

Word2Vec を用いた対立語に基づく漫才台本の自動生成

青木 哲[†] 梅谷 智弘[†] 北村 達也[†] 灘本 明代[†]

[†] 甲南大学知能情報学部 〒658-8501 兵庫県神戸市東灘区岡本 8-9-1

E-mail: †s1371001@s.konan-u.ac.jp, ††{umetani,t-kitamu,nadamoto}@konan-u.ac.jp

あらまし 近年、ロボット工学の著しい発展により、ロボットが人の生活に深く浸透してきている。しかし、人とロボットとのコミュニケーションは未だ円滑とは言いがたい。これまで我々は、人とロボットの円滑なコミュニケーションの実現を目標として、Web ニュースから漫才台本を自動生成しそれを演じる漫才ロボットを提案してきた。ここで自動生成する漫才台本は「つかみ」「本ネタ」「オチ」の三段構成からなっている。これまでの提案システムでは、「つかみ」は季節の挨拶のみを行っていたため、「つかみ」からニュースの内容に基づく「本ネタ」の間には大きなギャップがあるという問題点があった。そこで本研究ではこれら問題点を解決すべく、ニュースのカテゴリとそのカテゴリとの対立語を用いた新たな「つかみ」の台本生成を提案する。さらに、「本ネタ」で用いている対立ボケの手法も新たに提案する。

キーワード 漫才台本, 自動生成, Word2Vec, ロボット

1. はじめに

近年、動画サイトやマイクロブログなど、インターネットを利用した様々なコンテンツが増加し、メディアの多様化が進行している。それと対照的に、テレビやラジオなど、従来から存在するコンテンツは衰退の傾向がある。NHK 放送文化研究所が 2015 年に行った調査^(注1)によると、2005 年の調査から比較して、テレビは速報性などの評価が減少している一方、インターネットの速報性などの評価は増加している。その為、「電源を入れていけば受動的に多様な情報を得ることができる」といったメディアからの情報取得の利点が生かされにくくなっている。

一方近年、Pepper [1] やロボホン [2] 等ロボットが人の生活に急速に浸透してきている。しかしながら、人とロボットとのコミュニケーションは未だ円滑とは言いがたい。そこで神田ら [3] は、ロボット同士の対話観察を人が行うことで、ロボットに対して他の人間に対するように自然なコミュニケーションが可能になることを心理学的実験を通して実証している。

そこで我々は、テレビなどが持つ受動的に情報が得られるという利点と、ロボットと人の円滑なコミュニケーションの補完として、娯楽性の高い漫才に着目し、インターネット上のニュース記事から漫才の台本を自動生成し、それを演じる漫才ロボットの開発をしてきた [4]。

我々の提案する自動生成による漫才台本は、「つかみ」、「本ネタ」、「オチ」の三段構成からなる。つかみは挨拶を兼ねた最初の笑いとお本ネタへの話題提供を行っている。そして、本ネタは題材となったニュース記事の内容を読み上げてユーザに説明しながら、同時に様々なボケとツッコミを挿んで笑いをとっている。オチはまとめとして台本生成の題材となったニュース記事

の内容を 1 つのキーワードで簡潔に表現し、最後にそのキーワードをお題に自動生成した謎かけで笑いをとり締める。しかしながら、これまで提案してきた漫才台本のつかみは漫才の本文と関係のない季節の挨拶のみを行っているため、ニュース本文に基づく対話を生成している本ネタとの間で話が急激に変化し、本ネタの主題が明確になりにくいという問題が発生している。そこで本論文では、つかみの特徴である漫才の導入部であり最初のひと笑いである点に着目し、そのニュースが属するカテゴリを用いたオチによるつかみの生成手法を提案する。

本論文で提案するニュースのカテゴリを用いたボケには、安部 [5] の提案するおかしみの構図を参考にする。具体的には、そのニュースが属するカテゴリ（以下、主題カテゴリと呼ぶ）に対して、そのカテゴリに対する対立語を決定し、主題カテゴリと対立語をはき違える事によるボケの生成を行う。ここでいう主題カテゴリとは、漫才台本にするニュース記事のサイトにおいて、そのニュース記事が属する最小のカテゴリを示す。例えば、「錦織 フェデラーやマレーら『死のブロック』全豪 OP 組み合わせ」というニュースの場合はカテゴリは「テニス」となる。また、対立語の決定には、Wikipedia や Word2Vec を用いる。ここで生成したボケを「取り違えボケ」と呼ぶ。取り違えボケをつかみで行うことにより、ニュースの主題語を印象付け、つかみと本ネタの間で発生していた急激な話題変化が緩和され、ニュースの主題を明確にすることが期待できる。

