, (e_n, t_n) denotes the node accessed at step n and visible exclusively at step n , with $t_n < t_q$; (e_q, t_q, r_q) represents elements in the query $(e_q, r_q, t_q, ?)$ and is globally visible. The agent uses $s_0 = (e_q, t_q, e_q, t_q, r_q)$ as the initial state.

3.2.2 Actions

Define $\mathcal{A}$ as the action space, and A_n as the set of optional actions at step n , where $A_n = \{(r', e', t') \mid (e_n, r', e', t') \in F_t, t' < t_q, t_n < t_q\}$. Since a patient's disease progresses with each visit, the agent can select actions with timestamps subsequent to t_n during action selection. Therefore, in TKG reasoning task, the agent can select actions from the entire set of known facts.

3.2.3 Transition

Define a state transition function δ to map from state S_n to S_{n+1} during the state transition by the agent, such

that $S \times A \rightarrow S$, i.e., $S_{n+1} = \delta(s_n, a_n) = (e_{n+1}, t_{n+1}, e_q, t_q, r_q)$.

3.2.4 Rewards

A dual-channel aggregated reward function R is defined, integrating historical and non-historical diagnostic information, to guide the agent toward accurate action selection.

3.3 Dual-channel Reward Function

3.3.1 Historical Reward

The historical reward enabled the policy network to focus on historical diagnosis information. When the predicted fact was related to historical interactions, the model incorporated these known facts into the prediction of future facts, as expressed by Equation 1:

$$R_{his}(s_L) = I \setminus \{e_L = e_{\hat{e}} \wedge e_L \in H_q\} \cdot (1 + p_{his}(e_L)) \quad (1)$$

where H_q denotes the set of historical entities, i.e., all entities that appeared in the previous n visits. The historical occurrence frequency p_{his} of entity $e_L \in H_q$ is calculated to participate in the reward function computation.

3.3.2 Non-historical Reward

The non-historical reward did not consider the influence of historical information in the diagnosis process, allowing the predicted fact to be treated as a new potential fact unrelated to historical interactions, with the specific formula as Equation 2:

$$R_{nhis}(s_L) = I \setminus \{e_L = e_{\hat{e}}\} \quad (2)$$

where e_L is the predicted entity of the action finally selected by the agent, and $e_{\hat{e}}$ is the correct entity in the query quadruple.

3.3.3 Dual-channel Aggregation Reward

Formally, the dual-channel aggregated reward function performed a weighted integration of the historical and non-historical reward functions, which can be expressed as Equation 3:

$$R = \alpha \cdot R_{his} + (1 - \alpha) \cdot R_{nhis} \quad (3)$$

where α is a hyperparameter, whose value is provided by the event balance factor β in the dynamic weight allocator module mentioned in Section 3.4.4.

3.4 Policy Network

The policy network $\pi_\theta(a_n | s_n)$ was used to evaluate the action $a_n \in A_n$ selected by the agent, where θ denotes the model parameters. This network consists of four modules: the diagnosis embedding module, the dual-path encoding module, the action scoring module, and the dynamic weight allocator module.

3.4.1 Diagnosis Embedding Module

We began by randomly and uniformly initializing each fact $F_t = (e_s, r, e_o, t)$ within a low-dimensional dense vector space. Since the progression of a patient's disease was closely related to visit number, the entity and visit number were jointly represented as a diagnosis at the current step, i.e., $e_n^t = (e_n, t)$. Entities, relations, and timestamps were represented as $e \in R^{d_e}$, $r \in R^{d_r}$, $t \in R^{d_t}$, where d_e , d_r and d_t denote their embedding dimensions, respectively.

During the patient's visit process, diseases evolve gradually over time. For a prediction query $(e, r, q, r_q, ?, t_q)$, when the agent selected an action, $\Delta t = t_q - t_n$ was used to represent the distance between the query visit number t_q and the current visit number t_n , which was then encoded using an activation function. Thus, the representation of a diagnosis at the current step is as Equation 4:

$$e_n^t = [e_n; \sigma(w \Delta t + b)] \quad (4)$$

where w and b are learnable parameters, σ is an activation function, and $[\cdot]$ denotes the vector concatenation operation.

3.4.2 Dual-path Encoding Module

The dual-path encoding module consists of the LSTM historical diagnosis encoding and the MLP non-historical diagnosis encoding.

(1) LSTM Historical Diagnosis Encoding

For the patient's historical diagnosis information, an LSTM layer was used to encode such information as historical path embedding. The patient's historical diagnosis is represented as Equation 5:

$$h_n = (a_0, a_1, \dots, a_{n-1}) \quad (5)$$

where $a_n \in A_n$ denotes the action selected by the agent at step n , i.e., (r_n, e_n, t_n) .

The h_n is embedded into a continuous vector space R^{3d} and then encoded. The historical path embedding is as

Equation 6:

$$h_n^{his} = LSTM(h_{n-1}, E a_n) \quad (6)$$

where $E a_n$ denotes the embedding of a_n , and $h_0 = LSTM(0, E a_0)$.

(2) MLP Non-Historical Diagnosis Encoding

For the patient's non-historical diagnosis information, a multilayer perceptron was used to extract latent semantic features of non-historical diagnosis entities as non-historical path embedding. The representation of the patient's non-historical diagnosis information consists of a fully connected layer and an activation function, as Equation 7:

$$h_n^{nhis} = MLP(s; p; E a_n) = W2 \cdot ReLU(W1 \cdot (s; p; E a_n) + b1) + b2 \quad (7)$$

where $W1$, $W2$, $b1$ and $b2$ are learnable parameters, $E a_n$ denotes the embedding of a_n , and $[\cdot]$ denotes the vector concatenation operation.

3.4.3 Action Scoring Module

The policy network generated scores for each optional action and calculated state transition probabilities. This module computed state transition probabilities jointly based on the outputs of the dual-path encoding module.

To balance the influence of historical and non-historical diagnoses, the event balance factor β was introduced to perform weighted allocation on their action scores. Ultimately, the probability distribution P over candidate actions is derived according to Equation 8:

$$P = \beta \cdot P_{his} + (1 - \beta) \cdot P_{nhis} \quad (8)$$

where P_{his} is the state transition probability distribution of historical diagnoses, and P_{nhis} is the state transition probability distribution of non-historical diagnoses.

3.4.4 Dynamic Weight Allocation Module

In the dynamic weight allocator module, the event balance factor β was used to dynamically and interpretably adjust the weight allocation in both the dual-channel aggregation reward function and the action scoring module.

