

Temporal Relevance for Representing Learning over Temporal Knowledge Graphs

Bowen Song^{a,e}, Kossi Amouzouvi^{b,e,f}, Chengjin Xu^c, Maocai Wang^{a,*}, Jens Lehmann^{b,d,e} and Sahar Vahdati^{e,b}

^a *School of Computer Science, China University of Geosciences (Wuhan), Hubei, China*

E-mails: songbowen@cug.edu.cn, mcwang@cug.edu.cn

^b *Faculty of Computer Science, TU Dresden, Dresden, Germany*

E-mail: kossi.amouzouvi@tu-dresden.de

^c *International Digital Economy Academy, China*

E-mail: xuchengjin@idea.edu.cn

^d *Amazon, Germany*

E-mail: jlehmann@amazon.com

^e *Institute for Applied Informatics (InfAI), Dresden, Germany*

E-mail: vahdati@infai.org

^f *Department of Mathematics, KNUST, Kumasi, Ghana*

Abstract. Representation learning for link prediction is one of the leading approaches to deal with incompleteness problem of real world knowledge graphs. Such methods are often called knowledge graph embedding models which represent entities and relationships in knowledge graphs in continuous vector spaces. By doing this, semantic relationships and patterns can be captured in the form of compact vectors. In temporal knowledge graphs, the connection of temporal and relational information is crucial for representing facts accurately. Relations provide the semantic context for facts, while timestamps indicate their temporal validity. However, existing embedding models often overlook the intricate interplay between relational and temporal parts of the facts in temporal knowledge graphs. These models tend to focus on effectively representing individual components, consequently capturing only a fraction of the overall knowledge. Additionally, some relations in temporal facts are time-insensitive, while others are highly time-dependent. This complexity reduces the ability of temporal knowledge graph embedding models in accurately capturing these characteristics. To address these challenges, we propose a novel embedding model based on temporal relevance. This model operates within a complex space with real and imaginary parts to effectively embed temporal knowledge graphs. Specifically, the real part of the embeddings of our proposed model captures the semantic characteristics of facts by considering the importance of temporal information and relational information associated with each fact. Simultaneously, the imaginary part of the embeddings learns the connections between diverse elements, without predefined weights. Our approach is evaluated through extensive experiments on the link prediction task, where it majorly outperforms state-of-the-art models. The proposed model also demonstrates remarkable effectiveness in capturing the complexities of temporal knowledge graphs.

Keywords: Temporal knowledge graph, Representation learning, Knowledge graph embedding, Link prediction

1. Introduction

Knowledge graphs (KGs) as a formalism for knowledge representation and management, have emerged as underlying technology of many downstream AI tasks, including recommendation systems and question-answering

*Corresponding author. E-mail: mcwang@cug.edu.cn.

1 systems, and so on [1]. Example of such KGs are DBPedia[2], YAGO [3], and NELL [4]. Generally, knowledge 1
 2 graphs consist of a collection of triples representing real world facts, in the form of (s, r, o) , where s and o are subject 2
 3 and object entities respectively, r is the relation between them. Despite the large quantities of triples in KGs, they 3
 4 often suffer from incompleteness problem. One of the most popular methods to deal with this problem is by Knowl- 4
 5 edge Graph Embedding (KGE) models that are based on technique for predicting missing links between existing 5
 6 entities. KGEs operate by representing entities and relations from the symbolic KG into a low-dimensional vector 6
 7 space [5, 6], and then measure the plausibility of possible links. Performing link prediction is mainly to complete 7
 8 queries in the form of $(s, r, ?)$ or $(?, r, o)$ according to the score of the potential entity in the object position. 8
 9

10 Facts in the real world, often involve time, for which the representation in KGs need to be extended beyond 10
 11 triples (e.g., quadruples, quintuples), to include the temporal knowledge. These type of KGS are called temporal 11
 12 knowledge graphs (TKGs) such as ICEWS [7], GDELT [8] and Wikidata [9]. In TKGs designed with quadruples, 12
 13 the facts are represented with one timestamp in the form (s, r, o, τ) . In this representation, τ refers to the timestamp 13
 14 for the (s, r, o) fact. Similar to the static KGs, TKGs also have incompleteness problem which could be found by 14
 15 embedding models design to represent TKGs in vector space. These models are called Temporal Knowledge Graph 15
 16 Embedding (TKGE) and perform link prediction by posing queries $(s, r, ?, \tau)$ or $(?, r, o, \tau)$. These cases belong to 16
 17 the facts in real world such that potential subject or object entities are changing with time. 17

18 Among different types of TKGEs, Tensor-based models which employ tensor decomposition techniques, have 18
 19 demonstrated remarkable efficacy in representing temporal facts [10, 11]. As a remark, relations in TKGs serve 19
 20 as the connectors linking entities to form a fact, and time constitutes the temporal validity attribute of the fact. 20
 21 By decomposing the facts into tensors, quadruples are transformed into lower-dimensional embeddings, effectively 21
 22 mitigating complexity, and information from diverse attributes can be integrated to achieve comprehensive represen- 22
 23 tations. Different tensor decomposition methods, such as Canonical Polyadic (CP) decomposition [12] and Tucker 23
 24 decomposition [13], are leveraged in KGEs such as [10, 13, 14]. In these models, associations between entities, 24
 25 relations and timestamps are established through multiplication operations directly. However, generally in temporal 25
 26 facts, the connection between relations and semantics fundamentally diverges from the connection between time and 26
 27 semantics imposed by relations. Therefore, when learning over temporal facts, it is important to not only account 27
 28 for the semantic aspects of these facts but also recognize the importance of the temporal attribute for each individual 28
 29 fact. Additionally, some relations in temporal facts are time-insensitive, while others are highly time-dependent. 29
 30 However, such tensor-based models often prioritize considerations regarding how to accurately represent times- 30
 31 tamps or the entirety of quadruples. The methods to derive connections among these three components through 31
 32 straightforward multiplication fail to align with the authentic meaning of some temporal facts. Consequently, they 32
 33 tend to overlook the intricate interplay between time and semantics inherent in real-world facts. 33
 34

35 We introduce a novel model dubbed TRKGE, considering temporal relevance in temporal knowledge graph, 35
 36 which leverages the tensor decomposition method but facilitated to be aware of time importance. Temporal relevance 36
 37 is used to judge the importance of time in facts when learning their representation. For example, the semantics 37
 38 of facts formed by "visit" and "daughter of" have different temporal attributes, one is temporary and the other 38
 39 is permanent. Similar to other tensor decomposition models, our model is also built in the complex space. The 39
 40 real part focuses on capturing temporal relevance within the facts. It comprehensively incorporates embeddings of 40
 41 entities, relations, and timestamps. To ensure consistency between the transformation in the real part and complex 41
 42 multiplication, we employ rotation matrices in this part. These matrices effectively adjust entities based on their 42
 43 relationships, facilitating a specific understanding of the interplay between time and semantics. Furthermore, we 43
 44 introduce attention mechanisms in the real part to learn the temporal relevance within the facts. This attention 44
 45 mechanism enhances the model capacity to realize the significance of different elements in temporal facts. On the 45
 46 other hand, the imaginary part of our model utilizes half of the embeddings to learn the connections among diverse 46
 47 elements through direct multiplication. Then the imaginary part complements the real part in capturing complex 47
 48 relationships within temporal knowledge graphs. Experimental results underscore its performance in comparison 48
 49 to state-of-the-art baselines, thereby illustrating the ability of the proposed method in learning temporal relevance 49
 50 when representing temporal facts. 50
 51

