

対話型エージェントにおけるボケる返答機能の実装

呉 健朗¹ 鈴木 奨¹ 瀧田 航平¹ 堀越 和¹ 中辻 真² 宮田 章裕^{1,a)}

概要：

現代社会において対話型エージェントは、看護やショッピングなどの生活における様々なシーンへと普及し始めており、今後多くのユーザが利用すると考えられる。しかし、対話型エージェントとの会話は無機質なものになりがちであり、このようなエージェントに親しみを抱けないユーザは少なくないと思われる。この問題を解決するために、エージェントが対話の中でユーモアを交えることで、ユーザはエージェントに対して親しみを持てるのではないかという仮説に基づき、我々は、ユーザの入力に対し、ユーモアのある聞き間違いをして聞き返すボケをするエージェントを提案する。例えばこれは、“学会について教えて？”とエージェントに対して質問をすると、“え、合体？”と、エージェントが“学会”を“合体”と聞き間違いをして、聞き返すというものである。これによって、エージェントが対話の中でユーモアを交えられるようになり、ユーザはエージェントに対して親しみを感じられると考えられる。

Implementation of a Jocosely Replying Function for a Conversational Agent

KENRO GO¹ SHO SUZUKI¹ KOHEI TAKITA¹ NAGOMU HORIKOSHI¹ MAKOTO NAKATSUJI²
AKIHIRO MIYATA^{1,a)}

1. はじめに

現代社会において対話型エージェントは、看護やショッピングなどの我々の生活における様々なシーンへと浸透し始めており、将来的にはユーザの作業の自動化や、ユーザの友人になるといった活躍が期待される。しかし、その一方でエージェントとの無機質な対話に親しみを感ずることができないユーザも少なからずいると思われる。例えば、リビングルームにコミュニケーションロボットが導入されたシーンを考えたとき、ロボットが常に堅苦しい話ばかりしていたら、家族の一員にはなりにくいだろう。あるいは、介護施設で高齢者の話し相手になるロボットが、冗談1つも言わず、生真面目な会話ばかりしていたら、ロボットが施設の人気者になることは難しいだろう。このような問題を解決するためには、エージェントとユーザの間に親

和的な関係を築くことが必要であると言える。また、人同士の親和的な関係の形成には、笑いが欠かせないとされており [1]、ユーモアと親しみの関係性が窺える。

そこで我々は、エージェントが対話の中でユーモアを交えられるようにすることで、ユーザはエージェントに対して親しみを持てるのではないかという仮説を立てた。この仮説に基づき、認知科学の分野において支持されている不適合-解決モデルを参考に、我々はユーザの発言の一部に対してユーモアのある聞き間違いをして聞き返すボケをする対話型エージェントを提案する。これは、ユーザの発言のある単語に対して意味が離れており、音が近く、かつユーザが理解できる単語をボケとして出力するものである。今回実装したプロトタイプシステムでは、“aについて教えて？”とエージェントに対して質問をすると、“え、b？”と、ユーザの発言の中の a という単語に対して意味が離れており、音が近く、かつユーザが理解できる単語である b を出力することで聞き間違いを行っている。例えば、“学会について教えて？”とエージェントに対して質問をす

¹ 日本大学 文理学部
College of Humanities and Sciences, Nihon University

² NTT レゾナント株式会社
NTT Resonant Inc.

a) miyata.akihiro@nihon-u.ac.jp

ると、“え、合体?”とエージェントが聞き返す。このように、エージェントとのユーモアのある対話を実現することによって、ユーザはエージェントに対して親しみを感じられると考えられる。

本稿の貢献は次の通りである。

- ユーザがエージェントと対話を行うシーンにおいて、エージェントがユーザの発言の一部に対してユーモアのある聞き間違いをして聞き返すボケをするシステムを提案し、実装を行なったこと。

2. ユーモア発話を行うエージェントの研究事例

ロボットや対話システムがユーザに笑いを提供する技術は大きく分けて、一方的に話すエージェントとユーザと対話を行うエージェントに分けられる。

一方的に話すエージェントの例として、[2], [3], [4], [5]が挙げられる。[2]では、エージェントがユーザに笑い感情を誘起させる手段として大喜利が用いられている。ユーザの笑い感情が誘起することがセンサデータから確認された会話文をデータセットとして、機械学習のアプローチに基づいてユーザの笑い感情を誘起するロボットインタラクションモデルの生成を行っている。[3]では、Web ニュース記事本文の感情情報を抽出して感情を考慮し、おかしみの構造図に基づいた形式的なボケによる漫才台本自動生成手法を提案している。これらはいずれも、笑いを通してユーザとエージェントのコミュニケーションをより良いものにするという試みである。[4]では、ユーザの入力した単語の意味を利用することで、日常会話に応用できる置換型駄洒落を出力するシステムを構築している。[5]は、漫才形式の対話文の自動生成システムである。入力した文章から単語を選び、それを置換することでボケを生成し、漫才形式の対話を生成している。これらは、ある単語を別の単語に置換することでエージェントによる笑いの実現を目指している。

