

# Knowledge Graph の Embedding 手法の実装と 2 種類のデータセットによる検証

## 1 はじめに

近年, 人工知能技術は急速な発展を遂げている. その中でも人間の知識をグラフ構造で表現する Knowledge Graph (KG) [1] が注目を集めており, 人工知能の基盤技術としてさまざまな分野で活用されている. 近年よく用いられる大規模言語モデルでは単語間の関係性を人間が理解しやすい形で可視化することが難しいため, 意味情報を理解しやすい形で表すことが困難であるが, 知識構造を表現する KG を利用することで単語の意味情報を人間にわかりやすい形で可視化できる可能性がある.

KG の作成方法として TransE [2] や ComplEx [3] などの多くの Embedding 手法がある. 今後, これらの既存の Embedding 手法に各ノードのもつ属性を反映させて表現できる Graph Embedding 手法を適用して実装することを考えている.

本研究では, Graph Embedding 手法を適用する手法の調査として, 2 種類のデータセットを用いて既存の Embedding 手法の精度を検証する.

## 2 要素技術

### 2.1 Knowledge Graph

KG [1] とは, さまざまな知識を体系的に連結し, その関係をグラフ構造で表した知識のネットワークのことである. 図 1 に KG の例を示す. KG は, head, tail を要素に持つ entity 集合と, その entity 間の関係を表する relation を要素にもつ relation 集合によって構成されており, 図 1 のように entity をノード, relation をエッジとする有向グラフとして表される. また, KG にはグラフ構造とは異なる表現方法として (head, relation, tail) と表す 3 組構造 (triple) がある.

### 2.2 Embedding 手法

KG の Embedding 手法とは 3 組の埋め込み表現をニューラルネットワークを用いて手法ごとの Score 関数を最小化するように学習するモデルである. 基本的な手順は以下のようになる.

1. 3 組をベクトル空間や複素数空間, ハイパボリック空間を用いて埋め込み表現に初期化

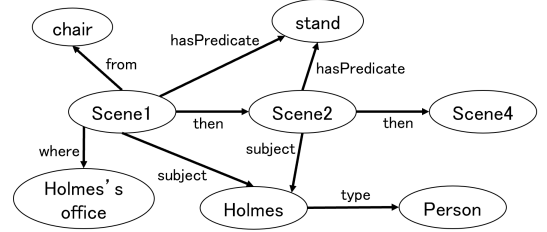


図 1: Knowledge Graph の例

2. Score 関数を最小化

3. head, relation が与えられたときの tail を予測

以下,  $\mathbf{v}_h, \mathbf{v}_r, \mathbf{v}_t \in \mathbb{R}^k$ ,  $\mathbf{w}_h, \mathbf{w}_r, \mathbf{w}_t \in \mathbb{C}^k$ ,  $\mathbf{e}_h, \mathbf{e}_t \in \mathbb{H}^k$  をそれぞれ head, relation, tail の埋め込み表現,  $b_h, b_t \in \mathbb{R}$  を head, tail におけるバイアスとする.

#### 2.2.1 TransE

KG の代表的な Embedding 手法として Translating Embedding (TransE) [2] が挙げられる. TransE では 3 組をベクトル空間に初期化し, (1) 式のユークリッド距離を表す Score 関数を最小化することで Embedding する.

$$\text{Score}(\mathbf{v}_h, \mathbf{v}_r, \mathbf{v}_t) = \|\mathbf{v}_h + \mathbf{v}_r - \mathbf{v}_t\| \quad (1)$$

#### 2.2.2 CP

Canonical Polyadic (CP) [4] はテンソル分解を挿入した Embedding 手法である. CP では, 3 組をベクトル空間に初期化し, (2) 式の  $S$  個の rank-1 テンソルの和, つまりテンソル分解を表す Score 関数を最小化することで Embedding する. なお,  $\otimes$  はテンソル積,  $\mathbf{u}$  は埋め込み表現であり,  $\mathbf{X} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$ ,  $s \in \{1, \dots, S\}$ ,  $\mathbf{u}_s^{(m)} \in \mathbb{R}^{N_m}$  とする.  $S$  は以下のテンソル分解が成立する最小の自然数である.

