

AI っち : Word2Vec を用いたなぞかけ作成支援システム

大原嶺† 松澤智史† 武田正之†

東京理科大学 理工学部 情報科学科†

1. はじめに

ユーモアの理解は人間の高度な知的活動である。近年の AI 技術の発展は著しく、以前は難しいとされていたクリエイティブ分野でも応用は進んでいるが、しかし、ユーモア関連での成果は比較的少ない。こういった研究は、直接的には人間の役に立たずとも、それ自体が人間の知的活動を模しているといった点で意義深いと考える。

2. 目的

本研究ではユーモア分野の中でも、ある程度パターンが定型化されているなぞかけに着目する。既存研究とは異なるアルゴリズムを用いることにより、熟練した人間同等のなぞかけ作成を支援するシステムを構築する。関連語の抽出部分には Word2Vec の学習モデルを用いる。

3. 基礎知識

3.1 なぞかけ

なぞかけとは、落語にルーツを持つ言葉遊びである。近年では、2010 年にお笑いタレントのねづっち氏がブームを起こした。ルールは、「お題」となる言葉 A に対して「A とかけて B と解く。その心はどちらも C/C'。」といった形式を満たす言葉 B, C, C' を考えるものである。

例「卒業論文とかけて手紙と解く。その心はどちらも、はいけい（背景/拝啓）が最初に必要でしょう。」

3.2 Word2Vec

Word2Vec[1]は、Tomas Mikolov らによって提案された、単語の意味をベクトル表現するための手法である。学習には、2 層からなる順伝播型ニューラルネットワークを用いる。前後の単語から対象単語が現れる条件付き確率を最大化して推測する CBow モデルと、出力層における周辺単語予測のエラー率の合計を最小化して推測する Skip-grams モデルがある。ベクトル化された 2 単語についてコサイン類似度を計算することで、それらの類似度を計算することができる。

4. 関連研究

前田らの研究[2]は、なぞかけを自動生成するためのアルゴリズムを確立し、データベース上に登録された語について実際になぞかけを作成するシステムを提案している。アルゴリズムの詳細を図 1 に示す。お題 A に対して心となる関連語 C を先に見つけ、C の音韻類似語 C' を抽出した後、解として C' の関連語 B を抽出する。



図 1: 関連研究のなぞかけ自動生成アルゴリズム

関連語を抽出する部分では、連想概念辞書[3]を用いている。辞書に掲載されている単語については人間らしいなぞかけを生成できているが、一方で、辞書にはない固有名詞や新語では生成することができない。なぞかけは固有名詞や時事ネタを題材にする機会も多く、これは多彩ななぞかけを作る上での障壁となる。本研究では、Word2Vec の学習モデルを用いてこの問題を解決する。

4. 提案手法

4.1 概要

関連研究の問題点を解決するための手法を提案する。以降、なぞかけ熟練者であるねづっち氏のなぞかけ同等の完成度を持つなぞかけの作成を支援する意味で、本システムを AI っちと呼ぶ。

4.2 Word2Vec による学習

日本語版 wikipedia の全記事を MeCab[4]と MeCab 用の辞書 mecab-ipadic-NEologd[5]を用いて形態素解析と分かち書きを行い、その後 Word2Vec により学習を行う。Word2Vec による学習時のパラメータについての詳細を表 1 に示す。

表 1: Word2Vec による学習時のパラメータ

モデル	Skip Gram
文脈窓	15
ベクトルサイズ	200
単語最低出現回数	50

4.3 AI っちのアルゴリズム

Word2Vec の学習モデルを用いてなぞかけを生成する部分のアルゴリズムは以下の通りである。

- (1) Word2Vec の学習モデルに入力としてお題 A を与える。A は Word2Vec の語彙に含まれるものに限定される。
- (2) A の類似語を上位から順に一定数抽出する。
- (3) Word2Vec の語彙の中からランダムに解 B を抽出し、B の類似語を上位から順に一定数抽出する。
- (4) A の類似語と B の類似語のなかから、音韻が同一となるものをマッチングする。
- (5) 見つかった組み合わせを心 C/C' として出力する。C/C' として、複数のパターンが出力される場合もある。

Althc : A support system for Nazokake using Word2Vec
 † Ryo Ohara Tomohumi Matsuzawa Masayuki Takeda
 Tokyo University of Science, Faculty of Science and
 Technology, Department of Information Science

図 2 に、入力として「探偵」を与えた場合の本アルゴリズムのイメージを示す。



図 2: 提案手法のなぞかけ自動生成アルゴリズム

5. 実験

5.1 概要

AI っちを用いて作成されたなぞかけを評価する被験者実験を行う。実験は、X 群、Y 群それぞれから抽出したお題が同一である 2 種類のなぞかけを混ぜて提示する。X 群には、なぞかけの作成経験が無く、なぞかけに関する特別な知識を持っていない被験者 A (20 代男性) が AI っちを用いて作ったなぞかけが含まれている。Y 群には、ねづっち氏が作ったなぞかけが含まれている。被験者 (男女 27 名、平均年齢 24.6 歳) は、提示されたなぞかけがどちらの群のものであるかは知らされないまま、どちらが面白かったかを判定する。