The query $q(s, p, ?, t)$ was encoded as $v_q = [s_q; p_q; t_q]$, where s_q is the embedding vector of the query entity, p_q is the embedding vector of the query relation, and t_q is the embedding vector of the visit number. v_q was input into a fully connected layer, and the weight β was generated through activation by the sigmoid function, with the representation of β as Equation 9:

$$\beta = \text{sigmoid}(W \cdot V_q + b) \quad (9)$$

When the query q focuses on historical diagnoses, β tends toward 1; conversely, when the query q focuses on non-historical diagnoses, β tends toward 0. The event balancing factor β enables the model to adaptively and interpretably adjust the decision-making basis according to the specific conditions of the current patient.

3.5 Training Procedure

The event balance factor β was jointly trained with the policy network $\pi_\theta(a_n | s_n)$. The dual-channel aggregation reward R was maximized via the REINFORCE algorithm[35], such that the jointly trained policy network can be expressed as $\pi_{\theta, \beta}(a_n | s_n)$.

With the search space length fixed as L , the joint training objective of β and the policy network were optimized by maximizing the expected reward over the training sample set F_{train} as Equation 10:

$$J(\theta, \beta) = \sum_{(s, p, o, t) \in F_{train}} \left[\sum_{a_1, \dots, a_L \sim \pi} R \right] \quad (10)$$

Optimization was performed with the REINFORCE algorithm by iterating over all quadruples in F_{train} and updating the parameters through the following stochastic gradient as Equation 11:

$$\nabla_{\theta, \beta} J(\theta, \beta) = E_{\pi_{\theta, \beta}} [\nabla_{\theta, \beta} \log \pi_{\theta, \beta}(a | s) \cdot R] \quad (11)$$

Where the value of β only participates in the calculation of the policy gradient and is treated as a constant during the calculation of the reward function, without involvement in the reward gradient calculation.

4 Experiments

4.1 Dataset Description

The CG-TKG dataset used in this work was derived from clinical diagnosis and treatment data of chronic gastritis treated by a chief physician at Guang'anmen Hospital, China Academy of Chinese Medical Sciences (from March 2009 to October 2022). The dataset includes 17,906 patients with a total of 38,360 visit records (the visit number per patient

ranges from 1 to 35). The data covers the following aspects:

(1) Symptoms and TCM diagnoses: 40,993 types of information, including chief complaints, present illness history, and TCM inspection, auscultation-olfaction, inquiry, and palpation (e.g., gastric distension, stomach pain, poor appetite, heartburn, acid regurgitation, white tongue coating, slippery pulse, etc.).

(2) Western medicine diagnoses: 119 types of chronic gastritis diseases, including chronic atrophic gastritis, chronic superficial gastritis, erosive gastritis, *Helicobacter pylori*-associated gastritis, etc.

(3) TCM syndrome diagnoses: 263 types of results, such as spleen deficiency with dampness-heat syndrome, liver-stomach disharmony syndrome, spleen-stomach disharmony syndrome, liver depression and spleen deficiency syndrome, spleen-stomach weakness syndrome, etc.

(4) Laboratory tests: A total of 59 test items (such as comprehensive biochemistry, urine routine, complete blood cell analysis, stool routine + occult blood), involving 139 indicators (such as white blood cells, red blood cells, urine protein, carbohydrate antigen 724, alpha-fetoprotein) and their results.

(5) Gastroscopy and pathological examinations: including 35 items (e.g., active inflammation, intraepithelial neoplasia, atrophy, gastric mucosa texture, gastric mucosa color) and 56 descriptive indicators (e.g., mild, moderate, normal, smooth and soft, rough mucosa).

The CG-TKG dataset was partitioned by visit number: we selected patient data with 4, 8, and 12 visits to construct three temporal subsets. For each subset, the first 3, 7, and 11 visits (from cases with 4, 8, and 12 total visits, respectively) serve as the training set, while the remaining visits are assigned to the validation and test sets. After manual verification by Traditional Chinese Medicine (TCM) experts, three dynamic chronic gastritis clinical datasets were finally obtained—CG-TKG-4, CG-TKG-8, and CG-TKG-12—and Table 1 presents the data statistics of the CG-TKG dataset.

Table 1: CG-TKG Dataset Statistics

Dataset	Entity Count	Relation Count	Training Set Size	Validation Set Size	Test Set Size	Timestamps	Timestamp Interval		Patient Count
							Mean	Variance	
CG-TKG-4	2304	132	161697	27206	27206	4	37.00	84.94	772
CG-TKG-8	1225	116	68357	4797	4797	8	38.06	88.90	140
CG-TKG-12	868	39	36569	3031	3031	12	30.45	62.25	54

4.2 Experimental Settings and Evaluation Metrics

RL4TKGR was initialized with uniform embeddings and implemented using PyTorch. The model parameters were set as follows: the entity embedding dimension was set to 100, the relation embedding dimension to 100, and the visit number dimension to 20. The Adam optimizer was used for parameter optimization with a learning rate of 0.001. During training, the batch size was set to 512.

Mean Reciprocal Rank (MRR) and Hits@1/3/10 were used to evaluate the model's performance in Temporal Knowledge Graph prediction tasks. The formulas and explanations of these evaluation metrics are as follows.

(1) MRR

MRR is defined as the average of the reciprocals of the ranks of the first relevant result for all queries, as shown in Equation 12:

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i} \quad (12)$$

where Q is the set of queries; rank_i is the rank of the first correct answer in the i -th query.

(2) Hits@k

Hits@k is defined as the hit rate of whether the correct answer appears in the top k results, as shown in Equation 13:

$$\text{Hits@k} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} 1(\text{rank}_i \leq k) \quad (13)$$

where $1(\cdot)$ is an indicator function that takes 1 if the condition in the parentheses is satisfied, and 0 otherwise; rank_i is the rank of the first correct answer in the i -th query.

4.3 Introduction to Baseline Models

Here we present the baseline models used for comparison. Static knowledge graph reasoning models include TransE[15], R-GCN[36], SACN[37], and DistMult[17], while the TKG reasoning models include RE-GCN[23], Cen[24], RE-Net[25], CeNet[26], and TiTer[21]. For fairness, all baseline models were configured with their optimal hyperparameters as reported in the original papers or tuned to achieve their best performance.

TransE: Proposed by Bordes *et al.* in 2013, its core idea was to model relations as translation operations in the embedding space, i.e., for a triple (h, r, t) , the geometric constraint $h + r \approx t$ should be satisfied. This model has few parameters, high training efficiency, and can be scaled to large-scale knowledge bases with millions of entities.

R-GCN: Proposed by Schlichtkrull *et al.* in 2018, it extended Graph Convolutional Networks (GCN) to multi-relational data. Its core innovation was assigning independent weight matrices to different relation types to capture local

structural dependencies through neighborhood feature aggregation.

SACN: Proposed by Shang *et al.* in 2019, the model consists of a Weighted GCN (WGCN) and a Conv-TransE decoder. WGCN differentiated the importance of edges by employing a relation weight matrix, whereas Conv-TransE leveraged convolutional layers instead of fully connected layers to capture semantic constraints among triples.

