SIG-SWO-044-03

知識グラフの補完における Translation-based Models の発展と課題

Development and Challenges of Translation-based Models for Knowledge Graph Completion

蛭子琢磨 [1,2*]　市瀬龍太郎 [2,1,3]
Takuma Ebisu[1,2]　Ryutaro Ichise[2,1,3]

[1] 総合研究大学院大学
[1] SOKENDAI
[2] 国立情報学研究所
[2] National Institute of Informatics
[3] 産業技術総合研究所
[3] National Institute of Advamced Industrial Science and Technology

Abstract: Knowledge graphs are useful for many artificial intelligence tasks. However, knowledge graphs often have missing facts. To populate knowledge graphs, the graph embedding models map entities and relations in a knowledge graph to a vector space and predict unknown triples by scoring candidates triples. Translation-based models are part of knowledge graph embedding models and they employ the translation-based principle. The principle can efficiently capture the rules of a knowledge graph, however TransE, the original translation-based model, has some problems. To solve them many extensions of TransE have been proposed. In this paper, we discuss such problems and models.

1 はじめに

知識グラフは，現実世界の知識をコンピュータが読み取りやすい形で記述する一つの手段である．YAGO [Suchanek 07] や，DBpedia [Auer 07] といった知識グラフは，質問応答システムや知識推論といった様々な AI タスクに利用されている．一部の知識グラフは，数百万のエンティティ，数千のリレーション，数十億の知識を含んでいるにもかかわらず，多くの知識が欠落している．そこで，知識グラフを自動的に補完するシステムが求められている．

知識グラフでは，知識は有向グラフの形で表される．それぞれのノードがエンティティを表し，ラベル付けられたエッジが 2 つのエンティティの間にその関係が存在することを示す．このエッジは，終点と始点の 2 つのエンティティとエッジのリレーションからなるトリプル (h, r, t) の形で表現されている．知識グラフ中のリレーションは，しばしば特別に強い関係性を持つ．例えば，国籍は，生まれた土地と密接に関わっており，後者から前者を予測することが可能である．こういった関係性を捉えるために，様々なモデルが開発されてきた．特に，知識グラフを特定の空間に埋め込むことで，新しいトリプルを予測するモデルは知識グラフ埋め込みモデル (knowledge graph embedding model) と呼ばれ，大まかに，translation-based モデル，bilinear モデル，neural network-based モデルの 3 つに分けられる．

TransE [Bordes 13] は，知識グラフ補完のための，最初の translation-based モデルである．TransE は，その効率性と単純性から注目を集めた．TransE は，(h, r, t) が知識グラフに存在している場合，エンティティとリレーションを $\boldsymbol{h}+\boldsymbol{r}=\boldsymbol{t}$ という式を満たすように，ユークリッド空間に埋め込む．ここで，$\boldsymbol{h}$ と $\boldsymbol{r}$ と $\boldsymbol{t}$ はそれぞれ，埋め込まれた h と r と t を表す．原理自体は非常に単純であるものの，知識グラフの構造を効率的に捉えることが可能である．しかし，TransE にはいくつかの問題点があることがこれまで指摘されてきた．これらの問題点の克服を試みたモデルとして，TransH [Wang 14]，pTransE [Lin 15a]，TorusE [Ebisu 18] などがあげられる．この論文では，TransE の課題と，その拡張モデルについて論じ，translation-based モデルに残された課題を示す．

*連絡先： 総合研究大学院大学複合科学研究科情報学専攻
東京都千代田区一ツ橋２丁目１－２
E-mail: takuma@nii.ac.jp

2 TransE

TransE [Bordes 13] は，最初の translation-based モデルである．TransE は単語の分散表現を得るためのモデルである skip-gram モデル [Mikolov 13] から着想を得て作られたモデルであり，TransE でもユークリッド空間に埋め込まれたエンティティの差がその間の関係を表すようにベクトル表現を学習する．つまり，TransE は，存在するトリプルに対して $\boldsymbol{h} + \boldsymbol{r} = \boldsymbol{t}$ という原理を採用している．この原理は，一階述語論理を捉えることが可能である．例えば，“$\forall e_1, e_2 \in E$, $(e_1, r_1, e_2) \rightarrow (e_1, r_2, e_2)$”というルールが存在すれば，$r_1$ と r_2 のベクトル表現が同じになるように学習し，“$\forall e_1, e_2 \in E$, $\{\exists e_3 \in E, (e_1, r_1, e_3) \wedge (e_3, r_2, e_2)\} \rightarrow (e_2, r_3, e_1)$”. というルールが存在すれば，$\boldsymbol{r_1} + \boldsymbol{r_2} = \boldsymbol{r_3}$ となるように学習する．

