Learning Knowledge Evolution with Time Duration from Finer-time-granularity Temporal Knowledge Graph

1st Shaochong Du
Department of Automation
Shanghai Jiao Tong University
Shanghai, China
dushaochong@sjtu.edu.cn

2nd Kang Chen
Department of Automation
Shanghai Jiao Tong University
Shanghai, China
chen1127@sjtu.edu.cn

3rd Hao Duan
Department of Automation
Shanghai Jiao Tong University
Shanghai, China
bunny$_duanhao@sjtu.edu.cn$

4th Shijie Yang
Department of Automation
Shanghai Jiao Tong University
Shanghai, China
young1215@sjtu.edu.cn

5th Hong Huo
Department of Automation
Shanghai Jiao Tong University
Shanghai, China
huohong@sjtu.edu.cn

6th Tao Fang
Department of Automation
Shanghai Jiao Tong University
Shanghai, China
tfang@sjtu.edu.cn

***Abstract*—How to represent the temporal information corresponding to each fact in temporal knowledge graphs (TKGs) effectively is always challenging. Most existing representation learning methods usually map the timelines of knowledges into multiple points and ignore the underlying continuous time property and periodic characteristics behind those dense timestamps in finer-time granularity temporal knowledge graphs. A novel Knowledge Evolution with Time Duration (KETD) model is proposed for knowledge representation of temporal knowledge graphs with finer-time granularities. It represents the entity embedding as a time-varying nonlinear function to find the specific continuous time property from dense timestamps. To the best of our knowledge, it is the first attempt to learn embeddings of time durations and combines them into embeddings of entities and relations to predict potential periodic facts. It also uses the temporal point process to capture the impact of historical facts on the current fact. The experiments on self-built finer-time granularity TKG of Satellite-to-Earth Communication (STEC) and two public datasets have demonstrated the superiority of KETD compared to some baseline approaches.**

***Index Terms*—temporal knowledge graph, representation learning, time duration, knowledge evolution**

I. Introduction

Knowledge graph (KG) has been widely used in many fields since it was proposed by Google in 2012 for its excellent ability to represent knowledge in the real world [1]–[3]. As the time goes by, some knowledges may change, while some knowledge may be only meaningful in a specific time period. Hence, the temporal knowledge graph (TKG) has emerged recently [4]–[6]. A fact is represented by a quadruple (s, r, o, τ) in the temporal knowledge graph instead of a triple in the knowledge graph, where s, r and o denote the subject entity, relation, and object entity, respectively, and τ is the time period $[\tau_b, \tau_e]$ that the knowledge is valid, and τ_b, τ_e are the begin time and end time of the fact, respectively. The duration τ can be taken as several timestamps, while the temporal knowledge graph can be viewed as several snapshots corresponding to these timestamps.

The representation learning methods of temporal knowledge graph usually map entities, relations, and timestamps into a vector space and obtain knowledge representations based on this vector space, and only timestamps are considered to represent temporal information of facts. They have been testified to be effective on temporal knowledge graphs with coarse-time granularity, such as YAGO11K [7] (one year) or ICEWS18 [8] (one day). However, in the real world there exist some domain specific temporal knowledge graphs with finer-time granularity (e.g., one hour, one minute, or one second). The finer the granularity is, the more timestamps are. The existing representation learning methods usually map the timelines of knowledges into multiple points, which will no longer be appropriate for representing such dense timestamps. For on one hand, dense timestamps may imply the continuous time property. On the other hand, the facts in temporal knowledge graphs with finer-time granularity may be repeatedly valid in a certain time duration and inherently have periodic characteristics. Therefore, how to represent the facts with dense timestamps and learn the time duration representations of facts in finer-time-granularity TKGs have drawn our great attention.

For knowledge representation of temporal knowledge graphs with finer-time granularities, a novel Knowledge Evolution with Time Duration (KETD) model is proposed. It represents the entity embedding as a nonlinear function over time, which

This study is supported by the National Key Research and Development Program of China (No.2018YFB0505000), the National Science and Technology Major Project (21-Y20A06-9001-17/18), and the Science Fund for Creative Research Groups of the National Natural Science Foundation of China (No. 61221003).

may find the specific continuous time property from dense timestamps. It also learns the embeddings of the time duration of facts and combines them into the embedding of entities and relations for mining potential periodic characteristics. Moreover, the temporal point process is also used to capture the impact of historical facts on the current fact. The contributions of this paper are as follows.

