# 摘要

教育大数据已经是国家大数据十大工程之一，《教育信息化2.0行动计划》明确指出要全面提高利用大数据支撑保障教育服务能力。当前，“互联网+教育”一方面建设了丰富的教育资源，同时产生了大量的学习过程数据。借助推荐技术充分开采上述数据，能有效刻画用户学习状态、能力和需求，提供个性化的学习服务。

针对如上需求，基于开源的MOOPer数据集提供的学习过程数据，设计并实现了在线学习个性化服务平台。能够通过基于LFM的推荐算法为学生提供个性化学习推荐服务与学习资源质量评价，通过机器学习为老师提供学生学习过程中的异常行为侦测，并采用Hadoop、Spark等分布式技术实现上述服务带来的大规模矩阵分布式计算优化，采用Vue和Echarts构建了教育大数据分析与可视化系统。

关键词：LFM算法；聚类算法；Hadoop；Spark；Vue

# 开头

近些年来，随着云计算和大数据的兴起，

分布式计算技术得到了广泛的应用和发展。分布式计算技术的出现主要是为了应对大规模数据处理的需求。通过将一个任务或一个应用程序拆分成多个子任务或子应用程序，分布到多个服务器上进行处理，从而实现数据和计算的分布式处理，避免单点故障，提高了系统的稳定性和可靠性，同时也提高了系统的可伸缩性和扩展性，可以大幅度减少计算时间，提高数据处理的效率。在推荐系统中，通过采用分布式计算技术，可以更好地处理大规模用户和物品之间的关系，提高推荐的准确性和实时性。因此，分布式系统已经成为了当前大规模数据处理和高并发访问的标配。

# 1 相关工作

## 数据集介绍

## 分布式计算设计

## 相似度计算方法

## 1.4 KNN算法

# 2 个性化学习支撑平台

## 2.1 个性化学习支撑服务

基于流行度的推荐

基于内容的推荐

基于协同过滤的推荐

实时推荐

## 2.2 学习异常行为检测

什么是学习异常行为，异常行为检测思路，检测效果常用性能指标，检测结论。多截图

线上学习因缺少面对面的直接交流，教师无法及时掌握学生的知识学习和能力达成情况，也无法得知学生的真实学习状态。如何基于学习过程数据及时准确地发现学生的异常学习行为，进而构建合理的在线学习评价机制，是提升在线学习的质量和效率的关键，也是在线学习的重要支撑服务之一。

学习异常行为检测分为学习行为异常检测和学习时间异常检测。学习行为异常检测是

Mooper提供了学生在课程测试中取得的最终成绩、测试次数、测试耗时等学习行为数据。通过聚类算法将具有相同特征的数据分为一类，由于一个班级的学生学习水平大致相当，在相同的环境中同一老师的教导下，分析学生在课程测试上的花费时间与课程成绩是否正常

精准度

## 2.3 课程评价词云

设计理由

算法设计：数据预处理-》权值提取-》

效果实现

# 3 结束语

# 参考文献