The Ineffectiveness of TKGE Models in Encoding Real-World Knowledge Graphs

Ryan Ong
Data Science Institute
Imperial College London
chuan.ong18@imperial.ac.uk

Jiahao Sun
Wealth Management
Royal Bank of Canada (London)
jiahao.sun@rbc.com

Ovidiu Serban
Data Science Institute
Imperial College London
o.serban@imperial.ac.uk

Yi-Ke Guo
Data Science Institute
Imperial College London
y.guo@imperial.ac.uk

Abstract

Temporal knowledge graphs (TKGs) have been rising in popularity in many industrial applications. However, for TKG-based applications to perform accurately, we need to have a reliable temporal knowledge graph embedding (TKGE) model to capture the semantic meanings of entities and the relationship between entities. This is possible when we have many standardised academic TKGs that are well-connected with popular entities. However, in real-world settings, these well-connected TKGs are hardly available. Instead, real-world TKGs are usually more sparse and filled with noisy and less popular entities, which makes it very challenging to use to train an accurate TKGE model. In this paper, we ran five different TKGE models on the TKGQA mergers and acquisitions (M&A) dataset to assess the effectiveness of TKGE models in encoding real-world TKGs. Specifically, we selected M&As because it's common for a well-known company to merge/acquire a less popular/unknown company and as such we can evaluate the effectiveness of TKGE models in encoding the less well-known companies. The results show that TKGE models are ineffective in encoding less popular/unknown entities in sparse KGs; given the lack of information on the entities, the TKGE models find distinguishing them in the embedding space challenging.

1 Introduction

Knowledge graphs (KGs) have been used to represent human knowledge and have become very popular in different industrial applications to solve complex problems. KGs usually refer to a static representation of entities, relations, and facts, meaning that they do not change with time. On the other hand, temporal knowledge graphs (TKGs) are knowledge graphs where, in addition to entities and relations, the facts are represented by adding time information to capture the validity period of the fact.

The key to KG-based applications is to find a reliable knowledge graph embedding (KGE) model for encoding the entities, relations, and facts into low-dimensional vectors; to capture both the semantic meanings and relationships between entities. In order to train a reliable KGE model, we would require a well-connected knowledge graph to ensure that we have enough information about each entity (and between entities). The most popular standardised KG/TKG datasets are WIKIDATA (Leblay and Chekol (2018)), ICEWS (Lautenschlager et al. (2015)), GDELT (Leetaru and Schrodt (2013)), etc.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

However, when it comes to utilising KGs for real-world applications, datasets are usually less available but, more importantly, less connected and informative. These settings make it very difficult to train an accurate KGE model to encode entities and relations since a) there are not enough facts about entities and b) there are fewer connections between entities. Without an accurate KGE model to represent knowledge, we will not be able to build reliable KG-based applications.

This paper ran experiments on five different temporal knowledge graph embedding (TKGE) models on the TKGQA dataset to assess the effectiveness of TKGE models on a real-world dataset. Specifically, we run the link prediction task, a common task to evaluate the reliability of TKGE models. The resulting MRR and Hits@k scores show that TKGE models are ineffective in encoding less-connected KGs, which are common in industrial applications.

2 Related Work

2.1 Temporal Knowledge Graph Completion and Question Answering Datasets

The two most popular temporal knowledge graph completion (TKGC) datasets are the ICEWS (Lautenschlager et al. (2015)) and the GDELT (Leetaru and Schrodt (2013)) dataset, followed by the WIKIDATA (Leblay and Chekol (2018)) dataset. Cai et al. (2022) provides detailed statistics of the TKGC datasets, highlighting that all TKGC datasets have a very dense and connected knowledge graph with at least a 1:10 entities to facts ratio. For example, GDELT has the smallest entity set of 500 but with the largest number of facts of around 3.4 million. This density gives TKGE models a robust dataset to train on and enables them to encode information properly.

