

# オートエンコーダとの同時学習による知識共有

高橋 諒<sup>1</sup> 田 然<sup>1</sup> 乾 健太郎<sup>1,2</sup>

<sup>1</sup> 東北大学 <sup>2</sup> 理研 AIP

{ryo.t, tianran, inui}@ecei.tohoku.ac.jp

## 1 はじめに

WordNet [1], YAGO [2], Freebase [3] などの**知識ベース**は、意味解析 [4], 情報抽出 [5], 質問応答 [6] などに広く応用されてきた。知識ベースはノードとしての**エンティティ**と異なるタイプのエッジとしての**関係**とで構成されるマルチリレーショナルグラフであり、**事実**を(ヘッドエンティティ, 関係, テールエンティティ)の三つ組で表現する。事実を完全網羅するのが難しいため、欠損されたエンティティを予測する**知識ベース補完**が盛んに研究されてきた [7]。これらの研究では、エンティティと関係を含む知識ベースの要素を連続ベクトル空間に埋め込むことで、格納された事実(三つ組)の背後に存在する統計的な規則性をモデル化し、記述漏れした事実を導き出す。

この連続空間埋め込みを行う際の研究課題の一つに、関係間の知識共有がある。直感的に、例えば「(人物が) (作品を) 監督した」と「(人物が) (作品を) 製作した」という関係はどちらも「人物」と「作品」を結びつく概念であるように、多くの関係はいくつかの概念を共有するので、このような直感を知識ベースのモデル化に取り入れることが望ましい。連続空間への埋め込みはこのような概念の共有を促すが、一方で知識データベースのモデル化の性質上、関係はエンティティ間を写像する演算として機能しなければならないので、関係を埋め込む際の選択肢は大幅に制限される。例えば、関係をエンティティベクトル間の線形変換としてモデル化するのは自然な方法であるが、この場合に関係はエンティティベクトルの次元の2乗分のパラメータを持ち、埋め込みによる概念共有の促進作用が非常に弱いと思われる。逆に、関係をエンティティベクトル間の平行移動としてモデル化する場合、関係の埋め込み次元はエンティティベクトルの次元と等しいが、この場合に概念共有の促進作用が非常に強いけれど、複雑な関係に対しては表現力が不足すると思われる [7]。この

ように、空間埋め込みに頼るだけでは、どの選択肢も関係の概念共有に最適であるとは限らない。

そこで本研究は、知識ベースの埋め込みを訓練する**と同時に**、関係に対するオートエンコーダを合わせて学習することで関係間の知識共有を促す柔軟な枠組みを提案し、評価実験を行なった。この手法は、関係のモデル化自体に何の制約も与えずに、知識共有の強さ度合いをオートエンコーダのコード長で調整できる利点を持つ。

知識ベース補完で一般的に用いられるベンチマークデータセットでの実験は、我々のモデルの最先端の性能を示す。関係に対応する行列同士の類似度を比較することで、似た関係は行列同士も似ることを観測し、オートエンコーダによるクラスタリングを効果的に使っていることを明らかにする。

## 2 関連研究

知識ベース補完のためにこれまで多くの埋め込みモデルが提案されてきた。このようなモデルの近年の包括的な調査は Nguyen (2017) [7] や Wang ら (2017) [8] を参照されたい。

この分野を草分けしたモデル TransE [9] は、関係を単純にエンティティベクトル間の平行移動としてモデル化したものである。しかし関係をこのように表現すると、一対一関係に対しては有効であるが、多対多関係のような複雑な概念に対しては表現力が不足していると思われる [7]。そこで、TransR [10] や STransE [11] などの拡張では、平行移動を行う前に、関係に応じてエンティティベクトルに線形変換を施すことで関係の多様性に対処したが、これによってまた関係に対するたくさんのパラメータ(行列)を導入することになり、関係間の知識共有は課題となった。

このような背景に踏まえ、Xie ら (2017) [12] の研究は我々とモチベーションを共有し、関係間の知識共有を明示的にモデル化した。具体的には、彼らのモデルは

