

# 短文テキストの意味分析と意味類似検索

小山 大介<sup>†</sup>    上之園 隼志<sup>†</sup>    杉田 祥太郎<sup>†</sup>    大島 裕明<sup>††</sup>    莊司 慶行<sup>†††</sup>

田中 克己<sup>†</sup>

<sup>†</sup> 福知山公立大学 情報学部 〒620-0886 京都府 福知山市 字堀

<sup>††</sup> 兵庫県立大学 社会情報科学部 〒651-2197 兵庫県 神戸市 西区 学園西町

<sup>†††</sup> 青山学院大学 理工学部 〒252-5258 神奈川県 相模原市 中央区 淵野辺

E-mail: <sup>†</sup>{32045031,32045009,32245045,tanaka-katsumi}@fukuchiyama.ac.jp, <sup>††</sup>ohshima@ai.u-hyogo.ac.jp,  
<sup>†††</sup>shoji@it.aoyama.ac.jp

**あらまし** 本論文では、菰句（キャッチコピー）、俳句、短歌といった短文テキストの「意味」分析を行う手法を提案する。具体的には、語の意味分析を行う Word2Vec の結果を用いて、短文の意味分析を行う 2 つの手法を提案する。1 つ目の手法は、2 つの文の意味類似度をそれぞれに出現する各語の意味ベクトルの相互比較に基づいて計算する方法であり、2 つ目の手法は、与えられた文に出現する各語の意味ベクトルを加減算することで元の文の意味ベクトルを合成する方法である。これらの手法を菰句、俳句・短歌に適用して、意味が類似する菰句や俳句・短歌を検索する実験を行ったので、その実験結果についても報告する。

**キーワード** 短文テキスト, 菰句（キャッチコピー）, 俳句, 短歌, 意味分析, 情報検索

## 1 はじめに

テキストの機械学習技術と言語処理技術の進展により、様々なテキスト情報の「意味」の分析が可能となりつつある。特に、語や文や文書の「意味の分散表現」をニューラルネットワークの機械学習で求める T.Mikolov らの Word2Vec 手法が著名である [1] [2]。

文や文書の「意味」を分析することで、類似の意味を持つ文・文書の検索や新たな生成が可能となり、社会に蓄積された大量のテキストデータの新しい活用にもつながる。

本論文では、菰句（キャッチコピー）、俳句、短歌といった短文テキストの「意味」分析を行う手法を提案する。具体的には、語の意味分析を行う Word2Vec の結果を用いて、短文の意味分析を行う 2 つの手法を提案する。1 つ目の手法は、2 つの文の意味類似度をそれぞれに出現する各語の意味ベクトルの相互比較に基づいて計算する方法であり、2 つ目の手法は、与えられた文に出現する各語の意味ベクトルを加減算することで元の文の意味ベクトルを合成する方法である。

これらの手法を菰句、俳句・短歌に適用して、意味が類似する菰句や俳句・短歌を検索する実験を行ったので、その実験結果についても報告する。

第 2 節では、関連研究を紹介し、本研究の位置づけを示す。第 3 節では、短文テキストの「意味」分析を行う手法を 2 つ提案する。第 4 節では、菰句（キャッチコピー）の意味分析と意味類似検索、第 5 節では、俳句・短歌の意味分析と意味類似検索について述べる。第 6 節は本論文のまとめであり今後の課題についても述べる。

## 2 関連研究

語の意味の分散表現をニューラルネットワークの機械学習で求める手法として、T.Mikolov らの Word2Vec 手法が著名である [1]。T.Mikolov らは、その後、この手法を拡張して、文や文書の意味の分散表現を求める手法も提案している [2]。また、語の意味は時間や時代によって変化する。このため異なる時期・時代に使用されてきた語の意味を機械学習で求める手法が Yating らによって提案されている [3]。

最近では、機械学習によって新たな文やテキストを生成する仕組みについても多くの研究がなされており、例えば、T. Zhang らは BERT を用いたテキスト生成手法を提案し、生成されたテキストを評価するために BERTScore というアルゴリズムを提案している [4]。