- To the best of our knowledge, the proposed KETD model is the first attempt to learn the embedding of the time duration of facts. The quadruple information $(s, r, o, [\tau_b, \tau_e])$ in the TKG is treated as $(s, r, o, [\tau_b, \tau_d])$, where τ_d is the duration of the fact, i.e., $\tau_e - \tau_b$, and the duration τ_d is combined into the space vector of entities and relations, which may be used to predict the periodic facts.
- For dense timestamps in finer-time-granularity TKGs, the proposed KETD model represents the entity embedding as a nonlinear function over time to find the specific continuous time property of facts.
- A new finer-time-granularity TKG of Satellite-to-Earth Communication (STEC) is constructed, and it has the one-second time granularity. Comprehensive experiments have been conducted on the new dataset and the public datasets, and the experimental results have demonstrated that KETD has outperformed the existing methods.

The rest of this paper is organized as follows. Section II is the related work. The proposed KETD model is elaborated in Section III. The experiment results are in Section IV. Finally, the conclusion is given in Section V.

II. Related Work

Temporal knowledge graph representation learning aims to learn low-dimensional vector representations of entities, relations, and timestamps based on the quadruples in temporal knowledge graphs, and has attracted a lot of research in recent years. Some approaches attempt to learn the timestamp representations, and incorporate the representations of timestamps into the vector space of entities and relations to obtain the score function. Based on the translation model TransE [9], which uses relation representations as translation vectors, TTransE [10] first adds an additional translation vector obtained by a timestamp representation, and then calculates the distance between the transformed subject entities and object entities to judge the score of the quadruple. HyTE [11] directly fuses the entity representation with the timestamp representation to obtain the new entity representation under different snapshots, and then uses TransE to learn the representation under each snapshot. TNTComplEx [12] extends the factorization model ComplEx [13] by adding temporal properties to a part of the relation representations, which can learn from both static and temporal facts in knowledge graphs. TA-DistMult [14] uses LSTM to embed temporal information into the predicate representation based on the factorization model (RESCAL [15], DistMult [16]), and then uses the score function of DistMult to learn representations.

However, all these methods consider entities and relations as separate and independent under different snapshots, and do not reflect the changes of entities and relations over time, resulting in inferior temporal representations. On the other hand, these methods are only considered TKGs with coarse temporal granularity, as the number of timestamps increase rapidly as the temporal granularity becomes finer. They may fail in representing entities and relations under all snapshots.

Some approaches attempt to learn the evolution of entities over time by representing entities in TKGs as diachronic entities, such that there is partial similarity between representations of the same entity at adjacent timestamp. DE-DistMult [17] represents one part of the entity embedding as a time-varying nonlinear function to represent the time-varying properties of the entity, while the other part does not vary with time to represent the static properties of the entity. Know-Evolve [18] uses the temporal point process to capture the impact of historical facts on future facts, and learn the dynamically evolving entity representations. EvoExplore [19] attempts to merge local and global structural evolution into a temporal knowledge graph representation, while using a time-varying nonlinear function to represent dynamic entities to learn the evolutionary process of a TKG. Other methods model temporal knowledge graph representation learning from different perspectives. RE-NET [20] employs a recurrent fact encoder module to encode historical facts and a neighborhood aggregator module to model concurrent facts to capture the information of graph structures. CyGNet [21] can identify repeated facts to predict future facts based on historical vocabulary or entire entity vocabulary.

However, these methods decompose a quadruple containing duration into multiple quadruples, which ignores the influence of the duration of a fact. Actually, the duration of a fact is usually associated with the triple (s, r, o). When the same fact is repeated, the duration of this fact may be more similar to the same historical fact. Therefore, the proposed KETD represents entities as dynamic entities and learns a representation of duration to explore the role of duration for periodic facts.

III. Proposed Method

In this section, the proposed KETD model is elaborated. The related definitions are given at first. The specific notations and descriptions are listed in Table I.

A. Related Definitions

1) Temporal Knowledge Graph: A temporal knowledge graph is defined as a directed graph $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{T}, \mathcal{D})$ containing timestamps and durations, where $\mathcal{E}$ and $\mathcal{R}$ are the entity set and relation set, respectively, and $\mathcal{T}$ and $\mathcal{D}$ are the begin time set and duration set of facts, respectively. Each fact in a TKG is represented by a quadruple $(s, r, o, [\tau_b, \tau_d])$, where $s, o \in \mathcal{E}$, $r \in \mathcal{R}$, $\tau_b \in \mathcal{T}$, and $\tau_d \in \mathcal{D}$.

2) Knowledge Evolution: Given a temporal knowledge graph $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{T}, \mathcal{D})$, Knowledge Evolution refers to the establishment process of relationships between entities as an evolutionary process over time, based on historical facts.

