Aligning Internal Regularity and External Influence of Multi-granularity for Temporal Knowledge Graph Embedding

Tingyi Zhang[1], Zhixu Li[2], Jiaan Wang[1], Jianfeng Qu[1](✉), Lin Yuan[1], An Liu[1], Lei Zhao[1], and Zhigang Chen[3]

[1] School of Computer Science and Technology, Soochow University, Suzhou, China
{tyzhang1,jawang1,lyuan}@stu.suda.edu.cn, {jfqu,anliu,zhaol}@suda.edu.cn
[2] Shanghai Key Laboratory of Data Science, School of Computer Science, Fudan University, Shanghai, China
zhixuli@fudan.edu.cn
[3] iFLYTEK Research, Suzhou, China
zgchen@iflytek.com

Abstract. Representation learning for the Temporal Knowledge Graphs (TKGs) is an emerging topic in the knowledge reasoning community. Existing methods consider the internal and external influence at either element level or fact level. However, the multi-granularity information is essential for TKG modeling and the connection in between is also under-explored. In this paper, we propose the method that **A**ligning-internal **R**egularity and external **I**nfluence of **M**ulti-granularity for **T**emporal knowledge graph **E**mbedding (ARIM-TE). In particular, to prepare considerate source information for alignment, ARIM-TE first models element-level information via the addition between internal regularity and the external influence. Based on the element-level information, the merge gate is introduced to model the fact-level information by combining their internal regularity including the local and global influence with external random perturbation. Finally, according to the above obtained multi-granular information of rich features, ARIM-TE conducts alignment for them in both structure and semantics. Experimental results show that ARIM-TE outperforms current state-of-the-art KGE models on several TKG link prediction benchmarks.

Keywords: Temporal knowledge graph · Representation learning · Multi-granularity information alignment

1 Introduction

Knowledge Graph (KG) has proved its powerful strength in various downstream tasks, such as recommender systems [18,44], question answering [6,42] and relation extraction [29,37]. Despite its great benefits, most of the KGs suffer from

T. Zhang and Z. Li—Equal contribution.

A. Bhattacharya et al. (Eds.): DASFAA 2022, LNCS 13247, pp. 149–164, 2022.
https://doi.org/10.1007/978-3-031-00129-1_10

incompleteness [36] due to the emergence of new facts. To alleviate this problem, Knowledge Graph Embedding (KGE) is often regarded as an effective approach for Knowledge Graph Completion (KGC) which focuses on deriving new facts based on existing ones. Specifically, KGE methods embed KG elements (i.e., entities and relations) into a low-dimensional vector space while preserving the original semantic and structural information [13]. Recently, Temporal Knowledge Graphs (TKGs) have gained increasing attention. In order to accurately present the temporal knowledge, TKGs include the valid time of each fact, e.g., the fact that Pierre Curie won the Nobel Prize is presented as (*Pierre Curie*, *wonPrize*, *Nobel Prize*, 1903).

Generally, Temporal Knowledge Graph Embedding (TKGE) methods can consider the important information from internal and external perspectives. The internal information models the regularity of evolution in TKG while the external information takes the randomness into consideration. Existing methods try to model the information from either element-level or fact-level: (i) Element-level modeling methods [7,8,10,41] focus on learning the evolving characteristics of the entities and relations in TKG. The external influence is usually modeled as element-level uncertainty. (ii) Fact-level modeling methods [4,20,32,45] aim at incorporating the regularity information or randomness of facts in structure and semantics along the time. However, the information from multi-granularity is integral for TKGE. Specifically, on the one hand, only modeling the important evolution characteristics of the elements, the vital influence from the neighboring facts would be neglected, e.g., the Tokyo 2020 Olympic Games postponed to 2021 due to the occurrence of COVID-19 pandemic. On the other hand, only considering the fact-level modeling would weaken the element-level evolution characteristics in representation, e.g., the Olympics are held every four years rather than a coincidence. Therefore, the alignment should be adopted to explore the semantic and structural characteristics of multi-granular information. Based on the well-modeled features in multi-granularity, alignment could effectively establish an informative and interactive mechanism in between to promote the modeling of multi-granular information and enhance the representation of the TKG. Although some recent methods [15,21] consider the element-level evolution characteristics and the fact-level concurrent or temporally adjacent facts. These methods aim at acquiring better representations in a single granularity. Besides, the external influence in multi-granularity is also ignored. To address the above issues, we propose the method that **A**ligning internal **R**egularity and external **I**nfluence of **M**ulti-granularity for **T**emporal knowledge graph **E**mbedding (ARIM-TE). ARIM-TE leverages the multi-granular information and explores consistence in between through semantic and structural alignment. In order to achieve fully alignment with abundant features, ARIM-TE models the internal regularity with external influence at both element-level and fact-level. Specifically, ARIM-TE first models element-level information as the addition between the internal regularity and external perturbation. Secondly, based on the information flowing from the element level, fact-level information is modeled as the fusion of the internal and external information by introducing the merge gate.

On the one hand, supported by the internal regularity at element level, the modeling of fact-level internal regularity includes the local property of temporal adjacent facts and the global structure revealed by the evolving local property in sequence. On the other hand, the external influence at fact level is modeled as a modification of its respective element-level external influence through introducing Gaussian distributed random perturbation so as to tackle with occasional facts. Finally, ARIM-TE conducts the information alignment in structure and semantics between the multi-granular information.