DistMult: Proposed by Yang *et al.* in 2015, it represented relations as diagonal matrices. Its advantage was low parameter count, enabling efficient capture of symmetric relations, but it cannot handle asymmetric relations.

RE-GCN: Proposed by Li *et al.* in 2021, the model integrated GCN, Gated Recurrent Unit (GRU), and static constraint components. It simulated entity evolution paths through a temporal gating mechanism and quantified node structural importance using the k-shell algorithm.

Cen: Proposed by Li *et al.* in 2022, it designed a complex evolution pattern learning framework that decouples the temporal evolution process into periodic and aperiodic patterns. It innovatively introduced a time-aware replication mechanism to dynamically reuse historical states for handling repeated events.

RE-Net: Proposed by Jin *et al.* in 2019, at its core, the model was composed of an RNN temporal encoder and a global neighborhood aggregator, where the former captured historical event sequences and the latter performed parallel processing of concurrent events within the same timestamp.

CeNet: Proposed by Xu *et al.* in 2023, the model used historical contrastive learning to distinguish key historical events through cross-temporal contrastive loss, and designed an entity drift module to address the representation shift problem.

TiTer: Proposed by Sun *et al.* in 2021, it introduced reinforcement learning into TKG reasoning, designed reward functions and agents, leveraging a Markov Decision Process (MDP) to condition on historical states and sample prospective trajectories.

4.4 Comparative Experiments

RL4TKGR was compared with the models in section 4.3. All baseline models used optimal performance parameters. The comparative experimental results are shown in Table 2.

Table 2: Comparative Experimental Results

Model	CG-TKG-4				CG-TKG-8				CG-TKG-12			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
TransE	24.34	1.17	31.13	39.41	23.60	1.89	29.45	37.59	24.92	10.44	32.36	36.08
R-GCN	13.39	6.21	14.51	27.34	15.72	7.21	15.33	28.51	14.15	6.31	14.25	31.17
SACN	20.28	11.80	20.99	36.95	22.91	18.76	27.13	39.35	21.89	9.93	25.27	39.08
DistMult	26.83	14.69	30.48	53.50	25.80	16.06	30.81	49.80	29.88	22.12	38.07	45.77
Cen	29.55	19.51	32.88	47.69	30.40	18.94	34.49	50.06	31.19	20.91	36.71	51.02
RE-GCN	30.98	21.60	34.50	49.22	31.55	20.84	35.69	51.19	32.13	23.34	38.23	53.59
RE-Net	31.89	22.40	34.55	50.17	32.49	21.07	37.45	53.09	35.38	25.67	38.90	55.07
CeNet	35.11	29.45	50.71	59.37	37.69	30.96	50.39	60.25	40.32	31.42	52.69	63.92
TiTer	34.02	31.48	35.40	39.98	50.56	<u>47.97</u>	52.64	53.93	49.55	42.05	56.57	62.36
RL4TKGR	<u>51.66</u>	<u>35.39</u>	<u>53.14</u>	<u>60.15</u>	<u>53.18</u>	45.48	<u>59.52</u>	<u>67.26</u>	<u>56.95</u>	<u>49.75</u>	<u>62.54</u>	<u>71.79</u>

Table 2 shows the comparison of prediction performance between RL4TKGR and various static knowledge graph reasoning models and TKG reasoning models on the three temporal datasets (CG-TKG-4/8/12). The underlined data represent the indicators ranking first, and the bolded data represent the indicators ranking second. Experimental results demonstrate that TKG reasoning models significantly outperformed static approaches overall, validating the importance of modeling temporal information for analyzing the evolutionary process of chronic gastritis.

RL4TKGR achieved better performance on all datasets and evaluation metrics, except that the Hits@1 metric on the CG-TKG-8 dataset was 2.19% lower than that of the TiTer model. Unlike the lenient metric Hits@10, Hits@1 metric requires the model to place the correct answer at the first rank. The 2.19% gap may reflect that TiTer has more accurate local reasoning capabilities for a small number of specific short-term events, while the global historical modeling of RL4TKGR may disperse the priority of these events. Compared with baseline models, RL4TKGR achieved an average improvement of 8.86% in MRR, 3.04% in Hits@1, 5.09% in Hits@3, and 5.52% in Hits@10. Especially on the long-term evolution dataset CG-TKG-12, RL4TKGR's Hits@10 reached 71.79%, significantly outperforming CeNet's 63.92%, which proves that the model can effectively capture the long-term development laws of chronic gastritis through modeling historical diagnosis and treatment information. Experimental findings substantiate that RL4TKGR, through the integration of temporal information and historical diagnostic features, was capable of accurately capturing the complex interaction patterns underlying disease progression.

4.5 Ablation Study

Ablation experiments were designed to verify the effectiveness of the historical diagnosis encoding module and the historical reward in RL4TKGR. The experiments removed the historical diagnosis encoding module (w/o e), removed the historical reward (w/o r), removed both simultaneously (w/o e&r), and evaluated their performance on the three datasets CG-TKG-4, CG-TKG-8, and CG-TKG-12. Table 3 reports the experimental results, where the full model (RL4TKGR) achieves significant improvements over all ablated versions across every dataset and evaluation metric, thereby validating the critical contribution of both modules.

Specifically, on the CG-TKG-4 dataset, the MRR of the complete model (51.66%) was 17.3%, 24.2%, and 54.1% higher than that of w/o e (44.05%), w/o r (41.59%), and w/o e&r (33.54%), respectively. This trend is consistent across the Hits@1, Hits@3, and Hits@10 metrics. Especially in Hits@10, the complete model (66.55%) was 85.4% higher than the worst-performing w/o e&r (35.89%), indicating that the integration of historical information significantly improved model capability. Similarly, the ablation experimental results on the CG-TKG-8 and CG-TKG-12 datasets showed the same trend as CG-TKG-4.

Comparing the impact of ablating a single module, the removal of the historical diagnosis encoding module (w/o e) caused slightly more damage to performance than the removal of the historical reward (w/o r). This indicates that the diagnosis encoding module played a more fundamental role in modeling patients' historical states. Nevertheless, when both modules were ablated simultaneously (w/o e&r), the performance exhibited a steep decline, suggesting that the historical reward substantially enhanced the effectiveness of the diagnosis encoding module and that the two components demonstrated a strong complementary relationship.