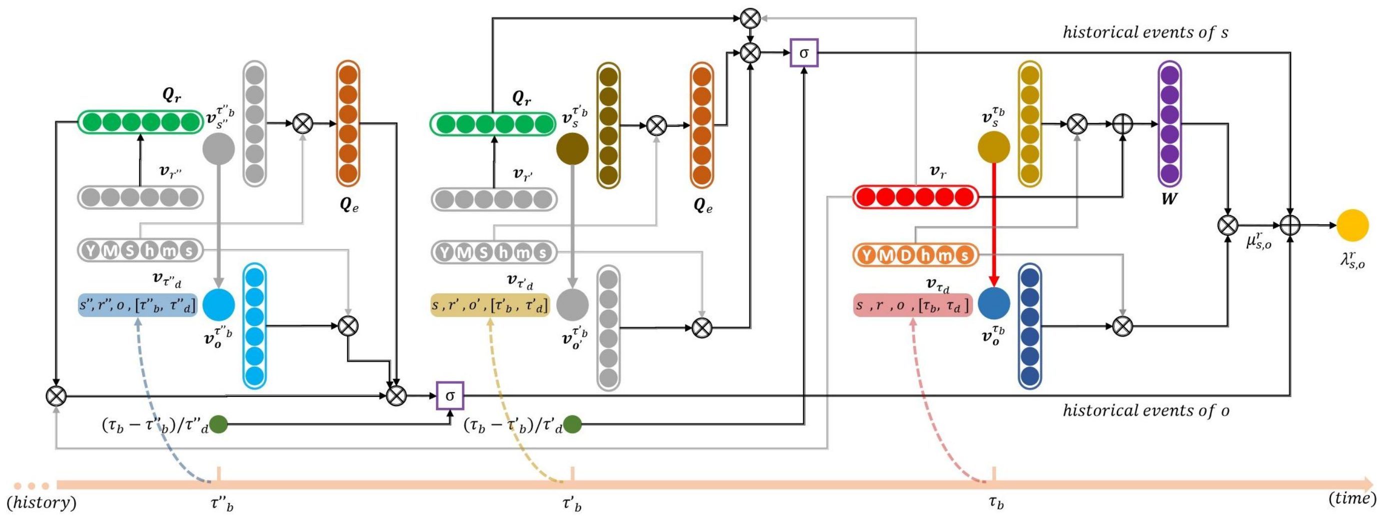

Fig. 1. Illustration of the intensity calculation process for each fact. From Eq. 2, the intensity of quadruple $(s, r, o, [\tau_b, \tau_d])$ occurring at time point t. $(s, r', o', [\tau'_b, \tau'_d])$, $(s'', r'', o, [\tau''_b, \tau''_d])$ are the historical facts for the subject entity s and object entity o, respectively.

TABLE I
NOTATIONS AND DESCRIPTIONS

Notation	Description
$\mathcal{G}$	Temporal knowledge graph
$\mathcal{E}, \mathcal{R}$	Entity set, relation set
$\mathcal{T}, \mathcal{D}$	Begin timestamp set, time duration set
q	A quadruple in temporal knowledge graph
S_F	A discrete sequence of facts
d	The dimension of embedding
$\boldsymbol{v}_e^\tau$	Representation of entity e at time τ
$\boldsymbol{v}_r$	Representation of relation r
$\boldsymbol{v}_{\tau_d}$	Representation of time duration τ_d
$\boldsymbol{W}$	Parameter matrix
$\boldsymbol{Q}_e, \boldsymbol{Q}_r$	Matrix for measuring similarity

The process can be denoted as a discrete sequence of facts $S_F = \{q_i\}_{i=1}^N = \{(s, r, o, [\tau_b, \tau_d])_i\}_{i=1}^N$, which is obtained according to the order of begin time, i.e., τ_b, where N is the number of facts in this TKG, q_i is the i_{th} fact in the TKG.

3) Temporal Knowledge Graph Representation Learning: The purpose of our knowledge representation learning is to learn three mapping functions $f_E : \mathcal{E} \to \mathbb{R}^d$, $f_R : \mathcal{R} \to \mathbb{R}^d$ and $f_D : \mathcal{D} \to \mathbb{R}^d$ to map the entities $\mathcal{E}$, relations $\mathcal{R}$, and time durations $\mathcal{D}$ in a TKG $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{T}, \mathcal{D})$ to a vector space, respectively. The mapping functions will capture the evolution of knowledge in the TKG over time.

B. Model Components

The proposed KETD model represents the entity embedding as a nonlinear function over time, and it uses the temporal point process to capture the impact of historical facts on the current fact. The temporal point process assumes that historical facts that occurred before the current moment τ affect the current fact.

Hawkes process [22] is used to model the facts in TKG. The intensity function of Hawkes process takes into account both the base intensity of the current fact itself and the influence of historical facts on future facts, which decays as the time goes by. And its intensity function is defined as follows:

$$\lambda(\tau) = \mu(\tau) + \sum_{\tau_i < \tau} g(\tau - \tau_i) \tag{1}$$

where $\mu(\tau)$ denotes the base intensity of the current fact, $g(\tau - \tau_i)$ is used to describe the decaying impact of historical facts as the time goes by.

