

Twitter から獲得した発話を用いる対話エージェントの構築

Building a Dialogue Agent using Utterances from Twitter

小新田 亮人 *¹ 稲葉 通将 *¹ 高橋 健一 *¹
 Makoto Koshinda Michimasa Inaba Kenichi Takahashi

*¹ 広島市立大学
 Hiroshima City University

This paper proposes a response method for non-task-oriented dialogue agent being able to speak broad range of topics. The method acquires sentences from Twitter. The sentences are filtered by rules and scored on the basis of training data. We acquire the high score sentences as candidate of utterances. Next, the utterances are ranked by learning to rank. The top rank utterance selects as a suitable utterance for user's utterance. Results of an experiment demonstrate that an agent based on the proposed method can conduct natural and enjoyable conversation compared to other dialogue agent.

1. はじめに

非タスク指向型対話エージェントはまだ人間と同等の対話性能を持っているとは言いがたく、改善すべき課題が多い。

非タスク指向型対話エージェントの課題の1つとして、応答する際に、文脈を考慮せず、直前のユーザの発話しか考慮できないという問題が存在する。このような問題を解決するため、我々の研究では、統計的手法を用いて発話を選択し、対話を行う統計的応答手法を提案した [稲葉 12]。しかし、この手法では、使用する発話候補は予めすべて人手で作成していた。人手による発話候補の作成はコストが非常に大きいことや、発話候補にない話題は扱えないという問題点が存在する。そこで我々は、Twitter から発話を獲得する手法を提案した [稲葉 14]。この手法では Twitter から発話を獲得するため、発話の作成のコストを削減できる。また、Twitter では広範囲にわたる話題をカバーしているため、幅広い話題に対応した発話候補集を獲得することが可能である。ただし、本手法は発話の生成のみを扱っており、どのような応答を返すかということの対象としていない。そこで本研究では、これらの手法を組み合わせることで、幅広い話題に対応可能な非タスク指向型対話エージェントを構築する。

統計的応答手法による応答を分析したところ、挨拶や質問に対する応答の際、対話が破綻するケースが多いことが判明した。そこで本研究では、この問題を解決するために予め決められたルールに基づいて応答するルールベース手法を組み合わせる。また、統計的応答手法では、対話エージェント側の発話のみを参照し、話題転換を行っており、不自然な話題転換になることも少なくない。そこで本研究では、人間が主導する話題転換を把握し、話題転換を行う話題転換を把握する手法を提案する。

2. 関連研究

多くの対話エージェントは手作業によって作成されたルール、この発話にはこう返すというペアのルールに従って応答を行う。対話エージェントは、こういったルールを大量に保持し、これらを用いて様々な応答を行う。これらのルールの作成用の言語として Artificial Intelligence Markup Language(AIML)

が知られている [Wallace 09]。

しかし、こういったルールを用いて応答を行う対話エージェントは扱える話題の範囲が限定されるため、近年では、Web上から取得した発話データを用いて応答を行う手法 [柴田 09] や Twitter から発話を獲得する研究 [Higashinaka 14]、ユーザの発話を持つ係り受け関係に関連する係り受け関係を Web から抽出し、これら関係と抽出した関係を含む Web 中の文の用例を組み合わせることで発話を生成する手法 [杉山 15] などの幅広い話題を扱えるようにする研究が行われている。

応答手法の研究では、Twitter などのメッセージに対して統計機械翻訳の手法を用いて応答を生成する手法 [Ritter 11] がある。Ritter らは、メッセージに対する応答として統計機械翻訳を用いた手法と類似度の高い文で応答する手法と比較した結果、統計機械翻訳を用いた手法のほうが優れた結果となったと報告している。その他にも、目黒らの研究 [目黒 14] では、応答手法を単独で用いるよりも複数の手法を組み合わせるほうが自然性などの向上が見られたことから、複数の手法を組み合わせることでより自然性の高い対話エージェントを構築できる可能性があるとして報告している。そこで本研究でもルールベース手法と統計的応答手法、話題転換を把握する手法を組み合わせ、対話エージェントを構築する。