STransE をベースとするが、関係に応じて施す線形変換は数少ない共通概念を表す行列の和で表せるというハードな制約をモデルに加えた。このような制約付きで学習するために、疎なアテンション機構によるアルゴリズムを提案し、関係に対する行列間のパラメータ共有を実現させた。これに対して我々の提案はハードな制約がないため、通常の高率勾配降下法 (SGD) による学習ができ、より柔軟なモデルであると言える。

知識ベース補完にはパス情報が有用であることも知られている [13,14] が、我々の同時学習手法は訓練スキームに特別な制約を入れていないので、パス情報を利用する訓練を取り入れることも簡単である。パス情報を使った訓練は、多段階推論によってより多くの欠損事実を補完する可能性を持つ。例えば、バラク・オバマがホノルルで生まれ、かつホノルルがアメリカ合衆国の都市であることを使えば、オバマの国籍がアメリカ合衆国であることを推測できるかもしれない。したがって、今回我々はパス情報も取り入れた訓練モデルを実装し、評価した。将来的には更に、Riedel ら (2013) [15] の研究のように、自然言語のテキストも使った同時学習で知識ベース補完を行いたい。

オートエンコーダは、高次元な入力データを一旦低次元のコードに変換してから、コードを使って元の入力データを復元するように訓練されたニューラルネットワークである。これまでその次元削減効果が注目され、教師なしの学習手法として事前訓練に使われたり [16]、分散表現の学習手法として単体で使われたり [17–19] してきたが、同時学習で知識共有を促す手段としてオートエンコーダを使ったのは、我々の知る限りこの研究が初めてである。

### 3 モデル

知識ベース補完のモデルとして、我々は Guu らのパス情報を取り入れた双線型モデル [13] と vecDCS [20] の訓練手法をベースにする。このモデルでは、エンティティを  $d$ -次元のベクトル、関係を  $(d \times d)$  の行列として表現する。学習は、一つのエンティティ  $h$  から出発し、いくつかの関係  $r_1, \dots, r_n$  からなるパスを経由して辿り着いたもう一つのエンティティ  $t$  に対して、エネルギー関数  $f(h, r_1, \dots, r_n, t)$  を最大化することでエンティティベクトル  $\mathbf{h}$ ,  $\mathbf{t}$  と行列  $\mathbf{M}_{r_1}, \dots, \mathbf{M}_{r_n}$  を推定する。エネルギー関数は

$$f(h, r_1, \dots, r_n, t) := \exp(\mathbf{t}^\top \mathbf{h} \mathbf{M}_{r_1}, \dots, \mathbf{M}_{r_n} \mathbf{t})$$

と定義し、知識ベースからこのようなデータが取れる尤度に相当する。推定時は、知識ベースからのデータと合わせ、ランダムに生成された  $k$  個の負例  $h, r'_1, \dots, r'_n, t'$  を使って

$$\frac{f(h, r_1, \dots, r_n, t)}{k + f(h, r_1, \dots, r_n, t)} \cdot \prod_k \frac{k}{k + f(h, r'_1, \dots, r'_n, t')} \quad (1)$$

を最大化する。これは、「 $(h, r_1, \dots, r_n, t)$  が正例で  $(h, r'_1, \dots, r'_n, t')$  が負例」であるイベントの尤度に相当する。

#### 3.1 オートエンコーダとの同時学習

ここで、行列  $\mathbf{M}_r$  を  $d^2$ -次元のベクトルとして見なした時に記号  $\mathbf{m}_r$  を使う。オートエンコーダは、まず行列  $\mathbf{A}$  によって  $\mathbf{m}_r$  を  $d^2$  より遥かに低い次元数  $l$  のコードベクトルに射影し、得られたコード  $\mathbf{A}\mathbf{m}_r$  に非線形変換の  $\text{ReLU}(\mathbf{A}\mathbf{m}_r)$  を掛けた後、行列  $\mathbf{B}$  をもって元の行列を復元しようとする：

$$\mathbf{m}_r \approx \mathbf{B} \text{ReLU}(\mathbf{A}\mathbf{m}_r)$$

復元したベクトルと  $\mathbf{m}_r$  とのコサイン類似度を最大化するために、エネルギー関数

$$g(\mathbf{m}_r) := \exp(\mathbf{m}_r \cdot \mathbf{B} \text{ReLU}(\mathbf{A}\mathbf{m}_r))$$

