

数据挖掘技术在教育领域的应用

郭 茹

(西安财经学院 统计学院, 西安 710061)

摘要:各大院校每年招收数千名学生,学习不同的课程。他们在招生时就向学生们收集信息并且储存在计算机上。懂得使用数据的好处对学校来说至关重要。数据可用于分类和预测学生将来的行为表现,辍学可能性等,对教师来说也是如此。因此,本文对数据挖掘技术在教育领域中的作用进行了探讨。此外,注重应用数据挖掘这一技术有助于学校提供更具有竞争力的课程给学生们,以提高他们的能力。

关键词:各大院校;教育领域;数据挖掘;应用数据挖掘技术

中图分类号: TP311.13 **文献标志码:** A **文章编号:** 1671-1807(2014)09-0121-04

近年来,数据挖掘引起了信息产业界的极大关注,其主要原因是存在大量数据,可以广泛使用,并且迫切需要将数据转换成有用的信息和知识。获取的信息和知识可以广泛用于各个方面,包括商务管理、生产控制、市场分析、工程设计和科学探索等。今天,许多行业使用数据挖掘来做出决策,并编制营销策略,做出客户细分以实现其商业目标。但很多大学忽略了数据挖掘技术。在教育领域,数据挖掘的应用是在全球竞争的形势下形成的新趋势。了解“数据挖掘技术”的任务、技术和应用是教育领域应用这一技术并获得发展的基础。因此,有必要研究数据挖掘技术在教育领域中的作用。

1 数据挖掘概念

数据挖掘是一个探索性的数据分析, Taniar 和 Rhaya^[1]定义数据挖掘为从数据中隐含的未知的潜在的和有用的信息中进行的非凡的提取过程。

数据挖掘的目的是将大量数据转化为有用的知识,从大量的数据里发现有用的模式,以确定重要的趋势和先前未知行为模式。总之,定义强调的重要性是数据挖掘作为通过将原始数据转化为有用的信息来建立模型的基本工具。教育部门可能会通过它获得有用的信息,有助于向学生提供更有用的课程和更好的学校环境。

2 数据挖掘过程

许多作者^[2]为数据挖掘制定了不同的步骤。但标准的数据挖掘过程所使用的主要模型是 CRISP-DM(跨行业标准数据挖掘流程)。主要包括商业理

解、数据解读、数据准备、建模、评估和应用。

2.1 商业理解

首先要在当前形势的评估下确定企业的目标,建立数据挖掘目标并在考虑业务目标的基础上确定项目计划。

2.2 数据解读

确定经营目标后,下一步就是数据收集和数据说明。从各种渠道收集数据并描述它的重要性,仔细检查数据质量并检查信息是否被遗漏。

2.3 数据准备

将收集到的数据通过筛选,不需要的数据被删除,用整理出来的数据进行建模和构建新数据。然后,数据被保存以备后续分析。此外,新数据可以根据需要并将其用于其他分析。

2.4 建模

用于测试数据有效性的方法例如分类、回归、聚类和关联规则等几个数据挖掘技术。贝叶斯、决策树、logistic 回归和支持向量机都是分类的工具。回归分析工具有线性回归、稳健回归和神经网络。分层和 k-均值聚类是工具,而解决他们所使用的工具是 Apriori。在这个阶段,模型用来进行评估设计的测试。

2.5 评估

使用上述方法和技术后,模型被建立并用来从业务目标角度对数据进行评估和分析。模型对企业目标的影响是我们必须理解并且进行再检查的,数据应用之前进行检查也避免了早期未被注意到的数据有

收稿日期:2014-06-20

作者简介:郭 茹(1987—),女,陕西礼泉县人,西安财经学院硕士生,研究方向:数据挖掘。

所遗漏。

2.6 应用

在这个阶段,报告和图形已经生成和被监测,进而做出基于数据制成的报告和图表的决议。

上述步骤是被许多行业认可和实行的。各个大学也可以申请 CRISP-DM 模型来实现其目标并实现盈利。

3 数据挖掘的优点

数据挖掘应用于金融、服务、银行、零售、医疗、制造、电信、食品工业、工程、电机行业,预测水,道路交通、制药、药物开发、求职者和保险行业,而教育部门还没有使用过。数据挖掘应用在教育部门的好处主要有以下几个方面:

