Can Pattern Learning Enhance Complex Logical Query Answering?

Yunjie He[1,2,3,*,†], Mojtaba Nayyeri[2], Bo Xiong[1,2], Yuqicheng Zhu[1,2,3],
Evgeny Kharlamov[3,4] and Steffen Staab[2,5]

[1]*The International Max Planck Research School for Intelligent Systems, Germany*
[2]*University of Stuttgart, Germany*
[3]*Bosch Center for Artificial Intelligence, Germany*
[4]*University of Oslo, Norway*
[5]*University of Southampton, UK*

Abstract

Logical patterns, such as symmetry and composition, have been proven to be beneficial in the knowledge graph completion task. However, their influence has been unexplored in first-order logical (FOL) query reasoning methods. In this work, we present an inductive bias for query embedding models, Pattern-aware Cone Embedding (PConE), to support learning and reasoning with logical patterns. PConE combines the advantages of cones and the rotation operator for powerful algebraic operations for pattern inference. Our experiments demonstrate how the capability to capture logical patterns positively impacts the results of query answering.

Keywords

Query answering, Knowledge graphs, Query embedding

1. Introduction

Knowledge Graphs (KGs) represent real-world facts as sets of triples of the form (head entity, relation name, tail entity). KGs can be stored in and queried efficiently by triple stores using query languages such as SPARQL under the assumption of completeness. However, when querying incomplete KGs, some triples are not explicitly available. Simply traversing the graph misses relevant results. Query embedding methods [1, 2, 3] are proposed to infer the missing links while reasoning FOL queries. These methods map queries and entities to a vector space to measure the plausibility of an entity being the answer of a query based on their distance. The considered queries allow the use of query variables for both head and tail positions, along with the negation (¬) of such constrained triple patterns, as well as the conjunction (∧) and disjunction (∨) of intermediate outcomes. An example of such queries is given in Figure 1.

The quality of query embedding approaches depends mainly on their ability to represent entities, relations, and queries in geometric spaces. In particular, relations in KGs may exhibit logical patterns, and effectively modeling these patterns relies on the geometry and the defined

ISWC 2023 Posters and Demos: 22nd International Semantic Web Conference, November 6–10, 2023, Athens, Greece
*Corresponding author.
yunjie.he@ipvs.uni-stuttgart.de (Y. He)

$$q = \phi(x) = \exists y: (AthleteWonTrophy(y, World\ Cup) \vee AthleteWonTrophy(y, Europe\ Cup)) \wedge \neg Teammate(C.Ronaldo, y) \wedge BirthPlace(x, y)$$

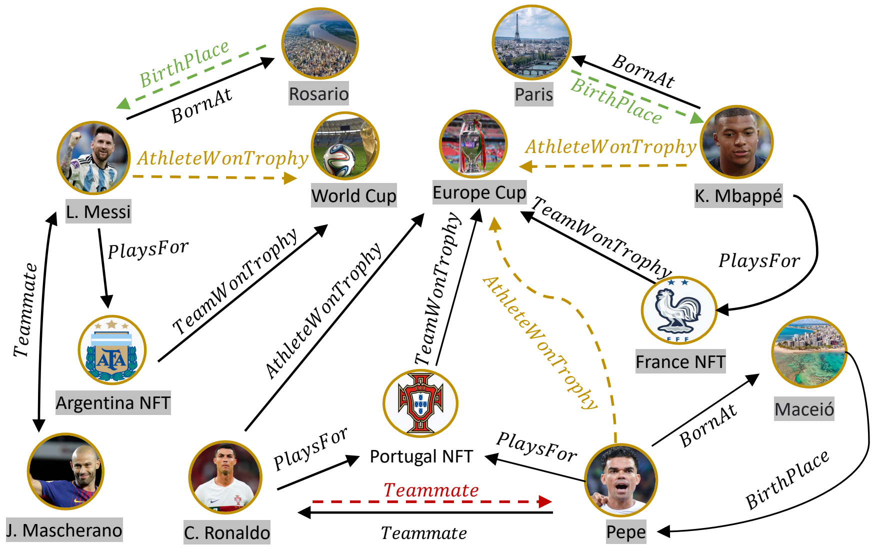

Figure 1: An example on how query answering over an incomplete KG is affected by logical patterns, with the absent edges labeled by dashed lines. The variables of the given FOL query cannot be directly extracted from the given facts.

operators. Examples of patterns found within KGs encompass symmetry (e.g., *Teammate*), inversion (e.g., *BornAt* and *hasBirthPlace*), and composition (e.g., when an athlete *PlaysFor* a team and *TeamWonTrophy*, it implies *AthleteWonTrophy*). As these logical patterns greatly influence the interplay of entities and relations, related work [4, 5] has demonstrated that an embedding's ability to support them improves its link prediction quality. However, similar accommodation of logical patterns in the embedding space is still lacking for query embeddings.

