





28部电影，超70万用户， 超200万条评分/评论数据

字段说明

字段	说明
movieid	电影 id (从 0 开始，连续编号)
title	英文名称
title_cn	中文名称

	movieid	title	title_cn
0	0	Avengers Age of Ultron	复仇者联盟2
1	1	Big Fish and Begonia	大鱼海棠
2	2	Captain America Civil War	美国队长3
3	3	Chinese Zodiac	十二生肖
4	4	Chronicles of the Ghostly Tribe	九层妖塔

数据集地址：https://pan.baidu.com/s/1c0yn3TlkzHYTdEBz3T5arA

原始数据集地址：

https://www.kaggle.com/utmhikari/doubanmovieshortcomments

11.yf\_dianping数据集：

24 万家餐馆， 54 万用户， 440 万条评论/评分数据

字段说明

字段	说明
restid	餐馆 id (从 0 开始，连续编号)
name	餐馆名称

	restid	name
210902	210902	NAN
124832	124832	NAN
26766	26766	曹锡制造(新苏天地店)
91754	91754	NAN

数据集地址：https://pan.baidu.com/s/1yMNVHLi6QYsGbJT7u51Nfg

原始数据集地址：http://yongfeng.me/dataset/

12.yf\_amazon数据集：

52 万件商品， 1100 多个类目， 142 万用户， 720 万条评论/评分数据

字段说明

字段	说明
productid	产品 id (从 0 开始，连续编号)
name	产品名称
catids	类别 id (从 0 开始，连续编号，从左到右依次表示一级类目、二级类目、三级类目)

	productid	name	catids
331420	331420	欧斐亚假发 女式 虎手癣 Q7602	802,143,996
130646	130646	YESO TQT 中佳 黑褐色斜挎包 巧玛 9411	1111,864,781
179686	179686	多斯特论柏江益母野餐	832,552,337
504123	504123	Tuscarora 逆斯卡源控 中佳 豹暗豹牌壳镜头巾 PSU3083	1111,522,720
387768	387768	我们的故事—一个北大男老和博的人生形态	832,519,599

原始数据集地址：http://yongfeng.me/dataset/

数据集地址：https://pan.baidu.com/s/1SbfpZb5cm-g2LmnYV\_aR8Q

13.Datahub数据中心：

包含文本分类、情感分析以及知识图谱的数据集

相关地址：http://www.datahub.ileadall42.com/data/list?category=2&parent\_category=1

14.知乎看山杯数据集：

数据集下载地址：https://pan.baidu.com/s/1qUr6IQn6DzrMlbaAUzSIQ

提取码：qb1w

15.AI\_challenger情感分析数据集：

数据集分为训练、验证、测试A与测试B四部分。数据集中的评价对象按照粒度不同划分为两个层次，层次一为粗粒度的评价对象，例如评论文本中涉及的服务、位置等要素；层次二为细粒度的情感对象，例如“服务”属性中的“服务人员态度”、“排队等候时间”等细粒度要素。

数据集下载地址：

https://github.com/nju161250102/AI\_challenger/tree/master/data

16.复旦中文文本分类语料库

数据链接https://pan.baidu.com/s/1833mT2rhL6gBMIM0KmyKg

密码：zyxa

情感分析数据集（正面10000条，负面5000条）07-17  
吸收了 谭松波 非平衡酒店评论语料库（7000条正面，3000条负面，有部分重复数据）...

ChnSentiCorp中文情感分析数据集04-09  
ChnSentiCorp 是一个中文情感分析数据集，包含酒店、笔记本电脑和书籍的网购评论。

评论 8 条 >

写评论



张三文的博文\_Lambda 热评 谢谢你

中文情感分析数据\_BrainEditor的博文\_情感分析数据11-16  
中文情感词汇本体可以用于解决多类别情感分类的问题,同时也可以用于解决一般的倾向...

基于Bert的中文情感分析代码及分析\_SurverDevin的博文\_Bert...11-12  
基于Bert的中文情感分析代码及分析 众所周知,Google 前段时间开源了他们的 BERT - B...

Sentiment Analysis 中文情感分析-数据集03-19  
该数据集包括了四个领域的中文评论: 笔记本电脑、汽车、相机和手机。评论属于 二分...

情感分析及数据集 最新发布 yingzi的技术博文 282  
由于情感可以被分类为离散的极性或尺度（例如，积极的和消极的），我们可以将情感...