2. Related Work

2.1. Static Knowledge Graph Embedding

In the last decade, various KGE models have been proposed for static knowledge graph completion. These models can generally be categorized into translation-based models, tensor decomposition models and neural models. Translation-based models treat a relation as a translation from the subject entity to the object entity. The scores in these models are typically calculated based on the distance or angle between the translated subject entity embeddings and object embeddings. TransE [15] treats relations as translations from subject entities to object entities, then calculates the distance between translated subject and object entities. RotatE [16] represents each entity and relation as an embedding in complex space, with a real part and an imaginary part. Relations are represented as angles, and subject entities are rotated by relations, the angles between them is the score of the model. Additionally, some researchers proposed models that embed entities and relations in non-Euclidean spaces. MURP [17] first use hyperbolic space to model hierarchical structures in KGs, which brought a big improvement. Later, AttH [18] bring both rotation and reflection into hyperbolic spaces to model various relational patterns. These models focus on capturing the translation relationship between entities and relations and perform well in capturing one-to-one and one-to-many relationships. Tensor decomposition models assume entities as vectors and relations as matrices, such as RESCAL [19], whose scores are calculated based on the product between these embeddings. But RESCAL cannot model complex structures, so ComplEx [12] was proposed, it represents entities and relations as complex embeddings consisting of both real and imaginary parts. The use of complex embeddings can capture both symmetric and asymmetric relationships in KGs. Compared with translation-based models, these models can capture more complex relationships. Besides, neural models employ neural architectures to train embeddings for entities and relationships for KG modeling, such as ConvKB [20], ConvE [21], and KBGAT [22]. While static knowledge graph embedding models have received significant attention over the years, they do not fully correspond to real-life situations. In real life, many facts are often time-dependent and static KGE models can not perform well in TKGs, therefore, there is a need for research on the temporal knowledge graph embedding model.

2.2. Temporal Knowledge Graph Embedding

Similar to the static KGs, temporal knowledge graphs are also modeled by the embeddings of entities, relations and timestamps. Many TKGE models are built following existing static KGE models. For example, TTransE [23], is an extension of TransE model, which translates entities by both relations and timestamps. HyTE [24] also borrowed the idea from TransH [25], but it projects entities and relations to different time-dependent planes rather than relation-dependent planes. Based on the static KGE model BoxE [26], BoxTE [27] extends the box embedding model to temporal knowledge graph. RotateQVS [28] models the temporal changes with rotation in quaternion vector space and uses a score function that is similar to TransE. Inspired by RotatE, TeRo [29] defines the temporal evolution of entity embedding as a rotation from the initial time to the current time in the complex space. TComplEx and TNTComplEx [14] are both developed based on the static KGE model ComplEx, which use additional tensors to represent timestamps, and then the score is obtained by complex multiplication of embeddings of entities, relations and timestamps. In addition, many new temporal knowledge graph embedding models have been proposed. ChronoR [10] treat relations and timestamps as rotation and scale. Rich information between temporal and multi-relational features in TKGs are captured by using high-dimensional rotations as transformation operators. TeLM [30] uses multivector representation and the geometric product to model entities, relations and timestamps. TLT-KGE [11] treats semantic and temporal information as different axes in complex or quaternion spaces, and a shared time window as well as a temporal-relational binding module are designed to establish the connection between different parts and timestamps. However, most of these works only focus on how to represent the flow of time or how to model timestamps, they ignore the real relationships between time and real facts. In our work, we propose the temporal relevance of entities and relations in facts. Temporal facts are represented by a method that is closer to the actual semantics, thereby improving the ability to learn features in temporal facts.

3. The TRKGE Model: Time Relevance in Temporal Knowledge Graph Embedding

Let us consider a temporal knowledge graph in which the facts are represented as quadruples (s, r, o, τ) , where s and o represent subject and object entities respectively, r represents the relation, and τ represents the timestamp. We define \mathcal{E} , \mathcal{R} and \mathcal{T} as the sets of entities, relations and timestamps, then $s, o \in \mathcal{E}$, $r \in \mathcal{R}$ and $\tau \in \mathcal{T}$. Temporal knowledge graph embedding models aim at completing such TKGs by learning embeddings of entities, relations, and timestamps. The score function of a TKGE model measures the likelihoods of quadruples, hence, new quadruples can be inferred, and their plausibility can be judged to complete the TKG. For temporal facts, relations have different time sensitivities. Therefore, when learning temporal information, the proportion of time information versus relational information in different facts is also needed to be considered. The proposed TRKGE model is capable of capturing these complexities of time relevance with reads to relations. The rest of this section focuses on the mathematical concepts required to understand, details theories and development of our model TRKGE.

3.1. Preliminaries and Background

As all geometric TKGE models, our model embeds entities in a continuous vector space and uses geometric transformation to preserve information between subject and object entities.

A vector space \mathbb{V} over a field \mathbb{K} is a set endowed with two operations, namely addition and scalar product, denoted by $+$ and \cdot respectively. These two operations must satisfy some axioms such as associativity, commutativity, existence identity and inverse elements, compatibility, and distributivity. In this work, we consider the 2 dimension complex vector space \mathbb{C}^2 over the field of real numbers \mathbb{R} . Vectors $\mathbf{s} \in \mathbb{C}^2$ are in the form $\mathbf{s} = (s_1, s_2)$ where $s_j = x_j + iy_j \in \mathbb{C}$, for $j = 1, 2$, and $i^2 = -1$. We define and denote by $\text{Re}(\mathbf{s}) = (x_1, x_2)$ the real part of \mathbf{s} and by $\text{Im}(\mathbf{s}) = (y_1, y_2)$ its imaginary part. Embeddings in high dimensional vector space are obtained by considering the Cartesian product of \mathbb{C}^2 . We use $\langle \mathbf{s}, \mathbf{s}' \rangle$ to denote the Hermitian inner product of two vectors \mathbf{s} and \mathbf{s}' . Its real part $\text{Re}(\langle \mathbf{s}, \mathbf{s}' \rangle)$, denoted by $\text{sim}(\mathbf{s}, \mathbf{s}')$, is a positive real vector.