However, when it comes to real-world datasets, this is an unrealistic environment since there will be many entities on which we have limited information that we would still be very interested in encoding. For example, in mergers and acquisitions (M&A), it is common for rising companies to be acquired by big companies. Nevertheless, we have very little detail on these rising companies, making it hard for the TKGE models to perform well under datasets with limited data about the entities.

Similar arguments about temporal knowledge graph question answering (TKGQA) datasets can be made. For example, both Saxena et al. (2021a) and Jia et al. (2018) built the CRONQUESTIONS and TEQUILA TKGQA datasets using templates that take advantage of entities with a vast amount of connections within the knowledge graph. This approach works very well on popular entities, but these methods will not be effective on entities with fewer connections or are less known in real-world applications.

Datasets that are closer to real-world applications are the PROPARA (Dalvi et al. (2018)) dataset, which consists of procedural text on scientific processes, the TextWorld KG (Adhikari et al. (2020)) dataset, a text-based games generated using TextWorld, and the NBAtransactions (Tang et al. (2019)) dataset, which consists of the transaction-news pairs in the NBA. However, these datasets are all restricted within a specific use case and domain, which makes it a dense knowledge graph built for a specific purpose.

2.2 Temporal Knowledge Graph Embeddings

Temporal knowledge graph embedding (TKGE) models encode entities, relations and time into embeddings and ultimately perform fact scoring. Most TKGE models are built on top of existing static KGE models by extending the KGE model or processing the temporal information before it is fed into a static KGE model. For example, TTransE (Leblay and Chekol (2018)) is an extension of TransE (Bordes et al. (2013)), where in addition to using relation as translation, it also uses time information to translate (similar to relation) the entities embeddings. Other extensions include TComplEx and TNTComplEx (Lacroix et al. (2020)), which are extension of the static ComplEx (Trouillon et al. (2016)) model and BoxTE (Messner et al. (2022)), which is an extension of the static BoxE (Abboud et al. (2020)) model.

Alternatively, we can process temporal information with the static model extension before feeding it into our static KGE model. For example, Goel et al. (2020) proposed diachronic embeddings to make entity and relation embeddings as a function of temporal information. This approach led to diachronic entity embeddings, meaning that the same entity will have different embeddings depending on the time point. The benefit of this method is that it is model-agnostic, meaning that we can feed the diachronic entity and relation embeddings into any static KGE models.

3 Experimental Settings

3.1 Methods

We selected 5 TKGE models; two extensions and three diachronic embeddings. For ComplEx, TComplEx, and TNTComplEx, we used the same hyper-parameters as Saxena et al. (2021b) to initialise the models; for DE-TransE, DE-DistMult, and DE-Simple, we used the hyper-parameters introduced in Goel et al. (2020):

1. ComplEx (Trouillon et al. (2016)) - represents entities and relations as complex vectors and has the following scoring function:

$$\begin{aligned}\phi(s,r,o) =& Re(< u_s, v_r, u_o^* >)\\ =& Re(\sum_{d=1}^{D} u_{sd} v_{rd} u_{od}^*)\\ =& < Re(u_s), Re(v_r), Re(u_o) >\\ &+ < Im(u_s), Re(v_r), Im(u_o) >\\ &+ < Re(u_s), Im(v_r), Im(u_o) >\\ &- < Im(u_s), Im(v_r), Re(u_o) >\end{aligned} \tag{1}$$

where Re(.) represents the real part, Im(.) represents the imaginary part, and * represents the complex conjugate.

2. TComplEx and TNTComplEx (Lacroix et al. (2020)) - extends ComplEx model with temporal information. The temporal information is represented as complex vectors and the new scoring function for TComplEx is:

$$\phi(s,r,o,t) = Re(< u_s, v_r, u_o^*, w_t >) \tag{2}$$

For TNTComplEx, they extended the TComplEx by introducing two representations for the relations; one that's time-sensitive and one that's time-insensitive, which gives us the TNTComplEx's scoring function:

$$\phi(s,r,o,t) = Re(< u_s, v_r * T, u_o^*, w_t > + < u_s, v_r, u_o^*, 1 >) \tag{3}$$

3. DE-TransE (Goel et al. (2020)) - same scoring function as TransE (Bordes et al. (2013)) but with diachronic entities embeddings instead.

