はじめに
進捗
おわりに

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

はじめに

はじめに
進捗
おわりに

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

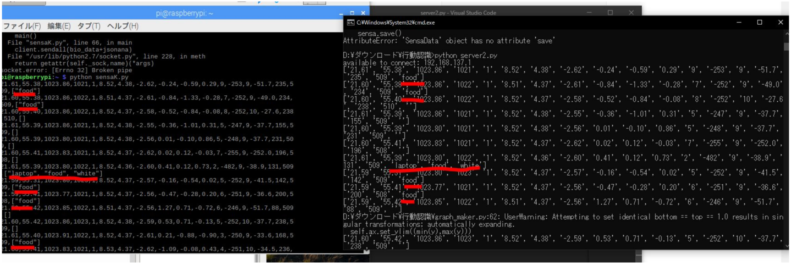
課題

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。

進捗

解決

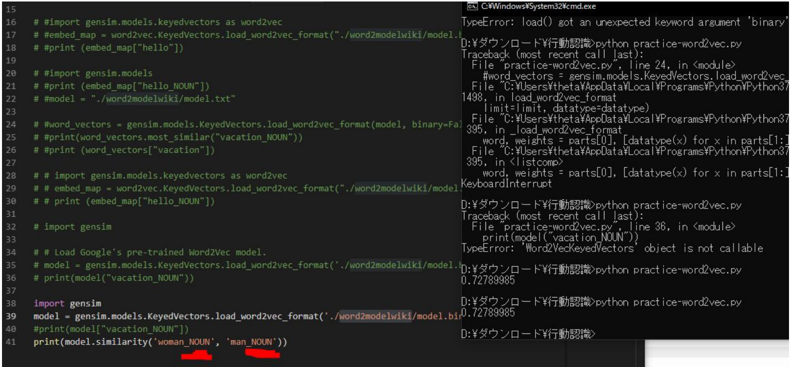
3 か月前のラズベリーパイのデータは生体データの取得、送信だけだったので 5 秒ごとにカメラで画像取得、画像認識 API を通してテキストの取得、生体データと合わせて PC に送信の部分をなんとか元に戻した。（なんで、データが書き換わったのか結局原因不明。）



進捗

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。



進捗

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-negativ
e300.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_any2vec.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.
```

進捗

はじめに
進捗
おわりに

未解決

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

進捗

はじめに
進捗
おわりに

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"
```

おわりに

はじめに
進捗
おわりに

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パイ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。