

# 源文件说明文档

## 武慧凯

Crawl.py : 用于爬取网页数据, 包括百度指数和电影基本信息。

Recognition.py : 用于处理百度索引指数的截图。

ml.py : 用于数据处理, 特征提取和分析, 模型选择和超参数确定, 以及模型的训练和测试。

predict.py : API, 读入特征文件, 输出票房信息。

## 汪春奇

story\_processing.py : 用于处理电影故事简介

actors\_processing.py : 用于处理电影演员信息

nn.py : 前向神经网络

rnn.py : 循环神经网络

## 李钢

extract\_sift\_feature.m : 用于提取海报图片 SIFT 特征

cluster.m : 用于对 SIFT 特征的聚类处理

bag\_of\_words.m : 用于生成海报图片的 BoW 表达

EuclideanDistance.m : 用于计算不同 SIFT 特征之间的相似性

main\_BoW.m : BoW 主程序

## 辛詰

pre.py : 基于三层前馈神经网络 ( 输入层 , 隐含层 , 输出层 ) , 构造数据集 ,  
基于训练集进行训练 , 完成预测