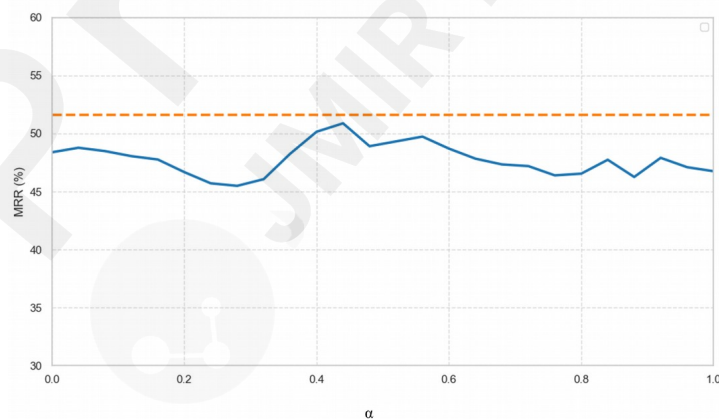
Table 3: Ablation Study Results

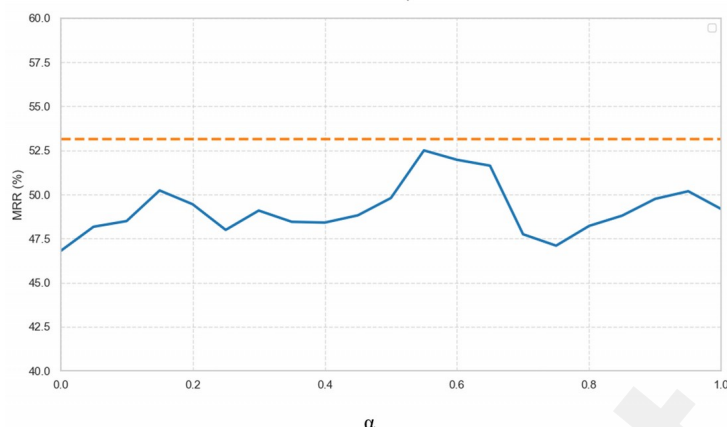
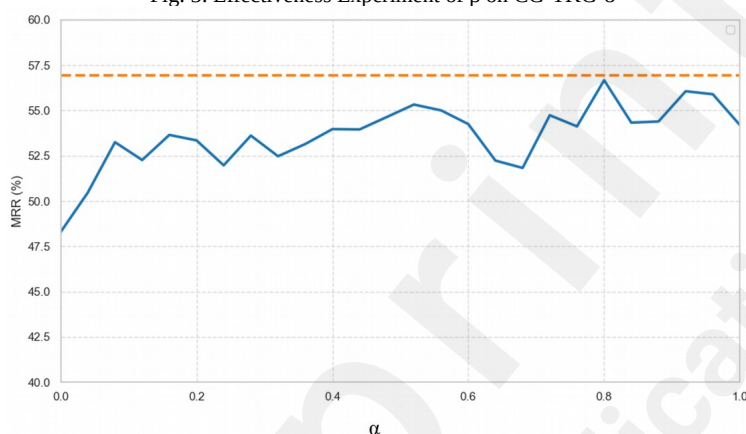
Model	CG-TKG-4				CG-TKG-8				CG-TKG-12			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
w/o e	44.05	40.45	45.25	56.44	50.53	41.91	55.95	64.86	54.30	44.93	63.54	67.22
w/o r	41.59	39.38	43.17	54.69	49.55	40.04	54.56	62.35	53.16	44.72	60.88	66.39
w/o e&r	33.54	32.35	34.05	35.89	34.49	31.52	36.56	39.52	42.95	36.51	44.76	57.66
RL4TKGR	51.66	44.49	55.74	66.55	53.18	45.48	59.52	67.26	56.95	49.75	62.54	71.79

4.6 Experiments on the Effectiveness of the Event Balancing Factor in the Reward Function

In the RL4TKGR model, the reward function calculated the weighted value of the historical reward and the non-historical reward using a hyperparameter α , where α was provided by the value of the event balance factor β . To verify the impact of the event balance factor β in the reward function on model performance, experiments were conducted on three different datasets: CG-TKG-4, CG-TKG-8, and CG-TKG-12. The hyperparameter α in the reward function was fixed to guide the learning of the policy network, and hyperparameter search was performed on each dataset with α ranging from 0 to 1. The experimental results are shown in Figs. 2~4, where the dashed line represents the MRR value of RL4TKGR for reasoning on the three datasets, and the polyline represents the MRR value of the hyperparameter experiment with a fixed α value. In CG-TKG-4, when α was fixed, the MRR value fluctuated between 45% and 51%, while the MRR value of the RL4TKGR model was 51.66%, which was significantly higher than the results of model training guided by reward functions with any fixed α value. The same trend was observed in datasets CG-TKG-8 and CG-TKG-12.

In summary, the application of the event balance factor β in the reward function has a significant effect on improving model performance. Especially when trained jointly with the policy network, β can adaptively adjust the weights of historical and non-historical information, thereby optimizing the overall performance of the model. Therefore, dynamically adjusting β was an effective mean to improve model performance.

Fig. 2: Effectiveness Experiment of β on CG-TKG-4

Fig. 3: Effectiveness Experiment of β on CG-TKG-8Fig. 4: Effectiveness Experiment of β on CG-TKG-12

4.7 Interpretability Explanation of the Event Balance Factor in the Reward Function

In the dynamic weight allocator module, the event balance factor β was used to dynamically and interpretably adjust the weight distribution between the dual-channel aggregation reward function and the action scoring module. When the query focuses on historical diagnosis information, β approaches 1; when the query focuses on non-historical diagnosis information, β approaches 0.

β enabled the model to automatically and interpretably adjust the source weights of decision-making bases according to the patient's specific current conditions. In the design of the dual-channel aggregation reward function, the value of α was directly provided by β . As shown in the hyperparameter experimental results in Figs. 2~4, the interpretability is explained as follows:

(1) Within CG-TKG-4, in which the maximum number of patient visits is four, the probability of disease progression is relatively high whereas historical correlations are weak. The model attained its highest MRR when α lied between 0.4 and 0.5, providing evidence of its interpretability.

(2) Within CG-TKG-8, in which the maximum number of patient visits is eight, the disease state tends to stabilize, the likelihood of progression decreases, and historical correlations become stronger. The model achieved its highest MRR when α lied between 0.5 and 0.6.

(3) Within CG-TKG-12, where the maximum number of patient visits is twelve, the disease state is the most stable, the likelihood of progression is the lowest, and historical correlations are the strongest. The model attained its highest MRR when α was approximately 0.8.

4.8 Case Analysis of Chronic Gastritis Diagnosis Prediction

Three visits of patient 49375 were selected as examples for analysis.

First visit: The visit time is June 22, 2013. The manifestations included occasional stomach pain after acute gastroenteritis, abdominal distension, poor belching, foreign body sensation in the pharynx, phlegm in the pharynx, normal appetite, average sleep quality with many dreams, and normal urination and defecation. The tongue was dark red with thin white coating, and the pulse was deep and thready. The diagnosis was chronic gastritis, and the prescribed medicines were Zhizhu Kuanzhong Capsules and Qingyan Tablets.

