**中国科学技术大学软件学院**

**工程实践**

**结题报告**

**项 目 名 称：基于 Spark 的景点推荐系统**

**成 员 名 单： 褚思源 林靖清 王炜康**

**导 师： 袁宇丹**

**工 程 领 域： 推荐系统**

**研 究 方 向： web开发、个性推荐**

**一、 简况**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **名称** | **中文** | 基于 Spark 的景点推荐系统 | | | |
| **英文** | sights recommendation system based on spark | | | |
| **项目组成员名单** | **姓名** | | **学号** | **项目中的分工** | **签 章** |
| 林靖清 | | SA19225242 | 主要负责推荐系统 |  |
| 王炜康 | | SA19225378 | 主要负责爬虫 |  |
| 褚思源 | | SA19225123 | 主要负责网站 |  |
|  | |  |  |  |
| **中 英 文 摘 要** | **随着英特网的快速发展，用户可以享受更加便利地服务。同时，面对海量数据，精准地获得目标信息也变得更加困难。一方面，用户面临对大量的信息束手无策，经常会迷失在大量的信息空间中；另一方面，网站也失去了与用户的联系，不能与用户建立稳固的合作关系。在这种背景下推荐系统（Recommender Systems）应运而生，它根据用户的趣爱好推荐符合用户兴趣爱好的对象。**  **本课题计划开发一个景点推荐系统。系统是一个基于SSM框架的web项目，通过Scrapy框架爬取景点数据，并通过Spark引擎进行数据处理、实现推荐算法。**  **其中后台管理系统对用户信息和景点信息进行管理，如添加删除景点信息和完成用户信息的完善。推荐系统通过在网站系统获取到用户的点击事件（如用户对某各景点的评分、用户的对某个景点详情页的浏览次数）记录信息，并根据根据该信息做出相应的处理，将推荐结果存入到数据库中，web 前端通过查询数据库将推荐的景点展示给用户。**  **With the rapid development of Internet, users can enjoy more convenient services. At the same time, in the face of massive data, it is more difficult to obtain target information accurately. On the one hand, users are at a loss to deal with a large amount of information and often get lost in a large amount of information space; on the other hand, the website also loses contact with users and cannot establish a stable cooperative relationship with users. In this context, recommender system arises. It recommends the objects according to the interests of users.**  **This project plans to develop a scenic spot recommender system. The system is a web project based on SSM framework. It crawls the data of scenic spots through scrapy framework, processes the data through spark engine, and realizes the recommendation algorithm.**  **The background management system manages user information and scenic spot information, such as adding and deleting scenic spot information and completing the perfection of user information. The recommender system obtains the user's click event (such as the user's rating for each scenic spot, the user's browsing times for the details page of a scenic spot) record information in the website system, and makes corresponding processing according to the information, stores the recommendation in the database, and the web front end displays the recommended scenic spot to the user through the query database.** | | | | |
| **主题词** | **中文** | **Spark引擎 景点推荐** | | | |
| **英文** | **Spark/Spot recommender** | | | |

**二、项目背景**

**1.阐述该选题的研究意义，或工程设计的价值和意义，国内外概况和发展趋势，选题的先进性和实用性，技术难度及工作量。**

随着Internet的发展，英特网为用户提供越来越多信息和服务的同时，其结构变得更加复杂，在网络上的海量信息中，用户及时地获得所需要的信息也更加困难。一方面，用户面临对大量的信息束手无策，经常会迷失在大量的信息空间中；另一方面，网站也失去了与用户的联系，不能与用户建立稳固的合作关系。在这种背景下推荐系统(Recommender Systems)应运而生，它根据用户的趣爱好推荐符合用户兴趣爱好的对象。推荐系统模拟商店销售人员向用户提供商品推荐，不仅帮助用户找到所需信息，而且通过推荐系统，网站可以有效保留用户，提高网站的点击率和用户的忠诚度。

从20世纪90年代中期推荐系统的概念出现至今，推荐系统在近几年得到了飞速发展，在电子商务、影音娱乐方面有着广泛应用。

基于以上原因，我们计划开发一个景点推荐系统。系统以网站的方式呈现，通过Scrapy框架爬取数据，并通过Spark引擎进行数据处理、实现推荐算法。

**2.参考文献**

[1] Davidson J, Liebald B, Liu J, et al. The YouTube video recommendation system[C]//Proceedings of the fourth ACM conference on Recommender systems. 2010: 293-296.

[2] 蒲姗姗,何燕.个性化学术资源推荐研究:现状、特点及展望[J].图书馆学研究,2019(16):9-17.

[3] 刘鲁,任晓丽.推荐系统研究进展及展望[J].信息系统学报,2008(01):82-90.

[4] 周俊,董晓蕾,曹珍富.推荐系统的隐私保护研究进展[J/OL].计算机研究与发展,2019(10):1-16[2019-11-15].http://kns.cnki.net/kcms/detail/11.1777.TP.20191025.1706.006.html

[5] McDonald D W, Ackerman M S. Expertise recommender: a flexible recommendation system and architecture[C]//Proceedings of the 2000 ACM conference on Computer supported cooperative work. 2000: 231-240.

