

文脈を考慮した対話のニューラルネットワークモデルの 日本語注文対話への適用

Applying a neural network model of a dialogue considering the context
to Japanese ordering dialogue

村山 友理 *¹
Yuri Murayama

小林 一郎 *¹
Ichiro Kobayashi

森田 武史 *²
Takeshi Morita

中野 有紀子 *³
Yukiko Nakano

山口 高平 *²
Takahira Yamaguchi

*¹お茶の水女子大学
Ochanomizu University

*²慶應義塾大学
Keio University

*³成蹊大学
Seikei University

There are many works on dialogue systems considering the context using a neural network. In this work, to realize a robot cafeteria, we applied a neural network model to a process of an ordering dialogue. We created a Japanese ordering dialogue dataset and conducted an experiment with it. As a result, we obtained a mean test error rate of low value.

1. はじめに

近年、文脈を考慮した対話における記憶装置付きニューラルネットワークモデルとして様々なモデルが提案されている [Graves 16, Graves 14, Weston 14, Sukhbaatar 15, Kumar 16, Vinyals 15, Yang 18]. 本研究では、ロボット喫茶店の実現を目指し、喫茶店における注文対話に取り組んでおり、注文対話の課題として、複数の人からの注文や注文の変更、注文の確認などが挙げられる。この課題を解くには客の発言の記録や、文脈を理解していることが必要であり、注文対話の処理にニューラルネットワークモデルを適用することを考えた。適用するにあたり、日本語の注文対話データセットを作成し、それについて実験を行なった。結果は、平均テスト誤り率として低い値を得ることができた。

2. 記憶装置付きニューラルネットワーク

Memory Networks [Weston 14] は、巨大メモリとメモリへの入出力操作を行う学習コンポーネントから構成されるメモリ付きニューラルネットワークである。メモリに文脈情報などを格納しておき、そこから質問に関する情報を連鎖的に読み出し利用することにより、文脈を考慮した質問応答を可能にしている。Memory Networks では、メモリに対して Hard Attention を用いており、それは微分不可能なので、学習データをアノテーションし、質問に関する情報の選択について別途学習させる必要がある。

End-To-End Memory Networks [Sukhbaatar 15] は、Memory Networks の質問に関する情報の選択において、微分可能な Soft Attention を用いたモデルである。それを用いることにより、パラメータを end-to-end に学習することができ、学習データをアノテーションする必要がなくなるため、より多様なデータを扱うことができる。

Dynamic Memory Networks [Kumar 16] は、End-To-End Memory Networks に対して入力文や質問文、メモリ内部のベクトルのエンコードに GRU を用いたモデルである。

Neural Turing Machine [Graves 14] は、読み書き可能な外部メモリを持つ LSTM である。メモリに対して Soft Attention を用いて読み出しと書き込みを行う。Differentiable Neural Computer [Graves 16] は、Neural Turing Machine のメモリ

制御に関するアクセスメカニズムをより改善したモデルであり、質問応答のタスクにおいて記憶装置付きニューラルネットワークの中で最も良い結果を達成している。本研究では、このモデルを採用し注文対話の実現を目指す。Differentiable Neural Computer について次章で詳細に説明する。

3. 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。Differentiable Neural Computer (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。

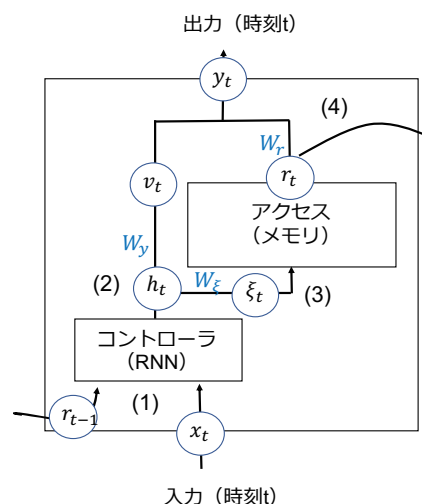


図 1: DNC の全体図

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、読み出しヘッドによってメモリの読み出しが行われる。

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。

連絡先: 村山友理, お茶の水女子大学大学院人間文化創成科学研究科
理学専攻情報科学コース小林研究室, 〒 112-8610 東京都文京区大塚
2-1-1, murayama.yuri@is.ocha.ac.jp