4. DE-DistMult (Goel et al. (2020)) - same scoring function as DistMult (Yang et al. (2015)) but with diachronic entities embeddings instead.

5. DE-SimplE (Goel et al. (2020)) - same scoring function as SimplE (Kazemi and Poole (2018)) but with diachronic entities embeddings instead.

The configuration of these models is based on the original papers by Saxena et al. (2021a) and Goel et al. (2020).

3.2 Dataset

The TKGQA dataset is a temporal update and question answering dataset based on financial news articles related to merger and acquisition (M&A). The TKGQA dataset consists of 14757 entities, 13 relations, and 1278 timestamps. In TKGQA dataset, facts are represented in the form of (head entity, relation, tail entity, start date, end date). There are 14777 facts, of which 12601 facts are used for training, 1088 facts for validation, and 1088 facts for testing.

Table 1: TKGE Results on TKGQA Dataset

	MRR	Hits@1	Hits@3	Hits@10
ComplEx	7.34	3.49	7.9	15.35
TComplEx	0.52	0.09	0.46	1.01
TNTComplEx	7.33	3.58	8.09	15.17
DE-SimplE	0.68	0.32	0.69	1.33
DE-DistMult	0.61	0.23	0.69	1.19
DE-TransE	**19.10**	**9.28**	**24.13**	**37.13**

4 Results and Discussions

Table 1 shows the TKGE results on the TKGQA dataset, split by the two TKGE models; the extension vs preprocessing the temporal information. As shown in Table 1, most TKGE models performed poorly on the TKGQA dataset, highlighting the ineffectiveness of these models when applied to real-world datasets that are less well-connected than standardised datasets. DE-TransE had the best results, achieving a filtered MRR score of 19.10 and outperforming TComplEx and TNTComplEx. This could be attributed by the fact that rather than having a time representation (TComplEx and TNTComplEx) that's independent to the entity and relation representations, diachronic embeddings (DE) captures the dynamic evolution patterns that exist in entities and relations by making the entity and relation representations dependent on time. This is very important in performing well on the TKGQA dataset since M&A's stages follow a sequential evolution.

In the appendix, we have included qualitative examples of different relation types to further analyse our two best-performing models, DE-TransE and TNTComplEx, and showcase the main challenge of encoding real-world datasets. For example, in the appendix, Table 2 showcases the top 10 link predictions on the subsidiary relation. Given the nature of the subsidiary, a well-known company can have many subsidiaries at any given moment, and the facts are unlikely to change, which means that the same input of head entity, relation, and time information can have many possible correct answers. These possible correct answers are bolded in each qualitative example.

In Table 2, despite the incorrect prediction of a specific subsidiary (Mercedes-Benz U.S. Int. (Q1921340)) of Mercedes-Benz Group, the top 10 predictions by TNTComplEx cover other possible subsidiaries of Mercedes-Benz Group, showing that the model managed to understand the subsidiary relation. However, it does not have enough information about the subsidiaries to distinguish between them since many of these subsidiaries lack public information and only exist in the knowledge graph as the subsidiary of a company. Thus, they have similar embeddings in the embedding space. Nevertheless, this finding of the TNTComplEx TKGE model is consistent with other qualitative examples provided in the appendix.

Regarding DE-TransE, Table 5 and Table 6 show that when dealing with unknown entities, DE-TransE tends to bias towards predicting the head entity/entities that are related to the head entity, which is incorrect. Specifically, with Table 5, most of the top 10 predictions by TransE are related to the head entity sap. Furthermore, table 3 and Table 4 show that DE-TransE faced a similar problem as TNTComplEx, where it is difficult to distinguish between entities that are related to the head entity due to a lack of information about each entity. Despite this confusion, DE-TransE still managed to score competitively on Hits@k, specifically Hits@10.