著者らは、本研究の先行研究として、語の意味の分散表現から、短文（俳句、短歌、菰句）の「意味」を計算するアルゴリズムおよび意味類似する短文検索の手法を提案した [5]。

## 3 提案手法

本節では、語の意味分析を行う Word2Vec の結果を用いて、文の意味分析を行う手法を二つ提案する。

- 語ペアの意味類似度合成法
- 語の意味ベクトル合成法

### 3.1 語ペアの意味類似度合成法

語ペアの意味類似度合成法は、2 つの文の意味類似度を、それぞれの文に出現する各語の意味ベクトルを相互比較して計算して合成する方法であり、手法としては、与えられたテキス

トと自動生成したテキスト間の類似度を計算する BERTScore 法 [4] に基づくものである。BERTScore 法は、元来、事前学習された BERT から得られるベクトル表現を利用して、与えられたテキストと生成されたテキスト間の類似度を計算する評価手法で、その手順は以下の通りである：

- (1) 生成テキスト (Candidate) と正解テキスト (Reference) を BERT に入力し、トークン (語) のベクトル表現を獲得する。
- (2) それらのベクトル表現を利用して、トークン (語) 間の  $\cos$  類似度行列を作成。
- (3) 最後に各トークン (語) に対して最大類似度を利用して、Precision, Recall, F 値を計算し、スコアとする。

本稿で提案する、語ペアの意味類似度合成法は、基本的には BERTScore 法と同一である。但し、BERTScore 法では入力データが生成テキストと正解テキストであるところを、語ペアの意味類似度合成法では入力データとして任意の 2 つのテキストとした点が異なっている。

語ペアの意味類似度合成法の手順は以下の通りである (図 1 参照)。

- (1) 短文テキスト集合を Word2Vec で学習し、各語の意味を、意味ベクトル (Word2Vec の中間層の値のベクトル) として獲得する。
- (2) 短文テキストのペア (対)  $(x, y)$  ( $x = (u_1 u_2 u_3 u_4 u_5)$ ,  $y = (v_1 v_2 v_3 v_4)$ ,  $(u_i, v_j)$  は語) に対して、語ペア毎のコサイン類似度をそれらのベクトル表現を利用して、語間の  $\cos$  類似度行列を作成する。
- (3) 最後に、語ペア毎の最大類似度を用いて、文  $x, y$  の類似度スコアとする。

語ペアの意味類似度合成法の長所は、文中の語の出現場所に関係無く文同士の類似度が求められる点であり、欠点は、2 文の語の意味ベクトル同士をすべて比較するため、計算量が大きくなることである。

### 3.2 語の意味ベクトル合成法

語の意味ベクトル合成法は、与えられた文に出現する各語の意味ベクトルを加減算することで元の文の意味ベクトルを合成する方法である。Word2Vec では、語の意味ベクトルの加減算が可能であり [1]、本稿で提案する語の意味ベクトル合成法も、この、語の意味ベクトルの加減算を利用したものである。

語の意味ベクトル合成法の手順は以下の通りである：

- (1) 短文テキスト集合を Word2Vec で学習し、語の意味を意味ベクトルとしてを獲得する (各語の意味は、Word2Vec の中間層の値のベクトルとして獲得する)。
- (2) 短文テキスト (文)  $s = w_1 w_2, \dots, w_n$  の意味ベクトル  $\text{meaning}(s)$  を、各語  $w_i$  の意味ベクトル  $\text{meaning}(w_i)$  を加算することで求める (図 2 参照)。

語の意味ベクトル合成法の長所は、計算量を増やすことなく、文中の語の出現場所に関係無く文の意味ベクトルが求められる点である。

## 4 惹句の意味分析と意味類似検索

惹句 (キャッチコピー) <sup>1</sup> とは主に商品・サービスや作品の広告等、何らかの告知や宣伝用に用いられ、謳い文句や煽り文句となる文章のことである。

惹句 (キャッチコピー) の分類は多く行われているが、手法による分類 (ベネフィット型、情報提供型、権威型、問いかけ型、切れ味重視型、対比構造型、常識の逆型、比喩型) <sup>2</sup> や、ジャンル・雰囲気・ターゲットによる分類 (ジャンル (飲食、企業、旅行等)、雰囲気 (印象的、面白い等)、ターゲット (男女、主婦・学生・社会人等) <sup>3</sup> などがある。

