

大規模テキストコーパスを用いた感情を考慮した返答発話生成ルールの自動生成手法

Emotion considering response generation rule generating method using large-scale text corpus

田淵 悠真
Yuma Tabuchi

広島市立大学大学院 言語音声メディア工学研究室
Language and Speech Research Laboratory, Graduate School of Information Sciences, Hiroshima City University

概要 我々の研究室では、これまで音響的特徴から推定した感情と発話文字列の両方を考慮して返答発話を返す自然言語対話システムを提案している。しかし自然言語対話システムにおいて返答発話生成ルールを増やす際、話者の感情を考慮した条件付け作業は膨大な時間がかかるとともに作業者の主観が入ってしまう。本研究では、返答発話がどのような話者感情のときに現れるかを自動分析することで、大規模テキストコーパスから自動で返答発話生成ルールを構築する手法を提案する。

1 はじめに

現在様々な分野で自然言語対話システムが普及しつつある。さらに近年は指示された作業を実施するタスク指向型対話だけでなく、雑談のような非タスク指向型の対話の実現についても研究が進められている。しかし従来の自然言語対話システムでは“字面でしか判断できない”という問題がある。例えばシステムからの「元気ですか?」という問いかけに対して、力なく「元気だよ…」と答えても、音声認識処理では「元気だよ」という文字列としてしか情報を得られないため、システムはユーザが元気なものとして対話を進めてしまう。しかし人間同士の対話の場合、たとえ「元気だよ」と答えていても、声の調子や表情などのノンバーバル情報からも相手の状態を判断し、「強がっているけど本当は元気じゃないな」というような判断をすることができる。

この問題を解決するため、我々はユーザの感情を推定し、推定した感情も考慮して返答発話を返すシステムを提案している[1]。このシステムは音響分析によりユーザ感情を推定し発話内容と推定感情の両方を考慮した対話を行うことができる。このシステムは返答の際に言語情報と感情推定の結果からデータベースを参照し、マッチしているものを返答発話として返す仕組みとなっている。したがって、このシステムで様々な日常会話を成立させるためには返答発話生成ルールの大幅な追加が必要不可欠である。

しかしながら、追加される返答発話候補は、こういった発言に対してこの返答発話を返すのかということのみならず、こういった感情の際にこの返答発話を返すのかということまで定義しなければならない。しかしこの作業にはかなりの作業コストがかかり、さらに作業者の主観が入ってしまう可能性が高いという問題がある。

そこで本研究では、大規模テキストコーパスから収集した返答発話候補の直前の話者感情の分類を行うことで、返答発話生成ルール追加を自動化することを目指す。そのためまず、先行発話に対しての返答発話を大量に収集し、ある返答発話がどのような発話状況においてよく用いられるのかを分析する。本論文における発話状況とは、話者の

感情状態(positive, negative, neutral)を表す。そして、大規模コーパスから収集した先行発話-返答発話対に対して、返答発話から推定される先行発話時の話者の感情を条件として追加し、返答発話生成ルールを自動作成する。

2 感情を考慮した自然言語対話システム [1]

図 1 に本研究で用いる自然言語対話システムの処理の流れを示す。まずマイクでユーザの先行発話を取得する。そして取得した音声から音響分析ツールで音響的な特徴を抽出する。その特徴を事前に学習済の機械学習器に入力することによってユーザの感情を 3 種類(positive, negative, neutral)に分類する。また音声認識エンジンを使用してユーザの先行発話を文字列に変換する。この文字列とユーザ感情の推定結果の両方を使って返答発話生成ルールとマッチングを行う。最後にマッチングしたルールに基づいて返答発話文字列を作成し、返答を音声で出力する。

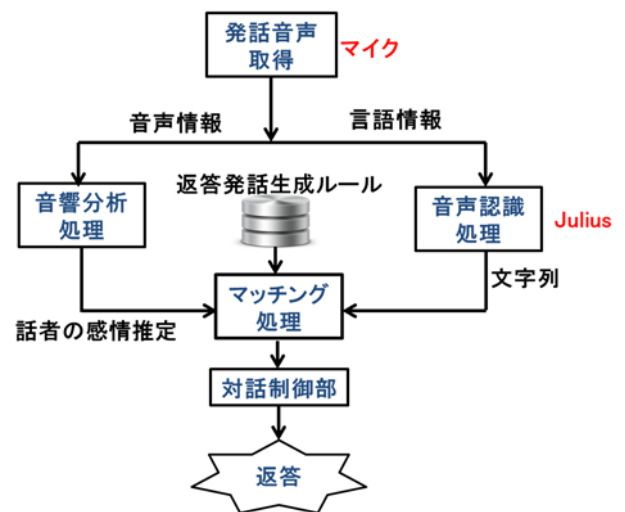


図 1 感情を考慮した自然言語対話システムの処理の流れ

本システムで用いる返答発話生成ルールは【先行発話 # システムの直前の発話 # 話者感情 # 返答】の形式で作成される。“先行発話”は直前のユーザの発話文字列を表す。“システムの直前発話”は先行発話の前にシステムが出力した発話を表す。“話者感情”は先行発話を音響分析して得られる感情の種類を表す。これにより、同じ字面のユーザ発話に対しても口調に合わせて異なるリアクションをとることができる。“返答”は条件に合致した場合にシステムが出力する返答の文字列を表す。

