International Journal on Semantic Web and Information Systems
Volume 19 • Issue 1

Active Temporal Knowledge Graph Alignment

Jie Zhou, Laboratory for Big Data and Decision, National University of Defense Technology, China*
Weixin Zeng, Laboratory for Big Data and Decision, National University of Defense Technology, China
Hao Xu, Laboratory for Big Data and Decision, National University of Defense Technology, China
Xiang Zhao, Laboratory for Big Data and Decision, National University of Defense Technology, China
https://orcid.org/0000-0001-6339-0219

ABSTRACT

Entity alignment aims to identify equivalent entity pairs from different knowledge graphs (KGs). Recently, aligning temporal knowledge graphs (TKGs) that contain time information has aroused increasingly more interest, as the time dimension is widely used in real-life applications. The matching between TKGs requires seed entity pairs, which are lacking in practice. Hence, it is of great significance to study TKG alignment under scarce supervision. In this work, the authors formally formulate the problem of TKG alignment with limited labeled data and propose to solve it under the active learning framework. As the core of active learning is to devise query strategies to select the most informative instances to label, the authors propose to make full use of time information and put forward novel time-aware strategies to meet the requirement of weakly supervised temporal entity alignment. Extensive experimental results on multiple real-world datasets show that it is important to study TKG alignment with scarce supervision, and the proposed time-aware strategy is effective.

KEYWORDS

Active Learning, Active Strategies, Entity, Knowledge Graph, Knowledge Management, Temporal Entity Alignment, Temporal Knowledge Graphs, Weakly-Supervised Temporal Entity Alignment

INTRODUCTION

Temporal knowledge graphs (TKGs) store entities, their relationships, and time information with a structured knowledge representation. They are widely used to facilitate downstream tasks in the field of artificial intelligence, such as recommender systems (Schall, 2015) and natural language understanding (Choudhury et al., 2022). A TKG stores knowledge in the form of (e_s, r, e_o, τ), where e_s is the subject entity, e_o is the object entity, r denotes the relation between entities, and τ represents a specific timestamp or a time interval with beginning time and ending time. Typical TKGs

DOI: 10.4018/IJSWIS.318339

*Corresponding Author

include ICEWS14 (García-Durán et al., 2018), DELT (Trivedi et al., 2017), and Wikidata (Erxleben et al., 2014) which contain temporal facts; for example, the triple $\left(Biden, \Pr esidentOf, USA\right)$ is valid only from January 2021.

Since most TKGs are developed independently, existing TKGs are often incomplete but complementary to each other. Thus, several TKG fusion approaches are put forward and attempt to combine several TKGs into a single and comprehensive one. As an important stage of TKG fusion, temporal entity alignment (TEA) is the task of detecting the equivalent entities (i.e., the entities that refer to the same object) from different TKGs.

The task of TEA is inherently challenging from at least the following aspects:

- Usage of Time Information: In TKGs, most of the events have specific time stamps or time intervals, such as $\left(Beijing, host, Olympic\ \ Games, 2008\right)$ and $\left(London, host, Olympic\ \ Games, 2012\right)$. If the observer neglecst the time information and only considers the relation $host$ and the object entity $Olympic\ \ Games$, they may mistakenly match $Beijing$ and $London$. Figure 1 provides a more specific example. Therefore, in TEA, it is critical to make good use of the time information.
- Heterogeneity: Since most different TKGs are constructed individually and obtain source data from various channels, the same entities in different graphs may have different relations and time information with other entities. Different TKGs may cover the different parts of factual events, which makes matching more difficult. For instance, in Figure 2, the two entities $\left\langle Olympia\cos\_F.C.\right\rangle$ and $Olympia\cos F\ \ C$ refer to the same entity in the real world, but it is hard to match them since they are associated with different events.

To address these issues, several approaches have been put forward. Particularly, the time-aware entity alignment approach based on graph neural networks (TEA-GNN) (Xu et al., 2021) first designs a time-aware GNN to cope with TEA, which exploits a time-aware mechanism to introduce the time information into entity embeddings. The time-aware entity alignment using temporal relational attention (TREA) (Xu et al., 2022), on the other hand, incorporates temporal embeddings to enrich the entity embeddings and achieves state-of-the-art performance. Nevertheless, they cannot fully tackle the aforementioned challenges brought by TEA. Besides, they also overlook another notable challenge of TEA:

- Lack of Labeled Data: The performance of the aforementioned TEA models heavily depends on the amount of labeled data (i.e., aligned entity pairs). When the amount of labeled data decreases, the accuracy of alignment drops sharply. Hence, in existing TEA literature, one of the prerequisites is the availability of sufficient aligned entity pairs. Unfortunately, such labeled data are always lacking in the real world. In the meantime, it requires excessive effort to obtain the manual annotation of aligned entity pairs. Thus, it calls for the study of TEA with scarce supervision. Currently, there is no study targeted at the weakly-supervised TEA.

Note. TKG1 is extracted from YAGO, while TKG2 is extracted from Wikipedia. The left entity <Jon Otsemobor> and the right entity “Richie Partridge” have the same relations and object entities. When ignoring the time information of TKGs, these two entities can be matched. Nevertheless, they are inequivalent due to their different time information in each quadruple.

Note. TKG1 is extracted from YAGO, while TKG2 is extracted from Wikipedia. The left entity <Olympiacos_F.C.> and the right entity “Olympiacos F.C.” refer to the same entity in the real world. This entity is a football team that has had many football players over years. Since different TKGs were constructed in different methods, they may cover various parts of factual events, which may make aligning reasoning more difficult.

