

意味類似度と発音の強弱に着目した韻の検索システムの検討

春原 匠冴[†]東京都立産業技術高等専門学校[†]横井 健[‡]東京都立産業技術高等専門学校[‡]

1 はじめに

韻を用いた文章はリズムがあり聞き心地が良く、印象に残りやすいため、歌詞やキャッチコピーなどにも多く利用されている。近年、単純な母韻の一致だけでなく、発音の工夫によって消える弱い音を利用した韻が使用されることが増えてきている。そのため、既存の韻の検索システムにおいて、意味類似度と発音の強弱を意識した韻の検索が課題となっている。そこで、本研究では発音の強弱と意味に着目して、発音の強弱と意味類似度を考慮した適切な韻の検索システムの開発を目指す。

2 提案手法

ブログ記事「ラップができる AI を作ろう」[1]を参考に以下の手順で韻を検索するシステムを提案する。

1. 国立国語研究所の松下言語学習ラボの日本語を読むための語彙データベース [2] を単語辞書として使用する。
2. 辞書内の単語をすべてカタカナに変換し、さらにカタカナの母音のみに変換する。その際に弱系である「ッ」、「ン」を省略し、長音の「ー」は1つ前の文字と同じ母音に置き換える。
3. 辞書内すべての母音に変換した単語と元の単語を異なる言語の韻検索に活用できる

Universal Sentence Encoder [3] を用いてベクトル化する。

4. 検索対象の単語も同様に母音に変換して2種類のベクトルを作成する。そして、母音のみのベクトルをもとにベクトル近傍探索ライブラリの faiss [4] を用いて、入力された単語と近いベクトルを持つ単語を見つけ出し韻検索を行う。
5. cosine 類似度を用いてベクトルの類似度を求めてから以下の式で韻同士の意味的類似度を考慮し、降順にソートして表示する。

$$\begin{aligned} \text{韻の類似度} &= 0.7 \times \text{母音の類似度} \\ &+ 0.3 \times \text{単語の類似度} \end{aligned}$$

「ラップができる AI を作ろう」では試しに母音の類似度 10 割で検索を行っていた。単語の類似度の比重を重くすると、母音の一致がかなりおろそかになってしまい韻ではない単語が多く検索されてしまう。そのため本研究では上記の式の比重とした。

3 実験

3.1 実験方法

既存の韻検索システム「韻ノート」[5] と本研究で提案したシステムで比較検証を行った。

3.2 実験結果

表1及び表2に既存システム「韻ノート」及び本研究の提案システムが表示した韻の検索結果をそれぞれ上から4件ずつ示す。

提案手法では検索対象単語「食べ物」に対して「タケノコ」や「揚げ物」など食べ物を連想さ

rhyme search system focusing on meaning similarity and pronunciation strength

[†] Shogo Sunohara, Tokyo Metropolitan College of Industrial Technology

[‡] Takeru Yokoi, Tokyo Metropolitan College of Industrial Technology

表1 既存システム (韻ノート)

検索対象単語	検索結果韻 1	検索結果韻 2	検索結果韻 3	検索結果韻 4
食べ物	安全保障	発展途上	カフェイン投与	三点構造
ボールペン	ボールペン	昇竜拳	共通点	小休憩
ホワイトボード	狼男	将来のこと	古代の咆哮	北海道放送

表2 提案システム

検索対象単語	検索結果韻 1	検索結果韻 2	検索結果韻 3	検索結果韻 4
食べ物	食べ物	タケノコ	揚げ物	タテヨコ
ボールペン	おおむね	底抜け	横付け	外付け
ホワイトボード	ホワイトボード	テキーラ	ニュートロスタット	マークウィス

せる言葉が優先的に検索されたが、既存システムでは「安全保障」や「発展途上」など食べ物とは関係ない言葉が検索されたことが分かる。

母音に着目すると、提案手法では「アエオオ」の完全一致のみ検索されたが、既存システムでは「アエオオ」に対して「アンエンオオー」や「アッエンオオー」、「アエインオウオ」などの本研究では無視してしまった「ン」、「ッ」について、さらに「カフェイン」を「カフェーン」と崩した発音ができることを想定した韻が検索されたことが分かる。

4 考察

提案手法では検索対象単語に対して意味の近い単語を検索するため、韻を用いた文章を作成しやすい単語を検索できる。しかし、単純な母音の一致のみ表示してしまうため弱い音など実際の発音の工夫によって消える音を考慮した検索ができないことが分かる。「ホワイトボード」など長い言葉では近い言葉が検索されなかったり、母音が全く一致していなかったりと検索元となる単語辞書にも課題があることが分かる。

5 まとめ

本研究は意味類似度と発音の強弱に着目した韻の検索システムの開発を行った。意味の近い

単語を優先的に検索することによって既存システムよりも文章がつながりやすい言葉を検索することに成功した。しかし発音の強弱について実際の発音において消える音を意識した検索が課題となった。

参考文献

- [1] wakamoto ryosuke. ラップができる ai を作ろう part.1 ~ 韻を検索する ~. https://tech.dentsusoken.com/entry/rapgpt_part1. 参照 Oct 20, 2023.
- [2] 松下達彦. 松下言語学習ラボ 日本語を読むための語彙データベース ver. 1.11. <http://www17408ui.sakura.ne.jp/tatsum/database.html>. 参照 Sep 22, 2023.
- [3] Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, et al. Universal sentence encoder. *arXiv preprint arXiv:1803.11175*, 2018.
- [4] Jeff Johnson, Matthijs Douze, and Hervé Jégou. Billion-scale similarity search with gpus. *IEEE Transactions on Big Data*, 7(3):535–547, 2019.
- [5] 株式会社スリーアウトチェンジ. 韻ノート. <https://in-note.com/>. 参照 Jan 12, 2024.