In this poster, we propose a novel method, PConE, to support the acquisition and representation of logical patterns for query answering. We describe its basic working mechanism and the evaluation strategy and present experimental evidence to demonstrate its efficiency in handling first-order logical queries with a lightweight structure (with only half the parameters of other baseline models).

2. PConE

PConE defines each relation as a rotation from the source entity set to the answer/intermediate entity set, where entities and entity sets are modeled as vectors and cones, respectively. Each cone $\mathbf{q}$ is parameterized by $\mathbf{q} = (\mathbf{h}_U, \mathbf{h}_L)$, where $|\mathbf{h}_U|_2 = \mathbf{1}$, $|\mathbf{h}_L|_2 = \mathbf{1}$ with $|\cdot|_2$ being the L2 norm, and $\mathbf{h}_U, \mathbf{h}_L \in \mathbb{C}^d$ represent the counter-clockwise upper and lower boundaries of the cone, such that $\mathbf{h}_U \equiv e^{i\boldsymbol{\theta}_U}, \mathbf{h}_L \equiv e^{i\boldsymbol{\theta}_L}$, where $\boldsymbol{\theta}_{\{U,L\}}$ represent the angle between the boundary and the axis, d is the embedding dimension. First-order logical operators, conjunction, disjunction, and

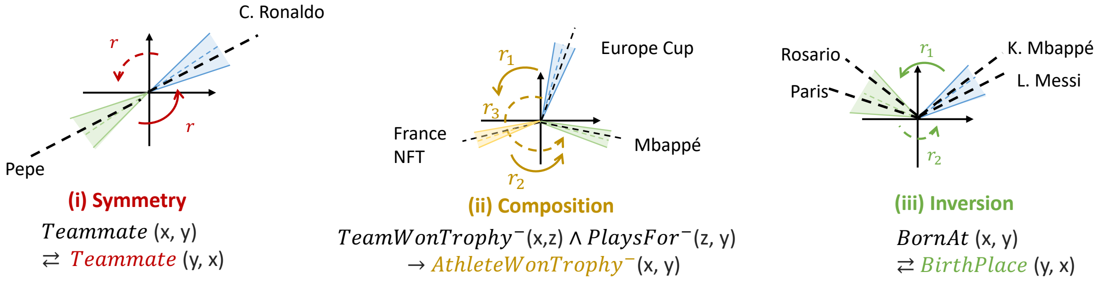

Figure 2: Illustration of PConE on query reasoning. Given the KG in Figure 1, PConE answers the query by deriving the following information with relational rotation: (i) *Pepe* is *C.Ronaldo*'s *teammate* (symmetric rotation); (ii) *Mbapp* is the athlete who has won the *Europe Cup* (compositional rotation); (iii) The birthplaces of *Messi*/*Mbapp* are *Rosario*/*Paris* (inversion rotation).

negation, are translated into geometric operators in the complex vector space. We derive the final query embedding by executing geometric operators on the selected entity sets along the computation graph. The model is trained to minimize the distance between the query cone embedding and the answer entity vector. The geometric operators are designed below.

Relational Transformation Given a set of entities and a relation, the transformation operator selects the neighboring entities by relation. Existing query embedding methods [2, 1, 6, 3] apply multi-layer perceptron networks to accomplish this task. They do not accommodate the learning of potential logical patterns which can help in reasoning logical queries. To capture diverse patterns, we represent each relation $\mathbf{r} \in \mathbb{C}^{2\times d}$ as a counterclockwise relational rotation on query embeddings about the origin of the complex plane such that $\mathbf{r} = (\mathbf{r}_U, \mathbf{r}_L)$, where $|\mathbf{r}_U| = \mathbf{1}$, $|\mathbf{r}_L| = \mathbf{1}$, and $\mathbf{r}_U, \mathbf{r}_L \in \mathbb{C}^d$. Given the query embedding $\mathbf{q} = (\mathbf{h}_U, \mathbf{h}_L)$ and a relation $\mathbf{r}$, the transformed query embedding $(\mathbf{h}'_U, \mathbf{h}'_L)$ is defined as $\mathbf{h}'_U = \mathbf{h}_U \circ \mathbf{r}_U, \mathbf{h}'_L = \mathbf{h}_L \circ \mathbf{r}_L$. Figure 2 illustrate the relational transformation process in the presence of various patterns.

Logical Operators Given the input of multiple entity sets modeled by cone embeddings, the intersection operator computes their intersections through a neural network-based permutation-invariant function. Given the cone embedding of a set of entities $\mathbf{q}$, the negation operator finds its corresponding negation $\mathbf{q}^\neg$ as the complement of the cone. In addition, given the input of multiple entity sets $\mathbf{q}_1, ..., \mathbf{q}_n$, the union operator finds the disjunction set as the union of multiple cone embeddings in the same complex plane.

3. Preliminary Results

We evaluate our model on a wider range of datasets and dataset splits in addition to existing query answering benchmark datasets to thoroughly assess how learning logical patterns affects query answering.