- 1)确定学生对课程选择的需求和喜好及专业性的选择;
- 2)确定学生的发展趋势模式;
- 3)预测学生的知识,档次,最终结果;
- 4)支持自动搜索数据;
- 5)使构建学生的个人资料变得更加容易;
- 6)帮助管理层了解更多学生信息。

4 数据挖掘任务

数据挖掘被不同行业进行应用,预测、分类、关联和集群是数据挖掘的最重要的任务。

4.1 预测

Sen 等学者^[3]于 2012 年建立的模型对中等教育的测试采用敏感性分析(决策树算法,支持向量机,神经网络与物流回归。他们认为,决策树分析是最好的预测方式,其次是支持向量机和神经网络。Logistic 回归是正确率最少的预测。总之,数据挖掘可以应用于预测学业成绩、学生的知识分类、性别、年龄、职业、学分的多少和动力不足的学生。

4.2 分类

将数据分成等量的组及分组时所使用的绝对变量被称为分类。分类可分为两种类型:受监督和无监督。当物体或箱子是事先知道的被称为监督分类,而无监督分类装置中的对象或事件并不预先已知的。

4.3 关联

关联的关键是找到其相似性,发现它们之间有趣的关联或关系的过程,它对整个数据进行了总结。先验和 GRI 算法主要用于关联规则。有学者指出,关联规则是发现的 if-then 规则,这意味着如果找到了一个变量,另一个变量将具有特定值。例如,选择营销的学生也可以选择金融作为专业。此外,选择 M. Com 课程的学生也将选择 MBA 课程。

4.4 聚类

聚类是对相似的对象进行分组。Rajamani 等^[4]定义了聚类是对物理或抽象对象为一类相似对象进行分类的过程。与分类和预测不同,聚类分析数据对象,不考虑已知的类标号。一般地,训练数据中不提供类标号,因为不知道从何开始。聚类可以产生这种标号。对象根据最大化类内的相似性、最小类间的相似性的原则进行聚类或分组。即,对象的聚类这样形成,使得在一个聚类中的对象具有很高的相似性,而与其他聚类中的对象很不相似。所形成的每个聚类可以看作一个对象类,由它可以导出规则。聚类也便于分类编制,将观察组织成类分层结构,类似的事件组织在一起。

5 数据挖掘技术

这里对最常见的数据挖掘技术进行了讨论,它带来了各种技术、共同探索模式,并可以从数据库中构建模型。数据挖掘技术可以用来获得有用的信息或有趣的知识。感知有用性和感知易用性是影响一个人的意图使用数据的因素。七种类型的数据挖掘模型主要有:关联、分类、聚类、预测、回归、序列的发现和可视化。图 1 给出了数据挖掘技术。

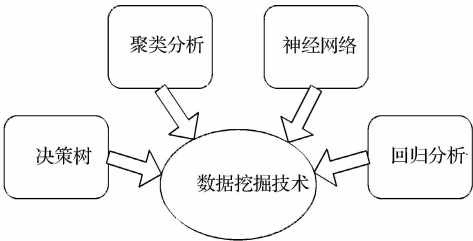


图 1 数据挖掘的技术

5.1 神经网络

这项技术可用于大型复杂的数据分类。它可以用来从课程角度选择学生,提高学生课程满意度,以及帮助学生选择专业。神经网络由神经元或“原型”来表示每个集群。输入的数据也代表了它的神经元与连接到原型的神经元。每一个这样的连接都有一个重点,这就是学会适应学习的过程。

5.2 决策树

决策树是一种可用于分类的大量数据和预测数据挖掘技术。决策树由“层次树”表示,其中每个叶指的是一个概念和包含概念的一个概括描述。它的节点通过分支连接,时间从左向右,每个分支代表一个决定或可能的事件。许多行业都用它来进行分类和预测客户行为。同样,高校可以利用这种技术来分类学生的表现、行为和期望。

5.3 回归分析

回归是一个强大的工具,可以预测性质变量之间和因果变量的可能值的关系。回归分析可以在数据挖掘中应用到预测学生的 GPA,对于专业的选择,投诉和满意度的期望,人口,地理,心理和行为变量。此外,简单线性回归和多元线性回归是回归分析中最常使用的技术。这些技术可以使用 SPSS 统计软件来完成。