Lastly, Table 7 showcase the errors made by DE-TransE on the test set, ranked by relations type. The results support our qualitative analysis, whereby the top four errors are from relation type owned by (P127), board member (P3320), subsidiary (P355), and owner of (P1830), all of which are relations that have the characteristics of a) multiple correct answers given the same input, and b) consists of many less popular entities, making it harder for the model to distinguish between them.

5 Conclusion

In this paper, we ran five TKGE models on the TKGQA M&A dataset to evaluate how effective TKGE models are in encoding less well-known entities in a sparse and less informative real-world TKG. Our results show that many TKGE models are ineffective in encoding less well-known entities since there is very little information about them regarding their attributes and connections with other entities. Furthermore, these experiments show that our TKGE models found distinguishing between the less well-known entities challenging, resulting in similar embeddings in the embedding space.

Although it is crucial to explore TKGE models with well-known entities and well-connected TKGs, we hope our result will push more research efforts toward making TKGE models more effective in real-world settings where TKGs are sparse and less informative. Future work includes investigating low-resource TKGE models that can better encode sparse and less informative real-world TKGs. Additionally, we can experiment with different data augmentation/expansion techniques to better differentiate less well-known entities and thus encode them into different embeddings.

References

Ralph Abboud, İsmail İlkan Ceylan, Thomas Lukasiewicz, and Tommaso Salvatori. 2020. BoxE: A Box Embedding Model for Knowledge Base Completion. In *Proceedings of the Thirty-Fourth Annual Conference on Advances in Neural Information Processing Systems (NeurIPS)*.

Ashutosh Adhikari, Xingdi Yuan, Marc-Alexandre Côté, Mikulás Zelinka, Marc-Antoine Rondeau, Romain Laroche, Pascal Poupart, Jian Tang, Adam Trischler, and William L. Hamilton. 2020. Learning Dynamic Knowledge Graphs to Generalize on Text-Based Games. *CoRR* abs/2002.09127 (2020). `https://arxiv.org/abs/2002.09127`

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Modeling Multi-Relational Data. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2* (Lake Tahoe, Nevada) *(NIPS'13)*. Curran Associates Inc., Red Hook, NY, USA, 2787–2795.

Borui Cai, Yong Xiang, Longxiang Gao, Heng Zhang, Yunfeng Li, and Jianxin Li. 2022. Temporal Knowledge Graph Completion: A Survey. *ArXiv* abs/2201.08236 (2022).

Bhavana Dalvi, Lifu Huang, Niket Tandon, Wen-tau Yih, and Peter Clark. 2018. Tracking State Changes in Procedural Text: a Challenge Dataset and Models for Process Paragraph Comprehension. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. Association for Computational Linguistics, New Orleans, Louisiana, 1595–1604. `https://doi.org/10.18653/v1/N18-1144`

Rishab Goel, Seyed Mehran Kazemi, Marcus Brubaker, and Pascal Poupart. 2020. Diachronic Embedding for Temporal Knowledge Graph Completion. In *AAAI*.

Zhen Jia, Abdalghani Abujabal, Rishiraj Saha Roy, Jannik Strötgen, and Gerhard Weikum. 2018. TEQUILA: Temporal Question Answering over Knowledge Bases. *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (2018).

Seyed Mehran Kazemi and David Poole. 2018. SimplE Embedding for Link Prediction in Knowledge Graphs. In *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Eds.), Vol. 31. Curran Associates, Inc. `https://proceedings.neurips.cc/paper/2018/file/b2ab001909a8a6f04b51920306046ce5-Paper.pdf`

Timothée Lacroix, Guillaume Obozinski, and Nicolas Usunier. 2020. Tensor Decompositions for Temporal Knowledge Base Completion. In *International Conference on Learning Representations*. `https://openreview.net/forum?id=rke2P1BFwS`

Jennifer Lautenschlager, Steve Shellman, and Michael Ward. 2015. ICEWS Event Aggregations. (2015).