Second visit: The visit time is July 13, 2013. The manifestations were disappearance of stomach pain, occasional abdominal distension, foreign body sensation in the pharynx, poor belching, phlegm in the pharynx, normal appetite, average sleep quality, improvement in dreaminess, and normal urination and defecation. The tongue was dark red with

thin white coating slightly yellowish, and the pulse was deep and thready. The diagnosis was chronic gastritis, and the prescribed medicines were Zhizhu Kuanzhong Capsules and Qingyan Tablets.

Third visit: The visit time is August 3, 2013. The manifestations were loose stools in the past week, occasional abdominal distension, poor belching, phlegm in the pharynx, normal appetite, average sleep quality with many dreams, and normal urination. The tongue was dark red with thin white coating, and the pulse was deep and thready.

Using RL4TKGR, based on the patient's historical diagnosis information and the symptom information from the third visit, the top 10 probabilistic reasoning results for the queries (*Patient 49375, prescribe medicine, ?, 3*) and (*Patient 49375, disease diagnosis, ?, 3*) were generated. As shown in Table 4, the bolded entities represent the actual results of the patient's third visit. Analysis shows that in both query cases, RL4TKGR ranked the correct result first in the prediction sequence—confirming its ability to predict the development path of chronic gastritis diagnosis and treatment, and demonstrating its effectiveness in capturing the complex relationships between temporal information and medical entities.

Ranking	(Patient 49375, prescribe medicine, ?, 3)	(Patient 49375, disease diagnosis, ?, 3)
	Predicted Entities	Predicted Entities
1	Zhizhu Kuanzhong Capsules	Chronic gastritis
2	Guben Yichang Tablets	Chronic atrophic gastritis
3	Simo Decoction Oral Liquid	Chronic superficial gastritis
4	Bazhen Yimu Pills	Erosive gastritis
5	Qingyan Tablets	Helicobacter pylori-associated gastritis
6	Xuhanting Capsules	Spleen deficiency and dampness-heat syndrome
7	Wuweizi Granules	Mild dysplasia of chronic gastritis
8	Qizhi Weitong Granules	Chronic gastritis with intestinal metaplasia
9	Xuhan Weitong Granules	Chronic gastritis with hiatal hernia
10	Jinghua Weikang Capsules	Liver stagnation and spleen deficiency syndrome

Meanwhile, the fourth visit result of patient 49375, with the visit time being September 21, 2013, shows significant improvement in the condition: alleviated abdominal distension, belching, reduced night sweats, general chills, disappearance of phlegm, normal appetite, average sleep, sometimes formed stools, and normal urination. Diagnosed with chronic gastritis. The patient's condition improved significantly after taking the prescribed drugs, and the further progression of chronic gastritis was controlled, further confirming the clinical effectiveness of the predicted drugs. These results indicated that RL4TKGR can effectively predict future disease status and recommend appropriate therapies, thus serving as a reliable data-driven tool to support clinical decision-making.

5 Conclusion

We investigated the temporal dependency of chronic gastritis progression and, based on CG-TKG constructed from traditional Chinese medicine diagnostic and treatment data, proposed RL4TKGR, a Reinforcement Learning-based Temporal Knowledge Graph Reasoning model. By designing a reinforcement learning policy network that exploited the associations among historical diagnosis data, the model achieved interpretable prediction of chronic gastritis diagnosis and treatment. Experimental results validated its effectiveness and demonstrate its superiority in TKG reasoning tasks for chronic gastritis.

RL4TKGR provided novel technical support for clinical diagnosis decision-making of chronic gastritis. Its TKG reasoning approach not only enhanced the accuracy of disease progression warning but also offered a transferable solution for temporal data analysis in other complex diseases. Future research can further optimize the model's adaptability in medical environments under temporal features, explore fusion mechanisms for multimodal data (such as imaging, text, physiological signals, etc.) to improve the model's perception and decision-making capabilities, or combine the powerful semantic understanding and generation capabilities of Large Language Models (LLMs) to deepen the knowledge reasoning process, expanding its potential applications in multi-disease joint prediction and the generation of personalized treatment strategies.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China (NO. 82374621) and supported by the Research Project of China Academy of Chinese Medical Sciences "Evidence Study Based on Multimodal Knowledge Graph Reasoning of the Idea of Treating Pre-disease in TCM (2023016)".

References

- [1] Banks M, Graham D, Jansen M, *et al.* British Society of Gastroenterology guidelines on the diagnosis and management of patients at risk of gastric adenocarcinoma [J]. Gut, 2019, 68(9): 1545-75.
- [2] Zhang Y, Li F, Yuan F, *et al.* Diagnosing chronic atrophic gastritis by gastroscopy using artificial intelligence[J]. Digestive and Liver Disease, 2020, 52(1): 34-40.

