

タスク指向対話における発話の分類と根拠抽出

福永 隼也[†] 西川 仁[†] 徳永 健伸[†] 横野 光[‡] 高橋 哲朗[‡]

[†]東京工業大学 情報理工学院 [‡]株式会社富士通研究所

fukunaga.s.ab@m.titech.ac.jp {hitoshi,take}@c.titech.ac.jp

{yokono.hikaru,takahashi.tet}@jp.fujitsu.com

1 はじめに

タスク指向対話システムの中には、そのタスクを達成するために、背後に用意された属性-値の表を埋めることが必要となるシステムが存在する。例えば、商品検索を行う対話システムは、ユーザの希望する商品を検索するために、商品への希望条件をユーザから聞き出す必要がある。この場合、商品の希望条件の一覧が属性-値の表であり、属性には「価格帯」などの条件名、値には「1,000 円台」などの具体的な条件が対応する。

このようなタスクを行う最も単純な対話システムは、用意された表の全ての属性について一つずつユーザに尋ねるシステムであるが、ユーザが質問された属性について知らない場合や、どのような値を指定すべきかわからない場合は値を埋めることができない。例えば、パソコンの商品検索においてシステムがメモリサイズについて質問したとしても、ユーザがメモリという概念を知らない場合や、自身の目的に照らしてどの程度のメモリが必要かわからない場合には具体的な値を指定することはできない。このような場合、ユーザは自分が達成したい目的を直接対話システムに伝えることがあり得る。例えば、動画編集をするためにパソコンを探しているユーザは、パソコンの仕様を伝える代わりに「動画編集をしたい」という自身の目的を伝えることが考えられる。このようなユーザの目的はあらかじめ定義された表中には存在しない可能性があり、その場合、表中の属性にしか対応できない対話システムは表を埋めることが出来ない。人間であれば、このような場合でも「動画編集を行うならばメモリが16GB必要」といった推論を行うことによって表中の属性に変換できる。福永ら (2017) は、不動産検索対話において、あらかじめ定義された検索条件以外の内容が無視できない割合で出現し、またそれらの多くが推論によって検索条件に変換できる内容であることを示している。

我々の目標は、タスク指向対話システムにおいて、あらかじめ定義された属性以外の内容を含むユーザ発話を、あらかじめ定義された属性-値のペアに変換することである。また、対話システム上で使用することを考慮すると、その変換が適切かどうかをユーザに確認する必要がある。そのためには、ユーザ発話のどの部分から判断して、その属性への変換を行ったかという根拠を抽出する必要がある。すなわち、ユーザ発話を入力として、属性-値ペアへの変換結果と、ユーザ発話中のどの部分に変換の根拠かという2種類の情報を出力することが我々の最終的な目標である。

この目標を達成するために、本稿では、サポートベクターマシン (SVM) による分類器によってユーザ発話をあらかじめ定義された属性に変換し、SVM で学習された素性の重みを用いて変換の根拠となる単語を抽出する。

不動産検索対話コーパスを用いた評価実験の結果、属性の変換に利用する素性および根拠抽出の対象としてユーザ発話のみを用いるのは不十分であること、また、変換の根拠として1単語のみでは不十分であることがわかった。

2 関連研究

本研究の目的は、1つのユーザ発話をあらかじめ定義された属性に変換することであり、変換先の属性は複数になることがあり得る。このように1つのデータが複数のクラスに所属することを許容する分類問題はマルチラベル分類問題と呼ばれる。対話中の発話に対するマルチラベル分類問題の関連研究として、Xu et al. (2017) はシンガポール観光案内対話コーパス中の発話の発話行為 (Speech Act) を推定している。このコーパスでは、1つの発話に複数の発話行為が付けられていることがあるため、発話行為の推定はマルチラベル分類問題となる。この研究では、各発話に対し、それぞれの発話行為が付けられるか否かの二値ベクトルを畳み込みニューラルネットワーク (CNN) によ

