HyTE: Hyperplane-based Temporally aware Knowledge Graph Embedding

Shib Sankar Dasgupta
Indian Institute of Science
Bangalore, India
s.s.dasgupta.iisc@gmail.com

Swayambhu Nath Ray
Indian Institute of Science
Bangalore, India
swayambhunath93@gmail.com

Partha Talukdar
Indian Institute of Science
Bangalore, India
ppt@iisc.ac.in

Abstract

Knowledge Graph (KG) embedding has emerged as an active area of research resulting in the development of several KG embedding methods. Relational facts in KG often show temporal dynamics, e.g., the fact *(Cristiano_Ronaldo, playsFor, Manchester_United)* is valid only from 2003 to 2009. Most of the existing KG embedding methods ignore this temporal dimension while learning embeddings of the KG elements. In this paper, we propose HyTE, a temporally aware KG embedding method which explicitly incorporates time in the entity-relation space by associating each timestamp with a corresponding hyperplane. HyTE not only performs KG inference using temporal guidance, but also predicts temporal scopes for relational facts with missing time annotations. Through extensive experimentation on temporal datasets extracted from real-world KGs, we demonstrate the effectiveness of our model over both traditional as well as temporal KG embedding methods.

1   Introduction

Knowledge Graphs (KGs) are large multi-relational graphs where nodes correspond to entities, and typed edges represent relationships among them. KGs encode factual beliefs in the form of triple (entity, relation, entity), e.g., *(Brussels, isCapitalOf, Belgium)*. Examples of a few KGs include NELL (Mitchell et al., 2018), YAGO (Suchanek et al., 2007), and Freebase (Bollacker et al., 2008). KGs have been found to be useful for a variety of tasks, viz., Information Retrieval (Kotov and Zhai, 2012; Xiong and Callan, 2015), Question Answering (Dong et al., 2015; Bordes et al., 2015; Yao and Durme, 2014), among others.

KG embedding has emerged as a very active area of research over the last few years, resulting

in the development of several techniques (Bordes et al., 2013; Nickel et al., 2016b; Yang et al., 2014; Lin et al., 2015; Trouillon et al., 2016; Dettmers et al., 2018; Guo et al., 2018). These methods learn high-dimensional vectorial representations for nodes and relations in the KG, while preserving various graph and knowledge constraints.

We note that KG beliefs are not universally true, as they tend to be valid only in a specific time period. For example, *(Bill Clinton, presidentOf, USA)* was true only from 1993 to 2001. KG beliefs with such temporal validity marked are called as temporally scoped. These temporal scopes are increasingly available on several large KGs, e.g., YAGO (Suchanek et al., 2007), Wikidata (Erxleben et al., 2014). The mainstream KG embedding methods ignore the availability or importance of such temporal scopes while learning embeddings of nodes and relations in the KGs. These methods treat the KG as a static graph with the assumption that the beliefs contained in them are universally true. This is clearly inadequate and it is quite conceivable that incorporating temporal scopes during representation learning is likely to yield better KG embeddings. In spite of its importance, temporally aware KG embeddings is a relatively unexplored area. Recently, a KG embedding method which utilizes temporal scopes was proposed in (Jiang et al., 2016). However, instead of directly incorporating time in the learned embeddings, the method proposed in (Jiang et al., 2016) first learns temporal order among relations (e.g., *wasBorn → wonPrize → diedIn*). These relation orders are then incorporated as constraints during the KG embedding stage. Thus, the embedding learned by (Jiang et al., 2016) is not explicitly temporally aware.

In order to overcome this challenge, in this paper, we propose *Hyperplane-based Temporally aware KG Embedding (HyTE)*, a novel KG em-

bedding technique which directly incorporates temporal information in the learned embeddings. HyTE fragments a temporally-scoped input KG into multiple static subgraphs with each subgraph corresponding to a timestamp. HyTE then projects the entities and the relations of each subgraph onto timestamp specific hyperplanes. We learn the hyperplane (normal) vectors and the representation of the KG elements distributed over time jointly. Our contributions are as follows.

- We draw attention to the important but relatively unexplored problem of temporally aware Knowledge Graph (KG) embedding. In particular, we propose HyTE, a temporally aware method for learning Knowledge Graph (KG) embedding.

- In contrast to previous time-sensitive KG embedding methods, HyTE encodes temporal information directly in the learned embeddings. This enables us to predict temporal scopes for previously unscoped KG beliefs.

- Through extensive experiments on multiple real-world datasets, we demonstrate HyTE's effectiveness.

We have made HyTE's source code and datasets used in the paper available at https://github.com/malllabiisc/HyTE

2 Related Work

Temporal fact and event extraction: Time, apart from being an information, also introduces a separate dimension to knowledge. Thus temporal scoping of relational facts is an imperative part of automatic knowledge graph construction and completion. T-YAGO (Wang et al., 2010) extracts temporal facts from semi-structured data like Wikipedia Infoboxes, and categories using only regular expressions. On the other hand, systems like PRAVDA harvests temporal information from free text sources using label propagation. CoTS (Talukdar et al., 2012b) uses integer linear program based approach to model temporal constraints and proposes joint inference framework with few seed examples.