In summary, our main contributions are as follows:

- We introduce a new TKGE model, namely ARIM-TE, which is the first to simultaneously consider element-level and fact-level information with their respective internal regularity and external influence.
- We propose the merge gate to learn the fact-level information by fusing the internal information with external gaussian distributed random perturbations based on corresponding element-level information.
- We craft the alignment between multi-granular information in structure and semantics to achieve an informative TKGE.
- Experimental results on link prediction task show that our ARIM-TE outperforms the state-of-the-art KGE models on ICEWS14 and GDELT.

2 Related Work

2.1 Static Knowledge Graph Embedding

Traditional KGE methods can be generally divided into two categories: translational distance methods and semantic matching methods [34].

Translational distance methods assume the relation as the translation from the head entity to the counterpart of the tail entity in embedding space. TransE [2] is a typical translation-based model, which cannot precisely deal with complex relations (e.g., 1-to-N, N-to-1 and N-to-N relations) due to its strong assumption. TransH [35] utilizes the hyperplanes to divide the facts based on different relations. TransR [22], TransD [11] and TranSparse [12] construct different projection functions to learn a better representation of KG elements.

Semantic matching methods such as tensor factorization approaches [17,27,28,33,43] model the KG as a three-way tensor and learn the representation of each element by tensor-decomposition. Other approaches [1,5,9,23,26,31] conduct semantic matching through various kinds of neural networks.

Different from the above methods which ignore the temporal information of facts, our paper mainly focuses on TKGE.

2.2 Temporal Knowledge Graph Embedding

TKGE methods model the information from internal and external perspectives. Some methods only pay attention to characteristics of elements. TA-DistMult [7] focuses on modeling time-aware representations of relations with a recurrent neural network. By Graph Convolution Neural Network, TeMP [38] models the structural information of entities within graph and integrates the information across time. TeRo [40] embeds the evolution of entities as rotations. ChronoR [30] models time and relation as the rotation transformation from head entity to tail entity. DE-SimplE [8] obtains the diachronic embeddings for relations and entities. DyERNIE [10] and HERCULES [25] pay attention to hyperbolic embeddings for TKG. DyERNIE leverages velocity vectors to learn dynamic entity representations on Riemannian manifolds. HERCULES utilizes temporal relations as the curvature of Riemannian manifolds. Rather than methods mentioned above, which only consider the internal information, ATiSE [41] reckons the information from external through including the randomness in additive time series decomposition.

Other methods pay attention to fact-level information. HyTE [4] projects each triple into its corresponding time hyperplane. ConT [24] learns a new core tensor for each timestamp. TTransE [20] models the temporal information as a translational vector in score function. TNTComplEx [19] and TeLM [39] utilize discrete timestamps and conduct 4th-order tensor factorization to obtain embeddings. CygNet [45] models the whole fact with copy mode and generation mode. The fact-level external influence is usually modeled with probabilistic, e.g. Know-Evolve [32] leverages temporal point process to model the occurrence of each fact.

Recent methods consider the importance of information in multi-granularity. Re-Net [15] models the joint probability for each fact based on the evolution characteristics of each element and the information on neighborhood aggregator. RE-GCN [21] embeds elements with evolved representations by considering the concurrent or temporally adjacent facts and the static property of entities in their name strings. However, they ignore the external influence in multi-granularity and the lack of alignment in structure and semantics would lead inconsistency in final representation. Different from the existing methods, ARIM-TE conducts semantic and structural alignment in multi-granularity with information which is riched of both internal and external features.

3 Problem Formulation

The notations in this paper are as follows: the lower-case letters denote the scalars, the boldface lower-case letters denote the vectors, and the boldface upper-case letters denote the matrices. Additionally, $\sigma(\cdot)$ is the sigmoid function, tanh is an activation function, $\circ$ denotes the Hadamard product. For vectors $\boldsymbol{v}_1 \in \mathbb{R}^{d_1}$ and $\boldsymbol{v}_2 \in \mathbb{R}^{d_2}$, $[\boldsymbol{v}_1; \boldsymbol{v}_2] \in \mathbb{R}^{d_1+d_2}$ is the concatenation operation.

The TKG $\mathcal{G}$ consists of temporal facts in the form of (h, r, t, τ), where $\tau \in \mathcal{T}$ represents the valid time of the fact and $\mathcal{T}$ is the set of timestamps. h, $t \in \mathcal{E}$, $r \in \mathcal{R}$, where $\mathcal{E}$, $\mathcal{R}$ denote the set of entities and relations, respectively. The TKG can be divided into several sub-KGs: $\mathcal{G} = \mathcal{G}_1 \cup \mathcal{G}_2 \cup \cdots \cup \mathcal{G}_K$ based on the valid time of the fact, K denotes the number of time steps in TKG. The sub-KG $\mathcal{G}_k$ consists of facts that are valid at time step k.

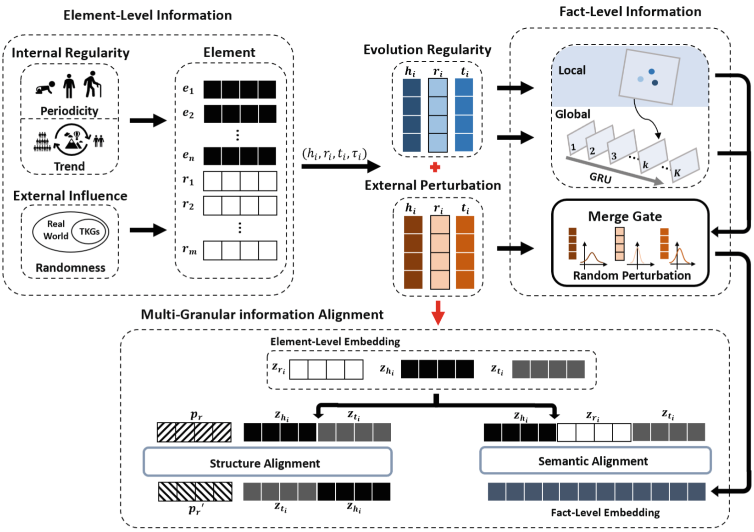

Fig. 1. The overview of our ARIM-TE, consists of element-level information modeling (Sect. 4.1), fact-level information modeling (Sect. 4.2) and multi-granular information alignment (Sect. 4.3).