$$\mathbf{X} = \sum_{s=1}^S \mathbf{u}_{hs}^{(1)} \otimes \mathbf{u}_{rs}^{(2)} \otimes \mathbf{u}_{ts}^{(3)} \quad (2)$$

#### 2.2.3 MuRE

Multi-Relational Euclidean (MuRE) [5] ではまず 3 組をベクトル空間に初期化し, (3) 式の Score 関数を最小化することで Embedding する. なお,  $\mathbf{R} \in \mathbb{R}^{k \times k}$  を対角関係行列とする.

$$\text{Score}(\mathbf{v}_h, \mathbf{v}_r, \mathbf{v}_t) = \|\mathbf{R}\mathbf{v}_h - (\mathbf{v}_t + \mathbf{v}_r)\|^2 - b_h - b_t \quad (3)$$

### 2.2.4 RotatE, ComplEx

複素数空間を用いる Relational Rotation in complex space (RotatE) [6], Complex Embedding (ComplEx) [3] という Embedding 手法がある. RotatE は relation を回転として表現できるという特徴をもつ. どちらの手法もまず 3 つ組を複素数空間に初期化し, それぞれ (4), (5) 式の Score 関数を最小化することで Embedding する. なお,  $\circ$  はアダマール積,  $\bar{w}$  は複素共役であり,  $\langle \mathbf{a}, \mathbf{b}, \mathbf{c} \rangle = \sum_k a_k b_k c_k$  とする.

$$\text{RotatE} : \text{Score}(\mathbf{w}_h, \mathbf{w}_r, \mathbf{w}_t) = \|(\mathbf{w}_h \circ \mathbf{w}_r) - \mathbf{w}_t\| \quad (4)$$

$$\text{ComplEx} : \text{Score}(\mathbf{w}_h, \mathbf{w}_r, \mathbf{w}_t) = \text{Re}(\langle \mathbf{w}_h, \mathbf{w}_r, \bar{\mathbf{w}}_t \rangle) \quad (5)$$

### 2.2.5 RotE, RefE, AttE

Rotation-based Embedding (RotE) [7], Reflection-based Embedding (RefE) [7], RotE と RefE を組み合わせた Attention-based Embedding (AttE) [7] という Embedding 手法がある. これらの手法では 3 つ組をベクトル空間に初期化し, それぞれ (6), (7), (8) 式の Score 関数を最小化することで Embedding する.

$$\text{RotE} : \text{Score}(\mathbf{v}_h, \mathbf{v}_t) = \|\text{Rot}(\Theta_r)\mathbf{v}_h - \mathbf{v}_t\| - b_h - b_t \quad (6)$$

$$\text{RefE} : \text{Score}(\mathbf{v}_h, \mathbf{v}_t) = \|\text{Ref}(\Phi_r)\mathbf{v}_h - \mathbf{v}_t\| - b_h - b_t \quad (7)$$

$$\text{AttE} : \text{Score}(\mathbf{v}_h, \mathbf{v}_t) = \|\text{Att}(\mathbf{q}_{\text{Rot}}, \mathbf{q}_{\text{Ref}}; \mathbf{a}_r)\mathbf{v}_h - \mathbf{v}_t\| - b_h - b_t \quad (8)$$

このとき,  $\text{Rot}(\Theta_r)$ ,  $\text{Ref}(\Phi_r)$ ,  $\text{Att}(\mathbf{q}_{\text{Rot}}, \mathbf{q}_{\text{Ref}}; \mathbf{a}_r)$  はそれぞれ (9), (10), (12) 式で定義される.  $\Theta_r := (\theta_{r,i})_{i \in \{1, \dots, \frac{d}{2}\}}$ ,  $\Phi_r := (\phi_{r,i})_{i \in \{1, \dots, \frac{d}{2}\}}$  を relation 特有のパラメータ,  $d$  を embedding 次元数,  $\mathbf{a}_r$  を注意ベクトルとする.

