

October 18, 2019

# 環境・生体ライフログからの 行動パターン解析における 分類精度の向上

沼田 賢一

富山県立大学 情報基盤工学講座

1. はじめに
2. セクション名
3. おわりに

October 18, 2019

## 前回の課題

英語版 wikipedia の word2vec トレーニング済みデータの読み込みでのエラーが解決できなかった

## 課題

ラズベリーパイ側のセンサーやカメラなどのデータを PC に送信するファイルが、3 か月前にファイルを受け継いだ時のデータに勝手に書き換えられていた。

## 解決

3 か月前のラズベリーパイのデータは生体データの取得、送信だけだったので 5 秒ごとにカメラで画像取得、画像認識 API を通してテキストの取得、生体データと合わせて PC に送信の部分をなんとか元に戻した。(なんで、データが書き換わったのか結局原因不明。)

```
File Edit View Window Help
File 'tensorflow.py', line 66, in main
client_session.run(data=[sonoma])
File 'tensorflow.py', line 238, in meth
return sess.run(self._sess.graph.get_operation_by_name('data'))
AttributeError: 'SonomaData' object has no attribute 'save'

D:\Vagrant\Python\env\python server2.py
available to connect: 192.168.137.1
192.168.137.1 5023 1023.86, "1022", "1", "8.52", "4.38", "-2.62", "-0.24", "-0.59", "0.29", "9", "-253", "9", "-51.7",
192.168.137.1 5023 1023.86, "1022", "1", "8.51", "4.37", "-2.61", "-0.84", "-1.33", "-0.28", "7", "-252", "9", "-49.0",
192.168.137.1 5023 1023.86, "1022", "1", "8.52", "4.37", "-2.58", "-0.52", "-0.84", "-0.08", "8", "-252", "10", "-7.2",
192.168.137.1 5023 1023.86, "1022", "1", "8.52", "4.38", "-2.55", "-0.01", "-0.31", "9", "-247", "9", "-37.7",
192.168.137.1 5023 1023.80, "1021", "1", "8.52", "4.38", "-2.98", "0.01", "-0.10", "0.88", "9", "-248", "9", "-37.7",
192.168.137.1 5023 1023.86, "1021", "1", "8.52", "4.37", "-2.62", "0.02", "0.12", "-0.03", "7", "-255", "9", "-252.0",
192.168.137.1 5023 1023.80, "1022", "1", "8.52", "4.38", "-2.60", "0.41", "0.12", "0.73", "2", "-482", "9", "-38.9",
192.168.137.1 5023 1023.80, "1022", "1", "8.52", "4.37", "-2.57", "-0.16", "-0.54", "0.02", "9", "-252", "9", "-41.5",
192.168.137.1 5023 1023.77, "1021", "1", "8.52", "4.37", "-2.96", "-0.47", "-0.28", "0.20", "6", "-251", "9", "-36.6",
192.168.137.1 5023 1023.85, "1022", "1", "8.51", "4.37", "-2.56", "1.27", "0.11", "0.72", "6", "-246", "9", "-51.7",
192.168.137.1 5023 1023.86, "1022", "1", "8.52", "4.38", "-2.59", "0.53", "0.71", "-0.13", "5", "-252", "10", "-37.7",
D:\Vagrant\Python\env\python server.py:82: UserWarning: Attempting to set identical bottom == top == 1.0 results in singular
transformations: automatically expanding
File "server.py", line 1023, in __init__
self._sess.run(self._sess.graph.get_operation_by_name('data'))
```

## 未解決

wikipedia のモデルの読み込みエラーは解決できたが、もっていた英語 wikipedia のモデル (Nordic Language Processing Laboratory から) では前処理として、単語とセットで PoS-tagged という品詞のタグがつけられているので実行するときに単語に品詞の区別をしてから実行しないといけない。

```

15
16 #import gensim.models.keyedvectors as word2vec
17 # embed_map = word2vec.KeyedVectors.load_word2vec_format("./word2modelwiki/model.
18 # #print (embed_map["hello"])
19
20 #import gensim.models
21 # #print (embed_map["hello_NOUN"])
22 # #model = "./word2modelwiki/model.txt"
23
24 # #word_vectors = gensim.models.KeyedVectors.load_word2vec_format(model, binary=False)
25 # #print(word_vectors.most_similar("vacation_NOUN"))
26 # #print (word_vectors["vacation"])
27
28 # # import gensim.models.keyedvectors as word2vec
29 # # embed_map = word2vec.KeyedVectors.load_word2vec_format("./word2modelwiki/model.
30 # # print (embed_map["hello_NOUN"])
31
32 # import gensim
33
34 # # Load Google's pre-trained Word2Vec model.
35 # # model = gensim.models.KeyedVectors.load_word2vec_format('./word2modelwiki/model.
36 # # print(model["vacation_NOUN"])
37
38 import gensim
39 model = gensim.models.KeyedVectors.load_word2vec_format('./word2modelwiki/model.bi
40 #print(model["vacation_NOUN"])
41 print(model.similarity('woman_NOUN', 'man_NOUN'))