5.4 聚类分析

客户分析指确定客户群具有相似特性,分割完整的数据使之成为一组集群,其中类别事先是不知道的。聚类分析可以用于制作标签,将这些对象都集中或分组,依据的原则是最大化类内和最小化类间的相似性。聚类也被称为分割,可以根据人口的统计变量(如性别,收入,年龄,学历,宗教,职业)、地理变量(如城市,国家,州,邮政编码,区)、心理统计变量(例如生活方式,个性和价值观)和行为变量(如福利)等进行细分。简言之,分类和回归是一种监督模型,而聚类分析是一种无监督模型,他们都用于将预测和分类的大数据转化为可用信息。因此,大学可以用聚类分析研究不同高校,学生,教师,行政人员,课程和考试。

6 数据挖掘在教育行业中的应用

数据挖掘在教育上的应用可以总结为四个方面:提高学生自身条件,发现知识结构,研究通过学习软件提供的教学支持与科学发现有关的学习和学习者。Dr. Mohd Maqsood Ali^[5]通过检测学生的参与和 LVS(视频直播)学习行为,发现学生使用聊天消息进行正面情绪,负面情绪和表达社会支持的传达。此外,聊天消息的有无和最终级的数量之间没有正相关关系。数据挖掘可用于报告和分析,帮助编制数据,有针对性的对学生进行策略营销。

此外,Pena 等人^[6]研究了应用数据挖掘技术在 web 教育领域的应用。以及在学生的建模,辅导,内容和评估领域的挖掘应用。他们认为,教育上的数据挖掘在帮助学生了解学习成果,寻找学生的行为和特征及对学生进行分类方面有巨大贡献。数据挖掘可以在教育部门的下列功能区域被应用。

6.1 学生报名预测

Aksenova 等^[7]在两类研究生间建立预测模型,预测国内的和归国留学生与其本科水平的比较。该模型基于人口,该地区的失业率,机构建设费和学费,家庭收入,过去招生院校的数据,数据挖掘与其他工具的帮助。他们的结论是数据挖掘可以很大程度地

应用于高等教育方面。

Kovacic, J. Zlatko^[8]是在入学学生数据(社会人口和环境变量)的基础上进行预测,使用如 CART 的数据挖掘技术。最后得出基于预注册信息对学生进行分类,有助于确定学生的学业下降过程,建议专业方向的选择,为他们提供咨询和指导方案,使他们学有所成的结论。

6.2 预测学生的剖析

Romdhane 等^[9]表明,数据挖掘可以建立客户模型描述他们的各种特定的习惯、需要和客户群行为。它为新客户预测其特殊需要。因此,数据挖掘可以根据入院时提供的学生信息帮助管理人员确定人口,地理和学生的消费心理等特点。神经网络技术,可用于识别不同类型的学生。此外,判别分析也可用于识别图形。回归分析,决策树和贝叶斯分类也可以应用。因此,聚类分析可以针对分段的学生进行不同的行为预测。

6.3 课程发展

Hsia 等^[10]的研究过程中针对喜好,完成率和注册专业使用数据挖掘算法,如决策树,关联分析。他们发现课程之间的相关性取决于登录者的专业类别。他们致力于研究在高等教育领域的课程建设和营销中数据挖掘的重要性。

6.4 学生投诉

Chen 等^[11]根据学生的投诉提出了“PARA”(P = 初步诊断, A = 高级诊断, R = 回顾与, A = 行动)模式的故障服务。数据挖掘技术来预测投诉类别和发展战略以改善服务。

6.5 结业

大学可以根据学生的满意度,学生的忠诚度和学生完成学业的程度了解学生对课程完成的程度。

6.6 学生确定目标

Woo 等^[12]定义的客户目标是“建设对特定客户战略的一个过程。”他们指出,可以用地图可视化方法进行客户定位。顾客地图有助于建立以客户为导向的战略。这是一个找到正确的目标客户的特征、价值和需求的方法。组织与客户定位的三个维度分别是:客户值(使用和行为),客户特征(人口和心理),以及客户的需求(投诉和满意度)。

6.7 图书馆设施

数据挖掘技术可用于图书馆,贷款账目,学生的报告,选择书籍,获得的信息。此外,聚类分析可以用于理解,例如,基于性别,年龄和学生成绩对书籍的选择和订货系统。

6.8 学生选课

各种因素影响学生选课和使用神经网络,如学生的课程量,课程成绩,课程类型,课程的时间和时间数,期末考试的时间和学生的需求。这些因素被用作神经网络的输入来进行建模。此外,分析并利用神经网络预测学生课程满意度,发现学生就读的课程和高分率、最终分级是最影响学生选课的满意度的因素。

