**湖 南 信 息 学 院**

**数据挖掘课程论文**

|  |  |
| --- | --- |
| **论文题目** | **数据挖掘在教育行业的应用** |
| **学生姓名** | **谢梦华** |
| **学号** | **201501420328** |
| **二级学院** | **电子信息学院** |
| **专业名称** | **软件工程** |
| **班级** | **15级软件工程三班** |
| **指导教师** | **张钰莎** |

二〇一八年 十 月 十八 日

**摘 要**

近三十年来，随着物联网、大数据、云计算等新兴的网络技术在智慧校园建设中的广泛应用，使得教育信息化迅速发展，导致教育数据迅速增长和教育大数据的出现。然后这些飞速增长的数据没有等到有效利用，为了解决此问题，文章利用数据挖掘技术对教育大数据进行分析研究，从中挖掘出有价值的知识。这些知识被广泛应用在指导学校的日常管理、优化教师教学模式、激发学生自主学习兴趣等方面，效果不错。

**关键词**：大数据时代；数据挖掘技术；应用

**ABSTRACT**

In the past 30 years, with the extensive application of emerging network technologies such as the Internet of things, big data and cloud computing in the construction of smart campus, education informatization develops rapidly, leading to the rapid growth of education data and the emergence of education big data.However, these rapidly growing data have not been effectively utilized. In order to solve this problem, this paper USES data mining technology to analyze and study education big data and mine valuable knowledge from it.This knowledge has been widely used to guide the daily management of schools, optimize the teaching mode of teachers, and stimulate students' interest in independent learning.

**Keywords:** Big data era;Data mining technology;application

**目 录**

**诚信明**……………………………………………………………………………………Ⅰ

**摘要**……………………………………………………………………………………Ⅱ

Abstract…………………………………………………………………………………Ⅲ**第一章 前言**…………………..…….…………………………………………………1

**第二章 大数据时代下的数据掘**…………………………………………………2

2.1 大数据的发展………………….…………………………………………………2

2.2数据挖掘的产生………………….………………………………………………2

2.3大数据时代下的数据挖掘…………………………………………………3

2.3.1 应用于学校日常的管理……………………………………………………..3

2.3.2应用于设置课程和试卷分析，使其更加合理化….………………………..4

2.3.3大数据在教育资源分配中的应用……………..….…………………………4

**结论**…………………………………………………………………………………5

**参考文献**………………………………………………………………………………6

**致谢**………………………………………………………………………………………7

第一章 前 言

随着网络通信和计算机信息技术的发展与应用，人类社会已经进入大数据时代。人们可以感知记录种类繁多和规模超大的网络信息数据，同时通过分析和处理这些海量数据，能够对蕴含其内的核心信息进行深度挖掘，得到更多的价值信息。这些知识已被应用于交通运输、银行保险、科技医疗各个行业，能够很好地指导和促进企事业的管理与发展。目前信息化技术的迅猛发展使教育信息化进入一个新阶段，随着网络学习和各种应用系统的普及推广，教育领域的关于教师和学生的各类相关数据量增长速度很快，这些教育大数据直接影响学校的日常管理、教学效果和师生的工作学习生活。智慧校园使大数据、云计算、物联网和数据挖掘等网络技术与学校教育深度融合，特别是数据挖。

技术在教育领域的广泛应用，便于从海量教育大数据中获取有利用价值的隐藏信息。通过分析研究这些来自于教师（科研能力、授课情况、个人信息等）和学生（基本资料、学习行为、课外活动、消费记录和社交圈子等）各个方面的有用的知识，将其合理应用到教育教学的多个环节，能够支持领导决策、改善教育质量、促进教育公平、指导学校的日常管理和教学活动等；同时也能够弥补数据挖掘在教育大数据领域中的应用研究这一空白。

第二章 大数据时代下的数据挖掘

## **2.1 大数据的发展（1980年－至今）**

“大数据”一词在1980年[美]著名未来学家阿尔文·托夫勒著的《第三次浪潮》书中将“大数据”称为“第三次浪潮的华彩乐章”。1997年美国宇航局研究员迈克尔·考克斯和大卫·埃尔斯沃斯首次使用“大数据”这一术语来描述20世纪90年代的挑战：模拟飞机周围的气流——是不能被处理和可视化的。数据集通常之大,超出了主存储器、本地磁盘，甚至远程磁盘的承载能力。称之为“大数据问题。”