また、これまで我々は、本ネタのボケの一種に対立語抽出による対立ボケ生成 [6] を行ってきた。ここでは対立語抽出時に Wikipedia の記事を基にしていたため、抽出した対立語の実際の使われ方を考慮しておらず、直感的に対立関係と認識できないものを対立語として提示するという問題点があった。実際、対立ボケの有用性を計る実験の結果では、「適している」「どちらかと言えば適している」という回答が全体の 32%であるのに対し、「適していない」「どちらかと言えば適していない」という回答が全体の 50%と、1.5 倍以上の数存在し、適切な対立

(注1) : https://www.nhk.or.jp/bunken/research/yoron/pdf/20150707_1.pdf

語が抽出されていないことが判明した。そこで、本論文ではこの問題点を解決するために、各単語の実際の使われ方を考慮し、それを基に対立語を抽出し、対立ボケの改善を行う。具体的には、対立語の抽出に Word2Vec [7] を使い、コーパスとして Twitter に投稿されたツイートを無作為に取得し、使用する。これにより、これまで考慮されていなかった実際の使われ方が考慮され、より直感に近い結果となる。

以下、第 2 章では本研究の関連研究に対して述べ、第 3 章では漫才台本自動生成器について、第 4 章は取り違えボケについて、第 5 章では実験とその考察について、第 6 章ではまとめと今後の課題について述べる。

2. 関連研究

本論文で提案する漫才台本の目的として、ニュース記事の理解支援がある。

ニュースの理解力向上に着目した研究として、北山ら [8] は、ユーザが閲覧しているニュースに関連した比較ニュースをアーカイブからの確に検索する検索方式を提案している。また、田中ら [9] は、ニュース記事中のエンティティを抽出し、それを用いて Web 上のリソースからエンティティ間の関係を示す情報を取得する手法を提案している。本研究は、取得したニュースそのものに娯楽性を付与して適用し、新たなコンテンツとしてユーザに提供する点で異なる。

Word2Vec を用いた研究は近年盛んに行われている。

野沢ら [10] は、レシピ投稿サイトの投稿を基にコーパスを作成し、それを用いて、例えば「ベーコン」の代わりに「ウインナー」のようにある食材の代替となる食材の抽出を行っている。本論文とは、似た意味を持つものを抽出するという点で共通しているが本論文では対象が食材に限定されていない。また、土田ら [11] は、Twitter に投稿されたテキストを基にコーパスを作成し、それを用いて、例えば「東京」における「東京タワー」は、「京都」における「京都タワー」のようにある地域のランドマークを別の地域のランドマークと対応付ける意味演算を行っている。

漫才に関する研究も近年行われている。

吉田ら [12] は、入力した文章から単語を選び取り、それを用いてボケを生成し、対話形式の文章を生成している。しかし、実際の漫才でいうところの「本ネタ」のみの生成である上、基となる文章を用意する必要がある点で本研究とは異なる。関 [14] は、漫才の分析により漫才の対話中に含まれる対話のズレと笑いの関係性を明らかにしている。林ら [13] は、ロボットによる漫才を新たな社会的受動メディアと位置付け、2 体のヒューマノイドロボットを用いた漫才を行っている。人間の漫才とロボット漫才との比較実験を行い、その結果エンターテインメントとしてのロボット漫才の有用性を示している。本論文では、漫才の内容である台本に着目し、対話を自動生成している点で異なる。

3. 漫才台本自動生成

漫才台本自動生成の手順を 図 1 に示す。我々が提案してき

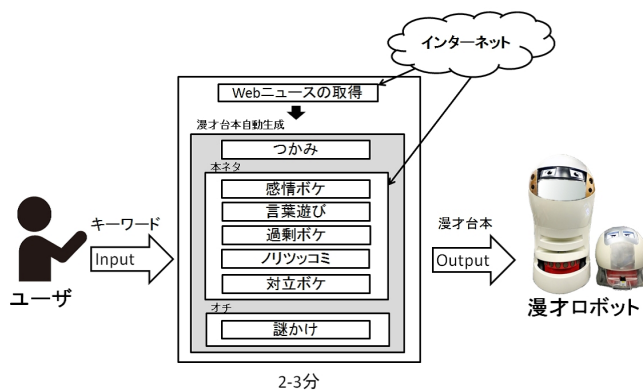


図 1 漫才台本自動生成の手順

た漫才台本生成システムは、ユーザがキーワードを入力すると、そのキーワードに関連したニュース記事を取得し、それを基に漫才を生成する。生成される漫才は、ボケ役とツッコミ役の二体のロボットが対話形式のしゃべくり漫才を行う形式になっている。

