

言い換えを利用した対話行為推定の汎化性能向上

塚原 裕史 内海 慶

株式会社デンソーアイティラボラトリ 研究開発グループ

{htsukahara, kuchiumi}@d-itlab.co.jp

1 はじめに

対話行為認識は、対話の分析において基礎となるタスクである。対話行為は、発話における意図を抽象化し、ラベルで表現したものである。対話行為は音声対話のみではなく、インスタントメッセージ、SNS や twitter などのソーシャルメディアのタイムライン上でのユーザ間のインタラクションを分析する上でも有効であると考えられる [1, 2, 3]。また対話システムにおいても、ユーザの意図を正しく理解し、その意図に応じて適切な応答を生成するためにも、対話行為による対話制御は非常に重要になる [4]。その他、音声認識、話題推定、自動翻訳などへの応用もある。

これまで対話行為推定には、変換ベース学習、事例ベース、決定木、ニューラルネットワーク、隠れマルコフモデルなどの機械学習の手法が用いられてきた [5, 6]。教師あり学習では、大量のラベル付きの学習データを用意する必要がある。しかし、データの入力空間全体に渡って十分な密度で学習データを用意することは容易ではない。仮に大量のデータを入手できたとしても、それらへのラベル付けには多くのコストを要する。

後者の問題については、学習データの一部にのみラベル付けを行い、半教師あり学習を行うという手法がある。自然言語処理の分野では、これまでこのようなアプローチがしばしば適用されてきた [7]。一方、画像認識などの問題では、各学習データに拡大・縮小、回転、オクルージョンなどの摂動を加えることで、見かけ上のデータ数を増大し、汎化性能を向上させるデータ拡張 (data augmentation) の方法が多く用いられているが、自然言語処理においては、このような試みはまだあまり利用されていないように思われる。これは自然言語では画像に比べて摂動の与え方が難しいことやどのような摂動を与えるべきかということ、適用する問題に大きく依存し、画像の場合のような汎用的な摂動がそれほど自明ではないことなどが考えられる。

これまでに筆者らは、クラウドソーシングを利用し、

テキストによる雑談データの収集とそれらへの固有表現や対話行為などのアノテーション付けを行ったテキスト雑談コーパスの構築を行ってきた [8, 9]。この雑談コーパスによって対話行為の学習を行った識別器を音声対話システムに適用したとき、たとえ音声認識による誤認識がない場合でも、話し言葉に特有の文末における言い回しや省略などにより、しばしば誤認識が発生した。そこで、本稿では書き言葉的な表現から話し言葉的な表現への言い換え変換によるデータ拡張をテキスト雑談コーパスに適用し、対話行為の学習を行うことで、音声対話にも適用可能な対話行為の識別器を構築することを試みる。

本稿の構成は以下の通りである。まず、2節で我々のテキスト雑談コーパスを用いた対話行為の学習、3節で書き言葉から話し言葉への言い換え変換、4節で言い換え変換を用いたデータ拡張を行った場合の汎化性能の違い、5節でデータ拡張を行わない場合と行った場合での音声対話における対話行為認識結果の違いについて述べる。6節では関連研究について簡単に触れる。最後に、7節で結論を述べる。

2 対話行為認識

2.1 テキスト雑談コーパス

本稿で用いるテキスト雑談コーパス [8, 9] には、表 1 に示すような 8 種類の対話行為がアノテーション付けされた約 2 万件の発話がある。これらの対話行為は、対話の詳細な分析には不十分ではあるが、最低次の近似的な分析が行えると考えられる最小限のものに簡略化されている。

インスタントメッセージによる対話と同様に [3]、我々が収集したテキスト雑談においても、ある話者が発話を投入後、相手がテキスト入力に遅延している場合などに、相手の応答を待たずに、同じ話者が複数の発話を続けて投入し、その後、相手が前の発話から順

