TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 19/24 pp1554–1565
DOI: 10.26599/TST.2023.9010142
Volume 29, Number 5, October 2024

A Temporal Knowledge Graph Embedding Model Based on Variable Translation

Yadan Han, Guangquan Lu*, Shichao Zhang, Liang Zhang, Cuifang Zou, and Guoqiu Wen

Abstract: Knowledge representation learning (KRL) aims to encode entities and relationships in various knowledge graphs into low-dimensional continuous vectors. It is popularly used in knowledge graph completion (or link prediction) tasks. Translation-based knowledge representation learning methods perform well in knowledge graph completion (KGC). However, the translation principles adopted by these methods are too strict and cannot model complex entities and relationships (i.e., N-1, 1-N, and N-N) well. Besides, these traditional translation principles are primarily used in static knowledge graphs and overlook the temporal properties of triplet facts. Therefore, we propose a temporal knowledge graph embedding model based on variable translation (TKGE-VT). The model proposes a new variable translation principle, which enables flexible transformation between entities and relationship embedding. Meanwhile, this paper considers the temporal properties of both entities and relationships and applies the proposed principle of variable translation to temporal knowledge graphs. We conduct link prediction and triplet classification experiments on four benchmark datasets: WN11, WN18, FB13, and FB15K. Our model outperforms baseline models on multiple evaluation metrics according to the experimental results.

Key words: knowledge graph; knowledge graph completion; variable translation; temporal properties; link prediction; triplet classification

1 Introduction

Knowledge graphs (KGs) are large-scale semantic network graphs in which nodes represent entities and edges represent relationships between them. Each edge in KGs corresponds to a fact, represented by a triplet (h, r, t), h and t represent the head and tail entities, respectively, and r represents the relationship. Nowadays, more and more large-scale knowledge graphs are constructed, such as Freebase[1], WordNet[2], YAGO[3], and NELL[4], they usually consist of a large number of facts in the real world and have been widely used in different fields, such as semantic analysis[5–7], node classification[8–11], text classification[12], and personalized diagnosis[13], etc. However, since KGs are constructed manually or semi-manually, the incomplete phenomenon of knowledge graphs is common. For example, there are about three million human entities in Freebase, but 71% of the entities have no birthplace information, and 91% of the entities have no education information[14]. Therefore, predicting missing entities or relationships in the triplets, which is called knowledge graph completion (KGC), has always been a vital issue. KGC aims to solve the data sparseness problem in KGs and improve their integrity.

• Yadan Han, Guangquan Lu, Shichao Zhang, Liang Zhang, Cuifang Zou, and Guoqiu Wen are with Key Lab of Education Blockchain and Intelligent Technology, Ministry of Education, Guangxi Normal University, Guilin 541004, China and also with Guangxi Key Lab of Multi-Source Information Mining and Security, Guangxi Normal University, Guilin 541004, China. E-mail: hanyadan@stu.gxnu.edu.cn; lugq@mailbox.gxnu.edu.cn; zhangsc@csu.edu.cn; zhangliangcsuer@gmail.com; 2335907631@qq.com; wenguoqiu2008@163.com.
* To whom correspondence should be addressed.
Manuscript received: 2023-06-20; revised: 2023-10-13; accepted: 2023-11-21

Knowledge representation learning is popularly used in knowledge graph completion tasks due to its high computational efficiency and low complexity. Generally, traditional translation-based methods (such as TransE[15] and TransH[16], etc.) all adopt the same translation principle $h+r\approx t$, such principle is too strict and cannot model complex entities and relationships (i.e., N-1, 1-N, and N-N). For example, for the 1-N relationship is_author_of, there are three triplets associated with it, i.e., (WilliamShakespeare, is_author_of, Hamlet), (WilliamShakespeare, is_author_of, Othello), and (WilliamShakespeare, is_author_of, Macbeth). As shown in Fig. 1a, considering the ideal embedding of $h+r\approx t$ in TransE, the entities Hamlet, Othello, and Macbeth will get the same embedding vector. To alleviate this problem, TransF[17] and TransE-DT[18] are proposed. TransF presents a flexible translation principle (FT): $h+r\approx \alpha t$ (or $t-r\approx \alpha h$), which extends the embedding range of entities and relationships to a line. As shown in Fig. 1b, considering the flexible translation principle $h+r\approx \alpha t$ in TransF, the entities Hamlet, Othello, and Macbeth will get the embedding vectors with the same direction but different magnitude. The principle of flexible translation proposed by TransF is universal and effective, but it limits the distribution of entity orientations. TransE-DT proposes a dynamic translation principle (DT) $(h+\alpha_h)+(r+\alpha_r)\approx(t+\alpha_t)$ to solve this problem. DT extends the embedding range of entities and relationships to a plane by introducing parameters α_h, α_r, and α_t. As shown in Fig. 1c, the ideal embedding of DT is that the embeddings of the three entities Hamlet, Othello, and Macbeth are flexible in magnitude and direction. However, we noticed that the DT model also has flaws, e.g., if h and r are given, $t+\alpha_t$ is a fixed vector. For example, if given WilliamShakespeare and is_author_of, Hamlet+α_{Hamlet}, Othello+α_{Othello}, and Macbeth+α_{Macbeth} are fixed vectors. Besides, these traditional translation principles are mostly used in static knowledge graphs. However, many facts in knowledge graphs are not static; they are usually only true for a certain period or timestamp. For example, the triplet (Bill_Clinton, president of, US) is true just from 1993 to 2001, and the triplet (Steve_Jobs, diedin, California) is true only on October 5, 2011.

To solve the above problems, this paper proposes a temporal knowledge graph embedding model based on variable translation, TKGE-VT. The proposed method further relaxes the constraint of the translation principle in DT and proposes a new variable translation principle (VT): $(h+\beta_h+\varphi)$ + $(r+\beta_r+\lambda)$ $\approx$ $(t+\beta_t+\alpha)$. In our model, the embedding range of entities and relationships is still a plane. Unlike DT, we do not strictly restrict $t+\alpha_t$ to be a fixed vector, but allow them to be in the same direction and the magnitude is flexible. As shown in Fig. 1d, the ideal embedding of VT is that the embeddings of the three entities of Hamlet, Othello, and Macbeth are on a plane, and the directions of Hamlet+β_{Hamlet}, Othello+β_{Othello}, and Macbeth+β_{Macbeth} are same, but the magnitudes are flexible. Moreover, the proposed method considers the temporal properties of entities and relationships in

