****

数据结构课程设计开题报告

|  |  |
| --- | --- |
| **题 目：** | 【深度资讯】新闻热点分析及个性化推荐系统 |
| **姓 名：** | 1．孙嘉祎 2. 黄欣灵 3. 沐思遥 |
| **学 号：** | **1.**202000800154 **2.**202000800082 **3.**202000800109 |

**2022 年 4 月 17 日**

课程设计开题报告

|  |
| --- |
| **1.本课题的研究意义和目标** |
| 信息过载，是我们在这个数字化时代的常态。现代社会，冗余的信息流，无时无刻不在被制造、传播到我们身边。浏览网站，网络信息量爆炸，用户极容易在纷杂的信息中难辨方向，找不到自己需要的信息。同样，在新闻领域中，每天的时事新闻多种多样，且不同新闻会引起不同用户群体的关注。如果不将新闻分类并设置个性化推荐，不同用户群体将很难在大量的新闻中获取自己感兴趣的新闻，即很难拥有较良好的阅览新闻体验，为了解身边事及天下事增添了难度。  基于这一实际问题，本次数据结构课程设计，我们小组将设计一个【深度资讯】新闻网站，除了新闻网站的基础功能即搜索、阅览新闻等外，我们同时将目光聚焦在新闻热点分析及个性化推荐功能上——新闻热点分析聚焦总体关注度最高的新闻，设置热点新闻栏目，不拘于分类，为所有用户提供服务，便于用户了解当今热点新闻；个性化推荐为不同用户群体设置不同推荐，便于不同群体用户获得自己更感兴趣的新闻，减少繁琐的搜索步骤，拥有更佳的阅读体验。  同时，此次课程设计同样锻炼了我们小组的合作能力，加强了我们对于数据结构相关知识点的学习理解，辅助我们更深入了解机器学习相关内容，也提高了我们结合实际运用知识的能力。 |
| **2. 具体研究内容** |
| 1.推荐系统的评价指标  由于推荐系统比较复杂，所以涉及到的评价指标也很多。当然，用户满意度最为的有效，因为这本来就是推荐系统的最终目标，但是奈何资源有限成本太高，推荐系统还依赖于其它客观评价指标。  (1)推荐准确度：这个参数可以离线计算所得，而且较为的客观，所以是各大研究论文算法最重要的参考指标。  (2)覆盖率：体现了挖掘算法对发掘长尾商品的能力。最简单的定义是，对所有用户推荐出的产品做并集，然后看这个出现的并集产品与总产品数中所占的比例，这种方式比较的粗线条，因为推荐系统中马太效应频繁，所以好的推荐算法应当是所有商品被推荐的几率差不多，都可以找到各自合适的用户，所以实际中会考虑信息熵、基尼系数等指标。  (3)多样性：因整个推荐系统涉及到的因素太多，如果只推荐用户一个类别的相似物品，失败风险比较的大，而且也难以实现整个推荐效益的最大化。  (4)新颖性：原理就是那些用户没有接触过、没有操作过的商品，或者是流行度比较低的商品，对用户来说是比较新鲜的物品，往往会有意外的效果。  (5)信任度：这个指标比较的主观，就是让用户信任推荐系统做出的推荐是有根据有理由的，以及推荐系统内部是如何运作的。  (6)健壮性：比如针对关联推荐算法，商户恶意下单提高产品的推荐频率，水军恶意评论等。  2.机器学习在推荐系统的应用  所谓的机器学习模型指使用机器学习相关算法来得到召回层结果的方法，例如NMF（Nonnegative Matrix Factor，非负矩阵分解）算法、LFM模型（Latent Factor Model，基于潜在隐因子模型）、FTRL（Follow The Regularized Leader）算法、PersonalRank算法等。  所谓的机器学习模型指使用机器学习相关算法来得到召回层结果的方法，例如NMF（Nonnegative Matrix Factor，非负矩阵分解）算法、LFM模型（Latent Factor Model，基于潜在隐因子模型）、FTRL（Follow The Regularized Leader）算法、PersonalRank算法等。  所谓的深度学习模型指使用深度学习相关算法来得到召回层结果的方法，包括深度神经网络（DNN）和Embedding等方法。  而在排序层中，我们也可以使用机器学习和深度学习两种方式来对召回结果进行排序，这里的排序一般采用点击率预估，并将预估得到的概率降序排序取前N得到最终的排序结果。  3.机器学习算法在工业生产中的应用  微服务是一种软件开发技术- 面向服务的体系结构（SOA）架构样式的一种变体，它提倡将单一应用程序划分成一组小的服务，服务之间互相协调、互相配合，为用户提供最终价值。每个服务运行在其独立的进程中，服务与服务间采用轻量级的通信机制互相沟通（通常是基于HTTP的RESTful API）。每个服务都围绕着具体业务进行构建，并且能够独立地部署到生产环境、类生产环境等。另外，应尽量避免统一的、集中式的服务管理机制，对具体的一个服务而言，应根据上下文，选择合适的语言、工具对其进行构建。微服务这种松耦合度的架构特别适用于机器学习算法在已有工业框架当中的部署与整合。 |

课程设计开题报告

|  |
| --- |
| **3.本课题的重点和难点** |
| 1. 机器学习算法选择。选择合适的机器学习算法进行热点分析及个性化推荐； 2. 应用程序架构。选择合适的架构以提高系统稳定性，提高并发度，规范化软件体系结构，简化开发流程； 3. 设计简洁便于操作的用户界面。设计用户界面美观简洁且便于用户使用，达到使用门槛低的要求。 |
| **4.进度安排** |
| 前期：学习项目的相关知识，如NLP、Deep FM、协同过滤等，初步建立解决问题的模型架构  中期：完成项目基本框架，做出可视化系统  后期：扩大数据量，进一步优化项目框架 |

课程设计开题报告

|  |
| --- |
| **5. 参考文献（论文和网址）** |
| 1. LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation. SIGIR 2020 【GNN-based】 2. FairRec: Fairness-aware News Recommendation with Decomposed Adversarial Learning Author(Institute): Fangzhao Wu(Microsoft) 3. Who You Would Like to Share With? A Study of Share Recommendation in Social E-commerce Houye Ji , Junxiong Zhu , Xiao Wang , Chuan Shi, Bai Wang , Xiaoye Tan , Yanghua Li , Shaojian He 4. [推荐系统工业界顶会论文总结——AAAI 2021 - 知乎 (zhihu.com)](https://zhuanlan.zhihu.com/p/467270467)我从这上面扒的，看着粘贴吧 5. <https://www.jiqizhixin.com/articles/2019-02-11-6> 个性化推荐系统实践应用 6. <https://blog.csdn.net/weixin_46605806/article/details/118642644> 新闻个性化推荐系统 |