

python实现word2vec训练结果bin文件转txt文件

经理让我把word2vec训练后得到的bin文件转为txt文件，目前还不知道txt文件用来干什么。其实word2vec训练语料时可以选择训练出bin文件或者txt文件，但是训练出bin文件时过程太漫长，我怕直接训练出txt文件也一样慢，所以还是自己想办法做这个事情了。

我用到了gensim，这个需要自己安装一下，我的电脑装这个还挺麻烦的。



```
# -*- coding: utf-8 -*-
import gensim
import codecs

def main():
    path_to_model = 'result.bin'
    output_file = 'file.txt'
    bin2txt(path_to_model, output_file)

def bin2txt(path_to_model, output_file):
    output = codecs.open(output_file, 'w', 'utf-8')
    model = gensim.models.KeyedVectors.load_word2vec_format(path_to_model, binary=True)
    print('Done loading Word2Vec!')
    vocab = model.vocab
    for item in vocab:
        vector = list()
        for dimension in model[item]:
            vector.append(str(dimension))
        vector_str = ",".join(vector)
```

公告

昵称：[右介](#)
园龄：[10个月](#)
粉丝：[2](#)
关注：[2](#)
[+加关注](#)

| | | | | | | |
|--------------|----|----|----|-------------------|----|----|
| < 2017年12月 > | | | | | | |
| 日 | 一 | 二 | 三 | 四 | 五 | 六 |
| 26 | 27 | 28 | 29 | 30 | 1 | 2 |
| 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 17 | 18 | 19 | 20 | 21 | 22 | 23 |
| 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| 31 | 1 | 2 | 3 | 4 | 5 | 6 |

搜索

找找看

谷歌搜索

常用链接

[我的随笔](#)
[我的评论](#)
[我的参与](#)
[最新评论](#)
[我的标签](#)

随笔分类

```
line = item + "\t" + vector_str
output.writelines(line + "\n") #本来用的是write()方法，但是结果出来换行效果不对。改成
writelines()方法后还没试过。
output.close()

if __name__ == "__main__":
    main()
```

分类: [Python](#)

好文要顶

关注我

收藏该文



右介

关注 - 2

粉丝 - 2

0

0

[+加关注](#)[« 上一篇：非常不错停用词表](#)[» 下一篇：gensim加载word2vec训练结果\(bin文件\)并进行相似度实验](#)posted @ 2017-05-31 10:13 [右介](#) 阅读(754) 评论(0) [编辑](#) [收藏](#)[刷新评论](#) [刷新页面](#) [返回顶部](#)注册用户登录后才能发表评论，请 [登录](#) 或 [注册](#)，[访问](#)网站首页。[【推荐】50万行VC++源码: 大型组态工控、电力仿真CAD与GIS源码库](#)[【促销】腾讯云技术升级10大核心产品年终让利](#)[C#\(1\)](#)[Kafka\(7\)](#)[Linux\(9\)](#)[MongoDB\(5\)](#)[MySQL\(7\)](#)[Neo4j\(1\)](#)[NLP\(6\)](#)[Python\(32\)](#)[Scala\(4\)](#)[Scrapy\(4\)](#)[Solr\(9\)](#)[Spark\(5\)](#)

随笔档案

[2017年12月 \(2\)](#)[2017年11月 \(1\)](#)[2017年10月 \(14\)](#)[2017年8月 \(6\)](#)[2017年7月 \(6\)](#)[2017年6月 \(20\)](#)[2017年5月 \(23\)](#)[2017年2月 \(1\)](#)

最新评论

1. Re:python调用jieba(结巴)分词 加入自定义词典和去停用词功能
@NER当然不是自己意识判断，如果一个词在原始分词词典中不存在，那么将永远不会在分词结果中出现，我认为这时就应该人为添加了。但是添加后分词工具也不一定每次只要遇到“云计算”三个字就会把他分出来，毕竟三.....

--右介

2. Re:python调用jieba(结巴)分词 加入自定义词典和去停用词功能

谢谢你的同看 我是新手 但是还想

【推荐】高性能云服务器2折起，0.73元/日节省80%运维成本

【新闻】H3 BPM体验平台全面上线



葡萄城报表
千万种报表 同一种选择

在线设计 报表
数据价值即刻体现

立即了解

最新IT新闻:

- 永辉超市：未来12个月内 腾讯无继续增持计划
 - 拓宽AR的娱乐价值，Snapchat推出首款桌面应用「Lens Studio」
 - 腾讯发现者揭秘：怎么应对TensorFlow的安全风险，修复有多难
 - 怼周鸿祎92女生再发声：我愿意跟周教主面对面对话
 - 「中国特斯拉」蔚来ES8的「汽车重新定义」和它的iPhone时刻
- » 更多新闻...



阿里云 告别高昂运维费用 云计算全面助力
40+款核心产品免费半年 再+8000津贴任意采购

立即申请

最新知识库文章:

- 以操作系统的角度述说线程与进程
- 软件测试转型之路
- 门内门外看招聘
- 大道至简，职场上做人做事做管理
- 关于编程，你的练习是不是有效的？

谢谢你的回复。我走利了。但是还是问问。这种结果好坏是根据自己意识来判断的？云计算应该分在一起还是应该“云”和“计算”？

--NER

3. Re:python调用jieba(结巴)分词 加入自定义词典和去停用词功能

@NER创建一个txt文件，把原本无法通过常规分词的词汇放在里面，有没有固定的文本结构我记不清了。然后用jieba.load_userdict("自定义词典.txt")语句加载。这种方式只适用于：“云.....

--右介

4. Re:python调用jieba(结巴)分词 加入自定义词典和去停用词功能

请问楼主，最后像云计算这样分词不准确的问题怎么解决的？

--NER

阅读排行榜

1. python调用jieba(结巴)分词 加入自定义词典和去停用词功能(12331)
2. python实现将文件夹内所有txt文件合并成一个文件(2182)
3. python连接mysql并插入数据(1523)
4. Python连接mongodb提取部分字段内数据并写入txt文件(1277)
5. MySQL实现删除数据左右空格 trim() 左空格ltrim() 右空格rtrim() (1226)

评论排行榜

1. python调用jieba(结巴)分词 加入自定义词典和去停用词功能(4)

推荐排行榜

» [更多知识库文章...](#)

1. [python调用jieba\(结巴\)分词 加入自定义词典和去停用词功能\(1\)](#)

Copyright ©2017 右介