Interaction between subject and object entities is done via a variety of geometric transformations. In the literature, many KGE models are built upon the rotation transformation. Similarly, our model uses rotation. However, there is a systematic difference between how the rotation transformation was used by other models. In fact a two-dimensional representation of a rotation of angle θ is defined by the two matrix

$$R(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

or the unit complex number $e^{i\theta}$. In the previous models, a vector $\mathbf{s} \in \mathbb{C}^2$ is rotated through the product $e^{i\theta_j} s_j$ or $R(\theta_j) \begin{pmatrix} x_j \\ y_j \end{pmatrix}$, $j = 1, 2$. In our model, the rotation is achieved through $R(\theta_1) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and $R(\theta_2) \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$. In other words, rotation transformations act on the real or imaginary parts ($\text{Re}(\mathbf{s})$, $\text{Im}(\mathbf{s})$) of vectors in \mathbb{C}^2 instead of acting on its projections (s_1, s_2) onto the complex plane \mathbb{C} . We extend this formalism into d dimensional vector space $\mathbb{V} = (\mathbb{C}^2)^d$ by considering the Cartesian product of \mathbb{C}^2 . The matrices $R(\theta_j)$ are combined to form block diagonal matrices $R(\Theta) = \text{diag}(R(\theta_1), \dots, R(\theta_d))$ which represent rotation transformations in high dimension. In this case, Θ is the set of rotation angles θ_j .

3.2. Entity, Relation, and Timestamp Embeddings: Temporal Relevance

We embed subject and object entities, relations, and timestamps in the TKG by d -dimensional vectors $\mathbf{s}, \mathbf{o}, \mathbf{r}, \tau \in \mathbb{V}$, respectively. More explicitly, any embedding has two embeddings $\text{Re}(\mathbf{z})$ and $\text{Im}(\mathbf{z})$ in $(\mathbb{R}^2)^d$, and so can be expressed as $\mathbf{z} = \text{Re}(\mathbf{z}) \oplus i\text{Im}(\mathbf{z})$. By \mathbf{z} , we refer to the embeddings $\mathbf{s}, \mathbf{o}, \mathbf{r}, \tau$, and by \oplus , we refer to the element wise addition. Similarly, we will use \odot to refer to element wise complex multiplication. We transform the real part of relations, r , and Timestamps, τ , connecting two entities, into high dimensional block diagonal rotation matrices $R(\Theta_r)$ and $R(\Theta_\tau)$ where $\Theta_r, \Theta_\tau \in \mathbb{R}^d \subset \mathbb{V}$. Thus, relations and timestamps are firstly embedded in \mathbb{V} , and their real parts are projected into block diagonal real matrix vector space.

To allow the model to selectively learn relational and temporal information based on temporal relevance, we transform subject entities according to relation and time embeddings, respectively. The real part of subject entity embeddings is used to learn the temporal relevance, and the imaginary part is used to learn the original relationships between different elements in facts. We transform real part of subject entity embeddings simultaneously by time and relation embeddings to yield $\mathbf{s}_\tau = \mathbf{R}(\Theta_\tau)\text{Re}(s)$, and $\mathbf{s}_r = \mathbf{R}(\Theta_r)\text{Re}(s)$. We thereafter use the attention mechanism to quantify the proportion of temporal information and relational information in facts.

The attention mechanism is proven to have a very significant role in deep learning models [18, 31]. Since the semantics of facts is intrinsically related to relations, the temporal relevance of the facts is also determined by the relations. The attention mechanism allows TRKGE to use relation-specific attention vector, defined as α , to compute two (positive real) attention coefficients, α_r and α_τ . The former is dedicated to capture relational information and the latter is focused on learning temporal information. Higher is α_τ associated to a relation, more relevant is the temporal information that relation conveys. The attention vector and coefficients are related by the equation below:

$$(\alpha_r, \alpha_\tau) = \text{softmax} \left(\frac{(\alpha^T \mathbf{s}_r, \alpha^T \mathbf{s}_\tau)}{\sqrt{d}} \right). \quad (1)$$

We recall that d is the dimension of the embedding vector space \mathbb{V} . Then, the transformed embedding, aware of temporal relevance, can be obtained by

$$\mathbf{s}_{TR} = \alpha_r \mathbf{s}_r \oplus \alpha_\tau \mathbf{s}_\tau. \quad (2)$$

3.3. TRKGE Model

The embeddings with temporal relevance, defined in the previous subsection, constitute the real part of the transformed subject entity embeddings, \mathbf{s}_f . Due to the biased learning of the attention mechanism, some relational or temporal information in the original facts may be lost. Thus, we select the imaginary part of the original subject embedding as the imaginary part of \mathbf{s}_f . The transformed subject entity embeddings are then defined as follows:

$$\mathbf{s}_f = \mathbf{s}_{TR} \oplus i\text{Im}(s). \quad (3)$$

The score function of the TRKGE model is defined as the real part of the Hermitian inner product of embeddings of transformed subject entities, relations, timestamps, and object entities. This is obtained as follows

$$f(s, r, \tau, o) = \text{sim}((\mathbf{r} \odot \tau) \odot \mathbf{s}_f, \mathbf{o}). \quad (4)$$

In Eq 4, the real part of embeddings focus on the temporal relevance, and the imaginary part is used to learn the relationships between different elements in quadruples. As semantics of facts is of great importance in representation learning that is mainly formed by relations, only learning the associations between four elements may lead to the loss of semantic information. So an additional relation embedding is designed to enhance the learning of the semantics of facts. Therefore the final score function is

$$f(s, r, \tau, o) = \text{sim}((\mathbf{r} \odot \tau \oplus \mathbf{r}_{sem}) \odot \mathbf{s}_f, \mathbf{o}). \quad (5)$$

where $\mathbf{r}_{sem} \in \mathbb{V}$ is the additional relation embeddings.

3.4. Loss Function

In our work, cross entropy loss is used to train the model with uniform negative sampling, where negative examples for (s, r, o, τ) are sampled from all possible quadruples generated by replacing subject or object entities. The

loss function to be minimized is

$$loss(G; \Phi) = \sum_{(s,r,\tau,o,\tau) \in G} \left(-\log \left(\frac{f(s,r,\tau,o)}{\sum_{o' \in \mathcal{E}} f(s,r,\tau,o')} \right) - \log \left(\frac{f(s,r,\tau,o)}{\sum_{s' \in \mathcal{E}} f(s',r,\tau,o)} \right) \right) \quad (6)$$

where Φ is the set of all the model parameters. o' is the object entity of the negative quadruple (s, r, o', τ) , and s' is the subject entity of the negative quadruple (s', r, o, τ) .

