

Twitter を用いた非タスク指向型対話システムの発話生成

Utterance Generation for Non-task-oriented Dialogue Systems using Twitter

稲葉 通将*¹ 神園 彩香*¹ 高橋 健一*¹
 Michimasa INABA Sayaka KAMIZONO Kenichi TAKAHASHI

*¹ 広島市立大学大学院情報科学研究科
 Graduate School of Information Sciences, Hiroshima City University

Recently, computerized dialogue systems are studied actively. Non-task-oriented dialogue systems that handle domain-free dialogues like chats are expected to be applied in various fields. This paper provides an utterance generation method for non-task-oriented dialogue systems. We search twitter data by topic words and acquire sentences. The sentences are filtered by rules and scored on the basis of training data. We acquire the sentences which have a high score as utterances. The results of an experiment demonstrate that the proposed method can generate appropriate utterances with a high degree of accuracy.

1. はじめに

近年、道案内や情報検索など、特定のタスク達成を目的としたタスク指向型対話システムだけではなく、タスク達成を目的とせず、人間と雑談を行う非タスク指向型対話システムの研究も活発になりつつある。しかし、非タスク指向型対話は話題に制限がないこと、また、目的がないため対話の展開を前もって予測することが困難であることなどから、自然な対話が可能な非タスク指向型対話システムの実現には課題も多い。

それらの課題のうち、本研究では発話生成を対象とする。既存の非タスク指向型対話システムの発話生成手法としては、ルール・テンプレートをを用いた手法、対話ログを用いた手法などが提案されている。ルール・テンプレートをを用いた手法を採用している対話システムの代表例が ELIZA[1] である。ELIZA はユーザの発言と応答ルールとの単純なパターンマッチングで動作する。応答としては、応答ルールで書かれた発話をそのまま使用するか、もしくはユーザの発言中の単語をルールに従って穴埋め式のテンプレートに埋め込み、応答を行う。一方、対話ログを用いた手法を採用している対話システムとしては Jabberwacky*¹ が挙げられる。Jabberwacky は大規模な対話ログのデータベースから、ユーザの発言に類似した発言を検索し、その発見した発言に対する応答をシステムの応答として用いている。ルール・テンプレートをを用いた手法では応答ルールの作成は全て人手で行う必要があることから、作成コストが極めて大きい。また、テンプレートを使用することから、発言のバリエーションが少ないという問題がある。一方、対話ログを用いた手法では、バリエーションの問題は無いものの、システムが扱える話題を増やすためには、その話題について話された十分な量の対話ログをあらかじめ用意する必要があり、そのコストは大きい。

そこで本研究では、入手が比較的容易である Twitter のデータを用いて任意の話題に関する発話を生成する手法を提案する。提案手法が生成する発話は、ある話題に関する事実、もしくは意見・感想を述べる発話とする。これは発話行為を示すタグである SWBD-DAMSL タグ [2] では Statement-non-opinion と

Statement-opinion に相当する。非タスク指向型対話が収録されている Switchboard コーパス [2] には SWBD-DAMSL タグが付与されており、コーパス内でそれぞれのタグが占める割合は Statement-non-opinion が 36%、Statement-opinion が 19% である。したがって、本手法により、非タスク指向型対話で使用される発話の約半分がカバーできることになる。

なお、本研究ではユーザの発言に対して対話システムがどのような応答を返すべきかという対話戦略・対話管理については対象とせず、応答候補となる発話の生成のみを扱う。

2. 発話生成手法

2.1 概要

本節では、Twitter データを用いて発話候補を自動生成する手法について述べる。本手法は、任意の話題を入力とし、その話題に関する発話を複数出力する。

本研究では、汎用性の高い発話生成を目指し、下記の条件全てを満たす発話を正解発話とする。

- ある話題に関する事実、もしくは意見・感想を述べた発話であること
- 日本語として適切であり、意味の通じる発話であること
- その発話だけで意味・意図が理解できること
- 限定された場所・時間でしか使用できない発話ではないこと

