

Twitter と word2vec を用いた時期に合った返答発話選択手法

木村 葵 目良 和也 黒澤 義明 竹澤 寿幸
 広島市立大学大学院情報科学研究科
 {my67009, mera, kurosawa, takezawa}@hiroshima-cu.ac.jp

1 はじめに

現在、道案内やチケット予約のようなタスク指向型対話システムだけではなく、雑談など対話そのものを目的とするような非タスク指向型対話システムへの注目が集まっている。しかし、ルールベース手法や統計的手法など既存の返答手法では返答が難しい発話が存在する。例えば「最近のカーブの調子はどうですか?」のように時期によって適切な返答が変わる入力発話の場合、ルールベース手法のように事前に固定の返答発話を用意しておくとき、時期を逸した返答になってしまう。また、統計的手法を用いても、用意される返答発話候補がデータベースを構築した時期に依存する、語の類似性を測る際に新語や流行語に対応できない、またその時期特有の語の共起関係を考慮できないなどの問題がある。

そこで本研究では、時期によって適切な返答が変わるような入力発話に対して、現在話題や流行を考慮した返答を行うための手法を提案する。新しい話題に対応するために、提案手法では Twitter から収集した過去一週間のツイートを対象として統計的手法を適用する。返答候補選択の際には、word2vec[1]を用いることで、辞書に登録されていない語にも対応する。word2vec を用いる利点としては、新しいコーパスを用意すれば新しい言葉にも対応できることや、概念的には遠い言葉だがよく共起する表現を拾えることが挙げられる。例えば、「スパゲッティ」と「フォーク」はそれぞれ食べ物と道具であるため概念的には遠いがよく共起する語であるため、入力発話と返答発話の類似性を測る際には有効な指標として使うことができる。

2 提案手法の処理の流れ

提案手法の処理の流れを図1に示す。

1. 任意の入力発話を受け取る

2. Twitter より収集した過去一週間分のツイート (Twitter データ) に対してフィルタリングを行い、残った発話を返答発話候補とする
3. 入力発話の話題語と意図タイプを判定する
4. (2)で収集した返答発話候補の話題語と意図タイプを判定する
5. 入力発話の意図タイプと返答発話候補の意図タイプの整合度を計算する
6. 入力発話と返答発話候補の話題語どうしの類似度を word2vec を用いて計算する
7. 整合度と類似度からスコアを計算する
8. 最もスコアの高い返答発話候補を入力発話の返答として返す

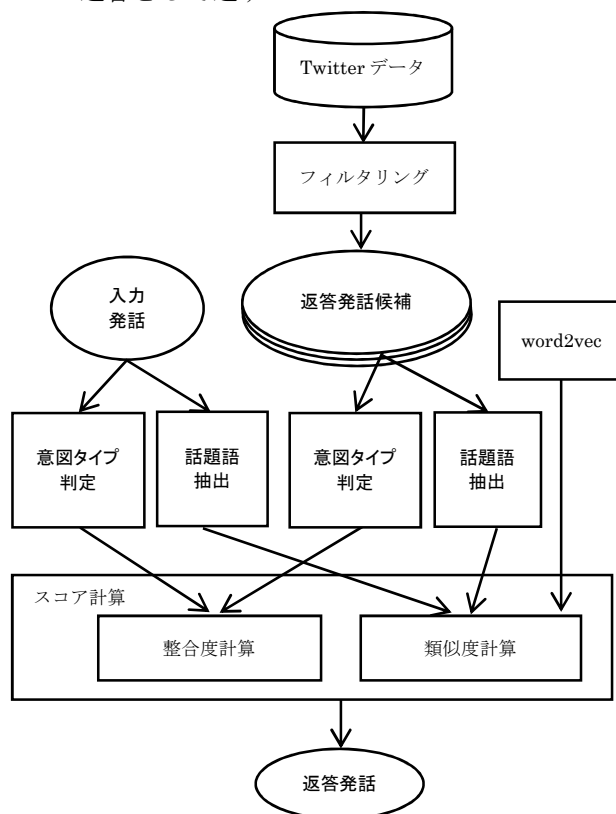


図1 提案手法の処理の流れ

3 返答発話候補の収集

まず、Twitter のパブリックタイムラインのデータを Streaming API[2]を利用することで収集する。使用する Twitter データは入力発話が入力された時

