

おかしみの構造図に基づく人名ボケの自動生成

11571098 原口 和貴 (灘本研究室)

あらまし: 本研究では, ロボットの特徴を活かした, おかしみの構造図に基づくボケとして, 新たに人名ボケの提案を行う. 具体的には, 入力された人名の持つ属性に対して誤った属性を抽出し, 取得した属性を対比することでおかしみを生じる, 人名ボケを提案する.

1. はじめに

近年, コミュニケーションロボットの発展に伴い, ロボットと人のコミュニケーションは重要になってきている. しかしながら, 人とロボットの円滑なコミュニケーションが行われていないのが現状である. そこで我々は, 人とロボットの円滑なコミュニケーションとして「笑い」に着目し, 漫才ロボットの研究^[1]をしてきた. 開発した漫才ロボットを図 1 に示す.



図 1 漫才ロボット

従来手法では, ニュース記事を基に漫才の台本を自動生成しているが, おもしろい漫才を生成しているとは言いがたい. ユーザに親しみのある話題として, 人々が知っている有名人を対象とすることにより, わかりやすくより面白い漫才ができると考えた. そこで本論文では, ロボットの特徴を活かした, おかしみの構造図^[2]に基づく人名ボケの提案を行う.

2. 提案手法

提案する手法では, ユーザにとって親しみ深い有名人を入力とする. 以下, 本文では有名人の名前を人名と呼ぶ.

2.1 人名ボケとは

提案する人名ボケの生成例を図 2 に示す. 人名ボケとは, ボケ役が入力された人名の属性好きだと述べる. これに対してツッコミ役が, その

ボケ	フィギュアスケートが好きやねん
ツッコミ	ほお, ほかにどなんがおるんや?
ボケ	羽生結弦, 浅田真央, 宇野昌磨, 高橋大輔, 安藤美姫
ボケ	千葉真一, 仲本工事, 田中理恵, 村上茉愛, 内村航平
ツッコミ	って, 途中からフィギュアスケートじゃなくて体操競技になってるやん!
ボケ	人工知能でこんなこともできます
ツッコミ	余計なことするな...

図 2 人名ボケの生成例

属性を有する他の人名を問う. ここで言う属性とは, 人物の職業など, その人物を特徴づけるものとする. 属性は, 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^[3]ことから, Wikipedia 記事の最初の 1 文を用いて取得する. 人名・芸名やグループ名, 身体的特徴は属性ではないとする.

次に, ボケ役は回答の途中で誤った属性の人名を答え, ツッコミ役がツッコミを入れ訂正する. 図 3 に人名ボケのおかしみの構造図を示す. 以下, 本文では入力された人名の属性を元属性, 間違えた人名の属性を誤属性と呼ぶ. ここで, 元属性と誤属性の乖離がおかしみの構造図における概念 A, B に相当し, これがおかしみを生じる. 人名ボケの生成手順を図 4 に示す.



図 3 人名ボケのおかしみの構造図

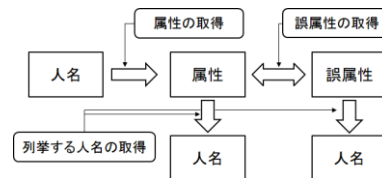


図 4 人名ボケの生成手順

2.2 誤属性の生成

人名ボケでは, 元属性から誤属性の生成を行う. この際, 元属性と誤属性の連想のしやすさを考慮する. 元属性と誤属性の対比を適切に行うため, 誤属性を以下のように定義する.

条件1. 元属性と共通の上位概念を持つ.

Wikipedia のカテゴリ構造のツリーにおいて, 元属性と兄弟関係にあり, 元属性の親カテゴリにはないカテゴリが親カテゴリに存在するものを, 誤属性として抽出する.

条件2. 文法上の使われ方が元属性とある程度似ている.