2002年在9/11袭击后,美国政府为阻止恐怖主义已经涉足大规模数据挖掘。前国家安全顾问约翰·波因德克斯特领导国防部整合现有政府的数据集，组建一个用于筛选通信、犯罪、教育、金融、医疗和旅行等记录来识别可疑人的大数据库。一年后国会因担忧公民自由权而停止了这一项目。2004年9/11委员会呼吁反恐机构应统一组建“一个基于网络的信息共享系统”，以便能快处理应接不暇的数据。“大数据”在云计算出现之后才凸显其真正价值，谷歌(Google)在2006年首先提出云计算的概念。

2007–2008年随着社交网络的激增，技术博客和专业人士为“大数据” 概念注入新的生机。“当前世界范围内已有的一些其他工具将被大量数据和应用算法所取代”。《连线》的克里斯·安德森认为当时处于一个“理论终结时代”。一些政府机构和美国的顶尖计算机科学家声称,“应该深入参与大数据计算的开发和部署工作,因为它将直接有利于许多任务的实现。”2008年9月《自然》杂志在推出了名为“大数据”的封面专栏。

**2.2 数据挖掘的产生**

数据挖掘（英语：Data mining），又译为资料探勘、数据采矿。它是数据库知识发现（英语：Knowledge-Discovery in Databases，简称：KDD)中的一个步骤。数据挖掘一般是指从大量的数据中通过算法搜索隐藏于其中信息的过程。数据挖掘通常与计算机科学有关，并通过统计、在线分析处理、情报检索、机器学习、专家系统（依靠过去的经验法则）和模式识别等诸多方法来实现上述目标。

数据挖掘利用了来自如下一些领域的思想：(1) 来自统计学的抽样、估计和假设检验，(2)人工智能、模式识别和机器学习的搜索算法、建模技术和学习理论。数据挖掘也迅速地接纳了来自其他领域的思想，这些领域包括最优化、进化计算、信息论、信号处理、可视化和信息检索。一些其他领域也起到重要的支撑作用。特别地，需要数据库系统提供有效的存储、索引和查询处理支持。源于高性能（并行）计算的技术在处理海量数据集方面常常是重要的。分布式技术也能帮助处理海量数据，并且当数据不能集中到一起处理时更是至关重要。

## 2.3 数据挖掘在教育行业的应用

**2.3.1 应用于学校日常的管理**

一、教育数据的挖掘，对管理部门、教师、学生和技术研发人员具有重要的意义。学校的教学管理数据库中记录着所有教师和学生的工作学习、科研活动、社会实践、处罚奖励等相关情况，领导利用数据挖掘技术对教学资源和管理数据进行深入的关联分析，找出师生各种常见行为或活动之间的内在隐蔽联系。并在管理中采取过程监控、风险预警、分类管理、趋势预测等措施，改变了以前定性和模糊的分析和评价老师课堂教学质量的不科学现象，能够改善学校目前的考核管理方式，实现智准确、高效管理学校的各项工作，为教学应用和学校发展提供有效科学的决策依据。在通常情况下，在教师平时的教学过程中采用讲授法、调查法、参观法、实验法、实习法、分组讨论法、计算机辅助教学法等多种教学方法来完成自己的教学任务。由此通过运用关联规则或回归线性分析数据挖掘等方法来分析研究相关的教育大数据，选择有利于学生知识吸收和教学需要的最佳授课方式。

二、又如采用数据挖掘技术通过智慧校园跨平台了解学习其他老师的优秀课程和教学方法，研究学生个性化学习工具、网络学习过程、兴趣爱好和学习行为，有利于掌握其学习规律和特点，能够为其推荐合适的学习资源，优化学习方法和改善开展自我导向的适应性学习和提高课堂教学质量。它为教育教学活动提供实时科学的信息数据，有助于形成关于教育教学的智慧决策，提供客观依据有利于教育教学活动的有效实施，尽可能地实现和完善教育教学活动的价值与功能。通过挖掘教育大数据，可以对学生的学习成绩、兴趣爱好、消费轨迹、行为记录及奖励处罚数据库等相关信息进行分析研究和处理，能够快速获取学生的鉴定结果，便于及时禁止和指正学生不良学习行为。这样既能减轻教师的工作量，又能够避免教师对学生先入为主的缺乏客观和公正的主观武断性评价。利用教育大数据的挖掘结果来科学评定学生的学习行为，其优点是合理地反馈学生信息、激发学生学习兴趣、发现学生的个性需求和实现因材施教。

