

雑談システムにおける Twitter データからの 統計的バックチャネル応答抽出手法

Statistical Backchannel Extraction from Twitter Data in Casual Dialogue System

福田 拓也*¹
Takuya Fukuda

若林 啓*²
Kei Wakabayashi

*¹筑波大学 図書館情報メディア研究科
Graduate School of Library, Information and Media Studies, University of Tsukuba

*²筑波大学 図書館情報メディア系
Faculty of Library, Information and Media Science, University of Tsukuba

The backchannel plays an important role in smooth communication. For dialogue system, performing appropriate backchanneling is a significant factor that makes more natural conversation. However, many existing dialogue systems have poor backchannel patterns and only can produce simple responses. In this paper, we propose a method to extract various backchannels that are suitable for user utterance with no restriction of the diversity of backchannels. We conduct an experiment that compares the proposed method with a naive method based on similarity and character length limit. The result shows that the proposed method achieves a higher evaluation score.

1. はじめに

近年、雑談システムに関する研究が活発になされており、注目を集めている。雑談システムとは特定のタスク達成を目的としない、人間との自然な会話の成立を目指した対話システムである。雑談システムの応答方法としてあらかじめ用意した発話候補データからシステムの応答として適切な文章を選択するものがある。発話候補データは Twitter データなどを利用することにより低コストで膨大な量のデータの獲得が可能である。

雑談システムの応答内容について、図 1 の「身長をもっと伸ばしたい」というユーザ発話に対する応答を考える。応答例の 1 つとして「伸ばすには成長ホルモンが特に重要なんだって」というような情報提供を行う生産的な応答が挙げられる。しかし、人間同士の会話ではこのような応答ばかりがやりとりされるのではなく、「あ〜分かる!」というような一見、非生産的な応答も多く用いられる。このような聞き手側が行う会話中の短いリアクションをバックチャネルと呼ぶ。このバックチャネルはコミュニケーションをより円滑で効率的なものにする働きがあり [堀口 88], 多様なバックチャネルをユーザ発話に合わせて柔軟に返答することは人間との自然な対話を実現するために重要な要素であると考えられる。

本研究では Twitter データから適切なバックチャネルを抽出する手法を提案する。これにより、ユーザ発話に合わせた多様なバックチャネル応答を抽出することを目指す。

2. 関連研究

バックチャネル応答を行う研究として、ルールに基づく適切なバックチャネル生成手法 [佐藤 01] がある。これは入力発話の表層構造と意味情報を獲得し、適切なバックチャネルを出力するものである。また、[山口 16] ではバックチャネルの形態と先行発話の統語構造や韻律的特徴の関係について分析し、バックチャネルの形態予測と生成を行っている。ここではバック

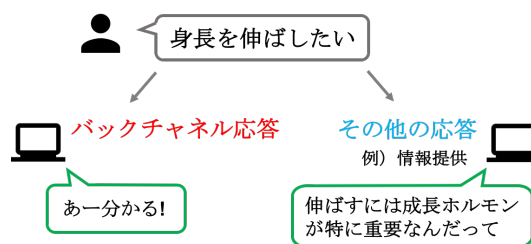


図 1: 応答パターン

クチャネルについて応答系感動詞や感情表出系感動詞を対象としている。他にもバックチャネル応答を考慮したシステムの研究として [井上 15] などがあるが、これらの研究で扱われるバックチャネルはいずれも単純なもので「なるほど」や「ほー」などユーザ発話内容に依存しない応答である。このような単純な応答はシステムがユーザの話に対して関心が無いという印象を与えてしまうことが考えられる。

これに対してより豊富なパターンを扱った研究として、[Inaba 16] は「大丈夫ですか」や「残念です」などのバックチャネルを返答クラスとして扱い、入力文がどの返答クラスに属するかというバックチャネルについての多値分類問題として定式化している。この研究ではあらかじめ 44 種類のバックチャネルを返答クラスとして定義している。しかし、例えばバックチャネルは「そうですね」などのように相手の発話内容にあまり関係なく汎用的に使えるものもあれば、「かつこいい!」など相手の発話内容についてある程度踏み込んだ限定的な使われ方をするものもあり、実際の会話で使用されるバックチャネル応答はニュアンスの差も含め多岐に渡る。このため、あらかじめバックチャネルの種類を限定しない方法が望ましい。このため本研究ではより豊富なパターンのバックチャネルの抽出のため、バックチャネルの種類を制限せず、入力文に対して発話候補データ中の各文章が適切なバックチャネルであるかどうかの 2 値分類を行う問題として扱う。

