Hybrid-TE: Hybrid Translation-based Temporal Knowledge Graph Embedding

Zhihao Wang
Shanghai Key Laboratory of Trustworthy Computing
East China Normal University, Shanghai, China
51184501158@stu.ecnu.edu.cn

Xin Li
Shanghai Key Laboratory of Trustworthy Computing
East China Normal University, Shanghai, China
xinli@sei.ecnu.edu.cn

***Abstract*—Knowledge graph embedding has become a promising method for knowledge graph completion. In this work, we propose Hybrid-TE, a hybrid translation-based temporal knowledge graph embedding, which combines two translation models TransD and HyTE for modeling both temporal and multi-relational facts. Benefiting from two underlying models, Hybrid-TE first builds entity and relation embeddings in separate vector space for modelling multi-relational facts, and then explicitly learns time information by translational embedding on time-specific hyperplanes. We observe that a simple combination of two models does not lead to a satisfactory predictive precision. We therefore propose to project a triplet to all time-specific hyperplanes on which it is temporally valid. Besides, we also explore extra negative relation samplings that differ from positive samplings in relations. We conduct extensive experiments with real datasets on link prediction, relation prediction and temporal scope prediction. Experiments show significant improvements over previous time-insensitive or time-aware models.**

***Index Terms*—knowledge graph embedding, translation model, temporal knowledge graph, knowledge graph completion**

I. Introduction

Knowledge graphs (KGs) such as WordNet ,Freebase and YAGO play an important role in many AI related applications like question answering, information retrieval and recommender systems. A knowledge graph is a directed graph that encodes real-world entities with different types and attributes as nodes and different types of relations on entities as edges. relational fact in KG is usually represented as a triplet (h, r, t), where h and t denote the head and tail entity, respectively, and r models the relationship between entities. Although a typical KG from the real-world contains millions of relational facts, they often suffer from imcompleteness [1]. KG completion aims to predict the most probable missing entities and relations. Since real-world KGs are huge and heterogeneous with containing multi-relational facts, tranditional methods based on symbolic and logical approach are neither scalable nor suitable for KG completion.

Recently, KG embedding becomes a promising method for link prediction. It attempts to learn low-dimensional embeddings in continuous vector space for each entity and relation in KGs, and uses a scoring function defined over the embeddings of entities and relations to measure the plausibility of a triplet. Among the representation learning methods, translation-based embedding gives a good tradeoff between model complexity and the state-of-the-art predictive accuracy. Originated from transE [2], translation-based methods model relations in KGs as translation operations on the embeddings of head and tail entities. These methods embed entities and relations into a vector space denoted by bold letters. It is expected that $\mathbf{h}+\mathbf{r} \approx \mathbf{t}$ holds in the embedding space when (h, r, t) is a valid triplet. TransE is suitable for modeling 1-to-1 relations, but has some flaws when dealing with reflexive, 1-to-N, N-to-1 and N-to-N relations. To fill the gap, a few more translation-based models, including TransH, TransR/CTransR, TransD, and et al., are proposed to further improve transE for effectively modeling various types of multi-relational KGs [3]–[5].

Most existing KG embedding methods have focused on static graphs. The relational facts in static KGs are supposed to be universally true. In fact, there are lots of temporal relational facts in KGs, e.g., $(Einstein, diedIn, Princeton)$ happened in 1955, and $(Obama, presidentOf, USA)$ was true only between 2008 and 2017. Time-aware KG embedding is relatively less studied and not well-understood. Among the learning representation methods of temporal KGs, HyTE and RE-NET are the state-of-the-art TKG embedding methods. HyTE is translation-based and models each timestamp as a hyperplane with translation operations on it. It explicitly learns time effects in the embedding representation of entities and relations. HyTE only handles N-to-1, 1-to-N, N-to-N type of relations casued by time. That is, it assumed that most relations are 1-to-1 at a specific time instance. We note that various types of multi-relational facts would still remain after projecting to time-specific hyperplanes. RE-NET is a novel recurrent event network based architecture that models temporal relations as events over time. It uses RNN as an event encoder to model temporal and multi-relational interactons between entities, and uses so-called neighborhood aggregators to model concurrent interactions within the same timestamp.

