**个人简历**

reasum_line7



基本资料

姓名： 周扬 性 别：男

出生年月：1992.4 户 口：杭州

学 历：硕士 联系电话：15868403006

电子邮箱：[434647589@qq.com](mailto:1050775173@qq.com)

reasum_line7教育背景

2015.09 ～ 2018.03 杭州电子科技大学 电子通信与技术 硕士

2010.09 ～ 2014.06 嘉兴学院 电子信息工程 学士

reasum_line7基本技能

熟悉python语言，其中：

熟悉基于python的网络爬虫，熟悉urllib模块，scrapy框架，能利用fiddler软件进行抓包分析，能利用phantomjs等对反爬措施比如，异步加载，随机ID生成，验证码等进行处理。能结合cookie进行爬虫等。会设置用户代理池等一些常见爬虫技能。

了解机器学习的一些基本算法，并能通过一些小规模数据实现各个算法的功能。包括：knn分类算法，决策树，朴素贝叶斯算法进行语句分类，logistic回归算法，支持向量机，基于单层决策树的adaboost算法，树回归，利用pca或者svd进行数据简化等等。

了解数据挖掘基本知识，了解自然语言处理相关知识以及情感分析相关。

reasum_line7英语技能

语言水平：通过英语四级考试。

计算机水平：省二级C

reasum_line7

个人项目一：购物网站物品图片，名字，评论数爬取

主要实现对购物网站比如，京东，当当，淘宝等购物网站物品数据的爬取。

主要功能的实现：

1.在爬取当当和淘宝的商品信息时候，其网页在翻页后呈现很明显的规律，很快提取到各个选项对应的数据或者图片网址。

2.爬取京东网时需要从大类-小类-具体类进行层层爬取。在抓取价格的过程中需要进行抓包获取到有价格的网址，分析后获得商品ID并结合之前抓包获得的网址进行爬取。

reasum_line7

个人项目二：豆瓣网个人日记爬取

主要功能的实现：

1.在scrapy中实现网站登录。

2.在登录过程对可能出现的验证码进行处理。

3.保存相关cookie信息，让在爬行豆瓣网中其他网页时通过其保留登录状态。

4.登录成功后，根据需要进行相关信息的爬取。

reasum_line7个人项目三：腾讯动漫，新浪微博图片爬取

下载腾讯动漫里某个连载漫画或者新浪微博某微博里的的图片到本地。

以触屏版的新浪微博爬取为例：

1. 通过phantomjs模拟用户行为对网页进行翻页，显示出异步加载的所需内容。
2. 下载已加载出所需内容的网页到本地，并查看其加载后的网页截图。
3. 对所需的内容进行分析，利用常规爬虫方法进行爬取。

reasum_line7

个人项目四：对垃圾邮件进行分类

1.利用pyhton实现朴素贝叶斯对垃圾邮件进行判别。

2.创建词表，并将输入文档转换成向量，训练函数将基于训练数据最终返回两个向量包含每个词汇分别在垃圾邮件和非垃圾邮件邮件中的概率以及垃圾邮件的所占的比例。

3.分类器将根据公式P(W|Ci)\*P(Ci)计算出该邮件是否属于垃圾邮件。

reasum_line7

个人项目五：机器学习算法的python实现

结合小规模数据实现一些机器学习基本算法包括：knn分类算法，决策树，朴素贝叶斯算法进行语句分类，logistic回归算法，支持向量机，基于单层决策树的adaboost算法，利用pca或者svd进行数据简化等等。

reasum_line7

实习经历：

在网易实习，从事游戏论坛贴吧的爬取到数据清洗到文本分析得出结论并写成报告的一个过程。设计文本处理（中英文）以及情感分析

reasum_line7

荣誉与奖励

本科期间多次获得奖学金。