連絡先: 筑波大学 図書館情報メディア研究科

〒 305-8550 茨城県つくば市春日 1-2

E-mail: s1621632@slis.tsukuba.ac.jp

3. 提案手法

提案手法の概要を図2に示す。入力ユーザ発言文 u とし、応答文の候補集合を S とする。本研究では、候補集合 S は Twitter データに含まれる Tweet から構成する。ただし、後述する特徴量抽出の条件を満たすため、 S はある Tweet に対する Reply であり、かつ別の Tweet によって Reply されているような Tweet の集合とする。それぞれの応答文候補 $i \in S$ について、ユーザ発言文 u に依存する特徴抽出関数 f を適用した結果得られるベクトルを $\mathbf{x}_i = f(i, u)$ とする。入力文に対する、応答文 i の適切なバックチャネルらしき表すスコアとして以下を定義する。

$$b(\mathbf{x}_i) = P(y = 1 | \mathbf{x}_i) \quad (1)$$

これは素性 \mathbf{x}_i が与えられた場合に応答文 i が適切なバックチャネルである確率を表す。提案手法では定義した $b(\mathbf{x}_i)$ の推定にロジスティック回帰を用いる。ロジスティック回帰は2クラス分類問題における一般化線形モデルの一種であり、以下の式として表される。

$$\text{logit}\{b(\mathbf{x}_i)\} = \log \frac{b(\mathbf{x}_i)}{1 - b(\mathbf{x}_i)} = \alpha + \beta^T \mathbf{x}_i \quad (2)$$

また、ロジスティック回帰の確率モデルを以下に示す。

$$b(\mathbf{x}_i) = \frac{\exp(\alpha + \beta^T \mathbf{x}_i)}{1 + \exp(\alpha + \beta^T \mathbf{x}_i)} \quad (3)$$

特徴抽出関数 f は、以下に説明する3つの素性を抽出し、3次元の実数値ベクトルを得る関数とする。まず、1つ目の素性は入力文と応答文の繋がりにある。あらかじめ Twitter データの Tweet j とその Tweet に対する Reply i のペアデータ集合 P を用意しておく。入力文 u が与えられたとき、それぞれのペアデータ $(j, i) \in P$ について、 u と j の TF-IDF ベクトルを作成し、そのコサイン類似度を計算する。このとき求めた類似度を、応答文候補 i の素性として用いる。判別の際、類似度を考慮することにより入力文と似た Tweet であるほど、そのペアである Reply は入力文に対してつじつまのあった応答が期待できると考える。

2つ目の素性は応答文の情報量である。バックチャネルの特徴の1つとして例えば情報提供を行う応答と比べて、文章としての情報量が少ないことが考えられる。このときバックチャネル候補文 i に含まれる各単語の IDF (Inverse Document Frequency) を計算し、それらの合計値を文章の情報量として求める。IDF は語の重要度を表す単純な指標であるが本研究では語の情報量として考え、重要な語ほど情報量も多いとする。これを素性として用いることで IDF の合計値が大きい文章であればバックチャネルらしくない文章とし、小さければバックチャネルらしい文章として判別できると考える。

最後に、応答文から期待される次の発言の文字数を用いる。これは2つ目までの素性を用いた場合では、システムの応答候補として図2の Reply の「にゃー」のように一見バックチャネルらしいがその応答に対してユーザが次にどのような応答をすればよいか困ってしまうような応答も適切な応答として判断してしまう。このような応答は会話の停滞をもたらす、システムの評価を下げてしまうことが考えられる。このため会話を停滞させないより良いバックチャネルとして、聞き役として相手の長い発言を期待できる応答が1つの観点であるとして、応答文の次の発言の平均文字数を素性として用いる。Tweet 集合 S の条件より、バックチャネル候補 $i \in S$ は、Twitter データ

において、別の Tweet に Reply されていなければならないが、ここで i に対する Reply の集合を R_i とする。 i に対する Reply $k \in R_i$ それぞれについて、記号や「www」などの自然言語でない表現を除いた文字数 L_k を求め、その平均 $\frac{1}{|R_i|} \sum_{k \in R_i} L_k$ を i の素性とする。これにより判別の際、平均文字数が少ない文章であれば良いバックチャネルではないものとして相手が返答に困るような応答をさけることが期待できると考える。表1は応答に対する次発言の平均文字数の例である。