て推定することで、SVMによるベースラインと比較してより良い精度を達成している。

また、機械学習による推定の根拠を人間が理解できる形で出力するという研究は、様々な分野において行われている。特に自然言語処理においては、Bahdanau et al. (2014) が機械翻訳において注意機構を取り入れ、翻訳先の各単語が出力される際、入力文中のどの単語に注目したかという情報を利用している。また、Lei et al. (2016) は、商品レビューから各評価項目のレーティングを推定するタスクを例として、入力テキスト中のどの部分がレーティングの推定の根拠となるかを選択する手法を提案し、注意機構をベースとしたモデルよりも高い精度での根拠抽出を達成している。

3 発話の変換と根拠抽出手法

本稿では、ユーザ発話中の単語を素性ととする SVM を用いた分類器によって対話中のユーザ発話をあらかじめ定義された属性に変換し、各属性ごとの変換根拠として、ユーザ発話に含まれる単語中で最も SVM の素性の重みが大きい単語を出力する手法を提案する。

3.1 属性への変換

発話をあらかじめ定義された属性に変換する手法として SVM を用いる。1 つの発話が複数の属性に変換される場合があるため、単一の分類器ではなく、各属性について、その属性へ変換されるか否かを分類する 2 値分類器を作成する。入力発話を各属性に対応する分類器で 2 値分類し、それぞれで変換されると判断された属性のリストを変換結果として出力する。

入力発話を表現する素性として、発話中に含まれる名詞、動詞、形容詞、副詞の見出し語のうち、コーパス中で 2 回以上出現する単語を用いる。また、数と固有名詞については、一般化のためにそれぞれ 'NUM'、'PROP' という記号に置換する。

SVM のカーネル関数には線形カーネルを用いる。これは、根拠抽出に素性の重みを使用するためである。

3.2 変換根拠の抽出

SVM では、分類を行う際に素性ベクトルのどの要素が分類に寄与するかという重みを学習し、その重みが大きい素性ほど分類結果に影響する。発話の変換においては、発話中の単語を素性として用いるため、重みの大きい素性に対応する単語が、その属性に変換されるか否かの分類に大きな影響を及ぼす。したがって、入力発話中の単語のうち、最も素性の重みが大きい単語が、その属性への変換の根拠だといえる。

変換の根拠抽出は以下の手順で行う。

1. 入力発話を各属性の SVM によって変換する
2. 変換結果中の各属性について、対応する SVM の素性の重みを取得する
3. 各属性について、入力発話中の単語のうち最も大きな重みを持つ単語を、その属性への変換の根拠として出力する

したがって、変換後の属性それぞれに対し 1 単語が根拠として出力される。

4 評価実験

4.1 データセット

評価実験に用いるデータとして、Takahashi and Yokono (2017) が作成した不動産検索対話コーパスを用いる。このコーパスは、不動産屋と客の役割を担う 2 名の作業者が、客の希望する物件を探すためにおこなう日本語テキストチャットの模擬対話を収集したコーパスである。客側には、10 種類用意されたプロフィールのうち 1 つが割り当てられ、割り当てられたプロフィールに合う条件の物件を希望する。また、不動産屋側は、実際に物件を検索し提示することせず、物件の検索に必要な情報が得られたと判断した時点で対話を終了する。コーパス中の総対話数は 986 対話、総発話数は 29,058 発話、1 対話あたりの平均発話数は 29.5 発話である。また、客側による発話（ユーザ発話）数は、14,487 発話である。

コーパス中の各発話には、福永ら (2017) によって、38 種類の物件検索条件と、それ以外の内容を表す「その他」タグがアノテーションされている。本稿では、38 種類の検索条件をあらかじめ定義された属性として考え、「その他」タグが付与されたユーザ発話を 38 種類の検索条件へ変換することを目的とする。そのためには、学習および評価のための正解データとして、「その他」タグが付与されたユーザ発話から検索条件のリストへのマッピングが必要となる。マッピングの作成には、「その他」タグが付与された発話に同様に付与されている具体的な内容の記述を用いる。この内容の記述は「住む人数」や「勤務地」など、その発話の内容に関する記述であり、同じ内容の発話には同じ記述が付与されている。「その他」タグが付与されたユーザ発話から検索条件のリストへのマッピングは、この内容の記述ごとに、著者の主観によって作成した。例えば、「住む人数」が多ければ、より多くの部屋が必要となるため「間取りタイプ」に関係する。したがって、「住む人数」が内容として付与されているユーザ発話を「間取りタイプ」という検索条件にマッピングする。

