Logic and Commonsense-Guided Temporal Knowledge Graph Completion

Guanglin Niu\textsuperscript{1}, Bo Li\textsuperscript{1,2}\textsuperscript{*}

\textsuperscript{1} Institute of Artificial Intelligence, Beihang University, Beijing, China
\textsuperscript{2} Hangzhou Innovation Institute, Beihang University, Hangzhou, China
\{beihangngl, boli\}@buaa.edu.cn

Abstract

A temporal knowledge graph (TKG) stores the events derived from the data involving time. Predicting events is extremely challenging due to the time-sensitive property of events. Besides, the previous TKG completion (TKGC) approaches cannot represent both the timeliness and the causality properties of events, simultaneously. To address these challenges, we propose a Logic and Commonsense-Guided Embedding model (LCGE) to jointly learn the time-sensitive representation involving timeliness and causality of events, together with the time-independent representation of events from the perspective of commonsense. Specifically, we design a temporal rule learning algorithm to construct a rule-guided predicate embedding regularization strategy for learning the causality among events. Furthermore, we could accurately evaluate the plausibility of events via auxiliary commonsense knowledge. The experimental results of TKGC task illustrate the significant performance improvements of our model compared with the existing approaches. More interestingly, our model is able to provide the explainability of the predicted results in the view of causal inference. The appendix, source code and datasets of this paper are available at https://github.com/ngl567/LCGE.

Introduction

Knowledge graph (KG) has been developed rapidly in recent years, which stores facts in the form of (subject, predicate, object). To further exploit the events involving time, temporal KG (TKG) represents each event as a quadruple (subject, predicate, object, time) where the time information can be formulated by a timestamp or time interval. For instance, an event (Barack Obama, Consult, Xi Jinping, 2014-11-11) in a TKG as shown in Figure 1 indicates that this event occurs on the definite date of 2014-11-11.

Temporal KG completion (TKGC) is an essential technique to predict whether some potential events missing in the TKG will occur, i.e., (Xi Jinping, Consult, Barack Obama, 2014-06-15) shown in Figure 1. Particularly, an event is only valid at a specific time namely the timeliness. The existing TKGC approaches can be classified into two categories: (1) the evolution-based models are capable of representing the causality among events to reason the future events, such as Know-Evolve (Trivedi et al. 2017) RE-NET (Jin et al. 2020) and CyGNet (Zhu et al. 2021). As the declaration of causality of events is shown in Figure 1, when two events occur in certain time order, one event has an impact on the other. The event occurring earlier is the reason and the event occurring later is the result. (2) The TKG embedding (TKGE) models, which this paper focuses on, evaluate the plausibility of events via scoring events with embeddings of entities and predicates together with timestamps, including TTransE (Leblay and Chekol 2018), HyTE (Dasgupta, Ray, and Talukdar 2018) and DE-SimplE (Goel et al. 2020). TKGE models regard the events that occur at different times are completely independent and these approaches predict the events at the known time.

However, the previous models face several challenges: (1) the existing TKGC models believe that any TKG simply contains events involving time but they ignore the long-term effective commonsense knowledge implied in the TKG. (2)
The evolution-based models struggle to reason about events with weak correlations to past events while the TKGE models are unable to exploit the causality among events. In summary, all the existing TKGC approaches cannot jointly represent the timeliness and causality of events. (3) Almost all the previous TKGC techniques are data-driven without explainability. Besides, StreamLearner (Omran, Wang, and Wang 2019b) is the only known approach that automatically mines temporal rules from TKGs. However, it merely explores the single pattern that all the atoms in the rule body are restricted at the same time but ignores the diverse temporal rule patterns among events.

To address the above challenges, we develop a novel and effective Logic and Commonsense-Guided Embedding (LCGE) model to represent events more adequately for improving the performance of TKGC. Concretely, we design a temporal rule-guided predicate embedding regularization for learning the causality property of events. Furthermore, a joint event and commonsense-based KG embedding strategy is proposed to score each event via learning the time-sensitive representation involving timeliness and causality as well as the time-independent representation in the view of commonsense. The main contributions of our work include:

- We design a temporal rule-guided regularization strategy to inject the causality among events into predicate embeddings. To the best of our knowledge, it is the first effort to introduce temporal rules into TKGE models.
- We model each event from the perspectives of both the time-sensitive representation and the commonsense, facilitating higher accuracy in predicting missing events.
- The experimental results on three benchmark datasets of TKGs illustrate the significant performance improvements of our model compared with several state-of-the-art baselines. More interestingly, our model could provide explainability via temporal rules.