TransE は，主に次の 3 つの要素から成る．

- 原理: TransE は知識グラフに存在する (h, r, t) について，$\boldsymbol{h} + \boldsymbol{r} = \boldsymbol{t}$ を満たすようにベクトル表現を学習する．ベクトル表現が，どれくらい原理に従っているかを測るために，スコア関数 f を用いる．通常，$\boldsymbol{h} + \boldsymbol{r} - \boldsymbol{t}$ の L_1 か L_2 ノルムが $f(h, r, t)$ として用いられる．この場合，$f(h, r, t) = 0$ であることは，$\boldsymbol{h} + \boldsymbol{r} = \boldsymbol{t}$ を正確に満たすことを示す.
- Negative Sampling: 知識グラフは，通常正例しか含まない．そこで，TransE では Negative Sampling によって負例を作成する．負例は，正例トリプルのヘッドエンティティかテイルエンティティをランダムに変更することによって得られる．TransE はこれらの負例について，そのスコアが大きくなるように学習する．
- 正則化: TransE にはベクトル表現が拡散することを防ぐために，正則化が必要である．TransE は正規化を正則化として用いている．つまり，全てのエンティティのベクトル表現の大きさは 1 である．言い換えると，全てのエンティティのベクトル表現は，超球面上に存在する．

TransE は margin loss を目的関数の定義に用いる．目的関数の定義は次の通りである：

$$\mathcal{L} = \sum_{(h,r,t) \in \Delta} \sum_{(h',r,t') \in \Delta'_{(h,r,t)}} [\gamma + f(h, r, t) - f(h', r, t')]_+ \quad (1)$$

ここで，Δ は正例集合，Δ' は負例集合であり，$[x]_+$ は x の正部分を取る関数，$\gamma > 0$ はマージンハイパーパラメータである．TransE は最急降下法を用いて学習を行う．

TransE は大きく分けて，次の 2 つの問題を持つことが指摘されてきた：

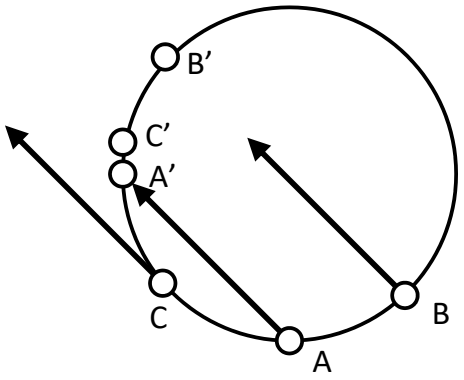

図 1: TransE より得られるベクトル表現のイメージ．ここで，(A, r, A'), (B, r, B') と (C, r, C') は真のトリプルであり，これらを球面上に上手く埋め込むことは難しい．

- 埋め込み可能性問題:　TransE を用いて，任意の知識グラフが整合性をもってユークリッド空間に埋め込めることは保証されていない．
- 原理と正則化の競合問題: 学習の際に，原理と正則化が競合し，原理を正確に満たすことが困難である．

「埋め込み可能性問題」を引き起こしている例としては，いくつかのリレーションを表現することが不可能であることがあげられる．例えば，TransE の原理では，対称関係 r を表そうことすると，$\boldsymbol{r} + \boldsymbol{r} = 0$ であることが求められるため，上手くベクトル表現ができない．ほかには，推移関係 r' においては，$\boldsymbol{r'} + \boldsymbol{r'} = \boldsymbol{r'}$ が要求され，これもまた問題となりうる．また，原理と正則化の競合例を，図 1 に示す．エンティティは球面 S^{n-1} 上に存在するが，ほとんどのエンティティ e とリレーション r について，$\boldsymbol{e} + \boldsymbol{r} \notin S^{n-1}$ が成り立つ．これまでに，これらの問題を解決するために，様々な拡張モデルの提案がなされてきた．

3 Translation-based モデルの発展

「埋め込み可能性問題」を解決するために，大きく分けて 2 つのアプローチが取られてきた．1 つ目は，リレーションごとにエンティティのベクトル表現を別の空間に写像する方法を用いた embedding-projection モデル，2 つ目は，TransE を埋め込み可能性の問題のないシンボリックな手法と組み合わせた方法である pTransE. そして，「原理と正則化の競合問題」は，埋め込み空間を変更したモデルである TorusE によって解決が試みられた．