Julien Leblay and Melisachew Wudage Chekol. 2018. Deriving Validity Time in Knowledge Graph. In *Companion Proceedings of the The Web Conference 2018* (Lyon, France) *(WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1771–1776. `https://doi.org/10.1145/3184558.3191639`

Kalev Leetaru and Philip A. Schrodt. 2013. GDELT: Global data on events, location, and tone. *ISA Annual Convention* (2013).

Johannes Messner, Ralph Abboud, and Ismail Ilkan Ceylan. 2022. Temporal knowledge graph completion using box embeddings. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 7779–7787.

Apoorv Saxena, Soumen Chakrabarti, and Partha Talukdar. 2021a. Question Answering Over Temporal Knowledge Graphs. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, Online, 6663–6676. `https://doi.org/10.18653/v1/2021.acl-long.520`

Apoorv Saxena, Soumen Chakrabarti, and Partha Talukdar. 2021b. Question Answering over Temporal Knowledge Graphs. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics*.

Jizhi Tang, Yansong Feng, and Dongyan Zhao. 2019. Learning to Update Knowledge Graphs by Reading News. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, Hong Kong, China, 2632–2641. `https://doi.org/10.18653/v1/D19-1265`

Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex Embeddings for Simple Link Prediction. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48* (New York, NY, USA) *(ICML'16)*. JMLR.org, 2071–2080.

Bishan Yang, Wen tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding Entities and Relations for Learning and Inference in Knowledge Bases. *CoRR* abs/1412.6575 (2015).

A Appendix

A.1 Qualitative Results with DE-TransE and TNTComplEx

Table 2: Qualitative example 1 - top 10 predictions from DE-TransE and TNTComplEx on subsidiary (P355) relation. **Bolded text** - the predicted entity is one of the correct answers given the head, relation, and time

['mercedes-benz_group', 'subsidiary (P355)', 'Mercedes-Benz U.S. Int. (Q1921340)', 2021.0, 7.0, 1.0]		
k = 10	DE-TransE	TNTComplEx
1	BMW Group Plant Dingolfing (Q796355)	**Mercedes-Benz AG (Q76754241)**
2	BMW Group Classic (Q23786344)	**Daimler Japan (Q30343913)**
3	Rolls-Royce Motor Cars (Q234803)	**Rolls-Royce Power Systems (Q833456)**
4	BMW South Africa (Q27889504)	**Mitsubishi Fuso Truck and B.C. (Q1190247)**
5	Mini (Q36024)	**Fujian Benz Automotive (Q1262515)**
6	BMW-Flugmotorenbau (Q107177309)	**car2go (Q1902)**
7	Alpina (Q203550)	**American LaFrance (Q202791)**
8	BMW i (Q796784)	**Stellantis North America (Q181114)**
9	Land Rover (Q35907)	**Daimler Truck North America (Q16834431)**
10	**Mercedes-Benz U.S. Int. (Q1921340)**	**Mercedes-Benz USA (Q1921337)**

Table 3: Qualitative example 2 - top 10 predictions from DE-TransE and TNTComplEx on owned by (P127) relation. **Bolded text** - the predicted entity is one of the correct answers given the head, relation, and time