3.1 漫才台本の構成

生成される漫才の構成を表 1 に示す。大まかに、最初の挨拶の「つかみ」、本文となる「本ネタ」、最後のまとめの「オチ」の三段構成となっている。

3.1.1 つかみ

つかみは、挨拶と漫才の導入部としての役割がある。始めに季節の挨拶とその際に行われる表情ボケで最初の一笑を取る。表情ボケとは、ロボットが話す内容に応じて感情を目の表情で表現するが、その表情が話す内容には明らかに不適切であるというボケである。本論文では、この「つかみ」を改善することで本ネタで用いられるニュース記事の主題を明確にすると同時に、一笑を取る。

3.1.2 本ネタ

本ネタは漫才の軸となる部分である。取得されたニュース記事を説明しながら、様々な掛け合いを行い、笑いを取っていく。漫才全体で占める割合が最も大きく、複数の掛け合いで構成される。ボケの挿入は、記事の構造をユーザが把握できるように、句読点による文単位で行い、1 文に付き最大で 1 ボケを挿入する。ただし、記事の最初の 1 文は記事の概要が顕著に表れるためボケの挿入は行わない。また、ボケが作成されない 1 文であっても、ツッコミは相槌を行うことで対話に参加する。「一つの漫才の中には、いろんな種類のボケが入っていた方がいい」[15] とのことから、本漫才にもできる限りボケの種類を増やすべきであり、その目的で複数の掛け合いが提案されている。掛け合いのパターンとしては「言葉遊びボケ」「ノリツッコミ」「過剰ボケ」「対立ボケ」がある。この内、対立ボケにおける対立語の抽出手法は本論文で新たな手法を提案する。

言葉遊びボケ

言葉遊びボケとは、ボケ役がニュース記事中の単語を別の単語と読み間違えるというボケである。例えば、「投票(とうひょう)」という単語を「凍傷(とうしょう)」という単語に読み間違えるというものである。ツッコミは、この読み間違えた語に対

表 1 漫才台本の構成

構成名称	役割	自動生成されるボケとツッコミ
つかみ	漫才のはじめの部分で、挨拶を兼ねた最初のひと笑いと、本ネタへの話題の提供を行う。	●表情ボケ ●取り違いボケ (提案手法)
本ネタ	漫才の本文となる部分で、取得したニュース記事を基に作成を行う。ニュースの内容を解説しながら様々なボケとツッコミで笑いを取る。	●言葉遊びボケ ●ノリツッコミ ●過剰ボケ ●対立ボケ
オチ	漫才の最後の部分で、記事のまとめを行い、最後に一笑を取る。	●謎掛け

ボケ	福原は彼ともたくさん話をし、どんな”電卓”でも応援してくれると言ってくれた。
ツッコミ	そうそう、電卓はホント高いなー
ツッコミ	…って、
ツッコミ	なんでやねん！電卓って！それは計算機の種類で電子(式)卓上計算機の略やろ！電卓ちゃうくて選択や！
ボケ	おっと、うっかりしてもうた
ツッコミ	大丈夫かいな…

図 2 ノリツッコミの生成例

しツッコミ、正しい語に訂正をすることで掛け合いが成立する。また、ツッコミとなる正しい単語への訂正には、その単語をタイトルとする Wikipedia の記事を利用することで、間違えた単語の説明を補足させる。補足情報として Wikipedia 記事の最初の 1 文がそのタイトルの概要を顕著に表していることから、その最初の 1 文を用いる。例えば、「凍傷」の説明には、「低温が原因で生じる皮膚や皮下組織の傷害」といった一文が抽出される。ノリツッコミ

ノリツッコミとは、通常のツッコミとは異なり、ツッコミ役が一度ボケに同調し(のっかり)、話題を展開した後に間違いを訂正した上で改めてツッコミを入れるという掛け合いである。ツッコミ役が一時的にボケ役に転じることからツッコミでありながら同時にボケにもなるという特徴がある。本研究では、言葉遊びボケに対して適応を行い、同調の部分にそのボケた単語の印象を用いている。例えば、ボケが「選択」を「電卓」とボケる。それに対し単語の印象を用いてツッコミが「電卓は高い」と同調する。その後すぐにツッコミが訂正を入れる。これによって生成された例を図 2 に示す。

過剰ボケ

過剰ボケとは、ボケ役がニュース記事中の数値の値を異常に多い値と読み間違えるというボケである。例えば「たこ焼き 600 円」という文を「たこ焼き 600 万円」のように読み間違えるというものである。