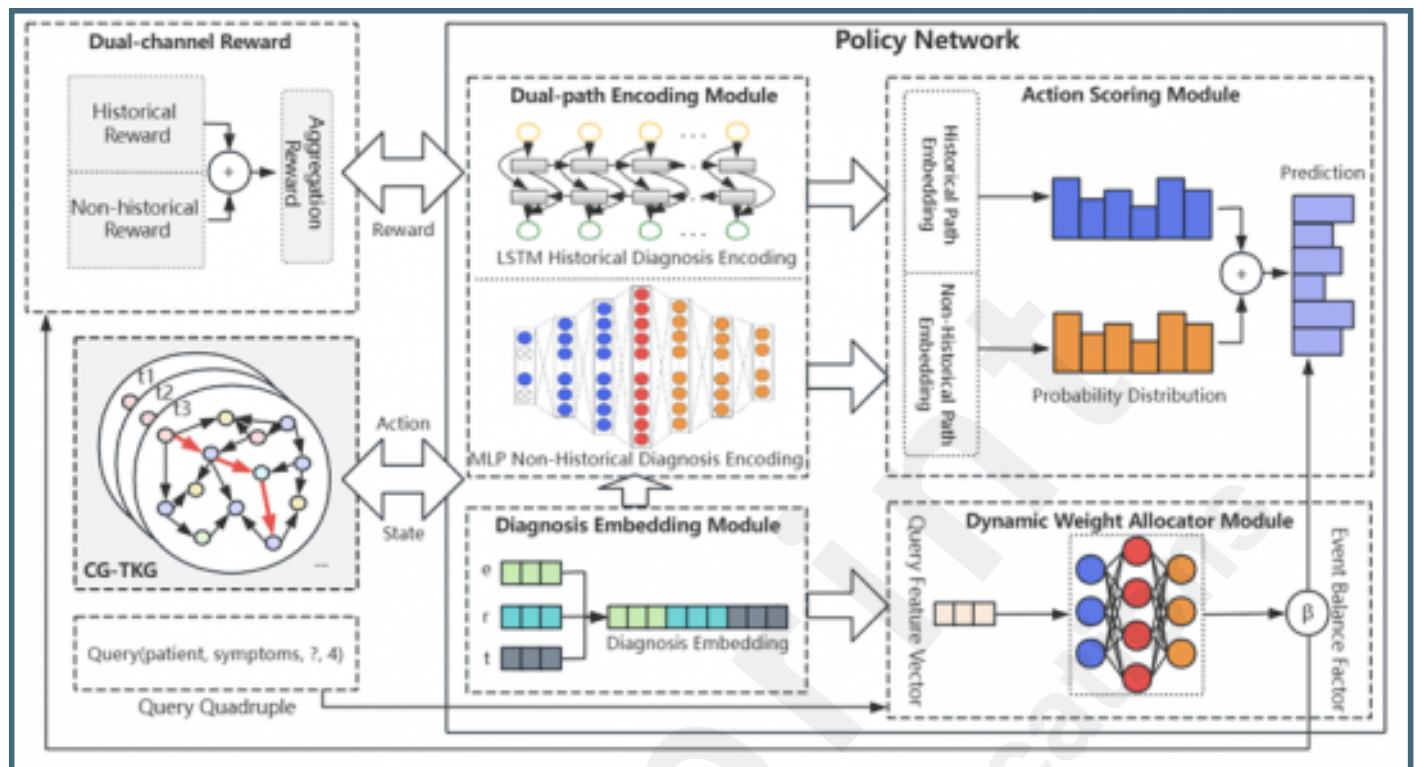
- [3] Turtoi D C, Brata V D, Incze V, *et al.* Artificial Intelligence for the Automatic Diagnosis of Gastritis: A Systematic Review[J]. *Journal of Clinical Medicine*, 2024, 13(16): 4818.
- [4] Luo J, Cao S, Ding N, *et al.* A deep learning method to assist with chronic atrophic gastritis diagnosis using white light images[J]. *Digestive and Liver Disease*, 2022, 54(11): 1513-1519.
- [5] Tao X, Zhu Y, Dong Z, *et al.* An artificial intelligence system for chronic atrophic gastritis diagnosis and risk stratification under white light endoscopy[J]. *Digestive and Liver Disease*, 2024.
- [6] Kim H H, Lim Y S, Seo S I, *et al.* A deep recurrent neural network-based explainable prediction model for progression from atrophic gastritis to gastric cancer[J]. *Applied Sciences*, 2021, 11(13): 6194.
- [7] Huang R J, Huang E S, Mudiganti S, *et al.* Risk of gastric adenocarcinoma in a multiethnic population undergoing routine care: an electronic health records cohort study[J]. *Cancer Epidemiology, Biomarkers & Prevention*, 2024, 33(4): 547-556.
- [8] Kim H H, Lim Y S, Seo S I, *et al.* A deep recurrent neural network-based explainable prediction Model for progression from Atrophic Gastritis to gastric cancer[J]. *Applied Sciences*, 2021, 11(13): 6194.
- [9] Jiang T, Liu T, Ge T, *et al.* Towards time-aware knowledge graph completion[C]//*Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. 2016: 1715-1724.
- [10] Yin Z, Kuang Z, Zhang H, *et al.* Explainable AI method for tinnitus diagnosis via neighbor-augmented knowledge graph and traditional Chinese medicine: Development and validation study[J]. *JMIR Medical Informatics*, 2024, 12: e57678.
- [11] Song F, Wang B, Tang Y, *et al.* Research of medical aided diagnosis system based on temporal knowledge graph[C]//*Advanced Data Mining and Applications: 16th International Conference*, 2020: 236-250.
- [12] Bourque V R, Poulain C, Proulx C, *et al.* Genetic and phenotypic similarity across major psychiatric disorders: a systematic review and quantitative assessment[J]. *Translational Psychiatry*, 2024, 14(1): 171.
- [13] Chen S, Wang J. A survey on temporal knowledge graphs-extrapolation and interpolation tasks[C]//*The international conference on natural computation, Fuzzy systems and knowledge discovery*. Cham: Springer International Publishing, 2022: 1002-1014.
- [14] Jiang T, Liu T, Ge T, *et al.* Encoding temporal information for time-aware link prediction[C]//*Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. 2016: 2350-2354.
- [15] Bordes A, Usunier N, Garcia-Duran A, *et al.* Translating embeddings for modeling multi-relational data [J]. *Advances in neural information processing systems*, 2013, 26.
- [16] Garcia-Duran A, Dumančić S, Niepert M. Learning sequence encoders for temporal knowledge graph completion[C]//*Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2018: 4816-4821.
- [17] Yang B, Yih S W, He X, *et al.* Embedding entities and relations for learning and inference in knowledge Bases[C]//*Proceedings of the International Conference on Learning Representations (ICLR) 2015*. 2015.
- [18] Goel R, Kazemi S M, Brubaker M, *et al.* Diachronic embedding for temporal knowledge graph completion[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. 2020, 34(04): 3988-3995.
- [19] Goel R, Kazemi S M, Brubaker M, *et al.* Diachronic embedding for temporal knowledge graph completion[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. 2020, 34(04): 3988-3995.
- [20] Han Z, Chen P, Ma Y, *et al.* xerte: Explainable reasoning on temporal knowledge graphs for forecasting future links[J]. *arXiv preprint arXiv:2012.15537*, 2020.
- [21] Sun H, Zhong J, Ma Y, *et al.* Timetraveler: reinforcement learning for temporal knowledge graph forecasting[C]//*Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. 2021: 8306-8319.
- [22] Liu R, Yin G, Liu Z, *et al.* Reinforcement learning with time intervals for temporal knowledge graph reasoning[J]. *Information Systems*, 2024, 120: 102292.
- [23] Li Z, Jin X, Li W, *et al.* Temporal knowledge graph reasoning based on evolutionary representation learning[C]//*Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*. 2021: 408-17.
- [24] Li Z, Guan S, Jin X, *et al.* Complex evolutionary pattern learning for temporal knowledge graph reasoning [J]. *arXiv preprint arXiv:220307782*, 2022.
- [25] Jin W, Jiang H, Qu M, *et al.* Recurrent event network: Global structure inference over temporal knowledge graph [J]. 2019.
- [26] Xu Y, Ou J, Xu H, *et al.* Temporal knowledge graph reasoning with historical contrastive learning[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. 2023: 4765-73.
- [27] Wang C, Zheng Z, Cai X, *et al.* A review of knowledge graph applications in the medical field[J]. *Journal of*