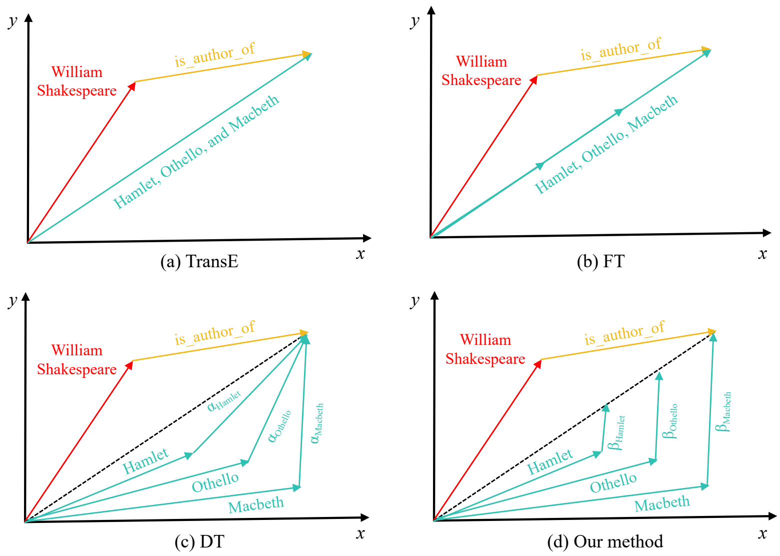

Fig. 1 Illustration of TransE, FT, DT, and our method.

triplets and applies the proposed principle of variable translation to temporal knowledge graphs (TKGs).

In order to verify the effectiveness and applicability of the model, we conducted link prediction experiments on WN18 and FB15K, as well as experiments on triplet classification on WN11, WN18, and FB13. The evaluation results show that TKGE-VT has significantly improved in MeanRank (MR), Hits@10, and accuracy (ACC) metrics.

The main contributions of this paper are as follows:

- This paper proposes a temporal knowledge graph embedding model based on variable translation — TKGE-VT. The model proposes a variable translation principle (VT), which relaxes the strict constraints of the translation principle in DT, and can better model triplet facts;
- The proposed model considers the temporal properties of triplet facts, adding temporal properties to the head entity, relationship and tail entity, respectively, and applies the proposed principle of variable translation to temporal knowledge graphs;
- We conduct link prediction and triplet classification experiments on four real knowledge graph datasets, and the experimental results show that our method outperforms classic models such as FT and DT.

2 Related Work

2.1 Static knowledge graph representation

Knowledge representation learning represented by translation-based methods has attracted extensive attention from many researchers. TransE[15] is the first proposed translation-based knowledge representation learning model. For each triplet (h,r,t), TransE regards tail entity t as the translation from head entity h by relation r, i.e., when triplet (h,r,t) holds, $h+r\approx t$. TransE performs well when dealing with 1-1 relationships but cannot handle complex and diverse relationships such as N-1, 1-N, and N-N relationships well. To alleviate this problem, TransH[16] proposed an improved translation model on the hyperplane. For the triplet (h,r,t), TransH projects the embeddings of head entity h and tail entity t into a relation-specific hyperplane by the relation-specific mapping matrix w_r. Both TransE and TransH project entities and relationships to the same vector space, but they belong to different types of objects. To this end, TransR[19] is proposed. TransR projects entities and relationships to entity space and relationship space, respectively, then defines a mapping matrix M_r to project entity embeddings into relationship space. In TransR, the mapping matrix is the same for all entities. However, different types of entities have different properties. TransD[20] is proposed based on this idea. Given a triplet (h,r,t), TransD defines two mapping matrices for the head entity and tail entity respectively: $\boldsymbol{M}_{rh}=\boldsymbol{l}_{r_\mathrm{p}}\boldsymbol{l}_{h_\mathrm{p}}+\boldsymbol{I}^{d\times k}$ and $\boldsymbol{M}_{rt}=\boldsymbol{l}_{r_\mathrm{p}}\boldsymbol{l}_{t_\mathrm{p}}+\boldsymbol{I}^{d\times k}$, where $\boldsymbol{l}_{h_\mathrm{p}}$, $\boldsymbol{l}_{t_\mathrm{p}}$, $\boldsymbol{l}_{r_\mathrm{p}}\in R^k$. In order to solve the problem of entity distribution optimization, TranSHER[21] leverages relation-specific translations between head and tail entities to relax the constraint of hyper-ellipsoid restrictions, which can provide more direct guidance on optimization by introducing an intuitive and straightforward relation-specific translation. To solve the problem that the TransE model cannot handle complex attributes well, TransP[22] proposes a knowledge graph embedding model based on entity attributes and relationship attributes, it introduces the idea of hyperplane projection to map the head entity and tail entity to the plane of a specific relationship to enhance the model's ability to handle complex relationships. However, these traditional translation-based models all adopt the same translation principle $h+r\approx t$, which is too strict and cannot model complex entities and relationships.

2.2 Temporal knowledge graph representation

Temporal knowledge graph is obtained by adding temporal information to the traditional knowledge graph. Compared with static knowledge graphs, it can better capture the time properties of triplet facts, so temporal knowledge graph completion has become a current research hotspot. In order to make knowledge graphs dynamic, t-TransE[23] adds a separate time dimension in triplet (h,r,t) to extend the fact representation in knowledge graphs into a quadruple (h,r,t,T), T represents the time dimension feature of the triplet. BoxTE[24] proposes a box embedding model for temporal knowledge graph completion (TKGC), building on the static knowledge graph embedding model BoxE[25], which allows to additionally capture inference patterns across time and model certain temporal relational information. ATiSE[26] incorporates time information into entity/relationship representations using Additive Time Series decomposition. Moreover, considering the temporal uncertainty during the evolution of entity/relationship representations over time, ATiSE maps the representations of temporal knowledge graphs into the space of multi-dimensional Gaussian distributions. EvoKG[27] captures the ever-changing structural and temporal dynamics in TKGs via recurrent event modeling and models the interactions between entities based on the temporal neighbourhood aggregation framework. Further, EvoKG achieves accurate modeling of event time using flexible and efficient mechanisms based on neural density estimation. TLT-KGE[28] models semantic information and temporal information as different axes of complex number or quaternion space. Meanwhile, two specific components carving the relationship between semantic and temporal information are devised to buoy the modeling. Inspired by diachronic word embeddings, R. Goel et al.[29] defines entity embedding as a function that takes entities and timestamps as input, which maintains time-aware features of entities at any time point. T-GAE[30] employs LSTM network to learn new time-aware relational embeddings to incorporate time information. Then, it utilizes these time-aware relational embedding and GATs considered as neighbourhood aggregators to learn the entity and relational features of the central entity neighbourhoods. In this paper, we also consider the temporal properties of triplet facts, different from the above methods, we learn the temporal properties of the head entity, relationship, and tail entity, respectively. Thus, we can represent the correct triplet for any timestamp instead of just representing triplet facts at a specific timestamp or period as in t-TransE. Finally, we apply the proposed principle of variable translation to temporal knowledge graphs.

