

# 記憶装置付きニューラルネットワークモデルによる文脈と構造化知識を用いた対話への取り組み

村山 友理<sup>†</sup>小林 一郎<sup>‡</sup><sup>†</sup>お茶の水女子大学大学院 人間文化創成科学研究科 理学専攻<sup>‡</sup>お茶の水女子大学 基幹研究院 自然科学系{<sup>†</sup>murayama.yuri, <sup>‡</sup>koba}@is.ocha.ac.jp

## 1 はじめに

現在、対話のアーキテクチャとして Seq2Seq[1] や Transformer[2] といった sequence to sequence なモデル (相手の発話を入力として応答となる発話を生成するモデル) が主流になっている。一方、より自然で知的な対話を行うには文脈理解や知識活用が必要であるが、そのための長期間におけるデータ保存の能力には限界があると議論されてきた。そこで、文脈情報などの長期の記憶を保持するために、End-To-End Memory Networks[3] や Differentiable Neural Computer (以下、DNC) [4] といった記憶装置付きニューラルネットワークモデルが提案されている。これらのモデルは記憶装置を付け加えたことにより従来のモデルに比べてより複雑な情報処理を行えるようになり、対話における文脈を踏まえた質問応答でも高い精度を実現している。本研究では、DNC を拡張し文脈と知識の両方を用いたモデルを提案する。一貫性があり、かつ大規模な知識を要求する質問とその答えて構成された会話のデータセットの一部を用いて実験を行なった。

## 2 Differentiable Neural Computer

従来のニューラルネットワークでは、変数やデータ構造の表現や、長期間におけるデータの保存に能力の限界があると議論されてきた。DNC は、外部メモリを用意し、それについて読み書きができるニューラルネットワークを構築することによって課題の解決を試みた。DNC の全体図を図 1 に示す。

DNC の外部メモリにはヘッドがついており、書き込みヘッドによってメモリの書き換え、読み出しヘッ

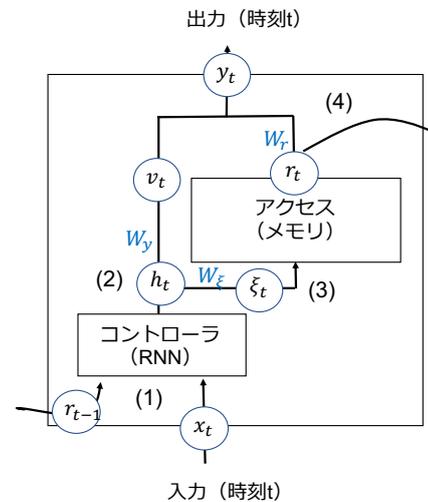


図 1: DNC の全体図

ドによってメモリの読み出しが行われる。1 タイムステップで行う処理の流れは以下の通りである。

1. DNC 内のコントローラ (RNN) は、毎時刻につき、入力  $x_t$  と前時刻でメモリから読み出した情報  $r_{t-1}$  を合わせて受け取り、 $h_t$  を出力する。
2.  $h_t$  を線形変換し、出力  $v_t = W_y h_t$  とメモリを制御するためのベクトル  $\xi_t = W_x h_t$  に分ける。
3.  $\xi_t$  によってメモリへの書き込みが行われ、メモリの状態が更新される。
4. メモリからの読み出しも行われ、読み出された情報  $r_t$  は RNN の出力に  $y_t = v_t + W_r r_t$  のように加算されると同時に、次時刻の RNN への入力に回される。

以上の処理を繰り返すことにより、メモリへの読み書き操作を取り込んだニューラルネットワークが実現さ

れる。メモリを付け加えたことにより、DNC はより複雑な情報処理を行えるようになり、文脈を踏まえた質問応答で高い精度を実現している。また、グラフ上の最短路探索問題やブロックパズルの問題など、様々なジャンルの問題を解くことができ、その汎用性の高さを期待させる。

