知識ベースとテキストの構成的同時学習

高橋 諒 [1,2]　井之上 直也 [1,2]　谷中 瞳 [2,1]　乾 健太郎 [1,2]
[1] 東北大学　[2] 理化学研究所
{ryo.t,naoya-i,inui}@ecei.tohoku.ac.jp　hitomi.yanaka@riken.jp

1 はじめに

知識ベース補完は，ある知識ベースにおける既知の事実から未知の事実を予測することを目的としたタスクであり，質問応答システム [12, 19] や生物医学ドメインにおける知識獲得 [3] などへの応用が進んでいる．Wikidata [16]，Freebase [1]，UMLS [8] などの典型的な知識ベースでは，実世界の事実を〈ヘッドエンティティ，関係，テールエンティティ〉の三つ組の形式で格納する（例：⟨*Chainer*, *developer*, *PFN*⟩）．このもとでは，知識ベース補完は ⟨*Chainer*, *used_for*, *?*⟩ のように欠けたエンティティを予測する問題として表現される．

既存研究の多くは知識ベースにおける推論を低次元ベクトル空間上の演算としてモデル化し，エンティティや関係の埋め込みを三つ組の集合から学習する [17]．このような知識ベース埋め込みのモデルは，知識ベースの局所的に密な部分を取り出した人工的なベンチマークデータセットで高い性能を示しているが，実世界の知識ベースに見られるような疎な状況での性能が限定的であることが知られている [11]．

より現実的な状況を想定して，知識ベース埋め込みの学習に追加の情報を統合する研究が進んでいる．一つは複数の関係の系列(**関係パス**)を考慮し,構成的な推論を可能にすることである [7, 14]. 例えば, ⟨*Tensorflow*, *developer*, *Google Brain*⟩ と ⟨*Google Brain*, *field_of_work*, *Machine Learning*⟩ から ⟨*Tensorflow*, *used_for*, *Machine Learning*⟩ を推論するために，関係のパス *developer/field_of_work* の埋め込みを *used_for* の埋め込みに近づくように学習する．もう一つはテキスト情報との統合である [13, 15, 18]．例えば，*developer* という関係がテキスト上に “*A is developed by B.*” として現れやすいことを捉えられれば，⟨*Chainer*, *developer*, *PFN*⟩ が知識ベースに格納されていない場合でも，“*Chainer is developed by PFN.*” と書かれたテキストを手がかりとして上の三つ組と組み合わせることで，⟨*Chainer*, *used_for*, *Machine Learning*⟩ を推論できる．これらの二つの方向性はこれまでの研究では主に独立に研究されてきたが，相補的であり，知識ベースの関係知識とテキストから抽出した関係知識を組み合わせることによって，推論のさらなる高度化の可能性が考えられる．

そこで本研究では，知識ベース埋め込みの学習において，関係パスの学習とテキスト情報の統合を同時に行うモデルを提案する．実験では，Wikidata に基づく新たなベンチマークデータセットを構築し，小規模ながら Wikipedia のテキスト情報を統合することで，知識ベース補完における予測性能の改善の可能性を示す．なお，提案手法のソースコードと実験に用いたデータセットは研究利用可能な形で公開予定である[*1]．

2 知識ベース補完

本研究では，知識ベースとしてエンティティ h, r と関係 r からなる三つ組 $\langle h, r, t\rangle$（例：⟨*Chainer*, *developer*, *PFN*⟩）の集合 $\mathcal{T}$ を考える．知識ベースはエンティティをノードとし，関係をエッジとした有向多重グラフともみなせる．知識ベース補完 (Knowledge Base Complition) とは，一部の三つ組が欠けた知識ベースが訓練データとして与えられたとき，訓練時には出現しない三つ組を補完することを目的としたタスクである．

このタスクに取り組む先行研究の多くは，三つ組の事実らしさを測るスコア関数を訓練する．例えば，最も広く知られているモデルの一つである TransE [2] は，2 つのエンティティ（以下，それぞれヘッドエンティティ，テールエンティティと呼ぶ）h, t と関係 r をそれぞれ d 次元ベクトル $\boldsymbol{h}, \boldsymbol{t}, \boldsymbol{r} \in \mathbb{R}^d$ で表現し，スコア関数

$$f(h, r, t; \Theta) := -\|\boldsymbol{h} + \boldsymbol{r} - \boldsymbol{t}\| \tag{1}$$