に回答を返すという非同期な対話が起る現象が少ないが見られた。主に、質問に対して回答を行う場合に発生し易く、他の場合にはほぼこのような非同期は見られなかった。これは質問の前の発話内容について強く関心があった場合に、先に前の発話へ応答し、質問に対しても応答義務を強く感じ、無視することはできず、後から質問への回答を返すために発生していると考えられる。

我々のテキスト雑談コーパスでは、このような非同期が発生していると考えられる箇所については発話の順序を入れ替えて同期したように補正してある。

ラベル	対話行為
:da1_greeting:	挨拶
:da2_yes_no_question:	真偽質問
:da3_yes_no_answer:	真偽回答
:da4_info:	情報提供／自己開示
:da5_change_topics:	話題転換
:da6_question:	質問
:da7_answer:	回答
:da8_feedback:	感想／意見

表 1: テキスト雑談コーパスにおいてアノテーション付けされている対話行為の種類 [8, 9].

2.2 素性

対話行為の認識には、主に表層的な情報、韻律的な情報、対話履歴などの情報が利用される。韻律情報は、音声対話では非常に重要であるが、テキスト対話では利用できない。本稿では表層素性として、bag-of-words, bag-of-bigrams を利用した。但し、単語として名詞、前置詞は対話行為推定にあまり寄与しないと考え、省略した。また、後で音声対話に応用することを考えて、句読点などの記号も省略した。また、バイグラムについては、これらの単語を排除した後に残った単語列を用いて構成した。対話履歴素性としては、直前の発話の話者と対話行為ラベルを用いた。話者素性については、現在の話者と直前の話者とが同じか否かの 2 値とした。

2.3 対話行為学習

対話行為の識別モデルには、決定木、サポートベクトルマシン、ロジスティック回帰の 3 つを考えた。また、これらの学習には Weka を利用した [10]。識別精度の評価には、10-分割交差検定を用い、Cohen κ 係数で

示してある。評価結果を表 2 に示す。サポートベクトルマシンとロジスティック回帰については、パラメタについてグリッドサーチを行って、もっとも結果が良かったものを示している。またサポートベクトルマシンについては、多項式カーネルと RBF カーネルとを比べて、RBF カーネルを用いた場合にもっとも良い結果が得られたので、その値を示している。対話行為数が少ないことから、どのモデルでも比較的良い値が得られているが、特にロジスティック回帰で良い結果が得られている。以降の議論では、識別器としてロジスティック回帰を用いる。

識別モデル	Cohen κ 係数
決定木	0.6564
サポートベクトルマシン	0.6693
ロジスティック回帰	0.7258

表 2: 対話行為の識別性能

3 話し言葉変換

本稿では書き言葉的な表現から話し言葉的な表現への言い換えについて考える。特に、フレーズ単位での言い換えと文末表現の言い換えについて考える。言い換えは、変換前の表現パターンとそれに一致する部分の置き換えとの組からなる規則を定義して行う。同じ変換前の表現パターンに対して複数の言い換えが可能な場合には、それぞれについて規則を定義する。また、今回、規則は人手で定義した。

フレーズ単位での言い換え規則の事例を表 3 に示す。フレーズ単位での言い換え規則は、あいさつやあいさつなど 8 種類のフレーズについて言い換え規則を定義した。

変換規則		
変換前	→	変換後
おはようございます	→	おはよう
おはようございます	→	おは
おはようございます	→	おっはー
どっちなのですか	→	どっちな
どっちなのですか	→	どっち

表 3: フレーズ単位での言い換え規則の事例

文末表現の言い換え規則の事例を表 4 に示す。規則の事例とその規則による変換事例を示した。文末表現

の言い換え規則については、173 種類の言い換え規則を定義した。

4 言い換え変換によるデータ拡張

言い換え変換では、与えられた発話文に対して、3 節で定義した言い換え規則のパターンで適用できるものがあるかどうかマッチングを行い、マッチした部分について表現の置き換えを行う。もし、どのパターンにもマッチしない場合には、何も変更は行わない。

