

ユーザーの発話意図理解に基づくインタビュー発話の生成に向けて

Towards generation of interview utterance based on user's utterance intention understanding

松好祐紀^{1*} 滝口哲也¹ 有木康雄¹
Matuyoshi Yûki¹ Takiguchi Tetsuya¹ Ariki Yasuo¹

¹ 神戸大学システム情報学研究科

¹ Graduate School of System Informatics, Kobe University, Department of Information Science

Abstract: The purpose of this study is to construct an interviewer agent to listen to how elderly people are concerned about social participation. In this paper, we describe the language understanding unit, dialogue management unit, language generation unit, which is the core of the interviewer agent. In the language understanding unit, the utterance intention of the user is extracted as a frame using the LSTM Encoder-Decoder and the keywords appearing in the utterance sentence are extracted as the slots. We plan to construct the dialog management unit with the deep neural network and the language generation unit with the LSTM Encoder-Decoder like the language understanding unit.

1 はじめに

本研究では、高齢者を対象に、「社会参加をどのように考えているか」について聞き取るインタビュアーエージェントの構築を目標としている。本稿では、インタビュアーエージェントの中核である、言語理解部、対話管理部、言語生成部に関して述べる。言語理解部では、LSTM Encoder-Decoder を用いてユーザーの発話意図をフレーム、発話文中に現れるキーワードをスロットとして推定する。対話管理部、言語生成部は現在検討している段階であり、対話管理部はディープニューラルネットワーク、言語生成部は言語理解部と同じく、LSTM Encoder-Decoder で構築を行う予定である。

本稿では、言語理解部に関しては、これまでに研究してきた「オセロゲーム中にユーザーを支援する質問応答システム」[1]において実装したモデルを用いて説明する。

2 提案システム

2.1 システム概要

想定しているシステムの構成としては、まず、ユーザーの質問が音声認識され、文字列の形で言語理解部

に渡される。そして、言語理解部で推定されたユーザーの発話意図を、対話管理部においてシステムの発話意図に変換する。最後に言語生成部において、システムの発話意図からシステム発話を生成する。

言語理解部では、ユーザー発話のフレーム、スロットを推定する。これまでの研究 [1] では、フレームはユーザーの大きな意図を表す「質問タイプ」として定義した。質問タイプは、理由、定義、場所など 15 種類とした。また、スロットは、質問文中のキーワードの種類（「X 打ち」なら「オセロ用語」など）を「質問キーワードクラス」として 13 種類定義した。

提案モデルは図 2 のような、LSTM Encoder-Decoder で実装した（以下、 $Model_{SLU}$ とする）。質問文を形態素に分解し、それぞれを one-hot な単語ベクトルに変換して $(x_1 \dots x_m)$ word embedding を行い、モデルに入力する。文末記号 $\langle eos \rangle$ が入力された時点の隠れ層 h_{qtype} から質問タイプの推定値を計算する。そして、 $\langle go \rangle$ が入力された後、質問キーワードクラス $(y_1, y_2 \dots)$ が順番に生成される [2]。図では省略しているが、Attention 機構 [1,3] をモデルに導入している。

$Model_{SLU}$ の特徴としては、デコーダでの各時間ステップでの入力に h_{qtype} を加えている。また、現在の時間ステップ t の出力 y_t の計算に、2 つ前の時間ステップ $t-2$ の隠れ層の出力 y_{t-2} も利用する。

*連絡先：神戸大学システム情報学研究科
〒657-8501 兵庫県神戸市灘区六甲台町 1-1 自然科学総合研究棟 3 号館 805 号室
E-mail: yuki.matsuyoshi@stu.kobe-u.ac.jp