Figure 1.
An illustrative example of temporal entity alignment

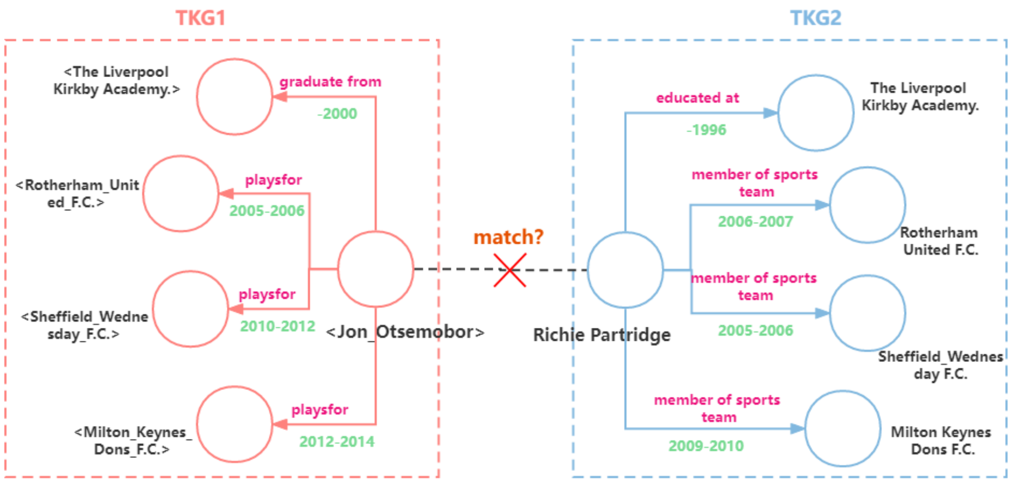
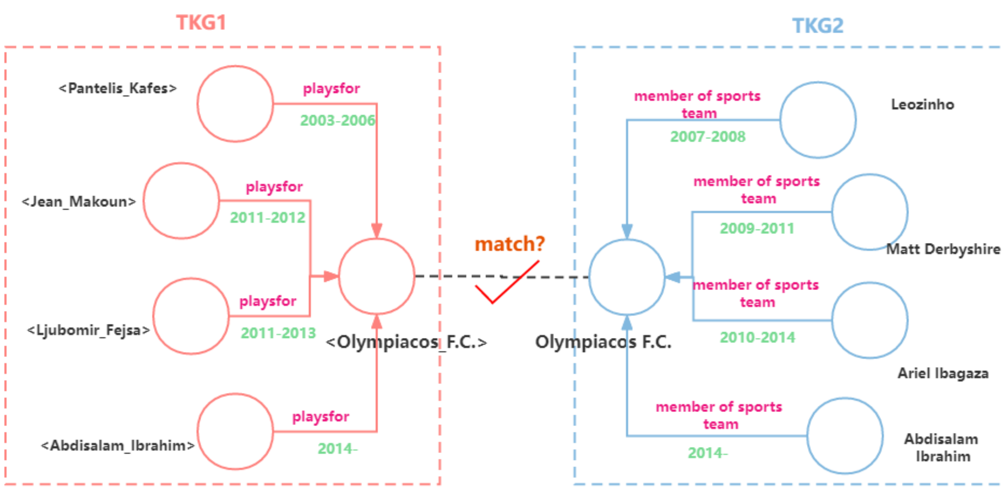


Figure 2.
An illustrative example of an aligned entity pair



In response to the aforementioned challenges, the authors first formally formulate the task of TEA under scarce supervision. Then, the authors propose to utilize active learning to mitigate the issue of limited labeled data. Active learning has been widely used in many weakly-supervised scenarios, such as disaster analysis (Said et al., 2021) and EA between general KGs (Zeng et al., 2021), which is an effective learning algorithm that can interactively query the oracle to label new data points with the desired outputs (Settles, 2009). Although traditional active EA solutions can be adapted to tackle the alignment between TKGs, time information has not been sufficiently utilized in these studies, which is the key element in TKGs. As thus, the authors develop new time-aware strategies incorporating temporal information to achieve better alignment performance under weak supervision signals. The backbone of our proposed model is a simple GNN model that considers structural and temporal information only, where time information is treated as the properties of links between entities.

To verify the proposed approach under scarce supervision, the authors conduct extensive experiments on real-world datasets extracted from the Intergrated Conflict Early Warning System (ICEWS), YAGO3, and Wikidata. The experimental results validate that it is indeed crucial to study weakly-supervised TEA, and time information is beneficial to the design of query strategies for active temporal alignment.

The main contribution of this article can be summarized as follows:

1. The authors are among the first attempts to formally formulate the weakly-supervised TEA problem.
2. The authors adopt the active learning framework to cope with the scarce supervision issue and devise time-aware query strategies to select the most valuable entities for labeling, which can provide useful signals for the subsequent alignment stage.
3. The authors conduct extensive experiments on real-world datasets, and the empirical results demonstrate the significance of studying weakly-supervised temporal alignment as well as the effectiveness of their proposed time-aware strategies.

The rest of this paper is organized as follows: The second section defines the task of weakly-supervised TEA and reviews some related work, the third section describes the proposed approach, including two new time-aware strategies, the fourth section presents the evaluation results and analysis, and the fifth and last section introduces the conclusion and future work.

BACKGROUND

In this section, the authors first formally formulate the problem of weakly-supervised TEA. Then, they introduce some related works.

Problem Formulation

TKGs store real-world information in the form of quadruples $\left(e_s, r, e_o, \tau\right)$. A TKG could be defined as $G=\left(E, R, T, Q\right)$, where E, R, and T represent the sets of entities, relations, and timestamps, respectively, and $Q \subset E \times R \times E \times T$ represents the set of factual quadruples. Although different TKGs are constructed from different sources, there are still many entity pairs referring to the same real-world object which may have many same timestamps. Formally, suppose G1 and G2 are two TKGs, denoted as $G_1 = \left(E_1, R_1, T_1, Q_1\right)$ and $G_2 = \left(E_2, R_2, T_2, Q_2\right)$. Noteworthily, timestamps in most TKGs are presented in Arabic numerals and have similar formats. Hence, timestamps in different TKGs can be easily aligned by manually unifying their formats, and the uniform time set is denoted as $T=T_1 \cup T_2$. $P=\left\{\left(e_{i1}, e_{i2}\right) \mid e_{i1} \in E_1, e_{i2} \in E_2\right\}$, which represents the set of prealigned seed pairs between two TKGs. The objective of TEA is to find new aligned temporal entity pairs based on these prealigned seeds.