3 Method

In this paper, we propose a temporal knowledge graph embedding model based on variable translation — TKGE-VT, which can better model complex and diverse entities/relationships. TKGE-VT is also a translation-based model, which adopts the variable translation principle to relax the strict constraint of the translation principle in DT. Besides, to consider the temporal properties of triplet facts, the proposed method extends the variable principle to temporal knowledge graphs.

3.1 Motivation

Knowledge representation learning methods are effective for knowledge graph completion tasks. Traditional translation-based methods adopt the translation principle of $h+r\approx t$ (or $h+r\approx \alpha t$). However, such translation principles are too strict and cannot correctly model complex entities and relationships. DT model adopts the dynamic translation principle and extends the range of entities and relationships embedding vectors to a plane, which relaxes the strict constraints of the traditional translation principle. However, we noticed that the DT model also has flaws, e.g., if the head entity h and relationship r are given, $t+\alpha_t$ is a fixed vector, as shown in Fig. 1c. To alleviate this issue, the proposed method further relaxes the constraint of the translation principle in DT, so that the model can better handle complex entities and relationships.

Furthermore, the triplet facts are valid just for a specific timestamp or period, that is, the facts in knowledge graphs all have temporal properties. However, we note that traditional translation principles are primarily used in static knowledge graphs and ignore the temporal properties of triplet facts. The temporal properties provide more helpful information about entities and relationships. Therefore, we consider applying the proposed principle of variable translation to temporal knowledge graphs.

3.2 TKGE-VT

3.2.1 Principle of variable translation

In order to solve the existing problems in DT model (i.e., the translation principle is too strict and ignores the temporal properties of triplet facts), we propose a temporal knowledge graph embedding model based on variable translation —TKGE-VT. The key intuition behind TKGE-VT is to provide a higher degree of freedom for the embeddings of entities and relationships. TKGE-VT proposes a variable translation principle to relax the strict constraint of the translation principle in DT. Specifically, if the head entity h and relationship r are given, we allow the embedding range of t to be a plane, and the magnitude of the embedding vector $t+\beta_t$ is flexible, α as a hyperparameter to adjust the magnitude of $t+\beta_t$ (see Fig. 2a), where β_t is the parameter vector of the tail entity. Similarly, if the tail entity t and relationship r are given, we allow the embedding range of h to be a plane, and the magnitude of the embedding vector $h+\beta_h$ is flexible, φ as a hyperparameter to adjust the magnitude of $h+\beta_h$ (see Fig. 2b), where β_h is the parameter vector of the head

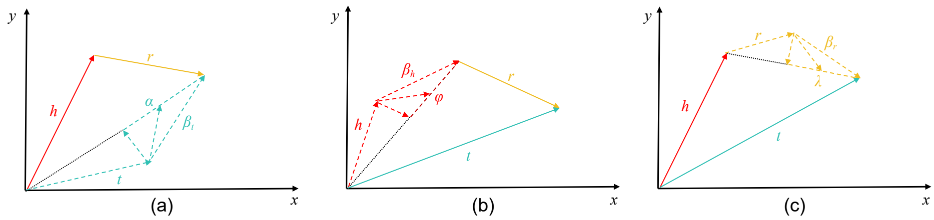

Fig. 2 Principle of variable translation.

entity; if the head entity h and tail entity t are given, we allow the embedding range of r to be a plane, and the magnitude of the embedding vector $r+\beta_r$ is flexible, λ as a hyperparameter to adjust the magnitude of $r+\beta_r$ (see Fig. 2c), where β_r is the parameter vector of the relationship. Therefore, the variable translation principle is defined as

$$(h+\beta_h+\varphi)+(r+\beta_r+\lambda)\approx(t+\beta_t+\alpha) \tag{1}$$

where φ, λ, and $\alpha \in \mathrm{R}^{m\times n}$. Correspondingly, the scoring function of TKGE-VT is

$$f_r(h,t)=\|(h+\beta_h)+(r+\beta_r)-(t+\beta_t)+\mu\|_{L_1/L_2} \tag{2}$$

where μ is a hyperparameter, and $\mu=\varphi+\lambda-\alpha$. L_1 and L_2 represent L_1-norm and L_2-norm, respectively.

For the above variable translation principle, we give the following theoretical proof:

(1) In the case that there are many triplets $(h,r_i,t)\in S$ with $i\in 1,2,\ldots,n$, we will get $r_1+\beta_{r_1}=t-h-\lambda_1,\ldots,r_n+\beta_{r_n}=t-h-\lambda_n$. In this way, $r+\beta_r$ is no longer a fixed vector, but a set of vectors with different magnitudes in the same direction.

(2) In the case that r is an 1-N relationship with triplets $(h,r,t_i)\in S, i\in 1,2,\ldots,n$, we will get $t_1+\beta_{t_1}=h+r-\alpha_1,\ldots,t_n+\beta_{t_n}=h+r-\alpha_n$. Therefore, $t+\beta_t$ is not a fixed vector, but a set of vectors with variable magnitudes in the same direction.

(3) In the case that r is a N-1 relationship with triplets $(h_i,r,t)\in S, i\in 1,2,\ldots,n$, we will get $h_1+\beta_{h_1}=t-r-\varphi_1,\ldots,h_n+\beta_{h_n}=t-r-\varphi_n$. In this way, $h+\beta_h$ is also not a fixed vector, but a set of vectors of different magnitudes in the same direction.

3.2.2 Time-aware embedding

The DT model projects entity and relationship vectors into the same semantic space for the static knowledge graphs. However, temporal property is the primary source of N-1, 1-N, and N-N relationships. For example, for 1-N relationships: (h,r) pairs may be associated with different tail entities t at different time points; similarly, for N-1 relationships: (r,t) pairs may be associated with different head entities h at different time points.