3.1 Embedding-projection モデル

いくつかのモデルは，エンティティのベクトル表現を，リレーションごとに定められた写像によって写すことで，モデルの表現力を上げることを提案した．それらのモデルと，トリプル (h, r, t) についてのスコア関数を以下にしめす．

- TransH[Wang 14]: $\|(\boldsymbol{h}-\boldsymbol{w}_r^{\mathrm{T}}\boldsymbol{h}\boldsymbol{w}_r)+\boldsymbol{r}-(\boldsymbol{t}-\boldsymbol{w}_r^{\mathrm{T}}\boldsymbol{t}\boldsymbol{w}_r)\|_i$, ここで，$\boldsymbol{w}_r$ は単位ベクトルである．このモデルは，リレーションに応じて，エンティティを超平面上へと射影する．
- TransR [Lin 15b]: $\|\boldsymbol{W}_r\boldsymbol{h}+\boldsymbol{r}-\boldsymbol{W}_r\boldsymbol{t}\|_i$ ここで，$\boldsymbol{W}_r$ は行列である．TransR は TransH の一般化であり，リレーションに応じて，エンティティを別のベクトル空間へと線形に写像する．
- STransE[Nguyen 16]: $\|\boldsymbol{W}_{r,1}\boldsymbol{h}+\boldsymbol{r}-\boldsymbol{W}_{r,2}\boldsymbol{t}\|_i$ ここで，$\boldsymbol{W}_{r,i}$ は行列である．STransE は TransR の一般化であり，ヘッドエンティティとテイルエンティティは異なる写像で写される．

このようなスコア関数の変更により，正例トリプルについて，そのスコアは 0(つまり最良) を取りやすくなる．その一方で，これにより，いくつかの問題が起きることが予見される．まず，計算時間の増大である．特に，TransR と STransE は行列積を用いており，この計算時間は，$O(n^2)$ である．これは TransE の良い点であった，計算時間の短さを損なう危険性がある．そして，最も懸念されることは，原理の変更の弊害である．スコア関数の変更は，原理の変更にほかならず，これらのモデルではより弱い原理に変更されている．したがって，従来の原理では，ある程度一階述語論理を捉えることが可能であったが，弱い原理ではそうすることが難しくなる．

具体的に「原理と正則化の競合問題」には触れられていないが，これらのモデルでは正則化の変更も行われている．TransH，TransR,STransE は，従来の球面上へのエンティティのベクトル表現の制限ではなく，L_1 や L_2 正則化，もしくは球内への制限を用いている．

3.2 pTransE

pTransE [Lin 15a] は，TransE に，パスから得られる情報を加味したモデルである．具体的には，正例 (h, r, t) について，h と t の間に，リレーションの列であるパス $p = (r_1, \ldots, r_i)$ のインスタンスが存在する場合に，$\boldsymbol{r} = \sum_1^i \boldsymbol{r}_{\boldsymbol{k}}$ が成り立つように学習する．スコア関数は以下の通りである．

$$f(h,r,t) = \|\boldsymbol{h}+\boldsymbol{r}-\boldsymbol{t}\| + 1/Z \sum_{(r_{1,k},\ldots,r_{i,k})\in P(h,t)} Pr(r|p)R(p|h,t)\|(\sum_1^i \boldsymbol{r}_{j,k})-\boldsymbol{r}\| \quad (2)$$

ここで，$R(p|h,t)$ は，h を始点として，p に沿ったランダムウォークが t にたどり着く確率，また，Z は $\sum_{p\in P(h,t)} R(P|h,t)$ である．

TransE に，パスの情報を加味することにより，TransE でうまく扱うことができないグラフでも正確にトリプルを予測することを可能とする．問題点としては，トレーニングにかかる時間が非常に長いことである．これは，2 つのエンティティ間に存在するパスの数が，長さについて指数的に増加するからである．そこで，元の論文では，パスの長さの上限を小さい値に定めている．

pTransE でも正則化の変更が行われており，エンティティとリレーションは球内へと制限されている．

3.3 TorusE

TorusE [Ebisu 18] は「原理と正則化の競合問題」の解決に取り組んだモデルである．TorusE は，n 次元 Torus $T^n = \underbrace{S^1 \times S^1 \times \cdots \times S^1}_{n}$ を埋め込み空間として用いることで，正則化を用いずにベクトル表現の学習を行うことを可能としている．T^n はリー群と呼ばれる，多様体と群の両方の構造を持つ空間の 1 つであり，その上には和や差が自然に定義されている．例えば，1 次元 Torus$T^1 = S^1$ 上の点は，その角度で表現することが可能である，和や差は角度の和と差によってなされる．したがって，TransE の原理をそのまま Torus 上で定義することが可能である．更に，Torus はコンパクト空間であるため，ベクトル表現の発散が起こり得ず，正則化は不要となる．