This work mainly focuses on translation-based KG embedding that can effectively model temporal and multi-relational KGs. We limit our focus to model temporal relational data that are valid within a time span. That is, a temporal relation is represented in the format of $(h, r, t, [\tau_s, \tau_e])$, where τ_s denotes the starting time and τ_e denotes the ending time, respectively. The heterogeneous time expressions can be represented in this way. For instance, $[\tau, \tau]$ can denote a specific time instance τ and $[\tau, \infty]$ can denote a time span starting since τ. Inspired by previous work, we propose a hybrid translation-based method that combines transD and HyTE. Hybrid-TE builds entity and relation embeddings in separate vector spaces following TransD, considering relations and entities usually have different semantic meanings. Then it models each sampled timestamp as a hyperplane and projects entities and relation to the hyperplane with translation operations over it.

We observed that a simple combination of the two models does not give us a satisfactory predictive performance. We therefore propose to sample and model all starting time involved in temporal triplets as hyperplanes and projects a triplet to all time-specific hyperplanes over which it is temporally valid. That is, given a hyperplane corresponds to the time stamp τ_h, $(h, r, t, [\tau_s, \tau_e])$ will be projected to this hyperplane if $\tau_h \in [\tau_s, \tau_e]$. Previous works like HyTE only project $(h, r, t, [\tau_s, \tau_e])$ to the hyperplane corresponding to τ_s. This simple trick turns out to be very important and significantly improves the predictive performance. Besides, since the learning target of translation models is to minimize the margin-based ranking loss defined over the golden and negtive triplet samplings in KGs, we explore the extra set of negative relation sampling that replaces randomly the relation in golden triplets with other different relations. It turns out to be very useful for an accurate relation prediction.

We conduct extensive experiments on benchmark datasets including YAGO and Wikidata, with performing the tasks of link prediction and temporal scope prediction. Experiments show significant improvements over previous time-insensitive or time-aware translation models. We also compare with RE-NET, the state-of-the-art neural network based embedding method for modeling temporal and multi-relational knolwlege graphs. Experiments show the obvious effectiveness of our method for temporal link prediction.

II. Related Work

We refer to [6] for a detailed survey of the recent KG embedding methods and limit our focus here to related work on time-aware KG embedding methods that are less explored.

There are some notable works on compositional representation learning for KGs. RESCAL is a bilinear model that uses a tensor factorization method for link prediction. Later a holograhpic embeddings HoLE was proposed to learn compositional vectore space models for KGs. HoLE improves the scalability of RESCAL by using correlation as the compositional operators.

There are some recent work on combining temporal information in the translational embedding space. Jiang et al. proposed a time-aware embedding method for translation models. They explored the happening time of relational facts and incorporate the temporal order information of time-sensitive relations in translation models [7], [8]. Trivedi et al. presented Know-Evolve that models the non-linearly evolution of facts as a multivariate point process, for reasoning over dynamic KGs [9]. Leblay and Chekol presented a method for TKG representation learning by using side information with refined scoring functions [10]. As aforementioned, a translation model called HyTE was proposed for TKG completion [11]. We note that many relational facts are beyond 1-to-1 after projecting to time-specific hyperplanes, which also directly motivates this work to model both temporally evolving dynamics and diversity of multi-relational data.

There are recent work on learning embedding representations of TKGs based on neural networks. García-Durán et al. utilized recurrent neural network (RNN) to learn embedding representations and conduct completion tasks of TKGs [12]. Their approach can be combined with exisiting embedding models and they showed that their models TA-TRANSE and TA-DISTMULT are robust and effective to tackle with sparse and heterogeneous temporal expressions. Recently, Jin et al. proposed RE-NET (Recurrent Event Network) for modeling temporal and multi-relational TKGs [13]. They explored a recurrent event encoder to capture interactions between entitities and relations, and used neighborhood aggregators to sumarze concurrent, multi-hop entity interactions within each timestamp. They showed the effectiveness of their model RE-NET by comparison with state-of-the-arts baselines for temporal link prediction including HoLE, TA-TRANSE, TA-DISTMULT and HyTE.

III. Our Method

In this section, we present our translation-based embedding method Hybrid-TE, which combines transD and HyTE for modeling both temporal and multi-relational KGs.