ユーザと対話を行うエージェントの例として [6], [7], [8], [9], が挙げられる。[6]では、ユーザの対話継続欲求を向上させるために対話システムによるユーモア発話の自動生成手法を提案している。ユーモア発話の自動生成には単語間類似度を用いている。[7], [8], [9]では、ユーザの入力に対して、あらかじめ入力した内容に結び付けられたユーモアを含む応答文を返していると思われる。

3. 研究課題

今後より普及していくと予想される対話型エージェントは、未だその対話の多くが無機質なものである。そのため、無機質な対話に親しみを持っていないユーザには、このようなエージェントは受け入れられないという問題が懸念される。[1]では親和的な関係の形成に笑いが欠かせないと

されており、ユーモアと親しみの関係性が窺える。このことから我々は、エージェントが対話の中でユーモアを交えられるようにすることで、ユーザは親しみを持つことができるのではないかと仮説を立てた。

笑いを通してユーザと良好なコミュニケーションを築こうとしている事例はいくつか存在する [2][3][4][5]。しかし、このような研究はエージェントによるボケをユーザに見せることで笑いの提供を試みているため、ユーザとエージェントの間で対話などの直接的なコミュニケーションは発生していない。

[6], [7], [8], [9] は対話中でのユーモアの表現を試みているが、[6]についてはユーザに対し、エージェントが突飛な発言を行っており、ユーザがエージェントの発話をユーモアとして受容しにくい場合がある。[7], [8], [9]については、ユーモア表現を行う場合、あらかじめユーザが入力する文と、その応答が決められていると思われるため、ユーザが入力できる文の自由度が低く、すぐに飽きられてしまう可能性がある。

以上のことから我々は、ユーザがエージェントに対して親しみを感じられるようにするために、ユーザがエージェントと対話を行うシーンにおいて、下記を研究課題として設定する。

研究課題

課題 1: エージェントが対話の中でユーモアを交えられるようにする

課題 2: ユーザが受容できるようなユーモアを含む応答文を生成する

課題 3: ユーザが入力できる内容を制限しないようにする

4. 提案方式

我々は、エージェントがユーザと対話をする時、ユーモアを交えられるようにするために、学術的・実践的それぞれの観点に基づきボケの生成を行った。学術的観点では、認知科学研究者の多くがユーモア生起に不可欠と主張している「不適合の認知」に着目した [10]。中でも漫才・落語・4コマ漫画のようなユーモアは「不適合-解決モデル」で説明できる [11][12][13][14][15]。「不適合-解決モデル」とは、通常はまったく異種であり、関連がないと思われる思考、概念、状況(不適合)が、納得できる方法で結び合わされる(解決される)ことでユーモアを生起させるとする理論である。実践的観点では [16] で定義されているボケの作り方に着目した。これは、ある音から連想される言葉の中から、できるだけ意味に差がある2つの言葉を選び、それらをタイミングに合わせて使うというものである。以上より、我々は、ユーザが単語を入力すると、あらかじめ用意した出力候補単語群の中から、入力した単語(以降、入力単語)に対し、意味が離れており、音が近く、ユーザが理

解できる単語を聞き間違いとして聞き返すボケをするエージェントを提案する。

これは、ユーザの入力単語とエージェントが出力する単語の意味を離すことで不適合を作り、その不適合が音が近いという関連性をもたせることによって解決することで、ユーモアを生み出すというものである。また、エージェントとの対話を、ユーザの発言に対してエージェントが聞き間違いとして聞き返すというシーンに設定することで、タイミングを合わせる必要をなくした。さらに、シーンを設定することで自然な文脈でボケることができるようにし、ユーザが理解できる単語を出力することでエージェントが突飛なボケをすることを防いだ。

この方式により、3章で述べた3つの研究課題が達成できると考えられる。課題1については、ユーザとの対話の中で、不適合-解決モデル、[16]を参考にボケ生成を行う。課題2については、あらかじめエージェントと対話するシーンを設定し、自然な文脈でボケる。さらに、出力候補単語を一般的によく使われる単語にすることで突飛なボケが生成されるのを防ぐ。課題3については、ユーザが入力を行った後にボケを生成するため、あらかじめ入力に対する返答を結びつける必要がない。そのため、ユーザは自由に入力を行える。