を定義し、最適化の際にランダムに生成された  $k$  個の負例  $\mathbf{m}_{r'}$  と合わせて

$$\frac{g(\mathbf{m}_r)}{k + g(\mathbf{m}_r)} \cdot \prod_k \frac{k}{k + g(\mathbf{m}_{r'})} \quad (2)$$

を最大化する。また、式(2)の最大化において  $\mathbf{m}_r$  に対する勾配も計算し、これと式(1)で計算された  $\mathbf{m}_r$  の勾配と合わせてパラメータ  $\mathbf{m}_r$  の更新を行う。オートエンコーダとの同時学習によって、 $\mathbf{m}_r$  が低次元のコードから「復元されやすい位置」、つまり類似した関係同士がクラスタしているような空間位置に動くことを期待される。また、全ての関係  $r$  に対して  $\mathbf{m}_r$  が同じ行列  $\mathbf{A}$ ,  $\mathbf{B}$  によってエンコード・デコードされるので、異なる  $\mathbf{m}_r$  同士が行列  $\mathbf{A}$ ,  $\mathbf{B}$  を介してパラメータを共有しているとの見方もできる。これによって、異なる関係間の知識共有が促されると思われる。

## 4 実験

### 4.1 設定

知識ベース補完の性能を評価するために 4 つのベンチマークデータセット上で実験を行う。各データセ

表1 実験に用いたデータセットの統計情報. #E と #R はそれぞれエンティティと関係の種類数を表す. #Train, #Valid, #Test はそれぞれ訓練, 開発, 評価セットに含まれる三つ組の数を表す.

Dataset	#E	#R	#Train	#Valid	#Test
WN18	40,943	18	141,442	5,000	5,000
FB15k	14,951	1,345	483,142	50,000	59,071
WN18RR	40,943	11	86,835	3,034	3,134
FB15k-237	14,541	237	272,115	17,535	20,466

ットの統計情報を表 1 に示す. WN18 (WordNet) と FB15k (Freebase) は Bordes ら (2013) [9] によって, FB15k-237 は Toutanova and Chen (2015) [21] によって, WN18RR は Dettmers ら (2017) [22] によってそれぞれ導入された<sup>\*1</sup>. WN18RR と FB15k-237 の開発・評価セットは訓練セットに含まれない語彙を持つため, このような語彙を含む事実は予め除外した.

知識ベース補完では, ヘッドエンティティ (またはテールエンティティ) と関係が与えられたとき, もう一方のエンティティを予測するモデルの性能を評価する. 例えば, 三つ組  $(h, r, t)$  における  $h$  と  $r$  が与えられたときのテールエンティティを予測するには, 知識ベースに含まれるエンティティ  $t'$  に対しエネルギー関数  $f_r(h, t')$  を計算し, その値に応じてエンティティをランク付けする. 予測された平均ランク (MR) とランクの逆数の平均 (MRR) と Hits@k (上位 k 位にランクされた正しいエンティティの割合) を報告する. ランク付けの対象とするエンティティを選ぶ手続きは, 先行研究と同様, Bordes ら (2013) [9] の *filtered* 設定に従う. すなわち, あるエンティティをランク付けするとき, 訓練, 開発, または評価セットで既存の三つ組の一部であることが分かっている全てのエンティティを削除する. これは, 評価対象のエンティティよりも他の正しいフィルターを高くランクするようなモデルにペナルティを課すことを回避するためである.

## 4.2 結果

WN18 と FB15k における評価結果を表 2 に, WN18RR と FB15k-237 における評価結果を表 3 に示す. 我々の提案モデルは WN18 と WN18RR 上で平均ランクの最先端の性能を達成している一方で, MRR と

<sup>\*1</sup> WN18 と FB15k は, 評価セットの事例のほとんどがヘッドエンティティとテールエンティティを反転させた事例を訓練セットに含むため, 一般的な知識ベースをモデル化する能力を評価するためのデータセットとして現在では推奨されていない [22].