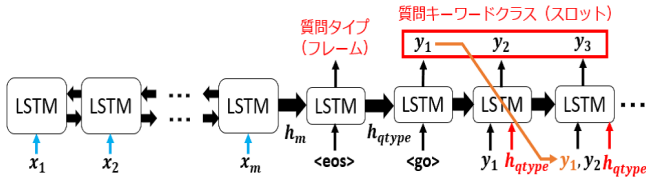


図 1: 提案モデル ($Model_{SLU}$) の概略図

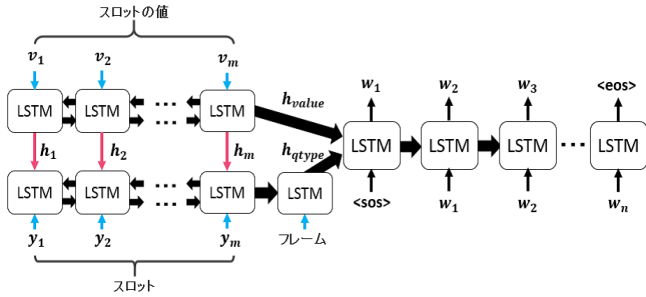


図 2: 提案モデル ($Model_{SLG}$) の概略図

2.2 言語生成部

言語生成部のモデルでは、システムの発話意図からシステムの発話を生成する。モデルは図 3 のような Encoder-Decoder で構築している ($Model_{SLG}$)。エンコーダは、発話意図のスロット ($y_1 \dots y_m$) とフレームをエンコードする部分と、スロットの値 ($v_1 \dots v_m$) をエンコードする部分に分かれる。デコーダでは、 $\langle \text{sos} \rangle$ が入力された後、システム発話の形態素 ($w_1, w_2 \dots$) が順番に出力される。現在、モデルの構築は完了し、[1] のデータセットを用いて実験を行っている。

3 実験

3.1 実験手法

言語理解部のモデルの実験を行った。データセットは、実際にオセロゲームをプレイする際に生じるユーザーの質問文データ集合 1788 分であり、各文に対して、質問タイプ、質問キーワードクラスを手手でアノテーションした。1788 文の内、1605 文を学習データ、183 文をテストデータとして使用した。また、質問タイプ、質問キーワードクラスをそれぞれ単独で推定するモデル ($Model_{QT}$, $Model_{QK}$)[1] も用意し、比較を行った。

3.2 実験結果

実験結果を表 1 に示す。質問キーワードクラスに関しては、単独のモデルに比べ $Model_{SLU}$ では推定率が向上したが、質問タイプに関しては、推定率の向上は見られなかった。質問キーワードクラスは質問タイプに比べて推定が複雑になるので、推定の信頼度が低くなる。よって、両者を組み合わせて推定すると、質問タイプの推定に影響が出たのではないかと考えられる。

表 1: joint モデル推定の質問タイプ推定に関する結果。(推定率の単位: %)

モデル	質問タイプ			質問キーワードクラスの推定率
	推定率	再現率	適合率	
$Model_{QT}$	85.3	0.81	0.83	-
$Model_{QK}$	-	-	-	81.4
$Model_{SLU}$	85.3	0.79	0.83	83.6

4 今後の取り組み

今後は、テーマをオセロゲームのユーザーサポートから、インタビュワーエージェントへと切り替える。言語理解部は、質問タイプ、質問キーワードクラスを、インタビュー対話におけるユーザー発話意図のフレーム、スロットに置き換えると、2.2 の $Model_{SLU}$ が利用できる。対話管理部では、ユーザーの発話意図、システムの発話意図、ともにフレーム、スロットの形式で定義するので、ユーザーの発話意図からシステムの発話意図への変換を行う。この変換を最初は DNN で、最終的には Deep Q-learning で構築する [4]。言語理解部では、システムの発話意図を基に、2.3 で述べた $Model_{SLG}$ を用いてシステムの発話生成を行う。

5 まとめ

本稿では、インタビュワーエージェント構築に向けて、これまでの研究を基に言語理解部を構築し、実験を行った。結果として、質問タイプの推定率が 85.3%、質問キーワードクラスの推定率が 83.6% という結果になった。今後は、この言語理解部のモデルをインタビュワーエージェントに転用し、対話管理部、言語生成部の構築に取り組む。

謝辞

本研究の一部は、JSPS 科研費 JP17K00236 の助成を受けたものである。

参考文献

- [1] 松好祐紀, 滝口哲也, 有木康雄. "Attention-based LSTM を用いた意図理解とキーワード抽出の統合による質問応答システム," 電子情報通信学会技術研究報告, Vol. 118, No. 198, pp. 9-14, 2018-08.
- [2] B. Liu and I. Lane, "Attention-Based Recurrent Neural Network Models for Joint Intent Detection and Slot Filling," in INTERSPEECH. 2016, pp. 685-689
- [3] Minh-Thang Luong *et al.* "Effective Approaches to Attention-based Neural Machine Translation," arXiv preprint arXiv:1508.04025, 2015
- [4] Pei-Hao Su *et al.* "Continuously Learning Neural Dialogue Management," arXiv preprint arXiv:1606.02689v1, 2016