データ拡張では、発話文にマッチするパターンがあった場合に、そのパターンについて定義されているすべての規則ごとに言い換え変換を行ったものを学習データに追加する。また、一つの発話文について複数のパターンがマッチする場合には、各マッチする部分ごとに、独立に規則による言い換えを行った結果を学習データに追加する。

このような言い換え変換によってデータ拡張を行った結果、学習データにおける発話数は仮想的に、約 2 万件から約 3 万件に、約 1.5 倍に拡大された。

表 5 に、データ拡張を行わない場合と行った場合における識別性能の違いを示す。識別器はロジスティック回帰である。データ拡張を行うことで、識別性能が向上していることがわかる。

ロジスティック回帰	Cohen κ 係数
データ拡張なし	0.7258
データ拡張あり	0.7685

表 5: データ拡張による識別精度の向上

5 音声対話への応用

テキスト雑談コーパスで学習した対話行為識別器を音声対話による雑談システムに対して適用した場合、音声対話では、テキスト対話に比べて、非常に発話が短いことに加えて、表層的な情報よりも、韻律的な情報の方が役割が大きいため、識別に韻律的な情報を素

性として利用していないことから、音声認識に誤りがないとしても、しばしば誤認識が発生する。特に、質問を行う発話においては文末で声が高くなるというような韻律的な特徴と同時に「ですか」のような文末表現が省略される傾向が高くなる。このような文末表現の省略はテキスト対話では発生しないことから、認識に失敗すると考えられる。

表 6 に、音声対話における対話行為推定結果の事例を示す。表における発話文は、音声対話システムにおけるユーザ発話の音声認識結果を示している。これらはどちらの発話でも、音声では文末で声が高くなっており、質問の発話であることがわかる。これらの発話文に対してテキスト雑談コーパスにおいてデータ拡張を行わないで学習した識別器による識別結果を見ると(表には発話の話者、直前の発話などの情報は省略してあるが、識別にはそれらを使っていることに注意)、質問ではなく、情報提示や感想などの間違った結果が認識されてしまっている。一方、データ拡張を行った識別器では正しく質問と認識されている。

6 関連研究

これまで自然言語処理の分野では、主にデータ拡張として、ラベル付きデータの数が十分でない場合に、ラベル無しデータを含めた半教師あり学習により、精度を向上する手法が主に考えられてきた。例えば、文書分類などの問題への適用がある [7]。

音声処理の分野では、音声の生成モデルを考えて、学習データに無いデータを補うという手法が考えられており、中国語の方言の発音をサポートベクトルマシンを用いて予測する問題などに応用されて、その有効性が報告されている [11]。

テキスト対話に対する対話行為の推定については、インスタントメッセージでの対話において、 n -グラムを用いた表層的な素性によりナイーブベイズ推定を利用した研究がある [3]。ここではタスクベースの対話が扱われており、対話行為は 12 種類からなる。素性は単語ユニグラム、バイグラム、トライグラムの場合

変換規則	
変換前 → 変換後	例
形容詞-ですね → 形容詞-ね	暑いですね → 暑ね
ですよ → ですね	好きですよ → 好きですね
ですか → ” (空文字列)	右ですか → 右

表 4: 文末表現の言い換え規則の事例

発話文	正解	データ拡張なし	データ拡張あり
手塚氏ってだれ	:_da6_question:	:_da4_info:	:_da6_question:
どれくらい打ってるの	:_da6_question:	:_da8_feedback:	:_da6_question:

