

報告書

1 今週の進捗

- 卒業研究の発表順について
- 卒業論文
- 学習時のマスク位置固定

2 卒業論文

先輩の修論, 卒論を参考にして書いている.

要素技術は詳しく書くため, 多くなってしまうがいいのか. BERT の説明に Transformer を使うため Transformer の説明が, Transformer の説明に Attention 機構を使うため Attention 機構の説明が必要になるが, それでいいのか.

提案手法は「MLM 学習時のマスク位置を一部固定したモデル」を書こうとしている.

3 学習時のマスク位置固定

MLM のマスク位置を tail の見出し語に固定して学習した結果を示す. tail の見出し語のみをマスクして学習する手法と, tail の見出し語を必ずマスクした状態で他のトークンを通常の MLM と同様に 15% の確率でマスクして学習する手法の 2 パターンの結果を示す. また, それぞれの手法においてテスト時の入力を tail の見出し語とその説明文の場合と, tail の見出し語のみの場合の 2 パターンにおいて示す. なお, MR (Mean Rank) は 300 個以内に予測されたもののみを用いて出した値である.

表 1: パラメータ

parameters	値
learning rate	5e-5
mask probability	0.15
batch size	32
max seq length	128
epoch	5, 20

MR は 300 位以内に予測結果が出たトリプルの個数がないため比較しにくい. そのトリプルの個数を確認しておく.

見出し語のみをマスクして学習後に見出し語とその説明文を入力としたテスト結果の MRR, Hits@1, Hits@3 は KG-BERT の文献値より良い結果が得られた.

見出し語のみをマスクして学習後に見出し語のみを入力としたテスト結果において KG-BERT を超える結果は得られなかった.

表 2: MLM を用いた実験結果

モデル	WN18RR				
	MR	MRR	Hits@1	Hits@3	Hits@10
KG-BERT [1] (文献値)	97	-	-	-	52.4
KG-BERT (再現実験)	117.77	0.25	12.41	29.44	51.85
見出し語のみマスク (見出し語とその説明文)	8.27	0.43	39.73	45.40	47.86
見出し語のみマスク (見出し語のみ)	39.31	0.13	9.54	15.34	19.87
他のトークン 15% (見出し語とその説明文)					
他のトークン 15% (見出し語のみ)					

参考文献

- [1] Liang Yao, Chengsheng Mao, and Yuan Luo. KG-BERT: BERT for knowledge graph completion. *CoRR*, Vol. abs/1909.03193, , 2019.