

1 はじめに

対話システムは、ユーザが持つ特定の目的を対話によって達成するタスク指向型対話システムと、対話そのものを目的とする雑談対話システムに大別される。対話状態追跡とは、ユーザの目的や要求を対話状態に反映し、適切に維持あるいは更新するタスクであり、タスク指向型対話システムの重要な要素である。

近年では、深層学習を用いた End-to-End 型の対話状態追跡モデルが高い性能を示しており、発話文から対話状態を直接出力することが可能となった [1, 2]。発話文を入力とするにあたり、対話履歴が対話の流れを捉えるために用いられている。従来研究では対話履歴として直近の数発話を入力とするが、計算量を抑えるために多くの発話文を入力しないことが多く、対話状態の推定に貢献する文が含まれないことがある。

本研究では、システムの発話意図を示す対話行為タグを用いて、長い対話履歴の中から対話状態追跡に重要だと思われるシステム発話を抽出する手法を提案する。対話行為タグによって各システム発話を意図ごとに分類し、特に状態に影響を与える発話を選択的に抜き出すことで、計算量を抑えつつ長い対話履歴を考慮することが可能であると期待できる。

対話システムの国際コンペティションである Dialogue System Technology Challenge 8 Track4 Dialogue State Tracking (DSTC8-Track4) [3] において、対話状態追跡の SGD データセットとベースラインモデルが公開されている。そのデータセットとモデルを用いて実験を行い従来手法と提案手法の比較を行う。

2 DSTC8-Track4

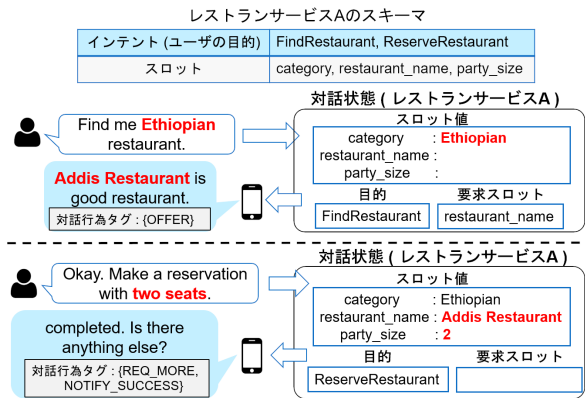


図 1: DSTC8-Track4 のイメージ図

DSTC8-Track4 は、新サービスや新ドメインへの対応を再学習なしで可能にする対話状態追跡モデルを作成することを目的としたタスクである。本タスクのイメージ図を図 1 に示す。

SGD データセットでは、レストラン案内や交通案内などのドメインごとに複数のサービスが存在する。そして、サービスごとにサービスを定義する枠組みであるスキーマが与えられ、対話状態を定義するために用いられる。スロットはユーザの要求を示すための属性で、インテントは目的である。対話状態は、ユーザの要求をスロットと値の組によって表すスロット値と、インテント、ユーザが値を要求しているスロットのリストである要求スロットで構成される。そして、ユーザが発話した際にそれまでの対話中に存在する情報を捉えて対話状態に反映する。また、システム発話には発話の意図を示す対話行為タグ、スキーマにはスロットと目的の意味をシステムが理解するための説明文が付与されている。そのため、入力として対話文、スキーマ、対話行為を用いて、対話状態を出力する。

3 対話行為タグを用いた重要対話履歴抽出

重要対話履歴とは、対話状態に反映されるスロット値を持つなど対話状態の推定に貢献する発話を指す。提案モデルでは DSTC8-Track4 で提供されているベースラインモデルに重要対話履歴抽出モジュールを加える。このモジュールは、入力文中にある前のターンのシステム発話が特定の対話行為タグを持つ場合に、その発話を保持する。既に発話を保持している場合は、新しい発話に更新する。そして、元々保持されていた発話を重要対話履歴として入力に加える。入力には、重要対話履歴、前のターンのシステム発話、現在のターンのユーザ発話の 3 発話とする。

4 評価実験

DSTC8-Track4 では、目的の推定精度を示す Active Intent Accuracy (AIA)、要求スロットの推定精度を示す Requested Slots F1 (RSF1)、スロット値の推定精度を示す Average Goal Accuracy (AGA) と Joint Goal Accuracy (JGA) を用いる。本実験ではこれらの指標に加え、学習時間も評価に用いる。

表 1: 提案手法と従来手法の比較

	AIA	RSF1	AGA
従来手法 (直近 4 発話入力)	0.974	0.948	0.844
提案手法 (OFFER)	0.946	0.947	0.845
	JGA	学習時間	
従来手法 (直近 4 発話入力)	0.553	26.3(h)	
提案手法 (OFFER)	0.565	20.7(h)	

提案手法では、ユーザにスロット値を提案する “OFFER” タグが事前に行った実験で最良の結果を示したため、それを用いている。また、従来手法は直近 4 発話入力としている。これら 2 つを比較した結果、JGA は提案手法の方が 1.2% 高い結果を示し、学習時間も削減している。ゆえに、提案手法はモデルの学習時間を減らすのと同時に、性能を向上させることが可能であるといえる。

5 むすび

本研究では、対話状態追跡における対話履歴の使い方として、システムの対話行為タグを用いて重要な対話履歴を抽出する手法を提案した。今回の手法は人間が選択的にどの対話行為タグを持つ発話に注目するかを決めている。この手法でも性能は向上するが、機械側で対話履歴中のどの発話に注目すべきかを推定可能な対話状態追跡モデルを検討したい。

参考文献

- [1] Nikola Mrkšić et al. Neural Belief Tracker: Data-Driven Dialogue State Tracking. arXiv preprint arXiv:1606.03777. 2017.
- [2] Bing Liu and Ian Lane. An End-to-End Trainable Neural Network Model with Belief Tracking for Task-Oriented Dialog. In Proceedings of Interspeech 2017, pages 2506–2510, 2017.
- [3] Seokhwan Kim et al. The eighth dialog system technology challenge. In Third NeurIPS workshop on Conversational AI: “Today’s Practice and Tomorrow’s Potential”, 2019.