対立ボケ

対立ボケとは、ツッコミ役が一旦ニュースの読み上げを中断させ、ボケ役に対しニュース中のある単語について知っているか問う。その際、ボケ役は別のもとの勘違いし、勘違いをしたものに関する説明をするというものである。例えば、まずツッコミ役が「サンパウロ」がどのようなものかとボケ役に問い、ボケ役はリオデジャネイロと勘違いし「カーニバルで有名」と答える。しかし、これは正しくないのでツッコミ役がツッコミを

ボケ	ブラジルで 5 日に開幕するリオデジャネイロ五輪で、世界最大規模を誇るサンパウロの日系社会が日本人選手への食材確保に一役買っている。
ツッコミ	なるほど
ツッコミ	…ところでサンパウロってどんなかわかっているか？
ボケ	あれやろ？カーニバルで有名なのやろ？
ツッコミ	違うわ！それはリオデジャネイロやろ。サンパウロは、グアルーリョス国際空港とかで有名なやつや！
ボケ	そうなんか、どっちも似たようなもんやろ
ツッコミ	完全に逆やろ…怒られるぞ！

図 3 対立ボケの生成例

入れる。これによって生成された例を図 3 に示す。

3.1.3 オチ

オチは、漫才を締めくくる部分である。本ネタの内容であるニュース記事の主題語を用いた謎かけを自動生成している。謎掛けとは「A とかけて B と解く。その心は、どちらも C(C') がつきものです」と言った形式で行われる一種の言葉遊びで、例えば「大谷選手とかけて野党と解く。その心は、どちらも投手(党首)がつきものです」というものである。

4. 取り違いボケの生成

4.1 取り違いボケの概要

我々の提案する取り違いボケは、ボケが主題カテゴリを対立語と取り違えた勘違いをし、対立語に関することを言い出しそれにツッコミがツッコミを入れるというものである。

取り違いボケの例を図 4 に示す。ここでは、主題カテゴリが「テニス」、対立語が「サッカー」である。まずはじめにボケが主題カテゴリについて述べる。次にツッコミがそれに対して誰が有名かと質問をする。それに対しボケは取り違えた対立語の関連人物について述べる。ここではサッカーに関連する人物「本田圭佑」が有名だと答える。そして、ツッコミが関連する人物が主題カテゴリの関連する人物でなく、対立語の関連する人物であると指摘し、それをニュースの主題語を用いて訂正する。本論文ではこの関連する語を関連語と呼ぶ。このように、主題カテゴリを取り違えて対立語の関連語を取り上げることにより、よりユーザに主題カテゴリを印象づけることが可能になると考える。

4.2 対立語の生成

取り違いボケの生成手順を図 5 に示す。取り違いボケの生成方法は、はじめに、ニュース記事のカテゴリを取得する。これ

表 2 ツイートの例

とりあえず 15 個注文。スバルくんお願い出て (; ㄥ ;)
ナルト読み終わった一面白い！
やぶースポーツのトップに須田が写真つきで出てたわ…
コンビニ出るとき、外出るおじいちゃん抜かしちゃったからドア持って待ってたら「お〜ありがとうありがとうどうもご親切に綺麗な人は心も綺麗だねえ」って言われて、その後その辺にいた綺麗なおじさんにナンパされて私忙しかった
進撃の新刊のハンジさん好きだー。

ボケ	最近テニスについて勉強しとんねん
ツッコミ	ほお〜、どんな人が有名？
ボケ	せやな、例えば 本田圭佑とかかな
ツッコミ	なんでやねん、それはテニスやなくてサッカーや
ツッコミ	テニスなら錦織圭とかが有名や
ボケ	そんなんどっちでもええやろ
ツッコミ	怒られるで
ツッコミ	ところで、テニスと言えば、地球でこんな話題あったの知っているか？

図 4 取り違えボケの例

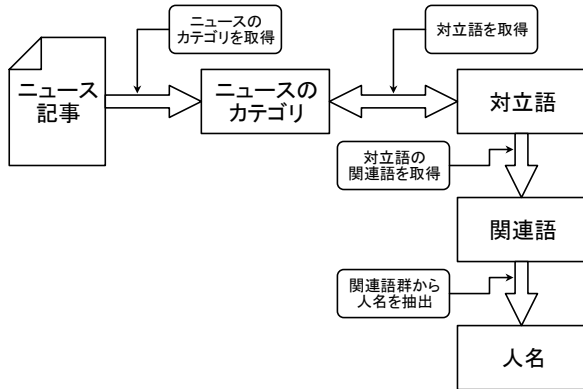


図 5 取り違えボケの手順

は Web ニュースサイトのディレクトリ構造を用いて取得する。取得したカテゴリが主題カテゴリとなる。次に、取得した主題カテゴリから対立語を取得する。

そして、対立語に対し関連語群を取得する。本論文では有名な人物であればそのカテゴリに関する知識があまりない人であっても認知していると考え、取得した関連語の候補から、人名のみを抽出し、関連語とする。最後にニュース記事のジャンル、主題語、対立語、関連語を用い、予め用意した雛形に当てはめて使用する。