Besides, a regularization term is added to the loss function to limit the complexity of the model, thus reducing the overfitting of the model. It can also improve the generalization ability of the model, so that the model can be better generalized to unknown data. Regularization is used separately for all embeddings. Because the relational part in the model is divided into two parts, and scores of two relation embeddings are calculated separately, we propose the following regularization.

$$\Omega^4(s, r, o) = \frac{1}{4} \left(\|s\|_4^4 + \|r\|_4^4 + \|r_{sem}\|_4^4 + \|o\|_4^4 \right) \quad (7)$$

During training, timestamp embeddings should behave smoothly over time to facilitate a better representation, and the embeddings of adjacent timestamps should be close together. For this, a temporal smoothness objective is used in the loss function.

$$\Lambda_\tau = \frac{1}{|\mathcal{T}|-1} \sum_{i=1}^{|\mathcal{T}|-1} \|\tau_{i+1} - \tau_i\|_4^4 \quad (8)$$

Therefore, the final loss function is:

$$loss(G; \Phi) = \sum_{(s,r,o,\tau) \in G} \left(-\log \left(\frac{f(s,r,o)}{\sum_{o' \in \mathcal{E}} f(s,r,o')} \right) - \log \left(\frac{f(s,r,o)}{\sum_{s' \in \mathcal{E}} f(s',r,o)} \right) \right) + \lambda_1 \Omega^4 + \lambda_2 \Lambda_\tau \quad (9)$$

where λ_1 and λ_2 are hyperparameters.

4. Experimental Setup

4.1. Datasets and Baseline models, and Evaluation Methodology

We compare our model with other baselines through temporal knowledge graph completion on three popular benchmarks namely ICEWS14, ICEWS05-15 and GDEL. The first two datasets are subsets of Integrated Crisis Early Warning System(ICEWS) dataset that contains information about international events. ICEWS14 collects the facts that occurred between January 1, 2014 and December 31, 2014, and ICEWS05-15 contain events from January 1, 2005 to December 31, 2015. GDEL is a subset of the The Global Database of Events, Language, and Tone (GDEL) that captures events and news coverage from around the world. This dataset includes conflict, cooperation, diplomatic, economic, humanitarian events and so on. GDEL is a very dense TKG, it has 3.4 million quadruples but only 500 entities and 20 relations. Table 1 is the statics of these three datasets.

In the experiments, baselines are chosen from both static KGE models and TKGE models. From the static KGE models, we use TransE [15], DisMult [32], QuatE [33]. We also compare our model with some state-of-the-art TKGE models, such as TTransE [23], HyTE [24], TA-DistMult [34], ATiSE [35], TeRo [29], RotateQVS [28], TCompleEx [14], TNTCompleEx [14], BoxTE [27], TLT-KGE(Complex) [11] and TLT-KGE(Quaternion) [11].

In this work, we perform the evaluations with a focus on link prediction task. This means to replace s and o with all entities in \mathcal{E} for each quadruple in the test set. Then the model scores all generated quadruples. In addition,

Table 1
Statistics for ICEWS14, Yago15k and GDEL T.

| | $ \mathcal{E} $ | $ \mathcal{R} $ | $ \mathcal{T} $ | N_{train} | N_{valid} | N_{test} | N_{total} |
|------------|-----------------|-----------------|-----------------|-------------|-------------|------------|-------------|
| ICEWS14 | 7128 | 230 | 365 | 72826 | 8963 | 8941 | 90730 |
| ICEWS05-15 | 10488 | 251 | 4017 | 386962 | 46092 | 46275 | 479329 |
| GDEL T | 500 | 20 | 366 | 2735685 | 341961 | 341961 | 3419607 |

the time-wise filter is used in experiments, which removes all the candidate entities that yield correct quadruples. Following metrics are used for comparison: Mean Reciprocal Rank (MRR), and H@n where $n \in \{1, 3, 10\}$. MRR is measured by $\sum_{j=1}^{n_t} \frac{1}{r_j}$, where r_j is the rank of the j -th test quadruple and n_t is the number of triples in the test set. H@n is the number of test quadruples ranked less than n.

4.2. Implementation Details

Main Experiment In this experiment, temporal knowledge graph completion is done for three datasets. On ICEWS14 and ICEWS05-15, we tune λ_1 and λ_2 from [0.005, 0.05]. And on GDEL T, they are tuned from [0.0000005, 0.00005]. Our model is training with a learning rate 0.1, and it is optimized by Adagrad. The dimensions used on three datasets are all set as 1500 and 2000. We compare the best performance of our model with others' best results reported in their papers. The hyperparameters we used in the experiments are shown in Table 2. To be fair, we also analyse and compare performance according to different number of parameters.

Table 2
Hyperparameters of TRKGE for ICEWS14, Yago15k and GDEL T.

| | dimension | epoch | batch size | λ_1 | λ_2 |
|------------|------------|-------|------------|-------------|-------------|
| ICEWS14 | 2000(1500) | 200 | 500 | 0.01 | 0.5 |
| ICEWS05-15 | 2000(1500) | 200 | 1000 | 0.01 | 2.5 |
| GDEL T | 2000(1500) | 400 | 2000 | 1e-6 | 1e-5 |

Ablation Study In this part, to demonstrate the representation capabilities of temporal and relational information in our model on different datasets, we establish two models, denoted as Rel-TR and Time-TR. In both models, we removed the attention mechanism. Rel-TR only use relation embeddings for transformations, while Time-TR solely utilize time embeddings for transformations. Both models are trained and evaluated with the same hyperparameters as the main experiment. In addition, we also evaluate the design of the imaginary part for learning the relationships between different elements and additional relation embeddings for enhancing semantic learning, namely Im-TR and Sem-TR.

Performance under Different Number of Parameters In this section, we compare our model with the best-performing baselines, TLT-KGE (Complex and Quaternion) and TNTComplEx. We trained these models in dimensions of 500, 1000, 1500, and 2000. TLT-KGE and TNTComplEx are trained by the optimal hyperparameters provided in their work. The variation of the models' performance with the number of parameters is shown in the form of a line graph.

Time-Wise Entity Filtering. In our experiments, we utilize time-aware filters [36], which are also employed by the baseline models. This filter differs from the static models as it incorporates timestamps during the process of filtering out entities, which is meaningful for temporal knowledge graph representation.

5. Experiment

5.1. Main Experiment

We performed extensive evaluations ensuring to reflect the strength of the proposed model. First, we report the performance of the TRKGE model and other baselines on the three aforementioned benchmark datasets. The exper-

Experimental results for these performance evaluations are shown in Table 3. The main result which can be seen in the table is that our model outperforms all other baseline models on 2000 dimensions in all the metrics. While on the ICEWS14 and ICEWS05-15 datasets, the TLT-KGE model performs better than other baseline models, our model still improves the results with a high margin. On the MRR metric, our model outperforms TLT-KGE model by a delta of 1.6%, on the ICEWS14 dataset and with a margin of 0.2% on the ICEWS05-15 dataset. On the GDELDT dataset, TNTComplEx model is the best baseline model. But our model TRKGE outperforms TNTComplEx by 4.1% on GDELDT. In 1500 dimensions that is the setting used by most of the baseline models, our model is performing better than any other model on the two ICEWS14 and GDELDT datasets. Its results are comparative with the best of state-of-the-art results on ICEWS05-15. As a sign of robustness, our proposed model shows a constant performance increase on all these datasets which are challenging for the upfront models.