$$\text{Rot}(\Theta_r) = \text{diag}(G^+(\theta_{r,1}), \dots, G^+(\theta_{r,\frac{d}{2}})) \quad (9)$$

$$\text{Ref}(\Phi_r) = \text{diag}(G^-(\theta_{r,1}), \dots, G^-(\theta_{r,\frac{d}{2}})) \quad (10)$$

$$G^\pm(\theta) := \begin{bmatrix} \cos(\theta) & \mp \sin(\theta) \\ \sin(\theta) & \pm \cos(\theta) \end{bmatrix} \quad (11)$$

$$\text{Att}(\mathbf{q}_{\text{Rot}}, \mathbf{q}_{\text{Ref}}; \mathbf{a}_r) = \alpha_{\mathbf{q}_{\text{Rot}}} \text{Rot}(\Theta_r) + \alpha_{\mathbf{q}_{\text{Ref}}} \text{Ref}(\Phi_r) \quad (12)$$

$$(\alpha_{\mathbf{q}_{\text{Rot}}}, \alpha_{\mathbf{q}_{\text{Ref}}}) = \text{Softmax}(\mathbf{a}_r^T \text{Rot}(\Theta_r)\mathbf{v}_h, \mathbf{a}_r^T \text{Ref}(\Phi_r)\mathbf{v}_h) \quad (13)$$

### 2.2.6 RotH, RefH, AttH

RotE, RefE, AttE をハイパボリック空間への拡張手法として Rotations in Hyperbolic space (RotH) [7], Reflections in Hyperbolic space (RefH) [7], Attention in Hyperbolic space (AttH) [7] という Embedding 手法がある. これらの手法では 3 つ組をハイパボリック空間に初期化し, それぞれ (14), (15), (16) 式の Score 関数を最小化することで Embedding する. なお,  $c$  は曲率である.

$$\text{RotH} : \text{Score}(\mathbf{e}_h, \mathbf{e}_t) = \frac{2}{\sqrt{c}} \tanh^{-1}(\sqrt{c} \|\text{Rot}(\Theta_r)\mathbf{e}_h \oplus^c \mathbf{e}_t\|) - b_h - b_t \quad (14)$$

$$\text{RefH} : \text{Score}(\mathbf{e}_h, \mathbf{e}_t) = \frac{2}{\sqrt{c}} \tanh^{-1}(\sqrt{c} \|\text{Ref}(\Phi_r)\mathbf{e}_h \oplus^c \mathbf{e}_t\|) - b_h - b_t \quad (15)$$

$$\text{AttH} : \text{Score}(\mathbf{e}_h, \mathbf{e}_t) = \frac{2}{\sqrt{c}} \tanh^{-1}(\sqrt{c} \|\text{Att}(\mathbf{q}_{\text{Rot}}, \mathbf{q}_{\text{Ref}}; \mathbf{a}_r)\mathbf{e}_h \oplus^c \mathbf{e}_t\|) - b_h - b_t \quad (16)$$

### 2.2.7 金田の手法

図 2, 3 にそれぞれ金田の手法 [8] のモデルの概略図, モデルを用いた学習の様子を示す. 図 2 のように KG の 3 つ組に対して Transformer によって埋め込み表現を作成し, それらを用いて未知の head, relation, tail を Multi Layer Perceptron (MLP) によって推測する. head と tail を区別するために head の埋め込み表現を MLP により非線形変換する. また, 図 3 のように埋め込み表現は, 3 つ組の一部の要素を [mask] に置き換え, その部分をモデルによって推論する中で学習する.