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 $\xi_t = W_\xi h_t$ に分ける。
3. ξ_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。これらのシステムは微分可能であるため、メモリの使い方も含めて end-to-end に学習することができる。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も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。

DNC の質問応答タスクでは bAbI データセット [Weston 15] を用いている。bAbI データセットは、人間の行う推論を真似た学習ができるように人工的に作られたデータセットであり、タスク毎に “John is in the playground. John picked up the football. Bob went to the kitchen. Where is the football? A:playground” のようなショートストーリーとそれについての質問と答えを含んだデータが 20 種類用意されている。公開されているデータセット *1 の en-10k サブセットでは、それぞれのタスクに対して、train 10,000 個、test 1,000 個のデータが提供されている。Graves らは DNC を用いた実験で、平均テスト誤り率 3.8% を達成したと報告している。本研究では、bAbI データセットを参考に日本語注文対話データセットを作成し、DNC を用いた注文対話を行う。

4. 日本語注文対話データセットの作成

本研究では、喫茶店での注文の場面を想定し、bAbI データセットの 20 タスクの内 10 タスクを参考に日本語の注文対話データセットを作成した。表 1 に作成した注文対話データセットの例を示す。尚、例では A, B, C とアルファベット順に並んでいるが、この順番はランダムにしている。

- Single Supporting Fact (一文参照)
bAbI データセットにおいて “Mary went to the bathroom. John moved to the hallway. Mary travelled to the office. Where is Mary? A:office” のように 1 文を見れば質問に対して答えられるようなタスクとなっており、本データセットでは「～の注文は?」という質問に対して該当する客の発言 1 文を見れば答えを出せるタスクとして用意した。
- Two Supporting Facts (二文参照) と Three Supporting Facts (三文参照)
Single Supporting Fact と同様に、それぞれ 2 文、3 文を参照すれば答えられるタスクとなっている。
- Conjunction (接続詞)
bAbI データセットでは接続詞の “and” を扱っており、本データセットでもそれに対応して “と” を扱うようにした。

- Yes No Questions (イエスノー質問)
質問に対して、はい/いいえで答えるタスクである。
- Counting (数え上げ)
bAbI データセットでは “Daniel picked up the football. Daniel dropped the football. Daniel got the milk. Daniel took the apple. How many objects is Daniel holding? A: two” のように人物が物を拾ったり落したりして最終的に持っている物の数を答えるタスクだが、本研究では注文の場面を想定しているため、料理の注文数を答えるタスクとした。
- Lists (列挙)
Counting タスクのように人物が物を拾ったり落したりし、最終的に持っている物のリストを答えるタスクであるが、本データセットでは「～の注文は?」という質問に加えて「全員の注文は?」といった質問にも答えるようにした。
- Simple Negation (否定)
否定表現を扱うタスクであり、本データセットでは注文の場面で想定される注文の変更を扱うとした。
- Basic Coreference (基本的共参照)
“Daniel was in the kitchen. Then he went to the studio. Sandra was in the office. Where is Daniel? A:studio” において代名詞 he が Daniel を指していることが分からないと答えられないようなタスクであり、注文対話においても「私もそれで」の “それ” が何を指すかを問うタスクとした。
- Compound Coreference (複合的共参照)
Basic Coreference における代名詞が複数の人物や物を意味する場合に、きちんと指し示すことができるかを試すタスクである。

本データセットでは、Counting タスク 408 個、Basic Coreference タスク 132 個、Compound Coreference タスク 264 個、他はそれぞれ 1,320 個作成し、総数 10,044 個の注文対話データセットを構築した。モデルの学習と評価にはそれらを 9 対 1 に分割して用いた。

5. 実験

5.1 実験設定

作成した注文対話データセットに対して Differentiable Neural Computer を用いて実験を行なった。表 2 に設定したハイパーパラメータの詳細を示す。7 万回学習を行ない、誤り率によってモデルを評価した。

5.2 実験結果

実験結果を図 2 に示す。

Yes No Questions タスクが最も誤り率が高かったが、学習が進むにつれて値は低くなっていった。次に誤り率が高かったのが Basic Coreference タスクであり、これも学習回数を重ねる毎に徐々に下がっていった。また、Simple Negation タスクは、1 万回学習を行なった時点での誤り率はやや高かったが、2 万回以降 0 になった。平均誤り率は 7 万回での評価において 0 になるまで緩やかに減少していった。