Table 3

Experimental results on ICEWS14, ICEWS05-15 and GDELDT. The results labeled with * are from the paper [28]. The results of TComplEx and TNTComplEx on GDELDT are produced from their original code. Other results are reported in their own paper.

| Model | ICEWS14 | | | | ICEWS05-15 | | | | GDELDT | | | |
|---------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | MRR | h@1 | h@3 | h@10 | MRR | h@1 | h@3 | h@10 | MRR | h@1 | h@3 | h@10 |
| TransE* | 28.0 | 9.4 | - | 63.7 | 29.4 | 9.0 | - | 66.3 | 11.3 | 0.0 | 15.8 | 31.2 |
| DistMult* | 43.9 | 32.3 | - | 67.2 | 45.6 | 33.7 | - | 69.1 | 19.6 | 11.7 | 20.8 | 34.8 |
| QuatE* | 47.1 | 35.3 | 53.0 | 71.2 | 48.2 | 37.0 | 52.9 | 72.7 | - | - | - | - |
| TTransE* | 25.5 | 7.4 | - | 60.1 | 27.1 | 8.4 | - | 61.6 | 11.5 | 0.0 | 16.0 | 31.8 |
| HyTE* | 29.7 | 10.8 | 41.6 | 65.5 | 31.6 | 11.6 | 44.5 | 68.1 | 11.8 | 0.0 | 16.5 | 32.6 |
| TA-DistMult* | 47.7 | 36.3 | - | 68.6 | 47.4 | 34.6 | - | 72.8 | 20.6 | 12.4 | 21.9 | 36.5 |
| ATiSE* | 55.0 | 43.6 | 62.9 | 75.0 | 51.9 | 37.8 | 60.6 | 79.4 | - | - | - | - |
| TeRo* | 56.2 | 46.8 | 62.1 | 73.2 | 58.6 | 46.9 | 66.8 | 79.5 | 24.5 | 15.4 | 26.4 | 42.0 |
| RotateQVS* | 59.1 | 50.7 | 64.2 | 75.4 | 63.3 | 52.9 | 70.9 | 81.3 | 27.0 | 17.5 | 29.3 | 45.8 |
| TComplEx | 61.9 | 54.2 | 66.1 | 76.7 | 66.5 | 58.3 | 71.6 | 81.1 | 38.5 | 30.5 | 41.2 | 53.8 |
| TNTComplEx | 60.9 | 52.1 | 66.0 | 77.4 | 66.9 | 58.4 | 72.2 | 82.2 | 38.9 | 30.7 | 41.6 | 54.6 |
| BoxTE(k=5) | 61.3 | 52.8 | 66.4 | 76.3 | 66.7 | 58.2 | 71.9 | 82.0 | 35.2 | 26.9 | 37.7 | 51.1 |
| TLT-KGE(Complex) | 63.0 | 54.9 | 67.8 | 77.7 | 68.6 | 60.7 | 73.5 | 83.1 | 35.6 | 26.7 | 38.5 | 53.2 |
| TLT-KGE(Quaternion) | 63.4 | 55.1 | 68.4 | 78.6 | 69.0 | 60.9 | 74.1 | 83.1 | 35.8 | 26.5 | 38.8 | 54.3 |
| TRKGE(1500) | 64.0 | 56.1 | 68.8 | 78.9 | 68.7 | 60.6 | 73.9 | 83.4 | 37.1 | 29.1 | 39.7 | 52.5 |
| TRKGE(2000) | 64.4 | 56.6 | 69.0 | 79.2 | 69.2 | 61.1 | 74.3 | 83.8 | 40.5 | 32.6 | 43.3 | 55.2 |

Table 3 also shows that most of the baseline models have different representation capabilities on the versions of ICEWS datasets and GDELDT. For example, TLT-KGE performs well on both ICEWS14 and ICEWS05-15, but its performance is mediocre on the GDELDT dataset. In contrast, the T(NT)ComplEx model performs well on GDELDT, but badly on the other two datasets namely ICEWS14 and ICEWS05-15. This observation can be explained based on the significant structural and statistical differences between those datasets and lack of models in handling those differences. We observed three aspects on these differences including:

- Density of Relations - This refers to the distribution of facts among relations. A dataset is considered denser if a higher number of relations have more facts associated with them.
- Temporal Relevance - This is determined by the number of timestamps in the dataset. A dataset with more timestamps indicates richer temporal information.
- Semantic Richness of Relations - This is determined by the number of facts associated with each relation. A dataset where most relations have a high number of facts indicates that the relations have rich semantics.

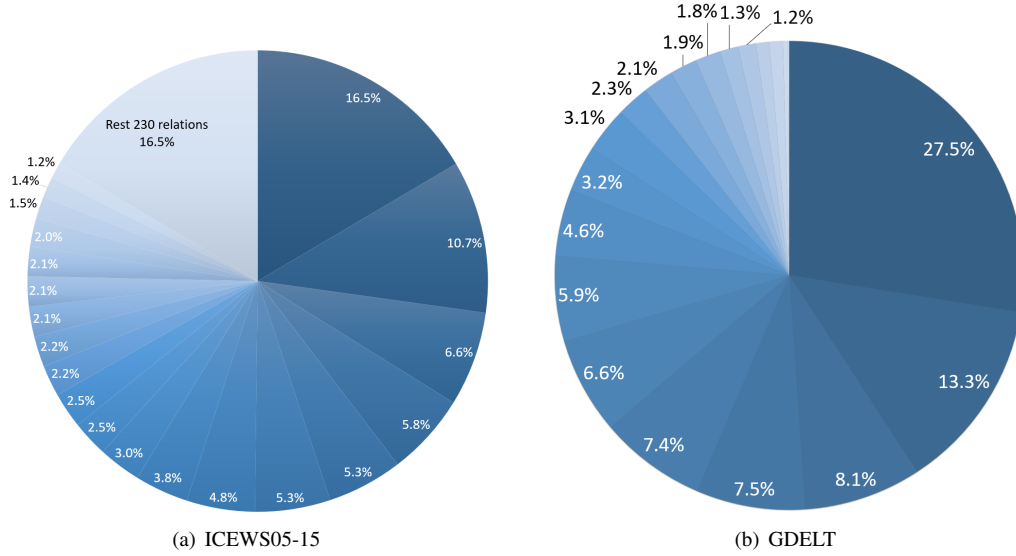


Fig. 1. Percentage of existing quadruples per each relations in two datasets.