The intensity function of each fact $(s, r, o, [\tau_b, \tau_d])$ in the TKG is defined according to Hawkes process as follows:

$$\begin{aligned} \lambda^r_{s,o}(\tau_b) =& \mu^r_{s,o}(\tau_b, \tau_d) + \\ & \theta \sum_{q_s \in H_s(\tau_b)} \alpha^{r^i}_{o^i}(\tau^i_b, \tau^i_d) k(\tau_b - \tau^i_b) + \\ & \theta \sum_{q_o \in H_o(\tau_b)} \alpha^{r^j}_{s^j}(\tau^j_b, \tau^j_d) k(\tau_b - \tau^j_b) \end{aligned} \tag{2}$$

where $\mu^r_{s,o}(\tau_b, \tau_d)$ denotes the base intensity of the fact. The last two lines of (2) are the historical impact of the subject and object entities. $H_s(\tau_b)$, $H_o(\tau_b)$ are the set of historical facts occurring at s or o before begin time τ_b, and $q_s = (s, r^i, o^i, [\tau^i_b, \tau^i_d])$, $q_o = (s^j, r^j, o, [\tau^j_b, \tau^j_d])$ are facts in the sets, respectively. $\alpha^{r^i}_e(\tau^i_b, \tau^i_d)$ denotes the excitation intensity coefficient of Hawkes process. $k(\tau_b - \tau^i_b)$ is the excitation function that decays with time, and the exponential decay function as:

$$k(\tau_b - \tau^i_b) = \exp(-\delta_t \cdot (\tau_b - \tau^i_b)) \tag{3}$$

where δ_t is a learnable decay rate. Thus the score function for each fact is divided into two parts, the base intensity and the influence of history, and θ is a learnable trade-off parameter. Fig. 1 illustrates the intensity calculation process for a quadruple $(s, r, o, [\tau_b, \tau_d])$ in KETD, where $(s, r', o', [\tau'_b, \tau'_d])$ and

Manuscript 736 submitted to 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Received April 20, 2023.

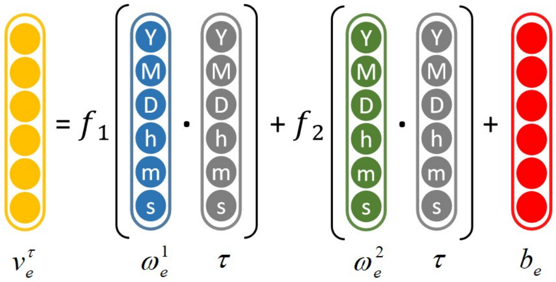

Fig. 2. Dynamic entity embedding at time τ

$(s'', r'', o, [\tau''_b, \tau''_d])$ are historical facts related to the subject and object entities in the current quadruplet that occurred before timestamp τ_b.

1) Base intensity: In a TKG, whether a relation is established between two entities is related to the duration of the fact, the semantics of the relation and the semantics of the both entities at the current moment. Thus, the base intensity is calculated through representations of entities, relations and time durations. Inspired by DE-DistMult [17], durations are combined into entity representations instead of timestamps to learn the impact of durations on facts. In particular, inspired by TransE [9], the relation is taken as a transformation from the subject entity to the object entity. However, TransE has difficulty dealing with one-to-many and many-to-one relations.

Thus, a matrix is utilized to obtain the base intensity of a quadruple $q = (s, r, o, [\tau_b, \tau_d])$ as follows:

$$\mu^r_{s,o}(\tau_b, \tau_d) = \left(\boldsymbol{v}_s^{\tau_b} \cdot \boldsymbol{v}_{\tau_d} + \boldsymbol{v}_r\right)^\top \boldsymbol{W} \left(\boldsymbol{v}_o^{\tau_b} \cdot \boldsymbol{v}_{\tau_d}\right) \tag{4}$$

where $\boldsymbol{W} \in \mathbb{R}^{d\times d}$ a matrix to measure the similarity, $\boldsymbol{v}_{\tau_d}, \boldsymbol{v}_r \in \mathbb{R}^d$ is the representation of relation r and time duration τ_d, respectively. $\boldsymbol{v}_s^{\tau_b}, \boldsymbol{v}_o^{\tau_b} \in \mathbb{R}^d$ is the representation of the subject entity s and the object entity o at time τ_b, respectively. Since the connections of the relations in TKG change with time, we establish the entities as diachronic entities $\boldsymbol{v}_e^\tau$, i.e., the representation of the entity is a nonlinear function that changes with time τ to learn the evolution of the entity as follows:

$$\boldsymbol{v}_e^\tau = f_1(\boldsymbol{\omega}_e^1 \cdot \tau) + f_2(\boldsymbol{\omega}_e^2 \cdot \tau) + \boldsymbol{b}_e \tag{5}$$

where $\boldsymbol{\omega}_e^1, \boldsymbol{\omega}_e^2 \in \mathbb{R}^d$ is the parameter to be trained. $f_1(\cdot)$ is a periodic function, which is used to learn the evolution of recurring facts. $f_2(\cdot)$ is a monotonic function, which is used to learn the evolution of non-periodic facts. $\sin(\cdot)$ and $\tanh(\cdot)$ are chosen as $f_1(\cdot)$ and $f_2(\cdot)$ respectively. Fig. 2 shows the calculation process of dynamic entity embedding at time τ.