Hits@k は先行研究と比べて低く留まった.

## 4.3 分析

オートエンコーダとの同時学習により, 異なる関係に対応する行列たちは互いにパラメータを共有するため, 意味的に似た関係はその行列同士も似るよう学習が進むはずである. 表 4 に FB15k において関係 *nominated\_for* に対応する行列と類似度が高い関係上位 4 件を示す. どちらのモデルにおいても, *nominated\_for* と意味的に類似した *honored\_for* がコサイン類似度の点で最も似た関係として学習されている. しかし提案モデルはオートエンコーダを用いないモデルに比べて, 他の意味的に類似しない関係との類似度を下げる方向に学習が進んでいる. 同様の傾向は表 5 に示した関係 *spouse* についても見られる. このように提案モデルは, オートエンコーダと同時学習することにより, オートエンコーダの意味的に類似した関係を集めるクラスタリング効果を知識ベース埋め込みに効果的に取り入れている.

## 5 おわりに

将来的には, Riedel ら (2013) [15] の研究に基づき, 知識ベース上の関係とテキスト上の関係の連続的な表現を同時に学習することを計画している.

## 謝辞

本研究は NTT コミュニケーション科学基礎研究所, そして JST CREST(課題番号: JPMJCR1301) の支援を受けて行ったものである.

## 参考文献

- [1] George A. Miller. Wordnet: A lexical database for english. *Commun. ACM*, 38(11):39–41, 1995.
- [2] Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. Yago: a core of semantic knowledge. In *Proceedings of the 16th International Conference on World Wide Web, WWW 2007, Banff, Alberta, Canada, May 8-12, 2007*, pages 697–706, 2007.
- [3] Kurt D. Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2008, Vancouver, BC, Canada, June 10-12, 2008*, pages 1247–1250, 2008.
- [4] Jonathan Berant, Andrew Chou, Roy Frostig, and Percy Liang. Semantic parsing on freebase from question-answer pairs. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1533–1544, 2013.
- [5] Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104(1):11–33, 2016.
- [6] Ben Hixon, Peter Clark, and Hannaneh Hajishirzi. Learning knowledge graphs for question answering through conversational dialog. In *NAACL*, 2015.
- [7] Dat Quoc Nguyen. An overview of embedding models of entities and relationships for knowledge base completion. *CoRR*, abs/1703.08098, 2017.
- [8] Quan Wang, Zhenqiang Mao, Bin Wang, and Li Guo. Knowledge graph embedding: A survey of approaches and applications. *IEEE Trans. Knowl. Data Eng.*, 29(12):2724–2743, 2017.
- [9] Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States.*, pages 2787–2795, 2013.
- [10] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu.

Method	WN18					FB15k				
			Hits					Hits		
	MR	MRR	@10	@3	@1	MR	MRR	@10	@3	@1
TransE [9]	251	-	89.2	-	-	125	-	47.1	-	-
TransH [23]	303	-	86.7	-	-	87	-	64.4	-	-
TransR [10]	225	-	92.0	-	-	77	-	68.7	-	-
STransE [11]	206	0.657	93.4	-	-	69	0.543	79.7	-	-
ITransF [12]	205	-	94.2	-	-	65	-	81.0	-	-
DistMult [24]	-	0.83	94.2	-	-	-	0.35	57.7	-	-
ComplEx [25]	-	0.941	94.7	-	-	-	0.692	84.0	-	-
ConvE [22]	504	<b>0.942</b>	<b>95.5</b>	<b>94.7</b>	<b>93.5</b>	64	0.745	87.3	80.1	67.0
Proposed	<b>193</b>	0.363	72.6	43.3	19.6	106	0.286	50.2	32.2	17.7

表2 WN18 と FB15k データセットにおける知識ベース補完評価結果.