The focus of this work is to study TEA with scarce supervision, which can be divided into two stages, namely, selecting and alignment. The former is to select informative entities from a pool of unlabeled entities μ for an oracle to label until the labeling budget B exhausts, which can provide more useful signals for the latter, which aims to exploit the labeled data to align the test entities.

Related Work

Knowledge Graph Representation

KG embedding (KGE) aims to embed entities and relations into a low-dimensional vector space and preserve the original knowledge (Zhao et al., 2020). A representative KGE model is TransE, which projects both entities and relations into the same vector space, assuming that $h+r\approx t$ for a triple $\left(h,r,t\right)$. In addition to traditional KGE models such as TransE and its variants (Nayyeri et al., 2021Lin et al., 2015), there are also some temporal KGE (TKGE) models (Lacroix et al., 2020; Sadeghian et al., 2021; Wu et al., 2020) which show that the combination of traditional embeddings and the time information is beneficial for the representation of TKGs. In the last few years, GNNs have become increasingly popular in many areas due to their ability to model non-Euclidean space. The graph convolutional network (GCN) is one of the most popular extensions of GNN, which can represent the embedding of entities by incorporating structural information of their neighborhoods via convolutional layers. The GCN and its variants have been widely utilized in EA. Some methods (Zeng et al., 2022) improve their capability of dealing with large-scale KG pairs. Some methods (Jin et al., 2020; Li et al., 2021; Zhang et al., 2022) consider the multigranularity information via the addition between internal regularity and external influence.

Temporal Entity Alignment

In the last few years, the task of EA has been widely studied. Most KG alignment methods aim to find equivalent entities across multiple KGs by measuring the similarities between entity embeddings, where the KGs are embedded into a unified vector space. Most existing EA models (Mao et al., 2020; Wu et al., 2019) are devoted to getting a better representation of the entities through KG embedding methods such as TransE and GCN. Recently, by incorporating time information into KGs, the temporal KG draws more attention, which adds to each fact the time validity interval of a specific event. TEA-GNN (Xu et al., 2021) first proposes to use time information to enhance the representation of entities in the task of TEA and successfully improve the accuracy of alignment. TREA (Xu et al., 2022) models the relation and time information of entities to enrich the embeddings of entities, which indeed improves the performance.

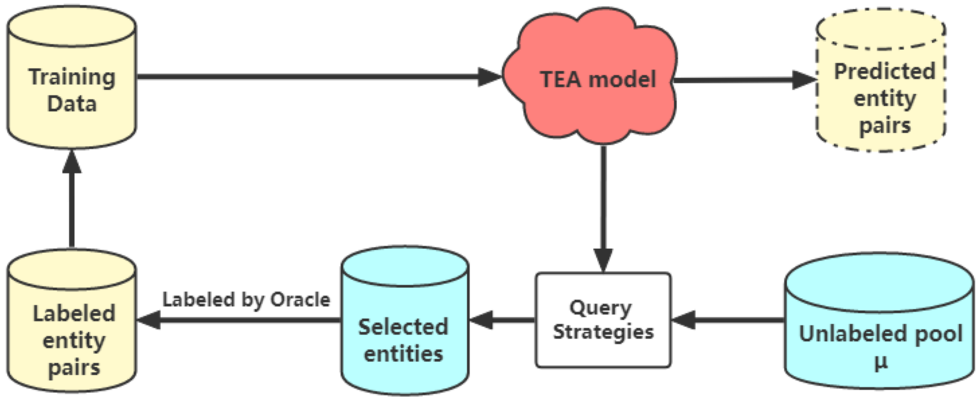
Active Learning

Active learning (also called “query learning” or “optimal experimental design” in the statistics) is a part of machine learning. The core of active learning is that, if the learning algorithm can learn to choose the most informational and effective entities, the model will get better performance with scarce training. Since data labeling costs much human labor, active learning draws more interest than ever before. Recently, some active learning approaches have been proposed and applied to EA, and, indeed, they have verified the effectiveness of active learning in the task of EA. Some works have exploited different active strategies to select informative instances (Berrendorf et al., 2021; Liu et al., 2021), and some (Zeng et al., 2021) have tried to use reinforced learning to obtain the best result from several strategies. Nevertheless, these methods cannot be directly used to tackle TEA, as the authors will verify in the following section.

MAIN FOCUS OF THE ARTICLE

To cope with TEA with scarce supervision, the authors adopt active learning to tackle the labeled data selection problem. The authors first introduce different active learning strategies, including novel time-aware query strategies, which are used to improve the alignment performance under scarce supervision. Then, the authors briefly introduce the backbone representation learning model used to generate the embeddings, which is not the focus of this work.

Figure 3.
The outline of active learning for weakly-supervised temporal entity alignment



Active Learning for Temporal Entity Alignment

Active learning can tackle weakly supervised TEA by selecting the entities it wants to learn from. By designing effective query strategies, the entities with the highest informativeness for labeling can be selected, and the annotated temporal entity pairs are added to the labeled data for subsequent training. These strategies enable active learning techniques to perform significantly better with fewer training data.