A. Temporal Knowledge Graph

A temporal knowledge graph (TKG) $\mathcal{K} = (\mathcal{E}, \mathcal{R}, \mathcal{D}^+_{\mathsf{time}})$ is a knowledge graph where $\mathcal{E}$ is a set of entities, $\mathcal{R}$ is a set of relations, and $\mathcal{D}^+_{\mathsf{time}}$ is a set of time-aware triplets. Each relational fact (h, r, t) is labeled with a time span $[\tau_s, \tau_e]$ in $\mathcal{D}^+_{\mathsf{time}}$ with $\tau_s \leq \tau_e$ during which the triplet (h, r, t) is temporally valid, where τ_s denotes the starting time and τ_e denotes the ending time, respectively. The time-aware triplet is jointly represented by $(h, r, t, [\tau_s, \tau_e])$. As aforementioned, heterogeneous time expressions can be uniformly represented in this format. We denote by $\mathcal{D}^+$ the set of relational facts that ignore time information in TKGs. That is,

$$\mathcal{D}^+ \stackrel{def}{=} \{(h, r, t) \mid (h, r, t, [\tau_s, \tau_e]) \in \mathcal{D}^+_{\mathsf{time}}\}$$

B. Sampling Time-Specific Hyperplanes

Hybrid-TE first samples a set $\mathcal{T}$ of time stamps, and when training the translation model, it models each time stamp as a hyperplane and projects a triplet to the corresponding hyperplane with translation operation on it. Following HyTE, Hybrid-TE samples the set of time stamps from the start time of each triplet in TKGs. That is,

$$\mathcal{T} \stackrel{def}{=} \{\tau_s \mid (h, r, t, [\tau_s, \tau_e]) \in \mathcal{D}^+_{\mathsf{time}}\}$$

Different from HyTE, given a timestamp $\tau \in \mathcal{T}$ and a triplet $(h, r, t, [\tau_s, \tau_e])$, Hybrid-TE will project the triplet to all time-specific hyperplanes associated with τ if $\tau \in [\tau_s, \tau_e]$ holds. In contrast, HyTE only projects a triplet $(h, r, t, [\tau_s, \tau_e])$ to the hyperplane associated with τ_s in their experiments. The simple

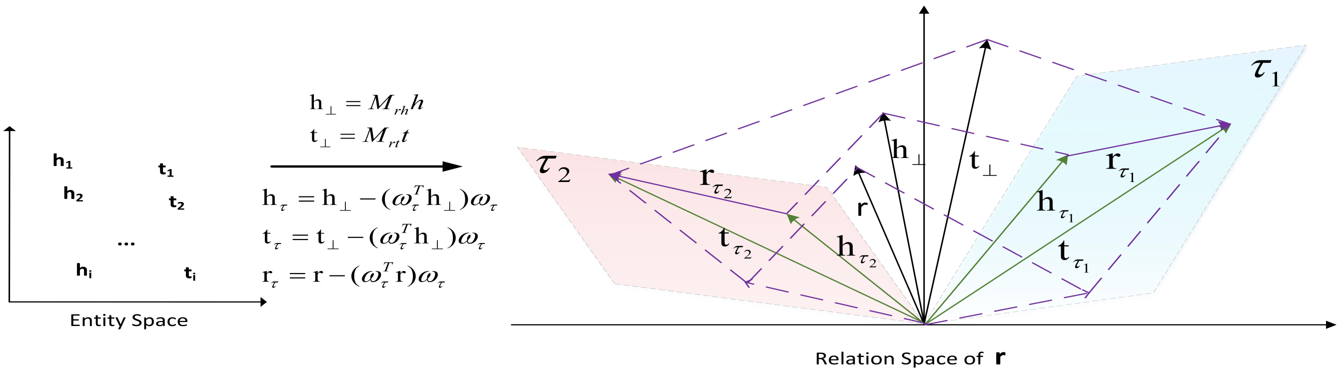

Fig. 1. Simple Illustration of Hybrid-TE

trick we troduced here turns out to be very useful for the predictive prediction for TKGs containing temporal relational facts valid within a time span. We denote by $\mathcal{D}_\tau^+$ the set of golden triplets that are valid at the time stamp τ given as follows, and by $\mathcal{D}_{x,\tau}^-$ the set of negative triplets with respect to some golden triplet x in $\mathcal{D}_\tau^+$.

$$\mathcal{D}_\tau^+ \stackrel{def}{=} \{(h,r,t,\tau) \mid \exists(h,r,t,[\tau_s,\tau_e]) \in \mathcal{D}_{\text{time}}^+ : \tau \in [\tau_s,\tau_e]\}$$

Since KGs only specify golden triplets without negative triplets, one has to choose a proper sampling method for taking the set of negative triplets. We give our sampling methods for $\mathcal{D}_{x,\tau}^-$ in the next subsection.