これまで我々是对立ボケにおいて対立語の抽出手法を提案してきた [4] [6]。既存手法の対立語の定義を以下に示す。

- 対立関係を持つ語同士は共通の上位概念を持つ
- 対立関係を持つ語同士の認知度が近似している

既存手法の対立語の生成の有用性を計る実験の結果では、適切な対立語が抽出されていないことが判明した。これは、単語の概念構造と認知度のみから対立語を求めているためと考えられる。

そこで本論文では、以下のように、改めて対立語を定義しそれに基づいた対立語の抽出手法を提案する。

(1) 対立関係を持つ語同士は文法上の使われ方が類似している

(2) 対立関係を持つ語同士は概念構造上互いに包含関係にない

以下、上記の定義による対立語の抽出手法を述べる。

条件 1

Word2Vec には、似た使われ方をする語は似たベクトルを生成するという特性がある。そこで、本論文では似た意味を持つ

単語は似た使われ方をすると考え、Word2Vec の学習結果を用いることで類似した単語の取得を行う。今回は話し言葉が大量に得られると言う点から、教師データとして TwitterAPI (注2) を用いて無作為に取得したツイートを用い、形態素解析器に MeCab (注3) を用いる。本研究では 281,412,044 ツイートを用いた。表 2 にその一部を示す。

条件 2

包含関係の抽出には、概念構造として Wikipedia のカテゴリ構造を用いる。一方の語がもう一方の語の子カテゴリである場合に概念構造上包含関係にあるとする。そのため、互いに互いの子カテゴリではない場合に概念構造上包含関係にないとして抽出する。

複数の語が対立語の候補として抽出された場合は、Word2Vec により学習された単語ベクトルの類似度がキーワードと最も高いものを対立語として抽出する。

4.3 関連語の抽出

関連語の抽出には、Wikipedia のカテゴリ構造を用いる。まず、対立語を Wikipedia のカテゴリ構造にマッピングする。そのマッピングされたカテゴリの子カテゴリである全てのページを取得し、その取得されたページを関連語の候補とする。

関連語の候補の中から、人名を関連語として決定する。このとき、人名は、我々の作成した人名辞書を用いる。人名辞書は Wikipedia のカテゴリ構造から生成している。

人名辞書の生成手順は以下の通りである。人名辞書を生成する為の Wikipedia のカテゴリ構造の概略図を図 6 に示す。図中の四角形がカテゴリ、丸がページ、実線の矢印が多数のページが取る概念構造の関係、破線の矢印が一部のページが取る概念構造の関係、バツ印が意図的に使用しない概念構造の関係を表している。まず、「生年」「没年」2つのカテゴリの子にある全てのページを取得する。この際、人名ではない様々なページが取得されるので除外を行う。まず、クローン羊のドリーなどの著名な動物は、「著名な動物」カテゴリの子カテゴリであるため、これを用いて除外する。ただし、「生年」「没年」2つのカテゴリも「著名な動物」カテゴリの子カテゴリであるため、この2つのカテゴリを通らずにたどり着いたページを対象とする。次に、バブル世代、宝塚歌劇団1期生などのページを除外する。具体的にはページ名にある特徴を用いて以下のように分けることが

(注2) : <http://apps.twitter.com>

(注3) : <http://taku910.github.io/mecab/>

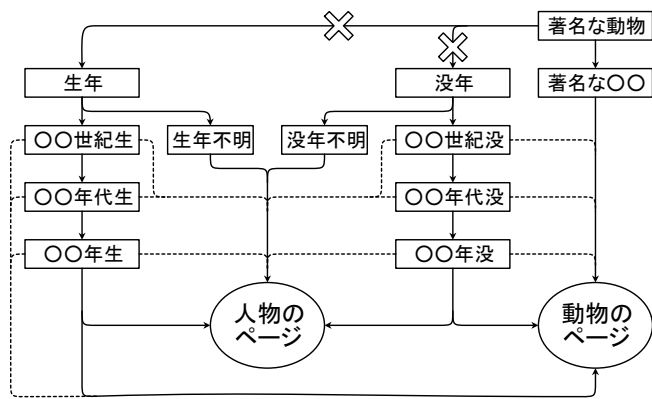


図 6 人名ページの抽出

できる。

- 人名に用いられることのない文字列が入っている
例: 「バブル世代」の「世代」
- アルファベット又はアラビア数字が用いられている
ただし、ミドルネームはその限りではない
例: 「宝塚歌劇団 1 期生」の「1」