を最大化するように埋め込みを学習する．

2.1 知識ベース埋め込みの構成的学習

より一般に，知識ベース埋め込みの学習では，関係 $r_1 \ldots r_l$ の合成 $r_1/\ldots/r_l$ を考慮することができる [7, 14]．あるエンティティ（ノード）h から出発し，知識グラフ上を l 回遷移した後に辿り着くエンティティを t と

[*1] https://github.com/reiyw/joint-text-kg

すると，TransE のスコア関数は

$$f(h, r_1/\ldots/r_l, t; \Theta) := -\|\boldsymbol{h} + \boldsymbol{r_1} + \cdots + \boldsymbol{r_l} - \boldsymbol{t}\| \quad (2)$$

のように拡張できる．このような知識グラフからサンプリングされたパスに対するスコア関数の最適化 (以下，compositional training) を行うことによって，知識グラフに内在する関係間の制約をデータから自動的に獲得できることが知られている [14]．例えば，知識ベース上で関係 *used_for* は統計的に *developer* と *field_of_work* を合成した関係とヘッドエンティティとテールエンティティを共有しやすい．この場合，*used_for* の関係埋め込み $\boldsymbol{r}_{used\_for}$ は *developer* と *field_of_work* を合成したベクトル $\boldsymbol{r}_{developer} + \boldsymbol{r}_{field\_of\_work}$ と近づくような力が働く ($\boldsymbol{r}_{developer} + \boldsymbol{r}_{field\_of\_work} \approx \boldsymbol{r}_{used\_for}$)．これは同時に，エンティティ埋め込みの空間的な配置に関する正則化をかけていることに相当する [7]．

2.2 テキストとの同時学習

クリーンだがスパースな知識ベースに欠けた情報を，ノイジーだが大規模なテキスト集合によって補うことで，新しい事実を推論する能力を強化することを目指して，知識ベース埋め込みの学習時にテキストの情報を統合する手法は活発に研究されている [13, 15, 18]．

知識ベース埋め込みの学習にテキスト情報を統合する際に重要となるアイデアは「テキスト中に出現するメンションペアは，どのような関係性をもって現れているか」を解決することにある．

メンション間の関係を記述するテキスト上の表現を textual relation と呼ぶことにする．例えば，*"Chainer is developed by PFN."* という文が与えられたとき，ここでの *Chainer* と *PFN* との間の関係は textual relation *"is developed by"* によって規定されていると考えられる．textual relation の具体的な表現形式については様々な議論がなされてきているが，主に関係抽出の分野では単に「メンション間の単語列」や「メンション間を結ぶ係り受けパス」などが広く用いられてきている [10]．

Toutanova [15] らは知識ベース埋め込みの学習データにテキストコーパスから抽出されたメンションペアと textual relation からなる三つ組（以下，textual triple）を追加し，知識ベース埋め込みの学習と同時に textual relation の表現を学習した．Toutanova らはまた，メンションペアを共有するような textual relation には共通の部分構造が存在することを見出した[*2]．メンションの埋め込みをエンティティの埋め込みと共有し，textual relation の部分構造を CNN を通して学習することによって，知識ベース補完の予測性能を改善することを示した．

本研究はこの同時学習の手法に compositional training を組み合わせることで，エンティティ埋め込みの空間的な配置に関する正則化をかけると同時に，textual relation の部分構造や知識ベースの関係へのマッピングを学習することを試みる．本研究のモデルと最も近いモデルとして Das [4] らのモデルが挙げられる．Das らのモデルも compositional training のアイデアと textual relation の学習を組み合わせたモデルであるが，関係の合成が非線形変換として表現されてしまうため，知識ベースの構造を反映したエンティティ埋め込みの配置を学習できないと考えられる．その結果，構成的な推論を行うために，知識ベースより関係パスを明示的にサンプルする必要があることに対し，我々は，関係の構成性を満たすようにエンティティの埋め込み空間を学習することで，推論時の関係パスのサンプリングが不要であるという点で異なる．