C. Sampling Negative Triplets

Following tranditional translation models and HyTE, we also explore two general kinds of methods for sampling negative triplets. One is time-agnostic negative sampling by taking all negative triplets that do not appear in $\mathcal{D}^+$, given as follows, for any triplet $x : (h,r,t)$,

$$\begin{aligned}\mathcal{D}_{x,\tau}^- \stackrel{def}{=} & \left\{(h',r,t,\tau) \mid h' \in \mathcal{E}, (h',r,t) \notin \mathcal{D}^+\right\} \\ & \cup \left\{(h,r,t',\tau) \mid t' \in \mathcal{E}, (h',r,t) \notin \mathcal{D}^+\right\} \\ & \cup \left\{(h,r',t,\tau) \mid \ r' \in \mathcal{R}, (h,r',t) \notin \mathcal{D}^+\right\} \end{aligned} \quad (1)$$

Another is time-dependent negative sampling by taking all negative samples that appear in $\mathcal{D}^+$ but not appear in $\mathcal{D}_\tau^+$, given as follows, for any triplet $x : (h,r,t)$,

$$\begin{aligned}\mathcal{D}_{x,\tau}^- \stackrel{def}{=} & \{(h',r,t,\tau) | h' \in \mathcal{E}, \\ & (h',r,t) \in \mathcal{D}^+, (h',r,t,\tau) \notin \mathcal{D}_\tau^+\} \\ & \cup \{(h,r,t',\tau) | t' \in \mathcal{E}, \\ & (h,r,t') \in \mathcal{D}^+, (h,r,t',\tau) \notin \mathcal{D}_\tau^+\} \\ & \cup \{(h,r',t,\tau) \mid r' \in \mathcal{R}, \\ & (h,r',t) \in \mathcal{D}^+, (h,r',t,\tau) \notin \mathcal{D}_\tau^+\} \end{aligned} \quad (2)$$

In the above definitions, the difference with previous approaches is that we further introduce a new set of negative triplets by replacing the relation in the golden triples randomly with other different ones. This simple trick turns out to be very useful for predictive precision on relation prediction.

D. Hybrid-TE

We propose a translation model Hybrid-TE that combines TransD and HyTE. Following TransD, each entity and relation is represented by two vectors. Given a triplet (h,r,t), the two vectors used are $\mathbf{h},\mathbf{h}_p,\mathbf{t},\mathbf{t}_p \in \mathbb{R}^n$ and $\mathbf{r},\mathbf{r}_p \in \mathbb{R}^m$, where the subscript p denotes the projection vectors. For each triplet (h,r,t), two mapping matrices $\mathbf{M}_{rh},\mathbf{M}_{rt} \in \mathbb{R}^{m\times n}$ are used to project entities from entity space to relation space:

$$\begin{aligned}\mathbf{M}_{rh} &= \mathbf{r}_p\mathbf{h}_p^\top + \mathbf{I}^{m\times n} \\ \mathbf{M}_{rt} &= \mathbf{r}_p\mathbf{t}_p^\top + \mathbf{I}^{m\times n}\end{aligned}$$

Let $\mathbf{h}_\perp = \mathbf{M}_{rh}\mathbf{h}$ and $\mathbf{t}_\perp = \mathbf{M}_{rt}\mathbf{t}$. Then following HyTE, we define for each sampled timestamp $\tau \in \mathcal{T}$,

$$\begin{aligned}\mathbf{h}_\tau &= \mathbf{h}_\perp - \left(w_\tau^\top \mathbf{h}_\perp\right) w_\tau \\ \mathbf{t}_\tau &= \mathbf{t}_\perp - \left(w_\tau^\top \mathbf{t}_\perp\right) w_\tau \\ \mathbf{r}_\tau &= \mathbf{r} - \left(w_\tau^\top \mathbf{r}\right) w_\tau\end{aligned}$$

which further projects a triplet to a time-specific hyperplane w_τ in the embedded relation space. We use the following scoring function:

$$f_\tau(h,r,t) = \|\mathbf{h}_\tau + \mathbf{r}_\tau - \mathbf{t}_\tau\|_{l1/l2}$$