## 2.3 データセット

ホームズデータセット<sup>1</sup>とは, ホームズの推理小説における凶器や犯人を推定する「ナレッジグラフ推論チャレンジ」<sup>2</sup>というコンペティションで公開されているデータセットである. 本研究で用いたデータセットはそのうちの 1 つである A Case of Identity (花婿失踪事件) のデータセットである.

WN18RR とは, 英語の大規模な語彙データベースである WordNet から triple を自動抽出して得られたデータセットの WN18 において, head と tail を入れ替えて得られる逆関係の triple を除去したデー

<sup>1</sup><https://github.com/KnowledgeGraphJapan/KGRC-RDF/tree/ikgrc2023>

<sup>2</sup><https://challenge.knowledge-graph.jp/2021/>

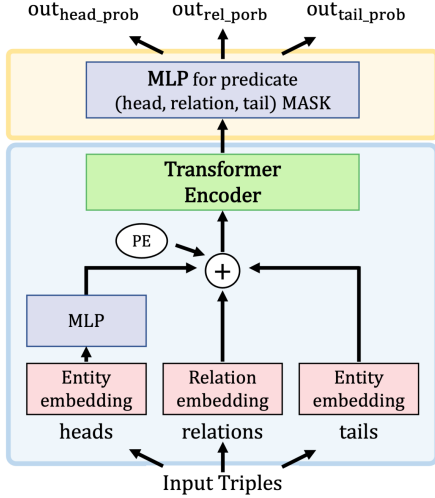


図 2: 金田の手法のモデル

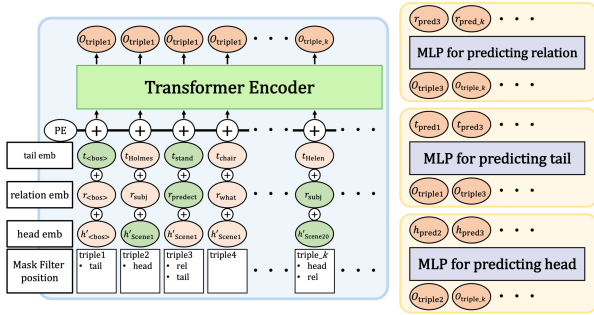


図 3: 金田の手法のモデルによる学習

タセットである. 表 1 に ホームズデータセットと WN18RR の例を示す. また, 表 2 にこの 2 つのデータセットにおける entity, relation, triple の数を示す.

### 3 実験

#### 3.1 評価指標

評価指標として Mean Reciprocal Rank (MRR) を用いた.  $|T|$  を triple 数,  $\text{rank}_i$  を適合要素のランクとすると, MRR は (17) 式で表される.

$$\text{MRR} = \frac{1}{|T|} \sum_{i=1}^{|T|} \frac{1}{\text{rank}_i} \quad (17)$$

さらに, 上位  $k$  個以内に正解の要素が出力されている割合を表す Hits@ $k$  ( $H@k$ ) を用いた. 本研究では  $k = 1, 3, 10$  で評価した. MMR,  $H@k$  はともに値が大きいとき推定精度が良いと判断される.

#### 3.2 実験概要

本研究では, 先に記述した既存の Embedding 手法と金田の手法に対してホームズデータセットと

表 1: ホームズデータセットと WN18RR の例

| head   | relation     | tail   |
|--------|--------------|--------|
| Scene1 | subject      | Holmes |
| Scene1 | hasPredicate | stand  |
| Scene1 | then         | Scene2 |
| Scene2 | subject      | Holmes |
| ...    | ...          | ...    |

| head          | relation                    | tail    |
|---------------|-----------------------------|---------|
| position      | hypernym                    | point   |
| rest          | derivationally related form | breath  |
| sophisticated | also see                    | worldly |
| brasov        | instance hypernym           | city    |
| ...           | ...                         | ...     |

