摘要：随着互联网的发展和普及，网络平台和论坛成为了股票投资者交流与讨论的重要平台，然后一些网络的评论信息对股票市场的影响日益显著。中国的股票市场处于刚刚发展的阶段，理性投资者较少，投资的决策很容易收到网络上的消息以及传闻的影响。在网络发展的现在，用户的评论数据量也越来越大，传统数据处理技术遇到了瓶颈，大数据相关技术为数据处理这方面提供了良好的支持。本文针对股票网站系统中，由于互联网“信息过载”造成的难以精确定位用户兴趣并提供产品推荐的问题，利用大数据平台，通过深入挖掘股票社中丰富的用户评论信息，开发产品特征提取算法，建立用户兴趣偏好模型，结合用户评论打分来改善传统协同过滤推荐的推荐准确性；通过一系列的处理，准确的挖掘出有用的信息，并向用户进行推荐。

关键字：

绪论

1.背景和意义

2.现状分析

3.目的和主要内容

4.中国股市简介

5.影响股市因素

5.1宏观因素

5.2微观因素

5.3市场因素

1.大数据相关技术概述

1.1 hdfs文件系统

1.1.1 概述

1.1.2 设计

1.1.3 特性

1.2 zookeeper协调服务

1.2.1 概述

1.2.2 选举机制

1.2.3 运行机制

1.3 Kafka消息系统

1.3.1 概述

1.3.2 架构

1.3.3 原理

1.3.4 执行流程

1.3.5 特性

1.3.6 使用场景

1.4 SparkStreaming流处理

1.4.1 概述

1.4.2 特性

1.4.3 执行流程

2.股评文本数据采集概述

2.1 数据爬虫概述

2.2 爬虫技术概述

2.3 Java爬虫Jsoup技术

2.3.1

2.3.2

3.文本数据分析相关技术与算法概述

3.1 分词

3.2 筛选

4.系统设计概述

4.1 系统模块概述

4.2 数据库设计概述

5.基于大数据平台的股评信息文本挖掘实现

5.1 股评数据爬取

2.2 消息系统中间键的搭建

5.2 大数据平台实时分词

5.3 词汇筛选

5.4 股票打分

5.5 结果处理以及股票推荐

结 论

附 录

参考文献

致 谢