The training Objective is given as follows:

$$\mathcal{L} = \sum_{\tau\in\mathcal{T}} \sum_{(x,\tau)\in\mathcal{D}_\tau^+} \sum_{(y,\tau)\in\mathcal{D}_{x,\tau}^-} \max\left(0, f_\tau(x) - f_\tau(y) + \gamma\right)$$

where γ is the margin separating golden triplets and negative triplets. Here the target is to minimize the margin-based ranking loss $\mathcal{L}$ following conventions in translation-based methods, subject to the following constraints: $\|w_\tau\|_2 = 1$ for each sampled timestamp $\tau \in \mathcal{T}$, $\|\mathbf{h}_\perp\|_2 \le 1$, $\|\mathbf{t}_\perp\|_2 \le 1$, $\|\mathbf{h}\|_2 \le 1$, $\|\mathbf{t}\|_2 \le 1$ and $\|\mathbf{r}\|_2 \le 1$ for each embedded entity and relation.

IV. Experiments and Analysis

A. Datasets

For KGs such as Wikidata [14] and YAGO [15], triplets are annotated with temporal meta-facts like *(#factID, occurSince, τ_s)* and *(#factID, occurUntil, τ_e)*, where τ_s and τ_e denotes the starting and ending time of some fact, respectively. There also exit event-based TKGs such as Integrated Crisis Early Warning System (ICEWS18) [16] and Global Database of Events, Language, and Tone (GDELT) [17].

TABLE I
DETAILS OF THE TWO DATASETS USED.

Datasets	Entity	Relation	Train/Valid/Test
Wikidata12K	12,554	24	32.5k/4k/4k
YAGO11K	10,623	10	16.4k/2k/2k

In this work, we are concerned with temporal relational facts uniformly labeled with time intervals. We use the same preprocessed dataset setup of Wikidata and YAGO proposed by Dasgupta et al. for evaluating HyTE on TKG embedding and completion [11]. The two datasets have been preprocessed to handle the issue of sparsity by recursively removing edges containing entity that only occurs once in the graph. Besides, top 10 most frequent temporally rich relations are extracted from the graph, and a subgraph is distilled such that each triplet in it is labeled with a time interval like during which the fact is temporally valid. The statistics of the two datasets is given in Table I. The dataset YAGO11k contains 20.4k triplets, 10,623 entities, and 10 kinds of relations. The dataset Wikidata12k contains 40.5k triplets, 12,554 entities, and 24 kinds of relations. The last column in Table I shows how each dataset is splitted as the training, validating, and testing sets.

B. Link Prediction

The task aims to predict the most probable missing head or tail entity in a test triplet [2], taking into account the time effects. For this task, we use time-agnostic sampling of negative triplets. The given golden triplet (h, r, t) is corrupted by randomly replacing h and t with a different entity such that the resulting corrupted triplet does not reside in the graph.

Evaluation protocol. We follow evaluation metrics proposed in [2]. We first compute scores of those corrupted triplets, and then rank them in an increasing order of their scores, and find the rank of the golden triplet in the order. The task is to calculate the mean rank (MR) over all the test golden triplets. We report results on two metrics including MR and the proportion of correct entities ranked in top 10 (Hits@10) in Table II, for comparing with translation based models. In Table IV, we report results compared with RE-NET. Following [18], we measure the performance with mean reciprocal rank (MRR) of all test queries and the proportion of correct entities ranked in top 1, 3 and 10 (Hits@1, Hits@3, Hits@10), respectively.

C. Relation Prediction

In this task, our purpose is to predict the relationship r between the head and tail entity, for a given golden triplet $(h, ?, t)$. Similar to link prediction, we replace the relation r in the golden triplet (h, r, t) randomly with different relations such that the resulting corrupted triplets do not belong to the knowledge base.

Evaluation protocol. Similar to link prediction, we report the metrics MR/MRR and the portion of correct entities ranked in top 1, 3, or 10 (Hits@1, Hits@3, Hits@10) to evaluate our results. Since the type of relations invovled in YAGO11k and Wikidata12k is few, we use the evaluation metric Hits@1 for comparison on head prediction in Table III. Following the conventions in [2], relation prediction is not included in computing MRR in Table IV when comparing with RE-NET.

D. Temporal Scope Prediction

For TKGs, it is not easy to ensure all of triplets are annotated with temporal meta-facts. It is an important problem to predict time annotations for those time-unknown triplets. The task here is to predict the possible time instance or time interval for the test golden triplet $(h, r, t, ?)$. The training phase is similar to that of link and relation prediction except that time-dependent negative sampling is used. Following HyTE, we only handle time granularity of years for the temporal scope prediction task.