### 3 提案手法

#### 3.1 提案モデル

DNC を拡張し知識を格納するメモリを追加することで、文脈と知識の両方を用いた応答の生成を目指す。提案モデルの全体図を図 2 に示す。

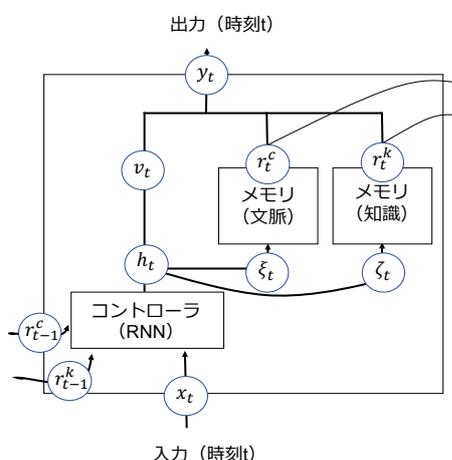


図 2: 提案モデルの全体図

DNC との主な違いは、コントローラへの入力に  $x_t$  と前時刻で文脈メモリから読み出した情報  $r_{t-1}^c$  に加えて前時刻で知識メモリから読み出した情報  $r_{t-1}^k$  を与える点と、コントローラからの出力  $v_t$  に文脈メモリから読み出された情報  $r_t^c$  と新たに知識メモリから読み出された情報  $r_t^k$  が合わさって  $y_t$  を生成する点である。知識メモリには構造化知識をあらかじめ格納したメモリを用い、書き換え操作は行わず、読み出し操作のみを行う。知識メモリの構築方法については後で述べる。

ここで、自然言語処理において知識を取り入れる際、形式的に記述された構造化知識は正しい結果を導くが、形式を重んじることから柔軟に利用することができないという問題がある。一方で、分散表現による知識表現では未知の知識項目に対しても柔軟に対応することができる。そこで、本研究では構造化知識を分散表現にして用いることで、正確に記述された知識を

ニューラルネットワークにおいて柔軟に活用することを考える。

また、知識を導入することの利点として以下の 3 点が挙げられる。

- 文脈情報からだけでは答えることができない、知識を問う質問に対して答えることができる。
- 曖昧な質問の内容を正確に特定することができる。
- 未知の語彙に対して柔軟に対応することができる。

#### 3.2 知識メモリの構築

本研究では知識として構造化知識、つまり、主語、述語、目的語の三つ組で表現されたデータ（例えば、「日本の首都は東京である。」は三つ組（日本、首都、東京）で表される）を扱う。DNC のモデルを用いて、三つ組の内 2 つを入力とし残りの 1 つを出力する（例えば、“日本”と“首都”を入力し“東京”を返す）ように学習し終わった後のメモリを知識メモリとして利用する。

## 4 実験

### 4.1 実験設定

Complex Sequential Question Answering データセット（以下、CSQA）[5] は一貫性があり、かつ大規模な知識を要求する質問とその答えで構成された会話のデータセットであり、以下の 4 つの能力を問うデータセットとなっている。

1. 複雑な自然言語質問文の解析
2. 発話における共参照と省略を解決するために文脈の利用
3. 曖昧な質問の明確化
4. 質問に答えるために知識グラフの関係するサブグラフの検索

約 20 万対話（合計 160 万ターン）から成り、知識には wikidata<sup>1</sup>を用いている。

今回の実験では、CSQA の会話データセットの内 264 対話（1 対話はおよそ 10~15 ターン）のみを用い

<sup>1</sup>[https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page)

た. wikidata については, 264 対話中に出現したエンティティとリレーションを少なくとも2つ含むトリプルを抽出し, その上で10回以上出現する語彙だけを含んだトリプルを選び, トリプルの総数は352,193, エンティティとリレーションの語彙数はそれぞれ59,721, 224 となった.

また, wikidata の embedding には TransE[6] を, 会話中の wikidata の語彙ではない単語には GloVe embeddings[7] を用いた.

表1に提案モデルにおけるハイパーパラメータの詳細を示す.

表 1: ハイパーパラメータの詳細

項目	仕様
LSTM のサイズ	256
バッチサイズ	1
学習率	1e-4
メモリの次元	256 × 64
読み出し回数	4
書き込み回数	1
最適化手法	RMSProp, momentum=0.9
重み初期化	$\sim N(0, 0.1)$

## 4.2 実験結果

提案モデルに対して, 上記のように CSQA の一部を用いて実験を行なった. 知識メモリには10万回学習したメモリを用いた. 3,000回学習を行ない, 応答として予測した単語と正解の単語が一致するかどうかの誤り率によって評価し, 85.81%であった. 図3に各タイムステップにおける文脈メモリへの書き込みのアテンションの重みを可視化した結果, 図4に文脈メモリからの読み出しの結果, 図5に知識メモリからの読み出しの結果を示す.