Model Overview

Active learning is a general approach that aims at getting as much performance gain as possible by labeling as few data as possible. Therefore, the design of active strategies is crucial to the performance of active learning methods. In this work, the authors mainly rely on pool-based active learning methods where the selected entities are drawn from a large pool of unlabeled data. At first, given the labeling budget B, an initial training set (i.e., also called the seeds) and a large pool of unlabeled data μ. This framework consists of multiple iterations. In each iteration, the authors first exploit specific active strategies to measure the informativeness of each entity in the pool μ, and hand top $b\left(b \ll B\right)$ entities to the oracle for labeling. Next, they add these newly labeled entity pairs to the training set and forwarded to the TEA model. In the specific model, with the help of these labeled entities, the authors could project different TKGs into a unified low-dimension embedding space, where the same entities become closer than the others and the misaligned entities get further and further away. Finally, the authors utilize the distance function and learned representation to conduct the final inference and get the result of alignment. When the budget B is exhausted, they stop the next iteration. Figure 3 shows the overview of this active learning framework.

Next, the authors provide a detailed description of several traditional query strategies, which they exploit in this work, and introduce their proposed strategies that consider the influence of temporal information.

Random

Random sampling is one of the most common sampling methods. Random sampling neglects the specific characteristics of the KGs and can fully explore the diversity of KGs. Assigning the same weight to the entities in the graphs, random sampling selects entities without any prior knowledge.

Degree Centrality

Since the entities in TKGs are not all discrete and also have a close connection with other entities, the authors consider the node with more connection with the others will play a more essential role in the graph and provides more useful information for the model training. Thus, the authors use degree centrality to characterize the centrality of entities, which is also the simplest centrality measure to compute. Formally, the degree centrality orders entities by the number of their directly connected edges. Obviously, the nodes with a higher degree are selected first.

PageRank Centrality

PageRank has been widely used in evaluating the importance of a Web page (Page et al., 1999), which is also adapted to measure the centrality of entities in a graph by considering their degrees as well as the influence of their neighbors. The authors leverage the PageRank centrality to get the most popular nodes in the graph.

Closeness Centrality

Closeness centrality indicates how close a node is to all other nodes in the graph, which is calculated as the average of the shortest path length from the node to every other node in the graph (Perez & Germon, 2016). Concretely, the closeness centrality of a node ***u*** is the reciprocal of the average shortest path distance to *u* overall *n-1* reachable nodes (Freeman, 1978):

$$C\left(u\right) = \frac{n-1}{\sum_{v=1}^{n-1} d\left(v,u\right)} \tag{1}$$

where $d\left(v,u\right)$ is the shortest-path distance between v and u, and $n-1$ is the number of nodes reachable from u.

Betweenness Centrality

Betweenness centrality is a widely used measure which can capture the role of a node in allowing information to pass from one part of the network to the other. It is calculated with the number of the shortest path (between any couple of nodes in the graphs) that passes through the target node (Perez & Germon, 2016). Betweenness centrality of a node v is the sum of the fraction of all-pairs shortest paths that pass through v.

$$c_B\left(v\right) = \sum_{s,t \in V} \frac{\sigma(s,t \mid v)}{\sigma\left(s,t\right)} \tag{2}$$

where V is the set of nodes, $\sigma\left(s,t\right)$ is the number of shortest $\left(s,t\right)$ –pairs, and $\sigma(s,t \mid v)$ is the number of those paths passing through some node v over than $s,t.If$ $s = t, \sigma\left(s,t\right) = 1$, and if $v \in s,t, \sigma(s,t \mid v) = 0$ (Brandes, 2008).

Time-Aware Centrality

Compared to the traditional KGs, TKGs have extra time information, such as timestamps and time intervals. As a time message is a supplement to the triplet information, the authors treat the time message as the attributes of the entities. Hence, time-aware centrality is influenced by the degree of centrality. Time information has different granularities, such as year, month, and day. These multigranularities will have a different influence on different graphs. For instance, in the DICEWS dataset, the time information is accurate to the day, which leads to a large number of time ids. To reflect the difference caused by different timestamps, the authors assign the weight by their degrees. The score function is defined as follows:

$$Score\left(e_i\right) = \sum_{e_j \in N_{e_i}, t \in T} \left(F_t\left(e_j\right) W\left(t\right) + \gamma\right) \tag{3}$$

where N_{e_i} is the set of neighbors of e_i, T is the set of time id, $W\left(t\right)$ means the specified weight of time, $F_t\left(e_j\right)$ denotes the number of facts that is related to time t, and γ is a hypermeter that represents the role of the traditional degree. Then, the top b entities from the unlabeled entities in the training set are selected for oracle to label.

Katz Centrality

Katz centrality (Katz, 1953) computes the relative influence of a node within a network by measuring the number of the immediate neighbors and also other nodes in the network that connect to the node under consideration through these immediate neighbors. The centrality of a node is based on the centrality of its neighbors, which is a generalization of the eigenvector centrality. The Katz centrality for node $\boldsymbol{i}$ is computed as follows:

$$x_i = \alpha \sum_j A_{ij} x_j + \beta \tag{4}$$

where A is the adjacency matrix of the graph with eigenvalues λ, the parameter β controls the initial centrality and $\alpha < \frac{1}{\lambda_{\max}}$.

Time Enhanced Centrality

Considering the nodes in the graph have a close connection with their neighbors, the centrality of their neighbors may have an essential impact on them. After further analysis of the YAGO-WIKI50K dataset, the authors found that more than 98% of triples are soccer teams and soccer players, which seemed like a domain KG in football. Since soccer teams have more connections than any soccer players, if only the nodes of the soccer team have been chosen will provide insufficient signals for models to learn. Hence, to tackle this domain question, the authors exploit the temporal centrality of neighbors, inspired by the Katz centrality, to enhance the time-aware centrality. The temporal centrality for node e_i:

$$\tau\left(e_i\right) = \alpha \sum_{e_j \in N_{e_i}} TA_{ij} e_j + \beta \tag{5}$$

where TA is the temporal adjacency matrix of KG that, if the fact quadruples of ei and ej has time information $TA_{ij} = 1$, else $TA_{ij} = 0$.