Evaluation protocol. We use the similar evaluation metric proposed by HyTE for temporal scope prediction. We project the test golden triplet $(h, r, t, [\tau_s, \tau_e])$ to all sampled time-specific hyperplanes in the training phase and then compute scores of the triplet on each hyperplane. Recall that we project a triplet to all time-specific hyperplanes on which the triplet is temporally valid during the training phase. We then rank them in an increasing order of their scores, take all ranks on those hyperplanes whose timestamps are within the given time interval $[\tau_s, \tau_e]$, and compute the mean rank of them. In contrast, HyTE only considers the lowest rank on the hyperplanes with their timestamps within the given interval.

E. Results and Analysis

Model parameters. For all the experiments, we take the embedding dimension size from {128, 256}, the value of margin γ among {1, 3, 5, 10}, and the learning rate of SGD in the range of {0.001, 0.0005, 0.0001}. We choose the batch size for each dataset as 20k. The best experimental results are obtained by using the following parameters: the embedding dimension $d = 128$, the margin $\gamma = 10$, and the learning rate is 0.0001. Here, we use the l_1 norm in the scoring function.

We compared our method on predictive accuracy with translation-based models in Table II, III and V, and the neural network based model RE-NET in Table IV, respectively. We experiment with the following evaluation settings:

- *basic model*: our model that combines TransD and HyTE;
- *negative relation sampling*: sampling negative triplets by replacing the relation in golden triplet randomly with different relations;
- *multi-hyperplane projection*: project each triplet to all time-specific hyperplanes on which it is temporally valid.

Link and relation prediction. Table II and III show the prediction accuracy of Hybrid-TE over other time-insensitive and time-aware translation models. Our model outperforms the state-of-the-art time-aware translation model HyTE on both link and relation prediction, and also performs much better

TABLE II

MEAN RANK (LOWER THE BETTER) AND HITS@10 (HIGHER THE BETTER) FOR DIFFERENT METHODS ON LINK PREDICTION. THE BEST RESULTS ARE IN BOLD. OUR PROPOSED METHOD HYBRID-TE OUTPERFORMS OTHER BASELINES AND METHODS.

Dataset	YAGO11k				Wikidata12k			
Metric	MeanRank		Hits@10(%)		Mean Rank		Hits@10(%)	
	tail	head	tail	head	tail	head	tail	head
TransD [5]	138	1208	35.4	13.2	346	562	25.7	14.1
TransH [3]	354	1808	5.8	1.5	423	648	23.7	11.8
HyTE [11]	107	1069	38.4	16.0	179	237	41.6	25.0
Hybrid-TE (basic model)	237	1267	28.5	13.2	404	641	20.2	10.2
Hybrid-TE (basic model with negative relation sampling)	181	1204	33.5	16.7	290	406	42.1	24.3
Hybrid-TE (basic model with multi-hyperplane projection)	116	193	53.4	43.3	202	306	57.1	49.7
Hybrid-TE	**92**	**184**	**59.8**	**47.4**	**89**	**131**	**69.2**	**57.4**

TABLE III

MEAN RANK (LOWER THE BETTER) AND HITS@1 (HIGHER THE BETTER) FOR DIFFERENT METHODS ON RELATION PREDICTION. THE BEST RESULTS ARE IN BOLD. OUR PROPOSED METHOD HYBRID-TE OUTPERFORMS OTHER BASELINES AND METHODS.

Dataset	YAGO11k		Wikidata12k	
Metric	Mean Rank	Hits@1(%)	Mean Rank	Hits@1(%)
TransD [5]	1.19	86.2	1.29	88.2
TransH [3]	1.53	76.1	1.4	88.1
HyTE [11]	1.23	81.2	1.13	92.6
Hybrid-TE (basic model)	1.24	85.3	1.31	89.3
Hybrid-TE (basic model with negative relation sampling)	1.14	89.7	1.08	94.1
Hybrid-TE (basic model with multi-hyperplane projection)	1.19	85.2	1.24	93.1
Hybrid-TE	**1.13**	**90.7**	**1.07**	**95.9**

TABLE IV

MEAN RECIPROCAL RANK (HIGHER THE BETTER) AND HITS (HIGHER THE BETTER) FOR DIFFERENT METHODS FOR LINK PREDICTION. THE BEST RESULTS ARE IN BOLD. OUR PROPOSED METHOD HYBRID-TE OUTPERFORMS OTHER BASELINES.