このうち、3点目の「その発話だけで意味が理解できる発話であること」は汎用性の高い発話が生成されることが望ましいことから設けた条件である。例を挙げれば、「○○の方が良いですね」はこの発話だけでは何に関して良いのか、ということがわからないためこの条件に反している。4点目の「極めて限定された時間でしか使用できない発話ではないこと」は発話の使用に際して、厳しい時間的制約が存在する発話は使用機会が少なく、望ましくないことから設けた条件である。例えば「今日から○○が発売されます」という発話は、その発売日以外に使用した場合不適切な発話となるため、時間的制約が存在する発話であるといえる。

提案手法では、まず入力された話題で Twitter データを検索し、話題を含む文を抽出する。その際、前処理としてルール

連絡先: 広島市立大学大学院情報科学研究科
 〒 731-3194 広島市安佐南区大塚東 3-4-1
 E-mail: inaba@hiroshima-cu.ac.jp

*¹ <http://www.jabberwacky.com>

に基づいたフィルタリングを行う。次に各文を教師データに基づいて点数付けし、その点数がしきい値以上の文を獲得する。最後に、後処理として語尾の変形を行い、発話として使用できる形で出力する。

本手法の基本的な考え方は、Twitter データを用いることにより、容易に大量のデータが入手できるという利点を生かし、発話として使用不可能な文を厳しい基準によって排除し、使用可能な文を精度よく獲得するというものである。したがって、フィルタリングルールは使用可能な文をも排除してしまう可能性のあるルールであっても、使用不可能な文を排除するのに有効なルールであれば採用する。また、点数付けについては、点数が下がりやすく上がりにくい非対称な手法により行い、少しでも使用不可能な可能性のある文を積極的に排除する。

2.2 Twitter からの文抽出

本節では、次節で述べる点数付け対象となる文の抽出手順について説明する。

まず、入力された話題で Twitter データを検索し、話題を含むツイートを含めて全て抽出する。この際、URL を含むツイートは除外する。次に、ツイートを句点および記号で文単位に分割し、その中で話題が含まれている文のみを抽出する。

次に、文を Mecab[3] を用いて形態素解析する。ここで、以下のフィルタリングルールのいずれかに該当する文は除去される。

ルール 1 単語数が 7 個以下の文 (ただし、話題は複数の単語からなっても 1 単語とみなす)

ルール 2 話題と名詞が連続している文

ルール 3 人名、代名詞が含まれている文

ルール 4 文頭が記号、助詞、助動詞、接続詞のいずれかの文

ルール 5 文末が格助詞、係助詞、接続助詞、並列助詞、名詞 (形容動詞語幹は除く) のいずれかの文

ルール 6 時間を限定する語、数値が含まれている文

ルール 7 不十分な比較が含まれている文

ルール 8 同一の単語が 2 個以上連続している文

フィルタリングルールは予備実験により、ヒューリスティックに決定した。

まずルール 2 であるが、話題と名詞が連続している場合、話題とはあまり関係ない文が抽出されることがあるために設けるルールである。例えば話題が「アメリカ」であるとき、「アメリカザリガニ」に関する文が抽出されることは望ましくない。

ルール 3 に該当する場合、その発話だけで意味・意図が理解できない文が多く抽出されるため設けたルールである。人名に関しては、有名ではない人名が含まれているとき、そのような場合が頻発する。代名詞はその先行詞が同一文中に含まれないことが多いためである。

ルール 4、ルール 5 に関しては、記号により文分割を行ったことによる、文の途中で不適切な分割が起こった文を除外するためのルールである。例えば話題語が「ボウリング」で、「二次会でボウリングは…やっぱしんどいなあ」というツイートから取得される文は「二次会でボウリングは」となる。これは、この発話だけで意味・意図が理解できないことから不正解発話であるが、本ルールにより文末が助詞-係助詞のため、除外される。

ルール 6 は極めて限定された時間でしか使用できない発話を除外するためのルールである。時間を限定する語とは「今日」や「来週」などの語を指す。

ルール 7 は「コーヒーよりも良いですね」のような比較対象が存在しない文を除外するためのルールである。具体的には、「方/が」、「ほう/が」のどちらかが文中に含まれている場合、「(名詞)/と」、「(名詞)/や」、「より」のいずれかが含まれていない場合は除外する (「/」は単語の区切り、括弧内は品詞を表す)。また、「より」が含まれている場合、「(名詞)/と」、「(名詞)/や」、「方/が」、「ほう/が」のいずれかが含まれていない場合は除外する。