Temporal Knowledge Graph Embedding aims at learning the low-dimensional representation of head entity $\boldsymbol{h} \in \mathbb{R}^{d\times 1}$, tail entity $\boldsymbol{t} \in \mathbb{R}^{d\times 1}$ and relation $\boldsymbol{r} \in \mathbb{R}^{d\times 1}$. The task of Temporal Knowledge Graph Completion is to derive new facts based on existing ones. In this paper, we focus on interpolation problem [16] which only considers the facts that are valid on the timestamp $\tau \in \mathcal{T}$ during link prediction. We adopt entity prediction task for evaluation.

4 Methodology

We propose a three-step method, which **A**ligning internal **R**egularity and external **I**nfluence of **M**ulti-granularity for **T**emporal knowledge graph **E**mbedding (ARIM-TE). Figure 1 illustrates the overview of our ARIM-TE. As the source of alignment, the multi-granular information is fully modeled with features of both internal regularity and external influence. In the first step, ARIM-TE obtains the periodicity and trend as internal evolution regularity, together with the external perturbation as external influence for each element (Sect. 4.1). In the second step, ARIM-TE utilizes merge gate to model the fact-level information as the fusion of internal regularity and external influence (Sect. 4.2). In the last step, ARIM-TE conducts the information alignment in structure and semantics with informative features in multi-granularity. Considering the form of the fact, ARIM-TE aligned element-level information based on their structure (i.e., (head, relation, tail)). Additionally, the information in multi-granularity is aligned semantically so as to enhance the representation of the TKG (Sect. 4.3).

4.1 Element-Level Information Modeling

The semantic information of each entity and relation in TKG is varied along the time. Some characteristics evolve regularly. For example, the growth of a person shows a trend and the scenic spots usually have peak season and off season. Compared to entities, the semantic information of relations evolves at a lower rate and is relatively more stable during a short period of time [8]. Consequently, representing the regular semantic evolution of each relation with a time-agnostic vector is sufficient. In addition to the internal evolution characteristics, ARIM-TE also measures the external influence of each element to alleviate the incompleteness problem of the TKG. The information out of TKG is modeled as the external perturbation. In this paper, we utilize addition operation to combine the characteristics in learning the element-level information.

$$\boldsymbol{\mu} = \begin{cases} \boldsymbol{\alpha}_e sin(\boldsymbol{\rho}_{s_e} \boldsymbol{\varphi}_\tau + \boldsymbol{\nu}_{s_e}) + \boldsymbol{\rho}_{u_e} \boldsymbol{\varphi}_\tau + \boldsymbol{\nu}_{u_e} & \text{if entity,} \\ \boldsymbol{\alpha}_r & \text{if relation.} \end{cases} \tag{1}$$

$$\boldsymbol{\delta} = \begin{cases} \boldsymbol{\beta}_e \boldsymbol{\zeta}_{\tau_e} & \text{if entity,} \\ \boldsymbol{\beta}_r \boldsymbol{\zeta}_{\tau_r} & \text{if relation.} \end{cases} \tag{2}$$

$$\boldsymbol{z} = \boldsymbol{\mu} + \boldsymbol{\delta}, \tag{3}$$

where $\boldsymbol{\varphi}_\tau$ represents the time embeddings, $\boldsymbol{\mu}$ denotes the evolution regularity of entity or relation. For entity e, $\boldsymbol{\alpha}_e$ denotes the periodicity feature, $sin(\cdot)$ models the periodic activation function which is parameterized by $\boldsymbol{\rho}_{s_e}$ and $\boldsymbol{\nu}_{s_e}$. $\boldsymbol{\rho}_{u_e}$ and $\boldsymbol{\nu}_{u_e}$ fit the semantic evolution trend for entities. Because the semantic information of each relation is relatively stable, $\boldsymbol{\mu}$ simplifies into $\boldsymbol{\alpha}_r$ when representing the relation r. Considering different semantic information stability of relation and entity, the external perturbation $\boldsymbol{\delta}$ for entity and relation includes the separate time variable $\boldsymbol{\zeta}_{\tau_e}$, $\boldsymbol{\zeta}_{\tau_r}$ and element-related characteristics $\boldsymbol{\beta}_e$, $\boldsymbol{\beta}_r$. Finally, the element-level embedding $\boldsymbol{z}$ is obtained by the combination of evolution regularity $\boldsymbol{\mu}$ and external perturbation $\boldsymbol{\delta}$. All of the vectors mentioned above have d dimensions.

After training with the score function $score_e$, the fact $f = (h, r, t, \tau)$ in TKG is initialized with element-level information.

$$score_e = \|\boldsymbol{z}_h + \boldsymbol{z}_r - \boldsymbol{z}_t\|. \tag{4}$$

$$\boldsymbol{z}_f = \boldsymbol{\mu}_f + \boldsymbol{\delta}_f = [\boldsymbol{\mu}_h; \boldsymbol{\mu}_r; \boldsymbol{\mu}_t] + [\boldsymbol{\delta}_h; \boldsymbol{\delta}_r; \boldsymbol{\delta}_t] = [\boldsymbol{z}_h; \boldsymbol{z}_r; \boldsymbol{z}_t]. \tag{5}$$

where $\boldsymbol{\mu}_f$ and $\boldsymbol{\delta}_f$ denotes the fact that initialized by evolution regularity and external perturbation, respectively.

