

対立語抽出に基づくWebニュースからの ロボット漫才台本自動生成手法の提案

甲南大学 知能情報部
真下 遼 灘本明代

背景



コミュニケーションロボット

共存



人

しかし

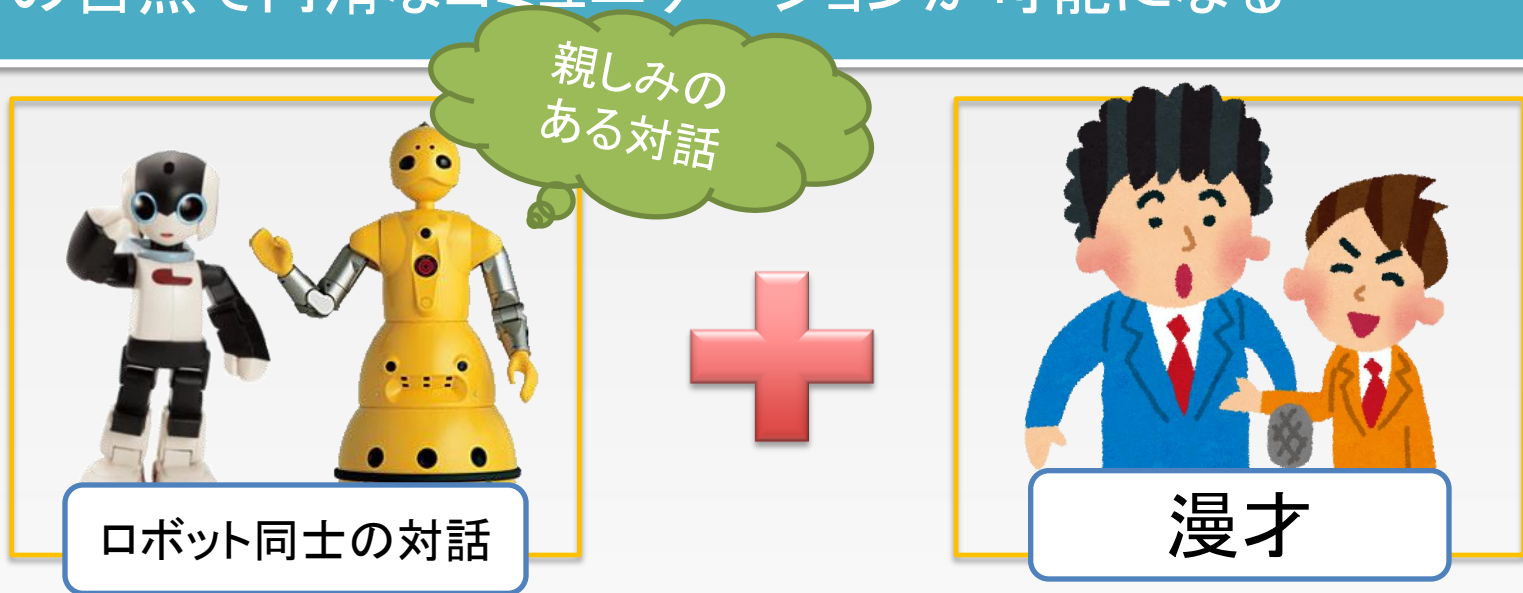
人とロボットのコミュニケーションは円滑に行われていない



人とロボットとの円滑なコミュニケーションの手助けとなるものが必要

背景

ロボット同士の対話観察を人が行うことで、人とロボットとの自然で円滑なコミュニケーションが可能になる



ロボット漫才

ロボット漫才

ロボットが漫才を演じることで
楽しいロボット同士の対話観察を
ユーザに提供

漫才の題材に時事ネタを取り入
れることでニュース離れの克服も
期待できる



アイちゃん
(ツッコミ)

なんでやねん!