表 6: 音声対話における対話行為推定結果の比較

について評価され、バイグラムでの精度がもっとも良く約 80%である。

人同士の雑談対話における対話行為の推定については、単語ユニグラム、バイグラムと文字トライグラム、クアドグラムを素性としてロジスティック回帰を用いた研究がある [12]。雑談での発話文は、とても短いものがあり、単語 n -グラムが素性に取り入れられているものと考えられる。ここで用いられた雑談コーパスは、1,259 件の対話からなり、67,801 件の対話行為アノテーションが含まれている。33 種類の対話行為が定義されている。1000 対話が学習に利用され、259 件の対話が評価用に利用され、約 61%の精度が得られている。

7 おわりに

本稿では、言い換え変換を利用したデータ拡張によって、対話行為推定の精度を向上する手法を提案し、その有効性をこれまでに筆者らが構築した対話行為アノテーション付きテキスト雑談コーパス [8, 9] を用いた学習によって示した。また、音声対話に特有の言い換えを用いたデータ拡張を行うことで、テキスト対話の学習データで学習した対話行為識別器であっても、音声対話の対話行為を正しく識別できるようになる可能性を示した。

今後、この対話行為推定を利用し、音声対話による雑談生成への応用を進めていきたい。

参考文献

- [1] Su Nam Kim, Lawrence Cavedon, and Timothy Baldwin. Classifying dialogue acts in multi-party live chats. In *Proceedings of the 26th Pacific Asia Conference on Language, Information, and Computation*, pp. 463–472, Bali, Indonesia, November 2012. Faculty of Computer Science, Universitas Indonesia.
- [2] Elina Zarisheva and Tatjana Scheffler. Dialog act annotation for twitter conversations. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 114–123, Prague, Czech Republic, September 2015. Association for Computational Linguistics.
- [3] Edward Ivanovic. Dialogue act tagging for instant messaging chat sessions. In *Proceedings of the ACL Student Research Workshop*, pp. 79–84, Ann Arbor, Michigan, June 2005. Association for Computational Linguistics.
- [4] Ehud Reiter and Robert Dale. *Building Natural Language Generation Systems*. Cambridge University Press, 2000.
- [5] Andreas Stolcke, Noah Coccaro, Rebecca Bates, Paul Taylor, Carol Van Ess-Dykema, Klaus Ries, Elizabeth Shriberg, Daniel Jurafsky, Rachel Martin, and Marie Meteer. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Comput. Linguist.*, Vol. 26, No. 3, pp. 339–373, September 2000.
- [6] Pavel Král and Christophe Cerisara. Dialogue act recognition approaches. *Computing and Informatics*, Vol. 29, No. 2, pp. 227–250, 2010.
- [7] Kamal Nigam, Andrew Kachites McCallum, Sebastian Thrun, and Tom Mitchell. Text classification from labeled and unlabeled documents using em. *Mach. Learn.*, Vol. 39, No. 2-3, pp. 103–134, May 2000.
- [8] 裕史塚原, 慶内海. オープンプラットフォームとクラウドソーシングを活用した対話コーパス構築方法. 言語処理学会第 21 回年次大会 (NLP2015), pp. 147–150, 2015.
- [9] Hiroshi Tsukahara and Kei Uchiumi. System utterance generation by label propagation over association graph of words and utterance patterns for open-domain dialogue systems. In *IN PACLIC-29*, 2015.
- [10] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The weka data mining software: An update. *SIGKDD Explor. Newsl.*, Vol. 11, No. 1, pp. 10–18, November 2009.
- [11] Chu-Cheng Lin and Richard Tzong-Han Tsai. A generative data augmentation model for enhancing chinese dialect pronunciation prediction. *Trans. Audio, Speech and Lang. Proc.*, Vol. 20, No. 4, pp. 1109–1117, May 2012.
- [12] Hiroaki Sugiyama, Toyomi Meguro, Ryuichiro Hishinaka, and Yasuhiro Minami. Open-domain utterance generation for conversational dialogue systems using web-scale dependency structures. In *Proceedings of the SIGDIAL 2013 Conference*, pp. 334–338, Metz, France, August 2013. Association for Computational Linguistics.