このモデルでは,「原理と正則化の競合問題」が完全に解決されている．また，埋め込み可能性の問題も一部解決されている．例えば，$T^1 = S^1$ において,180° に対応する点の和は 0° に対応する点である．つまり，$\boldsymbol{r}+\boldsymbol{r} = 0$ を満たすような $\boldsymbol{r} \neq 0$ が存在する．したがって，ユークリッド空間上への埋め込みでは表すことができなかった，婚姻関係などの一対一の対象関係を上手く表現できる．その一方で，その他の埋め込み可能性問題は解決されていない．

4 実験

これまでに述べたモデルは，link prediction タスク [Bordes 13] を用いて，評価がなされてきた．この章で

表 1: 各モデルの実験結果．これらの結果は，オリジナルの論文より良い結果が後から報告されている場合，その結果を示す．TransE の結果 [Ebisu 18]，TransR の結果 [Nickel 16]，TransH と pTransE の結果 [Xiao 16] はオリジナルの論文より良い結果が発表されているため，それらを引用する．

	WN18					FB15K				
	MRR		HITS@			MRR		HITS@		
Model	Filtered	Raw	*1*	*3*	*10*	Filtered	Raw	*1*	*3*	*10*
TransE	0.397	0.306	0.040	0.745	0.923	0.414	0.235	0.247	0.534	0.688
TransE (L_2 regularization)	0.435	0.314	0.081	0.782	0.942	0.512	0.245	0.337	0.652	0.784
TransH	–	–	0.313	–	0.920	–	–	0.248	–	0.644
TransR	0.605	0.427	0.335	0.876	0.940	0.346	0.198	0.218	0.404	0.582
STransE	–	–	–	–	0.934	–	–	–	–	0.797
pTransE	–	–	–	–	–	–	–	0.633	–	**0.846**
TorusE	**0.947**	**0.619**	**0.943**	**0.950**	**0.954**	**0.733**	**0.256**	**0.674**	**0.771**	0.832

表 2: データセットの詳細

Dataset	# Ent	# Rel	# Train	# Valid	# Test
WN18	40,943	18	141,442	5,000	5,000
FB15K	14,951	1,345	483,142	50,000	59,071

は，このタスクの結果を示し，得られた結果を考察することで，各々のモデルのアプローチを評価する．

4.1 データセット

2 つのデータセット，WN18 と FB15k[Bordes 13] を用いて実験は行われた．これらのデータセットは，それぞれ WordNet [Miller 95] と Freebase [Bollacker 08] から抽出されたものである．これらのデータセットのエンティティ数，リレーション数，トレーニングトリプル数，検証トリプル数そしてテストトリプル数を，表 2 に示す．

4.2 実験手順

それぞれのモデルは，学習データを用いて学習を行い，検証データを用いてハイパーパラメータの設定を行う．その後，それぞれのテストトリプルについて，ヘッドエンティティ(またはテイルエンティティ) をそれぞれのエンティティに置換した候補トリプルの作成を行う．候補トリプルを，モデルにより得られたベクトル表現とスコア関数を用いてランク付けする．このままでは，元のトリプルが，他の真のトリプルの影響により，不当に低くランク付けされるおそれがある．それを避けるために，ランキングから，学習データ，検証データ，テストデータ (元トリプル以外) に含まれるトリプルを削除する．このランキングを”filtered”ランキングと呼ぶ．このランキングを用いて，Mean Reciprocal Rank (MRR)，HITS@n を計算する．HITS@n は，元のトリプルが，n 位以内にランク付けされるクエリの割合である．

4.3 実験結果

実験結果を，表 1 に示す. まず TransE と TransE (L_1 もしくは L_2 正則化) を比べる．単純に，正則化を弱めただけであるが，これだけでも「原理と正則化の競合問題」の影響を和らげることができる．この結果，全てにおいて，L_1 もしくは L_2 正則化を用いたほうが，高い精度が得られる．この結果から,「原理と正則化の競合問題」が如何に TransE に影響を与えているかがわかる．

次に，embedding-projection モデルについて論じる．これらのモデルは，元の TransE と違って，L_1 もしくは L_2 正則化を採用しており，これからも恩恵を受けていることから，L_2 正則化を用いた TransE と比較することが妥当である．embedding-projection モデルは，WN18 の HITS@1 において，TransE より高い精度が得られているが，その他については TransE と同じぐらいか低い傾向がみられる．ここで，WN18 は全てのリレーションに対して逆リレーションが存在し，ほぼ全てのテストトリプルに対して逆トリプルがトレーニングセットに含まれることから，原理を弱めることで，明示的に逆トリプルが存在するような (もしくは長さ 1 のパスで辿れるような) クエリに対しては精度があがったと考える．逆に，原理の変更はパスをたどることを困難