### 4.1 惹句の語の意味分析

惹句の意味分析と意味類似検索として、まず、惹句の語の意味分析を分野毎に Word2Vec を用いて行った。具体的には、アルコール飲料 (ビール) に関する惹句、アルコール飲料 (ウイスキー/焼酎) に関する惹句、企業・求人・教育系に関する惹句をそれぞれ別々に収集し、収集した各々の惹句を Word2Vec で機械学習し、語の意味を分析した。語の意味分析にあたっては、分析対象とする語と意味的に近い語を Word2Vec の closest-to 関数を用いて求めた。

語の意味分析の結果としては、「意味的に近い」語としては、類似度は概ね低いものの、次のような特徴が明らかになった。

(1) 「ビール」と意味的に近い単語として「夢」、「愛」、「働く」、「嬉しい」など、若い働く世代に向けたメッセージが見られた (図 3 参照)。

(2) 一方で、ウイスキー/焼酎の惹句においては、「ウイスキー」と意味的に近い単語として「古い」、「甘い」、「大人」など、中高年層の世代に向けたメッセージが見られた (図 4 参照)。

(3) 企業・求人・教育系に関する惹句では、「勉強」と意味的に近い単語として「人生」、「授業」といった類似性の高い言葉が出力されていることから、意味分析が上手くできていると思われる。また、「仕事」と意味的に近い単語の中で「勉強」という言葉に対してマイナスの評価がされていることから、仕事と勉強は類似していないことが分かる (図 5、図 6 参照)。

以上のように単語の意味分析を Word2Vec により行い、意味的に近い単語の出力ができているが、出力結果の意味類似度の値が全体的に低いことが問題点として挙げられる。今後はこの値を少しでも良くなるように、データセットやソースコードの改善に取り組んでいきたい。

### 4.2 惹句の文の意味分析

本節では、惹句の「文」の意味分析について述べる。惹句の文の意味分析を行うために、前節で述べた、語ペアの意味類似度合成法 (BERTScore に基づく意味類似度計算法) を用いた。