3. Twitter からの発話の獲得手法

Twitter からの発話獲得手法 [稲葉 14] は、任意のキーワードを入力し、そのキーワードを含む発話を Twitter から獲得する。本手法における獲得すべき発話は「一発話だけで意味や意図が理解できる発話である」など4つの条件をすべて満たす発話を獲得すべき正解発話であると定義する。

本手法では、Twitter からキーワードを含む文を抽出し、その文をルールに基づいてフィルタリングを行う。なお、フィルタリングには「人名・代名詞を含む文の除去」など7つのルールを使用する。しかし、ルールだけは除外できない不正解発話が存在するため、文に対して教師データに基づいた点数付けを行い、点数の低い文を除外する。なお、教師データには発話の獲得条件にしたがって人手で判定を行った正解文と不正解文を用いる。

このような点数付けの結果、「iPhone の指紋認証便利だね」や「Mac はめっちゃ使いやすい」などの文を正解発話として獲得する。

連絡先: 広島市立大学大学院情報科学研究科知能工学専攻
 〒731-3194 広島県広島市安佐南区大塚東3丁目4-1
 E-mail:koshinda@cm.info.hiroshima-cu.ac.jp

4. 統計的応答手法

本章では、ランキング学習を用いた発話選択を行う統計的応答手法 [稲葉 12] について述べる。

まず、文脈を応答のために発話の選択を行う直前の発話から決めた発話数までの発話の集合と定義し、これを式 (1) で表す。

$$c = \{u_1, u_2, \dots, u_l\} \quad (1)$$

式 (1) の各要素 u_l は、文脈中に出現した発話を、 l は文脈中の発話数を表す。ただし、 u_1 は文脈 c における最後の発話、つまり、発話の選択を行う直前の発話を、 u_l は文脈 c における最初の発話を表す。また、本論文では便宜上、発話の選択を行う地点を u_0 とする。

次に、文脈 c に対する発話候補の集合を以下の式 (2) で定義する。

$$A_c = \{a_1^c, a_2^c, \dots, a_{|A_c|}^c\} \quad (2)$$

式 (2) において、 $|A_c|$ は発話候補数を、各要素 a_i^c は各発話候補を表す。この式 (2) に含まれる発話候補 a_i^c には、文脈 c に対する応答として適切な発話 (正解発話) と不適切な発話 (不正解発話) が混在する。つまり、発話選択とはこの発話候補の集合 A_c から正解発話の集合を見つけ出すことである。

本手法では、文脈 c に対する各発話候補 a^c の適切さ度合いを評価するために、 c と a^c の関係を以下の式 (3) で定義される n 次元素性ベクトルに変換する。

$$\Phi(c, a^c) = (x_1(c, a^c), x_2(c, a^c), \dots, x_n(c, a^c)) \in \mathbb{R}^n \quad (3)$$

式 (3) の $x_i(c, a^c)$ ($i = 1, 2, \dots, n$) は素性値を表す。この素性値は $\{0, 1\}$ の二値とした。本研究では、素性ベクトルにおける素性値を決定するために、11 種類の素性を用いた。

本手法では、得られた n 次元素性ベクトルの集合から適切な発話を選択するために、ランキング学習の 1 つである ListNet [Zhe 07] を用いる。ListNet は順位リストを確率分布で表現し、学習データとモデル間の分布の距離を最小化することにより、順位付けのための学習を行う。以下では、ListNet によるランキング学習の方法について説明する。

まず、ある文脈 c に発話候補の集合 A_c に対する評価 (スコア) のリストを式 (4) のように定義する。

$$Y_c = (y_1^c, y_2^c, \dots, y_{|A_c|}^c) \quad (4)$$

式 (4) のスコア y_i^c ($i = 1, 2, \dots, |A_c|$) は発話候補 a_i^c のスコアを示し、このスコアの大小により、発話候補の順位が決定する。例えば、対話の次の応答として自然であれば 3 点、やや不自然であれば 1 点、不自然であれば 0 点というようにスコアを手手で付与する。

