

マルチホップ推論に向けた機械読解の精度向上

福迫 太一 古川 翔大
鹿児島工業高等専門学校

1. はじめに

近年、自然言語処理分野において既存の文章を知識源とした質問応答システムを構築する試みが行われている。多段落にまたがった読解，すなわちマルチホップ推論を課題とする HotpotQA データセット [1] では，Clark と Gardner により提案された読解モデル [2] を再構成したものがベースラインモデルとして設定されている。本研究ではマルチホップ推論を要する機械読解の精度向上について検討する。具体的には，ベースラインモデルにおける RNN を LSTM に置き換え，その精度を評価する。

2. ベースラインモデル

Clark と Gardner による読解モデルは，ソース文と質問を入力し質問の回答に該当する範囲の先頭と末尾の単語を出力する。HotpotQA で再構成されたモデルではこの出力に加え，読解の根拠となる文と質問の種類の出力を行う。Attention の前後や出力層内部で RNN が用いられており，一定の長さの文章への対応が可能となっている。しかしながら，長すぎる文章では単語間の依存関係が適切に学習できなくなるという問題点が存在する。

3. 提案手法

本研究では，モデル内に配置される RNN の一部を LSTM に置き換えることにより，推論の精度の向上を図る。入力層，中間層，出力層の3箇所を RNN，LSTM を組み替え，計8通りの組み合わせについて推論精度を検討する。

4. 実験結果

表1の結果より，最も高い精度が得られた組み合わせは出力層のみに LSTM を使用したモデルである。また全ての箇所に LSTM が設定されたモデルでは推論精度の低下がみられた。一方で，図1のように長文読解が必要な問題などでは，最低限の数の文から必要な情報を抜き出した事例もみられた。

5. おわりに

機械読解モデルに LSTM を導入すると一文に対する読解能力が上昇するが，複数の文を参照する能力が低下するという特徴がある。本実験で最も推論の精度が得られたモデルでは，長文読解力と複数文の参照能力のバランスによる精度向上がみられたものと考えられる。今後は LSTM で複数文の参照ができるようなモデル実装を

表1 各組み合わせによるモデルの推論精度

※ em: 完全一致した回答の精度

f1: 部分一致した回答の精度

入力層	中間層	出力層	em	f1	sp_em	sp_f1	joint_em	joint_f1
RNN	RNN	RNN	46%	61%	23%	67%	12%	43%
RNN	RNN	LSTM	48%	62%	25%	69%	8%	33%
RNN	LSTM	RNN	43%	57%	20%	64%	8%	37%
RNN	LSTM	LSTM	41%	55%	20%	63%	14%	45%
LSTM	RNN	RNN	43%	56%	18%	61%	10%	39%
LSTM	RNN	LSTM	43%	57%	20%	65%	9%	37%
LSTM	LSTM	RNN	43%	57%	16%	61%	9%	37%
LSTM	LSTM	LSTM	40%	54%	17%	57%	10%	39%

Q: What studio album did Kanye West record with Roc-A-Fella Records and soul singer Dwele?

正答:

根拠

[Power (Kanye West song)] 2 文目，

[My Beautiful Dark Twisted Fantasy] 1 文目

回答

My Beautiful Dark Twisted Fantasy

[RNN, RNN, RNN] の場合:

根拠

[Graduation (album)] 1 文目，

[The College Dropout] 1 文目，

[My Beautiful Dark Twisted Fantasy] 1 文目，

[Watch the Throne, 0, Power (Kanye West song)]

1 文目

回答

Graduation

[LSTM, LSTM, LSTM] の場合:

根拠

[My Beautiful Dark Twisted Fantasy] 1 文目

回答

My Beautiful Dark Twisted Fantasy

図1 長文読解における質問応答のイメージ

検討する予定である。

参考文献

- [1] Zhilin Yang et al. HotpotQA: A Dataset for Diverse, Explainable Multi-hop Question Answering. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 2369–2380. 2018.
- [2] Christopher Clark and Matt Gardner. Simple and Effective Multi-Paragraph Reading Comprehension. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, pages 845–855. 2018.