

モジュール分割された日本語対話システムの作成

知能情報メディア主専攻 201611350 江畑 拓哉

指導教員 Claus Aranha (コンピュータサイエンス専攻)

櫻井鉄也 (コンピュータサイエンス専攻)

提出日 2018 年 10 月 17 日

1 序論

現在世界では様々な目的を持った対話システムが研究されている。主にそれは、ユーザを何らかの目標に導くためのタスク指向システム、対話そのものを目的とした非タスク指向システムの二種類に分かれている。またこの二つを組み合わせたシステムも開発されており、例えば Apple 社の Siri [2] がこれに該当する。

本研究では上記の二つを満たしなおかつより低コストなシステムを構築するため、様々な目的を持ったモジュールを作成し組み合わせる手法を提案する。

2 研究概要

2.1 研究目標

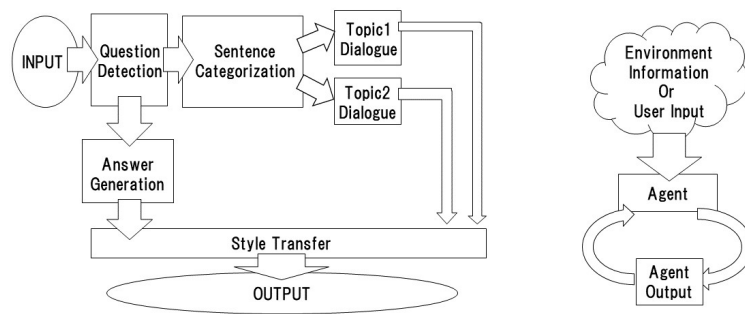


図 1 システム概図

本研究は上記のシステムを構築することを目的とする。

例えばそのエージェントの名前に関する質問が来た場合には、Question Detection を通り Answer Generation で自分の名前に関する文章を生成し、Style Transfer を経由して回答を出力する。また例えばゲームに関する入力の場合には Question Detection を通過し Sentence Categorization で ゲームに関する Topic Dialogue にアクセスし出力を得、Style Transfer を通して出力する。Style Transfer は口調変換であり、エージェントに様々なペルソナを持たせるために用いられる。そのペルソナはエージェントの外見などと同期することでより対話感を出すことが出来ると考えている。またエージェントは天候などの環境データ、これまでの対話データからそのステータスが変化することを想定しており、例えばあるモノに対する好感度はそれまでの対話から形成されるようにする。

2.2 利用するデータ

このシステムにおいて必要とするデータは主に二種類である。

- スタイル変換を行うためのデータ

このデータには、スタイルが付与されていない文と付与された文のペアを集めたものと、スタイルが付与されていない文の群と付与された文の群を用いるものがある。現在の進捗としてはデータ生成の問題から前者を用いている。

- 対話を行うためのデータ

入力文と出力文のペアを集めたものである。収集手法として、① 実際に手作りする ② アニメや演劇などから集める ③

SNS から集める を行い、特に後者 2 つについては有効なデータをフィルタリングする手法を研究している。

2.3 それぞれのモジュールに適用した手法

- Question Detection
入力文に対して LSTM を適用し質問番号を出力する手法より殆どの分類が行えることがわかった。
- Answer Generation
予め用意したテンプレートにエージェントのステータスから取り出した値を当てはめ生成する。
- Sentence Categorization
SCDV [5] のような手法や、sentiment analysis で使われる手法 [4] を検討している。また単語埋め込みとして、fasttext [3] や Word2Vec を採用したいと考えている。いずれも十分なサイズのデータを収集できておらず実験が出来ていないため、代替としてアニメの台本から入出力のペアを切り出した対話として成立しているデータ、一般的な対話を集めたデータを用いて RNN, CNN を用いた二値分類実験を行い、後者で 8 割 5 部ほどの検出をすることが出来た。
- Topic Dialogue
日本語でも Sequence to Sequence [7] のような手法を用いた様々な取り組みが行われているが、芳しい結果を得られたものは少ない。しかし公開されているデータセットを確認すると、英語のデータセットと比較して、余り有効なデータセットであるようには思えなかった (入力と出力のペアで意味が通らないものがある等) ため、自作の日常会話を集めたデータセットで Sequence to Sequence を用いた実験をした所、それなりの精度の実験結果を得ることが出来た。しかしデータ量が少ないため (1k 程度) これをいかに増やしていくかが今後の課題だ。
- Style Transfer
スタイルを付与されていないものと、付与されたもののペアをデータセットとして、Sequence to better Sequence [6] と、DAE [8] から着想を得たその一部変更したものと CopyNet [8] を用いて実験を行った。結果として前者 2 つはチューニングが難しい代わりに自然な文を生成することが出来、後者は比較的高速に文を生成できる他にある程度未知語に対応が出来ることがわかった。今後 Sequence to better Sequence に CopyNet を組み合わせたシステムを提案したいと考えている。
- データ収集について
手打ちに関しては誤字脱字や意味の通らない文を含まない理想的なデータになるように打ち込んでしまっている。アニメや演劇からの収集については、長すぎる文は必要な部分のみにし、専門的過ぎる文は削除している。こちらは著作権などの問題があるため、この対処を調べている。SNS からのデータの収集については Twitter を用いているが、ノイズの多いデータが混じっているため、そのデータクリーニングについて既存の手法 [9] を参考にしながら研究している。
- ステータス更新を行う手法
文章に関しては Word2Vec [1] や fasttext を用いた極性判定を用いて実験する予定だ。

References

- [1] Yoav Goldberg and Omer Levy. *word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method*. 2014. arXiv: 1402.3722.
- [2] Apple Inc. あなたの声を使って Siri にもっと色々頼みましょう. URL: <http://www.apple.com/jp/ios/siri>.
- [3] Armand Joulin et al. *Bag of Tricks for Efficient Text Classification*. 2016. arXiv: 1607.01759.
- [4] Yoon Kim. *Convolutional Neural Networks for Sentence Classification*. 2014. arXiv: 1408.5882.
- [5] Dheeraj Mekala et al. *SCDV: Sparse Composite Document Vectors using soft clustering over distributional representations*. 2016. arXiv: 1612.06778.
- [6] Jonas Mueller, David Gifford, and Tommi Jaakkola. "Sequence to Better Sequence: Continuous Revision of Combinatorial Structures". In: *Proceedings of the 34th International Conference on Machine Learning*. Ed. by Doina Precup and Yee Whye Teh. Vol. 70. Proceedings of Machine Learning Research. International Convention Centre, Sydney, Australia: PMLR, Aug. 2017, pp. 2536–2544. URL: <http://proceedings.mlr.press/v70/mueller17a.html>.
- [7] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. *Sequence to Sequence Learning with Neural Networks*. 2014. arXiv: 1409.3215.
- [8] P. Vincent et al. "Extracting and composing robust features with denoising autoencoders." In: (2008).
- [9] 稲葉 通将, 神園 彩香, and 高橋 健一. "Twitter を用いた非タスク指向型対話システムのための発話候補文獲得". In: 人工知能学会論文誌 29.1 (2014), pp. 21–31. DOI: 10.1527/tjsai.29.21.