3 Universal Graph

本研究では，知識ベースとテキストの同時学習の手法 [15] と，compositional training とを組み合わせたモデルを構築する．具体的にはまず，textual triple と KG triple とを統合した知識グラフ（以下，Universal Graph; UG）を構築する．そして，UG からサンプリングされるパスに基づいてモデルを学習する手法を提案する．

UG は知識ベースの三つ組 $\mathcal{T}_{\mathrm{KG}} = \{\langle h, r_{\mathrm{KG}}, t\rangle | h, t \in \mathcal{E}, r_{\mathrm{KG}} \in \mathcal{R}_{\mathrm{KG}}\}$ と textual triples $\mathcal{T}_{\mathrm{text}} = \{\langle h, r_{\mathrm{text}}, t\rangle | h, t \in \mathcal{E}, r_{\mathrm{text}} \in \mathcal{R}_{\mathrm{text}}\}$ の和集合からなる．ここで，$r_{\mathrm{KG}} \in \mathcal{R}_{\mathrm{KG}}$ は知識ベース上の関係，$r_{\mathrm{text}} \in \mathcal{R}_{\mathrm{text}}$ はテキストから抽出された関係である．さらに，r_{text} は単語列 $w_1, \ldots, w_{|r_{\mathrm{text}}|}$ からなる．UG からサンプリングするパス p には知識ベース上の関係とテキストから抽出された関係の両方が含まれうる：

$$p = r_1/\ldots/r_l, \quad r_1, \ldots, r_l \in \mathcal{R}_{\mathrm{KG}} \cup \mathcal{R}_{\mathrm{text}} \quad (3)$$

4 モデル

本研究では知識ベース埋め込みの最も基本的なモデルである TransE [2] から出発する．compositional

[*2] Toutanova らは textual relation としてメンション間の係り受けパスを採用したが，メンション間の単語列についても同様の議論が成り立つと考えられる．

	$\lvert\mathcal{E}\rvert$	$\lvert\mathcal{R}\rvert$	# 訓練	# 検証	# 評価
$\mathcal{T}_{\text{KG}}$([15])	14,541	237	272,115	17,535	20,466
$\mathcal{T}_{\text{text}}$([15])	13,937	2,740,000	3,978,000	0	0
$\mathcal{T}_{\text{KG}}$（本研究）	41,714	616	179,296	22,412	22,412
$\mathcal{T}_{\text{text}}$（本研究）	41,352	152,143	181,921	0	0

表1: データセットの統計量

training の枠組みに乗せるために，スコア関数として式 (2) を採用する．UG からサンプリングするパスには異なる種類の関係が含まれるため，次式にしたがって入力の r が知識ベース由来かテキスト由来かに応じて，固定長のベクトル $\boldsymbol{r} \in \mathbb{R}^d$ を計算する方法を変える：

$$\boldsymbol{r} = \begin{cases} \boldsymbol{R}(r), & r \in \mathcal{R}_{\text{KG}} \\ \text{Encoder}\,(r), & r \in \mathcal{R}_{\text{text}} \end{cases} \quad (4)$$

ここで，$\boldsymbol{R}$ は関係の埋め込みの lookup 関数，Encoder は任意の文エンコーダである．本稿では Encoder として k 層 BiLSTM を利用したが，BERT [5] などの事前学習済み言語モデルに基づく高性能な文エンコーダを使うことも可能であり，これは今後の課題とする．

訓練 学習データは UG 上のランダムウォークによって生成される．すなわち，$\mathcal{E}$ からサンプルされたヘッドエンティティ h から出発し，適当なパス $p = r_1/\ldots/r_{|p|}$ を経てテールエンティティ t に辿り着くことを N 回繰り返して学習データ $\{(h_i, p_i, t_i)\}_{i=1}^{N}$ を生成する．本研究では次の損失関数を最小化する．

$$\mathcal{L}(\Theta) = \sum_{i=1}^{N} \sum_{t^* \in \mathcal{N}(h_i, p_i)} \max(0, [\gamma + f^- - f^+]) \quad (5)$$

$$f^- = f(h_i, p_i, t^*; \Theta), \quad f^+ = f(h_i, p_i, t_i; \Theta) \quad (6)$$

ここで，t^* は (h_i, p_i, t_i) に対する負例，γ はマージンを表す．知識ベース埋め込みの研究では様々な損失関数が提案されている [17] が，他の損失関数との比較検討は今後の課題とする．