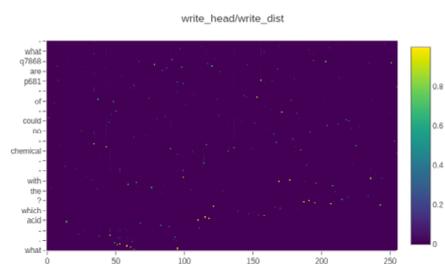


図 3: 各タイムステップにおける文脈メモリへの書き込みのアテンションの重みの可視化結果

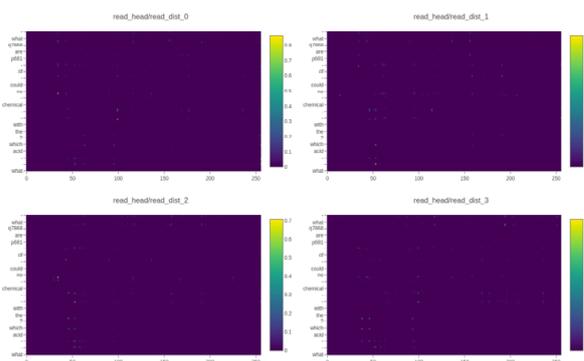


図 4: 各タイムステップにおける文脈メモリからの読み出しのアテンションの重みの可視化結果

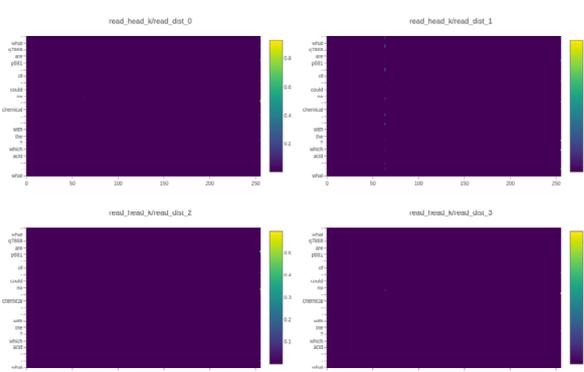


図 5: 各タイムステップにおける知識メモリからの読み出しのアテンションの重みの可視化結果

縦軸は各タイムステップ, つまり各単語に対応し, 横軸はメモリのスロット数256である. 読み出しの結果については, メモリからの読み出しは4回行うため, それぞれ示している. 書き込み操作についてはアテンションが強くかかっている部分はばらけており, メモリ全体を活用していた. 読み出し操作については文脈メモリの方は質問の後応答を生成するときに集中的に参照しており, 知識メモリはほとんど使われていないことが分かった. 読み出しにおいて, 書き込み時アテンションが当たっていた部分に強くアテンションがかかり, メモリの情報を用いている様子も見られた.

## 5 おわりに

DNC を拡張し, 文脈を捉えつつ構造化知識を活用できるモデルを提案した. 実験結果では, 応答を生成するのに知識メモリの情報あまり使われなかったが, 今回の実験では学習イテレーション数が3,000回と少なかったため, 学習回数を重ねるにつれて知識の効果

的な使い方を学習していくと思われる。今後の課題として、CSQA をすべて用いて実験を行いたい。また、提案モデルとオリジナルの DNC モデルの結果を比較したい。

representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543, 2014.

## 参考文献

- [1] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems*, pp. 3104–3112, 2014.
- [2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, pp. 5998–6008, 2017.
- [3] Sainbayar Sukhbaatar, Jason Weston, Rob Fergus, et al. End-to-end memory networks. In *Advances in Neural Information Processing Systems*, pp. 2440–2448, 2015.
- [4] Alex Graves, Greg Wayne, Malcolm Reynolds, Tim Harley, Ivo Danihelka, Agnieszka Grabska-Barwinska, Sergio Gomez Colmenarejo, Edward Grefenstette, Tiago Ramalho, John Agapiou, et al. Hybrid computing using a neural network with dynamic external memory. *Nature* 538(7626):471476, 2016.
- [5] Amrita Saha, Vardaan Pahuja, Mitesh M Khapra, Karthik Sankaranarayanan, and Sarath Chandar. Complex sequential question answering: Towards learning to converse over linked question answer pairs with a knowledge graph. arXiv preprint arXiv:1801.10314, 2018.
- [6] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Durán, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Neural Information Processing Systems 2013*, 2787–2795, 2013.
- [7] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. GloVe: Global vectors for word