[6] Errico J H, Sezan M I, Borden G R, et al. Collaborative recommendation system: U.S. Patent 8,949,899[P]. 2015-2-3.

[7] Shaya S A, Matheson N, Singarayar J A, et al. Intelligent performance-based product recommendation system: U.S. Patent 7,809,601[P]. 2010-10-5.

[8] Park M H, Hong J H, Cho S B. Location-based recommendation system using bayesian user’s preference model in mobile devices[C]//International conference on ubiquitous intelligence and computing. Springer, Berlin, Heidelberg, 2007: 1130-1139.

[9] Wei J, He J, Chen K, et al. Collaborative filtering and deep learning based recommendation system for cold start items[J]. Expert Systems with Applications, 2017, 69: 29-39.

**三、课题内容及具体实施方法**

|  |
| --- |
| **1．课题内容**  本课题主要进行基于spark的景点推荐系统的开发，分为三个部分：爬虫爬取数据，搭建旅游网站，为用户提供个性化推荐。   1. **系统需求分析**   在主网站中，用户只需在系统中对他们已有印象的景点进行打分，系统便可推荐其他景点。用户有以下功能性需求：  1.浏览推荐景点  2.景点评分  3.查看景点详  4.登录注册。  5.景点收藏  在后台管理系统中，管理员的功能性需求有：  1.登录  2.对景点进行管理，具有增、删、查、改权限。  3.对用户进行管理，具有增、删、查、改权限。  4.对评论进行管理，具有删、查权限。   1. **系统概要设计**   1.系统总体架构  在详细分析了系统需求之后，本文设计的个性化旅游景点推荐系统的总体架构如图1-1 所示。    图1-1系统总体架构  系统整体结构可分为表现层、接口层、服务层、计算层和存储层这五部分。用户通过表现层发出请求后，会调用接口层的接口来处理此请求，具体的业务逻辑处理由服务层提供。计算层由Spark提供离线计算，它是通过收集到的用户评分数据利用ALS推荐算法预测用户可能感兴趣的景点来为用户提供推荐服务。存储层提供数据的存储，数据库中存放着管理员信息、用户信息、景点信息、评论信、评分等信息。  2.模块结构与划分  根据需求分析以及系统总体结构设计，系统的总体功能结构如图1-2所示。系统主要分为后台管理系统和主网站两部分，主网站主要由首页、个人中心和景点详情页三个部分组成。首页主要为用户展示热门景点和最受欢迎景点以及为用户提供的个性化景点推荐，个人中心主要为用户提供个人信息的修改功能和查看修改自己的收藏景点和自己的评论，景点详情页为用户提供景点的详细情况以及打分和收藏功能。后台管理系统主要为网站的管理员提供了用户管理、景点管理、评论管理功能。    图1-2  3 数据库设计  本小节根据系统功能需求分析，对景点推荐系统的数据库进行介绍。在数据库概念设计中使用 E-R 图来描述整个数据库的模型，本系统主要包含的实体有：景点、用户、管理员和评论信息等，具体如图1-3所示的E-R图。     1. **采用的开发方法、环境，测试方案等**   爬虫  数据库：MongoDB、Redis  框架： Scrapy，使用 scrapy\_redis 和 Redis 实现分布式  网站  开发环境： IntelliJ IDEA、maven、git  软件架构： mysql、mybatis、springboot  后台管理系统  开发环境： IntelliJ IDEA、maven、git  软件架构： mysql、mybatis、springboot  推荐系统  开发环境： IntelliJ IDEA、maven、git  软件架构： spark、mysql   1. **技术难度及特色分析**   此课题的难点与特色为采用离线计算的个性化推荐。离线计算使用Spark-ALS推荐算法，数据来源为用户对景点的评分，根据用户与景点的评分矩阵进行矩阵分解，获得推荐结果。 |

**四、工作进度总结**

|  |
| --- |
| **应包括文献调研，理论分析，数值计算，理论分析，软硬件设计，仪器设备的研制和调试，撰写结题报告、技术论文等，要给出各个阶段的成果形式。**  2019.10-2019.11：选题，搜集有关资料，对课题背景意义、实用性和先进性与老师进行探讨，并完成开题报告的撰写  2019.11-2019.12：根据收集到的资料进行研读，学习相关技术并且完成初步方法的创写以及进行初步应用  2019.12-2020.02：实现若干功能，包括数据爬取框架的建立，数据库的创建以及数据的存储等基础模块  2020.02-2020.04：进行网站的前后端开发。  2020.04-2020.05：完成个性化推荐功能并测试。  2020.06-2020.07：撰写结题报告、结题答辩 |
| **成果及特色**  以截图形式展示：  图片包含 截图, 游戏机  描述已自动生成图片包含 截图, 游戏机  描述已自动生成图片包含 截图, 游戏机  描述已自动生成图片包含 游戏机, 截图  描述已自动生成图片包含 游戏机, 截图  描述已自动生成 |