A method for discovering temporal ordering among factual relations was proposed in (Talukdar et al., 2012a). The task of extracting temporally rich events and time expressions and ordering between them is introduced in TempEval

challenge (UzZaman et al., 2013; Verhagen et al., 2010). Various approaches (McDowell et al., 2017; Mirza and Tonelli, 2016) made for solving the task proved to be effective in other temporal reasoning tasks. Although we try to attend to a similar problem, the method proposed in this paper is more related to relational embedding learning paradigm than scoping temporal facts from the web.

Relational embedding learning methods: An enormous amount of research has been done in this field, especially for KG completion or link prediction task (Bordes et al., 2013). (Nickel et al., 2016a) provides a detailed review of the recent KG embedding learning methods. These can be broadly categorized into two different paradigms. TransE(Bordes et al., 2013), TransH(Wang et al., 2014), TransR (Lin et al., 2015), TransD (Ji et al., 2015) are the translational distance-based models. Here the main theme is to minimize the distance between two entity vectors where one of them is translated by a relation vector. The realm of matrix factorization based methods includes bilinear model RESCAL (Nickel et al., 2011), DistMult (Yang et al., 2014), HoIE (Nickel et al., 2016b). Some of the other notable models are Neural Tensor Networks(NTN) (Socher et al., 2013). We also provide some background on the traditional methods in section 3. However, the temporal dimension remains silent in all of these inference methods.

Link prediction through embeddings of the graph nodes and edges, are not only useful for inference over KG but also important for predicting incomplete pieces of the KG itself. Learning temporally steered embeddings is an important but proportionately less explored problem. Only some handful of methods have been proposed for this purpose. t-TransE (Jiang et al., 2016) learns time aware embedding by learning relation ordering jointly with TransE. They try to inflict temporal order on time-sensitive relations e.g. $wasBornIn \rightarrow wonPrize \rightarrow diedIn$. t-TransE does not use the time information directly, whereas we incorporate time directly in our learning algorithm. Another approach Know-Evolve (Trivedi et al., 2017) models the non-linear temporal evolution of KG elements using bilinear embedding learning method. They deploy recurrent neural network to capture non-linear dynamical characteristics of the embeddings. However, they restrict their domain to event-based interac-
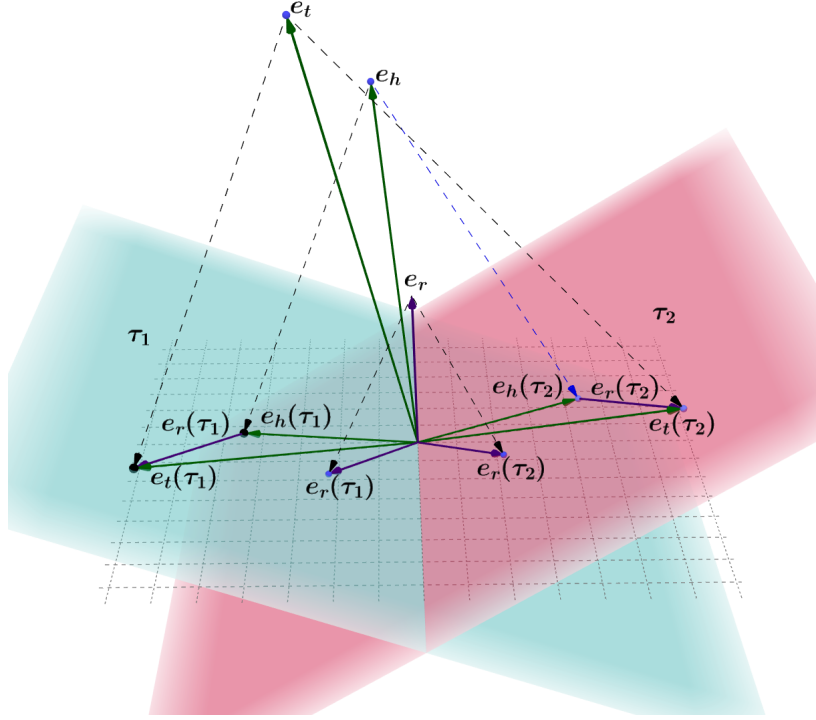
Figure 1: In the figure, the vectors e_h, e_r and e_t correspond to the triple (h, r, t) that is valid at time τ_1 and τ_2. $e_h(\tau_1), e_r(\tau_1)$ and $e_t(\tau_1)$ are the projections of this triple on the hyperplane corresponding to time τ_1 (similarly for time τ_2). Our method HyTE minimizes the translational distance, $\sum_i \|e_t(\tau_i) + e_r(\tau_i) - e_t(\tau_i)\|_1$, in order to learn the temporally aware representations of entities and relations in this triple.

tion type of datasets which are fairly dense in nature. Leblay and Chekol (2018) propose a method for temporal embedding learning using side information from the atemporal part of the graph. However, we use purely temporal KG to learn the temporally aware embedding.