表 1: 正解データにおける変換後の検索条件数に対する発話数

変換後の検索条件数	0	1	2	3	4	5	6	7	8	9	10	11	28	29	38
発話数	154	266	274	1,310	136	55	388	31	15	8	1	2	12	1	276

という具合である。

ユーザ発話 14,487 発話中「その他」タグが付与された 2,929 発話について、1 発話がいくつの物件検索条件に変換されるかを表 1 に示す。2,929 発話中、2 つ以上の検索条件に変換される発話は 2,509 発話 (85.7%) である。また、38 種類の検索条件全てに変換される発話が 276 発話 (9.4%) 存在しているが、これらは全て、不動産側の「他に何かご希望はありますか?」という内容の発話に対する、客側の「特にありません」という応答である。これらの発話は、全ての検索条件について、それ以上指定することがないということを示す発話であるので、全ての検索条件にマッピングしている。

データは 10 分割し、そのうちの 9 つを学習データ、残り 1 つを評価用データとして用いる。データの分割は、全発話をランダムにシャッフルした上で、正解データ中の属性の分布が偏らないように行った。学習には「その他」タグが付与されたユーザ発話だけでなく、全ての発話を使用した。評価は、評価用データ中の「その他」タグが付与されたユーザ発話のみに対して行った。

学習データ中の総発話数は 26,143 発話であり、そのうち「その他」タグが付与されているユーザ発話は 2,624 発話である。評価用データ中の総発話数は 2,915 発話であり、そのうち「その他」タグが付与されているユーザ発話は 305 発話である。

4.2 評価方法

検索条件への変換は、発話ごとに出力された推定の検索条件リストと、正解として付与されている検索条件リストとを比較し評価する。評価尺度として、正解の検索条件リストが空である 20 発話については、推定と正解が完全に一致した割合のみを、それ以外の 285 発話については、完全一致の割合に加え、精度、再現率、F 値を用いる。これは、正解が空の場合に精度と再現率の計算が出来ないためである。

根拠抽出については、各発話について根拠として出力された単語が、根拠と考えられるか否かを著者の主観によって評価する。評価の基準としては、発話に対し、「(根拠として出力された単語) ということは (対応する検索条件) に関係しますね」という応答が許容できるかどうかとする。評価の種類としては、その 1 単語のみで根拠として成り立つ (○)、1 単語のみで

は根拠とならないが発話中の前後の単語を加えることで根拠として成り立つ (△)、前後の単語を加えても根拠とならない (×) の 3 種類とする。

4.3 実験結果

4.3.1 検索条件への変換

正解の検索条件リストが空である 20 発話に対する推定と正解の完全一致の割合は、65.0%であった。また、それ以外の 285 発話についての精度、再現率、F 値の平均値、および推定と正解の完全一致の割合を表 2 に示す。

表 2: 正解が空でない発話についての評価結果

精度	再現率	F 値	推定と正解の完全一致の割合
67.7%	73.6%	67.1%	41.4%

4.3.2 根拠抽出

各検索条件についての根拠抽出の評価結果を、表 3 に○の割合の大きい順に 5 つ、表 4 に○の割合の小さい順に 5 つ示す。各検索条件について、上段の値は各評価が付けられた発話数、下段の値は全体に対する割合である。最も○の割合の大きい検索条件である「周辺環境」は、物件周辺のスーパーマーケットや小学校、病院などの施設についての検索条件であり、根拠として出力された単語には「治安」、「郊外」、「子供」が多かった。一方、表 4 に挙げた検索条件はいずれも全ての発話に対し出力された単語が根拠でないと判断されている。これらの検索条件に変換された発話のほとんどは先述した「特にありません」という発話であり、根拠としては「特に」という単語が出力されていた。また、最も△の割合が大きかった検索条件は「部屋の広さ」という条件であり、全体の 37.0%を△が占めていた。この検索条件において△だと判断された出力の多くは「人」という単語であった。これは単独では根拠として成り立たないが、直前の単語である「2」などの数字を含めることで物件に住む人数となるので、「部屋の広さ」の根拠として成り立つと判断した。