*1 <http://www.thespermwhale.com/jaseweston/babi/tasks1-20.v1-2.tar.gz>

表 1: 作成した注文対話データセットの例

Single Supporting Fact (一文参照) A: ハムサンド B: クロワッサン C: カルボナーラ A の注文 は? ハムサンド B の注文 は? クロワッサン C の注文 は? カルボナーラ	Counting (数え上げ) A: グレープフルーツ ジュース B: グレープフルーツ ジュース C: レモンティー グレープフルーツ ジュース の注文 数 は? 2 レモンティー の注文 数 は? 1
Two Supporting Facts (二文参照) A: ハムサンド B: クロワッサン C: カルボナーラ A: アップルパイ B: パナナブレッド C: チョコレート ケーキ A の注文 は? ハムサンド, アップルパイ B の注文 は? クロワッサン, パナナブレッド C の注文 は? カルボナーラ, チョコレート ケーキ	Lists (列挙) A: ボロネーゼ をシェア で A: ハムサンド B: クロワッサン C: カルボナーラ A の注文 は? ハムサンド, ボロネーゼ のシェア B の注文 は? クロワッサン, ボロネーゼ のシェア C の注文 は? カルボナーラ, ボロネーゼ のシェア 全員の注文 は? ハムサンド, クロワッサン, カルボナーラ, ボロネーゼ
Three Supporting Facts (三文参照) A: ハムサンド B: クロワッサン C: カルボナーラ A: アップルパイ B: パナナブレッド C: チョコレート ケーキ A: アップル ジュース B: グレープフルーツ ジュース C: カプチーノ A の注文 は? ハムサンド, アップルパイ, アップル ジュース B の注文 は? クロワッサン, パナナブレッド, グレープフルーツ ジュース C の注文 は? カルボナーラ, チョコレート ケーキ, カプチーノ	Simple Negation (否定) A: ハムサンド B: クロワッサン C: カルボナーラ B: クロワッサン ではなくて ハンバーガー にします A の注文 は? ハムサンド B の注文 は? ハンバーガー C の注文 は? カルボナーラ
Conjunction (接続詞) A: ハムサンド と アップルパイ と アップル ジュース B: クロワッサン と グレープフルーツ ジュース C: カルボナーラ A の注文 は? ハムサンド, アップルパイ, アップル ジュース B の注文 は? クロワッサン, グレープフルーツ ジュース C の注文 は? カルボナーラ	Basic Coreference (基本的共参照) A: グラタン B: 私 も それで C: ハムサンド A の注文 は? グラタン B の注文 は? グラタン C の注文 は? ハムサンド
Yes No Questions (イエスノー質問) A: アップルパイ B: パナナブレッド C: チョコレート ケーキ A の注文 は アップルパイ ですか? はい B の注文 は チョコレート ケーキ ですか? いいえ C の注文 は パナナブレッド ですか? いいえ	Compound Coreference (複合的共参照) A: グラタン と アイスティー B: 私 も それら で C: ハムサンド と アップル ジュース A の注文 は? グラタン, アイスティー B の注文 は? グラタン, アイスティー C の注文 は? ハムサンド, アップル ジュース

表 2: 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	RMSPProp, momentum=0.9
重み初期化	~N(0, 0.1)

5.3 考察

Yes No Questions タスクは質問に対して、本データセットにおいて唯一はい/いいえのどちらかを答えるものであるが、間違えた例の全てにおいて正解が「はい」のところを「いいえ」と誤答するなど、正解とは逆のものを答えてしまっていた。Basic Coreference タスクについては、データセットの数が最も少ないタスクであり、データセット数が学習に影響したものと考えられる。Simple Negation タスクは、間違えた例

の全てが正解が“ボロネーゼ”の場合であり、誤答として“ハンバーガー”と“ミート”が多く見られ、メインの料理の中で間違えていた。今回作成したデータセットは語彙数も 72 と少なく、かなり単純なものだったため、学習の難易度は比較的低かったと思われる。

6. おわりに

本研究では、喫茶店における注文対話の処理にニューラルネットワークモデルを適用した。日本語の注文対話データセットを作成し、Differentiable Neural Computer を用いて実験を行なった結果、平均テスト誤り率として低い値を得ることができた。