2) Influence of history: The occurrence of a fact in TKG is also related to the historical facts, which the subject or object entities are involved. For example, if a fact has happened many times in history, the probability of this fact occurring in the future will also be high. Since the historical facts of the subject entity $q_s = (s, r', e, [\tau'_b, \tau'_d])$ and the historical facts of the object entity $q_o = (e, r', o, [\tau'_b, \tau'_d])$ have an impact on the current fact $q = (s, r, o, [\tau_b, \tau_d])$. Inspired by EvoExplore [19], the impact of historical facts are represented by calculating the similarity between the relations and entities involved in the current fact and its related historical facts, and the excitation intensity coefficient is defined as follows:

$$\alpha_e^{r'}(\tau'_b, \tau'_d) = \Omega \cdot \mathrm{Sim}\left(\boldsymbol{v}_s^{\tau_b}, \boldsymbol{v}_o^{\tau_b}, \tau_d\right) \cdot k\left(\frac{\tau_b - \tau'_b}{\tau'_d}\right) \tag{6}$$

$$\Omega = \omega(r, r') \cdot \omega(x, e) \tag{7}$$

$$\mathrm{Sim}\left(\boldsymbol{v}_s^{\tau_b}, \boldsymbol{v}_o^{\tau_b}, \tau_d\right) = \left(\boldsymbol{v}_s^{\tau_b} \cdot \boldsymbol{v}_{\tau_d}\right)^\top \boldsymbol{Q}_e \left(\boldsymbol{v}_o^{\tau_b} \cdot \boldsymbol{v}_{\tau_d}\right) \tag{8}$$

where x represents the current target entity, i.e., when calculating the historical fact intensity of s, x is s, and e is the entity that has a relation r' with x. $\boldsymbol{Q}_e \in \mathbb{R}^{d\times d}$ is the matrix used to measure the degree of association between entities. $k(\cdot)$ is an excitation function that decays with time, and the longer the duration, the longer a historical fact lasts, the greater its impact on future facts. An exponential decay function is defined as follows:

$$k\left(\frac{\tau_b - \tau'_b}{\tau'_d}\right) = \exp\left(-\delta_d \cdot \frac{\tau_b - \tau'_b}{\tau'_d}\right) \tag{9}$$

where δ_d is a learnable decay rate. $\omega(x, e), \omega(r, r')$ are the importance of entity e and relation r' of the historical fact for entity x and relation r of the current fact, respectively, which are defined as follows:

$$\omega(x, e) = \frac{\exp\left[\mathrm{Sim}\left(\boldsymbol{v}_x^{\tau'_b}, \boldsymbol{v}_e^{\tau'_b}, \tau'_d\right)\right]}{\sum_{e' \in E_e(\tau_b)} \exp\left[\mathrm{Sim}\left(\boldsymbol{v}_x^{\tau'_b}, \boldsymbol{v}_{e'}^{\tau'_b}, \tau'_d\right)\right]} \tag{10}$$

$$\omega(r, r') = \frac{\exp\left(\boldsymbol{v}_r^\top \boldsymbol{Q}_r \boldsymbol{v}_{r'}\right)}{\sum_{r'' \in R_e(\tau_b)} \exp\left(\boldsymbol{v}_r^\top \boldsymbol{Q}_r \boldsymbol{v}_{r''}\right)} \tag{11}$$

where $E_e(\tau_b), R_e(\tau_b)$ are the set of subject or object entities and the set of relations, which are involved in the historical facts of entity x, respectively. $\boldsymbol{Q}_r \in \mathbb{R}^{d\times d}$ is the matrix used to measure the degree of association between relations.

C. Loss Function

By the above calculation, the intensity of each quadruple $q = (s, r, o, [\tau_b, \tau_d])$ in TKG is obtained. The sigmoid function $\sigma(\cdot)$ is used to ensure that the intensity is positive and constructs two types of negative samples $(e, r, o, [\tau_b, \tau_d])$ and $(s, r, e, [\tau_b, \tau_d])$ for each quadruple $(s, r, o, [\tau_b, \tau_d])$. In order for the correct entity to have a higher intensity than the other entities, the loss function is defined as follows:

$$L = -\sum_{q \in S_F} \frac{\sigma(\lambda^r_{s,o}(\tau_b, \tau_d))}{\sum_{e \in \mathcal{E}} \sigma(\lambda^r_{e,o}(\tau_b, \tau_d)) + \sigma(\lambda^r_{s,e}(\tau_b, \tau_d))} \tag{12}$$

Since the temporal knowledge graph contains a large number of entities, it is too expensive to compute all the negative sample intensities. To solve this problem, we use negative sampling to avoid computing the entire entity set. For each quadruple q, entities from the entity set are randomly selected to generate a negatively sampled entity set N^s for s and a

TABLE II
RESULTS (IN PERCENTAGE) ON ICEWS14-TD AND ICEWS18-TD.