関連語が複数有る場合、最も認知度の高い語を取り違えボケに用いる。認知度は Web 検索の検索結果数の数を用いる。しかしながら、取得された関連語の数が閾値より多い場合、はじめに書きかけのページと、概要に対立語が入っていないページを除外する。書きかけのページは、書きかけのページ専用のテンプレートが存在するため、それが用いられているページを書きかけのページとする。次に、除外後も関連語の数が閾値より多い場合、記事の文字数が多い順に上位から閾値の数までの語を対象にする。

5. 実験

提案手法の有用性を示すために、対立語の実験と取り違えボケの実験の 2 つの実験を行った。

5.1 対立語

実験条件

提案した対立語の抽出手法の有用性を示すため、評価実験を行った。被験者は 20 代の男性 8 人である。実験は既存手法によって抽出された対立語と、提案手法によって抽出された対立語をそれぞれキーワードと共に 25 対ずつ見せた。はじめに既存手法を用いて実験を行い、十分に日数を空けた上で提案手法を用いて実験を行った。キーワードは既存手法と提案手法で同じものを用いた。提示する順番は被験者によってランダムである。「対立語として適しているか」という設問に対し、「適している」「どちらかと言えば適している」「どちらとも言えない」「どちらかと言えば適していない」「適していない」「わからない」で評価を行った。提示したデータを表 3 に示す。

結果と考察

図 7 にそれぞれの解答の合計が全体の解答の合計に占める割合を示す。

設問に対し、「適している」「どちらかと言えば適している」という解答は既存手法では 32% 存在したのに対し、提案手法では

表 3 提示した対立語

キーワード	既存手法	提案手法
学校	家族	通学
先輩	医者	後輩
猫	主人公	イヌ
夏	少年	冬
綺麗	さくら	夕焼け
電話	レッド	呼出
マンガ	映画	小説
ダイエット	企画シングル	痩身
ラーメン	うどん	蕎麦
クリスマス	ハロウィン	ハロウィン
コーヒー	オレンジ	レモンティー
自転車	映画	原動機付自転車
サッカー	野球	野球
シンガーソングライター	坂本龍一	音楽家
ラッシュアワー	ダイ・ハード	芸人
夜	老人	昼
病院	診療所	点滴静脈注射
カレー	ゼリー	ハヤシライス
武器	チップ	防具
地獄	ふるさと	天国
温泉	博物館	ミョウバン
結婚式	卒業式	アヴェ・マリア
新幹線	秋田新幹線	各駅停車
ドイツ	フランス	イタリア
警察	消防	交番

52.5% 存在し、20 ポイント以上増加した。また、「どちらかと言えば適していない」「適していない」という解答は既存手法では 50% 存在したのに対し、提案手法は 22.5% 存在し、25 ポイント以上減少した。既存手法では半数が「どちらかと言えば適していない」「適していない」と解答したのに対し、提案手法では半数が「適している」「どちらかと言えば適している」と解答した。既存手法と提案手法で同じ対立語を提示した 2 対に対しては、結果に変化が見られなかったため、日数経過等の影響は出ていないと考えられる。

改善した例として、「猫」に対する対立語として「イヌ」が抽出されたものがある。これは犬、猫どちらも愛玩動物として飼育される身近な動物である為、話し言葉における出現パターンが類似し、結果として Word2Vec が似たベクトルを生成したと考えられる。悪化した例として、「温泉」に対する対立語として「ミョウバン」が抽出されたものがある。これは温泉や湯の花等の成分としてミョウバンが存在する場合があります、また概念構造上包含関係になかったため抽出されたと考えられる。