5. 実装

本章では、ユーザの入力単語に対して、どれくらい意味が離れているかを表す s_s (Semantic Score), どれくらい音が近いかを表す s_e (Edit distance Score), どれくらいユーザから理解されやすいかを表す s_f (Frequency Score) のそれぞれの算出方法と、その3つの Score をもとにボケを生成するまでの手順について述べる。

5.1 事前準備

Wikipedia 日本語記事全文を MeCab[17] を用いて形態素解析し、不要品詞を除去して分かち書きしたものをコーパスとした (960649 単語)。ここでの不要品詞とは IPA 品詞体系において、記号、助詞、助動詞、接続詞、副詞、連体詞、非自立、代名詞、接尾、数、サ変・スルと分類されるものを指す。

また、コーパスを元に読み方リストと言語モデルの作成を行った。読み方リストとは、コーパス内での出現回数が1000回以上かつ日本語である単語の標準形、MeCabで取得した標準形の読み方(カタカナ)、標準形の読み方をローマ字表記に変換したものから子音を削除して母音のみにしたもののリストである(図1)本研究における言語モデルとは、単語間の概念距離を計算できるように自然言語の単語をベクトル表現したものであり、word2vec[18]を用いてコーパスから獲得した。

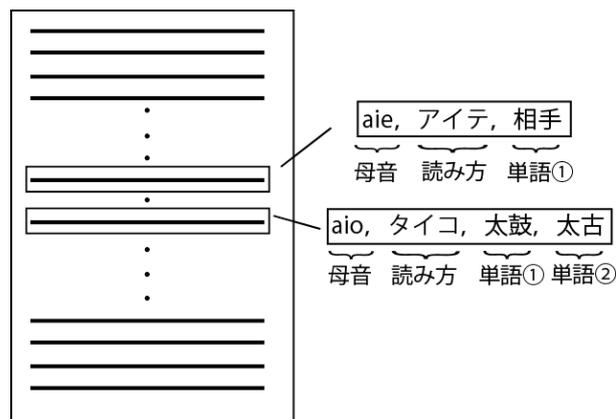


図 1: 読み方リスト

表 1: s_s が高い例・低い例

元の単語	s_s が高い単語	s_s が低い単語
こんにちは	硬化	こんばんは
食事	小惑星	食費

5.2 各 Score の計算

本節ではユーザがシステムに単語を入力した際の、出力候補単語それぞれの各 Score (s_s , s_e , s_f) の計算方法について述べる。

5.2.1 s_s : Semantic Score

ユーザの入力単語と出力候補単語がどれくらい意味が離れているかを表す指標として、言語モデルを用いて概念距離を測ることで s_s を算出した。入力単語と出力候補単語それぞれとの概念距離を測り、その距離が遠いほど高い s_s を与える(表1)。すなわち、 s_s は下記ようになる。

$$s_s = d_s \quad (1)$$

d_s は入力単語と出力候補単語の概念距離である。

5.2.2 s_e : Edit distance Score

ユーザの入力単語と出力候補単語の音がどれくらい近いかを表す指標として、編集距離を測ることで s_e を算出した。編集距離とは2つの文字列がどの程度異なっているかを示す距離であり、1文字の挿入・削除・置換によって一方の文字列をもう一方の文字列に変形するのに必要な手順の最小回数として定義される。入力単語と出力候補単語それぞれとの編集距離を測り、その距離が近いほど高い s_e を与える(表2)。ただし、今回編集距離を測る際には、より音を近づけるため、発声において重要な要素である母音のみを参照した。すなわち、 s_e は下記ようになる。

$$s_e = \frac{1}{1 + d_e} \quad (2)$$

d_e は入力単語と出力候補単語の編集距離である。

また、試行錯誤の過程で、入力単語との編集距離が同じである単語の中でも聞き間違いに聞こえる単語とそうでない単語があることが分かった。例えば、“戦闘機”という入

表 2: s_e が高い例・低い例

元の単語	s_e が高い単語	s_e が低い単語
情報	乗法	総合警備保障
花火	花見	安息日

表 3: s_f が高い例・低い例

s_f が高い単語	s_f が低い単語
日本 (862928 回)	詠嘆 (58 回)
放送 (458366 回)	毫碌 (53 回)