Model	ICEWS14-TD					ICEWS18-TD				
	MRR	Hit@1	Hit@3	Hit@5	Hit@10	MRR	Hit@1	Hit@3	Hit@5	Hit@10
TransE	16.25	11.63	17.97	21.21	24.56	16.91	12.31	17.97	21.14	25.68
DistMult	15.83	5.60	21.77	26.93	35.14	21.60	11.81	26.92	31.78	38.68
TTransE	12.35	9.54	12.81	14.68	17.54	13.31	9.61	13.93	17.10	20.15
HyTE	16.43	12.44	17.54	20.15	24.00	12.08	8.58	12.56	14.93	18.53
DE-DistMult	12.97	5.04	17.29	25.25	35.76	16.22	5.16	20.71	27.55	38.93
EvoExplore	23.01	11.56	25.62	36.31	47.38	23.01	10.82	26.18	35.57	48.45
KETD	**27.48**	**13.31**	**33.40**	**43.66**	**55.91**	**32.34**	**17.41**	**38.74**	**49.17**	**62.58**

The best and second best results are boldfaced and underlined respectively. The higher each metric is, the better the model performs.

TABLE III
RESULTS (IN PERCENTAGE) ON STEC.

Model	STEC				
	MRR	Hit@1	Hit@3	Hit@5	Hit@10
TransE	31.80	15.05	37.13	55.35	65.98
DistMult	35.14	19.47	38.81	56.78	69.40
TTransE	33.64	16.60	39.55	57.21	68.28
HyTE	34.74	17.72	40.05	59.58	71.39
DE-DistMult	27.82	8.89	35.07	52.67	64.43
EvoExplore	42.83	24.63	51.62	71.70	78.61
KETD	**48.25**	**30.04**	**57.96**	**73.82**	**87.38**

TABLE IV
INFORMATIONS OF DATASETS

dataset	$\lvert\mathcal{E}\rvert$	$\lvert\mathcal{R}\rvert$	δ	|train|	|valid|	|test|
ICEWS14-TD	12498	260	1d	129766	64883	64883
ICEWS18-TD	23033	256	1d	119556	14945	14944
STEC	105	3	1s	91307	1903	1902

negatively sampled entity set N^o for o, respectively. $|\cdot|$ is the length of the set. The modified loss function is as follows.

$$\begin{aligned}L = -\sum_{q\in S_F}\{\log\left[\sigma(\lambda^r_{s,o}(\tau_b,\tau_d))\right] + \\ \frac{1}{|N^s|}\sum_{e\in N^s}\log\left[\sigma(-\lambda^r_{o,o}(\tau_b,\tau_d))\right] + \\ \frac{1}{|N^o|}\sum_{e\in N^o}\log\left[\sigma(-\lambda^r_{s,o}(\tau_b,\tau_d))\right]\}\end{aligned} \tag{13}$$

IV. EXPERIMENTS

In this section, we use link prediction to compare KETD with existing baseline methods by using existing public datasets and a Satellite-to-Earth Communication Dataset (STEC) that we created to demonstrate its superiority. Table IV summarizes the statistics for these datasets and δ represents the time granularity of each dataset.

A. Dataset

Public datasets. Experiments are conducted on the public datasets ICEWS14 and ICEWS18. Since these datasets do not contain duration information, we process the data to obtain the corresponding duration of each fact.

The processing process is as follows: first, the whole dataset is made into a map, where the key is the triple (s, r, o) in the dataset and the value is the set of valid timestamps of the triple. Then, we sort the set of timestamps in the map and merge the consecutive timestamps (the difference between adjacent timestamps is less than the time granularity of the dataset) into $[\tau_b, \tau_d]$. Finally, each time information is combined with a triple to generate the new datasets ICEWS14-TD and ICEWS18-TD of quadruple $(s, r, o, [\tau_b, \tau_d])$.

STEC dataset. This dataset is a temporal dataset of the communication between the satellites and the ground stations obtained by using STK software. 26 satellites, 5 ground stations, 43 sensors, and 31 observed targets, for a total of 105 different entities and 3 types of relationships are considered. It has a temporal granularity of one second.