表 2: データセットの entity, relation, triple の数

| データセット                      | entity | relation | triple |
|-----------------------------|--------|----------|--------|
| A Case of Identity (花婿失踪事件) | 4,414  | 34       | 4,434  |
| WN18RR                      | 40,943 | 11       | 93,003 |

WN18RR の triple を入力として tail を予測し, その精度を評価した. 2 種類のデータセットにおける triple を 8 : 1 : 1 に分割し, それぞれ訓練データ, 検証データ, テストデータとした. なお, ホームズデータセットのデータ数は 4,434, WN18RR のデータ数は 93,003 である. 表 3, 4 に 11 個の既存手法のパラメータ, 金田の手法のモデルと訓練のパラメータを示す.

### 4 実験結果

表 5, 6 に 11 個の既存手法の実験結果, 金田の手法の実験結果を示す. 表 5 より, 精度が最も高い手法は A Case of Identity において AttH, WN18RR において RotH であることが分かった. しかし, MRR の数値を見ると WN18RR のほうが大きくなっている. これは, A Case of Identity より WN18RR のほうがデータ数が大きいことが原因であると考えられる. また, 表 6 より, A Case of Identity において表 5 のどの手法よりも良い結果となったが, WN18RR においては表 5 のどの手法よりも低い結果となった. これは, 金田の手法は KG における周辺情報を考慮する手法であり, WN18RR において周辺情報が意味のないものであることが原因であると考えられる.

### 5 まとめと今後の課題

本研究では 11 個の既存 Embedding 手法と金田の手法に対して 2 種類のデータセットを用いてその精度を検証した. 11 個の既存手法はデータ量の違いにより精度が変化し, 金田の手法は周辺情報を含む KG に特化していることが分かった.

今後の課題として, 使用するデータの量をデータセットの結合などにより増加させること, 既存の Embedding 手法に対して Graph Embedding 手法を適用することが挙げられる. また, 大規模言語モデ

表 3: 11 個の既存手法のパラメータ

| パラメータ                  | TransE, CP           | RotatE               | RotH                 |
|------------------------|----------------------|----------------------|----------------------|
|                        | MuRE, RotE           | ComplEx              | RefH                 |
|                        | RefE, AttE           |                      | AttH                 |
| 最大 epoch               | 200                  | 500                  | 500                  |
| embedding 次元数          | 128                  | 128                  | 128                  |
| batch size             | 100                  | 100                  | 100                  |
| negative sampling size | 1000                 | -                    | 50                   |
| regularization weight  | 0.0                  | $5.0 \times 10^{-2}$ | $5.0 \times 10^{-2}$ |
| 学習率                    | $1.0 \times 10^{-3}$ | $1.0 \times 10^{-2}$ | $1.0 \times 10^{-3}$ |
| 最適化関数                  | Adam                 | Adagrad              | Adam                 |

表 4: 金田の手法のモデルと訓練のパラメータ

| モデルのパラメータ        |  | 訓練のパラメータ            | 値                             |
|------------------|--|---------------------|-------------------------------|
| 最大系列数            |  | 最大 epoch            | 500                           |
| embedding 次元数    |  | 学習率                 | $1.0 \times 10^{-4}$          |
| Transformer ヘッド数 |  | head の MLP の学習率     | $5.0 \times 10^{-5}$          |
| Transformer レイヤ数 |  | relation の MLP の学習率 | $1.0 \times 10^{-6}$          |
| 活性化関数            |  | tail の MLP の学習率     | $1.0 \times 10^{-6}$          |
| batch size       |  | MASK の変換確率          | 0.15                          |
|                  |  | MASK を元の単語に置き換える確率  | 0.80                          |
|                  |  | MASK をランダムに置き換える確率  | 0.10                          |
|                  |  | 損失関数                | Focal loss ( $\gamma = 2.5$ ) |