3 Background: KG Embedding

In this section, we provide an overview of the existing methods for knowledge graph representation learning (Bordes et al., 2013), (Wang et al., 2014). Consider a KG $\mathcal{G}$ with a set of entities $\mathcal{E}$. The set of directed edges, $\mathcal{D}^+$ consists of triples (h, r, t), where the edge direction is from h to t and the edge label (also popularly known as relation) is r.

3.1 TransE and TransH

TransE (Bordes et al., 2013) is a simple and efficient translational distance model. It interprets the relation as a translation vector between head and tail entity vectors. Given two entity vectors $e_h, e_t \in \mathbb{R}^n$, it tries to map the relation as a translation vector $e_r \in \mathbb{R}^n$, i.e., $e_h + e_r \approx e_t$ for observed triple (h, r, t). So the distance based scoring func-

tion used for plausible triples is hereby,

$$f(h, r, t) = \|e_h + e_r - e_t\|_{l_1/l_2},$$

where, $\| \cdot \|_{l_1/l_2}$ is the l_1 or l_2-norm of the difference vector. $f(h, r, t)$ will be minimized for observed or correct triples. In order to differentiate between correct and incorrect triples, their TransE score difference is minimized using margin based pairwise ranking loss. More formally, we optimize

$$\sum_{x \in \mathcal{D}^+} \sum_{y \in \mathcal{D}^-} \max(0, f(x) - f(y) + \gamma),$$

with respect to the entity and relation vectors. γ is a margin separating correct and incorrect triples. $\mathcal{D}^+$ is the set of all positive triples, i.e., observed triples in KG. The negative samples are drawn randomly from the set -

$$\mathcal{D}^- = \{(h^{'}, r, t)|h^{'} \in \mathcal{E}, (h^{'}, r, t) \notin \mathcal{D}^+\}$$
$$\cup \{(h, r, t^{'})|t^{'} \in \mathcal{E}, (h, r, t^{'}) \notin \mathcal{D}^+\}.$$

TransE fails to model the many-to-one, one-to-many, many-to-many type of relations as it does not learn a distributed representation of entities when it is involved with many relations. In order

2003

to tackle these situations, TransH was proposed. TransH (Wang et al., 2014) models a relation r as a vector on a relation specific hyperplane and project entities associated with it on that particular hyperplane in order to learn distributed representation of the entities.

We notice that not only the role of the entities changes with time, but also the relationship between them changes. In this paper, we intend to capture this temporal behavior of the entities and relations and try to learn their embeddings accordingly. As discussed above, TransH (Wang et al., 2014) uses relation specific hyperplanes in order to prevent an entity from exhibiting identical characteristic when it is involved with different relations. Taking inspiration from the objective of TransH, we propose a hyperplane based method for learning KG representation distributed in time.

4 Proposed Method: HyTE

In this section, we present a detailed description of HyTE (Figure 1) which not only exploits the relational properties among entities but also uses the temporal meta-data associated with them.

4.1 Temporal Knowledge Graph

Usually knowledge graphs are treated as a static graph consisting of triples in form of (h, r, t). Adding a separate time dimension to the triple makes the KG dynamic. Consider the quadruple $(h, r, t, [\tau_s, \tau_e])$, where τ_s and τ_e denote the start and end time during which the triple (h, r, t) is valid. Unlike (Jiang et al., 2016), we incorporate this time meta-fact directly into our learning algorithm to learn temporal embeddings of the KG elements. Given the timestamps, the graph can be dismantled into several static graphs consisting of triples that are valid in the respective time steps, e.g., knowledge graph G can be expressed as $\mathbb{G} = \mathbb{G}_{\tau_1} \cup \mathbb{G}_{\tau_2} \cup \cdots \cup \mathbb{G}_{\tau_T}$, where τ_i, $i \in 1, 2, \cdots, T$ are the discrete time points.

We constructed this temporal component-graphs $(\mathbb{G}_\tau)$ from the quadruples by considering (h, r, t) to be a positive triple at each time point between τ_s and τ_e. Now, given a quadruple $(h, r, t, [\tau_s, \tau_e])$, we consider it to be a positive triple for each time point between τ_s and τ_e. So, we include (h, r, t) in each $\mathbb{G}_\tau$, where $\tau_s \leq \tau \leq \tau_e$. The set of positive triple corresponding to time τ is denoted as $\mathcal{D}_\tau^+$.

4.2 Projected-Time Translation

TransE considers entity and relation vectors in the same semantic space for a static graph. We observe that time is the main source of different many-to-one, one-to-many or many-to-many relations, e.g., (h, r) pair can be associated with different tail entity t at different points of time. Thus traditional methods fail to disambiguate them directly. In our time guided model, we want the entity to have a distributed representation associated with different time points.