Related Work

Traditional KGE Models

KGE technique aims to score the plausibility of facts via learning the entity and predicate embeddings. TransE (Bordes et al. 2013) models the interaction among a triple fact via regarding a predicate as the translation operation from subject to object. More advanced approaches represent predicates as rotation operations for modeling the symmetric and anti-symmetric predicates, such as RotatE (Sun et al. 2019), QuatE (Zhang et al. 2019) and DualE (Cao et al. 2021). RESCAL (Nickel, Tresp, and Kriegel 2011) conducts three-order tensor decomposition that introduces time information. To the best of our knowledge, it is the first effort to introduce the temporal rules into TKGE models.

Temporal KGE Models

Temporal KGE (TKGE) models extend the traditional KGE approaches by supplementing the time embeddings to represent the time-aware events. TTransE (Leblay and Chekol 2018) adds an extra time embedding in the translation-based score function to represent the timeliness of events. Motivated by the hyper-plane specific to the relation proposed in TransH (Wang et al. 2014), HyTE (Dasgupta, Ray, and Talukdar 2018) projects the entities and predicates into the hyper-plane of a specific time. DE-SimplE (Goel et al. 2020) leverages the diachronic entity embeddings to represent each entity in different timestamps. ATiSE (Xu et al. 2020b) learns the time-aware embeddings of entities and predicates to a Gaussian distribution for representing the time uncertainty. TeRo (Xu et al. 2020a) extends HyTE to learn the time-sensitive entity and predicate embeddings via rotation operation specific to various timestamps (Sun et al. 2019). TComplEx (Lacroix, Obozinski, and Usunier 2020) upgrades ComplEx to score each event via a fourth-order tensor decomposition that introduces time information.

Rule Learning

According to the symbolic characteristics of KGs, logic rules are naturally suitable for KG completion task. The Horn rule is a typical type of logic rule in the form of $a_1 \Leftarrow a_2 \land a_3 \land \cdots \land a_n$, in which $a_1$ denotes an atom in the rule head (namely head atom) and $a_2, \cdots, a_n$ are the atoms in the rule body (namely body atoms). Some effective rule learning algorithms are developed specifically for large-scale KGs relying on rule searching and rule quality evaluation, such as AMIE+ (Galárraga et al. 2015), ScaLeKB (Y.Chen, D.Z.Wang, and S.Goldberg 2016), RuLES (Ho et al. 2018), Anyburl (Meilicke et al. 2019), DRUM (Sadeghian et al. 2019), RLvLR (Omran, Wang, and Wang 2019a) and RNNLogic (Qu et al. 2021). However, these approaches are designed for static KGs rather than TKGs. StreamLearner (Omran, Wang, and Wang 2019b) is the only known algorithm to mine the temporal rules of which all the body atoms are restricted, simultaneously.

The Proposed LCGE Model

In this section, we firstly introduce the preliminaries (§). Then, we declare our developed temporal rule learning algorithm with diverse temporal rule patterns (§) and further propose the temporal rule-guided predicate embedding regularization (RGPR) (§). Afterward, the joint event and commonsense-based KGE mechanism (§) together with the overall optimization objective (§) are presented. The whole framework of our model is shown in Figure 2.

Preliminaries

Temporal Knowledge Graph. The temporal knowledge graph (TKG) is a set of events attached with time information. Each event is represented as a quadruple $(s, p, o, t)$, in which $s$ and $o$ are subject and object, $p$ denotes the predicate, and $t$ implies the timestamp or time interval. Particularly, an event with a time interval $[t_a, t_b]$ can be converted into two events with timestamps namely $(s, p, o, t_a)$ and $(s, p, o, t_b)$. Temporal Rule. A temporal rule is formulated as the conjunction of the atoms attached with time labels. In our work, we focus on the temporal rules in the form of

$$p_{n+1}(x, y, t) \Leftarrow p_1(x, z_1, t_1) \land \cdots \land p_n(z_{n-1}, y, t_n) \quad (1)$$

4570
where $p_i (i = 1, \cdots, n + 1)$ are the predicates, $x$, $y$ and $z_j (j = 1, \cdots, n - 1)$ denote the entity variables. Particularly, $t$ and $t_i (l = 1, \cdots, n)$ indicate the time variables that satisfy the constraint $t_1 \leq t_2 \leq \cdots \leq t_n \leq t$. A temporal rule signifies that the rule head will happen if the rule body holds. A grounding of the temporal rule is obtained by replacing the variables with specific entities and timestamps.