5 実験

5.1 実験設定

データセット Universal Graph の構築には，知識ベースとそれにエンティティリンキングされたテキストコーパスが必要となる．本研究では，Wikipedia の 720 記事に Wikidata へのエンティティリンキング[*3] がかけられた Linked Wikitext-2 [9] を元に UG を構築した．前述の通り，UG は知識ベース $\mathcal{T}_{\text{KG}}$ と textual triples $\mathcal{T}_{\text{text}}$ からなる．まず，知識ベース $\mathcal{T}_{\text{KG}}$ として Wikidata のサブセットを収集する．Wikidata RDF を $\mathcal{T}_{\text{WD}}$，Linked Wikitext-2 に含まれるエンティティ集合を $\mathcal{E}$ としたとき，$\{\langle h, r, t\rangle \in \mathcal{T}_{\text{WD}} | h, t \in \mathcal{E}\}$ からなる三つ組を収集した．次に，textual triples $\mathcal{T}_{\text{text}}$ として，文区切りを跨がない全てのメンションペア (h, t) について，それらの間の単語列を r_{text} として $\langle h, r_{\text{text}}, t\rangle$ を構成した．

評価方法 知識ベース $\mathcal{T}_{\text{KG}}$ を 8:1:1 の比率で訓練･検証･評価セットに分割し評価を行った．textual triple $\mathcal{T}_{\text{text}}$ を訓練セットに追加したデータをモデルの訓練に用いた．textual triple を訓練にのみ用いるのは Toutanova ら [15] と同様の設定である．これは，テキスト情報を既存の知識ベースに追加することで，知識ベースに欠けた三つ組をどれだけ補完できるかを評価していることを意味する．なお，分割時には検証・評価セットに訓練セットで未知の語彙が出現しないようにした．データセットの統計情報を表 1 に示す．

評価方法は Bordes ら [2] の filtered 設定に従う．エンティティが欠けた三つ組 $\langle h, r, ?\rangle$ が与えられ，三つ組 $\langle h, r, e\rangle$ が訓練・検証・評価セットに含まれるようなエンティティ $e \in \mathcal{E}$ を除いた全ての e に対しスコア $f(h, r, e)$ を計算する．計算されたスコアはゴールドの三つ組に対するスコア $f(h, r, t)$ と共に順位に変換され，ゴールドのエンティティ t に対する予測順位に基づいて評価指標を求める．予測順位の逆数の平均 (Mean Reciprocal Rank; MRR) と上位 k 位に順位付けられたゴールドのエンティティの割合（Hits@k）を報告する．

比較対象として次の 4 つのモデルを取り上げる: (1) TransE: 文エンコーダを使わず，r_{text} を 1 トークンとみなして埋め込みを学習する，(2) TransE+COMP: TransE に compositional training を導入する（Guu ら [7] とほぼ同一の設定），(3) UGTransE: r_{text} に対して文エンコーダを使う（式 (4)），(4) UGTransE+COMP: UGTransE に compositional training を導入する．

ハイパーパラメタの探索には Allentune [6] を用いた．検証セットで MRR が最大となるモデルの評価セットにおける各指標を報告する．

5.2 実験結果

評価結果を表 2 に示す．TransE に COMP を加えたことによる性能向上より，知識ベースの関係知識のみを用いた知識ベース補完において，compositional training が有効であることが確認できた．また，TransE から UGTransE への性能向上より，テキストから抽出した

[*3] もともと Wikipedia に含まれるアンカーリンクと，neural-el と追加のヒューリスティックによるリンキングが含まれる．

	H1	H10	MRR
TransE ($\mathcal{T}_{\mathrm{KG}}$ only)	13.1	32.5	.198
TransE	12.8	34.1	.200
TrasnE + COMP	13.6	**34.9**	**.208**
UGTransE	**13.8**	34.4	.207
UGTransE + COMP	12.5	32.5	.193

表2: 知識ベース補完の性能

関係知識を知識ベースに統合することの有効性が確認できた．これらの結果は，文献 [7, 14]，文献 [15] の結果とそれぞれ一貫するものである．

しかしながら，UGTransE に COMP を加えた場合，すなわち推論の構成的学習とテキスト情報を統合した場合には，さらなる性能の向上はみられなかった．原因として，UGTransE+COMP 以外のモデルについては，Allentune においてハイパーパラメタの実現値を性能向上が収束するまで十分に試行できたことに対して，UGTransE+COMP についてはまだ性能向上が見込める段階にあり，ハイパーパラメタの実現値をさらに試行することによる性能向上の可能性が示唆されている．