We represent time as a hyperplane i.e., for T number of time steps in the KG, we will have T different hyperplanes represented by normal vectors $w_{t_1}, w_{t_2}, \cdots, w_{t_T}$. Thus, we try to segregate the space into different time zones with the help of the hyperplanes. Now, triples valid at time τ (i.e., the sub graph $\mathbb{G}_\tau$) are projected onto time specific hyperplane w_τ, where their translational distance (TransE Section 3 our case) is minimized. To illustrate, in Figure 1, the triple (h, r, t) is valid for both time frame τ_1 and τ_2. Hence they are projected on hyperplanes corresponding to those times.

Now we compute the projected representation on w_τ as,

$$P_\tau(e_h) = e_h - (w_\tau^\top e_h)w_\tau,$$

$$P_\tau(e_t) = e_t - (w_\tau^\top e_t)w_\tau,$$

$$P_\tau(e_r) = e_r - (w_\tau^\top e_r)w_\tau,$$

where we restrict $\|w_\tau\|_2 = 1$.

We expect that a positive triple, valid at time τ, will have the mapping as $P_\tau(e_h) + P_\tau(e_r) \approx P_\tau(e_t)$. Thus, we use the following scoring function.

$$f_\tau(h, r, t) = \|P_\tau(e_h) + P_\tau(e_r) - P_\tau(e_t)\|_{l_1/l_2}.$$

We learn $\{w_\tau\}_{\tau=1}^T$ for each time stamp τ, along with the entity and relation embeddings. So, by projecting the triple into its time hyperplane, we incorporate temporal knowledge into the relation and entity embeddings, i.e., the same distributed representation will have a different role in different points in time.

Optimization : As mentioned in section 3.1 , we minimize the margin-based ranking loss.

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

where, $\mathcal{D}_\tau^+$ is the set of valid triples with time-stamp τ. The negative samples are drawn from the set of all negative samples, $\mathcal{D}_\tau^-$. We explored two different types of negative sampling:

- **Time agnostic negative sampling(TANS)** considers the set of all the triples that does not belong to the KG, irrespective of timestamps. More formally, for time step τ the negative samples are drawn from the set,

$$\mathcal{D}_\tau^- = \{(h^{'}, r, t, \tau)|h^{'} \in \mathcal{E}, (h^{'}, r, t) \notin \mathcal{D}^+ \} \cup \{(h, r, t^{'}, \tau)| t^{'} \in \mathcal{E}, (h, r, t^{'}) \notin \mathcal{D}^+\}. \quad (1)$$

- **Time dependent negative sampling(TDNS)** emphasizes on time. Along with the time agnostic negative sample, we add extra negative samples which are present in the KG but do not exist in the subgraph for a particular time stamp. Thus we draw negative samples from the set,

$$\mathcal{D}_\tau^- = \{(h^{'}, r, t, \tau)|h^{'} \in \mathcal{E}, (h^{'}, r, t) \in \mathcal{D}^+, (h^{'}, r, t, \tau) \notin \mathcal{D}_\tau^+\} \cup \{(h, r, t^{'}, \tau)| t^{'} \in \mathcal{E}, (h, r, t^{'}) \in \mathcal{D}^+, (h, r, t^{'}, \tau) \notin \mathcal{D}_\tau^+\}. \quad (2)$$

The above mentioned loss $\mathcal{L}$ is minimized subjected to the constraints.

$$\|e_p\|_2 \le 1, \forall\, p \in \mathcal{E}, \quad \|w_\tau\|_2 = 1, \forall\, \tau \in [T]$$

We enforce the first one by adding l_2- regularization of entity vectors with $\mathcal{L}$. We take care of the second constraint by normalizing the time embeddings viz., the hyperplane normal vectors after each update of stochastic gradient descent.

We perform link prediction as well as temporal scoping in order to show the effectiveness of HyTE. For link prediction, we train the model using the optimization procedure as described with time agnostic negative sampling (TANS, Equation 1). The temporal scoping task (Section 5.5) requires the time hyperplanes to be well structured in the embedding space. The time-dependent negative sampling (TDNS Equation 2) is more suitable in case of the temporal scoping problem.

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

Table 1: Details of datasets used. Please see Section 5.1 for details.

5 Experiments

We evaluate our model and compare with different state-of-the-art baselines based on Link prediction(Section 5.3,5.4) and Temporal scoping (Section 5.5). Evaluation metrics used are same as that of the traditional KG embedding method (Bordes et al., 2013) for link prediction task. For temporal scoping task, we present an evaluation criteria as none of the baselines are applicable for this task.

5.1 Datasets

Knowledge Graphs such as Wikidata (Erxleben et al., 2014) and YAGO (Suchanek et al., 2007) have time annotations on a subset of the facts. We extracted the temporally rich subgraph from them for testing our algorithm as well as the baselines.

YAGO11k: In the YAGO3 knowledge graph (Mahdisoltani et al., 2013), some temporally associated facts have meta-facts as *(#factID, occurSince, t_s), (#factID, occurUntil, t_e)*. The total number of time annotated facts containing both occursSince and occursUntil are 722,494. Out of them, we selected top 10 most frequent temporally rich relations. In order to handle sparsity, we recursively remove edges containing entity with only a single mention in the subgraph. This ensures a healthy connectivity within the graph. Finally, we obtain a purely temporal graph of 20.5k triples and 10,623 entities by following this procedure.