- Biomedical Engineering, 2023, 40(5): 1040-1044.
- [28] Carvalho R M S, Teixeira A S, Pesquita C. Building a Temporal Knowledge Graph for Electronic Health Records[C]//CEUR Workshop Proceedings. CEUR-WS, 2024, 3833.
 - [29] Geng S, Tao B, Liang G, *et al.* Temporal knowledge graph attention network for online doctor recommendation[C]// Proceedings of the 2023 8th International Conference on Intelligent Information Processing. 2023: 277-282.
 - [30] Chaturvedi R. Temporal knowledge graph extraction and modeling across multiple documents for health risk prediction[C]//Companion Proceedings of the ACM Web Conference 2024. 2024: 1182-1185.
 - [31] Postiglione M, Bean D, Kraljevic Z, *et al.* Predicting future disorders via temporal knowledge graphs and medical ontologies[J]. IEEE Journal of Biomedical and Health Informatics, 2024.
 - [32] Lippman S A. Dynamic programming and markov decision processes[M]//The New Palgrave Dictionary of Economics. Palgrave Macmillan, London, 1987: 1-7.
 - [33] Chaudhari P A, Khot S S. Alzheimer's disease prediction using CAdam optimized reinforcement learning-based deep convolutional neural network model[J]. Biomedical Signal Processing and Control, 2025, 108: 107968.
 - [34] Zhang H, Diao M, Zhang S, *et al.* Optimizing Vital Signs in Patients With Traumatic Brain Injury: Reinforcement Learning Algorithm Development and Validation[J]. Journal of Medical Internet Research, 2025, 27: e63847.
 - [35] Williams, Ronald J. Simple statistical gradient-following algorithms for connectionist reinforcement learning. Machine learning 1992, 8: 229-256.
 - [36] Schlichtkrull M, Kipf T N, Bloem P, *et al.* Modeling relational data with graph convolutional networks[C]//The semantic web: 15th international conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, proceedings 15.Springer,2018:593-607.
 - [37] Shang C, Tang Y, Huang J, *et al.* End-to-end structure-aware convolutional networks for knowledge base completion[C]//Proceedings of the AAAI conference on artificial intelligence.2019:3060-7.

