

数据挖掘在高校教务管理中的应用

杨亚菲

(国家开放大学 信息化部,北京 100039)

摘要:随着数据时代的到来,人们越来越重视数据的价值。目前已有许多先进数据处理技术应用于教育行业,其中数据挖掘对高校教务管理的研究受到广泛学者的关注。介绍了数据挖掘基本概念、数据挖掘实施步骤以及国际公认的经典数据挖掘算法,探讨了对于数据挖掘中分类算法、聚类算法以及关联分析在教务管理方面的应用,并阐述了对数据挖掘技术在教务管理方面应用的思考。

关键词:数据挖掘;教务管理;应用

中图分类号:G436 **文献标识码:**A **文章编号:**1672-7800(2019)06-0070-02

DOI:10.16735/j.cnki.jet.2019.06.025

在当今数据驱动的时代,数据是非常宝贵的资产。数据挖掘的核心是挖掘数据的商业价值,在大量数据中通过算法找到隐藏在背后的关键规律用以制定相关决策、指导相关业务工作。因此无论工作还是生活中的各种决策,因为有了数据挖掘的结果而变得更加省时、省力精准有效。随着高校将先进信息技术应用在教务管理的探索性实践发展,教务管理不仅可以规范高校相关管理工作,还可以在在一定程度上提高教学质量。教务管理系统可以说是高校中具有最丰富且最大规模数据的系统,利用数据挖掘相关技术获取教务管理数据中潜在的信息,对高校的发展起着至关重要的作用。

1 数据挖掘

1.1 概述

根据相关学者对数据挖掘的定义,数据挖掘可称为知识发现,它是在大量、随机、不完全、模糊的实际应用数据中提取有价值、潜在有用知识的处理过程^[1-2]。它更是一种强大的人工智能工具,通过分析数据库中多角度多维度的数据以发现有价值信息,并分类汇总数据之间的关系,运用此结果帮助制定和改进决策。

1.2 步骤

数据挖掘过程包括以下步骤:①数据获取,即根据已有需求有针对性地在数据源获取相关数据;②数据预处理,即规范格式、清洗数据、集成数据以进一步提高数据质量;③模型建立及优化,即选择和应用各种数据挖掘模型,并进行优化,以便得到更好的分类结果,此过程也可以视为选择合适的算法;④数据挖掘,即对经过处理的数据使用选择好的

模型进行数据挖掘,以获取有价值的规律或知识;⑤知识展现,即对数据挖掘的结果分析并评估,一般最后会进行可视化展示^[1-2]。

1.3 常用方法

国际权威学术组织 The IEEE International Conference on Data Mining (ICDM) 2006年12月在香港评选出了数据挖掘领域的十大经典算法:C4.5, K-Means, SVM, Apriori, EM, PageRank, AdaBoost, KNN, Naive Bayes, CART。根据主要用途可将上述十大经典算法分为分类、聚类、关联分析以及链接分析4类,以下就教务管理方面常用的挖掘算法进行阐述。

分类是基于未知类别数据的特征或属性,将其归类为已有类别。常用的分类算法有朴素贝叶斯以及决策树(C4.5, CART)。朴素贝叶斯在概率论原理指导下,基于贝叶斯定理与特征条件独立假设,对于待分类项求解在此项出现的条件下各个类别出现的概率,将要分类的项目分类为具有最高计算概率的类别^[5]。决策树又称判断树,它是一种以树形数据结构来展示决策规则和分类结果的模型,作为一种归纳学习算法,其重点是将看似无序、杂乱的已知实例,通过某种技术手段将它们转化成可以预测未知实例的树状模型^[6-7]。

聚类分析就是在未知的情况下根据数据样本之间一定的相似度将其划分为若干个不相区间,然后根据与已有数据样本区间的相似度来归类待划分数据^[4]。其中,k均值聚类算法(K-Means)比较经典,在k-Means算法中,对于给定的样本集,根据样本之间的距离将样本集划分为k个类。

关联分析是找大量数据集存在的关联性或相关性,描

收稿日期:2018-12-23

作者简介:杨亚菲(1991-),女,硕士,国家开放大学信息化部研究实习员,研究方向为计算机软件与理论。

述事物中某些属性同时出现的规律和规则,其中,Apriori 是经典的关联规则算法。Apriori 算法通过挖掘频繁项集来揭示对象之间的关联关系,其中频繁项集是指经常出现在一起的对象集合,关联规则隐含着对象之间可能存在很强的关系。

2 数据挖掘在教务管理系统中的应用

2.1 分类算法在教务管理中的应用

通过朴素贝叶斯分类算法监测学生考试成绩,分析学生考试成绩变化趋势^[5]。此外,将教务管理中学生的平时成绩和试卷成绩以及具体成绩细节作为数据挖掘对象,使用决策树算法进行分类分析,挖掘影响学生综合成绩的重要因素,以及这些因素与综合成绩之前的正负关系^[6-7]。基于此,不仅为教务管理者制定教学计划提供了科学依据,而且还可有针对性地指导学生高效率学习。