「語」の意味分析でも用いたビールの惹句、ウイスキー/焼酎

1 : <https://ja.wikipedia.org/wiki/キャッチコピー>

2 : <https://nakaeshogo.com/catch-copy-samples/>

3 : <https://catchcopy.make1.jp/>

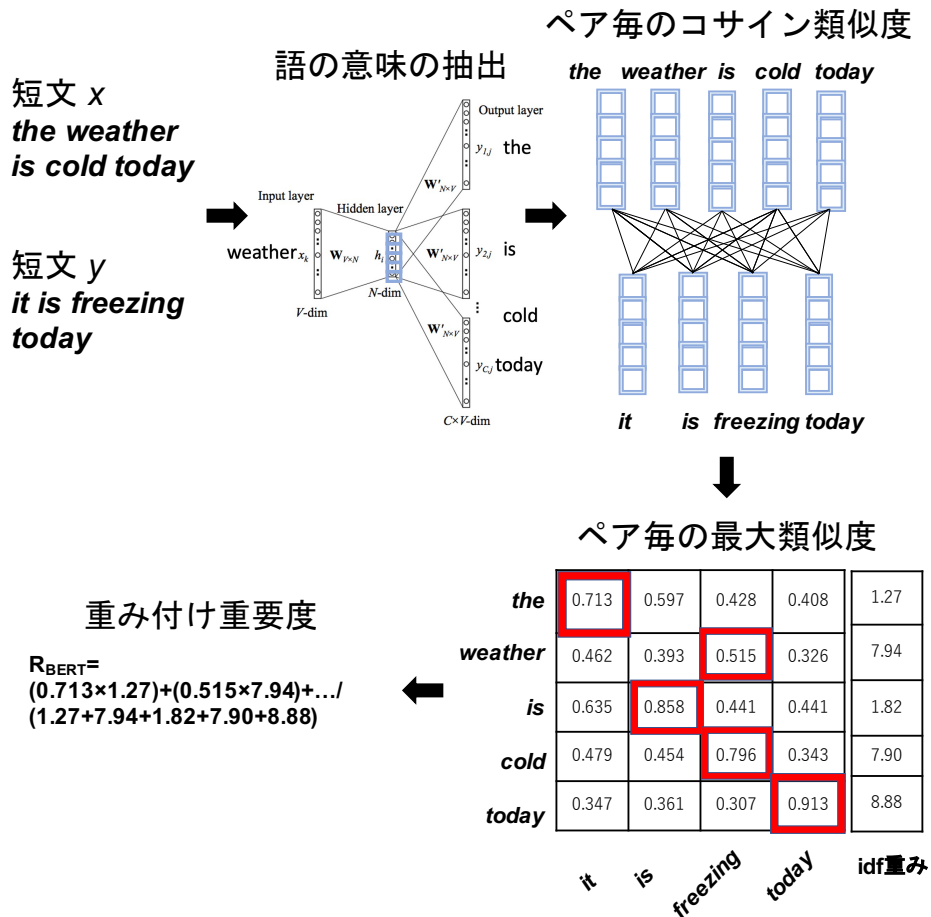


図 1 語ペアの意味類似度合成法による文の類似度計算

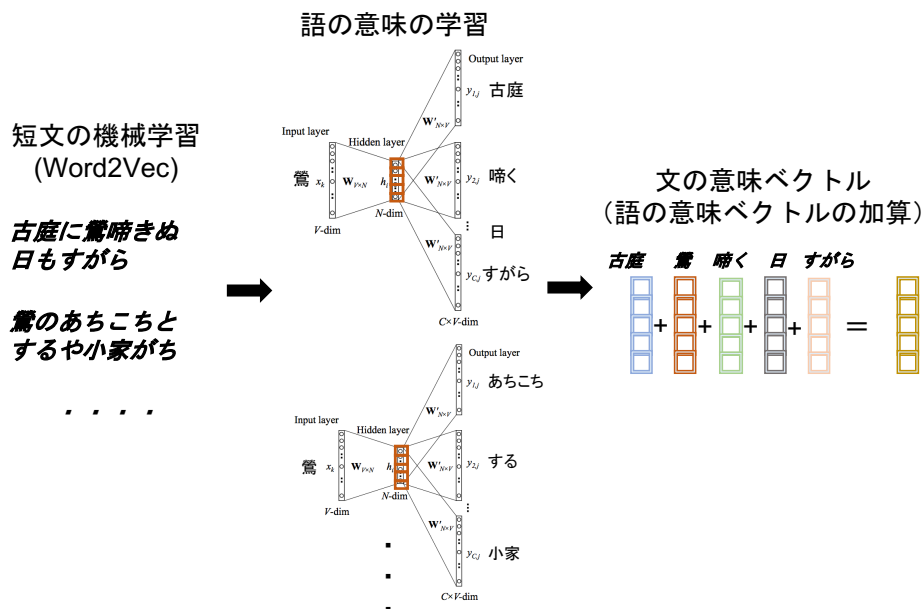


図 2 語の意味ベクトル合成法による文の類似度計算

の惹句、企業・求人・教育系の惹句について、「文」の意味分析を語ペアの意味類似度合成法を用いて行った。その結果は以下の通りである。

図 7 は、ビールに関する各惹句文に対して、意味類似する惹句文とその類似度を出力したものである。図 8 は、ウイスキー/

焼酎に関する各惹句文に対して、意味類似する惹句文とその類似度を出力したものである。図 9 は、企業・求人・教育系の各惹句文に対して、意味類似する惹句文とその類似度を出力したものである。

全体的に、精度良く、「意味類似」した文を出力できているも

```
> ## 「ビール」と意味的に近い語
> model %>% closest_to("ビール", 15)
word similarity to "ビール"
1   ビール      1.00000000
2     ある      0.18340305
3 キリンビール  0.17394367
4   うれしい    0.16854654
5     自慢      0.16353707
6     夢        0.15627759
7     愛        0.15509299
8     目        0.15203598
9     冷える    0.11853255
10    時間      0.10180118
11    働く      0.10173893
12   嬉しい     0.09717958
13    一番      0.09497306
14    いう      0.08594593
15     男       0.08211413
```