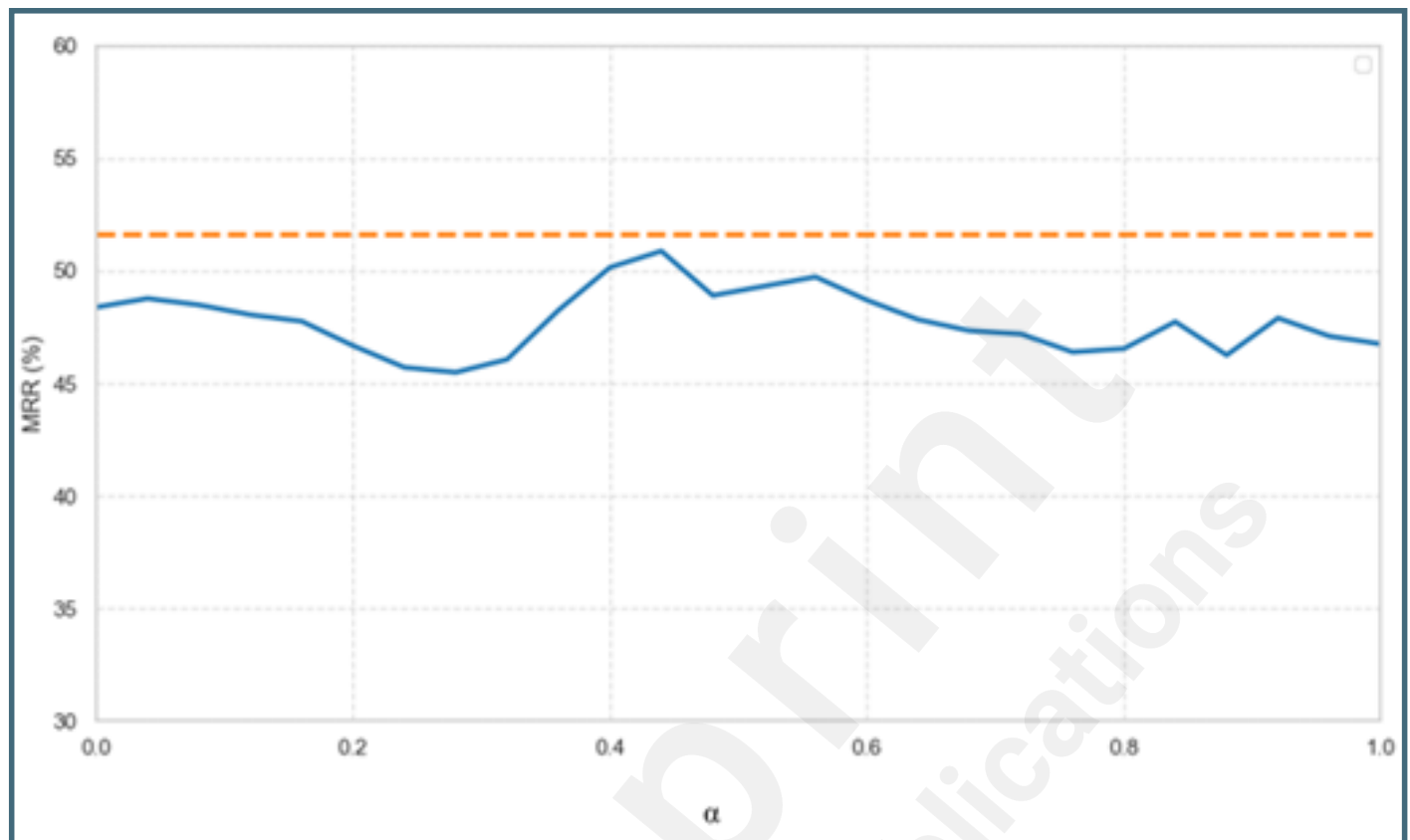
Supplementary Files

Figures

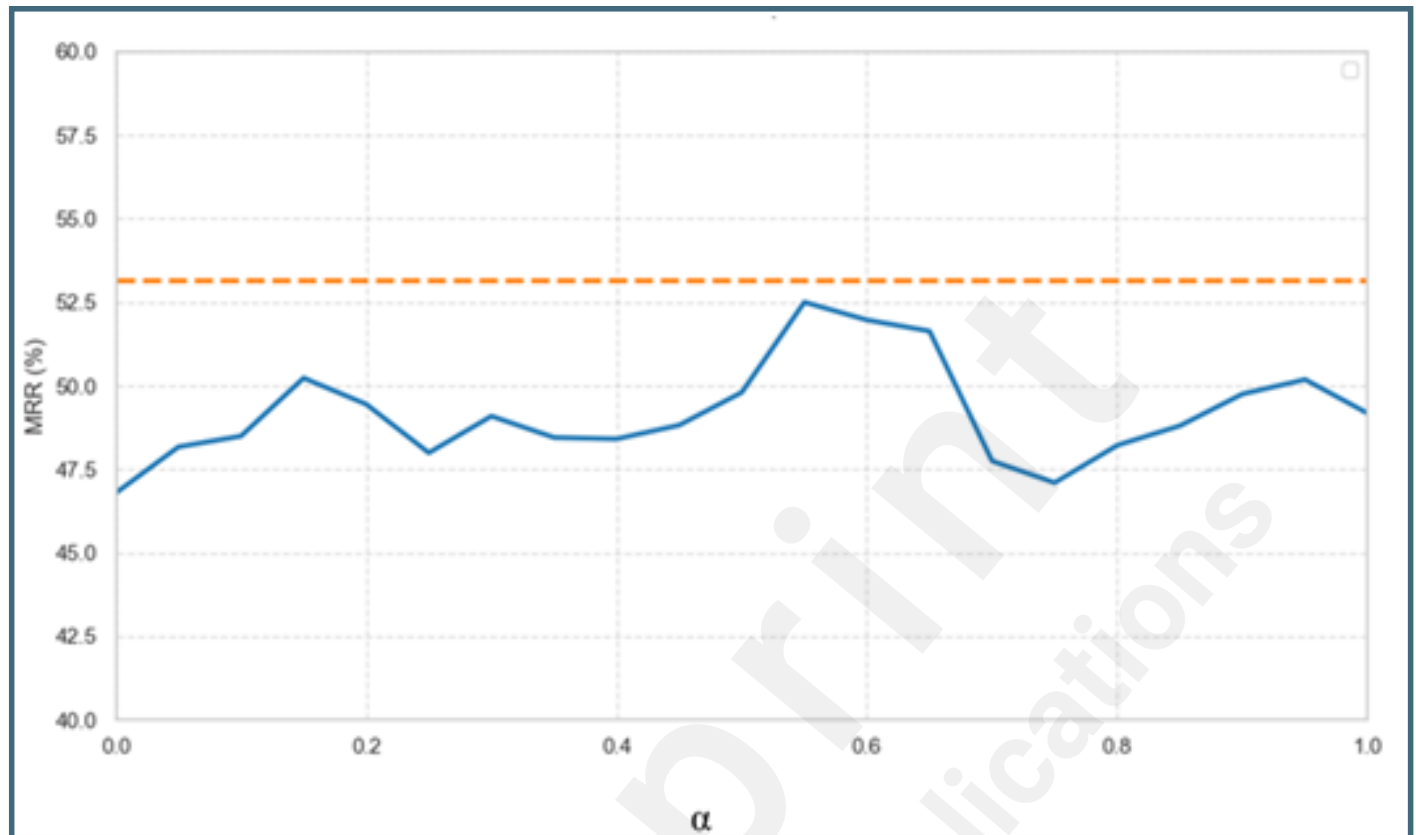
Framework of the proposed RL4TKGR.



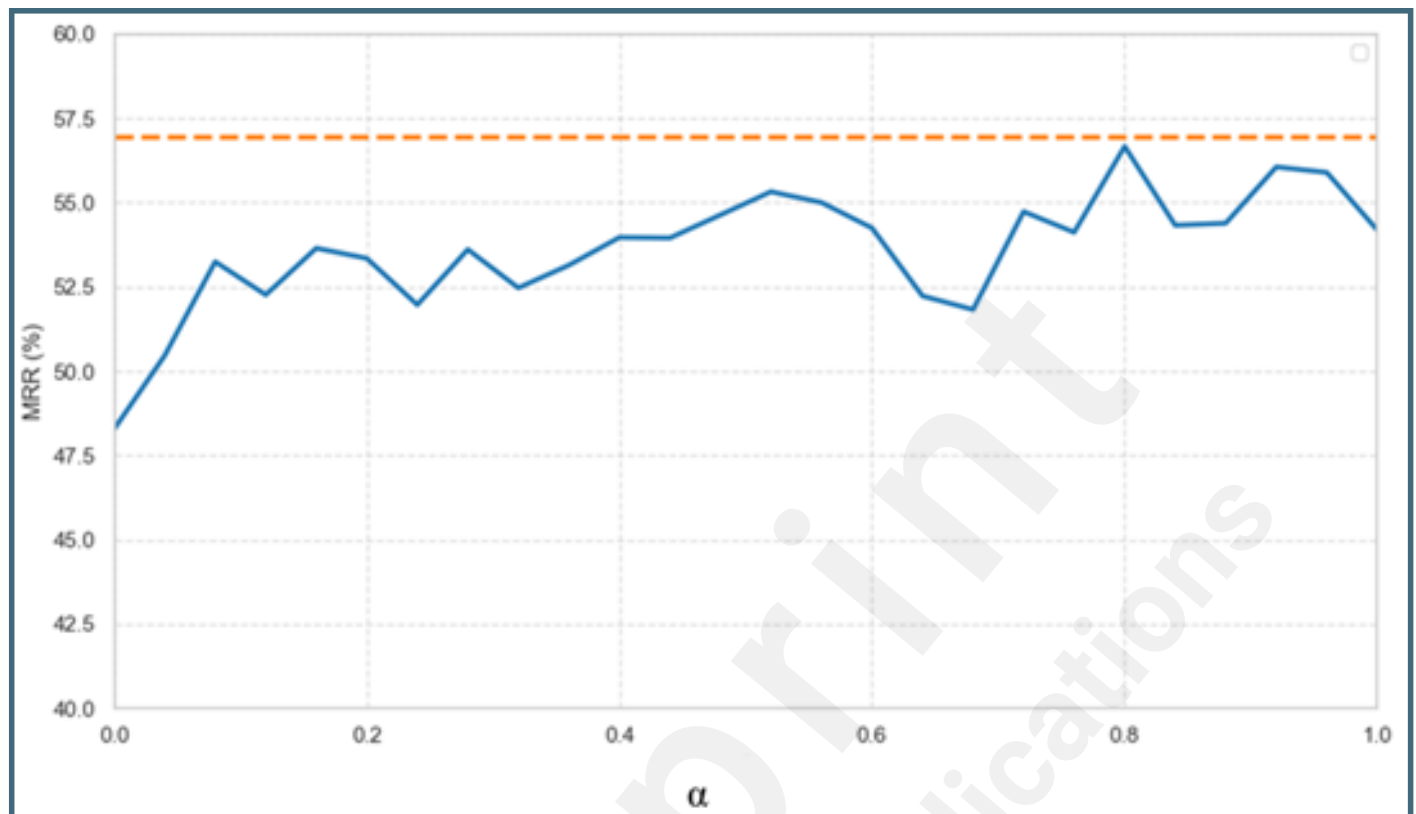
Effectiveness Experiment of ? on CG-TKG-4.



Effectiveness Experiment of ? on CG-TKG-8.



Effectiveness Experiment of ? on CG-TKG-12.



TOC/Feature image for homepages

Framework of the proposed RL4TKGR.