**Temporal Rule Learning**

We develop a novel static-to-dynamic strategy to mine the temporal rules with various patterns. In the static rule learning stage, we first convert all the quadruple events in training set into triples via masking the time information of each event as shown in Figure 2. Then, we obtain a global static KG (GSKG) consisting of all the triples and mine the static rules from the GSKG by any existing rule learning algorithm such as AMIE+ (Galárraga et al. 2015) or Anyburl (Melticke et al. 2019). Notably, temporal rules can be regarded as the extension of static rules with various temporal rule patterns.

**Formulation of Temporal Rule Patterns.** At the dynamic rule learning stage, five temporal rule patterns as shown in Figure 3 are well-designed according to the diverse temporal sequences among atoms as followings:

1. A length-1 rule where the two atoms have different timestamps: $p_2 (x, y, t + t_1) \iff p_1 (x, y, t)$.
2. A length-1 rule where the two atoms are valid at the same time: $p_2 (x, y, t) \iff p_1 (x, y, t)$.
3. A length-2 rule where the timestamps of the three atoms are different from each other: $p_3 (x, y, t + t_1 + t_2) \iff p_1 (x, z, t) \land p_2 (z, y, t + t_1)$.
4. A length-2 rule where the timestamps of the body atoms differ from that of the head atom: $p_3 (x, y, t + t_1) \iff p_1 (x, z, t) \land p_2 (z, y, t)$.
5. A length-2 rule for the three atoms to be valid at the same time: $p_3 (x, y, t) \iff p_1 (x, z, t) \land p_2 (z, y, t)$.

**Quality Evaluation of Temporal Rules.** On account of the static rules mined by the existing rule learning algorithms and the proposed five temporal rule patterns, we extend each static rule to be the corresponding candidate temporal rules according to the defined temporal rule patterns. To evaluate the quality of each candidate temporal rule, we firstly merge the events in the same time window into a sub-graph since some causally related events may occur in the same sub-graph or the adjacent sub-graphs. Then, we search for all the events satisfying the grounding of candidate temporal rules. Specific to a length-1 temporal rule in the general form $p_2 (x, y, t + T_1) \iff p_1 (x, y, t)$, it satisfies pattern (1) if $T_1 > 0$ and signifies pattern (2) if $T_1 = 0$. Thereby, following the definition of the evaluation criteria in previous rule learning models (Sadeghian et al. 2019), we propose support degree (SD), standard confidence (SC), and head coverage (HC) of temporal rules as follows:

$$SD = \frac{\#(a, b) : p_1 (a, b, t) \land p_2 (a, b, t), T_1 = 0}{\#(a, b) : \sum_{t_i \in [1, W_T]} p_1 (a, b, t) \land p_2 (a, b, t + t_i), T_1 > 0}$$

$$SC = \frac{SD}{\frac{\#(a, b) : p_1 (a, b, t)}{SD} \frac{\#(a, b) : p_2 (a, b, t)}{SD}, T_1 > 0}$$

$$HC = \frac{SD}{\frac{\#(a, b) : p_1 (a, b, t) \land p_2 (a, b, t + t_i), T_1 = 0}{SD}, T_1 > 0}$$

in which $W_T$ denotes the size of the pre-defined time window. The evaluation criteria of length-2 temporal rules are defined similarly to that of length-1 temporal rules and are provided in Appendix.
![Diagram](image-url)

Figure 3: A brief diagram of the proposed five temporal rule patterns. The dots indicate the variables. The red lines denote the predicates in the head atoms while the black lines represent the predicates in the body atoms.

We acquire the SC and HC of each candidate temporal rule by traversing the timestamp $t$ in accordance with the temporal rules’ evaluation criteria. The candidate temporal rules satisfying the thresholds of SC and HC are employed for the following regularization strategy in section . The entire algorithm of our proposed temporal rule learning module is presented in Appendix.