Table 4
Statistics of the number of quadruples associated with relations on ICEWS05-15 and GDELT.

| Dataset | ICEWS05-15 | | | | GDELT | | |
|-------------|----------------|--------------|------------|-------|------------------|----------------|---------|
| | 10000<n<100000 | 1000<n<10000 | 100<n<1000 | n<100 | 1000000<n<100000 | 10000<n<100000 | n<10000 |
| #relations | 12 | 27 | 84 | 128 | 10 | 9 | 1 |
| #quadruples | 318168 | 107146 | 32200 | 3815 | 2982818 | 434307 | 2482 |
| Ratio | 69.20% | 23.30% | 7.00% | 0.83% | 87.23% | 12.70% | 0.07% |

ICEWS14 and ICEWS05-15 have higher number of entities, relation types, and timestamps than GDELT which are shown in Table 1. While the GDELT dataset has only 20 relations, it is way denser than the other two datasets and have much larger data volume in terms of facts. We show the number of quadruples for different relations in Fig 1, and we also give the information about the number of quadruples formed by each relation in Table 4. As the statistics of the two datasets ICEWS14 and ICEWS05-15 are similar, we only show the case of ICEWS05-15. It is noting that in both ICEWS14 and ICEWS05-15, the top 10 relations in terms of quantity almost construct the 60% of the datasets. In ICEWS05-15, there are 212 relations, among which each relation is present in less than 1000 facts. Only 39 relations out of 212 is present in more than 1000 facts, which leads to less density, resulting in a weak expression ability of most relations. At the same time, this dataset has more than 4000 timestamps that makes the temporal information more relevant. Different from the ICEWS05-15 dataset, in GDELT, there are 10 relations that each of them built more than 100 thousand facts, and 9 relations that each of them built more than 10 thousand facts, so all relations are relatively fully used in different facts. Such a small number of relations and timestamps contribute to a large amount of data, indicating that both relations and time have rich semantics in this dataset.

The baseline models, TNTComplex and BoxTE, learn the semantics of relations and timestamps equally, allowing them to have good representation capabilities on GDELT. In contrast, TLT-KGE specifically models temporal information by using many extra embeddings related to time, leading to its strong performance on ICEWS05-15, but its normal performance on GDELT shows their design is limited. Based on the best results of our model on three datasets, it can be concluded that our model, by calculating temporal relevance and placing emphasis on learning temporal and relational information, demonstrates good adaptability to different types of datasets. Additionally,

with the increase in dimensions, the model’s representational power also improves. The performance of some well-performing baseline models in high dimensions will be discussed in the analysis of the model’s parameter quantity.

5.2. Ablation Study

We conducted an ablation experiment on the temporal relevance component. In order to do so, the attention mechanism was removed from the model design with a focus on learning temporal information and relational information. Rel-TR focuses on relational information, while Time-TR focuses on temporal information. Table 5 shows that Time-TR outperforms Rel-TR by 6% on ICEWS14 and 3% on ICEWS05-15, while these two models perform similarly on GDELDT.

Table 5

Impact of learning relational and temporal information is evaluated by Rel-TR and Time-TR, where Rel-TR only learns relational information and Time-TR only learns temporal information. Im-TR only learns relational and temporal information with temporal relevance, ignoring the original information of facts. Sem-TR removes the extra relation embedding for learning the semantics of facts

| Model | ICEWS14 | | | | ICEWS05-15 | | | | GDELDT | | | |
|----------|---------|------|------|------|------------|------|------|------|--------|------|------|------|
| | MRR | h@1 | h@3 | h@10 | MRR | h@1 | h@3 | h@10 | MRR | h@1 | h@3 | h@10 |
| Rel-TR | 60.4 | 50.7 | 66.7 | 77.7 | 67.7 | 59.0 | 73.4 | 83.3 | 38.2 | 30.0 | 40.9 | 53.9 |
| Time-TR | 64.1 | 56.3 | 68.6 | 78.6 | 69.0 | 61.2 | 73.9 | 83.3 | 38.3 | 30.5 | 40.9 | 53.2 |
| Im-TR | 63.5 | 55.5 | 68.1 | 78.0 | 68.4 | 60.4 | 73.4 | 83.2 | 40.3 | 32.5 | 43.0 | 55.2 |
| Sem-TR | 59.3 | 51.0 | 64.2 | 74.3 | 63.0 | 54.3 | 68.2 | 79.0 | 40.2 | 32.3 | 43.0 | 55.1 |
| Original | 64.4 | 56.6 | 69.0 | 79.2 | 69.2 | 61.1 | 74.3 | 83.8 | 40.5 | 32.6 | 43.3 | 55.2 |

As mentioned in the main experiment, in ICEWS14 and ICEWS05-15, time plays a significant role, and in GDELDT, both temporal and relational information are important. The results of Time-TR and Rel-TR align with these characteristics.

Furthermore, in order to test the design of the imaginary part in our original model which is used to save the original information of facts, Im-TR is proposed, where both real and imaginary parts learn relational and temporal information with temporal relevance.

On ICEWS14 and ICEWS05-15, Im-TR performs better than Rel-TR but worse than Time-TR., which means that Selectively learning information leads to the improved performance, but it still loses some information when compared with Time-TR and original model. On GDELDT it performs better than both Rel-TR and Time-TR and similarly to original model. As facts in GDELDT have rich temporal and relational information, the attention mechanism has a weak bias towards these two types of information, which is also shown in Fig 2(b). Therefore, Im-TR is similar with original one on GDELDT.

Sem-TR performs worse on both ICEWS14 and ICEWS05-15. From Fig 2(a), we can see on ICEWS datasets, the attention mechanism mainly focus on the temporal information, which leads to a massive lack of relational information, so extra relation embedding for enhancing the learning of the semantics of facts is important in our original design. Similar with Im-TR, attention mechanism dose not play an important role on GDELDT, so the lack of information is less, and its performance is also close to TRKGE.

Overall, the original model TRKGE has the best performance, which demonstrates the meaningfulness of bi-ased learning towards temporal and relational information. It not only improves the model’s performance but also increases its adaptability to different datasets.

5.3. Attention Value of Time

In order to visualize the emphasis on time, we show attention values on two datasets of ICEWS05-15 and GDELDT in Fig 2. In the left hand side, the Fig 2 (a), we can see that the attention values for the frequent three relations on

ICEWS05-15 are almost 1. This means the learning is entirely biased towards time, while the attention values for the less frequent relations are evenly distributed around 0.5. However, the most frequent 10 relations constitute 60% of the facts in ICEWS05-15. This indicates that the impact of the less frequent relations is not very high. Therefore, Time-TR model can effectively capture the majority of the facts but cannot represent the remaining facts well, resulting in slightly weaker performance compared to our model, TRKGE. In the case of the GDELDT dataset, where both relations and timestamps exhibit rich semantic information, the attention values are centered around 0.5 in the visualization. This aligns with the characteristics of this dataset in terms of time relevance relations which is also consistent with the performance of Rel-TR, Time-TR and Im-TR.

