

報告書

1 金田さんの実験

回そうとしたが、環境構築の時点でエラーが出て回せなかった。いろいろ試してみたけど無理だったので、回すのにまだ時間がかかりそう。

2 ホームズデータセットの実験

2.1 パラメータ

表 1: 実験パラメータ

パラメータ	値
最大 epoch 数	200
埋め込み次元	128
batch size	100
negative sampling size	1000
Regularization weight	0.0
margin γ	0.0
学習率	0.001

2.2 データセット

2.2.1 KG Challenge

ナレッジグラフ推論チャレンジにおいて公開されているデータセット。シャーロック・ホームズの小説の内容を用いている。

- A Case of Identity (花婿失踪事件)
- Abbey Grange (僧坊荘園)
- Crooked Man (背中の曲がった男) など

2.2.2 WN18RR

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。

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	トリプル数
A Case of Identity (花婿失踪事件)	4,414	34	4,434
WN18RR	40,943	11	93,003

2.3 結果

WN18RR, ホームズデータセットについて, 11 の model を用いて実験をする.

2.3.1 MRR (Mean Reciprocal Rank)

情報検索システムの評価指標の 1 つ. 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.

$$MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rank_i}$$

$|Q|$ は文書数, $rank_i$ は適合文書のランクを表している.

2.3.2 H@k

Hit k のこと. 予測したエンティティを順位付けしたときに, 上位 k 個以内に正解が含まれている割合を指す. 例えば「H@3」において, 推定したエンティティのうち上位 3 個以内に正解が含まれている割合を表している.

2.3.3 結果 1 (学習率 0.001)

前回の結果と同様.

表 3: 結果 (データセット: WM18RR)

model	MRR	H@1	H@3	H@10
TransE	0.192	0.029	0.305	0.483
CP	0.368	0.358	0.371	0.385
MurE	0.451	0.392	0.478	0.559
RotE	0.457	0.402	0.483	0.558
RefE	0.443	0.391	0.466	0.540
AttE	0.451	0.392	0.478	0.558
RotatE	0.000	0.000	0.000	0.000
ComplEx	0.000	0.000	0.000	0.000
RotH	0.028	0.018	0.029	0.047
RefH	0.025	0.015	0.025	0.043
AttH	0.025	0.016	0.025	0.044

2.3.4 下 5 つのモデルの結果

モデルごとにパラメータを変えて実行した.

表 4: モデルごとのパラメータ

パラメータ	RotatE	ComplEx	RotH	RefH	AttH
最大 epoch 数	500	500	500	500	500
埋め込み次元	128	128	128	128	128
batch size	100	100	100	100	100
negative sampling size	-1	-1	50	50	50
Regularization weight	0.05	0.05	0.0	0.0	0.0
学習率	0.01	0.1	0.0005	0.001	0.001

表 5: 結果（データセット：A Case of Identity）

model	MRR	H@1	H@3	H@10
RotatE[1]	0.244	0.180	0.286	0.333
ComplEx[2]	0.258	0.200	0.282	0.354
RotH[3]	0.269	0.200	0.304	0.381
RefH[3]	0.262	0.191	0.297	0.365
AttH[3]	0.278	0.218	0.309	0.363

参考文献

- [1] 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.
- [2] 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 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.
- [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.