Method	WN18RR					FB15k-237				
			Hits					Hits		
	MR	MRR	@10	@3	@1	MR	MRR	@10	@3	@1
DistMult [24]	5110	0.425	49.1	43.9	38.9	254	0.241	41.9	26.3	15.5
ComplEx [25]	5261	<b>0.444</b>	<b>50.7</b>	<b>45.8</b>	<b>41.1</b>	<b>248</b>	0.240	41.9	26.3	15.2
ConvE [22]	7323	0.342	41.1	36.0	30.6	330	<b>0.301</b>	<b>45.8</b>	<b>33.0</b>	<b>22.0</b>
Proposed	<b>1943</b>	0.197	45.6	26.3	6.7	261	0.216	38.4	21.6	13.5

表3 WN18RR と FB15k-237 データセットにおける知識ベース補完評価結果.

提案モデル	-autoencoder		
関係	類似度	関係	類似度
honored_for	0.416	honored_for	0.355
type_of_union	0.281	place	0.318
currency	0.266	educational_institution	0.316
time_zones	0.257	campuses	0.315

表4 FB15k において関係 nominated\_for に対応する行列と類似度が高い関係. -autoencoder はオートエンコーダとの同時学習を行わないモデルを表す. 類似度の指標にはコサイン類似度を用いた.

提案モデル	-autoencoder		
関係	類似度	関係	類似度
campuses	0.302	educational_institution	0.485
educational_institution	0.302	campuses	0.477
place	0.297	place	0.472
degree	0.276	team	0.421

表5 FB15k において関係 spouse に対応する行列と類似度が高い関係.

- Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA.*, pages 2181–2187, 2015.
- [11] Dat Quoc Nguyen, Kairit Sirts, Lizhen Qu, and Mark Johnson. Stranese: a novel embedding model of entities and relationships in knowledge bases. In *NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016*, pages 460–466, 2016.
- [12] Qizhe Xie, Xuezhe Ma, Zihang Dai, and Eduard H. Hovy. An interpretable knowledge transfer model for knowledge base completion. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long*

- Papers*, pages 950–962, 2017.
- [13] Kelvin Guu, John Miller, and Percy Liang. Traversing knowledge graphs in vector space. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015*, pages 318–327, 2015.
- [14] Yankai Lin, Zhiyuan Liu, Huan-Bo Luan, Maosong Sun, Siwei Rao, and Song Liu. Modeling relation paths for representation learning of knowledge bases. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015*, pages 705–714, 2015.
- [15] Sebastian Riedel, Limin Yao, Andrew McCallum, and Benjamin M. Marlin. Relation extraction with matrix factorization and universal schemas. In *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 9-14, 2013, Westin Peachtree Plaza Hotel, Atlanta, Georgia, USA*, pages 74–84, 2013.
- [16] Dumitru Erhan, Yoshua Bengio, Aaron Courville, Pierre-Antoine Manzagol, Pascal Vincent, and Samy Bengio. Why does unsupervised pre-training help deep learning? *J. Mach. Learn. Res.*, 11:625–660, March 2010.
- [17] Richard Socher, Jeffrey Pennington, Eric H. Huang, Andrew Y. Ng, and Christopher D. Manning. Semi-supervised recursive autoencoders for predicting sentiment distributions. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 151–161, Edinburgh, Scotland, UK., July 2011. Association for Computational Linguistics.
- [18] Carina Silberer and Mirella Lapata. Learning grounded meaning representations with autoencoders. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 721–732, Baltimore, Maryland, June 2014. Association for Computational Linguistics.
- [19] Ivan Titov and Ehsan Khoddam. Unsupervised induction of semantic roles within a reconstruction-error minimization framework. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1–10, Denver, Colorado, May–June 2015. Association for Computational Linguistics.
- [20] Ran Tian, Naoaki Okazaki, and Kentaro Inui. Learning semantically and additively compositional distributional representations. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1277–1287, Berlin, Germany, August 2016. Association for Computational Linguistics.
- [21] Kristina Toutanova and Danqi Chen. Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality*, pages 57–66, 2015.
- [22] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2D Knowledge Graph Embeddings. In *Advances in Neural Information Processing Systems 30*, 2017.
- [23] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada.*, pages 1112–1119, 2014.
- [24] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding Entities and Relations for Learning and Inference in Knowledge Bases. In *Proceedings of the 3rd International Conference on Learning Representations*, pages 1–12, 2015.
- [25] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In *Proceedings of the 33rd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, pages 2071–2080, 2016.