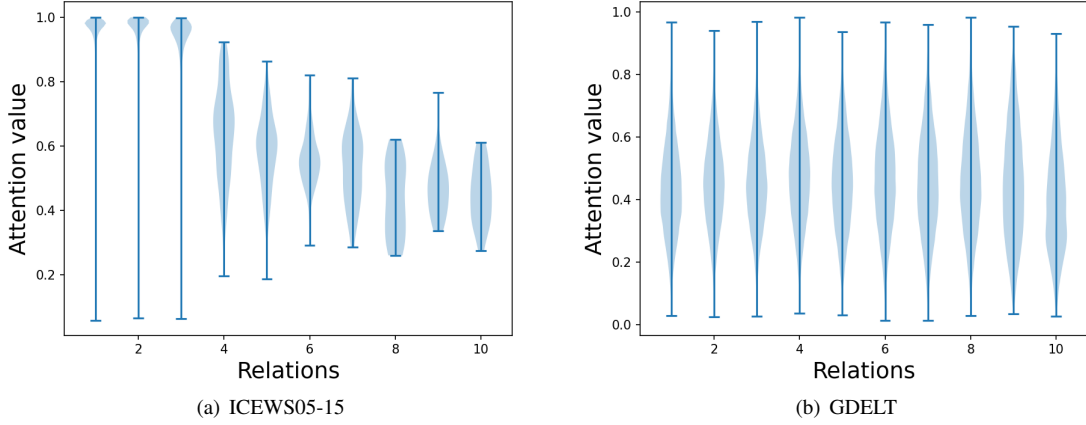


Fig. 2. Attention values on ICEWS05-15 and GDELDT. In the left (a), the first 3 values are from the most frequent 3 relations, the last 4 values are from the least frequent relations, and the middle 3 values are randomly selected. In the right (b), because of the small number of relations and memory limitations, the values are selected from the middle 10 relations.

5.4. Analysis on Number of Parameters

In this section, we performed an experiment to compare TNTCompEx, TLT-KGE and our model in terms of the number of parameters to reflect their efficiency. Table 6 provides a reference for the number of parameters of each model. It is evident that the performance gain of our model is not with higher parameter count than the other leading models, emphasizing its efficiency.

Table 6
Model parameter count for TRKGE, TNTCompEx and TLT-KGE. For TLT-KGE, w is the number of shared time windows.

| Model | Number of Parameters |
|-----------|--|
| TNTCompEx | $2d(\mathcal{E} + 4 \mathcal{R} + \mathcal{T})$ |
| TLT-KGE | $2d\left(\mathcal{E} + 2 \mathcal{R} + 3 \mathcal{T} + \frac{ \mathcal{T} }{w}\right)$ |
| TRKGE | $2d(\mathcal{E} + 5 \mathcal{R} + \mathcal{T})$ |

These three models were evaluated on dimensions of 500, 1000, 1500, and 2000, and line graphs were plotted and shown in Fig 3, to represent the model's performance with the number of parameters in four dimensions. As the Mean Reciprocal Rank (MRR) can indicate the overall effectiveness of a model, we correlated MRR with the number of parameters in the the graphs to illustrate the performance of different models under varying parameter counts. From Fig 3, it can be observed that TNTCompEx performed poorly on ICEWS14 and ICEWS05-15 but excelled on GDELDT due to its design for simultaneously learning relational and temporal information. TLT-KGE performed well

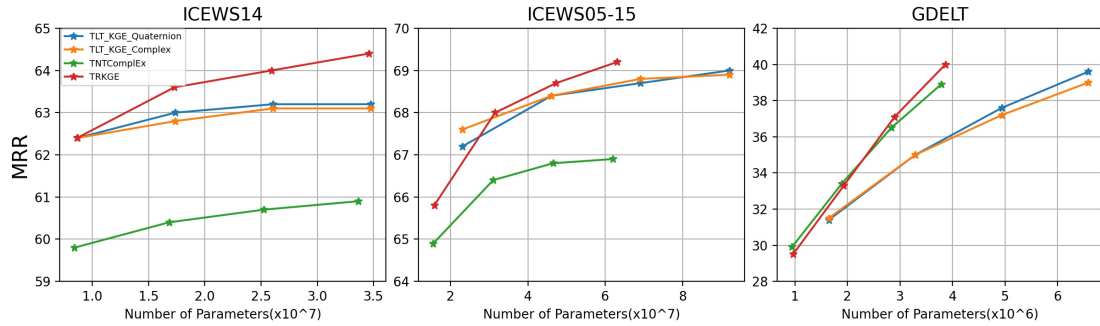


Fig. 3. Performance under different numbers of parameters. Circles represent 500 dimensions, triangles represent 1000 dimensions, stars represent 1500 dimensions, and squares represent 2000 dimensions

on ICEWS14 and ICEWS05-15 but poorly on GDELT, since it is designed to capture more temporal information, which results in more parameters, especially in the case of ICEWS05-15, where its number of parameters is much more than TNTComplEx and our model. From the graphs, it is obvious that our model achieves better results on ICEWS14 with similar parameter counts. On ICEWS05-15, our model can achieve similar optimal results with only half of parameters of TLT-KGE. On GDELT, both our model and TNTComplEx effectively learn rich information from relations and timestamps and achieve better results with much fewer parameters than TLT-KGE. The line graphs on all three datasets indicate that our model can efficiently and effectively represent temporal facts well in different types of TKGs.

6. Conclusion

By decomposing embeddings into real and imaginary parts, we have ensured that both temporal and relational facets are well-represented. The real part plays a crucial role in preserving temporal relevance of relational knowledge, guaranteeing that time-sensitive relations are not marginalized. In contrast, the imaginary component captures the nature of the original information, preventing potential information loss. This distinction and simultaneous handling of these components empower our model to preserve more complex characteristic in TKGs. Our experimental results serve as a testament to the model's proficiency. By excelling in link prediction tasks and outperforming existing state-of-the-art models, it underscores its capability to cater to the dynamics of TKGs. The novelty presented in this paper is not just in the creation of an advanced embedding model but also the demonstration that understanding and effectively representing the interplay between time and relationships in knowledge graphs can bring significant improvements. This serves as a new direction for future research in refining temporal relevance measures or exploring other complex space representations. In conclusion, as TKGs continue to grow in relevance, models like ours will be instrumental in ensuring that the temporal and relational information they contain is comprehensively captured and utilized. We believe that our proposed model sets a new benchmark in TKG embedding

References

- [1] L. Bellomarini, E. Sallinger and S. Vahdati, Knowledge graphs: the layered perspective, in: *Knowledge Graphs and Big Data Processing*, Springer, Cham, 2020, pp. 20–34.
- [2] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P.N. Mendes, S. Hellmann, M. Morsey, P. Van Kleef, S. Auer et al., Dbpedia—a large-scale, multilingual knowledge base extracted from wikipedia, *Semantic web* 6(2) (2015), 167–195.
- [3] F.M. Suchanek, G. Kasneci and G. Weikum, Yago: a core of semantic knowledge, in: *Proceedings of the 16th international conference on World Wide Web*, 2007, pp. 697–706.
- [4] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E.R. Hruschka and T.M. Mitchell, Toward an architecture for never-ending language learning, in: *Twenty-Fourth AAAI conference on artificial intelligence*, 2010.
- [5] L. Bellomarini, E. Sallinger and S. Vahdati, Reasoning in knowledge graphs: An embeddings spotlight, in: *Knowledge Graphs and Big Data Processing*, Springer, Cham, 2020, pp. 87–101.