4.2 Fact-Level Information Modeling

In TKG, the knowledge is presented in the form of fact. Though the important evolution characteristics of the elements is obtained after element-level information modeling, the truth that some facts may have correlations or causation relationships in between is still under-explored. The external influence out of the TKG should also be concerned at fact level. Therefore, the merge gate is proposed to combine the external influence with internal local property under the guidance of the global structure. For internal regularity modeling at fact level, ARIM-TE considers the local property of neighboring facts at corresponding time step based on the element-level evolution regularity. The evolving property along the time facilitates the modeling of global structure in TKG which supervise the local property at each time step. The external influence at fact-level is modeled as Gaussian distributed random perturbation based on the corresponding element-level information.

Local Property Encoding. Indeed, the occurrence of each fact would somehow influence the upcoming events. The ignored relevance between facts would lead the imprecise representation modeling. Though some events would have a long-lasting influence on the others, such as the advent of electricity which changed the way people live. The facts that happened within a short time period are more likely to have strong dependencies. Consequently, ARIM-TE adopts local property sequence to aggregate the information of neighboring facts at different time steps. According to the valid time τ, the TKG is divided into sequence $G_k(k \in \{1, 2, 3, ..., K\})$ of length K without overlaps. The specific semantic information of neighboring facts at time step K is modeled on a local hyperplane $\boldsymbol{o}_k$ with its corresponding property projection vector $\boldsymbol{\omega}_k$. For a fact (h, r, t, τ) initialized with element-level regularity characteristics $\boldsymbol{\mu}$ at time step k, the local property is encoded as:

$$\begin{aligned} \boldsymbol{o}_k(h) &= \boldsymbol{\mu}_h - (\boldsymbol{\omega}_k^T \boldsymbol{\mu}_h \boldsymbol{\omega}_k), \\ \boldsymbol{o}_k(r) &= \boldsymbol{\mu}_r - (\boldsymbol{\omega}_k^T \boldsymbol{\mu}_r \boldsymbol{\omega}_k), \\ \boldsymbol{o}_k(t) &= \boldsymbol{\mu}_t - (\boldsymbol{\omega}_k^T \boldsymbol{\mu}_t \boldsymbol{\omega}_k), \end{aligned} \tag{6}$$

Global Structural Modeling. The modeling of local influence only focuses on the information of neighboring facts at corresponding time step, separately. However, some facts would have long-lasting effects. Therefore, ARIM-TE adopts Gated Recurrent Unit (GRU) [3] to obtain the global structure information of the TKG across the time. During global structure learning, ARIM-TE models the local property at each time step with new-coming facts as well as the vital information in history.

$$\boldsymbol{h}_k = \mathrm{GRU}(\boldsymbol{o}_k, \boldsymbol{h}_{k-1}). \tag{7}$$

With the property projection $\boldsymbol{\omega}_k$ at time step k, the hidden representation is updated to $\boldsymbol{h}_k$ which contains structural information up to time step k. In order to keep a consecutive sequence, the hidden representation $\boldsymbol{h}_k$ should be close to the property projection vector $\boldsymbol{\omega}_{k+1}$ on the hyperplane $\boldsymbol{o}_{k+1}$ of the time step $k+1$. ARIM-TE introduces auxiliary loss to guide the modeling of local property in sequence with TKG's global structure.

$$loss_{aux} = \frac{1}{T-1}\sum_{k=1}^{T-1} \|\boldsymbol{h}_k - \boldsymbol{\omega}_{k+1}\|_2^2. \tag{8}$$

Internal and External Influence Merging. Because of the limited coverage of the TKG, the external information should also be considered in order to supplement the information internal. The external influence $\boldsymbol{n}_f$ for fact is modeled as Gaussian distributed random perturbation $\boldsymbol{\epsilon}$ based on the element-level external influence $\boldsymbol{\delta}_f$.

$$\boldsymbol{n}_f = \boldsymbol{\delta}_f \boldsymbol{\epsilon} = [\boldsymbol{\delta}_h \boldsymbol{\epsilon}_h; \boldsymbol{\delta}_r \boldsymbol{\epsilon}_r; \boldsymbol{\delta}_t \boldsymbol{\epsilon}_t] \tag{9}$$

Through the merge gate, the internal regularity in local and global would further merged with external influence through gate mechanism, e.g. a fact(h, r, t, τ):

$$\begin{aligned} \boldsymbol{a}_f &= [\boldsymbol{o}_k(h);\, \boldsymbol{o}_k(r);\, \boldsymbol{o}_k(t)] \\ \boldsymbol{m}_n &= \sigma(\boldsymbol{U}_n \boldsymbol{a}_f + \boldsymbol{W}_n \boldsymbol{n}_f), \\ \boldsymbol{m}_a &= \sigma(\boldsymbol{U}_a \boldsymbol{a}_f + \boldsymbol{W}_a \boldsymbol{n}_f), \\ \tilde{\boldsymbol{m}}_h &= \tanh(\boldsymbol{U}_h \boldsymbol{a}_f + \boldsymbol{W}_h (\boldsymbol{m}_a \circ \boldsymbol{n}_f)), \\ \boldsymbol{m}_f &= (1 - \boldsymbol{m}_n) \circ \tilde{\boldsymbol{m}}_h + \boldsymbol{m}_n \circ \boldsymbol{n}_f. \end{aligned} \tag{10}$$

$\boldsymbol{a}_f$ represents internal regularity of the fact which encoded with corresponding local property. Specifically, the local property in sequence is originally fused with global structural of TKG through training with auxiliary loss. $\boldsymbol{m}_n$ focuses on keeping the necessary external influence of the fact. $\boldsymbol{m}_a$ decides the requirement of internal information. Similar to GRU, we introduce a hidden state $\tilde{\boldsymbol{m}}_h$ in merge gate. Finally, the fact-level representation $\boldsymbol{m}_f$ is calculated by mixing internal regularity and external influence of TKG. $\boldsymbol{U}_n, \boldsymbol{U}_a, \boldsymbol{U}_h, \boldsymbol{W}_n, \boldsymbol{W}_a, \boldsymbol{W}_h$ are the weight matrices of the merge gate.