図 3 Word2Vec による惹句の語（ビール）の意味分析

```
> ## 「ウイスキー」と意味的に近い語
> model %>% closest_to("ウイスキー", 15)
word similarity to "ウイスキー"
1   ウイスキー  1.00000000
2     時間      0.22267955
3     甘い      0.18571291
4     古い      0.18422132
5     空っぽ    0.17173993
6     焼酎      0.16060487
7     持つ      0.14379578
8     人間      0.13519134
9     大人      0.13329644
10    遊ぶ      0.12673224
11     心       0.11478993
12    いい      0.09857394
13   新しい     0.09422216
14   過ぎる     0.08942732
15     やる     0.08875796
```

図 4 Word2Vec による惹句の語（ウイスキー）の意味分析

```
> ## 「勉強」と意味的に近い語
> model %>% closest_to("勉強", 15)
word similarity to "勉強"
1   勉強      1.00000000
2     人       0.18098324
3   </s>      0.15718161
4   人生      0.11518333
5     ある     0.11376453
6   知性      0.08960722
7   授業      0.05062243
8   残る      0.03276301
9     出る    -0.03842678
10  大人      -0.04778447
11   なる     -0.05300477
12  仕事      -0.06522216
13  自分      -0.09467874
14   する     -0.13147458
15   腹       -0.18615375
```

図 5 Word2Vec による惹句の語（企業・求人・教育系）の意味分析 1

のと考えられるが、この「意味類似」には大きく 2 種類あると考える。

1 つ目は「意味」の類似である。文が類似しているということとは、その文の意味が似ているということであり、最も分かりやすい類似であると考えられる。実際に、ビールの惹句である「おつかれ生です。」と意味的に近いものとして「まごころ冷えてます。」という文が出力されているが、どちらも相手を労い、ビールを飲んでもらうという気持ちが込められた文であり、意味が似ているものと考えられる。その他にも、ウイスキーの惹句である「ワクをハミ出せ、二十歳の可能性。」と「可能性にむかっ

```
> ## 「仕事」と意味的に近い語
> model %>% closest_to("仕事", 15)
word similarity to "仕事"
1   仕事      1.00000000
2   出る      0.236046321
3   </s>      0.121626129
4   する      0.075807571
5   大人      0.059599544
6   人        0.038698137
7   腹        0.012375998
8   人生      0.006043187
9   ある      -0.013469846
10  なる      -0.015803941
11  顔        -0.021663865
12  知性      -0.036416078
13  自分      -0.043003334
14  勉強      -0.065222157
15  残る      -0.148551724
```

図 6 Word2Vec による惹句の語（企業・求人・教育系）の意味分析 2

て、汗をかけ。」という 2 つの文も、「可能性に対して何か行動を起こす」という点において意味が類似していると考えられる。

2 つ目は「手法」の類似である。例えば、ウイスキーの惹句である「歳を重ねるほど、父と重なっていく。」と「離れるほど、私に近づいてくる。」という 2 つの惹句は、「〇〇するほど、～なる。」のように言葉の使い方や手法が似ていると判断することができる。

このように文の類似度計算について実験を行ったところ、「類似している」ということに対して 2 つの種類（文の意味類似と文の手法類似）が存在することが分かった。このことから、ジャンルによって使われている単語や手法は同じものが多いのではないかと予想される。また、本来の BERTScore は生成テキストと正解テキストの類似度を測るものであったが、今回の実験結果は精度が高いことから、任意の 2 つの文の類似度を測ることに応用できる可能性を示したことも本研究の成果と考える。今後は、語ベアの意味類似度合成法を利用して、今回扱わなかったジャンルや手法が使われた惹句についても、意味類似度計算を行っていききたい。