- [6] M. Ali, M. Berrendorf, C.T. Hoyt, L. Vermue, M. Galkin, S. Sharifzadeh, A. Fischer, V. Tresp and J. Lehmann, Bringing light into the dark: A large-scale evaluation of knowledge graph embedding models under a unified framework, *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021).
- [7] E. Boschee, J. Lautenschlager, S. O'Brien, S. Shellman, J. Starz and M. Ward, ICEWS Coded Event Data, Harvard Dataverse, 2015. doi:10.7910/DVN/28075.
- [8] K. Leetaru and P.A. Schrodt, Gdelt: Global data on events, location, and tone, 1979–2012, in: *ISA annual convention*, Vol. 2, Citeseer, 2013, pp. 1–49.
- [9] D. Vrandečić and M. Krötzsch, Wikidata: a free collaborative knowledgebase, *Communications of the ACM* **57**(10) (2014), 78–85.
- [10] A. Sadeghian, M. Armandpour, A. Colas and D.Z. Wang, Chronor: Rotation based temporal knowledge graph embedding, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35, 2021, pp. 6471–6479.
- [11] 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: *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 2022, pp. 2529–2538.
- [12] T. Trouillon, J. Welbl, S. Riedel, É. Gaussier and G. Bouchard, Complex embeddings for simple link prediction, in: *International conference on machine learning*, PMLR, 2016, pp. 2071–2080.
- [13] I. Balažević, C. Allen and T.M. Hospedales, Tucker: Tensor factorization for knowledge graph completion, *arXiv preprint arXiv:1901.09590* (2019).
- [14] T. Lacroix, G. Obozinski and N. Usunier, Tensor Decompositions for Temporal Knowledge Base Completion, in: *International Conference on Learning Representations*, 2020. <https://openreview.net/forum?id=rke2P1BFwS>.
- [15] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston and O. Yakhnenko, Translating embeddings for modeling multi-relational data, *Advances in neural information processing systems* **26** (2013).
- [16] Z. Sun, Z.-H. Deng, J.-Y. Nie and J. Tang, RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space, in: *International Conference on Learning Representations*, 2019. <https://openreview.net/forum?id=HkgEQnRqYQ>.
- [17] I. Balažević, C. Allen and T. Hospedales, Multi-relational Poincaré Graph Embeddings, in: *Advances in Neural Information Processing Systems*, 2019.
- [18] I. Chami, A. Wolf, D.-C. Juan, F. Sala, S. Ravi and C. Ré, Low-Dimensional Hyperbolic Knowledge Graph Embeddings, in: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Online, 2020, pp. 6901–6914. doi:10.18653/v1/2020.acl-main.617. <https://aclanthology.org/2020.acl-main.617>.
- [19] M. Nickel, V. Tresp and H.-P. Kriegel, A three-way model for collective learning on multi-relational data, in: *Icml*, 2011.
- [20] D.Q. Nguyen, T.D. Nguyen, D.Q. Nguyen and D. Phung, A Novel Embedding Model for Knowledge Base Completion Based on Convolutional Neural Network, in: *Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 2018, pp. 327–333.
- [21] T. Dettmers, P. Minervini, P. Stenetorp and S. Riedel, Convolutional 2d knowledge graph embeddings, in: *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32, 2018.
- [22] D. Nathani, J. Chauhan, C. Sharma and M. Kaul, Learning Attention-based Embeddings for Relation Prediction in Knowledge Graphs, in: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, 2019.
- [23] J. Leblay and M.W. Chekol, Deriving validity time in knowledge graph, in: *Companion Proceedings of the The Web Conference 2018*, 2018, pp. 1771–1776.
- [24] S.S. Dasgupta, S.N. Ray and P. Talukdar, Hyte: Hyperplane-based temporally aware knowledge graph embedding, in: *Proceedings of the 2018 conference on empirical methods in natural language processing*, 2018, pp. 2001–2011.
- [25] Z. Wang, J. Zhang, J. Feng and Z. Chen, Knowledge graph embedding by translating on hyperplanes, in: *Proceedings of the AAAI conference on artificial intelligence*, Vol. 28, 2014.
- [26] R. Abboud, I.I. Ceylan, T. Lukasiewicz and T. Salvatori, BoxE: A Box Embedding Model for Knowledge Base Completion, in: *Proceedings of the Thirty-Fourth Annual Conference on Advances in Neural Information Processing Systems (NeurIPS)*, 2020.
- [27] J. Messner, R. Abboud and I.I. Ceylan, Temporal knowledge graph completion using box embeddings, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36, 2022, pp. 7779–7787.
- [28] K. Chen, Y. Wang, Y. Li and A. Li, RotatEQVS: Representing Temporal Information as Rotations in Quaternion Vector Space for Temporal Knowledge Graph Completion, in: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 5843–5857. doi:10.18653/v1/2022.acl-long.402. <https://aclanthology.org/2022.acl-long.402>.
- [29] C. Xu, M. Nanyeri, F. Alkhoury, H.S. Yazdi and J. Lehmann, TeRo: A time-aware knowledge graph embedding via temporal rotation, *arXiv preprint arXiv:2010.01029* (2020).
- [30] C. Xu, Y.-Y. Chen, M. Nanyeri and J. Lehmann, Temporal Knowledge Graph Completion using a Linear Temporal Regularizer and Multi-vector Embeddings, in: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2021, pp. 2569–2578.
- [31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L.u. Kaiser and I. Polosukhin, Attention is All you Need, in: *Advances in Neural Information Processing Systems*, Vol. 30, I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, eds, Curran Associates, Inc., 2017. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.
- [32] B. Yang, W.-t. Yih, X. He, J. Gao and L. Deng, Embedding entities and relations for learning and inference in knowledge bases, *arXiv preprint arXiv:1412.6575* (2014).

- [33] S. Zhang, Y. Tay, L. Yao and Q. Liu, Quaternion Knowledge Graph Embedding, *arXiv preprint arXiv:1904.10281* (2019).
- [34] A. García-Durán, S. Dumančić and M. Niepert, Learning sequence encoders for temporal knowledge graph completion, *arXiv preprint arXiv:1809.03202* (2018).
- [35] C. Xu, M. Nayyeri, F. Alkhoury, H. Yazdi and J. Lehmann, Temporal knowledge graph completion based on time series gaussian embedding, in: *International Semantic Web Conference*, Springer, 2020, pp. 654–671.
- [36] R. Goel, S.M. Kazemi, M. Brubaker and P. Poupart, Diachronic Embedding for Temporal Knowledge Graph Completion, in: *AAAI*, 2020.
- [37] Z. Cao, Q. Xu, Z. Yang, X. Cao and Q. Huang, Geometry Interaction Knowledge Graph Embeddings, in: *AAAI Conference on Artificial Intelligence*, 2022.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51