Dataset	YAGO11k				Wikidata12k			
Metric	MRR	Hits@1(%)	Hits@3(%)	Hits@10(%)	MRR	Hits@1(%)	Hits@3(%)	Hits@10(%)
RE-NET	10.44	5.90	10.98	18.66	36.40	30.31	37.92	47.98
Hybrid-TE	**39.21**	**41.72**	**46.07**	**52.49**	**54.54**	**53.57**	**56.76**	**63.12**

TABLE V

MEAN RANK (LOWER THE BETTER) FOR DIFFERENT METHODS FOR TEMPORAL SCOPE PREDICTION. THE BEST RESULTS ARE IN BOLD. OUR PROPOSED METHOD HYBRID-TE OUTPERFORMS HYTE.

Methods	YAGO11k	Wikidata12k
HyTE	9.88	17.6
Hybrid-TE (basic model)	12.5	21.4
Hybrid-TE	**5.7**	**6.5**

for head and relation prediction on both datasets than HyTE. From the results of experimenting with different evaluation settings, we found that, a simple combination of TransD and HyTE is far from satisfactory on predictive accuracy to handle both temporal and various multi-relational data. In fact, results said that it is often even worse than TransD or HyTE alone by a simple combination of two models.

By experimenting basic model with negative relation sampling, the accuracy of link prediction is slightly improved yet is still sometime worse than other baselines; the accuracy of relation prediction is much improved and becomes better than HyTE. By experimenting basic model with multi-hyperplane projection, the accuracy of link prediction is significantly improved though the accuracy of relation prediction is slightly improved. As shown in the last low of these two tables, experiments on basic models with applying all the above tricks give us the best results.

The results compared with RE-NET is reported in Table IV. Here we compared the accuracy of two models on link prediction using MRR and top hits of correct entities. Since we are concerned with temporal datasets annotated with time intervals, we experimented RE-NET with our two datasets. Results showed a much better predictive accuracy of Hybrid-TE against RE-NET on temporal link prediction.

Temporal scope predication. Table V reported the performance comparison on temporal scope prediction with HyTE. We observed that experiments with only basic model would not give us a better predictive accuracy on this task. By applying basic model with various aforementioned tricks, we obtain a much better accuracy than HyTE.

Example. We give some examples in Table VI to illustrate when our model is more precise than HyTE on link prediction. We show that Hybrid-TE can better handle 1-to-N and N-to-1 muti-relational data than HyTE.

V. CONCLUSION AND FUTURE WORK

In this work, we present Hybrid-TE, a translation-based method for modeling both temporal and multi-relational KGs, which combines two translation models transD and HyTE. To boost predictive accuracy, we further propose to project each triplet to all time-specific hyperplanes on which it is

TABLE VI

EXAMPLES OF 1-TO-N AND N-TO-1 TYPE OF RELATIONS REMAINING AFTER PROJECTING TO TIME-SPECIFIC HYPERPLANES. EACH TIME STAMP ATTACHED TO A TRIPLET WOULD INDICATE A DIFFERENT MISSING HEAD OR TAIL OF IT.

Types of Relation	Testing Triplets	Time Stamps
1-N	John_McCarthy_(computer_scientist), hasWonPrize, **?**	[1971,2017], [1990,2017]
	Igor_Angulo, playsFor, **?**	[2003,2017], [2010,2011], [2003,2017]
N-1	**?**, wasBornIn, Stockholm	[1940,1940], [1935,1935], [1960,1960]
	?, isAffiliatedTo, Independent_politician	[2000-2002], [2009-2017], [1991-1995]

TABLE VII

QUALITATIVE RESULTS ON LINK PREDICTION FOR EXAMPLES IN TABLE VI. THE CANDIDATE MISSING ENTITIES ARE SORTED IN A DESCENDING ORDER OF THEIR SCORES. CORRECT ONE IS IN BOLD.