...SIMS:中文多模态情感分析数据集与细粒度的模态注释\_Yang SiChe...11-21  
CH-SIMS: A Chinese Multimodal Sentiment Analysis Dataset with Fine-grained Annotat...

中文文本情感分类及情感分析资源大全\_Paulzhao6518的博文11-21  
摘要 20世纪初以来,文本的情感分析在自然语言处理领域成为了研究的热点,吸引了众多...

中文文本分类项目数据集.rar08-08  
“复旦大学计算机信息与技术国际数据库中心自然语言处理小组”提供的小样本中文文本分...

Bert中文文本分类全网最详细教程+源码+数据集 热门推荐 cinjus的博文 1万+  
概述 Bert 是当前最先进的自然语言处理工具之一。最起码要会使用这个强大的工具。为...

中文情感分析语料库\_noter16的博文\_nlp情感分析语料库11-17  
中文情感分析的语料库非常少,这五个中文语料库是我在网上的搜集的。 数据集1:情感语...

中文情感分析任务数据集 LiQZ的博文 2291  
1 NLPCC 1.1 NLPCC-2012 下载网址: http://ccai.ccf.org.cn/conference/2012/pages/pag...

文本分类数据集汇总 buglesxu的博文 9923  
Dataset Classes Type Samples Best Method Performance AG News 4 Topic Train:1200...

可用于文本分类的中文数据集 m0\_50563633的博文 812  
可用于文本分类的中文数据集(已经标注好类) 博客的连接: https://blog.csdn.net/alip39f/...

机器学习之文本分类（附带训练集+数据集+所有代码） 天才幻想家 9183  
我本次对4类文本进行分类（（所有截图代码和数据集最后附带免费下载地址））主要...

情感分析常见数据集以及任务进展 xw2338的博文 3168  
情感分析 原文: https://github.com/sebastianruder/nlp-progress/blob/master/english/se...

机器学习-教育客服相关场景下的50组中文普通话对话文本语料-机器...08-12  
机器学习-教育客服相关场景下的50组中文普通话对话文本语料-机器学习数据处理必看-...

中文文本情感分析案例和数据集 07-08  
中文文本情感分析案例 课程配套程序 该文件夹是本次课程配套的程序，打来即可使用 ...

ocr中文数据集\_文本分类（情感分析） 中文数据...weixin\_39674366的博文 1565  
来自我的博文: https://blog.csdn.net/alip39blog.csdn.net/1.THUCNews数据集: THUCN...

最新最全自然语言处理之情感分析常用数据集 Tiger数学之家的博文 1万+  
本文介绍情感分析领域最常见的一些数据集。 目录 TOC (toc) SemEval SemEval-2014 ...

NLP实战 -中文文本分类问题: THUCNews数据集 qq\_36047533的博文 6088  
中文文本分类问题: THUCNews数据集1 THUCNews数据集与预处理1.1 数据集下载1.2 ...

中文文本情感分类实战 (weibo\_senti\_100k为...千万利器,莫过于你的信念 2156  
中文文本情感分类数据准备 数据准备 使用jieba分词 import jieba data\_path = "sources/...

文本分类数据集\_PU-Learning/文本分类/文本聚... weixin\_39797780的博文 969  
本文为PU-Learning/文本分类/文本聚类/情感分析相关研究提供部分常用数据集下载地址...

Sentiment Analysis 中文情感分析-数据集03-19  
该数据集包括了四个领域的中文评论: 笔记本电脑、汽车、相机和手机。评论属于 二分...

“相关推荐”对你有帮助么？

非常没帮助 没帮助 一般 有帮助 非常有帮助

©2022 CSDN 皮肤主题: 精致技术 设计师: CSDN官方博客 返回顶部

关于我 招贤纳士 商务合作 寻求报道 400-660-0108 kefu@csdn.net 在线客服 工作时间 8-30-  
们 士 作 道 公安备案号11010502030143 京ICP备19004658号 京网文〔2020〕1039-165号 经营性网站备案信息  
北京互联网违法和不良信息举报中心 家长监护 网络110报警服务 中国互联网举报中心 中国 Copyright 下载  
账号管理规范 版权与免责声明 版权申诉 出版物许可证 营业执照 ©1999-2022北京创新乐知网络技术有限公司



魔与刀 码龄5年 暂无认证

31 6万+ 67万+ 12万+ 原创 周排名 总排名 访问 等级  
1142 53 111 17 619  
积分 粉丝 获赞 评论 收藏



私信 关注

微博主文章



热门文章 关于文本分类（情感分析）的中文数据集汇  
总 41201