図 2 に返答発話生成ルールを用いた対話例を示す。【おはよう # # positive # いい天気だね】と【おはよう # #

negative # 元気そうじゃないね。どうしたの?】の2つのルールがあるため、ユーザがシステムに悲しい声で「おはよう」と言った場合は2番目の返答発話生成ルールが適用されシステムは「元気そうじゃないね。どうしたの?」と返答する。次にユーザがシステムの返答発話に対して悲しい声で「大丈夫」と答えると6番目の返答発話生成ルールにマッチングするためシステムは「そうか、何か楽しい音楽を流そうか?」と返答する。

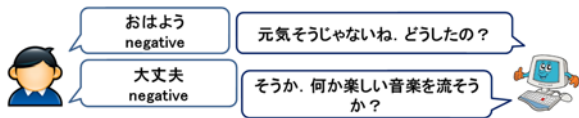


図2 返答発話生成ルールを用いた対話例

表1 返答発話生成ルール

No.	返答発話生成ルール選択条件			返答
	先行発話	システムの直前発話	感情	
1	おはよう		positive	いい天気だね
2	おはよう		negative	元気そうじゃないね。どうしたの?
3	試験に落ちたよ	元気そうじゃないね。どうしたの?	negative	今度もっと頑張ろう
4	大丈夫		neutral	何が大丈夫なの?
5	大丈夫	元気そうじゃないね。どうしたの?	positive	良い日を過ごしてください
6	大丈夫	元気そうじゃないね。どうしたの?	negative	そうか、何か新しい音楽を渡そうか?

3 返答発話生成ルールの自動生成手法

2章で説明した自然言語対話システムには同じ文字列の発話音声に対しても感情条件によってシステムが返答する発話が異なるという特徴がある。しかしこのようなルールを作成する際、発話文字列だけでは推定できない話者の感情状態の情報が必要となる。先行発話-返答発話対の収集にWeb文書のようなテキストコーパスの使用を想定した場合、このような感情状態の推定のためには、顔文字やエモーションが本文に付加されている必要がある。しかし、一般的なWebサイトに限らず、ブログやTwitterログなどを使うとしても、このような情報が付与されていることは多くない。LineやSMSなどのチャットログでは比較的多く見られるが、プライバシーの観点から、大規模にデータが公開/提供されることは期待できない。とはいえ感情を人手で推定するとしたら、作業者の主観が入ってしまうことや、大量の作業コストがかかるといった問題が生じる。さらに我々の用いるシステムでは、例え同じ入力文字列からでも話者感情によって返答を変える必要があるため、大量の返答発話生成ルールのバリエーションが必要となる。そこで本論文では大規模テキスト対話コーパスを用いた感情条件付き返答発話生成ルールの自動生成手法を提案する。

図3に提案手法の手順を示す。なお本提案手法の前提として、「ある感情状態にある相手の先行発話に対する返答としてよく用いられている返答発話は、同じ感情状態で発せられる別の先行発話に対する返答としても有効である」と仮定する。

手順1: 大規模テキスト対話コーパスから、大量の先行発話-返答発話対を収集する。そして、同じ返答発話に対する先行発話をまとめて収集する。図3の例では、返答発話 r1 に対する先行発話として、U1, U2, U3, ..., Um が収集されている。返答発話文字列を同じとみなすかどうかの判定には様々な手法がある。本研究で用いる手法の候補としては、内容語とその係り受けの構造が一致する発話、類似度やレーベンシュタイン距離などの基準、述語と否定の有無での判別などを想定している。

手順2: テキスト表現から感情推定が可能な先行発話に対してのみ、先行発話の話者感情を推定する。テキスト表現からの感情推定手法には様々なものがあるが、本研究では感情極性辞書を用いた感情語の有無による手法を想定している。図3の例では、感情推定が可能な先行発話 (U2, U3, ..., Um) に対して、それぞれ positive, negative, neutral のいずれかの感情が推定されている。また U1 の感情はテキストからは推定できなかったものとする。

手順3: 先行発話の話者感情を集計し、最も多い感情タイプを返答発話に対する先行発話の一般的な話者感情とする。図3の例では、返答発話 r1 は negative な先行発話に対してよく用いられると判断される。この時点で生成される返答発話生成ルールは【##negative#r1】である。

手順4: 感情推定が出来なかった先行発話について、返答発話生成ルールを作成する。このときの先行発話の話者感情は手順3で推定した感情とする。図3の例では、【U1##negative#r1】というルールが生成される。

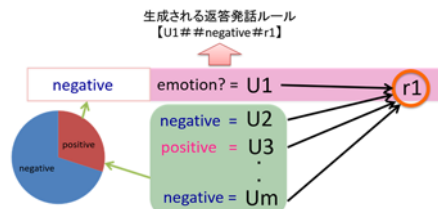


図3 返答発話生成ルールの生成手順

本手順の具体例を示す。U1が「大丈夫」、U2が「疲れたよ」、U3が「明日はデートだ」、Umが「もうだめだ」、r1が「元気出していこう」という先行発話-返答発話対が収集された場合、r1「元気出していこう」は一般的にnegativeな相手に発せられると判断される。U1「大丈夫」という先行発話はpositiveな感情でもnegativeな感情でも発せられるが、上記の傾向からnegativeな「大丈夫」への返答として【大丈夫##negative#元気出していこう】という返答発話生成ルールが作成される。

4 今後の課題

今後は良質なコーパスの選定や収集方法の検討、3章で述べた返答発話文字列が同じかどうかの判定をどう行うのかといったことへの検討が必要である。

参考文献

[1] Takumi.Takahashi, et al., "Natural Language Dialog System Considering Speaker's Emotion Calculated from Acoustic Features," Proc. of IWSDS2016, 2016.