次に、式 (3) で定義した素性ベクトル $\Phi(c, a^c)$ の集合を式 (5) のように定義する。

$$X_c = (\Phi(c, a_1), \Phi(c, a_2), \dots, \Phi(c, a_{|A_c|})) \quad (5)$$

この Y_c と X_c のペアを正しい順位付けを行う学習データとして使用する。ここで、素性ベクトル $\Phi(c, a_i)$ に対してスコア付けを行うための関数 f を用意する。この関数 f は、以下の式 (6) で定義される。なお、以下では素性ベクトル $\Phi(c, a_i)$ を Φ_i と記す。式 (6) の関数 f により得られた素性ベクトル Φ_i のスコアリストを式 (7) のように定義する。

$$f_\omega(\Phi_i) = \langle \omega, \Phi_i \rangle \quad (6)$$

$$Z_c = (f(\Phi_1), f(\Phi_2), \dots, f(\Phi_{|A_c|})) \quad (7)$$

この式 (6) における ω はパラメータベクトルを表し、 $\langle \cdot, \cdot \rangle$ は内積を表す。つまり、素性ベクトル Φ_i に対するスコア付けは素性ベクトル Φ_i とパラメータベクトル ω の内積で求めることができる。

本手法において、正しく順位付けを行うことは、式 (7) のスコアリストにおける各素性ベクトル Φ_i の順位を式 (4) のスコアリストにおける各素性ベクトル Φ_i の順位に近づけるということである。これを数式で表すと、式 (8) のように定義される。

$$\sum_{Y_c \in C} L(Y_c, Z_c) \quad (8)$$

この L は listwise 損失関数であり、この式 (8) の値を最小にすることがこの学習における目的となる。

ListNet では、この式 (8) を最小化するために、式 (4)、(7) の各スコアリストを以下の式 (9) を用いて確率分布に変換する。この式 (9) は式 (4) のスコアリストの確率分布を表した式である。

$$P_{Y_c}(\Phi_i) = \frac{b^{y_i^c}}{\sum_{j=1}^{|A_c|} b^{y_j^c}} \quad (9)$$

この式 (9) を用いて確率分布へ変換した式 (4)、(7) の確率分布をそれぞれ $P_{Y_c}(\Phi_i)$ 、 $P_{Z_c(f_\omega)}(\Phi_i)$ とすると、式 (8) の損失関数 L はクロスエントロピーを用いて以下の式 (10) で表すことができる。

$$L(Y_c, Z_c(f_\omega)) = - \sum_{i=1}^{|A_c|} P_{Y_c}(\Phi_i) \log(P_{Z_c(f_\omega)}(\Phi_i)) \quad (10)$$

パラメータ ω は最急降下法で求めることができる。

発話候補の順位付けは、関数 $f_\omega(\Phi_i)$ と最終的に得られたパラメータ ω を用いて各発話候補に対し、スコア付けすることにより行う。

5. 対話エージェントの構築

本研究で構築する対話エージェントは、図 1 で示すような流れで処理し、対話を行う。まず、対話開始時に対話エージェントは「こんにちは！」という挨拶と「よく行く飲食店はあるかな？」のような設定した 10 種類の質問の中から 1 つの質問を選び、これらを「こんにちは！よく行く飲食店はあるかな？」と組み合わせた発話で対話を開始する。次に、それに対する応答からキーワードを抽出する。キーワードには株式会社はてなが公開しているはてなキーワード一覧ファイル^{*1}にある語句を用いる。もしキーワードがあれば、それを用いて Twitter から発話を獲得する。なお、キーワードの抽出は、発話に対して Gosen [Gosen] を用いた形態素解析を行い、抽出する。そして、獲得した発話候補集から前章で述べた統計的応答手法を用いて発話の選択を行う。

以降は、エージェントの応答に対する対話中の発話から適切な応答手法を判定し、応答を決定する。ユーザの発話が設定したルールと一致すれば、ルールベース手法を用いて応答を行い、設定したルールには一致しないが対話中の発話に話題転換を示す語句などが含まれていれば、提案する話題転換を把握する手法を用いて応答を行う。設定ルールにも一致せず、話題転換を示す語句も含まれていなければ、統計的応答手法を用いて応答を行う。