ゴン太
(ボケ)

ロボット漫才の流れ

ユーザがお題を入力

お題に関連したニュース記事を
Webから取得

漫才の台本を自動生成

ロボットが漫才を行う

漫才の構成(三段構成)

つかみ

漫才の導入部分
挨拶と一笑い

表情ボケ

本ネタ

漫才の主軸となる部分
ニュース記事を読み上げて紹介しながら様々なボケとツッコミで笑いを取る

言葉遊び
フリツッコミ
過剰ボケ
対立ボケ

オチ

まとめと最後の笑い

謎かけ

対立ボケ

ニュース記事だけでは得られない知識の拡大
対立関係は直感的に把握しやすくボケとして明確

ボケ

「2020年夏季五輪の開催都市は**東京**が選ばれた」

あれやろ？
近畿地方に属する日本の都道府県の一つやろ？

そうなんか、まあでも似たようなもんやろ

ツッコミ

なるほど
ところで**東京**ってどんなかわかてるか？

対立関係

お前それ**大阪**と勘違いしてるやろ。
東京は、日本の関東地方南部に位置し、1869年2月11日以来、日本の首都が置かれている都市、地域や！

どこがやねん、怒られるで！

対立関係の特徴

対立関係を持つ語同士は

- 共通の上位概念を持つ
- 語の意味が対照的である
- 実際の対戦経験がある
- 属性の属性値が近似している
- 言語使用域が近似している

対立関係の定義

対立関係を持つ語同士は共通の**上位概念**を持つ

- 東京 ⇔ 大阪 : 日本の都市
- 野球 ⇔ サッカー : 球技

対立関係を持つ語同士の**認知度**は互いに近い

- **野球 ⇔ サッカー : 野球と同じくらいメジャー**
- 野球 ⇔ フットサル : 野球程普及していない

「東京」の上位概念語

共通の上位概念を持つ語
同位語

対立関係を持つ語同士は共通の上位概念を持つ

「東京」の上位概念

日本の都市

ミリオンモンキーズに登場する実在都市

就航都市

就航地

グループ

過去に存在した店舗

オリジナルアルバム

曲

シングル

収録曲

登場人物

作品

Wikipedia
の階層構造
より

上位概念語の重み

- 上位概念語にはそれぞれ重みを持たせる
- 少数の概念だけに含まれる上位概念語の方がより重要性が高い
 - 日本の都市 下位概念語数: 24語
 - 作品 下位概念語数: 120257語

重要

$$\text{重み} = \log \frac{\text{その概念語が持つ下位概念語の数}}{\text{全ての概念語の数}}$$

同位語にキーワードと共通の上位語の重みを加算する

「東京」の上位概念語と重み

「東京」の上位概念	下位概念語数	重み
ミリオンモンキーズに登場する实在都市	9	12.6
日本の都市	24	11.7
就航都市	240	9.4
就航地	258	9.3
グループ	333	9.1
過去に存在した店舗	347	9.0
オリジナルアルバム	1100	7.8
曲	3717	6.7
シングル	19519	5.0
収録曲	35981	4.4
登場人物	95662	3.4
作品	120257	3.2

「東京」の同位語と重み

「東京」の同位語	重み
ニューヨーク	35.1
シドニー	34.2
ロンドン	33.9
パリ	33.9
居酒屋	31.8
バンクーバー	30.7
北京	30.7
ひまわり	30.6
上海	26.3
大阪	24.5

認知度比

対立関係を持つ語同士の認知度は互いに近い

- 認知度 \doteq 検索結果数
- キーワードと同位語それぞれの検索結果数の比率から対立語を取得する

$$\text{認知度比} = 1 - \frac{\mathit{abs} (|X| - |Y|)}{\mathit{max} (|X|, |Y|)}$$

|X|: キーワードの検索結果数

|Y|: 対立候補語の検索結果数

「東京」の同位語と重み

「東京」の同位語	検索結果数	認知度比
東京	324000000	1.00
ニューヨーク	16100000	0.05
シドニー	2750000	0.01
ロンドン	16200000	0.05
パリ	17600000	0.05
居酒屋	40700000	0.17
バンクーバー	2430000	0.01
北京	625000000	0.52
ひまわり	9050000	0.03
上海	130000000	0.41
大阪	470000000	0.69

「東京」の対立語抽出結果

「東京」の同位語	重み	重み正規化	認知度比	相乗平均
ニューヨーク	35.1	100%	5%	22.4%
シドニー	34.2	97%	1%	9.8%
ロンドン	33.9	97%	5%	22.0%
パリ	33.9	97%	6%	24.1%
居酒屋	31.8	90%	17%	39.1%
バンクーバー	30.7	87%	1%	9.3%
北京	30.7	87%	45%	62.5%
ひまわり	30.6	87%	3%	16.6%
上海	26.3	75%	51%	61.8%
大阪	24.5	70%	60%	64.8%