ルール 8 は「すごいいいいいい」や「嫌だ嫌だ嫌だ嫌だ」のように、強調のために同じ単語を連続させる表現が含まれる文を除外するためのルールである。

2.3 文の点数付け

2.3.1 助詞以外の点数

本手法では、文中の単語に点数付けを行い、その点数を用いて文の点数を計算する。単語の点数は教師データを用いて計算する。教師データは 2.1 節で述べた基準に従って人手で判定を行った複数の正解発話と不正解発話からなる。

点数として、その単語の正解発話における出現確率と不正解発話における出現確率の比を用いる。ただし、単語は表層形ではなく基本形に直したものをを用いる。単語 w の点数 x_w の計算式を以下に示す。

$$x_w = \frac{A}{B} \div \frac{C}{D} \quad (1)$$

式中の A は教師データにおいて、単語 w の正解発話全体における出現回数、 B は正解発話全体の総単語数、 C は単語 w の不正解発話全体における出現回数、 D は不正解発話全体の総単語数である。したがって、点数が 0 に近ければ近いほど不正解発話に出現しやすい単語であることを示し、1.0 より大きければ大きいほど正解発話に出現しやすい単語であることを示す。

ただし、出現回数が n 回未満の場合、点数は付与しない。実験では $n = 5$ を用いた。また、話題に関しても点数を付与しない。

2.3.2 助詞の点数

助詞の点数については、助詞単体では点数の意義が小さいと考え、助詞とその前後の品詞を用いた 3-gram を用いる。助詞の点数についても他の品詞と同様に式 1 により、3-gram の出現回数を用いて計算する。

2.3.3 文の点数

文の点数は、以下の式により計算する。

$$\text{SentenceScore} = \prod_{w \in W} f_{\alpha}(x_w)$$

$$f_{\alpha}(x_w) = \begin{cases} x_w & (x_w < \alpha) \\ \alpha & (x_w \geq \alpha) \end{cases}$$

ここで、 W は点数が付与された文中の単語の集合を表し、 w は各単語を表す。また、 x_w は教師データから計算される w の点数である。 α は点数の上限を決定するパラメータである。実験では $\alpha = 1.40$ を用いた。

この式は、文の点数は α より小さい単語の点数はそのまま掛け合わされ、 α 以上の単語の点数の場合、 α が掛け合わされるということを意味する。このように、本手法では点数に上限

表 1: 文の点数の計算例

単語	点数付け対象	単語の点数
ココア	-	-
は	話題, は, 副詞-助詞類接続	1.08
とても	とても	7.84
美容	-	-
に	名詞-一般, に, 形容詞-自立	2.69
良い	良い	6.73
らしい	らしい	4.25
よ	助動詞, よ, EOS	2.64

表 2: 実験で使用した話題

YouTube	Yahoo	Facebook
かぼちゃ	大根	ゴーヤ
ANA	JAL	au
ユニクロ	Gap	H&M
iPad	Xperia	楽天銀行
ワールドカップ	はやぶさ	スカイツリー
ハートロッカー	告白	アバター
岡村隆史	板野友美	大島優子
ニコニコ動画		

を与えることにより、文の点数が下がりやすく上がりにくい点数付けを実現する。

表 1 に「ココアはとても美容に良いらしいよ」(話題: ココア)の点数付けの例を示す。この例では、「ココア」は話題のため、「美容」は教師データにおける出現回数が 5 回未満であったため、それぞれ点数が付与されていない。単語の点数に着目すると、1.40 点以上が 5 つと 1.08 点が 1 つなので、この文の点数は、 $1.40^5 * 1.08 = 5.81$ となる。

2.4 語尾の変形

最後に、文の語尾を変形する。語尾はですます調(例: つけ麺はおいしいです)よりも、「～ですよね」「～ますよね」「～でしょうか」のような語りかけ口調(例: つけ麺はおいしいですよね)の方が、ユーザに親近感を与え、対話も盛り上がりやすいと考えられる。したがって、語尾は可能な限り語りかけ口調に変形する。ただし、語りかけ口調にすると日本語として不自然になる場合もあるため、必要に応じてですます調にも変形する。