### The Designed RGPR Mechanism

Based on the generated temporal rules signifying the causality among atoms, we define a time transfer operator $T$, which ensures all the atoms in a rule are represented at the same time to calculate their correlations. Take the temporal rule $p_3(x, y, t + t_1 + t_2) \iff p_1(x, z, t) \land p_2(z, y, t + t_1)$ for an instance, projecting the body atom $p_1(x, z, t)$ into the same time window of the head atom requires time transfer operation twice. Furthermore, the temporal rule-guided predicate embedding regularization $G$ is proposed to inject the causality implied in the temporal rules into predicate embeddings corresponding to each temporal rule pattern:

1. $p_2(x, y, t + t_1) \iff p_1(x, y, t)$:
   
   $$G = \| (T \circ p_{r1}) - p_{r2} \|$$

2. $p_2(x, y, t) \iff p_1(x, y, t)$:
   
   $$G = \| p_{r1} - p_{r2} \|$$

3. $p_3(x, y, t + t_1 + t_2) \iff p_1(x, e, t) \land p_2(y, e, t + t_1)$:
   
   $$G = \| (T \circ (T \circ p_{r1}) \circ (T \circ p_{r2}) - p_{r3}) \|$$

4. $p_3(x, y, t + t_1) \iff p_1(x, e, t) \land p_2(y, e, t)$:
   
   $$G = \| (T \circ p_{r1}) \circ (T \circ p_{r2}) - p_{r3} \|$$

5. $p_3(x, y, t) \iff p_1(x, e, t) \land p_2(y, e, t)$:
   
   $$G = \| p_{r1} \circ p_{r2} - p_{r3} \|$$

where $p_{r1}$, $p_{r2}$ and $p_{r3}$ indicate the embeddings of predicates $p_1$, $p_2$ and $p_3$, respectively. $\circ$ denotes Hadamard product. Particularly, we theoretically prove that the designed $G$ could model the causality among events in Appendix.

### Joint Event and Commonsense-Based KGE

We take full advantage of the long-term validity of commonsense to evaluate the plausibility of an event accurately since some events that go against commonsense will never happen. Thus, we model each event from the perspective of both time-sensitive and time-independent representations.

To learn the time-sensitive representation of events, inspired by TKGE model TComplEx (Lacroix, Obozinski, and Usunier 2020), we learn the timeliness of each event embedded with the timestamp via four-order tensor decomposition. Besides, the causality among events can be represented via our RGPR mechanism together with the subject and object embeddings. Given an event quadruple $(s, p, o, t)$, the time-sensitive score function is defined as

$$E_1(s, p, o, t) = Re \left( s^T \text{diag}(p_t + p_r) \bar{o} \right)$$

$$= Re \left( \sum_{i=1}^{d} |s_i| \cdot |p \odot t + p_r|_i \cdot |o|_i \right)$$

(10)

where $s \in \mathbb{C}^d$, $p \in \mathbb{C}^d$, $o \in \mathbb{C}^d$ indicate the embeddings in the $d$-dimension complex vector space with regard to $s$, $p$ and $o$, respectively. $\bar{o}$ is the conjugate of $o$. Particularly, $p_t = p \odot t$ denotes the predicate embedding constrained by the timestamp $t$. Meanwhile, $p_r$ is the causality representation of the predicate $p$ learned by our RGPR mechanism. From Eqs. 5-9, if some events hold, the other events that are causally relevant to these events would have higher scores according to the regularization of predicate embeddings. $[x]_i$ denotes the $i$-th value of the complex vector $x$. Based on the developed score function in Eq. 10, we could jointly represent the timeliness and the causality of events, facilitating a more sufficient time-sensitive representation of events.