## 5 俳句・短歌の意味分析と意味類似検索

俳句・短歌という短文テキストの意味の抽出を、Word2Vec で計算された語の意味を加減算することで求める「語の意味ベクトル合成法」（前節）を用いて行った。

### 5.1 俳句・短歌データの収集と形態素解析

青空文庫や関連 Web サイトから、次のような実験に用いる俳句・短歌データを収集した。

- 芭蕉発句全集 1066 句（伊藤洋作）<sup>4</sup>
- 一茶発句全集約 2 万句（小林一郎編）<sup>5</sup>
- 和歌と俳句（与謝蕪村）<sup>6</sup>
- みだれ髪（与謝野晶子）（青空文庫）<sup>7</sup>

4 : <https://www2.yamanashiken.ac.jp/~itoyo/basho/haikusyu/Default.htm>

5 : <http://www.janis.or.jp/users/kyodoshi/issaku.htm>

6 : <http://www5c.biglobe.ne.jp/~n32e131/haiku/buson411.html>

7 : [http://www.aozora.gr.jp/cards/000885/files/51307\\_ruby\\_47032.zip1](http://www.aozora.gr.jp/cards/000885/files/51307_ruby_47032.zip1)





```
> model
A VectorSpaceModel object of 1003 words and 100 vectors
      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
</s>  0.004002686  0.004419403 -0.003830261 -0.003278046  0.001366577  0.003021088
する -0.188971728 -0.130228966  0.363412470 -0.203882977 -0.044897571 -0.257167935
花    -0.345784932 -0.182079196 -0.051740151 -0.084464587  0.058473658 -0.201170817
春    -0.199579641  0.225326672 -0.016330827 -0.063737392  0.295237631 -0.114961855
月    -0.350502521  0.049638912 -0.044909962  0.296696246  0.035704248  0.230843857
夜    -0.156838492 -0.131099701  0.125658587 -0.305767447  0.091929294 -0.209837198
秋    -0.147907704  0.031554800  0.368649095  0.369016230 -0.239375383 -0.190454021
ふる -0.178795800  0.309624612  0.326081425  0.096827611  0.015396807  0.012696579
ない -0.054425057 -0.286158442  0.222785905 -0.148025438  0.082069248  0.014903516
見る -0.357508093 -0.207425162  0.023792153  0.086095028  0.135355636 -0.106827073
```

図 12 与謝蕪村の俳句の語の意味の機械学習 (Word2Vec で機械学習した結果 (model))

```
> ## 意味的に近い単語を求める
> model %>% closest_to("花")
word similarity to "花"
1 花 1.00000000
2 よしの 0.8239078
3 柚 0.8118728
4 椎 0.7863986
5 そば 0.7841020
6 いばら 0.7815711
7 すだく 0.7734632
8 蔓 0.7683786
9 桜 0.7644134
10 ふす 0.7618870

> model %>% closest_to(~"五月雨"+"家")
word similarity to "五月雨" + "家"
1 旅 0.9665826
2 床 0.9646886
3 曾我 0.9595311
4 簀 0.9594445
5 鏡 0.9580128
6 零 0.9513634
7 またぐ 0.9500562
8 あける 0.9499454
9 魔 0.9482862
10 家中 0.9464029

> model[["五月雨"]] %>% cosineSimilarity(model[["家"]])
[1,] 0.7167073
```