変形は文の単語と品詞情報を用いたルールベースで行う。変形ルールはヒューリスティックに決定したものをを用いた。

3. 評価実験

3.1 実験設定

提案手法の有効性を確認するため、評価実験を行った。

使用した Twitter データは 2010 年 1 月から 12 月までの 12 ヶ月間収集した約 15 億ツイートである。発話を生成する話題は、Google 年間検索ランキング 2010^{*2}における 9 項目の各トップ 3 のキーワード 25 個とした^{*3}。表 2 に実験で使用した話題を示した。

教師データは発話生成に使用した Twitter データと同じデータを用いて作成した。教師データは大学生 1 名が任意の話題

表 3: 上位 10 発話の評価結果

手法	正解率
提案手法	94.8 %
点数付け手法のみ	80.4 %
フィルタリング + SVM	88.0 %
SVM (ベースライン)	85.0 %

で Twitter データを検索し、ツイート中の話題を含む文について正解発話と不正解発話の判定を行うことで作成した。作成した教師データにおける話題の種類は 59、正解発話数は 1866 個、不正解発話数は 6581 個の計 8447 個である。なお、教師データにおける話題は、表 2 と同一の語、および同義語は含まれていない。

性能比較のため、以下の手法による発話文の獲得も行った。

1. 点数付け手法のみ
2. フィルタリング + SVM
3. SVM(ベースライン)

まず 1. は、ルールによるフィルタリングを行う前の文に対して本論文で提案した点数付け手法を適用したものである。2. は提案手法と同様のフィルタリングを行った後、提案手法の点数付けの代わりに SVM を用いたものである。3. はベースライン手法であり、フィルタリングを行わず、SVM のみを用いたものである。SVM の実装として LIBSVM[4] を用い、正解発話クラスに所属する事後確率を点数として用いた。SVM の学習に用いたデータは、提案手法で点数付けに用いた教師データと同一とし、使用した素性は提案手法において点数付け対象となる要素(助詞以外の単語の基本形、助詞とその前後の品詞の 3-gram)とした。カーネルは線形カーネルを用いた。

各手法を用いて生成された発話のうち、各話題について文の点数の上位 10 件、計 250 発話を人手で評価した。評価は大学生 3 名が行い、多数決によって最終的な評価を決定した。3 名の評価の κ 値は 0.79 であり、かなりの一致が確認された。

3.2 実験結果

表 3 に各話題について、文の点数上位 10 件を評価した結果を示す。表の正解率は、250 発話のうち、人手によって正解発話と判定された発話の割合を示す。

表中の「提案手法」と「点数付け手法のみ」の正解率を比較すると、提案手法が優れていることが確認できる。ここから、ルールによるフィルタリング手法の有効性が確認できた。また、「提案手法」と「フィルタリング + SVM」を比較した場合でも、提案手法のほうが優れている。よって、点数付け手法についても有効性が確認できた。同様に「ベースライン」よりも「提案手法」のほうが正解率が高い。以上のことから、提案手法の有効性が示されたといえる。

一方で、「点数付け手法のみ」と「SVM(ベースライン)」を比較した場合、ベースラインのほうが正解率が高いという結果となった。これは、第 2.2 節で述べたフィルタリングルールのうち、ルール 8 が適用できなかった影響が大きい。例えば、「iPad ほししいいいいいいいいいい」に点数付けを行うと、複数回出現する「いい」の全てに点数が付与されることになる。「いい」の点数は 1.85 点であり、結果として文の点数は非常に大きいものとなる。「点数付け手法のみ」にはこのような発話が多く含まれていたため、正解率が低くなる結果となった。

*2 <https://sites.google.com/site/japanzgeist/>

*3 項目間で 2 件重複あり

表 4: 生成された発話の例

話題	発話
Yahoo	日本人の中で検索エンジンと言えば Yahoo なんですね
ゴーヤ	ゴーヤって高血圧や糖尿病の改善にとっても効果のある野菜なんです
au	au は、シンプルなデザイン路線だとセンスのよさが光ります
H&M	H&M のバッグや靴はレディースよりメンズの方がかわいいものがある気がします
Xperia	Xperia いい感じなんだけど、やっぱりおサイフケータイではないのでしょうか
スカイツリー	スカイツリーって本当に高いんですね
アバター	アバターは内容も面白いけど、世界観が綺麗な作品だと思います
板野友美	AKB では板野友美が可愛いと思います
ニコニコ動画	ニコニコ動画って表現の場でもあるんですよね