B. Evaluation Method

The quality of TKG representation learning is evaluated by link prediction, which aims to predict the missing knowledge in the TKG $((s, r, ?, [\tau_b, \tau_d])$ or $(?, r, o, [\tau_b, \tau_d]))$. For each quadruple $q = (s, r, o, [\tau_b, \tau_d])$ in the test set, we use all entities as candidate entities, get the intensity scores of the quadruple of candidate entities and rank them in descending order. Finally, the correct entity rank r_s and r_o are used as the evaluated metric.

Based on the above, two popular types of metrics are used to evaluate the performance of the model. One is Mean Reciprocal Rank (MRR), defined as MRR $= \frac{1}{2\times|\text{test}|}\sum_{q\in\text{test}}\left(\frac{1}{r_s}+\frac{1}{r_o}\right)$, and the higher the MRR, the better the model performance. The other is Hit@K, defined as

Manuscript 736 submitted to 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Received April 20, 2023.

$\text{Hit@}K = \frac{1}{2\times|\text{Test}|}\sum_{q\in\text{Test}}[\mathbb{K}(r_s \le K) + \mathbb{K}(r_o \le K)]$, where $\mathbb{K}(cond)$ is 1 if $cond$ holds and 0 otherwise.

C. Performance Comparison

We compare KETD with static and temporal knowledge graph representation learning methods. Including the static knowledge graph representation learning methods, TransE and DistMult, the temporal knowledge graph representation learning methods, TTransE, HyTE, DE-DistMult, and EvoExplore.

For the baseline methods, we used the hyperparameter settings from the corresponding papers and trained the same number of epochs as our model for comparison. For the proposed KETD model, the Adam optimizer is used. The number of epochs is set to 200, the batch training size is set to 512, the learning rate is set to 0.001, and the dropout value is set to 0.4. The dimensions represented $d \in \{50, 100, 200\}$, the number of negative samples $n \in \{3, 5, 10\}$, and the length of historical information $h \in \{1, 3, 5, 7\}$. The results of the comparison experiments are shown in Table II and Table III.

Based on the results in Table II and Table III, we can obtain the following findings. **1**. TTransE and DE-DistMult failed to outperform the static methods TransE and DistMult on ICEWS14-TD and ICEWS18-TD datasets, where the periodicity feature is less obvious, probably because of the less reasonable way of modeling temporal information used on these datasets. In contrast, the temporal representation learning method has better performance than the corresponding static method in the STEC dataset, where the periodicity feature is obvious. This result demonstrates that the inclusion of temporal information helps to obtain a more accurate representation for knowledge with distinct periodicity. **2**. We find that modeling TKGs using the temporal point process has better performance when processing the quadruples (s, r, o, τ) in the dataset into new quadruples $(s, r, o, [\tau_b, \tau_d])$. This is because the temporal point process can capture the effect of historical facts of subject and object entities on the knowledge embedding representation. **3**. It can be seen that KETD significantly outperforms the baseline approach and achieves the most advanced results in terms of MRR on all datasets. First, KETD uses a diachronic entity representation of entities based on a temporal point process, which helps to capture the evolution of entities over time. Second, KETD models the time duration, which captures the effect of duration on periodic facts to provide accurate prediction results.

D. Ablation Study

In order to evaluate the impact of each component of KETD on performance, the following ablation experiments are conducted, and Table V gives the experimental results for different variants of KETD on the STEC dataset. *w/o dynamic* indicates that dynamic entities are not used, *w/o history* indicates that the impact of historical information is not used, *w/o d-embedding* indicates that duration embedding is not used, and *w/o d-decay* indicates that the decay of historical facts over time is not used in (6), i.e., the decay of the impact of historical facts is only related to the begin time. From the experimental results, it can be seen that all these components have a positive effect on the overall model performance improvement.

TABLE V
RESULTS (IN PERCENTAGE) BY DIFFERENT VARIANTS OF KETD ON STEC.

Model	STEC				
	MRR	Hit@1	Hit@3	Hit@5	Hit@10
w/o dynamic	42.28	22.26	51.24	73.01	85.63
w/o history	26.76	9.89	27.86	53.98	64.61
w/o d-embedding	45.50	27.05	55.35	73.26	82.28
w/o d-decay	35.31	14.05	45.83	67.35	80.66
KETD	48.25	30.04	57.96	73.82	87.38

V. CONCLUSION

In this paper, a novel Knowledge Evolution with Time Duration (KETD) model is proposed, and it is a first attempt to combine the time duration of an fact into the representation of a TKG. It also is based on a temporal point process to learn a representation of the knowledge graph with a finer time granularity through base intensity and historical information. For the base intensity, a time-varying nonlinear function is used to describe dynamic entities. For historical information, the impact of historical facts are represented by calculating the similarity between the current fact and the historical facts involved in its subject and object entities. The experimental results demonstrate the superiority of KETD compared with the baseline methods. In the future, we will improve the interpretability of the model by combining effective durations with a more rational approach.

