

nl-4. Word2Vec の例

(自然言語処理入門)

URL: <https://www.kkaneko.jp/ai/nlp/index.html>

金子邦彦



謝辞



- 次のサイトの記事, ソースコード, データを使用しています

<https://aial.shiroyagi.co.jp/2017/02/japanese-word2vec-model-builder/>

- コマンドで, データのダウンロードと展開

```
wget http://public.shiroyagi.s3.amazonaws.com/latest-ja-word2vec-gensim-model.zip
```

```
7z x latest-ja-word2vec-gensim-model.zip
```

- プログラム

```
from gensim.models.word2vec import Word2Vec
```

```
mpath = 'word2vec.gensim.model'
```

```
m = Word2Vec.load(mpath)
```

```
def sim(w):
```

```
    for i in m.wv.most_similar(positive=[w]):
```

```
        print(i)
```

```
sim('楽しみ')
```

```
In [1]: from gensim.models.word2vec import Word2Vec
...: mpath = 'word2vec.gensim.model'
...: m = Word2Vec.load(mpath)
...: def sim(w):
...:     for i in m.wv.most_similar(positive=[w]):
...:         print(i)
...:
...: sim('楽しみ')
```

('楽しく', 0.8377718925476074)
('楽しい', 0.8144588470458984)
('遊び', 0.8099892139434814)
('気軽', 0.797997236251831)
('お客さん', 0.7635385394096375)
('自慢', 0.7526886463165283)
('笑顔', 0.7431896328926086)
('のんびり', 0.7385178804397583)
('大切', 0.7302102446556091)
('お喋り', 0.729663610458374)

```
In [2]:
```

```
In [3]: sim('晴れ')
('雨', 0.8238000869750977)
('小雨', 0.7920332551002502)
('快晴', 0.7777684926986694)
('降る', 0.766666054725647)
('晴天', 0.7512480020523071)
('寒い', 0.7504522800445557)
('晴れる', 0.7483686208724976)
('曇天', 0.7421630024909973)
('曇り', 0.7403557300567627)
('真冬', 0.7319853901863098)
```

```
In [4]:
```

```
In [4]: sim('徳川家康')
('豊臣秀吉', 0.9480844140052795)
('織田信長', 0.92942875623703)
('毛利輝元', 0.9221905469894409)
('伊達政宗', 0.9190208911895752)
('政宗', 0.9181203842163086)
('上杉景勝', 0.9160507917404175)
('秀吉', 0.9116554260253906)
('徳川氏', 0.9086129069328308)
('今川義元', 0.9033724069595337)
('信玄', 0.8999422788619995)
```

```
In [5]:
```

```
In [3]: sim('晴れ')  
( '雨', 0.8238000869750977 )  
( '小雨', 0.7920332551002502 )  
( '快晴', 0.7777684926986694 )  
( '降る', 0.766666054725647 )  
( '晴天', 0.7512480020523071 )  
( '寒い', 0.7504522800445557 )  
( '晴れる', 0.7483686208724976 )  
( '曇天', 0.7421630024909973 )  
( '曇り', 0.7403557300567627 )  
( '真冬', 0.7319853901863098 )
```

```
In [4]:
```

Word2Vec を用いて「晴れ」に意味の似ているものを検索