Wikidata12k: We extracted this temporal Knowledge Graph from a preprocessed dataset of Wikidata proposed by (Leblay and Chekol, 2018)[1]. We followed a similar procedure as described in YAGO11k. Here also, we distill out the subgraph with time mentions for both start and end. We ensure that no entity has only a single edge connected to it. We select top 24 frequent temporally rich relations for this case, which resulted in 40k triples with 12.5k entities. The dataset is almost double in size with respect to YAGO11k.

---

[1]https://staff.aist.go.jp/julien.leblay/datasets/

5.2 Method Compared

For evaluating the performance of our algorithm, we compare against the following methods:

- **t-TransE** (Jiang et al., 2016): This method uses a temporal ordering of relations to model knowledge evolution in the temporal dimension. They regularize the traditional embedding score function with observed relation ordering with respect to head entities.

- **HolE** (Nickel et al., 2016b): We consider this method as a representative of the state-of-the-art in non-temporal KG representation learning.

- **TransE** (Bordes et al., 2013): This is a simple but effective translation based model. We build HyTE on top of TransE and demonstrate the gains over this method.

- **TransH** (Wang et al., 2014): This method models each relation as different hyperplanes on which the translation operations are carried out. Our proposed method, HyTE also modifies TransE in a similar fashion by treating the timestamps as hyperplanes.

- **HyTE** : Our proposed method. Please see Section 4 for more details.

5.3 Entity Prediction

The task here is to predict the missing entity, given an incomplete relational fact with its time. We experimented with both YAGO11K and Wikidata12k dataset. Training is done in perspective of both head and tail prediction. More formally, for generation of negative sample from a correct triple (h, r, t, τ), we split them in two parts - $(h, r, ?, \tau)$ (for tail entity prediction) and $(?, r, t, \tau)$ (for head entity prediction). In this task, we follow the TANS (Equation 1) procedure for generating negative samples, i.e., for each of tail and head query terms we randomly replace an entity such that newly generated triple is not observed in the graph, for eg, we sample t' such that $t' \in \mathscr{E} \setminus t$ and $(h, r, t', \tau) \notin D_\tau^+$.

Ranking Protocol: For a test triple (h, r, t, τ), we generate corrupted triples by replacing tail entity (for tail prediction) or head entity (for head prediction) with all possible entities. Filtered protocol, proposed by (Bordes et al., 2013), says that the corrupted triples must not be a part of the graph

itself. To illustrate, given a test triple (h, r, t) for tail prediction task, we compute scores for the candidate set $C(h, r) = \{(h, r, t' : \forall t' \in \mathscr{E})\} \setminus (Train \cup Test \cup Valid) \cup (h, r, t)$. We rank all the triples in $C(h, r)$ in the increasing order of their score and find the rank of the actual triple (h, r, t). We report the mean rank over all the test queries (MR) and proportion of correct entities in top 10 rank (Hits@10).

5.4 Relation Prediction

The aim of this task is to predict the relation between two entities, i.e, for a given time-stamped triple with missing relation $(h, ?, t, \tau)$, we predict the relation r. For evaluation, we corrupt the triples with all possible relations and report the rank of the actual relation. We report Hits@1 for this task as the number of relations are quite less in numbers for both the datasets, 10 and 24 for YAGO11k and Wikidata12k respectively. Please note that we do not train our model separately for this task, rather we report the values obtained by the exact same model used for head and tail entity prediction.

The main motivation for this task is to deal with the relational conflict between two entities at a particular time-scope. For example, given the year 1992, a person 'X' and a city 'Y', one would like to know if he/she was $bornIn$ or $diedIn$ that city in that year. Through the explicit use of temporal information during training, we find that our method HyTE outperforms the baseline methods in both the datasets (as shown in 6.1)

5.5 Temporal Scope Prediction

Given the scarcity of time annotations of the KG facts, predicting time for atemporal part of the KG is an important problem. Unlike the previous baseline methods, our model can predict the time scope for a given triple. In order to perform better in this task, we want the hyperplanes to be well separated even after maintaining consistency with the positive triples. To incorporate this nature during training, we use the time-dependent negative sampling technique (TDNS Equation 2). Rest of the training procedure remains the same as the link prediction task. The model is trained on the same train split used for link prediction. In this task, we predict the time interval or the time instance τ for a given test triple $(h, r, t, ?)$. We project the relation and the entities of the triple on all the time hyperplanes and check the plausibility of that test triple on each

Dataset	YAGO11K				Wikidata12K			
Metric	Mean Rank		Hits@10(%)		Mean Rank		Hits@10(%)	
	tail	head	tail	head	tail	head	tail	head
Trans-E (Bordes et al., 2013)	504	2020	4.4	1.2	520	740	11.0	6.0
TransH (Wang et al., 2014)	354	1808	5.8	1.5	423	648	23.7	11.8
HolE (Nickel et al., 2016b)	1828	1953	29.4	13.7	734	808	25.0	12.3
t-TransE (Jiang et al., 2016)	292	1692	6.2	1.3	283	413	24.5	14.5
HyTE	**107**	**1069**	**38.4**	**16.0**	**179**	**237**	**41.6**	**25.0**

Table 2: Mean Rank (lower the better) and Hits@10(higher the better) for different methods for Entity Prediction task. Proposed method HyTE outperforms all the traditional approaches. The better performance of HyTE over t-TransE can be attributed to the fact that it incorporates time directly. Please see Section 6.1 for details.

