



# PowerAI 人工智能马拉松编程大赛

Hackathon for Finance Industry

## 作品展示-第十三组

( 大赛三等奖 )

**CSDN**

Power Systems



# 作品展示

## 作品展示（算法/代码等）

```
BASE_DIR = 'D:/python36/Scripts/cnn' # 这里是指当前目录
MODEL_DIR = BASE_DIR + '/gensim_model/20170626' # model路径
TEXT_DATA_DIR = 'D:/python36/Scripts/cnn/data_fenci_label' # 输入路径
TRAIN_TEXT_DATA_DIR = 'D:/python36/Scripts/cnn/train_test_data_fenci' # 输入路径
TEXT_TEST_DATA_DIR = 'D:/python36/Scripts/cnn/train_test_data_fenci' # 输入路径
MAX_SEQUENCE_LENGTH = 600 # 每个文
MAX_NB_WORDS = 5000 # 整体词库字
EMBEDDING_DIM = 100 # 词向量的维
VALIDATION_SPLIT = 0.1 # 这里用作
print('Indexing word vectors.')
new_model = gensim.models.Word2Vec
print(new_model)
print(new_model.similarity('股票', '期货'))
# second, prepare text samples and
print('Processing text dataset')
texts = [] # 存储训练样本的list
labels_index = {} # 词到词编号
labels = [] # 存储训练样本, 类别
catalog_dict = {0: '基金', 1: '车险', 2: '意外险', 3: '家庭财产保险', 4: '理财', 5: '其他', 6: '家庭财产保险', 7: '理财', 8: '其他'}
#catalog_dict = {0: '肯定', 1: '否'}
#1 for intension
#catalog_dict, file_list, texts, labels_index, labels
(TEXT_DATA_DIR, 1, catalog_dict)
catalog_dict, file_list, texts, labels_index, labels
(TEXT_DATA_DIR, 0, catalog_dict)
file_list_test, texts_test = load_data(TEXT_TEST_DATA_DIR)
file_list_test_train, texts_test_train = load_data(TRAIN_TEXT_DATA_DIR)
labels_index=catalog_dict.keys()
```

```
encoding=utf-8
import glob
import os
import sys
import jieba
import os
from gensim.models import word2vec
# 词向量词典
import gensim
class MySentences(object):
    def __init__(self, dirname):
        self.dirname = dirname
    def __iter__(self):
        for fname in os.listdir(self.dirname):
            for line in open(os.path.join(self.dirname, fname), encoding='utf-8'):
                yield line.split()
sentences = MySentences('D:/python36/Scripts/cnn/data_fenci/categoryall') # a memory-friendly iterator
model = gensim.models.Word2Vec(sentences, size=100, window=5)
model.save('D:/python36/Scripts/cnn/20170626')
model.similarity('股票', '期货')
```

## 作品描述

实现思路：  
我们使用jieba分词，gensim训练词向量词典，利用神经网络预测模型（卷积神经网络特征提取），使用了python开发语言，keras+tensorflow深度学习框架，以及GPU计算。

## 个人评价



### ◆ 对提交作品评价：

由于前期我们在深度学习领域技术与理论储备较少，我们对本次比赛提交的作品不是很满意，我们使用了cnn训练了预测模型，模型的训练效果与第一名差距较小，但是最后的预测效果较差，总结原因是前面对测试数据集的ETL，词向量的训练可能有问题，所以导致结果不理想。并且可能CNN+RNN更适合本次业务应用场景。

### ◆ 对大赛评价：

首先感谢主办方辛勤的付出，本次大赛我们感受到主办方的专业，敬业，热情，大赛给参赛团队提供的不止是展示技术，技能的平台，通过这次比赛，我们更多的开阔了视野，结实了优秀的行业前辈，并且学习到了行业发展的前沿研究知识,所以感谢主办方,希望这样的活动能够继续办下去。

## 团队风采



从左到右依次为：  
吴迪,陆浩,胡少迪,何琨,  
以上四位主要从事数据建模,数据分析  
工作  
(中间两位为带教老师)  
王琨文,孙威,赵彪  
以上三位主要从事系统开发工作