To eliminate the ambiguity in the above issue, we consider that both entities and relationships all have temporal properties and learn time-aware embeddings of them, respectively. Specifically, we add a separate time dimension T to the triplet to make the knowledge graphs dynamic and represent triplet facts with temporal annotations by quadruples (h,r,t,T). We use $(h,r,t,[T_s,T_e])$ to denote the fact that h and t has relation r during the time interval $T=[T_s,T_e]$, where T_s and T_e denote the start and end time during which the triplet (h,r,t) is valid. For some facts that happened at a certain time and did not last, we have $T_s=T_e$; for some facts that do not end yet, we represent T as $T=[T_s,+\infty]$. Unlike t-TransE, we incorporate this time meta-fact directly into our learning algorithm to learn temporal embeddings of the knowledge graph elements. Given the timestamps, the dynamic graph can be dismantled into several static graphs consisting of triplets that are valid in the respective time step, e.g., knowledge graph can be expressed as $G=G_{T_1}\cup G_{T_2}\cup\cdots\cup G_{T_i}$, where $T_i, i\in 1,2,\ldots,$ T are the discrete time points, and $G_{T_1},G_{T_2},\ldots,G_{T_i}$ represent static graphs at time points $T_1,T_2,\ldots,T_n$, respectively. We use TransE to learn time-aware embeddings of head entities, relationships and tail entities, and denote them as h_T, r_T, and t_T, respectively. Therefore, our method can represent the correct triplet for any timestamp, instead of just representing triplet facts at a specific timestamp or period like previous models.

Therefore, the joint temporal properties' head entity, relationship, and tail entity can be denoted as $h+h_T$, $r+r_T$, and $t+t_T$. In order to apply the proposed variable translation principle to the temporal knowledge graph, we incorporate the learned time-aware embeddings of head entities, relationships and tail entities into the variable translation principle, respectively. The variable translation principle incorporating entity-aware embedding of entities and relationships can be expressed as $(h+h_T+\beta_h+\varphi)+(r+r_T+\beta_r+\lambda)\approx(t+t_T+\beta_t+\alpha)$. Thus, for the given embeddings of h and r, tail entity t is expressed as $t+t_T+\beta_t+\alpha$ (see Fig. 3a); for the given embeddings of t and r, head entity h is expressed as $h+h_T+\beta_h+\varphi$ (see Fig. 3b); for the given embeddings h and t, relationship r is expressed as $r+r_T+\beta_r+\lambda$ (see Fig. 3c). Therefore, we design the scoring function of TKGE-VT as

$$f_r(h,t) = \|(h+h_T+\beta_h)+(r+r_T+\beta_r)-(t+t_T+\beta_t)+\mu\|_{L_1/L_2} \tag{3}$$

3.3 Training

In order to improve training speed and model prediction accuracy, it is necessary to construct negative triplets from positive triplets. Given a positive triplet (h,r,t), traditional methods randomly select an entity or relationship in knowledge graphs to replace the entity/relationship in (h,r,t) to obtain negative triplets. However, the random negative sampling method usually generates low-quality negative triplets and only effectively handles 1-1 relationships. When dealing with 1-N, N-1, and N-N relationships, it is possible to mark the original positive triplet as the negative triplet. To avoid this issue, we use the probabilistic method to construct negative samples, i.e., we replace h and t with different probabilities. Meanwhile, when replacing entities in triplets to generate negative triplets, we choose entities with similar semantics to improve the ability of the model to distinguish entities.

3.3.1 Replace head and tail entities by probabilistic method

For the relationship r, the number of corresponding head entities and tail entities is often unbalanced, and this will generate some false negative samples using the traditional random negative sampling method. For example, for a positive triplet (Beijing, City_of, China), replacing the head entity may generate a negative triplet (Shanghai, City_of, China), and it is still a correct triplet, we call such triplet false negative triplet. In order to reduce the probability of generating false negative samples, we set different replacement probabilities for h and t, respectively according to relationship types, that is, we replace h with a higher probability for 1-N relationships and replace t with a higher probability for N-1 relationships. For example, the relationship "Gender" has many head entities, while the tail entities are only male and female, we are more likely to get a natural negative triplet by replacing the tail entity.

Given a relationship r and all positive triplets associated with it during model training, we first calculate the average number of tail entities per head entity: tqh, and the average number of entities per head entity: hqt. Then we adopt the probability $\mathrm{tqh}/(\mathrm{tqh}+\mathrm{hqt})$ to replace the head entities and the probability $\mathrm{hqt}/(\mathrm{tqh}+\mathrm{hqt})$ to replace the tail entities. This sampling method reduces the probability of false negative samples and greatly reduces the computational complexity of the model.

We stipulate that if tqh $<$ 1.5 and hqt $<$ 1.5, r is regarded as 1-1; if tqh $>$ 1.5 and hqt $>$ 1.5, r is regarded as N-N; if tqh $\geqslant$ 1.5 and hqt $<$ 1.5, r is regarded as 1-N; if tqh $<$ 1.5 and hqt $\geqslant$ 1.5, r is regarded as N-1.

3.3.2 Select entities based on semantic similarity

In knowledge graphs, we find that entities with the same type tend to be distributed in a close area in vector space. For example, for the relationship "lives in", its corresponding head entity is often name of person and tail entity is often name of area, name of person will be concentrated in one area and name of area will be concentrated in another area. However, some head and tail entities may gather in close area, as shown in Fig. 4, distinguishing entities gathered in the same area is often difficult. Therefore, during negative sampling, we select entities with similar semantics for replacement so that the model can better distinguish entities.

Generally speaking, the higher entity similarity denotes that the semantics of entities are closer. Therefore, we judge the similarity of entities based on their semantic similarity. Semantic similarity is usually

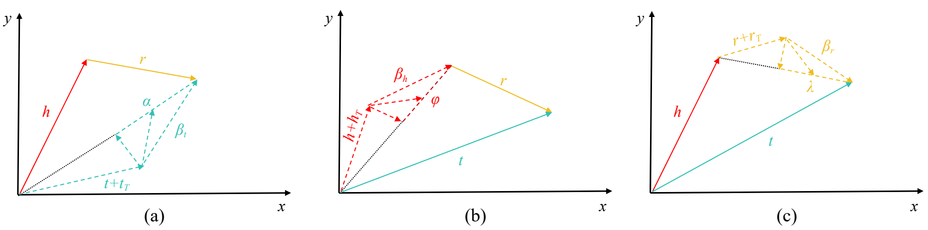

Fig. 3 Simple illustration of TKGE-VT.