['twitter,_inc.', 'owned by (P127)', 'BlackRock (Q219635)', 2021.0, 7.0, 1.0]		
k = 10	DE-TransE	TNTComplEx
1	Capital Ventures International (Q104626435)	**Morgan Stanley (Q334204)**
2	Whisper Systems (Q7993932)	**Elon Musk (Q317521)**
3	Dustin Moskovitz (Q370217)	Whisper Systems (Q7993932)
4	twitter,_inc.	Mimi Alemayehou (Q17489752)
5	Eduardo Saverin (Q312663)	Omid Kordestani (Q3352033)
6	Mark Rachesky (Q16734836)	Fei-Fei Li (Q18686107)
7	John C. Malone (Q3181165)	Martha Lane Fox (Q6509367)
8	Newport Trust (Q65084181)	Bret Taylor (Q4961964)
9	Amundi (Q2844522)	fabula ai
10	**BlackRock (Q219635)**	mobility intelligence

Table 4: Qualitative example 3 - top 10 predictions from DE-TransE and TNTComplEx on owner of (P1830) relation. **Bolded text** - the predicted entity is one of the correct answers given the head, relation, and time

['boeing', 'owner of (P1830)', 'Boeing Commercial Airplanes (Q8793)', 2021.0, 7.0, 1.0]		
k = 10	DE-TransE	TNTComplEx
1	Insitu (Q6038180)	**United Space Alliance (Q1541150)**
2	**Boeing Commercial Airplanes (Q8793)**	**Inventory Locator Service (Q6059912)**
3	Boeing Spain (Q28974597)	**Boeing International HQ (Q961568)**
4	McDonnell Douglas (Q201815)	**Boeing Plant 1 (Q4937150)**
5	H. J. Heinz Company (Q850324)	Boeing Capital (Q4937128)
6	Boeing Canada (Q890180)	embraer s.a.
7	Boeing Australia (Q4937112)	embraer sa
8	Alteon Training (Q2552512)	Steve Mollenkopf (Q18634617)
9	American Express (Q194360)	Larry Kellner (Q6490609)
10	Boeing Aircraft Holding Company (Q4937115)	Lynn Good (Q16866325)

Table 5: Qualitative example 4 - top 10 predictions from DE-TransE and TNTComplEx on success_acq relation. **Bolded text** - the predicted entity is one of the correct answers given the head, relation, and time

['sap', 'success_acq', 'signavio', 2021.0, 7.0, 1.0]		
k = 10	DE-TransE	TNTComplEx
1	sap	**emarsys**
2	SuccessFactors (Q7632539)	**plaut**
3	SAPD Processing (Australia) (Q28975959)	**mk data**
4	Business Objects (Q1017587)	**qualtrics**
5	Gigya (Q5560364)	**callidus software inc**
6	Sybase (Q259966)	**sap_concur**
7	SAPD Processing (United States) (Q28975953)	Hasso Plattner (Q71074)
8	KXEN Inc. (Q6341142)	Dietmar Hopp (Q62543)
9	Concur Technologies (Q5159089)	Klaus Tschira (Q95452)
10	SAP ČR (Q54968398)	SAP ČR (Q54968398)

Table 6: Qualitative example 5 - top 10 predictions from DE-TransE and TNTComplEx on success_merger relation. **Bolded text** - the predicted entity is one of the correct answers given the head, relation, and time

['randgold_resources', 'success_merger', 'barrick_gold', 2021.0, 7.0, 1.0]		
k = 10	DE-TransE	TNTComplEx
1	randgold_resources	**barrick_gold**
2	**barrick_gold**	google
3	dish network corp <dish.o>	sony
4	s.n	hitachi
5	nextel_communications	ppf_(company)
6	gsk consumer india	the_walt_disney_company
7	dish network corp.	hasbro
8	fox business charlie gasparino	lvmh
9	gasparino	adobe_inc.
10	vodafone mobile services ltd (vmsl)	mastercard

Table 7: DE-TransE - Number of errors on test set by relations

Relations	**Number of Errors**
owned by (P127)	477
board member (P3320)	268
subsidiary (P355)	229
owner of (P1830)	151
expecting_acq	51
expecting_merger	30
considering_acq	24
success_acq	23
considering_merger	15
business division (P199)	15
success_merger	13
terminated_merger	10
terminated_acq	2