本章では、応答に用いる 2 つの応答手法、ルールベース手法と話題転換を把握する手法について述べる。

*1 <http://developer.hatena.ne.jp/ja/documents/keyword/misc/catalog>

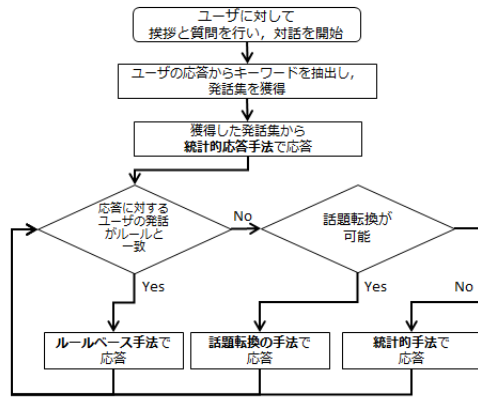


図 1: 応答手法の処理の概要

表 1: 設定したルールの例

ユーザーの発話 (正規表現)	対話エージェントの応答
あなた*名前*何*	僕の名前は KELDIC だよ
趣味	僕の趣味はみんなとお話することだよ

表 2: 設定した話題転換を示す語句の正規表現の例

正規表現
^ところで*
。ところで*
別な話
別の話

5.1 ルールベース手法

前章の統計的応答手法を用いることによって、対話エージェントはユーザーの発話から学習し、応答をすることが可能となった。しかし、統計的応答手法では、ユーザーが質問した時の次の発話選択において対話が破綻するケースが多いことが判明した。そこで本研究では、挨拶や質問などに対する応答手法としてルールベース手法を用いる。

本手法では、表 1 に示すような正規表現を用いたルールを作成した。ユーザーの発話がこのルールに一致した場合、対話エージェントは、設定したルールに基づいて応答を行う。

5.2 話題転換を把握する手法

人間同士の対話では同じ話題について話し続けていくのではなく、話題を転換させながら話を続けていく。統計的応答手法では、対話エージェントの発話のみを参照し、話題転換の把握を行っているため、ユーザーから話題転換をしようとしても話題転換がうまくいかなかったり、不自然な話題転換になることも少なくない。そこで本研究では、対話中の発話に対して形態素解析を行い、話題の中心となる単語を探し、その単語をキーワードとして新たな発話を Twitter から獲得する話題転換を把握する手法を提案する。本手法では、次に説明する「話題の転換を示す語句が使用された時」と「対話が一定数経過した時」の 2 つのパターンを用いて対話中の発話からキーワードを取得する。

(1) 話題の転換を示す語句が使用された時

ユーザーの発話に話題転換を示す語句が含まれている場合、その発話に対して形態素解析を行い、キーワードを獲得する。なお、話題転換を示す語句が含まれているか否かの判定には表 2 のように設定した正規表現を用いて判定する。

(2) 対話が一定数経過した時

ユーザーと対話エージェントの発話が合計で 8 以上となった時の直前のユーザーの発話、10 以上となった時の直前のユーザーと対話エージェントからキーワードを抽出する。なお、語句の取得タイミングは経験的に決定した。

表 3: 話題転換用のキーワードをユーザーの発話から取得する例

発話数	話者	発話
5	対話エージェント	iPhone は使いやすく最高だよ
6	人	iPhone が使いやすいというのはよく聞きます。
7	対話エージェント	iPhone はどのスマホよりもいいよ
8	人	う～ん、 Android もいいと思いますよ？
9	対話エージェント	(「Android」をキーワードとして獲得した発話候補集から発話を選択し、応答)

対話の発話数が 8 以上となった時の例を表 3 に示す。表 3 の発話数 8 の時のユーザーの発話の中に太字で示した「Android」というキーワードが含まれているため、「Android」に関する発話候補集を Twitter から獲得し、その中から発話を選択することで話題の転換を完了する。