今回、bAbI データセットを参考に注文対話を作成したが、今後の課題として、実際の注文対話のコーパスを収集して多様な注文対話に対応できるようにしていきたいと考えている。また、実際の喫茶店で想定される、料理に関する質問などは、文脈だけから答えることは難しいため、知識を導入することによって、より複雑な対話理解に取り組んでいきたい。

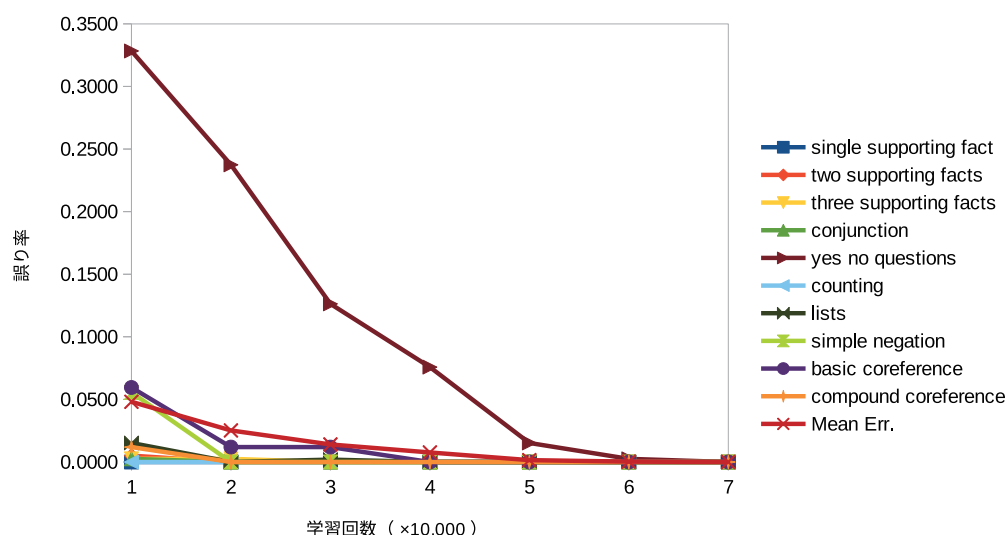


図 2: 実験結果

謝辞

本研究は、科学技術振興機構 (JST) 戦略的創造研究推進事業 (CREST) 「実践知能アプリケーション構築フレームワーク PRINTEPS の開発と社会実践」 (JPMJCR14E3) の支援によって実施した。

参考文献

- [Graves 16] Alex Graves, Greg Wayne, Malcolm Reynolds, Tim Harley, Ivo Danihelka, Agnieszka Grabska-Barwińska, Sergio Gómez Colmenarejo, Edward Grefenstette, Tiago Ramalho, John Agapiou, et al. 2016. Hybrid computing using a neural network with dynamic external memory. *Nature* 538(7626):471476.
- [Graves 14] Graves, A., Wayne, G. Danihelka, I. Neural Turing machines. Preprint at <http://arxiv.org/abs/1410.5401> (2014).
- [Weston 14] Weston, J., Chopra, S. Bordes, A. Memory networks. Preprint at <http://arxiv.org/abs/1410.3916> (2014).
- [Sukhbaatar 15] Sainbayar Sukhbaatar, Jason Weston, Rob Fergus, et al. 2015. End-to-end memory networks. In *Advances in neural information processing systems*. pages 2440-2448.
- [Kumar 16] Ankit Kumar, Ozan Irsoy, Peter Ondruska, Mohit Iyyer, James Bradbury, Ishaan Gulrajani, Victor Zhong, Romain Paulus, and Richard Socher. 2016. Ask me anything: Dynamic memory networks for natural language processing. In *International Conference on Machine Learning*. pages 1378-1387.
- [Vinyals 15] Vinyals, O., Fortunato, M., Jaitly, N. Pointer networks. In *Advances in Neural Information Processing*

Systems Vol. 28 (eds Cortes, C et al.) 2692-2700 (Curran Associates, 2015).

- [Yang 18] Hyochang Yang, Sungzoon Cho, et al. 2018. Finding remo (related memory object): A simple neural architecture for text based reasoning. *arXiv preprint arXiv:1801.08459*.
- [Weston 15] Weston, J., Bordes, A., Chopra, S. Mikolov, T. Towards AI-complete question answering: a set of prerequisite toy tasks. Preprint at <http://arxiv.org/abs/1502.05698> (2015).