各被験者の解答の傾向として、提案手法の回答全体の評価が既存手法から全員改善しているという点がある。これは、全体の抽出精度が改善した結果と考えられる。

5.2 取り違えボケ

実験条件

我々の提案する「取り違えボケ」の有用性を示すため、評価実験を行った。被験者は 20 代の男女 6 人である。実験は既存手

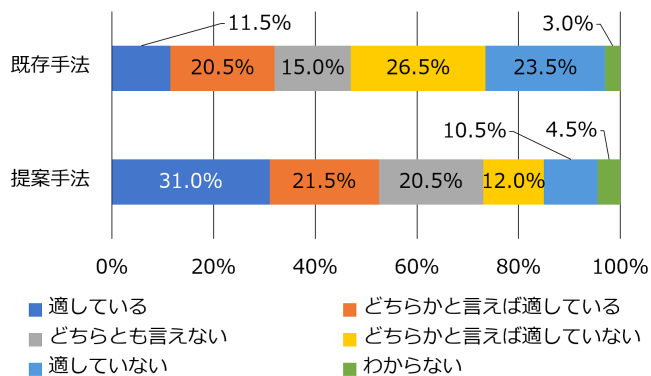


図 7 対立語のアンケート結果

法の季節の挨拶と表情ボケによるつかみの漫才と、提案手法の取り違えボケを用いたつかみの漫才の2つを被験者に見せた。提案手法のつかみは図4に示したものをを用いた。「主題が明確か」「対話は自然か」「漫才が面白いか」という3つの設問に対し、「言える」「どちらかといえば言える」「どちらとも言えない」「どちらかといえば言えない」「言えない」の5段階で評価を行った。図8に設問「主題が明確か」の結果を、図9に設問「対話は自然か」の結果を、図10に設問「漫才が面白いか」の結果を示す。

結果と考察

「主題は明確か」という設問に対しての回答は、既存手法は「どちらとも言えない」を選んだ被験者が17%いたのに対し提案手法はこの項目を選んだ被験者がいなかった。また、「言える」と答えた被験者が67%から83%に16ポイント増加したため、提案手法により主題が明確となり、ある程度改善したと言える。これは、つかみの段階で主題カテゴリを含む対話を行うことによりニュースのカテゴリを印象付けることができたためと考えられる。「対話は自然か」という設問に対しての回答は、既存手法は被験者の半数が「どちらかといえば言える」と回答したのに対し、提案手法は被験者の67%が「言える」と回答した。また、「言える」「どちらかといえば言える」と解答した被験者は、既存手法では66%に対し、提案手法では83%と、17ポイント増加した。さらに既存手法では「どちらかといえば言えない」を選んだ被験者がいたのに対し、提案手法では「どちらかといえば言えない」と「言えない」を選んだ被験者はいなかったため、提案手法により対話がより自然になり、改善したといえる。これは、既存手法で発生していたつかみと本ネタとの間に存在していたギャップを、提案手法はつかみの段階で主題カテゴリを対話に持ち出すことにより埋めることができたためと考えられる。「漫才は面白いか」という設問に対しての回答は、「どちらかといえば言える」の回答が33%から50%に17ポイント増加したため、漫才の面白さがある程度改善したと言える。これは、つかみの変更によりよりわかりやすい掛け合いとなったためと考えられる。

6. まとめと今後の課題

本論文では、つかみに用いる新たなボケであり、主題を明確

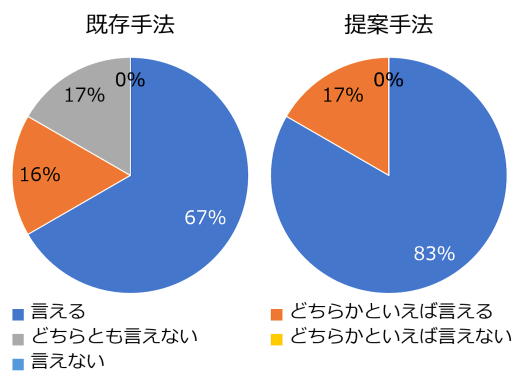


図 8 取り違えボケの実験結果「主題が明確か」

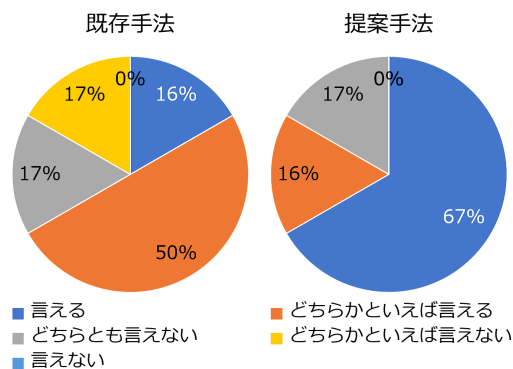


図 9 取り違えボケの実験結果「対話は自然か」

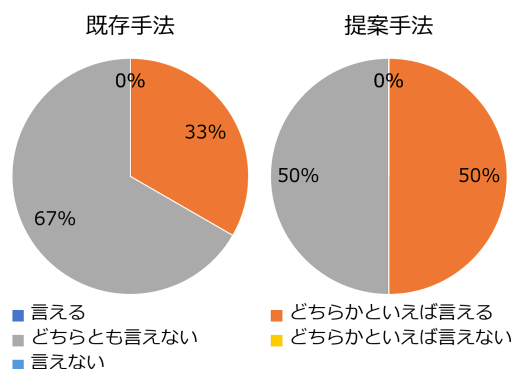


図 10 取り違えボケの実験結果「漫才が面白いか」

にすることができる新たなボケである取り違えボケの提案と、それに対する評価実験を行った。これにより漫才のつかみにより漫才の本ネタに用いられているニュースの主題が明確となり、対話がより自然になったことを確認した。また、それと同時に Word2Vec を用いて新たな対立語の抽出手法を提案し、それに対する評価実験を行った。これにより対立語の抽出精度が改善したことを確認した。