2.2 聚类算法

对教务管理中学生对教师教学测评结果和教师授课信息进行聚类分析,挖掘教师教学水平评价的关键影响因素,分析学生就相关课程对教师教学质量关注的重要指标^[4]。基于此分析,一方面教师可以有针对性地调整授课侧重点,提高教师教学水平;另一方面可以作为衡量教师教学质量以及教师绩效考核的重要量化参考依据,进而促进高校教师的团队管理。

2.3 关联分析

对教务管理积累的历史选课数据以及已经取得的成绩为对象进行关联分析,针对学生的选课倾向与学习课程后掌握情况、考试成绩等信息,挖掘课程之间相关性和课程与成绩之间相关性等有用信息构建相关模型^[8]。基于此,可以实现个性化课程推荐,为后续学生有针对性地调整培养方案提供科学依据,进一步完成个性化人才培养目标。此外,可为以后课程具体内容设计方面提供细节参考,帮助教师有的放矢地准备授课内容,提高整体教学质量。

3 对数据挖掘技术在教务管理中应用的思考

3.1 应用范围需要扩展

目前数据挖掘对教务管理方面的应用研究大多是针对学生选课以及成绩方面,而教务管理的业务范围并不仅限于此,还包括招生、学籍、科研、毕业、就业等多方面。只有将数据挖掘技术全面应用于教务管理各个方面,才能更加充分地完善教务管理系统,提高教务管理水平。此外,还可以为教学中自动问答功能以及个性化推荐功能提供基础性支持。所以,研究数据挖掘算法,探索合适的挖掘模型,以处理教务各业务方面数据是今后努力的方向。

3.2 数据安全需要引起重视

在应用数据挖掘处理教务相关数据的过程中,处理的

数据主体主要来自教务系统数据库,数据源头包含学生和老师的个人隐私信息。针对个人信息安全保护问题,国家明确规定了相关法律条文并发布实施。所以在获取教务管理数据时,应该尽量避免敏感数据。如果研究中不可避免需要一些隐私数据,则需要对数据采取匿名化或去标识化等保密措施。将高度关注个人信息的敏感性,防止个人隐私泄露,防止对个人安全造成威胁作为对待数据的原则。

3.3 加强对教务系统的完善

大多学者通过对教务某业务方面的数据采用数据挖掘技术进行试验分析,最后根据实验结果获取数据与业务相关的规律,用以支持制定教务管理相关决策或提升教学水平。然而,研究的最终目标不能局限于此,应该在相关研究成果的基础上,结合教务系统实际使用情况对其功能进行升级完善,将研究结果转化为教务系统中相应的功能模块或是辅助教务管理的功能模块,由此使得数据挖掘针对教务管理的应用成果能更加直观地呈现于实际管理工作中。

4 结语

随着信息化的推进,高校教务管理系统不断升级、功能不断完善、业务不断规范,教务管理方面的信息数据量也不断增加。如何正确科学使用教务管理方面的数据是每所高校管理者甚至是高层必须深思的问题,因为这决定着学校如何优化教学资源、标准化教学管理、提升教学水平,甚至决定着整个教育行业的发展走向。数据挖掘技术为在纷繁复杂的海量教务管理数据中挖掘知识提供了很好的工具,虽然已经有一些学者就此方面进行了研究,但是数据挖掘并没有在高校教务管理方面得到广泛应用,因此仍有较大的发展和创新空间,有必要进行深入研究。

参考文献:

- [1] ABDULMOHSEN A. Data mining in education [J]. International Journal of Advanced Computer Science and Applications, 2016(6):456-461.
- [2] HAN J W, KAMBER M. 数据挖掘概念与技术[M]. 北京:机械工业出版社, 2007.
- [3] 陈颖. 数据挖掘在教务系统中的应用分析[J]. 电子世界, 2018(9):197.
- [4] 施立珊,李就娣,莫卓达. 聚类分析在高校教务系统中的应用研究[J]. 电脑知识与技术, 2016(13):18-20.
- [5] 任喜峰. 基于朴素贝叶斯分类的考试成绩监测机制研究[J]. 统计与决策, 2007(22):163-164.
- [6] 商俊燕,陆兵,柏倩然. 决策树 C4.5 算法在学生成绩分析中的应用[J]. 微型电脑应用, 2015(4):43-44.
- [7] 史英杰,鲁晓丽. 基于决策树理论的学生成绩分析系统模型构建[J]. 科技展望, 2015(29):290.

(编辑:叶璐)