図 13 与謝蕪村の俳句の機械学習に基づく意味類似語の検索

| "五月雨"の意味類似語 (コサイン類似度)       |                             | "家"の意味類似語 (コサイン類似度)       |                           | "五月雨"+"家"とのコサイン類似度              |                                 |
|-----------------------------|-----------------------------|---------------------------|---------------------------|---------------------------------|---------------------------------|
| model %>% closest_to("五月雨") | model %>% closest_to("五月雨") | model %>% closest_to("家") | model %>% closest_to("家") | model %>% closest_to("五月雨"+"家") | model %>% closest_to("五月雨"+"家") |
| 1 五月雨 1.0000000             | 1 五月雨 1.0000000             | 1 旅 0.9665826             | 1 旅 0.9665826             | 1 五月雨 0.9047742                 | 1 五月雨 0.9047742                 |
| 2 ばた 0.9305354              | 2 骨 0.9261883               | 2 床 0.9646886             | 2 床 0.9646886             | 2 空 0.8658650                   | 2 空 0.8658650                   |
| 3 旅 0.9158569               | 3 簀 0.9261225               | 3 曾我 0.9595311            | 3 曾我 0.9595311            | 3 簀 0.8721785                   | 3 簀 0.8721785                   |
| 4 いそぐ 0.9069318             | 4 魔 0.9252293               | 4 簀 0.9594445             | 4 簀 0.9594445             | 4 鏡 0.8574107                   | 4 鏡 0.8574107                   |
| 5 曾我 0.9059942              | 5 古手 0.9228442              | 5 鏡 0.9580128             | 5 鏡 0.9580128             | 5 降りる 0.8569468                 | 5 降りる 0.8569468                 |
| 6 田舎 0.9040563              | 6 松明 0.9223364              | 6 零 0.9513634             | 6 零 0.9513634             | 6 光 0.8565671                   | 6 光 0.8565671                   |
| 7 飛ぶ 0.9029714              | 7 ふさぐ 0.9215659             | 7 またぐ 0.9500562           | 7 またぐ 0.9500562           | 7 隠れる 0.8509847                 | 7 隠れる 0.8509847                 |
| 1 五月雨 1.0000000             | 1 家 1.0000000               | 1 五月雨 0.9047742           | 1 五月雨 0.9047742           |                                 |                                 |
| 2 空 0.9220809               | 2 簀 0.9303183               | 2 空 0.8658650             | 2 空 0.8658650             |                                 |                                 |
| 3 藤 0.7499267               | 3 住み替る 0.9324174            | 3 簀 0.8721785             | 3 簀 0.8721785             |                                 |                                 |
| 4 望 0.7380657               | 4 いろいろ 0.9193118            | 4 簀 0.8574107             | 4 簀 0.8574107             |                                 |                                 |
| 5 光 0.7331254               | 5 代 0.9141818               | 5 降りる 0.8569468           | 5 降りる 0.8569468           |                                 |                                 |
| 6 桑 0.7189779               | 6 坊 0.8846717               | 6 光 0.8565671             | 6 光 0.8565671             |                                 |                                 |
| 7 日 0.7186242               | 7 重 0.8814018               | 7 隠れる 0.8509847           | 7 隠れる 0.8509847           |                                 |                                 |

図 14 蕪村/芭蕉の俳句の機械学習に基づく意味類似語の検索

語"花", 2語"五月雨+家"に対して意味類似語を検索している。検索結果は, 意味類似語と類似度のペアという形で示されている。また, 2語を与えて, その語間の意味類似度をコサイン類似度で求めることも可能であり, "五月雨"と"家"とのコサイン類似度も出力している。

意味類似語の検索性能は, 機械学習した俳句テキストに依存する点に注意が必要である。例えば, 機械学習した俳句データが与謝蕪村の俳句の場合と芭蕉の俳句の場合では, 意味類似する語が異なる, 図 14 のその 1 例を示す。

### 5.3 俳句・短歌の文の意味分析と意味類似文の検索

3.2 節で提案した「語の意味ベクトル合成法」に基づいて, 俳句・短歌の文の意味分析と意味類似文の検索を行った。俳句・短歌の文の意味分析と意味類似文の検索の手順は以下の通りである:

(1) 短文テキスト集合 (例えば, 与謝蕪村の俳句テキスト集合) を Word2Vec で学習し, 語の意味を意味ベクトルとして獲得する (各語の意味は, Word2Vec の中間層の値のベクトルとして獲得する)。

(2) 各文 (例えば, 1 つの俳句)  $s = w_1 w_2, \dots, w_n$  の意味

上位何件?: 10  
五月雨を集めて早し最上川  
五月雨の降り残してや光堂  
類似度: 0.9922054237530693  
海は晴れて比叡降り残す五月哉  
類似度: 0.9880728919757348  
五月雨や龍頭あぐる番太郎  
類似度: 0.9853719156700066  
五月雨に鶴の足短くなれり  
類似度: 0.9833700530716274  
氷苦く偃鼠が喉をうるほせり  
類似度: 0.981645416174487  
初秋や畳みながらの蚊屋の夜着  
類似度: 0.9803578010232169  
髪生えて容顔青し五月雨  
類似度: 0.978935413207665  
寒菊や粉糠のかかる臼の端  
類似度: 0.977161049466941  
徒歩ならば杖突坂を落馬かな  
類似度: 0.975224317257702  
五月の雨岩檜葉の緑いつまでぞ  
類似度: 0.9751709032537185

