

情報工学実験 II 課題 Word2Vec

1 課題 1

vector size を 3 から 10 に変更して実行した。また、選択した単語は "maps", "think", "tired" であり, CBoW, skip-gram の 2 つにおいて学習を行った。さらに, "country + rabbit - nothing - cats" としてこの結果を示した。

1.1 CBoW

表 1, 2, 3, 4 に CBoW における上記の単語に似ている単語の情報と分散表現ベクトルの加減算の結果をそれぞれ示す。

表 1: "maps" に似ている単語

順位	1	2	3	4	5
単語	curtsey	end	theyll	dreamy	tired
類似度	0.819	0.662	0.656	0.609	0.604

表 2: "think" に似ている単語

順位	1	2	3	4	5
単語	saying	with	alice	on	said
類似度	0.852	0.795	0.772	0.768	0.756

表 3: "tired" に似ている単語

順位	1	2	3	4	5
単語	anything	pegs	stupid	heap	flashed
類似度	0.828	0.826	0.754	0.737	0.691

1.2 skip-gram

表 5, 6, 7, 8 に skip-gram における上記の単語に似ている単語の情報と分散表現ベクトルの加減算の結果をそれぞれ示す。

表 4: "country + rabbit - nothing - cats" の結果

順位	1	2	3	4	5
単語	ought	watch	air	passed	im
類似度	0.819	0.727	0.725	0.7024	0.7017

表 5: "maps" に似ている単語

順位	1	2	3	4	5
単語	curtsey	end	theyll	dreamy	tired
類似度	0.819	0.662	0.656	0.609	0.604

表 6: "think" に似ている単語

順位	1	2	3	4	5
単語	saying	with	alice	on	said
類似度	0.852	0.795	0.772	0.768	0.756

表 7: "tired" に似ている単語

順位	1	2	3	4	5
単語	saying	with	alice	on	said
類似度	0.852	0.795	0.772	0.768	0.756

表 8: "country + rabbit - nothing - cats" の結果

順位	1	2	3	4	5
単語	ought	watch	air	passed	im
類似度	0.819	0.727	0.725	0.7024	0.7017

2 課題 2

2.1 実行結果

表 9, 10, 11 に ”船” ”泳ぐ” ”深い” の単語に似ている単語の情報をそれぞれ示す.

表 9: ”船” の結果

順位	1	2	3	4	5
単語	停泊	出航	貨物船	出港	乗船
類似度	0.785	0.772	0.771	0.745	0.739

表 10: ”泳ぐ” の結果

順位	1	2	3	4	5
単語	泳ぎ	泳ぎ出す	泳ぎ回る	立ち泳ぎ	遊泳
類似度	0.830	0.775	0.708	0.685	0.667

表 11: ”深い” の結果

順位	1	2	3	4	5
単語	奥深い	浅い	掘り下げる	深み	造詣
類似度	0.682	0.612	0.572	0.554	0.541

2.2 考察

日本語において, その単語を含むものや同じ部首, 似た意味の部首をもつ単語が上位の類似単語に選択されていると考えられる. また, 意味的に似ている単語も選択されているため, 意味解釈も行われていると考えられる.

さらに, 課題の英語の場合において, 日本語と比べると正確な類似度が得られていない. これは, 英語の学習に用いた単語数が少なかったためであると考えられる.