Word2Vecとスマートスピーカーを用いた単語連想システムの検討

新田　洸平　　岩佐　英彦
近畿大学工業高等専門学校 総合システム工学科

1.はじめに

我々は、スマートスピーカーにある単語を話しかけると連想される単語を返す単語連想システムの構築を目指している。新田らは、性質の異なる複数のコーパスから Word2Vec[2]のモデルを複数作成することよって様々な観点から類似した単語を求められることを示した[1]。しかしこの手法では、求められる類似語は与えた単語と同じ品詞のものばかりになり意外性に欠ける傾向があるという問題がある。

本稿では単語の品詞を考慮し、意外性のある単語連想を行える仕組みについて報告する。また、この単語連想のシステムをスマートスピーカーを用いて実現するための仕組みを示す。

2.手法と実験内容

インターネットから取得できる Wikipedia のテキストコーパス[3]を用いて、Word2Vec の学習モデルの構築を行った。

このモデルに対して、名詞：犬、動詞：歩く、形容詞：美しいを与え、それぞれ類似度の高い順に単語を出力し、それを品詞ごとに分類したものを結果として得た。

3.実験結果及び検討

Wikipediaコーパスから学習したモデルに単語を与えたとき、単に類似度の高い順に出力した場合の結果と、その結果を形態素解析ツールMeCab[4]によって品詞を分類して出力した場合の結果を下記に示す。

[入力単語：犬]

・類似度順

犬 = オオカミ, 柴犬, 猫, キツネ, 猟犬, 豚, オルトロス, パグ, ウサギ, イヌ

・品詞別の類似度順

名）オオカミ, 柴犬, キツネ, 猟犬, パグ
動）馴らす, 捕る, 飼う, 繕う, 舐める
形）かしこい

[入力単語：歩く]

・類似度順

歩く = 走る, 歩き回る, 駆け抜ける, 走り抜ける, 走り回る, 駆け回る, 楽しむ, 眺める, 登る, 見る

・品詞別の類似度順

名）山歩き, 夜道, 雪景色, 畦道, 練り歩き
動）走る,歩き回る,駆け抜ける, 走り抜ける, 走り回る
形）薄暗い, 険しい, 楽しい, 狭い, 懐かしい

[入力単語：美しい]

・類似度順

美しい = 美しく, すばらしい, 素晴らしい, 醜い, 気高く,心地よく,愛らしい, 暗い, 麗しい, 小さな

・品詞別の類似度順

名）野趣, 裸身, 貴婦人, 雪景色, 睡蓮
動）眠れる, 輝く, 映える, 戯れ, 色づく
形）美しく,すばらしい,素晴らしい,醜い,気高く

この結果より、類似度だけの場合には入力語と同じ品詞の単語に偏っているのに対して、品詞を分けることにより入力語との関連性を保ちながら、意外性のある単語が出力できることが確認できた。

4.スマートスピーカーを用いた単語連想の仕組

Google 社の AIY Voice Kit を用いて構築する対話システムを図1に示す。システムは、音声認識APIを用いて発話文を認識し、その結果を単語連想システムへ入力し、得られた結果を発話APIに与えるという流れで構成されている。

図 1 単語連想システムの構成図

5.今後の課題

本稿で提案した品詞を考慮して類似語を出力する仕組みと、複数のモデルを用いて類似語を出力する仕組みを組合せて、多様な観点から意外性のある類似語を出力できるシステムを構築する予定である。また、スマートスピーカーを用いた対話型のシステムを構築する。

参考文献

[1] 新田洸平, 岩佐英彦：複数の Word2Vec モデルを用いた単語連想システムの検討,第23回高専シンポジウムin神戸講演要旨集(2018).
[2]T.Mikolov, K.Chen, G.Corrado, and J.Dean.: Efficient Estimation of Word Representations in Vector Space. In ICLR Workshop.(2013).
[3]https://dumps.wikimedia.org/jawiki/latest/
[4]T.Kudo, K.Yamamoto, Yuji Matsumoto: Applying Conditional Random Fields to Japanese Morphological Analysis, Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), pp.230-237 (2004.)

Copyright © 2018 IEICE