三、根据学生的个人信息、学习成绩、网上学习轨迹、性格特点、知识结构等相关信息，挖掘出学生的基本特征，宏观指导和微观帮助学生不断修正其学习行为。学生不再局限于本校的某位老师的课堂教学，可以自主选择全球范围内的相关课程的著名学者的优秀课件，根据自己的时间和需求来自由学习，这样不但提高了学习效率，而且培养了学习兴趣。有利于教师通过比较事先制定的学生行为标准和实际通过数据挖掘技术对学生个性特征的分析结果，可以指导学生完善人格、修正自己的学习行为和提高学习能力，有利于学生综合素质的全面发展。

**2.3.2 应用于设置课程和试卷分析，使其更加合理化**

一、学生在校学习的过程中，课程合理安排的先后顺序非常重要，因为基础课程没有学的话，那么后继课程的学习就无法进行。即使是同一个年级的学生学习同一门课程，但由于授课老师、班级学习风气和自身的基础等原因，导致最终的学习成绩差别很大。通过数据挖掘的时间序列和关联分析等方法，仔细分析存放学校教学数据库中的往届学生各个学科的试卷和考试成绩，挖掘出这些海量教育数据的有价值的信息并分析这些数据的回归性和相关性性质，寻找其中的有用规律和影响学生学习成绩的重要因素，以此合理安排新生的课程。考试是教学活动的一个重要环节，能够检验教学效果。

二、虽然考试成绩能够反映教学效果，但无法说明影响成绩的具体因素和影响教学的直接原因，不能促进教学发展。加之试题的质量也能影响学生考试分数的高低，探索有效评价试题覆盖知识点全面度和难易度等质量的方法非常重要。如果在试卷分析过程中采用数据挖掘的关联规则，教师通过学生每道试题的实际得分情况便能分析出试题难易度、相关度和区分度等技术指标，就能较为合理地评价试题质量，实事求是地检查其的教学效果和学生对知识的掌握的具体情况，便能指导其今后的教学活动。

**2.3.3 大数据在教育资源分配中的应用**

一、教师的教学和学生的学习活动能够实现的基础就是教学资源，以前通过教师的自主研发和教育主管部门的配发来建设教学资源，而教师的自主研发极易出现资源技术含量低、可用性差和重复率高的弊端，政府配发的资源只能满足大部分师生的需要，无法满足个性化需求。挖掘教育大数据为教育资源的建设提供了新的思路，为教学资源库的构建提供技术支持，让优质资源的判定有据可依。它使广大师生能够方便地使用和共享存储在云端的教育资源数据，还能对大量非结构化的数据资源进行分析，挖掘出隐藏的有用信息，享受满足自身个性需求的数据资源，能够避免教学资源的重复建设和优质资源的浪费。

二、在教育大数据中借助数据挖掘研发人员发现智慧校园中各应用系统的使用频 率和相关内容，然后根据师生需求优化系统的操作方式和用户界面，不断完善系统以便提高服务质量。挖掘结构化和系统化的教育大数据，将MOOC，游戏学习等新模式和云计算、虚拟技术、3D 打印、网络计算、引擎开发等新技术应用于智慧校园建设中，对教学过程的实时监控和分析研究来保障教学质量。

**结 论**

随着教育信息化的推广应用，使学校多年的教学管理活动积累了大量的非结构化数据。为了合理有效地充分利用这些教育大数据，本文通过数据挖掘技术对教育大数据进行深度的分析研究，并将其应用到了教学管理、教师授课、学生学习等教育教学的各个环节，有利于优化教学管理、提高教学质量和推动教学改革。但是此类研究国内还不成熟，需要研究者投入更多的精力，突破技术瓶颈和应用限制。

**参考文献**

[1] 马秀麟，衷克定，刘立超.从大数据的视角分析学生评教的有效性.中国电化教育，2014（28）：1-2.

[2] 程称.大数据分析，2014（3）：1-2.

[3] 吉根林，赵斌.面向大数据的时空数据挖掘综述.南京师大学报，2014（4）：3.

[4] 王晓妮，韩建刚.数据挖掘在教育大数据中的应用研究，2018（25）：3-4.