4.3 Multi-granular Information Alignment

After element-level information learning, ARIM-TE learns the element-level internal evolution regularity and measures the external influence as external perturbation. Further, fact-level information is obtained through the merge gate which fuses the internal regularity with the external random perturbation based on the element-level information within facts. In order to explore the consistence of the information in multi-granularity, ARIM-TE conducts interactive information alignment in structure and semantics.

Considering the facts are denoted with relations and entities in order (i.e., the head entity, the relation and the tail entity), the structure of the facts with different types of relations are ignored through the simple combination of element-level information. Consequently, ARIM-TE aligns each type of relation in the form of fact with its related entity pairs in order (h, t) and in reverse (t, h), respectively:

$$\boldsymbol{q} = \boldsymbol{p}_r[\boldsymbol{z}_h; \boldsymbol{z}_t], \tag{11}$$

$$\boldsymbol{q}^{'} = \boldsymbol{p}_r^{'}[\boldsymbol{z}_t; \boldsymbol{z}_h], \tag{12}$$

$$\boldsymbol{s}_f = [\boldsymbol{p}_r; \boldsymbol{p}_r^{'}; \boldsymbol{q}; \boldsymbol{q} - \boldsymbol{q}^{'}]. \tag{13}$$

The entity pairs are presented with element-level information $\boldsymbol{z}$. $\boldsymbol{q}$, $\boldsymbol{q}^{'}$ represents the corresponding relation-specific structural information with entity pairs in order and in reverse. The information in multi-granularity should be semantic consistence since the fact is the combination of its elements. To align the semantic information of the fact at element-level $\boldsymbol{z}_f$ and fact-level $\boldsymbol{m}_f$, ARIM-TE conducts the introspective alignment instead of simple concatenation.

$$\boldsymbol{c}_f = [\boldsymbol{z}_f; \boldsymbol{z}_f \circ \boldsymbol{m}_f; \boldsymbol{z}_f - \boldsymbol{m}_f; \boldsymbol{m}_f]. \tag{14}$$

The operations $\boldsymbol{z}_f - \boldsymbol{m}_f$ and $\boldsymbol{z}_f \circ \boldsymbol{m}_f$ are element-wise subtraction and multiplication which are targeted at capturing contradiction and amplifying signals, respectively. With the concatenation of structure alignment $\boldsymbol{s}_f$ and semantic alignment $\boldsymbol{c}_f$ information, the plausibility of each potential fact is measured through a multi-layer perceptron (MLP) with the learnable parameters θ.

$$score_f = MLP([\boldsymbol{c}_f; \boldsymbol{s}_f]; \theta). \tag{15}$$

4.4 Model Learning

We adopt negative sampling strategy which randomly replace the head entity or tail entity for each positive fact $f = (h, r, t, \tau) \in \mathcal{G}$. In model learning, we build the query of head entity $(?, r, t, \tau)$ and tail entity $(h, r, ?, \tau)$, then construct the candidate set $S_{f,h}$ and $S_{f,t}$, respectively. Each candidate set consists of the target key and a number of entities selected by negative sampling. Finally, the cross-entropy loss for fact $f \in \mathcal{G}$ with head query and tail query is formulated as:

$$\begin{aligned} loss_{ce} = -(\sum_{f \in \mathcal{G}} log \frac{exp(score(f))}{\sum_{h^{'} \in S_{f,h}} exp(score(h^{'}, r, t, \tau))} \\ + log \frac{exp(score(f))}{\sum_{t^{'} \in S_{f,t}} exp(score(h, r, t^{'}, \tau))}). \end{aligned} \tag{16}$$

where $score(\cdot)$ could be changed in different learning steps. ARIM-TE adopts $score_e$ in element-level information learning. With $score_f$ the total loss in latter learning process is formulated as:

$$loss = loss_{ce}(score_f) + \gamma loss_{aux}. \tag{17}$$

Specifically, $loss_{aux}$ is utilized to assist the modeling of local property with global structure, γ denotes the trade-off hyper-parameter for the auxiliary loss.

5 Experiments

5.1 Datasets

We evaluate our model on two public datasets, i.e., ICEWS14 [7] and GDELT [32]. ICEWS14 is a common benchmark in TKG evaluation which is selected from the public dataset Integrated Crisis Early Warning System (ICEWS). GDELT is the subset of the Global Database of Events, Language, and Tone which was extracted by Trivedi. ICEWS14 includes the facts that happened in 2014 and GDELT retains the facts that occurred from April 1, 2015 to March 31, 2016. Each fact in datasets is annotated with its corresponding valid timestamp. e.g., (*Barack Obama*, *make a visit*, *France*, 2014.02.12). The statistics of each dataset is shown in Table 1.

Table 1. Statistics of experimental datasets.