「野球」の対立語抽出結果

同位語	重み	重み正規化	認知度比	相乗平均
サッカー	78.1	100%	99%	99.5%
バレーボール	78.1	100%	37%	60.8%
バスケット	55.2	71%	36%	50.6%
ソフトボール	44.8	57%	0%	0.0%
アイスホッケー	40.0	51%	1%	7.1%
テニス	34.1	44%	55%	49.2%
フットサル	33.6	43%	6%	16.1%
ゴルフ	30.4	39%	78%	55.2%
巨人の星	26.6	34%	1%	5.8%
カーリング	20.7	26%	1%	5.1%

対立ボケ

ボケ

「2020年夏季五輪の開催都市は**東京**が選ばれた」

あれやろ？

近畿地方に属する日本の都道府県の一つやろ？

そうなんか、まあでも似たようなもんやろ

ツッコミ

なるほど
ところで**東京**ってどんなかわかっているか？

Wikipediaの最初の一文

お前それ**大阪**と勘違いしてるやろ。
東京は、日本の関東地方南部に位置し、1869年2月11日以来、日本の首都が置かれている都市、地域や!

どこがやねん、怒られるで！

実験

- 目的

- 提案手法によりキーワードから取得される対立語の考察を行う

- キーワード

- YahooのWebニュース記事の記事タイトルを形態素解析して単語を取得しWikipediaに項目が存在する単語をキーワードに設定

結果と考察

キーワード	対立語
日本	中国
アメリカ	フランス
イチロー	松井秀樹
バラク・オバマ	ジョージ・ワシントン
与党	野党
チェス	オセロ
インターネット	新聞
マイクロソフト	東芝
マクドナルド	吉野家
エベレスト	阿里山
車	祖父
結婚	四季

キーワードに関して概ね関連性の高い語が取得された

認知度を考慮することにより上位概念語が共通の語でも一意に対立語らしい語を決定できた

上位概念語が適切でない

対立語候補の同位語の取得がWikipedia記事に依存している

他の対立関係の特徴も加味し対立関係の定義を追加してより対立語らしい語を取得する

動画

エンタメ

エンタメ総合

音楽

映画

ゲーム

アジア・韓流

滝川クリステル、流行語大賞受賞に感謝！7年後の東京五輪開催にも願い

シネマトゥデイ 2013年12月3日(火)10時34分配信



オリンピック招致活動の最終プレゼンテーションに参加した滝川クリステル - Ian Walton / [Getty Images Sport](#)

フリーアナウンサーの滝川クリステルが、「お・も・て・な・し」が「2013ユーキャン新語・流行語大賞」年間大賞を受賞したことへの感謝をオフィシャルブログにつづった。その一方で、「お・も・て・な・し」が生まれるきっかけとなった東京オリンピックの開催は7年後。「決して一過性の言葉として忘れてしまうのではなく、7年後には、大輪の花を咲かせ、その後もこの『おもてなし』という文化がずっと続いていくことを願っています」と願いを込めている。

滝川クリステル、こける…フォトギャラリー

滝川は年間大賞が発表された2日夜、1月以上ぶりにブログを更新すると、「本日、『おもてなし』が2013ユーキャン新語・流行語『年間大賞』に選ばれました。」と感謝。続けて、受賞は自分だけのものではなく、招致に関わった全員のものであると強調すると、「招致メンバーとして声を掛けていただいたこと、そして、この言葉を世界の舞台上で発することができたこと、まさに奇跡だったような気がします」と改めて大舞台を振り返った。

「日本人が元々持っている、大事な大事な心の機微がなしえる『おもてなし』を改めて再認識し、自信を持たれたことが何よりもうれしいことです」という滝川。7年後、そしてその後も「お・も・て・な・し」の文化が続くことを願っている。

滝川は9月7日にアルゼンチン・ブエノスアイレスで行われたオリンピック招致活動の最終プレゼンテーションで「お・も・て・な・し」とスピーチ。開催都市が東京に決定したこともあり、「お・も・て・な・し」は一躍ブームとなっていた。(編集部・福田麗)

まとめと今後の課題

- まとめ

- 人とロボットの円滑なコミュニケーションの実現を目的にロボット漫才のための漫才台本自動生成手法を提案
- 対立語に着目した対立ボケの生成手法を提案した。

- 課題

- ロボット漫才の有用性の確認
- 対立語の定義についての追求