HyTE	Hybrid-TE
Kyoto_Prize, **Turing_Award**	**Turing_Award**, Kyoto_Prize
IJCAI_Award_for_Research, **National_Medal_of_Science**	**National_Medal_of_Science**, IJCAI_Award_for_Research
Athletic_Bilbao, **Spain_national_20_football_team**	**Spain_national_20_football_team**, Athletic_Bilbao
Apollon_Smyrni_F.C., **CD_Numancia**	**CD_Numancia**, Apollon_Smyrni_F.C.
Gimnàstic_Tarragona, **Spain_national_19_football_team**	**Spain_national_19_football_team**, Gimnàstic_Tarragona
John_W._Brunius, **Princess_Elisabeth**	**Princess_Elisabeth**, John_W._Brunius
Barbro_Oborg, **Bibi_Andersson**	**Bibi_Andersson**, Barbro_Oborg
Per_Teodor_Cleve, **Anders_Ekborg**	**Anders_Ekborg**, Per_Teodor_Cleve
Carole_Keeton_Strayhorn, **Virgil_Goode**	**Virgil_Goode**, Carole_Keeton_Strayhorn
Carole_Keeton_Strayhorn, **Oscar_Goodman**	**Oscar_Goodman**, Carole_Keeton_Strayhorn
Aleksandar_Tomov_(politician), **Sergey_Shoygu**	**Sergey_Shoygu**, Aleksandar_Tomov_(politician)

temporally valid with negative relation samplings. We show through extensive experiments that Hybrid-TE outperforms the state-of-the-art relational embedding models.

ACKNOWLEDGMENT

The authors would like to thank anonymous referees for useful comments. Xin Li is corresponding author. This work is partially supported by Ministry of Science and Technology (Grant No. 2018YFC0830400), Shanghai Pujiang Program (Grant No. 17PJ1402200), and the National Natural Science Foundation of China (Grant No. 61802126).

REFERENCES

[1] D. Chen, R. Socher, C. D. Manning, and A. Y. Ng, "Learning new facts from knowledge bases with neural tensor networks and semantic word vectors," *arXiv preprint arXiv:1301.3618*, 2013.

[2] 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 26*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2013, pp. 2787–2795.

[3] Z. Wang, J. Zhang, J. Feng, and Z. Chen, "Knowledge graph embedding by translating on hyperplanes," in *Twenty-Eighth AAAI conference on artificial intelligence*, 2014.

[4] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, "Learning entity and relation embeddings for knowledge graph completion," in *Twenty-ninth AAAI conference on artificial intelligence*, 2015.

[5] G. Ji, S. He, L. Xu, K. Liu, and J. Zhao, "Knowledge graph embedding via dynamic mapping matrix," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2015, pp. 687–696.

[6] Q. Wang, Z. Mao, B. Wang, and L. Guo, "Knowledge graph embedding: A survey of approaches and applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 12, pp. 2724–2743, 2017.

[7] 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*, 2016, pp. 2350–2354.

[8] T. Jiang, T. Liu, T. Ge, L. Sha, B. Chang, S. Li, and Z. Sui, "Towards time-aware knowledge graph completion," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. Osaka, Japan: The COLING 2016 Organizing Committee, Dec. 2016.

[9] R. Trivedi, H. Dai, Y. Wang, and L. Song, "Know-evolve: Deep temporal reasoning for dynamic knowledge graphs," in *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ser. ICML'17, 2017, pp. 3462–3471.

[10] J. Leblay and M. W. Chekol, "Deriving validity time in knowledge graph," in *Companion Proceedings of the The Web Conference 2018*, ser. WWW '18, 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.

[12] 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 26*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2013, pp. 2787–2795.

[13] C. Esteban, V. Tresp, Y. Yang, S. Baier, and D. Krompaß, "Predicting the co-evolution of event and knowledge graphs," in *2016 19th International Conference on Information Fusion (FUSION)*, 2016, pp. 98–105.

[14] F. Erxleben, M. Günther, M. Krötzsch, J. Mendez, and D. Vrandečić, "Introducing wikidata to the linked data web," in *The Semantic Web – ISWC 2014*, P. Mika, T. Tudorache, A. Bernstein, C. Welty, C. Knoblock, D. Vrandečić, P. Groth, N. Noy, K. Janowicz, and C. Goble, Eds. Cham: Springer International Publishing, 2014, pp. 50–65.

[15] 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*, ser. WWW '07, 2007, pp. 697–706.

[16] E. Boschee, J. Lautenschlager, S. O'Brien, S. Shellman, J. Starz, and M. Ward, "ICEWS Coded Event Data," 2015.

[17] K. Leetaru and P. A. Schrodt, "Gdelt: Global data on events, location, and tone," *ISA Annual Convention*, 2013.

[18] W. Jin, C. Zhang, P. Szekely, and X. Ren, "Recurrent event network for reasoning over temporal knowledge graphs," *CoRR*, vol. abs/1904.05530, 2019.