ルへの KG の組み込みと意味情報の可視化の手法についても検討していく。

## 参考文献

- [1] 川村隆浩, 江上周作, 田村光太郎, 外園康智, 鶴飼孝典, 小柳佑介, 西野文人, 岡嶋成司, 村上勝彦, 高松邦彦, 杉浦あおい, 白松俊, 張翔宇, 古崎晃司. 第 1 回ナレッジグラフ推論チャレンジ 2018 開催報告 —説明性のある人工知能システムを目指して—. 人工知能 34 巻 3 号, 2019.
- [2] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, Vol. 26. Curran Associates, Inc., 2013.
- [3] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Eric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International*

表 5: 11 個の既存手法の実験結果

|         | A Case of Identity |              |              |              | WN18RR       |              |              |              |
|---------|--------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
|         | MRR                | H@1          | H@3          | H@10         | MRR          | H@1          | H@3          | H@10         |
| TransE  | 0.241              | 0.173        | 0.288        | 0.345        | 0.192        | 0.029        | 0.305        | 0.483        |
| CP      | 0.200              | 0.140        | 0.243        | 0.297        | 0.368        | 0.358        | 0.371        | 0.385        |
| MuRE    | 0.259              | 0.198        | 0.286        | 0.340        | 0.451        | 0.392        | 0.478        | 0.559        |
| RotE    | 0.260              | 0.198        | 0.286        | 0.356        | 0.457        | 0.402        | 0.483        | 0.558        |
| RefE    | 0.267              | 0.203        | 0.293        | 0.372        | 0.443        | 0.391        | 0.466        | 0.540        |
| AttE    | 0.266              | 0.205        | 0.300        | 0.356        | 0.451        | 0.392        | 0.478        | 0.558        |
| RotatE  | 0.244              | 0.180        | 0.286        | 0.333        | 0.435        | 0.407        | 0.444        | 0.487        |
| ComplEx | 0.258              | 0.200        | 0.282        | 0.354        | 0.414        | 0.389        | 0.425        | 0.460        |
| RotH    | 0.269              | 0.200        | 0.304        | <b>0.381</b> | <b>0.486</b> | <b>0.441</b> | <b>0.503</b> | <b>0.571</b> |
| RefH    | 0.262              | 0.191        | 0.297        | 0.365        | 0.474        | 0.428        | 0.493        | 0.558        |
| AttH    | <b>0.278</b>       | <b>0.218</b> | <b>0.309</b> | 0.363        | 0.475        | 0.428        | 0.494        | 0.567        |

表 6: 金田の手法の実験結果

|          | A Case of Identity |       |       | WN18RR |        |        |
|----------|--------------------|-------|-------|--------|--------|--------|
|          | H@1                | H@3   | H@10  | H@1    | H@3    | H@10   |
| head     | 0.872              | 0.985 | 0.999 | 0.641  | 0.759  | 0.796  |
| relation | 0.929              | 0.975 | 0.998 | 0.481  | 0.961  | 0.999  |
| tail     | 0.475              | 0.577 | 0.676 | 0.0071 | 0.0136 | 0.0195 |

*Conference on Machine Learning*, Vol. 48 of *Proceedings of Machine Learning Research*, pp. 2071–2080, New York, New York, USA, 20–22 Jun 2016. PMLR.

- [4] Timothée Lacroix, Nicolas Usunier, and Guillaume Obozinski. Canonical tensor decomposition for knowledge base completion, 2018.
- [5] Ivana Balazevic, Carl Allen, and Timothy M. Hospedales. Multi-relational poincaré graph embeddings. *CoRR*, Vol. abs/1905.09791, , 2019.
- [6] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. *CoRR*, Vol. abs/1902.10197, , 2019.
- [7] Ines Chami, Adva Wolf, Da-Cheng Juan, Frederic Sala, Sujith Ravi, and Christopher Ré. Low-dimensional hyperbolic knowledge graph embeddings. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 6901–6914, Online, July 2020. Association for Computational Linguistics.
- [8] 金田燎弥. Story knowledge graph を用いた物語の内容推定モデルの提案と応用. 修士論文, 2023.