	$\mathcal{E}$	$\mathcal{R}$	$\mathcal{T}$	Train	Valid	Test
ICEWS14	7,128	230	365	72,826	8,941	8,963
GDELT	500	20	366	2,735,685	341,961	341,961

5.2 Evaluation Metrics and Baselines

We evaluate our ARIM-TE on link prediction task. The link prediction task refers to answer two kinds of queries (i.e., $(?, r, t, \tau)$ and $(h, r, ?, \tau)$) generated from each fact (h, r, t, τ) in the test set. Take the head entity prediction $(?, r, t, \tau)$ as an example, we score and rank all potential entities in the filtered setting [2] which filters the entities according to the facts in TKG, since other entities except for the target head entity h may also link the query as a valid fact in TKG. We follow a similar approach for the tail entity prediction $(h, r, ?, \tau)$ to get the rank of the target entity. We report the Hit@n to show the proportion that the target entity in test set is included in the top-n of the filtered candidate list. Usually, the n is set to 1, 3 and 10. We also provide the Mean Reciprocal Rank (MRR) calculated by averaging the reciprocated rank of the target entity for each query. We compare our ARIM-TE with previous state-of-the-art methods including three static KGE models which ignore the temporal information: TransE [2], DistMult [43], SimplE [17]. We also select several competitive temporal KGE methods: ConT [24], TTransE [14], HyTE [4], TA-DistMult [7], DE-SimplE [8], ATiSE [41], TNTComplEx [19], TeRo [40] and TeMP [38].

5.3 Implementation Details

Our ARIM-TE is implemented using PyTorch. The time granularity in the experiment is set to month so as to alleviate the unbalance issue on each time step. Following the experimental set-up in DE-SimplE, the dimension of embeddings

Table 2. Comparison of different models on ICEWS14 and GDELT. The best results among all models are in bold.

Model	ICEWS14				GDELT			
	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10
TransE	0.280	9.4	–	63.7	0.113	0.0	15.8	31.2
DistMult	0.439	32.3	–	67.2	0.196	11.7	20.8	34.8
SimplE	0.458	34.1	51.6	68.7	0.206	12.4	22.0	36.6
ConT	0.185	11.7	20.5	31.5	0.144	8.0	15.6	26.5
TTransE	0.255	7.4	–	60.1	0.115	0.0	16.0	31.8
HyTE	0.297	10.8	41.6	65.5	0.118	0.0	16.5	32.6
TA-DistMult	0.477	36.3	–	68.6	0.206	12.4	21.9	36.5
DE-SimplE	0.526	41.8	59.2	72.5	0.230	14.1	24.8	40.3
ATiSE	0.545	42.3	63.2	75.7	–	–	–	–
TNTComplEx	0.607	51.9	65.9	77.2	–	–	–	–
TeRo	0.562	46.8	62.1	73.2	–	–	–	–
TeMP-GRU	0.601	47.8	68.1	82.8	0.275	19.1	29.7	43.7
TeMP-SA	0.607	48.4	**68.4**	**84.0**	0.232	15.2	24.5	37.7
ARIM-TE	**0.624**	**56.3**	65.1	74.1	**0.503**	**42.9**	**53.1**	**64.6**

d is 100. We choose the Adam optimizer in the training process. In GDELT, the negative ratio is 5 along the whole process. The batch size is 4096 at element level, then drops to 1024 with the purpose of improving the generalization performance. The learning rate is 1e-3 at element level, then drops to 3e-4. Considering the datasets with different sizes, in ICEWS14 the negative ratio is 500 at element level, then increase to 1000. The batch size is 512 at element level. Same as in GDELT, we adopt a smaller batch size of 128 in latter learning. The learning rate for ICEWS14 is 3e-4 in the whole training process. The trade-off hyper-parameter γ for global structural loss is set to 10 for the two datasets. The two-layer MLP with 1024 and 512 hidden sizes is chosen in information alignment. The dropout rate is tuned from $\{0.0, 0.2, 0.4\}$.

5.4 Results

In this section, we report the performance of our ARIM-TE and compare it with previous state-of-the-art models on two TKG datasets: ICEWS14 and GDELT. The best link prediction evaluation results of each baseline model and our ARIM-TE are shown in Table 2. Table 2 shows that ARIM-TE outperforms the baseline methods on two datasets and achieves SOTA performance on GDELT. On GDELT, ARIM-TE has great improvements in link prediction of 22.8% in MRR, 23.8% in Hit@1, 23.4% in Hit@3 and 20.9% in Hit@10 over the best baseline method. On ICEWS14, ARIM-TE gets the improvement of 4.4% in Hit@1 and

1.7% in MRR compared to the best baseline method, which confirms the importance of multi-granular information in getting accurate representation. Compared to ICEWS14, GDELT has denser training data on each snapshot. Our ARIM-TE fully learns the local property and the global structure on GDELT with its denser training data. Additionally, with more informative interactions between entities in GDELT, the element-level information is explored adequately. The alignment of expressive multi-granularity information significantly enhances the representation of the TKG and greatly improves the link prediction performance on GDELT compared to all the baselines. Though the data in ICEWS14 is relatively sparse, ARIM-TE still gets more accurate link prediction performance which implies that our ARIM-TE could effectively model the information internal and external with alignment of multi-granular information.

5.5 Ablation Study

To better understand our ARIM-TE, we run experiments on GDELT with several variants. The power of multi-granular information is measured through the variant ARIM-TE-E which only considers the element-level information. With the purpose of validating the effectiveness of two different components in element-level evolution regularity modeling, we construct the variants which only models trend ARIM-TE-ET or periodic characteristics ARIM-TE-ES. The performance of global structure and external influence is measured in the variant ARIM-TE-FG and ARIM-TE-FO, respectively. We also test the effectiveness of merge gate in variant ARIM-TE-FM which simply models the combination of the time property internal and external with sum. Finally, we measure the power of alignment in variant ARIM-TE-A which simply combine the multi-granularity information through concatenation. The variant ARIM-TE-AS only considers the alignment in semantics to test effectiveness of the structure alignment.