The final score function is as follows:

$$Score\left(e_i\right) = \sum_{e_j \in N_{e_i}, t \in T} F_t\left(e_j\right) W\left(t\right) + \mu \cdot \tau\left(e_i\right) \tag{6}$$

where the former is initial time-aware centrality and the last denotes the temporal centrality of neighbors. μ is a hyper-parameter to balance the weight.

Representation Learning

The authors apply their active TEA strategy to a traditional representation learning model, which is a representative widely spread of neural EA models, mainly considering structured information and attribute information in the graph.

In this section, the authors introduce the backbone model that generates the embeddings for alignment. The active learning strategies are exploited over the representation learning models. This framework utilizes a traditional EA model, which is a representative widely spread of neural EA models, mainly considered structured information and attribute information in the graph.

Specifically, the authors adopt the widely used GCN-Align (Wang et al., 2018) model, which can encode information of the neighborhood of a node as a real-valued vector. Given two KGs G_1 and G_2 in different KGs or different languages, a set of aligned entity pairs between them, the proposed active learning approach selects informational entities from the training set based on GCN-based entity embeddings. GCN-Align could embed entities from different KGs into a unified vector space, where equivalent entities are expected to be close than the others. Also, the alignment result is mainly dependent on a predefined distance function through entity embedding.

The inputs to a GCN are feature vectors of nodes and the adjacency matrix of the KG. The goal is to learn a function of features over the input graph and produces a node-level output. In the task of TEA, there are two main assumptions: (1) Equivalent entities in different temporal graphs tend to have similar timestamps or intervals; (2) equivalent entities also play the same important role in the different temporal graphs and have more common neighbors. GCNs can project entities into the unified vector space and let the same entities be close to each other by combining the structure information and time information,

In this work, the authors exploit GCNs to encode both structure information and temporal information and embed the entities into the unified vector space. Let H_s, H_t denotes the structure and temporal feature vector matrices of all entities; the convolutional computation is as follows:

$$\left[H_s^{(l+1)};H_t^{(l+1)}\right] = \sigma\left(D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\left[H_s^{(l)}W_s^{(l)};H_t^{(l)}W_t^{(l)}\right]\right) \tag{7}$$

where $W_s^{(l)}$ and $W_t^{(l)}$ denote the weight matrices for structure features and temporal features in the l-th layer, respectively, $\left[;\right]$ is the concatenation of two matrices, and σ represents the activation function, which is chosen as $\mathrm{Re}LU\left(\cdot\right) = \max\left(0,\cdot\right)$.

EXPERIMENT

In this section, the authors first introduce several existing popular temporal datasets, then they illustrate the proposed new temporal dataset, and finally they experiment settings. Subsequently, the authors empirically conduct experiments to evaluate different active strategies for weakly-supervised TEA on three temporal datasets.

Datasets and Settings

The authors utilize three different temporal datasets, namely, DICEWS, YAGO-WIKI-50K, and YAGO-WIKI-20K, which were extracted from ICEWS (Birch & Muchlinski, 2017), YAGO (Suchanek et al., 2007), and Wikidata (Erxleben et al., 2014), to explore the influence of time information on alignment accuracy. The detailed statics of these temporal datasets are listed in Table 1.

ICEWS05-15 is originally extracted from ICEWS and contains events occurring between 2005 and 2015. It is noteworthy that time information in ICEWS consists in timestamps and specific to a certain day, such as (Barack Obama, Make statement, China, 2014-04-26). DICEWS is built from ICEWS05-15 in a similar way to the construction of DFB datasets (Zhu et al., 2017). The two KGs of DICEWS are all extracted from ICEWS05-15, and their overlap ratio of them is up to 50%.

YAGO-WIKI50K and YAGOWIKI20K are all extracted from Wikidata and YAGO, which contain time information as a time interval with beginning time and ending time, such as (Derek Showers, playsFor, A.F.C. Bournemouth, 1977, 1979) that denotes the fact that Derek Showers plays for a football team called A.F.C. Bournemouth from 1977 to 1979. YAGO-WIKI50K contains about 50,000 entity pairs and both facts in the dataset own time information. Meanwhile, YAGO-WIKI20K is a hybrid dataset with 20,000 entity pairs and contains both temporal and nontemporal facts, where 17.5% of YAGO facts and 36.6% of Wikidata facts are nontemporal.

New Dataset

Through careful observation and analysis, the authors find that each entity in DICEWS has an average degree of more than 32, and the degree of some entities can be up to 400. Since the situation is too clustered and definitely different from the real world, the authors discrete the temporal datasets by decreasing the number of triples and degrees, in the meanwhile, keeping the distribution as much as possible.

Table 1 lists statistics of all datasets. Ent, Rels, Time, Qua, and Aligns represent the number of entities, relations, timestamps, fact quadruples, and aligned entity pairs, respectively, in the datasets. In each KG pair, 70%, 10%, and 20% of the gold pairs are used for testing, validation, and training, respectively. Since the authors study TEA with the limit of aligned pairs, they keep 200 seed entity pars as the initial training set for DICEWS and YAGO-WIKI20K, and 1000 seed entity pairs for YAGO-WIKI50K. Then, with the help of different active strategies, the authors select the entities from the rest of the training set for annotation and add them to the initial training set to generate a new one until the budget runs out.

Implementation Environments

The authors implement their method using GCN-Align, and maintain the initial hyper-parameters on all datasets. The default configuration is as follows: Embedding dimension $k = 100$, learning rate $lr = 20$, margin $\gamma = 3$, and the dropout rate is 0. The authors conduct all experiments on a single GeForce GTX 3090 GPU with 24GB RAM.