力単語に対し，“倦怠期”，“関東”という単語の各母音のみを考慮した編集距離はともに 2 であるが，前者は聞き間違いに聞こえ，後者はそうでないと感じられる人が多かった。我々は，多くの試行を行った結果，入力単語と出力単語の頭文字の母音と末尾の文字が一致する場合は，聞き間違いに聞こえやすいことに気付いた。これに基づき， s_e が高い場合でも，入力単語と先頭の文字の母音・末尾の文字が異なる単語は出力候補単語から除いた。

5.2.3 s_f : Frequency Score

出力候補単語がどれくらいユーザから理解されやすいかを表す指標として，Wikipedia コーパス内での各単語の出現回数を数えることで s_f を算出した。各出力候補単語の出現回数が多いほど高い s_f を与える (表 3)。すなわち s_f は下記のようになる。

$$s_f = \log f \quad (3)$$

f は出力候補単語のコーパス内での出現回数である。ただし，このとき単語の出現回数はべき分布に従うため，ごく一部の単語の出現回数が極端に大きい。これらの単語が最終的な総合 Score に与える影響が大きくなりすぎないように，出現回数の対数をとったものを s_f とした。また，同様の理由から s_f の最大値を 4.82 (出現回数換算で 660610 回) とした。この値は，コーパス内からランダムに選んだ 1000 単語を対象に出現回数の調査を行い，他の単語よりも圧倒的に出現回数の多かった単語を元に決定した。

一方，極端に出現回数が少ない単語に関しても，ユーザに理解されない可能性が高いと考え，出現回数が 1000 回を下回る単語は出力候補単語から除いた。

5.3 出力単語の決定

人名や地名などは，ごく一部の有名なもの以外はユーザに理解されない可能性が高いと考え，MeCab で人名・地域と判定される単語は出力候補単語から除いた。さらに，同様の理由から日本語以外も出力候補単語から除いた。また，エージェントの返答がオウム返しになってしまうと (表 4) ユーモアとして受容されない可能性が考えられる。そのため，入力単語を単にひらがな，あるいはカタカナ表記にしたものは出力候補単語から除いた。以上より，ユーザか

表 4: 返答がオウム返しになっている例

入力単語	出力単語
人参	にんじん
りんご	リンゴ

らの入力単語に対する出力候補単語となる条件は下記の通りである。

- 頭文字の母音と末尾の文字が入力単語と一致する。
- コーパス内での出現回数が 1000 回以上である。
- MeCab で人名・地域と判定されない。
- 日本語である。
- 入力単語を，ひらがな・カタカナ表記にしたものではない。

ユーザの入力単語に対し，上記の条件を満たすコーパス内の出力候補単語の 3 つの $\text{Score}(s_s, s_e, s_f)$ を算出し，それぞれ正規化処理，重み付けを行ってから合算したものを，出力候補単語それぞれの s_t (Total Score) とする。すなわち s_t は下記のようになる。

$$s_t = w_s s_s + w_e s_e + w_f s_f \quad (4)$$

w_s, w_e, w_f は重み係数である (本稿では重み係数は全て 1.0 とした)。

全ての出力候補単語の s_t を算出し， s_t が最も大きい値となる単語を出力単語とする。また，入力単語に対し，条件を満たす出力候補単語がない場合や，入力単語について，標準形が MeCab で判別できない，言語モデルに登録されていないといった場合には，ボケを正しく生成できない。そのため，ボケを正しく生成できた場合と明確に区別するために，ボケを正しく生成できない場合には，“What?” と出力するようにした。ここで英語を用いたのも，出力候補単語から英語を除いているため，正しく生成できた場合と区別が付きやすくなるためである。

6. 実験

提案手法の有効性を確認するために，我々は，提案方式が生成したボケに対してユーザはユーモアを感じるか，また，それによってユーザはエージェントに対して親しみやすくなるのかについて，現在検証実験を行っている。

7. おわりに

本稿は，普及した対話型エージェントとの無機質な対話に親しみを持たず，ユーザがエージェントを受け入れられないという問題の解消を狙ったものである。この問題を解決するために，まず，ユーモアと親しみの関係性に着目し，エージェントにユーモアのある対話をさせることで，ユーザは親しみを持つことができるのではないかという仮説を立てた。そして，この仮説に基づき，ユーモアを作成する手法として，認知科学領域で支持されている不適合-解決