6. 評価実験

6.1 実験設定

提案手法を性能を評価するために、提案手法を実装した対話エージェントと被験者 8 人が対話を行った。また、性能比較のために、NTT ドコモの雑談対話 API^{*2}と WOZ(Wizard of Oz)法による対話エージェントとも対話を行った。NTT ドコモの雑談対話 API は対話における話題と文脈を認識し、大規模発話データの中から発話を選択し応答する機能を WebAPI として提供している。WOZ 法は提案手法と同一の発話候補集の中から人手で発話を選択し、応答する。ただし、発話候補集の中に応答として適切な発話が存在しない場合、人手で発話を作成し、それを応答として用いる。

また、本実験では提案手法を実装した対話エージェントは対話データとそのデータに対する 10 個程度の発話候補がセットになった学習データを 1507 個用いて学習を行った。

各エージェントとの対話は、エージェントとユーザーの発話が計 20 発話になるまで対話を行い、その後、被験者にアンケートを行い、「非常に良い: 5」、「良い: 4」、「普通: 3」、「悪い: 2」、「非常に悪い: 1」の 5 段階で「対話の自然さ」と「対話の盛り上がり」について評価してもらった。

6.2 実験結果

対話後に行ったアンケートによる結果から対話の自然さ、盛り上がりに対するそれぞれの平均評価値は図 2, 3 のような結果となった。また、提案手法を実装した対話エージェントの対話は表 4 のような結果となった。

図 2 の結果で示すように、対話の自然さの平均評価値は提案手法 (2.50) は雑談対話 API(1.63) よりも高い結果となったが、WOZ 法 (4.63) には及ばない結果となった。

また、図 3 の対話の盛り上がりに関する平均評価値を示す結果でも、提案手法 (2.88) は雑談対話 API(2.13) よりも高い結果となったが、WOZ 法 (4.50) には及ばないという対話の自然さの結果と同じ傾向の結果となり、これらの結果から、提案手法は雑談対話 API より優れた対話を行うことが出来たということがわかった。

6.3 考察

評価実験では、提案手法のうち、ルールベース手法が使用された例はなかった。しかし、対話エージェントの「アンパンマンの本気が可愛すぎる」という発話に対して、ユーザーが「どういところが可愛いの?」と応答した時、うまく応答できないという問題が存在した。これは、前章の表 1 のように設定し

*2 https://dev.smt.docomo.ne.jp/?p=docs.api.page&api_docs-id=3

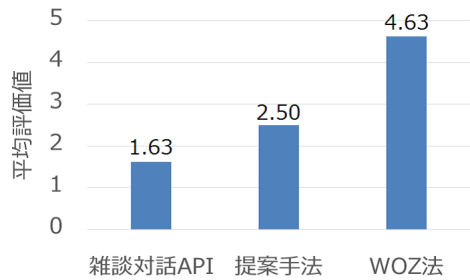


図 2: 評価実験における対話の自然さに対する平均評価値

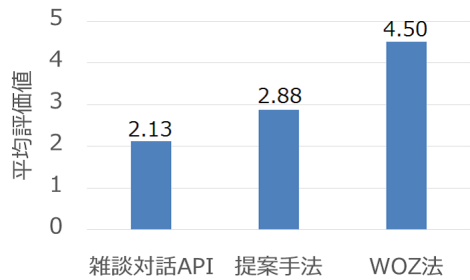


図 3: 評価実験における対話の盛り上がりに対する平均評価値

表 4: 提案手法を実装した対話エージェントとの対話例

発話数	話者	発話
⋮	⋮	⋮
3	対話エージェント	新しい iPhone はなかなか快適だよ
4	人	新しい iPhone ってことは iPhone6 のことかな? 自分にはあれは大きすぎて使いにくそう
5	対話エージェント	iPhone の小さい画面だとめんどくさい
6	人	大きすぎるほうがやだなあ。4s ぐらいのサイズがちょうどよかった
⋮	⋮	⋮