表 4 に各話題について正解発話の一部を示した。ここから、話題に関する事実や意見を述べた発話が正しく生成されていることがわかる。

3.3 考察

提案手法によって生成された発話の誤りの原因について調査したところ、以下が見られた。

- 誤字・脱字が存在する場合

提案手法により「Xperiaってなんだと思ったら DoCoMo の新しいスマートフォンなのではないですか」という発話が生成されたが、発話中の「スマートフォン」は「スマートフォン」が正しい。誤字・脱字については訂正手法がこれまでに提案されていることから [5], それらの手法を適用することで、訂正を行うことが可能である。

- 一文では意味・意図が不明な場合

提案手法により「ANA だと国内線でも言いますよね」という発話が生成された。しかし、この発話単体では何を言うのか、ということが読み取れないため不正解発話と判定された。この問題は名詞句の省略により発生していることから、ゼロ照応解析などにより、省略されている語を補うことが必要である。

- 語尾が特殊な場合

提案手法により「ユニクロのいいところはカラーとサイズの多さにあると思うウサですよ」という発話が生成されたが、不正解発話と判定された。この発話は語尾の変形前は「ユニクロのいいところはカラーとサイズの多さにあると思うウサよ」であり、「ユニクロのいいところはカラーとサイズの多さにあると思います」と変形できていたら正解発話であった。語尾変形のルールはヒューリスティックに決定したものであり、この例のように網羅できていないものもあることから、ルールのさらなる拡充が必要である。

4. むすび

本研究では、非タスク指向型対話システムの発話の自動生成を目的とし、Twitter データから任意の話題に関する発話を生成する手法を提案した。提案手法は、まず Twitter を任意の話題で検索し、取得した文をルールによりフィルタリングを行う。次に、文に点数付けし、点数がしきい値以上の文を発話として獲得する。実験の結果、発話として使用可能な発話が高精度で生成可能であることが確認できた。

今後は、提案手法により生成した発話と、我々が以前提案した発話選択法 [6] を用いて対話システムを実装し、実際にユーザと対話することにより評価を行う予定である。その際、定量的な評価法 [7] を用いた自然な対話が可能かどうかの評価だけではなく、楽しい対話が可能かといった主観的な評価も行う予定である。

参考文献

- [1] J. Weizenbaum. ELIZA-a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, Vol. 9, No. 1, pp. 36–45, 1966.
- [2] Daniel Jurafsky, Elizabeth Shriberg, and Debra Biasca. Switchboard swbd-damsl shallow-discourse-function annotation coders manual, draft 13. *University of Colorado, Boulder Institute of Cognitive Science Technical Report*, pp. 97–02, 1997.
- [3] Taku Kudo. Mecab: Yet another part-of-speech and morphological analyzer. <http://mecab.sourceforge.net/>, 2005.
- [4] Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol. 2, No. 3, p. 27, 2011.
- [5] 荒木哲郎, 池原悟, 塚原信幸, 小松康則, 田川崇史, 橋本憲久. m 重マルコフ連鎖モデルを用いた日本語の誤字・脱字・誤挿入誤り文字列の検出と訂正法. 電子情報通信学会論文誌. D-II, 情報・システム, II-パターン処理, Vol. 83, No. 6, pp. 1516–1528, 2000.
- [6] 稲葉通将, 平井尚樹, 鳥海不二夫, 石井健一郎. 非タスク指向型対話エージェントのための統計的応答手法. 電子情報通信学会論文誌 D, Vol. 95, No. 6, pp. 1390–1400, 2012.
- [7] 磯村直樹, 鳥海不二夫, 石井健一郎. HMM による非タスク指向型対話システムの評価. 電子情報通信学会論文誌 D, Vol. 92, No. 4, pp. 542–551, 2009.