表 3: 根拠抽出の評価結果 (上位 5 つ)

検索条件	○	△	×	合計
周辺環境	92 (60.5%)	18 (11.8%)	42 (27.6%)	152
エリア	58 (52.7%)	7 (6.4%)	44 (40.4%)	109
室内設備	38 (48.7%)	5 (6.4%)	35 (44.9%)	78
部屋の位置	24 (46.2%)	0 (0%)	28 (27.6%)	52
目的地からの時間	26 (35.6%)	8 (11.0%)	39 (53.4%)	73

表 4: 根拠抽出の評価結果 (下位 5 つ)

検索条件	○	△	×	合計
瑕疵保証	0 (0%)	0 (0%)	27 (100%)	27
収納	0 (0%)	0 (0%)	28 (100%)	28
駅利便性	0 (0%)	0 (0%)	30 (100%)	30
日当たり・採光	0 (0%)	0 (0%)	30 (100%)	30
物件種別	0 (0%)	0 (0%)	30 (100%)	30

5 考察

5.1 検索条件への変換

正解が空の 20 発話中で推定を誤った例は 7 例存在し、そのうち 4 例は「はい、よろしくお願いします。」という、不動産屋側からの質問や確認に対する肯定の発話であった。この発話だけからどの検索条件に変換すべきかを判断することは難しいといえる。同様に、残りの 285 発話中で推定を誤った 118 例について、そのうち 35 例 (29.7%) は、ユーザ発話のみからでは変換すべき検索条件を判断することが難しい例であった。これは、対話がひとりの話者によるものではなく、二人の話者のインタラクションであるという性質によるものである。したがって、検索条件へ変換を行う際には、素性としてユーザ発話のみを用いるのではなく、その直前のエージェント (不動産屋) の発話も利用すべきであると考えられる。

5.2 根拠抽出

実験結果で述べたように、「2 人」など人数を表す単語列であれば根拠として成り立つにも関わらず、1 単語のみという制限によって単独では根拠として成り立たない (△) と判断されることがあった。このことから、根拠として 1 単語のみを用いるのは不十分である

といえる。

また、表 4 で×だと評価された発話のほとんどは「特にありません」という内容の発話であることは先に述べたが、これらの発話が各検索条件に変換されると判断するためには、直前の不動産側による「他に何かご希望はありますか？」という内容の発話が必要である。このように対話においては、ユーザ発話だけではなく、その直前のエージェント (不動産屋) の発話に根拠が現れる場合が存在するため、根拠を抽出する対象を広げる必要がある。

6 おわりに

本稿では、あらかじめ定義された属性以外の内容を含むユーザ発話をあらかじめ定義された属性に変換し、その根拠を抽出することを目的として SVM を用いた手法を提案した。不動産対話コーパスを用いた評価実験の結果、属性の変換に利用する素性および根拠抽出の対象はユーザ発話のみを用いるのは不十分であること、また、変換の根拠として 1 単語のみでは不十分であることがわかった。

今後は、テキストを入力とした推定問題における根拠抽出の最先端の手法である Lei et al. (2016) などを、我々のタスクに適用し、対話の性質を考慮した手法を提案することを目指す。

参考文献

- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.
- Tao Lei, Regina Barzilay, and Tommi Jaakkola. Rationalizing Neural Predictions. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 107–117, 2016.
- Tetsuro Takahashi and Hikaru Yokono. Two persons dialogue corpus made by multiple crowd-workers. In *Proceedings of the 8th International Workshop on Spoken Dialogue Systems (IWSDS)*, 2017.
- Guanghao Xu, Hyunjung Lee, Myoung-Wan Koo, and Jungyun Seo. Convolutional Neural Network using a threshold predictor for multi-label speech act classification. In *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp. 126–130, 2017.
- 福永 隼也, 徳永 健伸, 西川 仁, 高橋 哲朗, 横野光. 対話行為スロット自動生成のための話題タグアノテーション. 言語・音声理解と対話処理研究会 (第 81 回) 第 8 回対話システムシンポジウム, pp.100–101, 2017.