たルールに一致しなかったために起こった問題であると考えられるため、この問題に対処するには質問文のルールを増やすべきであると考えられる。

また、話題転換を把握する手法では、評価実験において表 5 のように適切でない話題転換を行っているという問題が発生していた。表 5 では、発話数 10 までは「ぶよぶよ」という話題で話していたのに、途中から別の話題に切り替わっていることがわかる。この例では、発話数 9 の発話に含まれる「難易度」がキーワードであったため、「難易度」を含む発話を獲得してしまい、「難易度」を次の話題として話題転換を行っているが、この例では「ぶよぶよ」で話を続けるのが適切であると考えられる。しかし、話題転換を把握する手法では、話題転換する前の話題は終わった話題と見なしているため、再び同じ話題を獲得することができない。したがって、この問題の解決策としては、同じ話題の獲得制限をなくすことや、キーワードの取得条件を変更するといったことが考えられる。

7. おわりに

本研究では、幅広い話題に対応可能な対話エージェントとして Twitter から発話を獲得する手法とランキング学習を用いて応答を決定する統計的応答手法を組み合わせた。また、単純に組み合わせるだけではいくつかの問題点が存在するため、それに対する対処法を提案した。提案手法の性能評価のために、手法を実装した対話エージェントと人間が対話を行う評価実験を行ったところ、提案手法は雑談対話 API よりも優れた対話を行うことが可能であることを確認した。

表 5: 変な話題転換の一例

発話数	話者	発話
⋮	⋮	⋮
7	対話エージェント	ぶよぶよは意外と話題豊富だよ
8	被験者	ゲーム性だけでなく、キャラクターも個性豊かだからね
9	対話エージェント	キーボードでぶよぶよは難易度高すぎるのだね
10	被験者	俺は割りと慣れてるけど、苦手ならコントローラーを買ってきた方がいいね
11	対話エージェント	生物はテストごとに難易度が大きく変わるのが怖い
⋮	⋮	⋮

今後は、ルールベース手法でより多くの質問に対応するために、ルールを増やしたり、話題転換を把握する手法におけるキーワードの獲得条件の変更などを行う必要があると考えられる。

参考文献

- [稲葉 12] 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, 電子情報学会論文誌 Vol.95, No.6, pp.1390-1400(2012)
- [稲葉 14] 稲葉通将, 神園彩香, 高橋健一: Twitter を用いた非タスク指向型対話システムのための発話候補文の獲得, 人工知能学会論文誌 29 巻, 1 号, pp.21-31(2014)
- [Wallace 09] R.S. Wallace: The anatomy of A.L.I.C.E., *Parsing the Turing Test*, pp.181-210(2009)
- [柴田 09] 柴田雅博, 富浦洋一, 西口友美: 雑談自由対話を実現するための WWW 上の文書からの妥当な候補分選択手法, 人工知能学会論文誌 24 巻, 6 号, pp.507-519(2009)
- [Higashinaka 14] Higashinaka, R., Kobayashi, N., Hirano, T., Miyazaki, C., Meguro, T., Makino, T., Matsuo, Y.: Syntactic Filtering and Content-based Retrieval of Twitter Sentences for the Generation of System Utterances in Dialogue Systems, *Proceedings of 5th International Workshop on Spoken Dialog Systems*, pp.113-123(2014)
- [杉山 15] 杉山弘晃, 目黒豊美, 東中竜一郎, 南泰浩: 任意の話題を持つユーザ発話に対する係り受けと用例を利用した応答文の生成, 人工知能学会論文誌 30 巻, 1 号, pp.183-194(2015)
- [Ritter 11] Alan Ritter, Colin Cherry, William B. Dolan: Data-Driven Response Generation in Social Media, in *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp.583-193(2011)
- [目黒 14] 目黒豊美, 杉山弘晃, 東中竜一郎, 南泰浩: ルールベース発話生成と統計的発話生成の融合に基づく対話システムの構築, *The 28th Annual Conference of the Japanese Society for Artificial Intelligence*(2014)
- [Gosen] lucene-gosen - Japanese analysis for Apache Lucene, <https://code.google.com/p/lucene-gosen/>
- [Zhe 07] Zhe Cao, Tao Qin, Tie-Yan Liu, Ming-Feng Tsai, Hang Li: Learning to rank: From Pairwise Approach to Listwise Approach., *Proceeding of the 24th International Conference on Machine Learning*, pp.129-136(2007)