To learn the time-independent representation of commonsense associated with events, the timestamp in each event is masked to convert the event quadruple $(s, p, o, t)$ into the factual triple $(s, p, o)$. Motivated by some typical commonsense KGs such as ConceptNet (Speer, Chin, and Havasi 2017), commonsense is represented as two concepts linked by a predicate. Therefore, we score each event in the view of commonsense via the learnable concept and predicate embeddings together with the proposed time-independent score function based on commonsense:

$$E_2(s, p, o) = Re \left( \sum_{i=1}^{k} |s_c|_i \cdot |p_c|_i \cdot |o_c|_i \right)$$

(11)

where $s_c \in \mathbb{C}_k$, $p_c \in \mathbb{C}_k$, $o_c \in \mathbb{C}_k$ represent the concept embeddings in the $k$-dimension complex vector space with regard to subject $s$, predicate $p$ and object $o$, respectively. Particularly, $k$ should be set smaller than $d$ to enhance the abstract feature of entity concept embeddings.
Evaluation Protocol. On the consideration that the amount of entities in any TKG is much more than that of predicates, predicting entities is more challenging than predicting predicates. Thus, the temporal KG completion task usually focuses on entity prediction such as predicting entities is more challenging than predicting predicates. To achieve the prediction results, each entity \( e_i \) in the TKG namely the candidate entity is filled in the object position to reconstruct a candidate event quadruple \((s, p, e_i, t)\). Then, the score for evaluating the plausibility of each candidate event is obtained from the time-sensitive and time-independent representations according to Eq. 10 and Eq. 11, which is defined as

\[
E_{pred}(e_i) = E_1(s, p, e_i, t) + \lambda \cdot E_2(s, p, e_i)
\]
Table 3: Temporal KG completion results on three datasets. H1, H3 and H10 represent Hits@1, Hits@3 and Hits@10, respectively. Bold values are the best results and the second-to-best results are underlined in all the models. APG and RPG indicate the absolute performance gains and the relative performance gains achieved by our model compared with the best-performing baseline TeLM. APG and RPG can be calculated by
\[
\text{APG} = \frac{R\text{ours} - R\text{baseline}}{R\text{baseline}},
\]
\[
\text{RPG} = \left(\frac{R\text{ours} - R\text{baseline}}{R\text{baseline}}\right) \times 100\%.
\]

### Baselines
We select two types of state-of-the-art baseline models for comparison:

1. Some typical KGE models without time information, including TransE (Bordes et al. 2013), DistMult (Yang et al. 2015), ComplEx (Trouillon et al. 2016), RotatE (Sun et al. 2019) and QuatE (Zhang et al. 2019).
2. The previous well-performed TKGE models, including TTransE (Leblay and Chekol 2018), HyTE (Dasgupta, Ray, and Talukdar 2018), ATiSE (Xu et al. 2020b), TeRo (Xu et al. 2020a), TComplEx (Lacroix, Obozinski, and Usunier 2020) and TeLM (Xu et al. 2021).

### Metrics and Implementation Details
We rank all the candidate events according to their scores calculated by Eq. 15 in descending order. Then, the rank of the correct event of the i-th test instance is defined as ranki. We employ the commonly-used metrics for evaluating the results of TKGC:

1. The reciprocal mean rank (MRR) of the correct events, which can be calculated by
\[
\text{MRR} = \frac{1}{N} \cdot \sum_{i}^{N} \frac{1}{\text{rank}_i}
\]
where \(N\) is the total amount of test instances.
2. The ratio of the correct events ranked in the top n (Hits@n), which is calculated by
\[
\text{Hits}@n = \frac{1}{N} \cdot \sum_{i}^{N} I(\text{rank}_i \leq n)
\]
where the value of function \(I(\text{rank}_i \leq n)\) is 1 if \(\text{rank}_i \leq n\) is true. \(n\) is usually set as 1, 3 or 10. Note that the higher MRR and Hits@n indicate better performance.

In specific, the score of each event with a time interval in wikidata12k is achieved by averaging the individual scores of two events with endpoint timestamps of the time interval. Besides, the open-source rule learning tool AMIE+ (Galárraga et al. 2015) is utilized to mine the static rules for its convenience and good performance. We conduct all the experiments in Pytorch and on a GeForce GTX 2080Ti GPU. The batch size is set as 1024. The thresholds of SC and HC in our temporal rule learning algorithm are both fixed to 0.1 on all the datasets. We tune all the other hyper-parameters by grid search on the validation sets.

### Cases of Temporal Rules
Some cases of temporal rules mined by our temporal rule learning algorithm from ICEWS05-15 are exhibited in Table 2. We observe that each temporal rule has a confidence level, which corresponds to the weight of the rule-guided predicate embedding regularization derived from this rule. These temporal rules are friendly for humans to understand, which benefits the explainability of the predicted results.