REFERENCES

[1] X. Wang, X. He, Y. Cao, M. Liu, and T.-S. Chua, "Kgat: Knowledge graph attention network for recommendation," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 950–958. [Online]. Available: https://doi.org/10.1145/3292500.3330989

[2] X. Huang, J. Zhang, D. Li, and P. Li, "Knowledge graph embedding based question answering," ser. WSDM '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 105–113. [Online]. Available: https://doi.org/10.1145/3289600.3290956

[3] P. Chen, Y. Lu, V. W. Zheng, X. Chen, and B. Yang, "Knowedu: A system to construct knowledge graph for education," *IEEE Access*, vol. 6, pp. 31 553–31 563, 2018.

[4] T. Jiang, T. Liu, T. Ge, L. Sha, S. Li, B. Chang, and Z. Sui, "Encoding temporal information for time-aware link prediction," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin, Texas: Association for Computational Linguistics, Nov. 2016, pp. 2350–2354. [Online]. Available: https://aclanthology.org/D16-1260

[5] Z. Li, X. Jin, W. Li, S. Guan, J. Guo, H. Shen, Y. Wang, and X. Cheng, "Temporal knowledge graph reasoning based on evolutional representation learning," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 408–417. [Online]. Available: https://doi.org/10.1145/3404835.3462963

[6] A. Sadeghian, M. Armandpour, A. Colas, and D. Z. Wang, "Chronor: Rotation based temporal knowledge graph embedding," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 7, pp. 6471–6479, May 2021. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/16802

[7] 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*. New York, NY, USA: Association for Computing Machinery, 2007, p. 697–706. [Online]. Available: https://doi.org/10.1145/1242572.1242667

[8] E. Boschee, J. Lautenschlager, S. O'Brien, S. Shellman, J. Starz, and M. Ward, "ICEWS Coded Event Data," *Harvard Dataverse*, vol. 12, 2015. [Online]. Available: https://doi.org/10.7910/DVN/28075

[9] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *Advances in Neural Information Processing Systems*, vol. 26. Curran Associates, Inc., 2013.

[10] J. Leblay and M. W. Chekol, "Deriving validity time in knowledge graph," in *Companion Proceedings of the The Web Conference 2018*. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2018, pp. 1771–1776.

[11] 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*. Brussels, Belgium: Association for Computational Linguistics, Oct.-Nov. 2018, pp. 2001–2011.

[12] T. Lacroix, G. Obozinski, and N. Usunier, "Tensor Decompositions for temporal knowledge base completion," *arXiv e-prints*, p. arXiv:2004.04926, Apr. 2020.

[13] T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, and G. Bouchard, "Complex embeddings for simple link prediction," in *International conference on machine learning*, ser. Proceedings of Machine Learning Research, vol. 48. New York, NY, USA: PMLR, 20–22 Jun 2016, pp. 2071–2080.

[14] A. García-Durán, S. Dumančić, and M. Niepert, "Learning sequence encoders for temporal knowledge graph completion," *arXiv e-prints*, p. arXiv:1809.03202, Sep. 2018.

[15] M. Nickel, V. Tresp, and H.-P. Kriegel, "A three-way model for collective learning on multi-relational data," in *Proceedings of the 28th International Conference on International Conference on Machine Learning*, 2011, p. 809–816.

[16] B. Yang, W.-t. Yih, X. He, J. Gao, and L. Deng, "Embedding entities and relations for learning and inference in knowledge bases," *arXiv e-prints*, p. arXiv:1412.6575, December 2014.

[17] R. Goel, S. M. Kazemi, M. Brubaker, and P. Poupart, "Diachronic embedding for temporal knowledge graph completion," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, Apr. 2020, pp. 3988–3995.

[18] R. Trivedi, H. Dai, Y. Wang, and L. Song, "Know-evolve: Deep temporal reasoning for dynamic knowledge graphs," in *international conference on machine learning*, ser. Proceedings of Machine Learning Research, vol. 70. PMLR, 06–11 Aug 2017, pp. 3462–3471.

[19] J. Zhang, S. Liang, Y. Sheng, and J. Shao, "Temporal knowledge graph representation learning with local and global evolutions," *Knowledge-Based Systems*, vol. 251, p. 109234, 2022.

[20] W. Jin, M. Qu, X. Jin, and X. Ren, "Recurrent event network: Autoregressive structure inference over temporal knowledge graphs," *arXiv e-prints*, p. arXiv:1904.05530, Apr. 2019.

[21] C. Zhu, M. Chen, C. Fan, G. Cheng, and Y. Zhang, "Learning from history: Modeling temporal knowledge graphs with sequential copy-generation networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 5, May 2021, pp. 4732–4740.

[22] A. G. Hawkes, "Spectra of some self-exciting and mutually exciting point processes," *Biometrika*, vol. 58, no. 1, pp. 83–90, 04 1971.