これらの素性を用いてユーザ発言に対する Reply のバックチャネルスコアを計算し、最もスコアの高いものを応答として出力する。

Reply	Average (next reply length)
ウホッ	7.07
んちゅー	8.36
なるほど	19.3

表1: 次発言の平均文字数の例

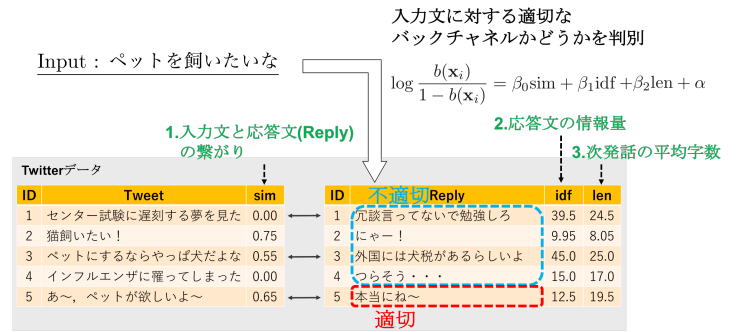


図2: 提案手法の概要

4. 評価実験

実験では提案手法が出力したバックチャネルが入力に対して適切かどうかを Tweet の類似度に基づく手法と比較し、評価する。

4.1 実験用データ

4.1.1 発言候補データと前処理

今回の実験では収集期間が2012/04/01~2012/04/07の Tweet-Reply データを使用した。このとき Tweet または Reply にハッシュタグや URL を含むデータは文章のコンテキストがより複雑になってしまうため除いた。さらにどちらかが「。」や「？」などの記号のみの文章であった場合も不適切な文章として除いた。以上によって収集した、Tweet が紐づけられた約1,500万件の Reply データを発言候補データとして用いた。

4.1.2 モデルの学習用データ

モデルの学習用データとしてまず上述の発言候補データとは別に Tweet データを用意する。そしてその中からあらかじめ人手で異なる話題を含んだ文章になるように選択する。今回は入力発言の件数を100とした。これらを学習用データの入力発言とする。次に選択した各入力発言と、発言候補データ中の Tweet との類似度が上位10000件のデータから、3章で述べた各素性について値が最も高い Reply と低い Reply を選び、その際の素性値の組を学習データとして採用する。これにより

1つの入力に発話につき6件の学習データが用意される。以上により、計600件の学習データについて第1著者が入力発話とReplyを見て、Replyが適切なバックチャネルかどうかを、適切ならば正例：1、不適切ならば負例：0としてラベル付けを行った。

4.2 実験方法

今回の実験では1000件の評価データを作成し、それをユーザの発話とみなす。評価データは学習用データと同様に異なる話題を含んだ文章になるよう選択したもので、このとき評価データとして不快な内容の文章は除いた。

評価データに対して、提案手法は素性1の値の上位10000件のデータからバックチャネルスコア(式3)が最も大きい候補文を出力する。また、比較手法はナイーブな手法として文字数制限をかけた中でTweetの類似度が最も高いReplyを出力するものとして設定した。比較手法の文字数の上限は10文字とした。

実験は手法ごとに出力した評価データに対する応答を「○：適切」、「▲：適切とも不適切とも言えない」、「×：不適切」の三段階で被験者に評価してもらう。このとき、評価の一致率についてFleissによるk係数[Fleiss 71]を適用すると0.398となり、低い一致であった(poor agreement)。この結果を受け、以下の手順で手法の評価値を求める。

- 被験者全員が同じ評価をつけたデータ以外を評価時のノイズを考慮するため除去
- ○を2点、▲を1点、×を0点として各データのスコアに変換
- 全てのデータのスコアの平均値を、手法の評価値として使用

実験はクラウドソーシングサービスのLancers*1を利用し、1つのデータに対して3名に評価をしてもらう。これにより手法ごとに計3,000件の実験データを用いて評価する。

4.3 結果と考察

実験の結果を図3に示す。提案手法では3名全員が「○：適切」としたデータは273件(0.592%)、「▲：適切とも不適切とも言えない」が19件(0.041%)、「×：不適切」が169件(0.367%)となった。また、比較手法では3名全員が「○：適切」としたデータは260件(0.526%)、「▲：適切とも不適切とも言えない」が7件(0.014%)、「×：不適切」が227件(0.460%)となった。

次に、手法ごとに平均評価値を表2に示す。提案手法では平均評価値は1.226、比較手法では1.067であった。2項検定を行なった結果、提案手法と比較手法の評価値の差は有意であった(p値=0.012)。このことから、提案手法では比較手法よりも適切なバックチャネル応答を行うことが期待できると考える。表3は提案手法が適切な応答ができた例である。