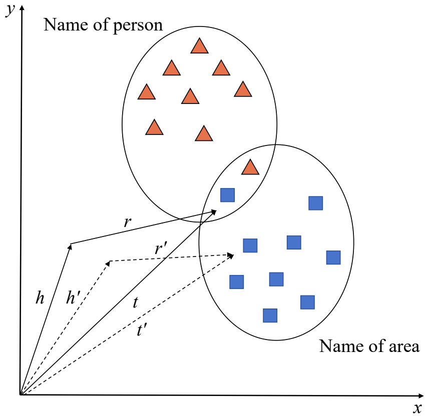

Fig. 4 Examples of entities that are not easily distinguishable.

represented by the vector similarity in distributed representation models[31], defined as

$$\operatorname{dis}(h, h') = \sqrt{\sum_{j=1}^{k} \left(h_j - h'_j\right)} \tag{4}$$

where h_j represents the head entity of the triplet that exists in the knowledge base, and h'_j represents the head entity selected to replace when corrupting the positive triplet to generate the negative triplet. The smaller value of dis(,) represents the higher similarity of entities. Therefore, given a positive sample (h,r,t), when replacing h to generate a negative triplet $(h^{'},r,t)$, h' is chosen to ensure $\operatorname{dis}(h,h^{'})$ is the smallest. Similarly, when replacing t to generate a negative sample $(h,r,t^{'})$, we choose t' to guarantee $\operatorname{dis}(t,t^{'})$ is the smallest.

During model training, to encourage discrimination between positive and negative triplets, we adopt the following margin-based ranking loss function as the training objective:

$$L = \sum_{(h,r,t)\in S} \sum_{(h',r,r')\in S'} \max\left(f_r(h,t) + \gamma - f_r(h',t'), 0\right) \tag{5}$$

In Eq. (5), S is the set of positive triplets, $S^{'}$ is the set of negative triplets, $\max(x,y)$ denotes return the maximum value of x and y, and γ is the margin of positive and negative triplets. Therefore, the goal of the loss function is to distinguish the correct triplets from the wrong triplets as much as possible.

4 Experiment

We evaluate the proposed method on link prediction and triplet classification tasks. In this section, we introduce the experimental implementation in detail, then present and analyze the experimental results.

4.1 Dataset

In the experiment, we evaluate the proposed TKGE-VT for link prediction and triplet classification tasks on the benchmark datasets, including WN18, FB15K, WN11, and FB13. FB15K and FB13 are two subsets of Freebase, a large-scale KG containing many knowledge facts. WN11, and WN18 are subsets of WordNet, a database featuring lexical relations between words. The details of the above datasets are shown in Table 1.

4.2 Link prediction

Given a triplet $(?,r,t)$ or $(h,r,?)$, the link prediction task aims to predict the missing h or t in the triplet. The task emphasizes the ranking of the correct head entity or tail entity, rather than just finding the best one. In this task, we select WN18 and FB15K as evaluation datasets. Besides, to prove that our model can handle complex relationships better, we also performed complex relationships experiment on FB15K.

Evaluation metrics. Similar to traditional translation-based methods, we report two evaluation metrics: MeanRank (MR) and Hits@10, where MeanRank represents the mean rank of correct triplets and Hits@10 represents the proportion of correct triplets ranked in the top 10. The above two evaluation metrics are consistent with the purpose of the link prediction task, so we consider these two evaluation metrics. A good model should achieve lower MR and higher Hits@10.

Implimentation. In the experiment, we selected the learning rate α in {0.0001, 0.001, 0.005, 0.01}, margin γ in {0.25, 0.5, 1.5, 2, 4, 4.5}, embedding dimension k in {50, 100, 150, 200}, hyperparameter μ in {0, 0.1, 0.5, 1, 2}, random number u in {0.0001, 0.001}, and batch size B in {20, 75, 120, 1200, 4800, 9600}. The validation set determines the optimal parameters. In the "unif" setting, the optional configurations are: α = 0.0001, $\gamma = 5$, $k = 100$, $\mu = 0.1$, $B = 4800$ on WN18; α = 0.0001, γ = 4, k = 200, μ = 0.1, B = 4800 on FB15K. In the "bern" setting, the optional configurations are: α =

Table 1 Dataset statistics.

Dataset	Ent	Rel	Train	Valid	Test
WN18	40 943	18	141 442	5000	5000
FB15K	14 951	1345	483 142	50 000	59 071
WN11	38 696	11	112 581	2609	10 544
FB13	75 043	13	316 232	5908	23 733

0.0001, $\gamma = 4$, $k = 200$, $\mu = 0.1$, $B = 4800$ on WN18; $\alpha = 0.001$, $\gamma = 4$, $k = 200$, $\mu = 0.1$, $B = 4800$ on FB15K. For both datasets, all training triplets will be trained for 500 epochs. In order to be consistent with the implementation of the baseline model, we conducted each set of experiments ten times and took the average of these ten results as the final results.

Experimental results and analysis. Table 2 shows the link prediction results. From Table 2, we can see that our model does not perform as well as the baseline models on WN18. This may be due to the limited number of relationships in WN18, and our model's advantage in handling complex relationships is not reflected. However, TKGE-VT achieves state-of-the-art performance on both MeanRank and Hit@10 on FB15K. Among them, on the Hits@10 metric, TKGE-VT has increased by 3.1% and 7.7% respectively on FB15K compared with TransE-DT. This may be because FB15K contains more complex relationships (i.e., 1-N, N-1, and N-N), and our model can better adapt to complex datasets, which shows that TKGE-VT can better handle complex relationships.

To further verify the ability of TKGE-VT to handle complex relationships, we conducted a complex relationships experiment on FB15K. Table 1 shows that FB15K contains 1345 relationships, in which 26.2% are 1-1 relationships, 22.7% are 1-N relationships, 28.3% are many-to-one relationships and 22.8% are many-to-many relationships. Therefore, FB15K is considered a large dense dataset. In this experiment, we use the optimal parameter combination on FB15K to test the scores under 1-1, 1-N, N-1, and N-N relationships, respectively. As can be seen from Table 3, the proposed model achieves state-of-the-art performance on Predicting Left under 1-N, N-1, N-N relationships and Predicting Right. Among them, on Predicting Left, Hit@10 of 1-N relationships reached 96.9%; on Predicting Right, Hit@10 of N-1 relationships reached 95.6%, which further proves that TKGE-VT can improve the performance of the model in handling complex relationships.

4.3 Triplet classification

For a triplet (h,r,t), the goal of triplet classification is to determine whether it is correct, this is one of the knowledge graph completion tasks. In this task, we need to set a relation-specific threshold σ_r, which is obtained when the accuracy of triplet classification on the validation set is maximized. For a given triplet (h,r,t), if the score obtained by the scoring function is less than the set threshold σ_r, then the triplet will be classified as positive. Otherwise, it is a negative triplet. In the experiment, we selected three public datasets, WN11, FB13, and FB15K, the details are shown in Table 1.

Evaluation metric. For triplet classification task, we use accuracy (ACC) as the evaluation metric. ACC represents the proportion of positive triplets and

Table 2 Link prediction results.