Evaluation Metrics

For each entity in the test set, GCN will compute the following distance measure between it and all target entities:

Table 1.
Statistics of temporal entity alignment datasets

Dataset	Ent1	Ent2	Rels1	Rels2	Time	Qua1	Qua2	Aligns
DICEWS	9,517	9,537	247	246	4,017	307,552	307,553	8,556
DICEWS-S	9372	9424	214	212	4010	67,194	67,194	8310
YAGO-WIKI50K	49,629	49,222	11	30	245	221,050	317,814	49,172
YAGO-WIKI20K	19,493	19,929	32	130	405	83,583	142,568	19,462

$$D\left(e_i, v_j\right) = \beta \frac{f\left(h_s\left(e_i\right), h_s\left(v_j\right)\right)}{d_s} + \left(1-\beta\right)\frac{f\left(h_t\left(e_i\right), h_t\left(v_j\right)\right)}{d_t} \tag{8}$$

where $\boldsymbol{f}\left(\boldsymbol{x}, \boldsymbol{y}\right) = \left\| \boldsymbol{x} - \boldsymbol{y} \right\|_1$, $\boldsymbol{h_s}\left(\cdot\right)$ and $\boldsymbol{h_t}\left(\cdot\right)$ denote the structure embedding and temporal embedding of an entity, respectively; ds and dt are dimensionalities of structure embeddings and temporal embeddings; β is a hyper-parameter that balances the importance of two kinds of embeddings. For each entity, the authors rank the target entities ascendingly by their distance and adopt Hits@1 as the primary evaluation metric.

Results and Analysis

To verify whether time information is beneficial for TEA, the authors conduct experiments with or without time information in temporal datasets (Table 2). SE and TE denote the structural embedding and temporal embedding. It can be observed that time information definitely enhances the representation of entities and leads to an improvement in aligned accuracy. Table 2 shows that, compared to initial embedding built only from structural embedding, the temporal embedding obtains an improvement of 14.05% and 10.03% regarding Hits@1 on two TKG datasets, respectively.

Active Learning Study

Since the temporal datasets have different entity pairs, the authors provide different settings of the budget that up to 10% of the whole entity pairs for them. Table 3 compares the TEA results of different active strategies on DICEWS-S datasets with different budgets. The limit of the number of entities in DICEWS results in the instability of the structure or KGs. From the results, most of the strategies perform significantly better than random sampling, only the Katz is worse than it. The proposed $time\quad aware$ strategy is one of the best-performing active strategies. Hence, those centralities which can catch the structure better can achieve better performance within a certain range; for example, closeness centrality can achieve the optimal results when the budget is 400, but, when the budget grows, it becomes backward.

Table 4 shows the results of YAGO-WIKI50K. Noteworthy is that random sampling is better than some traditional centralities such as $Degree$ and $PageRank$. The authors find that YAGO-WIKI50K is more like a domain KG where the relationship is single and most quadruples are soccer players and soccer teams. When the budget is small, random sampling can avoid selecting some nodes in a single type and fully explore the diversity of KGs. To cope with the fact that some types of nodes will own more connections than the others (e.g., soccer teams), the authors incorporate the temporal centrality of the neighbors to enhance the time-aware centrality, and the proposed $time\ enhanced$ outperforms the others.

Sensitivity Analysis

Considering not all facts in the temporal knowledge have time information, the authors conduct a sensitivity analysis on YAGO-WIKI20K, where 17.5% of YAGO facts and 36.6% of Wikidata facts

Table 2.
Comparison of with/without temporal embedding

Dataset	SE	$SE + TE$
DICEWS-1K	29.01	43.06
YAGO-WIKI50K-5K	51.95	61.98

Table 3.
Hits@1 results of GCN under different budgets on DICEWS-S

Strategy	400	600	800
GCN(Random)	18.46	23.33	26.63
+Degree	21.32	26.80	30.70
+PageRank	21.49	26.97	30.39
+Closeness	22.35	26.78	29.48
+Betweenness	21.01	25.87	29.48
+Katz	19.37	22.67	25.31
+Time-aware	21.94	27.02	31.32

Table 4.
Hits@1 results of GCN on YAGO-WIKI50K with different budgets

Strategy	2000	3000	4000	5000
GCN(Random)	53.22	60.94	66.03	70.23
+Degree	51.34	61.67	67.42	71.40
+PageRank	48.85	57.22	66.34	71.19
+Katz	53.67	62.06	66.98	70.02
+Closeness	54.35	60.09	64.09	66.73
+Betweenness	54.32	63.35	68.19	71.55
+Time-aware	51.75	61.78	67.94	71.68
+Time enhanced	55.87	64.09	69.10	72.08

are nontemporal. Table 5 shows the TEA results of different active strategies on YAGO-WIKI20K. Since the relation becomes complex and some exclusive nodes appear, random sampling cannot grasp the effective nodes in the graph, and the accuracy is much lower than other strategies. It can seem that the proposed $time\ \ aware$ and $time\ enhanced$ always achieve the best performance.

On the whole, the authors conclude that the proposed strategies that incorporate time information achieve the optimal or suboptimal performance in TEA between temporal datasets, and traditional node centrality such as $Betweenness$ still effective for active learning in weakly-supervised TEA.

Further Experiment

In this subsection, the authors aim to demonstrate that the proposed active learning framework can be applied to the state-of-the-art temporal graph embedding model to improve their performance under scarce labeled data.

Concretely, the authors apply the active strategies on TEA-GNN (Xu et al., 2021), a state-of-the-art temporal embedding model for TEA. Table 6 shows the TEA results on YAGO-WIKI20K. It can seem that the proposed $time\ \ aware$ and $time\ enhanced$ always achieve the best performance.

