

## 報告書

### 1 金田さんの実験の実装

金田さんのコードを用いて 3 つの model に関して実験を回した. 表 1 に実験パラメータを示す.

表 1: 実験パラメータ

パラメータ	値
埋め込み次元	32
batch size	100
Regularization weight	0.05
最適化関数	Adagrad
学習率	0.1

表 2 にそれぞれのデータセットのエンティティ数とリレーション数を示す.

WN18RR とは, WordNet のサブセットである WN18 から作成されたリンク予測のデータセットである. WN18 は 18 のリレーションと 40,943 のエンティティから構成されており, 多くのテキストトリプルにおいてトレーニングセットからトリプルを逆転させることで得られる. これにより, WN18RR データセットは評価データセットに逆関係のテスト漏洩がないことを確保するために作成されたものである.

FB237 とは, FB15k から作成されたリンク予測のデータセットである. FB15k には 1,345 のリレーション, 14,951 のエンティティ, および 592,213 のトリプルが含まれている. しかし, 多くのトリプルが逆関係であり, トレーニングからテストや検証に漏洩が生じる. そこで, FB15k-237 は, Toutanova と Chen (2015) によって作成され, テストと評価のデータセットに逆関係のテスト漏洩がないことを保証するために作成された.

表 2: データセット

データセット	エンティティ数	リレーション数
WN18RR	40,943	11
FB237	14,505	237

MRR (Mean Reciprocal Rank) とは, 情報検索システムの評価指標の 1 つである. MRR では, ある検索結果において最初に見つかった適合文書のランクの逆数をスコアとしており, こうして得たスコアの平均をとったものが MRR 値となる. つまり, MRR 値が大きいほど良い結果であることが示される.

また, H@k とは, 予測したエンティティを順位付けしたときに, 上位 N 個以内に正解が含まれている割合を示す. 例えば「H@3」において, 推定したエンティティと正解のエンティティとの距離が 3 未満であるデータの割合を表している. つまり, 「H@1」はエンティティの推定が成功したデータの割合となる. なお, H@1 に含まれるデータは H@3 にも含まれている.

表 3 は model (ComplEx, RotE, RotH) におけるそれぞれの結果である.

ComplEx [1] モデルとは, ナレッジグラフ埋め込み手法の一つであり, 複素数を用いて埋め込みをする. ナレッジグラフの head と relation, tail の複素数空間上の表現ベクトルをそれぞれ  $v_h, v_r, v_t$  としたとき, 正例の 3 つ組では,  $v_h + v_r = v_t$  の関係が成り立つように学習する. 1 なお, regularizer において N3 を使用している.

RotE [2] モデルとは, オイラーの定理を用いてリレーションを回転で表現するように学習するモデルのことである.

RotH [3] モデルとは,

表 3: test 結果

model	dataset	epoch	MRR	H@1	H@3	H@10
ComplEx-N3	WN18RR	100	0.407	0.384	0.413	0.448
RotE	WN18RR	75	0.455	0.416	0.471	0.528
RotH	FB237	160	0.314	0.225	0.344	0.494

## 感想

ソースコードを見てパラメータや内容の理解することが難しい。

まだ RotH モデルや RotE モデルの詳しいことについては調べられていない。

## 参考文献

- [1] 鵜飼孝典. Wikidata を利用したナレッジグラフ埋め込みの精度向上. 人工知能学会第二種研究会資料, Vol. 2021, No. SWO-055, p. 07, 2021.
- [2] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. Rotate: Knowledge graph embedding by relational rotation in complex space, 2019.
- [3] 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.