Method	WN18				FB15K			
	MeanRank		Hit@10 (%)		MeanRank		Hit@10 (%)	
	Raw	Filt	Raw	Filt	Raw	Filt	Raw	Filt
TransE[15]	263	251	75.4	89.2	243	125	34.9	47.1
TransH[16]	401	388	73.0	82.3	212	87	45.7	64.4
DistMult[32]	987	902	79.2	93.6	224	97	51.8	82.4
TranSparse[33]	223	211	80.1	93.2	190	82	53.7	79.9
STransE[34]	219	206	80.9	93.4	219	68	51.6	79.7
TransE-RS[35]	385	371	80.4	93.7	161	63	53.2	72.1
TransE-DT[18]	228	216	76.2	88.4	212	80	50.7	72.1
TransAt[36]	169	157	**81.4**	**95.0**	185	82	52.9	78.2
OPTransE[37]	211	199	79.2	91.7	141	53	51.0	69.9
RPJE[38]	205	183	79.1	91.1	186	50	51.5	70.3
ERDERP[39]	258	246	79.9	93.2	189	54	49.1	71.1
TransP[22]	160	144	78.7	91.7	151	57	50.9	71.0
TransE-ETF(max)[40]	**137**	**125**	77.2	88.8	149	47	53.7	74.9
TKGE-VT (unif)	316	299	78.3	93.2	172	**24**	53.2	**83.2**
TKGE-VT (bern)	353	336	78.4	92.7	**140**	37	**53.8**	79.8

Table 3 Hit@10 of each type of relations in FB15K.

(%)

Method	Predicting Left (Hits@10)				Predicting Right (Hits@10)			
	1-1	1-N	N-1	N-N	1-1	1-N	N-1	N-N
TransE[15]	43.7	65.7	18.2	47.2	47.2	19.7	66.7	50.0
TransH[16]	66.8	87.6	28.7	28.7	65.5	39.8	83.3	67.2
TranSparse[33]	87.1	95.8	44.4	81.2	87.5	57.0	94.5	83.7
STransE[34]	82.8	94.2	50.4	80.1	82.4	56.9	93.4	83.1
TransE-RS[35]	87.4	96.3	35.3	71.7	86.5	44.2	95.4	75.2
TransE-DT[18]	84.7	94.5	34.1	72.1	83.1	41.7	93.8	74.6
OPTransE[37]	**93.1**	93.4	55.0	80.8	90.8	57.4	91.7	81.3
TransMS[41]	91.4	95.9	44.9	78.5	91.6	54.1	93.6	82.0
RPJE[38]	92.2	96.0	54.4	81.6	91.1	73.9	91.3	83.3
ERDERP[39]	78.1	86.3	49.5	70.3	79.1	51.7	85.6	73.3
TransP[22]	87.4	92.6	44.2	66.5	85.9	45.5	92.2	76.4
TKGE-VT (unif)	91.4	**96.9**	**67.4**	81.7	91.2	**74.9**	94.5	**84.4**
TKGE-VT (bern)	92.4	96.6	57.2	**81.8**	**93.0**	60.1	**95.6**	84.3

negative triplets correctly predicted by the model to the positive triplets and negative triplets in the training set, a better model should have a higher ACC value. Since triplet classification aims to determine whether a given triplet is correct and consistent with the ACC metric, we use ACC as the evaluation metric for triplet classification. ACC is defined as follows:

$$\text{ACC} = \frac{T_{\text{pos}} + T_{\text{neg}}}{N_{\text{pos}} + N_{\text{neg}}} \tag{6}$$

where T_{pos} is the number of correctly predicted positive triplets, T_{neg} is the number of correctly predicted negative triplets, N_{pos} and N_{neg} denote the number of positive triplets and negative triplets in the training set, respectively.

Implimentation. In the SGD process, we chose the learning rate α in {0.001, 0.01, 0.1}, margin γ in {1, 2, 4, 4.5, 5, 10}, embedding dimension k in {20, 50, 100, 200}, hyperparameter μ in {0, 0.1, 0.5, 1, 2}, random number u in {0.0001, 0.001}, and batch size B in {20, 120, 480, 960, 4800}. We obtain the optimal configuration parameters by the accuracy on the validation set. The optional configuration on WN11 is: $\alpha = 0.001$, $\gamma = 10$, $k = 100$, $\mu = 0.1$, $B = 4800$, and L_1 as dissimilarity; the optimal configuration on FB13 is: $\alpha = 0.001$, $\gamma = 5$, $k = 200$, $\mu = 0.1$, $B = 4800$, and L_1 as the dissimilarity; the optimal configuration on FB15K is: $\alpha = 0.001$, $\gamma = 5$, $k = 100$, $\mu = 0.1$, $B = 120$, and L_1 as the dissimilarity.

Experimental results and analysis. Table 4 lists the experimental results of triplet classification. From Table 4, we can see that the proposed method—TKGE-VT achieves optimal performance on WN11, FB13, and FB15K. Compared with TransE-DT, TKGE-VT improved by 1.2%, 6.7%, and 10.4% on WN11, FB13, and FB15K, respectively. Among them, we find that the performance improvement of TKGE-VT on FB15K is the most significant, which indicates that TKGE-VT is effective in dealing with complex relationships.

Table 4 Triplet classification accuracy of different models.

(%)

Method	WN11	FB13	FB15K
TransE[15]	75.8	81.5	79.7
TransH[16]	78.8	83.8	87.7
TranSparse[33]	86.8	87.5	88.5
STransE[34]	86.4	89.1	83.2
TransE-RS[35]	85.3	83.0	81.9
TransE-FT[17]	86.4	82.1	90.5
TransE-DT[18]	86.6	85.3	83.3
TransAt[36]	88.2	89.1	—
OPTransE[37]	82.3	87.2	90.5
RPJE[38]	84.7	—	91.3
ERDERP[39]	—	—	91.2
TransE-ETF(max)[40]	85.7	85.5	—
TKGE-VT(this paper)	**87.4**	**90.5**	**93.7**

5 Conclusion and Future Work

In this paper, we present TKGE-VT, a temporal knowledge graph embedding model based on variable translation. TKGE-VT proposes a variable translation principle and can better capture the complex and diverse entities/relationships in KGs. In addition, we add temporal properties for entities and relationships respectively, and apply the proposed principle of variable translation to temporal knowledge graphs. We performed extensive experiments on four popular datasets for link prediction and triplet classification, and experiment results show that our method outperforms classic models (such as FT and DT).