5.2 なぞかけの表現付加部分について

AI っちは「A とかけて B と解く。その心はどちらも C/C'。」の形式を満たす B, C, C' を提案するものだが、実際のなぞかけには C/C' の後に表現付加をする場合がある。今回は、X 群のなぞかけを作ってもらう際に、表現付加の部分は自身で考えるように説明した上で実験を行った。

5.3 手順

- (1) ねづっち氏がインターネット上で公開しているなぞかけ [6] から新しい順に 10 個選び、Y 群とする。なお、選出する際に、お題の言葉が AI っちの語彙に含まれていることを確認している。
- (2) Y 群に含まれる 10 個のなぞかけと同一のお題で、被験者 A にそれぞれ 5 個ずつ、AI っちを用いてなぞかけを作らせる。
- (3) 被験者群 B には 10 対 20 個のなぞかけが提示され、それぞれ「妥当性・納得感」「意外性・独創性」「面白さ」を 5 段階評価、「どちらが面白いか」を 2 択でを回答する。

5.4 実験に用いたなぞかけの例

X 群の例 (下線部は利用者が付加) : 「店員とかけて支点と解く。その心はどちらも、かいてん(開店/回転)に必要でしょう。」「足湯とかけて攻撃と解く。その心はどちらもせんとう(銭湯/戦闘)につきものでしょう。」

Y 群の例 : 「店員とかけてそれを面倒がる客と解く。その心はどちらも、かわそう(買わそう/かわそう)とされているでしょう。」「足湯とかけてスピード違反と解く。その心はどちらもふっとばす(フットバス/吹っ飛ばす)でしょう。」

5.5 結果

被験者 27 名の評価の平均値を図 3 に示す。

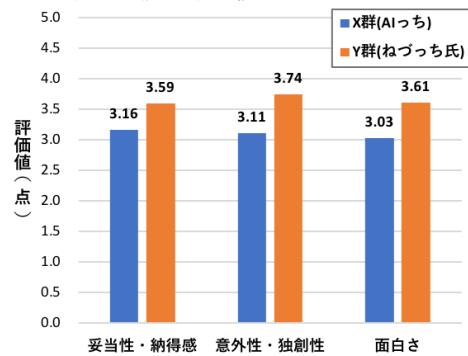


図 3: 被験者 27 名の評価の平均値

6. 考察と展望

関連研究の語彙 (1657 語) と比べて、AI っちの語彙 (131194 語) は大きく上回っているが、先行研究の面白さ (平均 2.91 点) を超えている。また、なぞかけ経験のないシステム利用者は 10 題すべてのお題でなぞかけを作ること的成功しており、妥当性・納得感の項目では平均 3.16 点である。以上より、なぞかけ作成支援システムとしての役割は果たせていると言える。

一方で、調査したすべての項目において Y 群の評価値の方が高くなっており、ねづっち氏のなぞかけには及んでいない。要因として、意外性・独創性の面で大きく劣っている点が挙げられる。ねづっち氏のなぞかけは関連語抽出のバリエーションが豊富 (足湯とフットバスなど) であったり、関連語 2 つをなぞかけに取り入れているなど、なぞかけ作成の上で幅広いテクニックを用いている。今後は、なぞかけ作成における様々なテクニックを取り入れ、多彩ななぞかけを提案すること、現在は利用者に委ねられている表現付加部分まで含めて提案することが望まれる。

参考文献

- [1] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, Jeffrey Dean, "Distributed Representations of Words and Phrases and their Compositionality", *Advances in Neural Information Processing Systems* 26, 2013
- [2] 前田実香, 鬼沢武久, 「言葉の関連性を用いたなぞかけ生成とその評価」, 筑波大学大学院博士課程 システム情報工学研究科修士論文, 2005
- [3] 慶応義塾環境情報学部石崎研究室編, 連想概念辞書 (2004 年度版第 1 版)
- [4] 工藤拓. Mecab: Yet another part-of-speech and morphological analyzer.
- [5] 佐藤 敏紀, 橋本 泰一, 奥村 学, 「単語分かち書き辞書 mecab-ipadic-NEologd の実装と情報検索における効果的な使用方法の検討」, 言語処理学会 第 23 回 年次大会 発表論文集, 2017
- [6] ねづっちチャンネル
https://www.youtube.com/channel/UC1-P_10oiJ0a_30_s6c6KMA (2019 年 1 月 10 日 最終確認)