点から過去一週間分とした。ただし、収集するのは「@ (ユーザ名)」の付いたツイートだけに限り、通常をつぶやきやリツイートを除く。収集する返答発話を「@」の付いたものに限定したのは、Twitter における「@」の付いたツイートは返信 (リプライ) を意味し、入力発話に対する返答になりやすいと考えたためである。また単語数が 4 未満もしくは 25 を越える文、URL、# (ハッシュタグ)、絵文字や括弧を含むツイートも除外する。URL、ハッシュタグ、絵文字、括弧を含むツイートを除去するのは、本提案手法が音声対話システムを想定した手法であり、発音しない語を取り除くためである。

次に、誤字脱字や使用される状況が限定されるツイートを除外するため、稲葉 [3] を参考にフィルタリングルールを設ける。以下に本手法で設定したフィルタリングルールのうち、稲葉らの手法を参考にしたルール 1, 2, 3, 4 を示す。

なお、フィルタリングルールには品詞情報を使用するものがあるが、本論文では形態素解析に MeCab[4] を使用した。

ルール 1. 助詞あるいは助動詞から始まる文

このような状況は誤字や脱字が原因で起こることが多かったために除外対象とする。

ルール 2. 文末以外に終助詞が含まれている文

終助詞が文末以外に出現する状況は、句点と読点の打ち間違えや句点の打ち忘れなどのミスが原因のことが多かったため除外対象とする。

ルール 3. 接続詞から始まる文

接続詞から始まる発話は文脈を限定する文であるため、返答として汎用性が低いと考えられるため、除外する。

ルール 4. 人名、代名詞が含まれている文

人名や代名詞が含まれている文は、それだけでは意味や意図が理解できないことが多い。具体的には名詞-固有名詞-人名、名詞-代名詞を持つ文を除外する。

また、以下に本論文オリジナルのルール 5, 6 を示す。

ルール 5. 名詞を全く含んでいない文

提案手法では入力発話中の名詞と返答発話候補中の名詞の類似度を返答選択に用いるため、名詞を含まない返答発話候補は処理できない。

なお、“ん” のような非自立名詞は名詞とはカウントしない。

ルール 6. Twitter 用語が含まれている文

「フォロー」、「リブ」、「DM」などの Twitter 用語を含む文は、Twitter 上での会話という極めて限定された状況でしか使われなかった文が多く見られたため除外する。

以上のフィルタリングルール 1~6 をパスしたツイートのみを返答発話候補として用いる。

4. 返答発話候補のスコア付け

返答発話候補の妥当性を評価するスコア付けは横川 [5] を参考に「入力発話と返答発話候補の整合度」と「入力発話と返答発話候補の話題語どうしの類似度」を用いて求める (式 1)。整合度と類似度の算出方法については次節以降で述べる。

$$\text{スコア} = \frac{\text{整合度} + \text{類似度}}{2} \quad (1)$$

4.1 入力発話と返答発話候補の整合度

本研究では、入力発話の発話意図と返答発話候補の発話意図をそれぞれ求め、入力発話の発話意図に対して返答発話候補の発話意図がどれだけ整合がとれているかを整合度としてスコア付けに用いる。

本研究では、横川[5]や林[6]を参考にして、3 種類の入力発話意図タイプと 7 種類の返答意図タイプを定義した。各意図タイプは、対象発話に含まれるキーワードの有無によって判定する。入力発話意図タイプ判別のためのキーワードは林らの手法を参考に決定し、返答発話意図タイプのキーワードは手動で意図タグ分類を行った予備実験の結果を参考に決定した。入力発話意図タイプと判別のためのキーワードを表 1 に、返答発話意図タイプとキーワードを表 2 に示す。なお、複数の発話意図タイプのキーワードが同時に含まれる場合は、表中で上にある発話意図タイプを優先する。

また、入力発話意図タイプと返答発話意図タイプの組み合わせに対して、表 3 のように整合度を設定した。例えば“意見”の入力発話は、“意見”や“同意”と言った意図の返答を期待していると考えられるため、整合度を高く設定している。逆に、“経験”、“疑問”、“挨拶”といった意図の返答は整合度を低く設定している。いずれの回答タイプにも当てはまらなかったものは未知として定義し、全て整合度 0.50 と設定する。

表 1 入力発話意図タイプとキーワード

発話意図タイプ	キーワード
経験	る方, ことはありますか?, どうでしたか?
提案	オススメ
意見	どちら, あなた, どう思いますか?