モデルや [16] のボケの作り方を参考に、ユーザが単語を入力すると、その入力単語に対し、意味が離れており、音が近く、ユーザの理解できる単語を、聞き間違いとして聞き返すボケをするエージェントを提案し、そのプロトタイプシステムを構築した。今後は提案システムの改善に向け、まずは、生成されるボケの面白さの向上を図りたい。この問題は最終的な Score 算出時の重み係数の値変更によって改善が期待される。また、単語のみならず、文の入力に対してもボケて返すことができるエージェントの考案にも着手したい。文の入力に対するボケの作り方として、例えば、入力文の中から聞き間違えると一番面白い単語を特定し、その単語を聞き間違えて返事をするといった方法が考えられる。さらに、今回は日本語に限定した実装・実験しか行っていないが、他国語への適応も行いたい。現在、ごく一部の処理（母音・子音処理等）のみ日本語を前提としているが、基本的には言語非依存のアルゴリズムであるため、比較的少ない手間で他国語対応は実現可能であると考えている。

本研究の期待される活用法として、ユーザとの良好なコミュニケーションがパフォーマンスや継続利用率の向上につながる場面での活用が期待される。例えば、NTT レゾナント社の教えて goo[19] にはユーザからの質問を AI が自動で応答する機能がある。このような質問掲示板で、AI がボケた返答をしてユーザに親しみを抱かせるといった活用法が考えられる。

参考文献

- [1] 井上宏: 「笑い学」研究について, 笑い学研究, No. 9, pp. 3-15 (2002).
- [2] 伊勢崎隆司, 小林明美, 望月崇由, 山田智広: 笑い感情を誘起するロボットインタラクションの検討, 情報処理学会研究報告グループウェアとネットワークサービス (GN), Vol. 2017-GN-100, No. 7, pp. 1-5 (2017).
- [3] 真下遼, 梅谷智弘, 北村達也, 灘本明代: 文の感情を考慮した漫才ロボット台本自動生成手法の提案, DEIM Forum 2015 F4-4 (2015).
- [4] 中谷仁, 岡夏樹: ロボットの日常会話におけるユーモア生成の試み, 人工知能学会 2009 年全国大会論文集, 1J1-Os2-5 (2009)
- [5] 吉田裕介, 萩原将文: 漫才形式の対話文自動生成システム, 日本感性工学会論文誌, Vol. 11, No. 2, pp. 265-272 (2012).
- [6] 藤倉将平, 小川義人, 菊池英明: ユーモア発話の自動生成における単語間類似度導入によるユーモア受容性の向上, HAI シンポジウム 2014 (2014).
- [7] Apple Inc.: Siri, <https://www.apple.com/ios/siri/> (Last visited on 2017/12/20).
- [8] SoftBank Robotics Corp.: Pepper, <https://www.softbank.jp/robot/> (Last visited on 2017/12/20).
- [9] Google Inc.: Google Assistant, <https://assistant.google.com/> (Last visited on 2017/12/20).
- [10] Martin, R. A.: The Psychology of Humor, Elsevier Academic Press (2007).
- [11] Coulson, S., & Williams, R. F.: Hemispheric Asymmetries and Joke Comprehension, *Neuropsychologia*, Vol.43, Issue 1, pp.128-141 (2005).
- [12] Samson, A. C., Hempelmann, C. F., Huber, O., & Zysset, S.: Neural Substrates of Incongruity-Resolution and Nonsense Humor, *Neuropsychologia*, Vol.47, Issue 4, pp.1023-1033 (2009).
- [13] Shultz, T. R.: The Role of Incongruity and Resolution in Children's Appreciation of Cartoon Humor, *Jnl. Experimental Child Psychology*, Vol.13, Issue 3, pp.456-477 (1972).
- [14] Suls, J. M.: Cognitive Processes in Humor Appreciation, In *Handbook of Humor Research*, Vol.1: Basic issues, pp.39-57 (1983).
- [15] 伊藤大幸: ユーモアの生起過程における論理的不適合及び構造的不適合の役割, *認知科学*, Vol. 17, No. 2, pp. 297-312 (2010).
- [16] 織田正吉, 野村雅昭: シャレ・ダジャレ学事始 (ことはじめ)(第 19 回研究会), *笑い学研究*, No. 6, pp. 55-67 (1999).
- [17] MeCab: Yet Another Part-of-Speech and Morphological Analyzer, <http://taku910.github.io/mecab/> (Last visited on 2017/12/20).
- [18] Tomas Mikolov, Kai Chen, Greg Corrad, Jeff Frey, Dean: Efficient Estimation of Word Representations in Vector Space, In *Proceedings of Workshop at ICLR* (2013).
- [19] 教えて goo, <https://oshiete.goo.ne.jp> (Last visited on 2017/12/20).