6 おわりに：今後の課題

今後の課題として，今回は文エンコーダとして素朴なモデルを用いたが，大規模コーパスで事前訓練された最先端の文エンコーダ（BERT [5] など）を用いることが考えられる．また，学習されたエンティティ，および関係の埋め込み空間を分析し，所望の性質が得られているのかを検証する予定である．

謝辞

本研究は，JST, CREST, JPMJCR1301 の支援，及び JSPS 科研費 JP18J20936 の助成を受けたものである．

参考文献

[1] Kurt D. Bollacker et al. “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*. 2008, pp. 1247–1250.

[2] Antoine Bordes et al. “Translating Embeddings for Modeling Multi-relational Data”. In: *Advances in Neural Information Processing Systems 26*. Ed. by C. J. C. Burges et al. Curran Associates, Inc., 2013, pp. 2787–2795.

[3] Qin Dai et al. “Distantly Supervised Biomedical Knowledge Acquisition via Knowledge Graph Based Attention”. In: *Proceedings of the Workshop on Extracting Structured Knowledge from Scientific Publications*. Minneapolis, Minnesota: Association for Computational Linguistics, June 2019, pp. 1–10.

[4] Rajarshi Das et al. “Chains of Reasoning over Entities, Relations, and Text using Recurrent Neural Networks”. In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. Valencia, Spain: Association for Computational Linguistics, Apr. 2017, pp. 132–141.

[5] Jacob Devlin et al. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, June 2019, pp. 4171–4186.

[6] Jesse Dodge et al. “Show Your Work: Improved Reporting of Experimental Results”. 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)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 2185–2194.

[7] 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*. Lisbon, Portugal: Association for Computational Linguistics, Sept. 2015, pp. 318–327.

[8] Donald A. B. Lindberg, Betsy L. Humphreys, and Alexa T. McCray. “The Unified Medical Language System.” In: *Methods of information in medicine* 32 4 (1993), pp. 281–91.

[9] Robert L. Logan IV et al. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Florence, Italy: Association for Computational Linguistics, July 2019.

[10] Sachin Pawar, Girish Keshav Palshikar, and Pushpak Bhattacharyya. “Relation Extraction : A Survey”. In: *ArXiv* abs/1712.05191 (2017).

[11] Jay Pujara, Eriq Augustine, and Lise Getoor. “Sparsity and Noise: Where Knowledge Graph Embeddings Fall Short”. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Copenhagen, Denmark: Association for Computational Linguistics, Sept. 2017, pp. 1751–1756.

[12] Delai Qiu et al. “Machine Reading Comprehension Using Structural Knowledge Graph-aware Network”. 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)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 5895–5900.

[13] Sebastian Riedel et al. “Relation Extraction with Matrix Factorization and Universal Schemas”. In: *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Atlanta, Georgia: Association for Computational Linguistics, June 2013, pp. 74–84.

[14] Ryo Takahashi, Ran Tian, and Kentaro Inui. “Interpretable and Compositional Relation Learning by Joint Training with an Autoencoder”. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics, July 2018, pp. 2148–2159.

[15] Kristina Toutanova et al. “Representing Text for Joint Embedding of Text and Knowledge Bases”. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, Sept. 2015, pp. 1499–1509.

[16] Denny Vrandečiundefined and Markus Krötzsch. “Wikidata: A Free Collaborative Knowledgebase”. In: *Commun. ACM* 57.10 (Sept. 2014), pp. 78–85.

[17] Quan Wang et al. “Knowledge Graph Embedding: A Survey of Approaches and Applications”. In: *IEEE Trans. Knowl. Data Eng.* 29.12 (2017), pp. 2724–2743.

[18] Zhen Wang et al. “Knowledge Graph and Text Jointly Embedding”. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1591–1601.

[19] An Yang et al. “Enhancing Pre-Trained Language Representations with Rich Knowledge for Machine Reading Comprehension”. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, July 2019, pp. 2346–2357.