取り違えボケに対する今後の課題は、主題が人名以外であった場合の生成手法の提案が挙げられる。取り違えボケでは、ニュースの主題が人名であることが多いため、主題カテゴリの人名はニュースの主題を用いている。しかし、ニュースの主題が常に人名とは限らず、人名以外が主題語となった場合、提案手法では生成不能に陥る。このような場合に適応可能な改善

手法が必要となる。また、既存手法である季節の挨拶との共存も今後の課題となる。これは、季節の挨拶を行わずにつかみを行った場合、季節の挨拶に存在した利点が活かされなくなるためである。

対立語に対する今後の課題は、Word2Vecのベクトルの類似度において上限と下限の閾値の設定、及び学習時のパラメータの最適化が挙げられる。これは、対立語の抽出手法の改善によって、多くの語に対して抽出結果が改善されたが、一部の語に対してはかえって悪化したためであり、提案手法において、これらの値は設定しておらず、直感的に対立語と思えない結果が抽出されたと思われるためである。また、学習時に用いる教師データの大きさの最適化も今後の課題である。今回は教師データとして2億8千万ツイート余を用い、パラメータは標準値のまま学習を行ったが、短期間でこれほどの量のツイートを取得するのは現実的ではないと同時に先程述べたようにより抽出精度も改善の余地があるため、より少ないツイートを用いてより適切な学習をする必要がある。

謝 辞

本論文の一部はJSPS科研費26330347, 16K07973及び、私学助成金（大学間連携研究補助金）の助成によるものである。ここに記して謹んで感謝の意を表す。

文 献

- [1] SoftBank <http://www.softbank.jp/robot/>
- [2] SHARP <https://robohon.com/>
- [3] T. Kanda, H. Ishiguro, T. Ono, M. Imai, and R. Nakatsu. Development and evaluation of an interactive humanoid robot "robovie". In Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on, pages 1848–1855. IEEE, 2002.
- [4] 真下遼, 梅谷智弘, 北村達也, 灘本明代 “Web ニュースからの漫才台本自動生成を用いたコミュニケーションロボット” WebDB Forum 2014
- [5] 安部達雄 “漫才における「フリ」「ボケ」「ツッコミ」のダイナミズム” 早稲田大学大学院文学研究紀要 第3分冊 日本文学演劇映像美術史日本語日本文化, Vol.51, No.28, pp. 69-79, 2006
- [6] 真下遼, 灘本明代 “対立語抽出に基づく Web ニュースからの漫才ロボット台本自動生成手法の提案”, DEIM Forum 2014 C2-4
- [7] Word2Vec <https://code.google.com/p/word2vec/>
- [8] 北山大輔, 角谷和俊 “ニュースアーカイブのための”コンテンツ構成順序を用いた比較ニュース検索” 日本データベース学会 letters, Vol. 6, No. 1, pp. 169-172, 2007
- [9] 田中祥太郎, ヤフトアダム, 田中克己 “ニュース記事の理解支援のための背景知識抽出と補完” 情報処理学会研究報告 Vol. 2014- DBS - 159No, 17
- [10] 野沢健人, 中岡義貴, 山本修平 “word2vec を用いた代替食材の発見手法の提案” 電子情報通信学会技術研究報告. DE, データ工学 114(204), 41-46, 2014-09-03
- [11] 土田崇仁, 遠藤雅樹, 加藤大受, 江原遥, 廣田雅春 “Word2Vecを用いた地域やランドマークの意味演算” DEIM Forum 2016 H5-1
- [12] 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 (Special Issue) pp.265-272 (2012)
- [13] 林宏太郎, 神田崇行, 宮下敬宏, 石黒浩, 萩田紀博 “ロボット漫才 社会的受動メディアとしての2体のロボットの利用” 日本ロボット学会誌, Vol. 25, No. 3, pp. 381-389, 2007

- [14] 関綾子敏 “おかしみの生成における言語操作の構造-漫才の資料として-” 『早稲田日本語研究』, 早稲田大学国語学会, 2002 年.
- [15] 元祖爆笑王,-「漫才入門 ウケる笑いの作り方全部教えます」, リットーミュージック, 2011 年.