```

```

C:\Windows\System32\cmd.exe
TypeError: load() got an unexpected keyword argument 'binary'
D:\ダウンロード\行動認識\python practice-word2vec.py
Traceback (most recent call last):
  File "practice-word2vec.py", line 24, in <module>
    #word_vectors = gensim.models.KeyedVectors.load_word2vec
  File "C:\Users\theta\AppData\Local\Programs\Python\Python37
1498, in load_word2vec_format
    limit=limit, datatype=datatype)
  File "C:\Users\theta\AppData\Local\Programs\Python\Python37
395, in _load_word2vec_format
    word, weights = parts[0], [datatype(x) for x in parts[1:]
  File "C:\Users\theta\AppData\Local\Programs\Python\Python37
395, in <listcomp>
    word, weights = parts[0], [datatype(x) for x in parts[1:]
KeyboardInterrupt
D:\ダウンロード\行動認識\python practice-word2vec.py
Traceback (most recent call last):
  File "practice-word2vec.py", line 36, in <module>
    print(model["vacation_NOUN"])
TypeError: 'Word2VecKeyedVectors' object is not callable
D:\ダウンロード\行動認識\python practice-word2vec.py
0.72789985
D:\ダウンロード\行動認識\python practice-word2vec.py
0.72789985
D:\ダウンロード\行動認識

```

## 未解決

次に、小野田さんからもらった英語 Googlenews からのモデルでやってみたが、`ValueError: array is too big; 'arr.size * arr.dtype.itemsize' is larger than the maximum possible size.` というエラーが出た。(調べたらメモリ不足が原因のエラー)

```
D:\ダウンロード\行動認識>python practice-word2vec.py
Traceback (most recent call last):
  File "practice-word2vec.py", line 39, in <module>
    model = gensim.models.KeyedVectors.load_word2vec_format('./GoogleNews-vectors-negative300/GoogleNews-vectors-negative300.bin', binary=True)
  File "C:\Users\theta\AppData\Local\Programs\Python\Python37-32\lib\site-packages\gensim\models\keyedvectors.py", line 1498, in load_word2vec_format
    limit=limit, datatype=datatype)
  File "C:\Users\theta\AppData\Local\Programs\Python\Python37-32\lib\site-packages\gensim\models\utils_arv2vec.py", line 349, in load_word2vec_format
    result_vectors = zeros((vocab_size, vector_size), dtype=datatype)
ValueError: array is too big; 'arr.size * arr.dtype.itemsize' is larger than the maximum possible size.
```

## 未解決

麻生さんと相談して自分でモデルを作ることになりましたが、wikipedia の圧縮ファイルが重すぎて解凍できない。



## 未解決

他の wikipedia のモデルも探していて、新しく見つけたモデルで実行してみたがボキャブラリーが少なくて使い物にならない。

```
D:\ダウンロード\行動認識>python practice-word2vec.py
Traceback (most recent call last):
  File "practice-word2vec.py", line 26, in <module>
    print(word_vectors['vacation'])
  File "C:\Users\theta\AppData\Local\Programs\Python\Python37-32\Lib\site-packages\gensim\models\keyedvectors.py", line
353, in __getitem__
    return self.get_vector(entities)
  File "C:\Users\theta\AppData\Local\Programs\Python\Python37-32\Lib\site-packages\gensim\models\keyedvectors.py", line
471, in get_vector
    return self.word_vec(word)
  File "C:\Users\theta\AppData\Local\Programs\Python\Python37-32\Lib\site-packages\gensim\models\keyedvectors.py", line
468, in word_vec
    raise KeyError("word '%s' not in vocabulary" % word)
KeyError: 'word 'vacation' not in vocabulary'
```

## まとめ

- ① 今回のようにデータが消えたりすることもあるのでラズベリーのファイルも含めて全部のデータをバックアップした。
- ② 他の word2vec のモデルを探したり作ろうとしたりしたがどれもうまくいかない

## 今後の課題

- 1 中間発表までに時間がないので今のところ唯一できている POS タグ付きのモデルで品詞の区別の処理をしてベクトル化をするか検討する。