

報告書

1 今週の進捗

- 研究会のまとめ
- 学習がされているか

2 研究会のまとめ

2.1 スライドについて

- v_w と $v_{w'}$ で区別が必要
- (A, B, C), (A, B, C') が訓練データ, (A, B, C'') がテストデータということを分かりやすく説明

2.2 研究内容について

- ランクの平均が評価指標として必要
Mean Rank (MR) という指標はあるが, 今回は 100 位までしか予測結果を出していないため MR を使用しなかった. しかし, 100 位まで出力した予測結果だけの MR を求めることはできる.
- 知識自体の意味情報なしとは TransE と BERT でどう違うのか
(A, B, C) と (A, B, C') があつたとき, C と C' の単語は似ていると考えられる. TransE では単語をベクトルに変換したあとの学習でそのベクトルが変化するため, 元々の単語の情報が失われる. そのため, C と C' の元のベクトルは似たベクトルであっても似ていないベクトルに変化する. BERT では C と C' は似た分散表現をとり変化することがないため, 単語の情報を保持したまま予測が可能となる.
- 訓練データ (A, B, C), (A, B, C') に対してテストデータ (A, B, C'') の C'' を予測できるのか (難しいタスク?) BERT では C, C', C'' は似た単語で似た分散表現をとると考えられるため, 予測できてほしい. それぞれのエンティティの分散表現を確認する必要がある.
 - Hits@1 が 13% もあることに疑問
 - Filtered が上がっていないのがおかしい
 - 学習がきちんとされているか

3 学習がされているか

3.1 訓練データ

訓練データをテストデータとしてテストした結果と, 同じ学習モデルを用いてテストデータでテストした結果を示す.

訓練データをテストデータとしたときの結果から学習が進んでいないことがわかった. epoch 数を増やしても同様の結果が得られた.

表 1: 訓練データをテストデータとしてテストした結果

モデル	WN18RR						
	MR	MR 100	MRR	Hits@1	Hits@3	Hits@10	Hits@100
訓練	20397	11.15	0.19896	0.130	0.236	0.325	0.440
訓練 (Fillter)	20396	10.99	0.20298	0.133	0.240	0.327	0.440
テスト	21418	12.53	0.18877	0.125	0.220	0.301	0.419
テスト (Fillter)	21418	12.42	0.19090	0.127	0.223	0.302	0.419

4 今後したいこと

- MLM を用いた実験の改良
- ナレッジグラフ推論チャレンジのデータセットの適用

参考文献