表 2 返答発話意図タイプとキーワード (一部)

発話意図タイプ	キーワード
挨拶	お疲れ, すみません, ありがとう, おはようなど
経験	だった, た (助動詞)
疑問	? (クエスションマーク)
要求・提案	オススメ, お願い, 欲しい, たい (助動詞), う (助動詞), よう (助動詞), 命令形 e, 命令形 ro
意見	得意, 苦手, 上手, 下手, 思う, 泣く, 面白い, 好きなど
同意	そうだ, 賛成, 同意
拒否	反対, 違う

表 3 入力発話意図タイプと返答発話意図タイプの組み合わせによる整合度

		返答発話意図						
		意見	経験	同意	拒否	疑問	要求提案	挨拶
入力発話意図	経験	0.50	0.75	0.25	0.25	0.25	0.50	0.25
	提案	0.75	0.50	0.25	0.25	0.25	0.75	0.25
	意見	0.75	0.25	0.75	0.50	0.25	0.50	0.25

4.2 入力発話と返答発話候補の類似度

入力発話と返答発話候補の類似度の計算は、入力発話と返答発話候補に含まれる話題語どうしの類似度を計算することで求める。話題語は名詞であり、その発話を象徴付ける語である。話題語間の類似度の計算には word2vec[1]を用いる。word2vec の学習コーパスには、発話が入力された時点から一定期間過去にさかのぼってインターネットから収集したブログデータを用いる。コーパスにブログ記事を用いたのは、話し言葉で書かれていることが多いことや、比較的新しい言葉が書かれていることが多いためである。なお、word2vec でベクトルが作成されていない語との類似度は全て 0 とする。

5 評価実験

本研究で定義した 3 種類の入力発話意図に対応する入力発話に対して提案手法を適用し、出力された返答発話候補リストの上位の発話の妥当性に

ついて評価する。実験に用いた入力発話を表 4 に示す。word2vec の学習コーパスとしては、Cluweb[7]で収集した 2009 年 3 月 27 日から 2014 年 8 月 28 日までの間に収集したブログ記事を用いた。返答発話候補は、2015 年 1 月 2 日の 0 時から遡って一週間分 StreamingAPI で収集したツイートをフィルタリングしたものを用いた。フィルタリング後の返答発話候補は 4496 個である。なお本実験では、各入力発話の話題語は事前に人手で設定した。返答発話候補の話題語は最初に登場した名詞とした。

各入力発話に対して提案手法を適用することで作成した返答発話候補リスト上位 5 件を表 5, 6, 7 に示す。

表 4 評価実験に用いた入力発話

入力発話	発話意図	話題語
(1)昨日観た映画はどうでしたか?	経験	映画
(2)先週貸した漫画はどうでしたか?	経験	漫画
(3)昨日の野球の試合はどうでしたか?	経験	野球
(4)最近お勧めの映画は何ですか?	提案	映画
(5)おススメの漫画などがありますか?	提案	漫画
(6)最近お勧めの野球選手は誰ですか?	提案	野球
(7)最近の映画についてどう思いますか?	意見	映画
(8)おススメの漫画などがありますか?	意見	漫画
(9)最近の野球についてどう思いますか?	意見	野球

表 5 入力発話 1 に対する返答発話候補リスト

順位	返答発話	スコア	採点
1	映画はじまったらもう目あくわ	0.875	-1
2	映画もたくさんある。	0.75	-2
3	映画館の音響で聞く発射音たまらんす	0.75	2
4	再放送したら録画するのにww	0.629	1
5	ニュースで見た	0.626	-1

表 6 入力発話 4 に対する返答発話候補リスト

順位	返答発話	スコア	採点
1	映画はじまったらもう目あくわ	0.75	-3
2	映画もたくさんある。	0.75	0
3	映画館の音響で聞く発射音たまらんす	0.75	-1
4	映画館に招いてくれてありがとうございます!	0.625	-3
5	映画化もだけど向こう側の人たちで犠牲者が別にでいていそうだ..	0.625	-3