Case Study

In order to study the effect of the proposed strategies on other strategies, the authors illustrate some cases that GCN-Align predicts from the test sets of YAGO-WIKI50K (B=3000) and DICEWS-S (B=800) with different strategies. In Table 6, the authors provide an example to show that the proposed

Table 5.
Hits@1 results of GCN on YAGO-WIKI20K with different budgets

Strategy	400	800	1200	1600
GCN(Random)	12.95	20.83	26.97	30.43
+Degree	17.93	32.03	37.12	40.60
+PageRank	17.29	26.31	34.41	38.96
+Katz	17.07	28.28	32.95	36.58
+Closeness	19.95	29.00	33.89	37.22
+Betweenness	17.05	27.79	34.02	38.46
+Time-aware	18.50	32.45	37.88	41.10
+Time enhanced	21.20	32.54	37.50	40.88

Table 6.
Hits@1 results of TEA-GNN (and the active learning-enhanced variants) on YAGO-WIKI20K with different budgets

Strategy	400	800	1200	1600
TEA-GNN (Random)	52.01	54.22	56.19	58.06
+Degree	54.01	57.07	58.62	59.82
+PageRank	51.25	54.70	57.33	58.94
+Katz	54.06	57.08	58.09	59.48
+Closeness	52.86	55.48	58.91	59.53
+Betweenness	52.19	56.33	58.24	59.64
+Time-aware	54.79	57.29	58.82	60.52
+Time enhanced	54.44	57.61	59.74	60.89

Table 7.
An example of different alignment predictions between $Time \quad Aware$ and other strategies

Entities to be aligned	<Jamie_Waite> (in E_1 of YAGO-WIKI50K (B=3000)
Predictions	Query by $time \quad aware$: Jamie Waite; Query by other centrality: Alex Revell
Similar quadruples of aligned entities in Q1 and Q2	(Jamie_Waite, playsfor, Cambridge_United_F.C., [2005, 2006], (Jamie_Waite, playsfor, Rotherham_United_F.C., [2003,2004], (Jamie_Waite, playsfor, Braintree_Town_F.C., [2004, 2004], (Jamie_Waite, playsfor, Kettering_Town_F.C., [2004, 2005], … (in Q1) (Alex Revell, member of sports team, Cambridge United F.C., [2000, 2004]), (Alex Revell, member of sports team, Braintree Town F.C.,[2004, 2006]), (Alex Revell, member of sports team, Kettering Town F.C, [2003, 2003]), (Alex Revell, member of sports team, Rotherham United F.C., [2011, 2015]), … (in Q2)

$time \quad aware$ strategy gives different predictions with consideration of training data, introducing more signals in temporal information. In YAGO-WIKI50K (B=3000) dataset, for the entity $Jamie \ waite$, with the help of the training data which the author queried by other strategies, the first retrieved entity in E2 is $Alex\ \mathrm{Re}vell$. Since the two soccer players have played for four same teams, the model wrongly aligns two different entities, while they have played in the four same teams in different periods of time. Besides, with the training data queried by $timeaware$, the model can give the right alignment, which also considers more time information.

CONCLUSION

Since labeling data costs too much human labor, the authors first utilize active learning and propose time-aware strategies to incorporate temporal information to tackle TEA with scarce supervision. In each iteration, active learning first selects the top b valuable entities by different active strategies to be labeled. Then, with the help of these newly labeled entity pairs, the model can learn to generate more accurate entity representation and achieve better performance. Experimental results on several TKG datasets indicate the effectiveness of time information and the proposed time-aware strategies. Besides, the authors also demonstrate the proposed active learning framework and active strategies can be applied to state-of-the-art TEA models to tackle TEA with scarce supervision.

For future work, the authors will try to integrate other types of information, such as relation information, into the overall framework to further enhance the performance.

ACKNOWLEDGMENT

This work was partially supported by National Key R&D Program of China No. 2020AAA0108800, NSFC under grant Nos. 62272469 and 61872446.

COMPETING INTERESTS

All authors of this article declare there are no competing interests.

REFERENCES

Berrendorf, M., Faerman, E., & Tresp, V. (2021). *Active learning for entity alignment*. Lecture Notes in Computer Science, 12656. doi:10.1007/978-3-030-72113-8_4

Birch, S., & Muchlinski, D. (2017). The dataset of countries at risk of electoral violence. *Terrorism and Political Violence*, *32*(2), 217–236. doi:10.1080/09546553.2017.1364636

Brandes, U. (2008). On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, *30*(2), 136–145. doi:10.1016/j.socnet.2007.11.001

ChoudhuryS. R.RogersA.AugensteinI. (2022). *Machine reading, fast and slow: When do models "understand" language?* https://arxiv.org/abs/2209.07430

Erxleben, F., Günther, M., Krötzsch, M., Mendez, J., & Vrandečić, D. (2014). Introducing wikidata to the linked data web. In *The Semantic Web–ISWC 2014: 13th International Semantic Web Conference*, *Riva del Garda, Italy*, *October 19-23, 2014*. *Proceedings, Part I 13* (pp. 50-65). Springer International Publishing. doi:10.1007/978-3-319-11964-9_4

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239. doi:10.1016/0378-8733(78)90021-7

García-Durán, A., Dumančić, S., & Niepert, M. (2018). Learning sequence encoders for temporal knowledge graph completion. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 4816-4821). doi:10.18653/v1/D18-1516

Jin, W., Qu, M., Jin, X., & Ren, X. (2020). Recurrent event network: Autoregressive structure inferenceover temporal knowledge graphs. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 6669-2283). doi:10.18653/v1/2020.emnlp-main.541

Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, *18*(1), 39–43. doi:10.1007/BF02289026

LacroixT.ObozinskiG.UsunierN. (2020). *Tensor decompositions for temporal knowledge base completion*. https://arxiv.org/abs/2004.04926