Dataset	YAGO11K		Wikidata12K	
Metric	Mean Rank	Hits@1(%)	Mean Rank	Hits@1(%)
Trans-E (Bordes et al., 2013)	1.7	78.4	1.35	88.4
TransH (Wang et al., 2014)	1.53	76.1	1.4	88.1
HolE (Nickel et al., 2016b)	2.57	69.3	2.23	83.96
t-TransE (Jiang et al., 2016)	1.66	75.5	1.97	74.2
HyTE	**1.23**	**81.2**	**1.13**	**92.6**

Table 3: Mean Rank (lower the better) and Hits@1 (higher the better) for different methods for Relation Prediction Task. Proposed method HyTE outperforms all the traditional approaches. Even though t-TransE imposes implicit relation ordering, HyTE beats it with a high margin. Please see Section 6.1 for details.

Negative Sampling	YAGO11K	Wikidata12k
TANS (Equation 1)	14.0	29.3
TDNS (Equation 2)	**9.88**	**17.6**

Table 4: The predicted Mean Rank (lower the better) for temporal Scoping. The number of classes are 61 and 78 for YAGO11K and Wiki-data12k respectively. The results depict the effectiveness of TDNS. Please see Section 6.2

of them. For evaluation, we order the time frames in increasing order of their plausibility score for that particular triple. Now, we select the rank of the time(τ) associated with the test triple. If the associated time is an interval, we consider the lowest rank among the times in between the interval.

6   Results

Implementation details: For all the methods, we have kept batch size b = 50k on both the datasets. The dimensions of the embeddings (d) are varied in the range $\{64, 128, 256\}$. The margins(η) for all the methods are chosen from the set $\{1, 2, 5, 10\}$. Learning rate used for SGD, $lr \in \{0.01, 0.001, 0.0001\}$.

The best configuration is chosen by corresponding lowest MR on the validation set. For both YAGO11k and Wikidata12k, we obtained $d = 128$, $\eta = 10$, $lr = 0.0001$ using ℓ_1-norm in the scoring function.

Both YAGO11k and Wikidata12k contain time annotations to the granularity of days. For the temporal scoping task, we only deal with year level granularity by dropping the month and date information. Timestamps are then treated as 61 and 78 different intervals for YAGO and Wikidata respectively. The main motive behind having time classes is to distribute the time annotations in the KG uniformly. For example, less frequent year mentions are clubbed into same time class but years with high frequency forms individual classes. We take care of the unbalance that may occur in terms of number of triples in a particular interval by applying a minimum threshold of 300 triples per interval during construction. To illustrate, in Wikidata there are classes like 1596-1777, 1791-1815 with a large span as the events occurring on those points of time are quite less in KG. The years like 2013, 2014 being highly frequent are self-contained.

6.1   Performance analysis & comparison

The obtained results for different tasks are based on the above mentioned hyperparameters.

Link Prediction: The results reported in Table 2 demonstrate the efficacy of HyTE. We observe that our model outperforms the traditional state-of-the-art link prediction model HolE (Nickel et al., 2016b) by a significant margin in both the datasets. We also show a large boost in performance over TransE (Bordes et al., 2013). This significant gain empirically validates our claim that including temporal information in a principled

Test quadruples	TransE	HyTE
Gordon_Carroll, ?, Baltimore,[1928, 1928]	diedIn, **wasBornIn**	**wasBornIn**,diedIn
S.Laubenthal, ?, Washington.,[2002, 2002]	wasBornIn,**diedIn**	**diedIn**, wasBornIn
Eugene_Sander, ?, Cornell_Univ.,[1959, 1965]	worksAt,**graduatedFrom**	**graduatedFrom**, isAffiliatedTo
Ernesto_Maceda, ?, Nacionalist_Party, [1971,1987]	isMarriedTo, diedIn	**isAffiliatedTo**,diedIn

Table 5: Example of Qualitative Results on relation prediction. The order of prediction is in descending order. Correct one is in **bold**. Please refer Section 6.2 for details.

fashion helps to learn richer embeddings of the KG elements. We notice that HolE is performing significantly poor in terms of MR but it exceeds the other baselines in Hits@10 by a large margin.

Again, in comparison with the temporal model t-TransE (Jiang et al., 2016), HyTE proves to be effective. t-TransE performs better than TransE and HolE due to its implicit time incorporation through relation ordering. HyTE with its direct inclusion of time in the relation-entity semantic space outperforms all of them.