表 7 入力発話 7 に対する返答発話候補リスト

順位	返答発話	スコア	採点
1	映画化もだけど向こう側の人たちが犠牲者が別にでていそうだ・・・	0.875	2
2	映画もたくさんある。	0.75	2
3	映画館の音響で聞く発射音たまらんす	0.75	3
4	映画館に招いてくれてありがとうございます！	0.625	-2
5	映画はじまったらもう目あくわ	0.625	-1

作成された返答発話候補リストの妥当性を評価するため、入力発話に対する返答発話の妥当性を [-3, 3] の整数値で主観評定した。比較対象として、ランダムに返答発話を選択した結果についても同様に主観評定した。表 4 のテスト用入力発話 9 件に対して提案手法とランダム法を適用して得られた発話候補リスト上位 5 件の採点値の平均を入力発話意図タイプごとにまとめて比較したものを表 8 に示す。

表 8 各手法上位五件の妥当性平均値

入力発話意図	提案手法	ランダム
経験 (入力発話 1~3)	0.1	-2.6
提案 (入力発話 4~6)	-1.3	-2.5
意見 (入力発話 7~9)	0.6	-1.9
全体 (入力発話 1~9)	-0.2	-2.3

表 8 より、入力発話“提案”は、妥当性平均値が負値となり、不適切な返答を多く選択する結果となった。入力発話“提案”、“意見”においては、上位 5 件の平均採点値が正の値を取ったため、適切な返答を選択できたと言える。

提案手法によって選択された返答発話の誤りの原因について分析したところ、以下の理由が見られた。

- 一文では意味・意図が不明な発話
表 6 の「映画はじまったら目あくわ」のように、その発話単体では何を意味するのかが読み取れなかったため不適切な返答と判定された。この問題は名詞句の省略により発生していると考えられるため、ゼロ照応解析などにより省略されている語を補うことや、もしくはこのような返答発話候補自体をフィルタリングルールを用いて除去することが必要である。
- 具体的な内容を必要とする意図の発話
“提案”はお勧めの作品名などある程度具体的な内容を示した返答をする必要があるため、

不適切な返答と判定される例が多くあった。この問題を解決するためには、話題語を具体的な作品名に変換するなどして対処する必要がある。

6 おわりに

本論文では、時期によって適切な返答が変わるような入力発話に対して現在の話題や流行を考慮した返答を行うため、過去一週間のツイートから返答候補を収集し、新語に対応した返答候補ランキングを作成する手法を提案した。返答候補のスコアは、整合度と類似度を用いた。整合度は 3 種類の入力発話意図タイプと 7 種類の返答意図タイプの組み合わせにより求め、類似度は入力発話と返答発話候補に含まれる話題語どうしの類似度を計算することで求めた。実験の結果、“経験”の意図タイプの入力発話に対しては、上位 5 件の妥当性平均値が 0.1 という結果が得られた。また、“意見”に対しては 0.6 という結果が得られた。

今後の課題として、入力発話意図タイプ“経験”“提案”の精度向上のための新しいルールの設定や、実際にシステムを構築しての対話印象の評価を行う予定である。

参考文献

- [1] T.Mikolov et al., “Distributed representations of words and phrases and their compositionality,” In Proceedings of NIPS, 2013.
- [2] Twitter, The Streaming APIs | Twitter Developers, [https://dev.twitter.com/streaming/overview/\(2016/01/10アクセス\)](https://dev.twitter.com/streaming/overview/(2016/01/10アクセス))
- [3] 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獲得,” 人工知能学会論文誌, Vol.29, No.1, pp.21-31, 2014.
- [4] T.Kudo, Mecab: Yet another part-of-speech and morphological analyzer, [http://taku910.github.io/mecab/\(2016/01/10アクセス\)](http://taku910.github.io/mecab/(2016/01/10アクセス))
- [5] 横川祐太, 白井清昭, “多様な質問を受け付ける質問応答システムの回答選択手法”, 言語処理学会, 第 21 回年次大会 発表論文集, pp.163-p166, 2015.
- [6] 林秀治, 山本和英, 質問意図による QA サイト質問文の自動分類, 信学技報, Vol.113, No.83, NLC2013-10, pp.51-56, 2013.
- [7] Clueweb, [http://lemurproject.org/clueweb12/\(2016/01/10アクセス\)](http://lemurproject.org/clueweb12/(2016/01/10アクセス))