Model Performance Table 1 summarizes the performance of all methods on answering various query types in two benchmark datasets WN18RR [7] and NELL [3]. On nearly all query

Dataset	Model	1p	2p	3p	2i	3i	pi	ip	2u	up
	Q2B	22.4	4.6	2.3	25.6	41.2	13.2	11.0	2.9	3.4
WN18RR	BetaE	44.1	9.8	3.8	57.2	76.2	32.6	17.9	7.5	5.3
	LinE	45.1	12.3	6.7	47.1	67.1	24.8	14.7	8.4	6.9
	ConE	46.8	14.5	9.3	59.0	83.9	33.6	18.7	10.0	9.8
	PConE	**50.9**	**17.6**	**9.9**	**70.5**	**89.0**	**38.9**	**29.6**	**18.4**	**14.0**
	Q2B	42.7	14.5	11.7	34.7	45.8	23.2	17.4	12.0	10.7
NELL	BetaE	53.0	13.0	11.4	37.6	47.5	24.1	14.3	12.2	8.5
	ConE	53.1	16.1	13.9	40.0	50.8	**26.3**	17.5	15.3	11.3
	PConE	**54.5**	**17.7**	**14.4**	**41.9**	**53.0**	26.1	**20.7**	**16.5**	**12.8**

Table 1

MRR results (%) of PConE and baseline models on answering logical queries on datasets NELL and WN18RR. The best statistic is highlighted in bold, while the second best is highlighted in underline.

types, PConE consistently outperforms all baseline approaches.

Respective Influences of Logical Patterns
To better study the specific impact of PConE on queries involving logical patterns, a more in-depth analysis is made of the query answering dataset NELL. We categorize the test dataset into five categories, *Inverse*, *Symmetry*, *Composition*, *Joint*, and *Others*, based on the relations involved in the queries. The subgroup *Joint* is a conjunctive set of queries that involve all three logical patterns. The category *Others* corresponds to queries that do not involve any of these logical patterns. Figure 3 shows the average performances of PConE and neural baseline model on these subgroups. It is observed that PConE outperforms the neural baseline model on queries that had logical patterns, especially inverse relations. However, PConE does not generalize as well to queries that were not influenced by logical patterns compared to the baseline model.

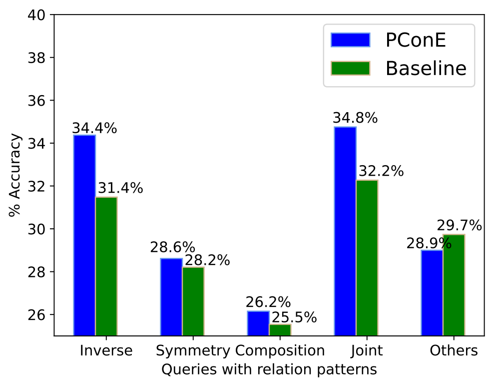

Figure 3: Average performances of PConE and Baseline model (ConE) over query subgroups with different logical patterns.

4. Conclusion

Logical patterns in complex query answering remains understudied. To the best of our knowledge, PConE is the initial study to investigate how logical patterns improve logical query reasoning. On the other hand, due to natural geometry features, the relational rotational projection can only be used to cone embedding. We will develop more generic and effective ways to improve relation pattern learning in complex query reasoning.

5. Acknowledgement

The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Yunjie He, Bo Xiong and Yuqicheng Zhu.

References

[1] H. Ren, W. Hu, J. Leskovec, Query2box: Reasoning over knowledge graphs in vector space using box embeddings, in: ICLR, OpenReview.net, 2020.
[2] Z. Zhang, J. Wang, J. Chen, S. Ji, F. Wu, Cone: Cone embeddings for multi-hop reasoning over knowledge graphs, in: NeurIPS, 2021, pp. 19172–19183.
[3] H. Ren, J. Leskovec, Beta embeddings for multi-hop logical reasoning in knowledge graphs, in: NeurIPS, 2020.
[4] Z. Sun, Z. Deng, J. Nie, J. Tang, Rotate: Knowledge graph embedding by relational rotation in complex space, in: ICLR (Poster), OpenReview.net, 2019.
[5] M. Nayyeri, C. Xu, Y. Yaghoobzadeh, S. Vahdati, M. M. Alam, H. S. Yazdi, J. Lehmann, Loss-aware pattern inference: A correction on the wrongly claimed limitations of embedding models, in: PAKDD (3), volume 12714 of *Lecture Notes in Computer Science*, Springer, 2021, pp. 77–89.
[6] W. L. Hamilton, P. Bajaj, M. Zitnik, D. Jurafsky, J. Leskovec, Embedding logical queries on knowledge graphs, in: NeurIPS, 2018, pp. 2030–2041.
[7] Z. Huang, M. Chiang, W. Lee, Line: Logical query reasoning over hierarchical knowledge graphs, in: KDD, ACM, 2022, pp. 615–625.