Relation prediction: Again, in this scenario, we show improvement over baselines. We hypothesize that time scope information helps to disambiguate among relations e.g. traditional methods like TransE or HolE will confuse between relations like wasBornIN, diedIn. Where time information surely helps to resolve that conflict. From Table 3, we validate this claim. In Section 6.2, we also demonstrate some qualitative result in favor of our assertion.

Temporal scoping of facts: We report the rank of correct time instance of the triple. If the triple scope is an interval of time, we consider the lowest rank that corresponds to the time within that interval. The ranks are reported in table 4 for both the datasets. The baseline model t-TransE (Jiang et al., 2016) is not applicable here as it does not use the time meta-facts directly. We also observe that the HyTE hyperplanes form a sequential map in the space. We discuss it with details in subsection 6.2.

6.2 Qualitative results

Table 5 contains some of the qualitative analyses for the relation prediction task. We mention some of the cases where transE is confusing between the temporal relations like *wasBornIn* and *diedIn*. Consider the second example in Table 5, where transE wrongly predicts *wasBornIn* . HyTE predicts *diedIn* as it has a prior knowledge that S.Laubenthal "was born in 1943", "created Excalibur on 1973" from the training data. As the query year is 2002, our method comes to such
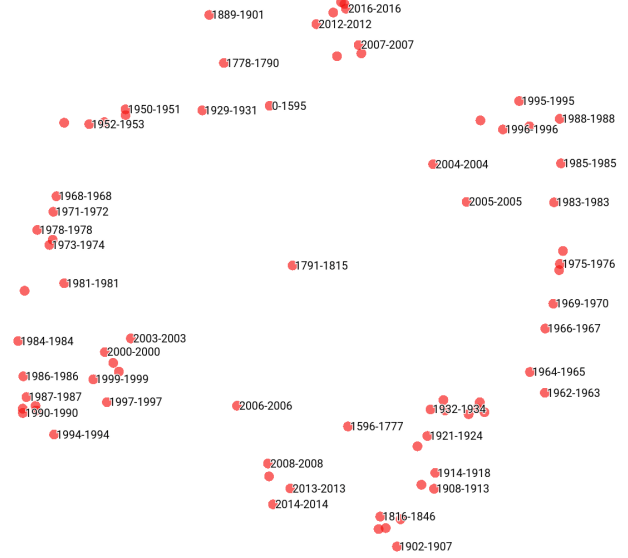

Figure 2: The figure illustrates 2-d PCA projection of the 128 dimensional time embeddings which are obtained after training HyTE for temporal scoping task. We observe that the trained time representations are forming natural clusters and ordering. Please ref Section 6.2

a conclusion through its relative temporal ordering. We see many examples of this kind, where HyTE is naturally learning some relation ordering in parallel with the temporal direction. We observe many type inconsistency in relation prediction for our model, e.g., for this fact *(Lauren Miller, wasBornIn, Lakeland,Florida,[1982,1982])*, our model predicts *isMarriedTo*. This can be attributed to the fact that we do not impose any type related constraints in our model. We will look forward to incorporating type and temporal constraints within our model as future work.

In Figure 2, we show a 2-d PCA projection of the 128-dimensional normal vectors of the hyperplanes. These vectors are trained for the temporal scoping task with time-dependent negative sampling (Equation 2). This figure demonstrates the ability of HyTE to structure the hyperplanes in the entity-relation space according to the data. Also, note that we do not regularize the model with any ordering constraint, but it learns the temporal ordering as well as the clustering from the

data itself. We hypothesize that this phenomenon emerges due to TDNS (Equation 2). However, in case of link prediction, we notice that the extra samples are affecting performance as they originate from the KG itself.

7 Conclusion

We propose HyTE, a hyperplane-based method for learning temporally aware knowledge graph embeddings. Our method exploits temporally scoped facts of KG to perform link prediction as well as prediction of time scopes for unannotated temporal facts. Through extensive experiments on real-world datasets, we demonstrate effectiveness of HyTE over both traditional and time aware embedding methods. In future, we would like to incorporate type consistency information to further improve our model and also integrate HyTE with open-world knowledge graph completion (Shi and Weninger, 2018). We are hopeful that our proposed temporal representation learning algorithm will motivate further research on temporal KG embedding learning.

Acknowledgments

References

Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: A collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, SIGMOD '08, pages 1247–1250, New York, NY, USA. ACM.

Antoine Bordes, Nicolas Usunier, Sumit Chopra, and Jason Weston. 2015. Large-scale simple question answering with memory networks. *CoRR*, abs/1506.02075.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 2787–2795. Curran Associates, Inc.

Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2d knowledge graph embeddings. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, Louisiana, USA, February 2-7, 2018*.

Li Dong, Furu Wei, Ming Zhou, and Ke Xu. 2015. Question answering over freebase with multi-column convolutional neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers*, pages 260–269.

