

## 基于 gensim 包训练 NLP 分词模型

[https://github.com/Jennifer1996/Projects/blob/master/codes/test\\_gensimmodels.py](https://github.com/Jennifer1996/Projects/blob/master/codes/test_gensimmodels.py)

1. 目的：熟悉 NLP WordEmbedding 的概念和实现
2. 方式：基于 gensim 包和中文维基语料
3. 运行结果：训练模型后生成

```
word_embedding_128
word_embedding_128.trainables.syn1neg.npy
word_embedding_128.wv.vectors.npy
```

使用模型验证“男人”与“女人”，“人”与“狗”的相似度，分别如下：

```
20 # 使用模型
21 #items = model.most_similar(u'中国')
22 # for item in items:
23 #     print item[0],item[1]
24
25 print model.similarity(u'男人', u'女人')
26 print model.similarity(u'人', u'狗')
```

```
C:\Users\1\Anaconda2\python.exe D:/123/Data/codes/testgensimmodels.py
0.90379655157
0.291435660498
Process finished with exit code 0
```