In future work, we intend to make further improvements to TKGE-VT. In this paper, we only combine the variable translation principle with temporal information. Therefore, we will consider incorporating the proposed variable translation principle with more multi-source information such as entity description, type information, and relation path, etc. Furthermore, in addition to the common tasks of link prediction and triplet classification, we will consider applying the proposed model to tasks such as semantic parsing and relation extraction, etc.

Acknowledgment

The work was supported partly by National Natural Science Foundation of China (Nos. 62372119 and 62166003), the Project of Guangxi Science and Technology (Nos. GuiKeAB23026040 and GuiKeAD20159041), the Innovation Project of Guangxi Graduate Education (No. YCSW2023188), Key Lab of Education Blockchain and Intelligent Technology, Ministry of Education, Guangxi Normal University, Guilin, China, Intelligent Processing and the Research Fund of Guangxi Key Lab of Multi-source Information Mining & Security (Nos. 20-A-01-01 and MIMS21-M-01), Open Research Fund of Guangxi Key Lab of Human-machine Interaction and Intelligent Decision (No. GXHIID2206), the Guangxi Collaborative Innovation Center of Multi-Source Information Integration and the Guangxi "Bagui" Teams for Innovation and Research, China.

References

[1] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, Freebase: A collaboratively created graph database for structuring human knowledge, in *Proc. 2008 ACM SIGMOD Int. Conf. Management of data*, Vancouver, Canada, 2008, pp. 1247–1250.

[2] G. Miller, WordNet: A lexical database for English, *CACM*, vol. 38, no. 11, pp. 39–41, 1995.

[3] F. M. Suchanek, G. Kasneci, and G. Weikum, Yago: A core of semantic knowledge, in *Proc. 16th Int. Conf. World Wide Web*, Banff, Canada, 2007, pp. 697–706.

[4] T. Mitchell, W. Cohen, E. Hruschka, P. Talukdar, B. Yang, J. Betteridge, A. Carlson, B. Dalvi, M. Gardner, B. Kisiel, et al., Never-ending learning, *CACM*, vol. 61, no. 5, pp. 103–115, 2018.

[5] G. Q. Lu, J. C. Li, and J. Wei, Aspect sentiment analysis with heterogeneous graph neural networks, *Information Processing & Management*, vol. 59, no. 4, p. 10, 2953.

[6] G. Lu and J. Huang, Learning representation from concurrence-words graph for aspect sentiment classification, *Comput J*, vol. 64, no. 7, pp. 1069–1079, 2021.

[7] F. Huang, X. Li, C. Yuan, S. Zhang, J. Zhang, and S. Qiao, Attention-emotion-enhanced convolutional LSTM for sentiment analysis, *IEEE Trans. Neural Netw. Learning Syst.*, vol. 33, no. 9, pp. 4332–4345, 2022.

[8] J. Li, G. Lu, Z. Wu, and F. Ling, Multi-view representation model based on graph autoencoder, *Inf. Sci.*, vol. 632, pp. 439–453, 2023.

[9] J. Gan, R. Hu, Y. Mo, Z. Kang, L. Peng, Y. Zhu, and X. Zhu, Multigraph fusion for dynamic graph convolutional network, *IEEE Trans. Neural Netw. Learning Syst*., pp. 1–12, 2022.

[10] L. Peng, Y. Mo, J. Xu, J. Shen, X. Shi, X. Li, H. T. Shen, and X. Zhu, GRLC: graph representation learning with constraints, *IEEE Trans. Neural Netw. Learning Syst*., pp. 1–14, 2023.

[11] Y. Mo, Y. Chen, Y. Lei, L. Peng, X. Shi, C. Yuan, and X. Zhu, Multiplex graph representation learning via dual correlation reduction, *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 12, pp. 12814–12827, 2023.

[12] G. Lu, J. Gan, J. Yin, Z. Luo, B. Li, and X. Zhao, Multi-task learning using a hybrid representation for text classification, *Neural Comput. Appl.*, vol. 32, no. 11, pp. 6467–6480, 2020.

[13] Y. H. Zhu, J. B. Ma, C. A. Yuan, and X. F. Zhu, Interpretable learning based Dynamic Graph Convolutional Networks for Alzheimer's Disease analysis, Information Fusion, vol. 77, 2022, pp. 53–61.

[14] R. West, E. Gabrilovich, K. Murphy, S. Sun, R. Gupta, and D. Lin, Knowledge base completion via search-based question answering, in *Proc. 23rd Int. Conf. World wide web*, Seoul, Republic of Korea, 2014, pp. 515–526.

[15] A. Bordes, N. Usunier, A. Garcia-Durán, J. Weston, and O. Yakhnenko, Translating embeddings for modeling multi-relational data, *Adv. Neural Inf. Process. Syst.*, vol. 2, pp. 2787–2795, 2013.

[16] Z. Wang, J. Zhang, J. Feng, and Z. Chen, Knowledge graph embedding by translating on hyperplanes, in *Proc. Twenty-Eighth AAAI Conf. Artif. Intell.*, Québec City, Canada, 2014, pp. 1112–1119.

[17] J. Feng, M. Huang, M. Wang, M. Zhou, Y. Hao, and X. Zhu, Knowledge graph embedding by flexible translation, in *Proc. Fifteenth Int. Conf. Principles of Knowledge Representation and Reasoning*, Cape Town, South Africa, 2016, pp. 557–560.

[18] L. Chang, M. Zhu, T. Gu, C. Bin, J. Qian, and J. Zhang, Knowledge graph embedding by dynamic translation, *IEEE Access*, vol. 5, pp. 20898–20907, 2017.

[19] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, Learning

entity and relation embeddings for knowledge graph completion, in *Proc. 29th AAAI Conf. Artif. Intell.*, Austin, TX, USA, 2015, pp. 2181–2187.

[20] G. Ji, S. He, L. Xu, K. Liu, and J. Zhao, Knowledge graph embedding via dynamic mapping matrix, in *Proc. 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int. Joint Conf. Natural Language Processing* (*Volume 1*: *Long Papers*), Beijing, China, 2015, pp. 687–696.

[21] Y. Li, W. Fan, C. Liu, C. Lin, and J. Qian, TranSHER: Translating knowledge graph embedding with hyper-ellipsoidal restriction, in *Proc. 2022 Conf. Empirical Methods in Natural Language Processing*, Abu Dhabi, United Arab Emirates, 2022, pp. 8517−8528.

[22] Y. Ma, G. Altenbek, and X. Wu, Knowledge graph embedding via entity and relationship attributes, *Multimed. Tools Appl.*, vol. 82, no. 28, pp. 44071–44086, 2023.