Table 3. Results for different variants of our ARIM-TE on GDELT. The best results among all models are in bold.

Variant	MRR	Hit@1	Hit@3	Hit@10
ARIM-TE	**0.503**	**42.9**	**53.1**	**64.6**
ARIM-TE-E	0.166	0.0	26.8	42.0
ARIM-TE-ET	0.484	40.8	51.1	63.0
ARIM-TE-ES	0.483	40.8	51.0	62.9
ARIM-TE-FG	0.482	40.6	50.9	62.8
ARIM-TE-FO	0.461	38.3	48.8	61.2
ARIM-TE-FM	0.448	36.9	47.4	60.1
ARIM-TE-A	0.431	35.1	45.6	58.4
ARIM-TE-AS	0.472	39.6	49.7	61.9

The results in Table 3 show that the multi-granular information greatly improves the performance of our ARIM-TE. ARIM-TE gains 42.9% Hit@1 and 33.7% MRR improvements over the variant ARIM-TE-E. The alignment in semantics and structure would significantly improve the performance with 7.8% in Hit@1. The external information at fact level would improve the performance with 4.2% in MRR and 4.6% in Hit@1 than ARIM-TE-FO. The well-combination of the fact-level information internal and external would boost its performance. The merge gate effectively fuses the information with the improvements of 6.0% in Hit@1 and 5.5% in MRR than simply add them together in ARIM-TE-FM. The structure alignment is also important and improves the performance of TKG with 3.3% in Hit@1. The lack of either trend or seasonal characteristics in evolution regularity modeling would weaken the performance of TKGE. Besides, ARIM-TE could better learn the evolving structure of TKG across time with more accuracy through considering the global influence in TKG.

6 Conclusion

We propose a novel model ARIM-TE based on multi-granular information with alignment in structure and semantics to enhance the performance of representation for TKGs. To prepare for an effective alignment, we consider element-level and fact-level information from both perspectives of internal regularity and external influence. Moreover, fact-level information is modeled by the message flow from its respective elements and is further fused by an elaborated merge gate. Experimental results indicate the effectiveness and superior performance of our ARIM-TE on several TKG benchmarks.

Acknowledgement. This research is supported by the National Key R&D Program of China (No. 2018AAA-0101900), the National Natural Science Foundation of China (Grant No. 62072323, 62102276), the Natural Science Foundation of Jiangsu Province (No. BK20191420, BK20210705, BK20211307), the Major Program of Natural Science Foundation of Educational Commission of Jiangsu Province, China (Grant No. 19KJA610002, 21KJD520-005), the Priority Academic Program Development of Jiangsu Higher Education Institutions, and the Collaborative Innovation Center of Novel Software Technology and Industrialization.

References

1. Bordes, A., Glorot, X., Weston, J., Bengio, Y.: A semantic matching energy function for learning with multi-relational data. Mach. Learn. **94**(2), 233–259 (2013). https://doi.org/10.1007/s10994-013-5363-6
2. Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: Neural Information Processing Systems (NIPS), pp. 1–9 (2013)
3. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling. In: NIPS 2014 Workshop on Deep Learning (2014)

4. Dasgupta, S.S., Ray, S.N., Talukdar, P.: HyTE: hyperplane-based temporally aware knowledge graph embedding. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 2001–2011 (2018)
5. Dettmers, T., Minervini, P., Stenetorp, P., Riedel, S.: Convolutional 2D knowledge graph embeddings. In: Thirty-Second AAAI Conference on Artificial Intelligence, pp. 1811–1818 (2018)
6. Dong, L., Wei, F., Zhou, M., Xu, K.: Question answering over Freebase with multi-column convolutional neural networks. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 260–269 (2015)
7. Garcia-Duran, A., Dumančić, S., Niepert, M.: Learning sequence encoders for temporal knowledge graph completion. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 4816–4821 (2018)
8. Goel, R., Kazemi, S.M., Brubaker, M., Poupart, P.: Diachronic embedding for temporal knowledge graph completion. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 3988–3995 (2020)
9. Guo, L., Sun, Z., Hu, W.: Learning to exploit long-term relational dependencies in knowledge graphs. In: International Conference on Machine Learning, pp. 2505–2514 (2019)
10. Han, Z., Chen, P., Ma, Y., Tresp, V.: DyERNIE: dynamic evolution of Riemannian manifold embeddings for temporal knowledge graph completion. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 7301–7316 (2020)
11. Ji, G., He, S., Xu, L., Liu, K., Zhao, J.: Knowledge graph embedding via dynamic mapping matrix. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 687–696 (2015)
12. Ji, G., Liu, K., He, S., Zhao, J.: Knowledge graph completion with adaptive sparse transfer matrix. In: Thirtieth AAAI Conference on Artificial Intelligence, pp. 985–991 (2016)
13. Ji, S., Pan, S., Cambria, E., Marttinen, P., Yu, P.S.: A survey on knowledge graphs: representation, acquisition, and applications. IEEE Trans. Neural Netw. Learn. Syst. 1–21 (2021). https://doi.org/10.1109/TNNLS.2021.3070843
14. Jiang, T., et al.: Towards time-aware knowledge graph completion. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pp. 1715–1724 (2016)
15. Jin, W., Qu, M., Jin, X., Ren, X.: Recurrent event network: autoregressive structure inference over temporal knowledge graphs. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pp. 6669–6683 (2020)
16. Kazemi, S.M., et al.: Representation learning for dynamic graphs: a survey. J. Mach. Learn. Res. **21**, 1–73 (2020)
17. Kazemi, S.M., Poole, D.: Simple embedding for link prediction in knowledge graphs. In: Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, pp. 4289–4300 (2018)
18. Kumar, S., Zhang, X., Leskovec, J.: Predicting dynamic embedding trajectory in temporal interaction networks. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1269–1278 (2019)
19. Lacroix, T., Obozinski, G., Usunier, N.: Tensor decompositions for temporal knowledge base completion. In: 8th International Conference on Learning Representations (2020)