Fredo Erxleben, Michael Günther, Markus Krötzsch, Julian Mendez, and Denny Vrandeǎić. 2014. Introducing wikidata to the linked data web. In *Proceedings of the 13th International Semantic Web Conference - Part I*, ISWC '14, pages 50–65, New York, NY, USA. Springer-Verlag New York, Inc.

Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. 2018. Knowledge graph embedding with iterative guidance from soft rules. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, Louisiana, USA, February 2-7, 2018*.

Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. 2015. 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)*, volume 1, pages 687–696.

Tingsong Jiang, Tianyu Liu, Tao Ge, Lei Sha, Sujian Li, Baobao Chang, and Zhifang Sui. 2016. Encoding temporal information for time-aware link prediction. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2350–2354. Association for Computational Linguistics.

Alexander Kotov and ChengXiang Zhai. 2012. Tapping into knowledge base for concept feedback: Leveraging conceptnet to improve search results for difficult queries. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM '12, pages 403–412, New York, NY, USA. ACM.

Julien Leblay and Melisachew Wudage Chekol. 2018. Deriving validity time in knowledge graph. In *Companion Proceedings of the The Web Conference 2018*, WWW '18, pages 1771–1776, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI'15, pages 2181–2187. AAAI Press.

Farzaneh Mahdisoltani, Joanna Biega, and Fabian M Suchanek. 2013. Yago3: A knowledge base from multilingual wikipedias. In *CIDR*.

Bill McDowell, Nathanael Chambers, Alexander Ororbia II, and David Reitter. 2017. Event ordering with a generalized model for sieve prediction ranking. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 843–853, Taipei, Taiwan. Asian Federation of Natural Language Processing.

Paramita Mirza and Sara Tonelli. 2016. Catena: Causal and temporal relation extraction from natural language texts. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 64–75. The COLING 2016 Organizing Committee.

T. Mitchell, W. Cohen, E. Hruschka, P. Talukdar, B. Yang, J. Betteridge, A. Carlson, B. Dalvi, M. Gardner, B. Kisiel, J. Krishnamurthy, N. Lao, K. Mazaitis, T. Mohamed, N. Nakashole, E. Platanios, A. Ritter, M. Samadi, B. Settles, R. Wang, D. Wijaya, A. Gupta, X. Chen, A. Saparov, M. Greaves, and J. Welling. 2018. Never-ending learning. *Communications of the ACM*, 61(5):103–115.

Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. 2016a. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104(1):11–33.

Maximilian Nickel, Lorenzo Rosasco, and Tomaso Poggio. 2016b. Holographic embeddings of knowledge graphs. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, pages 1955–1961. AAAI Press.

Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. 2011. A three-way model for collective learning on multi-relational data. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ICML'11, pages 809–816, USA. Omnipress.

Baoxu Shi and Tim Weninger. 2018. Open-world knowledge graph completion. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, Louisiana, USA, February 2-7, 2018*.

Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. 2013. Reasoning with neural tensor networks for knowledge base completion. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 926–934. Curran Associates, Inc.

Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. Yago: A Core of Semantic Knowledge. In *16th International Conference on the World Wide Web*, pages 697–706.

Partha Pratim Talukdar, Derry Wijaya, and Tom Mitchell. 2012a. Acquiring temporal constraints between relations. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 992–1001. ACM.

Partha Pratim Talukdar, Derry Wijaya, and Tom Mitchell. 2012b. Coupled temporal scoping of relational facts. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM '12, pages 73–82, New York, NY, USA. ACM.

Rakshit Trivedi, Hanjun Dai, Yichen Wang, and Le Song. 2017. Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In *Proceedings of the 34th International Conference on Machine Learning*.

Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *International Conference on Machine Learning (ICML)*, volume 48, pages 2071–2080.

Naushad UzZaman, Hector Llorens, Leon Derczynski, James Allen, Marc Verhagen, and James Pustejovsky. 2013. Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations. In *Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, volume 2, pages 1–9.

Marc Verhagen, Roser Sauri, Tommaso Caselli, and James Pustejovsky. 2010. Semeval-2010 task 13: Tempeval-2. In *Proceedings of the 5th international workshop on semantic evaluation*, pages 57–62. Association for Computational Linguistics.

Yafang Wang, Mingjie Zhu, Lizhen Qu, Marc Spaniol, and Gerhard Weikum. 2010. Timely yago: Harvesting, querying, and visualizing temporal knowledge from wikipedia. In *Proceedings of the 13th International Conference on Extending Database Technology*, EDBT '10, pages 697–700, New York, NY, USA. ACM.

Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, AAAI'14, pages 1112–1119. AAAI Press.

Chenyan Xiong and Jamie Callan. 2015. Query expansion with freebase. In *Proceedings of the 2015 International Conference on The Theory of Information Retrieval*, ICTIR '15, pages 111–120, New York, NY, USA. ACM.

Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2014. Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575*.

Xuchen Yao and Benjamin Van Durme. 2014. Information extraction over structured data: Question answering with freebase. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, June 22-27, 2014, Baltimore, MD, USA, Volume 1: Long Papers*, pages 956–966.