今回、適切な応答について提案手法と比較手法の割合に大きな差は見られなかったが、不適切な応答については提案手法の方が割合は小さかった。これは提案手法の方が明らかに不適切であるような応答をしにくいことを示している。

5. おわりに

本研究ではバックチャネルの種類を制限せず、ユーザ発話に合わせた多種多様なバックチャネルを抽出する手法を提案し

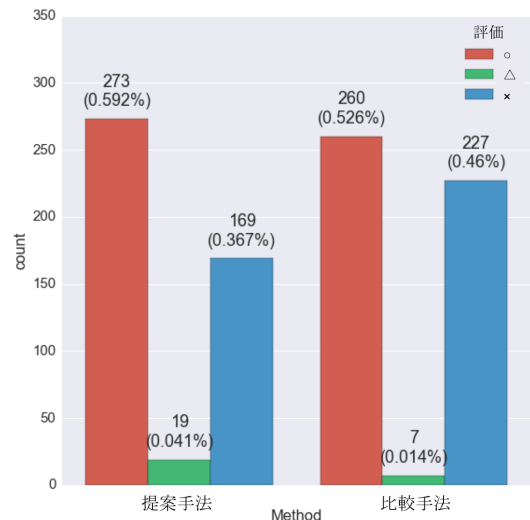


図3: 実験結果

手法	平均評価値	標準誤差	件数
提案手法	1.226	0.954	461
比較手法	1.067	0.992	494

表2: 各手法の平均評価値 (p値=0.012)

た、その際に入力文と応答文の繋がりや文章の情報量、応答文から期待される文字数を素性とし、これらを用いた適切な応答を目指した。このとき応答を行うタイミングは所与として、さらに笑いやうなづきなどの非言語行動については扱わないものとした。

実験では入力文に対して適切なバックチャネルを出力できたかを、類似度と文字数制限に基づくナイーブな手法と比較した。結果は被験者全員が同じ評価をしたデータについて提案手法の方が有意に高い評価を得ることができた。

課題としては、今回はTweet-Replyのペアデータはつじつまのあった応答がなされていると仮定している。だが、データ中には一見不自然に感じるハイコンテキストな応答も存在しており、このようなデータについて提案手法の性能は大きく影響されてしまう。このため、そのようなデータをできるだけ除くなどする必要があると考える。また、バックチャネルには相手の話の補完や繰り返し、さらには言い換え表現としての応答パターンも考えられる。しかし、このような応答表現に対して提案手法は基本的に文章としての情報量が少ないものがより適切なバックチャネルとしており、今後はTweetとReplyについて例えばTweetとReplyが似た文章かなど、どのような関係にあるかを考慮する必要がある。

評価データ	システムの応答
渋谷の雰囲気は結構好きではある。臭いけど	それな
身長と目しかいいとこなくて泣いた	まじですかw
俺頭悪すぎクソワロタ wwwwww	詳細kwsk
やはりイタリア人女性はいいな!	間違いない!
自分が情けなくて泣いてしまいそう。	大丈夫か

表3: 出力された応答例

*1 <http://www.lancers.jp>

謝辞

本研究の一部は、JSPS 科研費（課題番号 16H02904）の助成によって行われた。

参考文献

- [Fleiss 71] Fleiss, J., et al.: Measuring nominal scale agreement among many raters, *Psychological Bulletin*, Vol. 76, No. 5, pp. 378–382 (1971)
- [Inaba 16] Inaba, M. and Takahashi, K.: Backchanneling via Twitter Data for Conversational Dialogue Systems, in *The 18th International Conference on Speech and Computer (SPECOM-2016)*, pp. 148–155 (2016)
- [井上 15] 井上 昂治, 河原 達也: 自律型アンドロイド Erica のための音声対話システム, 言語・音声理解と対話処理研究会, Vol. 75, pp. 21–24 (2015)
- [佐藤 01] 佐藤 康将, 目良 和也, 市村 匠, 山下 利之, 相沢 輝昭, 吉田 勝美: 肯定/否定意図を検出するチャットシステムのためのあいづち生成手法, ファジィシステムシンポジウム講演論文集, Vol. 17, pp. 509–512 (2001)
- [山口 16] 山口 貴史, 井上 昂治, 吉野 幸一郎: 傾聴対話システムのための言語情報と韻律情報に基づく多様な形態の相槌の生成 (「フィールド研究とインタラクション」および一般), 言語・音声理解と対話処理研究会, Vol. 76, pp. 49–54 (2016)
- [堀口 88] 堀口 純子: コミュニケーションにおける聞き手の言語行動 (聴解の指導<特集>), 日本語教育, No. 64, pp. p13–26 (1988)