20. Leblay, J., Chekol, M.W.: Deriving validity time in knowledge graph. In: Companion Proceedings of the the Web Conference 2018, pp. 1771–1776 (2018)
21. Li, Z., et al.: Temporal knowledge graph reasoning based on evolutional representation learning. In: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2021, pp. 408–417 (2021)
22. Lin, Y., Liu, Z., Sun, M., Liu, Y., Zhu, X.: Learning entity and relation embeddings for knowledge graph completion. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 2181–2187 (2015)
23. Liu, Q., et al.: Probabilistic reasoning via deep learning: neural association models. arXiv preprint arXiv:1603.07704 (2016)
24. Ma, Y., Tresp, V., Daxberger, E.A.: Embedding models for episodic knowledge graphs. J. Web Semant. 100490 (2019)
25. Montella, S., Rojas-Barahona, L.M., Heinecke, J.: Hyperbolic temporal knowledge graph embeddings with relational and time curvatures. In: Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, pp. 3296–3308 (2021)
26. Nathani, D., Chauhan, J., Sharma, C., Kaul, M.: Learning attention-based embeddings for relation prediction in knowledge graphs. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 4710–4723 (2019)
27. Nickel, M., Rosasco, L., Poggio, T.: Holographic embeddings of knowledge graphs. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 1955–1961 (2016)
28. Nickel, M., Tresp, V., Kriegel, H.P.: A three-way model for collective learning on multi-relational data. In: Proceedings of the 28th International Conference on International Conference on Machine Learning, pp. 809–816 (2011)
29. Riedel, S., Yao, L., McCallum, A., Marlin, B.M.: Relation extraction with matrix factorization and universal schemas. In: Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 74–84 (2013)
30. Sadeghian, A., Armandpour, M., Colas, A., Wang, D.Z.: ChronoR: rotation based temporal knowledge graph embedding. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 6471–6479 (2021)
31. Schlichtkrull, M., Kipf, T.N., Bloem, P., van den Berg, R., Titov, I., Welling, M.: Modeling relational data with graph convolutional networks. In: Gangemi, A., et al. (eds.) ESWC 2018. LNCS, vol. 10843, pp. 593–607. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93417-4_38
32. Trivedi, R., Dai, H., Wang, Y., Song, L.: Know-evolve: deep temporal reasoning for dynamic knowledge graphs. In: Proceedings of the 34th International Conference on Machine Learning, pp. 3462–3471 (2017)
33. Trouillon, T., Welbl, J., Riedel, S., Gaussier, E., Bouchard, G.: Complex embeddings for simple link prediction. In: Proceedings of The 33rd International Conference on Machine Learning, pp. 2071–2080 (2016)
34. Wang, Q., Mao, Z., Wang, B., Guo, L.: Knowledge graph embedding: a survey of approaches and applications. IEEE Trans. Knowl. Data Eng. **29**(12), 2724–2743 (2017)
35. Wang, Z., Zhang, J., Feng, J., Chen, Z.: Knowledge graph embedding by translating on hyperplanes. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 1112–1119 (2014)
36. West, R., Gabrilovich, E., Murphy, K., Sun, S., Gupta, R., Lin, D.: Knowledge base completion via search-based question answering. In: Proceedings of the 23rd International Conference on World Wide Web, pp. 515–526 (2014)

37. Weston, J., Bordes, A., Yakhnenko, O., Usunier, N.: Connecting language and knowledge bases with embedding models for relation extraction. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp. 1366–1371 (2013)
38. Wu, J., Cao, M., Cheung, J.C.K., Hamilton, W.L.: TeMP: temporal message passing for temporal knowledge graph completion. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 5730–5746 (2020)
39. Xu, C., Chen, Y.Y., Nayyeri, M., Lehmann, J.: Temporal knowledge graph completion using a linear temporal regularizer and multivector embeddings. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 2569–2578 (2021)
40. Xu, C., Nayyeri, M., Alkhoury, F., Yazdi, H.S., Lehmann, J.: TeRo: a time-aware knowledge graph embedding via temporal rotation. In: Proceedings of the 28th International Conference on Computational Linguistics, pp. 1583–1593 (2020)
41. Xu, C., Nayyeri, M., Alkhoury, F., Yazdi, H., Lehmann, J.: Temporal knowledge graph completion based on time series gaussian embedding. In: International Semantic Web Conference, pp. 654–671 (2020)
42. Xu, K., Reddy, S., Feng, Y., Huang, S., Zhao, D.: Question answering on Freebase via relation extraction and textual evidence. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 2326–2336 (2016)
43. Yang, B., Yih, S.W.T., He, X., Gao, J., Deng, L.: Embedding entities and relations for learning and inference in knowledge bases. In: Proceedings of the International Conference on Learning Representations (ICLR) 2015 (2015)
44. You, J., Wang, Y., Pal, A., Eksombatchai, P., Rosenberg, C., Leskovec, J.: Hierarchical temporal convolutional networks for dynamic recommender systems. In: The World Wide Web Conference, pp. 2236–2246 (2019)
45. Zhu, C., Chen, M., Fan, C., Cheng, G., Zhang, Y.: Learning from history: modeling temporal knowledge graphs with sequential copy-generation networks. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 4732–4740 (2021)