[23] T. Jiang, T. Liu, T. Ge, L. Sha, S. Li, B. Chang, and Z. Sui, Encoding temporal information for time-aware link prediction, in *Proc. 2016 Conf. Empirical Methods in Natural Language Processing*, Austin, TX, USA, 2016, pp. 2350–2354.

[24] J. Messner, R. Abboud, and I. I. Ceylan, Temporal knowledge graph completion using box embeddings, *Proc. AAAI Conf. Artif. Intell.*, vol. 36, no. 7, pp. 7779–7787, 2022.

[25] R. Abboud, İ. İ. Ceylan, T. Lukasiewicz, and T. Salvatori, BoxE: A box embedding model for knowledge base completion, arXiv preprint arXiv: 2007.06267, 2020.

[26] P. Shao, D. Zhang, G. Yang, J. Tao, F. Che, and T. Liu, Tucker decomposition-based temporal knowledge graph completion, *Knowl. Based Syst.*, vol. 238, p. 107841, 2022.

[27] N. Park, F. Liu, P. Mehta, D. Cristofor, C. Faloutsos, and Y. Dong, EvoKG: Jointly modeling event time and network structure for reasoning over temporal knowledge graphs, in *Proc. Fifteenth ACM Int. Conf. Web Search and Data Mining*, Virtual Event, 2022, pp. 794–803.

[28] F. Zhang, Z. Zhang, X. Ao, F. Zhuang, Y. Xu, and Q. He, Along the time: Timeline-traced embedding for temporal knowledge graph completion, in *Proc. 31st ACM Int. Conf. Information & Knowledge Management*, Atlanta, GA, USA, 2022, pp. 2529–2538.

[29] R. Goel, S. M. Kazemi, M. Brubaker, and P. Poupart, Diachronic embedding for temporal knowledge graph completion, *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 4, pp. 3988–3995, 2020.

[30] X. Hou, R. Ma, L. Yan, and Z. Ma, T-GAE: A timespan-aware graph attention-based embedding model for temporal knowledge graph completion, *Inf. Sci.*, vol. 642, p. 119225, 2023.

[31] B. An, X. Han, L. Sun, and J. Wu, Triple classification based on synthesized features for knowledge base, https://www.semanticscholar.org/paper/Triple-Classification-Based-on-Synthesized-Features-Bo-Xianpei/190496c1f145bf8e66aeb7e1d0c9acfd48cc33b4, 2016.

[32] B. S. Yang, W. T. Yih, X. D. He, J. F. Gao, and L. Deng, Embedding entities and relations for learning and inference in knowledge bases, in *3rd Int. Conf. Learning Representations*, San Diego, CA, USA, 2015.

[33] G. Ji, K. Liu, S. He, and J. Zhao, Knowledge graph completion with adaptive sparse transfer matrix, in *Proc. 30th AAAI Conf. Artif. Intell.*, Phoenix, AZ, USA, 2016, pp. 985–991.

[34] D. Q. Nguyen, K. Sirts, L. Qu, and M. Johnson, STransE: A novel embedding model of entities and relationships in knowledge bases, arXiv preprint arXiv: 1606.08140, 2016.

[35] X. Zhou, Q. Zhu, P. Liu, and L. Guo, Learning knowledge embeddings by combining limit-based scoring loss, in *Proc. 2017 ACM on Conf. Information and Knowledge Management*, Singapore, 2017, pp. 1009–1018.

[36] W. Qian, C. Fu, Y. Zhu, D. Cai, and X. He, Translating embeddings for knowledge graph completion with relation attention mechanism, in *IJCAI'18*: *Proc. 27th Int. Joint Conf. on Artif. Intell.*, Stockholm, Sweden, pp. 4286–4292, 2018.

[37] Y. Zhu, H. Liu, Z. Wu, Y. Song, and T. Zhang, Representation learning with ordered relation paths for knowledge graph completion, in *Proc. 2019 Conf. Empirical Methods in Natural Language Processing and the 9th Int. Joint Conf. Natural Language Processing* (*EMNLP-IJCNLP*), Hong Kong, China, 2019, pp. 2662–2671.

[38] G. Niu, Y. Zhang, B. Li, P. Cui, S. Liu, J. Li, and X. Zhang, Rule-guided compositional representation learning on knowledge graphs, *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 3, pp. 2950–2958, 2020.

[39] L. Lin, J. Liu, F. Guo, C. Tong, L. Zu, and H. Guo, ERDERP: Entity and relation double embedding on relation hyperplanes and relation projection hyperplanes, *Mathematics*, vol. 10, no. 22, p. 4182, 2022.

[40] W. Chen, S. Zhao, and X. Zhang, Enhancing knowledge graph embedding with type-constraint features, *Appl. Intell.*, vol. 53, no. 1, pp. 984–995, 2023.

[41] S. Yang, J. Tian, H. Zhang, J. Yan, H. He, and Y. Jin, TransMS: knowledge graph embedding for complex relations by multidirectional semantics, pp. 1935−1942, 2019.

Guangquan Lu received the BEng degree in computer science from Guangxi Normal University, Guilin, China in 2011, and the PhD degree in philosophy from Sun Yat-sen University, China, in 2019. He is currently an associate professor with Key Lab of Education Blockchain and Intelligent Technology, Ministry of Education, Guangxi Normal University, Guilin, China. His research interests include deep learning, data mining, and graph representation learning for agricultural machine learning.

Shichao Zhang is a China National Distinguished Professor with Central South University, China. He received the PhD degree from Deakin University, Australia. His research interests include data mining and big data. He has published 90 international journal papers and over 70 international conference papers. He is a CI for 18 competitive national grants. He serves/served as an associate editor for four journals.

Cuifang Zou received the BEng degree from Guangxi Normal University, China in 2019. She is currently pursuing the MEng degree at School of Computer Science and Engineering / School of Software, Guangxi Normal University, China. Her research interests include node classification and link prediction.

Yadan Han received the BEng degree from Henan University of Urban Construction, China in 2021. She is currently pursuing the MEng degree at School of Computer Science and Engineering / School of Software, Guangxi Normal University, China. Her research interests include knowledge representation learning and knowledge graph completion.

Liang Zhang received the BEng degree from Central South University, China in 2021. He is currently pursuing the MEng degree at School of Computer Science and Engineering / School of Software, Guangxi Normal University, China. His research interests include graph representation learning and topology-imbalance learning in graph neural networks.

Guoqiu Wen received the PhD degree in 2022 and now is employed by Guangxi Normal University as a lecture. Her research interests include machine learning, feature learning, graph learning, and data mining.