にすることを示唆していたが，そのために比較的長いパスを辿らなくてはならないデータついての精度は下がっているのではないかと考える．このことを鑑みると，提案された embedding-projection モデルのような方法で原理を変更することが，Link Prediction に必ずしも良い影響を与えるとは言えない．

pTransE は，パスの情報を用いることによって，FB15k において高い精度を出している．WN18 の結果は報告されていないが，このデータセットは，先程述べた通り，逆関係を多く含んでおり，FB15k よりシンボリックな操作に向いたデータセットであるため，こちらも上手く扱えるであろう．

TorusE は，埋め込み可能性の問題はほぼ触れておらず，「原理と正則化の競合問題」の改善が主であるが，ほぼ全てにおいて最も良い結果を出している．これは正則化の競合がいかに問題であったかを示すと同時に，WN18 と FB15k は従来の原理を用いても，比較的埋め込みやすい対象であることを示している．

5 まとめ

TransE が提案されてから長らく時間が立つものの，直接的な「埋め込み可能性の問題」の解決の糸口は見えていない．しかし，パスの情報というシンボリックな手法を取り入れることで，精度を大きく上げることが可能であることが示されている．その一方で，「原理と正則化の競合問題」は完全に解決された．

これからの translation-based モデルの発展には，埋め込み可能性の問題をどう対処していくかが課題である．

謝辞

本研究の一部は，新エネルギー・産業技術総合開発機構 (NEDO) の支援により実施されたものである．

参考文献

[Auer 07] Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., and Ives, Z. G.: DBpedia: A Nucleus for a Web of Open Data, in *Proceedings of the 6th International Semantic Web Conference and the 2nd Asian Semantic Web Conference*, pp. 722–735 (2007)

[Bollacker 08] Bollacker, K., Evans, C., Paritosh, P., Sturge, T., and Taylor, J.: Freebase: A Collaboratively Created Graph Database for Structuring Human Knowledge, in *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, pp. 1247–1250 (2008)

[Bordes 13] Bordes, A., Usunier, N., García-Durán, A., Weston, J., and Yakhnenko, O.: Translating Embeddings for Modeling Multi-relational Data, in *Advances in Neural Information Processing Systems*, pp. 2787–2795 (2013)

[Ebisu 18] Ebisu, T. and Ichise, R.: TorusE: Knowledge Graph Embedding on a Lie Group, in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence* (2018)

[Lin 15a] Lin, Y., Liu, Z., Luan, H., Sun, M., Rao, S., and Liu, S.: Modeling Relation Paths for Representation Learning of Knowledge Bases, in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 705–714 (2015)

[Lin 15b] Lin, Y., Liu, Z., Sun, M., Liu, Y., and Zhu, X.: Learning Entity and Relation Embeddings for Knowledge Graph Completion, in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, pp. 2181–2187 (2015)

[Mikolov 13] Mikolov, T., Chen, K., Corrado, G., and Dean, J.: Efficient Estimation of Word Representations in Vector Space, *CoRR*, Vol. abs/1301.3781, (2013)

[Miller 95] Miller, G. A.: WordNet: A Lexical Database for English, *Commun. ACM*, Vol. 38, No. 11, pp. 39–41 (1995)

[Nguyen 16] Nguyen, D. Q., Sirts, K., Qu, L., and Johnson, M.: STransE: a novel embedding model of entities and relationships in knowledge bases, in *Proceedings of NAACL HLT 2016*, pp. 460–466 (2016)

[Nickel 16] Nickel, M., Rosasco, L., and Poggio, T. A.: Holographic Embeddings of Knowledge Graphs, in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pp. 1955–1961 (2016)

[Suchanek 07] Suchanek, F. M., Kasneci, G., and Weikum, G.: Yago: a core of semantic knowledge, in *Proceedings of the 16th International Conference on World Wide Web*, pp. 697–706 (2007)

[Wang 14] Wang, Z., Zhang, J., Feng, J., and Chen, Z.: Knowledge Graph Embedding by Translating on Hyperplanes, in *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, pp. 1112–1119 (2014)

[Xiao 16] Xiao, H., Huang, M., and Zhu, X.: From One Point to a Manifold: Knowledge Graph Embedding for

Precise Link Prediction, in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pp. 1315–1321 (2016)