### Experimental Results
The comparison results between our LCGE model with both the traditional KGE models and the existing TKGE models
are provided in Table 3. The results of all the baseline models on ICEWS14 and ICEWS05-15 are directly taken from the original paper of TeLM (Xu et al. 2021), and the experimental results of all baselines on Wikidata12k are obtained by running the source code corresponding to each baseline model. We observe that our proposed LCGE model outperforms all the baselines consistently and significantly on all the datasets. Particularly, compared with the best-performing baseline model TeLM, our approach obtains substantial performance improvements. The difference between TeLM and our model LCGE shown in Table 2 is statistically significant under the paired t-test at the 99% significance level. In terms of the absolute performance gains (APG), our model achieves 30.0%/23.4%/33.6% as to MRR and 37.1%/30.4%/31.6% as to Hits@1 on ICEWS14/ICEWS05-15/Wikidata12k. With respect to the relative performance gains (RPG), our model achieves 48.0%/34.5%/101.2% as to MRR and 68.1%/50.7%/136.8% as to Hits@1 on ICEWS14/ICEWS05-15/Wikidata12k.

In particular, even though TeLM achieves merely 0.545 and 0.599 as to Hits@1 on ICEWS14 and ICEWS05-15, the performances of LCGE are both higher than 0.9 on these datasets. This result illustrates that our model guarantees satisfying accuracy on the challenging TKGC task. Besides, the recent TKGE baselines and our model LCGE all outperform the traditional KGE models, which emphasizes the significance of learning the time-sensitive representation of events for TKGC task. Furthermore, the more superior performance of our model compared with the TKGC baselines demonstrates the effectiveness of evaluating the plausibility of each event from the perspectives of both time-sensitive representation and time-independent commonsense.

### Ablation Study

To verify the effectiveness of each module in our scheme, we observe the performances of two ablated models: (1) eliminating the RGPR module from the whole model (-RGPR), and (2) removing the time-independent score (-TIS) during both training and inference stages. The results of the ablation study as shown in Table 4 demonstrate that eliminating RGPR or TIS both has an obvious impact on the performance of the whole model. Besides, omitting TIS shows more performance drop, which suggests the vital role of predicting events in the view of commonsense. Furthermore, the number of high-quality temporal rules is limited, leading to relatively less performance improvement derived from temporal rules. The result on Wikidata12k is not presented since the number of high-quality temporal rules mined from Wikidata12k is too small to verify the effectiveness of the ablated model -RGPR. Thus, how to mine more temporal rules would be studied in-depth in future research.

### Case Study

It is noteworthy that our approach could employ the symbolic temporal rules to declare the inference process explicitly and enhance the explainability that all the existing TKGE models are lack of. On account of a TKGC query (China, Make a Visit, ?) as shown in Figure 4, our model predicts the entity South Korea with the highest score. More particularly, a temporal rule Host a visit(x, y, t + t) \iff Make a visit(x, y, t) 0.45

Figure 4: A case study of TKGC with explainability via the causality among events with a temporal rule.

### Conclusion

In this paper, we propose a novel and effective logic and commonsense-guided embedding model LCGE for TKGC task. As far as we can be concerned, our model is the first to learn both the time-sensitive representation of events involving timeliness and causality as well as the time-independent commonsense representation, simultaneously. Specifically, the causality among events could be learned via our temporal rule learning algorithm and the temporal rule-guided predicate embedding regularization. The experimental results of the TKGC task on three datasets illustrate the significant performance improvements obtained by our model LCGE compared with the state-of-the-art baselines. More interestingly, our model could provide the explainability of results according to causality inference via temporal rules.
Acknowledgements
This work was partially supported by Zhejiang Science and Technology Plan Project (No. 2022C01082), the National Natural Science Foundation of China (No. 62072022, 61772054) and the Fundamental Research Funds for the Central Universities.

References
Cao, Z.; Xu, Q.; Yang, Z.; Cao, X.; and Huang, Q. 2021. Dual Quaternion Knowledge Graph Embeddings. In Proceedings of the AAAI Conference on Artificial Intelligence, 6894–6902.