Li, Z., Jin, X., Li, W., Guan, S., Guo, J., Shen, H., Wang, Y., & Cheng, X. (2021). 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* (pp. 408-417). doi:10.1145/3404835.3462963

Lin, Y., Liu, Z., Sun, M., Liu, Y., & Zhu, X. (2015). Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the AAAI Conference on Artificial Intelligence* (vol. 29, pp. 2181-2187). doi:10.1609/aaai.v29i1.9491

Liu, B., Scells, H., Zuccon, G., Hua, W., & Zhao, G. (2021). ActiveEA: Active learning for neural entity alignment. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (pp. 3364-3374). doi:10.18653/v1/2021.emnlp-main.270

Mao, X., Wang, W., Xu, H., Wu, Y., & Lan, M. (2020). Relational reflection entity alignment. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (pp. 1095-1104). doi:10.1145/3340531.3412001

Nayyeri, M., Xu, C., Yaghoobzadeh, Y., Vahdati, S., Alam, M. M., Yazdi, H. S., & Lehmann, J. (2021). *Loss-aware pattern inference: A correction on the wrongly claimed limitations of embedding models* (Vol. 12714). Advances in Knowledge Discovery and Data Mining. doi:10.1007/978-3-030-75768-7_7

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999, November 11). *The PageRank citation ranking: Bringing order to the Web.* Stanford InfoLab Publication Server. http://ilpubs.stanford.edu:8090/422/

Perez, C., & Germon, R. (2016). Graph creation and analysis for linking actors: Application to social data. In *Automating open source intelligence* (pp. 103–129). Syngress. doi:10.1016/B978-0-12-802916-9.00007-5

Sadeghian, A., Armandpour, M., Colas, A., & Wang, D. Z. (2021). Chronor: Rotation based temporal knowledge graph embedding. In *Proceedings of the AAAI Conference on Artificial Intelligence* (vol. 35, pp. 6471–6479). doi:10.1609/aaai.v35i7.16802

Said, N., Ahmad, K., Conci, N., & Al-Fuqaha, A. (2021). Active learning for event detection in support of disaster analysis applications. *Signal, Image and Video Processing*, *15*(6), 1081–1088. doi:10.1007/s11760-020-01834-w

Schall, D. (2015). Overview social recommender systems. In Social Network-Based Recommender Systems (vol. 1159, pp.380-386). doi:10.1007/978-3-319-22735-1_1

Settles, B. (2009). *Active learning literature survey*. University of Wisconsin-Madison. https://minds.wisconsin.edu/handle/1793/60660

Suchanek, F. M., Kasneci, G., & Weikum, G. (2007). Yago: A core of semantic knowledge. In *Proceedings of the 16th International Conference on World Wide Web (WWW '07)* (pp. 697–706). Association for Computing Machinery. doi:10.1145/1242572.1242667

Trivedi, R., Dai, H., Wang, Y., & Song, L. (2017). Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In *Proceedings of the 34th International Conference on Machine Learning (ICML'17)* (vol. 70, pp. 3462-3471). Association for Computing Machinery. https://dl.acm.org/doi/10.5555/3305890.3306039

Wang, Z., Lv, Q., Lan, X., & Zhang, Y. (2018). Cross-lingual knowledge graph alignment via graph convolutional networks. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp, 349-357). doi:10.18653/v1/D18-1032

Wu, J., Cao, M., Cheung, J. C., & Hamilton, W. L. (2020). 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). doi:10.18653/v1/2020.emnlp-main.462

Wu, Y., Liu, X., Feng, Y., Wang, Z., Yan, R., & Zhao, D. (2019). Relation-aware entity alignment for heterogeneous knowledge graphs. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence* (pp. 5278-5284). doi:10.24963/ijcai.2019/733

Xu, C., Su, F., & Lehmann, J. (2021). Time-aware graph neural network for entity alignment between temporal knowledge graphs. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (pp. 8999-9010). doi:10.18653/v1/2021.emnlp-main.709

Xu, C., Su, F., Xiong, B., & Lehmann, J. (2022). Time-aware entity alignment using temporal relational attention. In *Proceedings of the ACM Web Conference 2022* (pp. 788-797). doi:10.1145/3485447.3511922

Zeng, W., Zhao, X., Tang, J., & Fan, C. (2021). Reinforced active entity alignment. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management* (pp. 2477-2486). doi:10.1145/3459637.3482472

Zeng, W., Zhao, X., Li, X., Tang, J., & Wang, W. (2022). On entity alignment at scale. *The VLDB Journal*, *31*(5), 1009–1033. doi:10.1007/s00778-021-00703-3

Zhang, T., Li, Z., Wang, J., Qu, J., Yuan, L., Liu, A., Zhao, L., & Chen, Z. (2022). *Aligning internal regularity and external influence of multi-granularity for temporal knowledge graph embedding* (Vol. 13247). Database Systems for Advanced Applications., doi:10.1007/978-3-031-00129-1_10

Zhao, X., Zeng, W., Tang, J., Wang, W., & Suchanek, F. (2020). An experimental study of state-of-the-art entity alignment approaches. IEEE Transactions on Knowledge and Data Engineering, 34. doi:10.1109/TKDE.2020.3018741

Zhu, H., Xie, R., Liu, Z., & Sun, M. (2017). Iterative entity alignment via joint knowledge embeddings. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*. doi:10.24963/ijcai.2017/595

Jie Zhou is a graduate student at the National University of Defense Technology and and his research interests mainly on entity alignment.

Weixin Zeng received the PhD degree from National University of Defense Technology (NUDT), China, in 2022. He is currently working at NUDT and his research interests include on knowledge graph and data mining.

Hao Xu received his PhD from National University of Defense Technology. His main research interests include mobile social network, knowledge graph construction, and information retrieval.

Xiang Zhao is a full professor at National University of Defense Technology, China and his research interests include knowledge graphs, advanced data management, and information retrieval.