Word2Vec^[4]を用いることで使われ方の類似した語を抽出する. 今回は松田ら^[5]が公開している日本語 Wikipedia エンティティベクトルを用いる. 元属性との Cos 類似度が 0.7 以下の

語を誤属性として抽出する。

条件3. 列挙する人名の認知度がある閾値以上である。

人名ボケでは属性に基づき人名を列挙する。列挙する人名の認知度が低い場合、ユーザに親しみある漫才とは言えない。ここで、認知度は Google 検索結果件数とする。

誤属性の列挙する人名の認知度がある閾値以上の場合、誤属性として抽出する。

2.3 列挙する人名の決定

取得した各属性から、列挙する人名を決定する。属性を有するすべての人物の Wikipedia ページを取得する。次に以下の式を用いて各人物の Score を算出する。

$$\text{Score} = \text{PageRank} + \log(\text{閲覧数})$$

Score 上位 20 件の人名に対して Google 検索結果件数を求め、検索結果件数が多い順に列挙する人名として決定する。

3. 実験と考察

提案手法の有用性を示すため、評価実験を行った。実験は 20 代の男女 6 名に対し行った。「大坂なおみ」を入力とし、従来手法により生成された漫才と、提案手法の人名ボケを用いた漫才の 2 つを被験者に見せた。

ここで、システムが提示した人名ボケにおける大坂なおみの属性はテニス、誤属性はゴルフであり、列挙した人名は大坂なおみ、セリーナ・ウィリアムズ、ノバク・ジョコビッチ、ロジャー・フェデラー、杉山愛、倉本昌弘、池田勇太、片山晋呉、松山英樹、石川遼である。

実験結果を表 1 に示す。

表 1 実験結果

設問	どの部分がボケかわかったか	
	わかった	わからなかった
提案手法	83%	17%
従来手法	67%	33%

「人名ボケがわかったか」という設問に対して、83%がわかったと回答した。従来手法に用いられるボケでは 33%がわからなかったと答えていることから、わかりやすいボケが生成できたと考えられる。また、わかったと回答した 5 名に「人名ボケはおもしろかったか」という設問を行い、「とてもおもしろい」「おもしろい」「どちらとも言えない」「つまらない」「とてもつまらない」の 5 段階で評価した。結果を図 5 に示す。40%が「おもしろい」、60%が「どちらとも言えない」と回答した。「どのような点がおもしろかった/つまらなかったか」という設問に対して、「誤属性のことをよく知らないためわからな

かった」という回答があった。このことから、元属性や誤属性自身の認知度を考慮するひつようがあると考えられる。

提案手法のわかりやすさに関する自由記述欄に「漫才の流れがわかりやすかった」とあり、このことから提案手法の漫才はわかりやすいと言える。

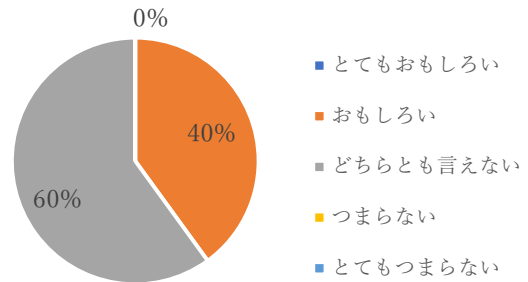


図 5 人名ボケはおもしろかったか

4. まとめと今後の課題

本論文では、よりわかりやすく面白い漫才台本を生成するために、人々が知っている有名人を対象とした漫才台本の自動生成を提案し、評価実験を行った。これにより人名を用いることで、漫才の内容がよりわかりやすくなり、よりおもしろい漫才となったことを確認した。

今後の課題として、お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。提案手法ではユーザにとって親しみ深い話題として人名を入力としているが、実際には人名以外をお題とすることも多々ある為、このような場合に適用可能な手法が必要となる。

参考文献

- [1] 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” Web DB Forum 2014
- [2] 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科紀要 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
- [3] Kotaro Nakayama, Takahiro Hara, and Shojiro Nishio. Wikipedia mining for an association web thesaurus construction. WISE 2007, Vol. 4831 of WISE 2007, pp.322-334, 2007.
- [4] Tomas Mikolov, Wen-tau Yih, Geoffrey Zweig. Efficient estimation of word representations in vector space. 2013
- [5] 鈴木正敏, 松田耕史, 関根聡, 岡崎直観, 乾健太郎 “Wikipedia 記事に対する拡張固有表現ラベルの多重付与” 言語処理学会第 22 回年次大会, 2016