図 15 芭蕉の意味類似する俳句の検索

共に, 我が臆病な自尊心と, 尊大な羞恥心との所為である  
しかし, それは臆病な自尊心とでもいうべきものであった  
類似度: 0.9997692029627594  
己の場合, この尊大な羞恥心が猛獣だった  
類似度: 0.9996161884497581  
勿論, 曾ての郷党の鬼才といわれた自分に, 自尊心が無かったとは云わない  
類似度: 0.9994923179340705  
実は, それが殆ど羞恥心に近いものであることを, 人々は知らなかった  
類似度: 0.9994915216646922  
人々は己を倨傲だ, 尊大だといった  
類似度: 0.9994112788997511

図 16 山月記の意味類似する文の検索

ベクトル meaning(s) を, 各語  $w_i$  の意味ベクトル meaning( $w_i$ ) を加算して求める

(3) 文 s の意味ベクトルと類似する意味ベクトルをもつ文  $s'$  を検索する

図 15, 図 16 に, 芭蕉の意味類似する俳句の検索結果, 山月記の文に意味類似する文の検索結果を示す。

## 6 まとめ

本論文では, 惹句 (キャッチコピー), 俳句, 短歌といった短文テキストの「意味」分析を行う手法を提案した。具体的には, 語の意味分析を行う Word2Vec を用いて, 短文の意味分析

を行う2つの手法を提案した。1つ目の手法は、2つの文の意味類似度をそれぞれに出現する各語の意味ベクトルの相互比較に基づいて計算する方法であり、2つ目の手法は、与えられた文に出現する各語の意味ベクトルを加減算することで元の文の意味ベクトルを合成する方法である。これらの手法を惹句、俳句・短歌に適用して、意味が類似する惹句や俳句・短歌を検索する実験を行ったので、その実験結果についても報告した。

今後の課題として次のような事項がある。

- 文の意味分析を行うアルゴリズムの計算時間の改善が必要である。特に、語ペアの意味類似度合成法は入力される文数の二乗に比例する計算量が必要であり、大量のデータを学習する場合には本手法の改善が必要である。
- 惹句（キャッチコピー）文はその表現方法も多様であり、特に、本稿では扱えなかった対比構造型や常識の逆型や比喻型の惹句の意味分析手法の開発が必要である。
- 俳句・短歌の意味の獲得や意味類似検索をおこなうためには、俳句の作法をより考慮して意味分析を行うことが必要である。例えば、ものの対比的配置（構造）と比喻による思いの表現となっている俳句等<sup>8</sup>については、意味の獲得手法の大幅な改良が必要と考えている。

## 謝 辞

本研究は、令和4年度科研費基盤研究(B)(一般)「機械学習による情報の意味獲得と意味類似情報の検索・生成」(代表：田中克己、課題番号22H03905)によるものである。

## 文 献

- [1] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, Vol. 26. Curran Associates, Inc., 2013.
- [2] Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. *Proc. of 31st Int'l. Conf. on Machine Learning*, pp. 1188–1196, 2014.
- [3] Yating Zhang, Adam Jatowt, Sourav S. Bhowmick, and Katsumi Tanaka. The past is not a foreign country: Detecting semantically similar terms across time. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 28, No. 10, pp. 2793–2807, 2016.
- [4] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. Bertscore: Evaluating text generation with bert. In *arXiv preprint arXiv:1904.09675*, 2019.
- [5] 小山大介, 上之園隼志, 杉田祥太郎, 田中克己. 短文テキスト（俳句・短歌・キャッチコピー）の意味分析と意味類似検索. 東海関西データベースワークショップ2022, p. <https://sites.google.com/mil.doshisha.ac.jp/dbws2022>, 2022年9月.

---

8：一茶の次の2句の意味は同じであることを判別できるであろうか：「雀の子 そこのけそこのけ御馬が通る」, 「寝返りをするぞそこのけきりぎりす」