6.9 教师的教学表现

采用逐步回归和数据挖掘技术的决策树可以预测导师的态度,员工状态,学生出勤率和学生对教学效能的反馈。

6.10 学生成绩

利用数据挖掘技术,Kumar 和 Uma^[13]研究了学生在使用过程中的表现,如朴素贝叶斯和决策树可以根据学生编号和标记。此外,他们认为数据挖掘过程可以对教师的教学水平进行分类,这有助于提高高等教育体系的水平,进而可以达到提高学生水平的效果。

6.11 学生辍学

Massa 和 Puliafito^[14]通过使用马尔科夫链——一个新的研究大学辍学问题的数据挖掘技术。他们的结论是同质群体的行为可以通过马尔可夫链进行研究。它可以被用来寻找集群中具有不同漏失风险程度相关联的学生。

6.12 学生关系管理

数据挖掘技术是管理学生的重要工具。数据挖掘技术可以用来寻找和支持学生中能够达到的获利能力的系统。Maqsood 将 CRM 定义为一种商业战略的整合,整合过程中,技术通过获取数据来提供最有利益和无利益的客户增值服务,并通过提供客户忠诚度计划来保持与客户间更好的关系。

7 结论

本文探讨数据挖掘在教育方面的应用。数据挖掘在一个行业内可以有各种应用,例如预测和分类的客户和聚类客户的特征并准备营销策略,通过对客户的细分来实现盈利。同样,大学也可以应用数据挖掘技术,预测学生从入学到各种课程的选择。数据挖掘可通过对学生进行分类和聚类来进行特征人口统计,心理和行为变量的预测。数据挖掘也可以通过使用 IF-THEN 规则来应用。此外,它可以根据 GPA 的高低在学生中分布来预测学生的学业完成情况。它也可用于对辍学学生,学生的学业成绩,教师的表现,以及学生的投诉率的预测。

参考文献

- [1] TANIAR DAVID, RAHAYA T WENNY. Parallel Data Mining in (ed.) Data Mining: A Heuristic approach[M]. Abass Hussein, Ruhul Sarker and Charles Newton, Idea Group Inc, 2002; 261—289.
- [2] GIUDICI, PAOLO. Applied Data Mining: Statistical Methods for Business and Industry[G]. John Wiley and Sons Ltd. , 2013.
- [3] SEN, BAHA, UCAR, et al. Predicting and analysing secondary education placement-test scores: A data mining approach [J]. Expert System with Applications, 2012, 39: 9468—9476.
- [4] RAJAMANI, KARTHICK, SUNG, et al. Extending the applicability of association rules[C]//The proceedings of Third Pacific Asia Conference, PAKDD 99, Beijing, China, April 26—28, 1999.
- [5] DR MOHD MAQSOOD ALI. Role of Data Mining In Education Sector[J]. International Journal of Computer Science and Mobile Computing, 2013(2): 374—383.
- [6] PENA ALEJANDRO, DOMINGUEZ R, et al. Education Data Mining: A sample of review and study case[J]. Academic World and Education and Research Center, 2009(2): 118—139.
- [7] AKSENOVA S STRETLANA, ZHANG DU, et al. Enrolment prediction through Data Mining[C]. IEEE Conference Publications, 2006; 510—515.
- [8] KOVACIC J ZLATKO. Early prediction of student success: Mining students' enrolment data[C]//proceedings of informing science and IT education Conferences (InSITE). 2010: 647—665.
- [9] ROMDHAE L B, FADHEL N, AVEB B. An Efficient approach for building customer profiles from business data[J]. Expert System with Applications, 2010, 37: 1573—1585.
- [10] HSIA, TAI-CHANG, SHIE, et al. Course Planning of extension education to meet market demand by using data mining technique: an example of Chinkuo Technical University in Taiwan[J]. Expert System with Applications, 2008, 34: 596—602.
- [11] CHEN SHIE AN-JIN, YU CHANG-HIS. A Customer-oriented Organisational diagnosis model based on data mining of customer complaints databases[J]. Expert System with Applications, 2012, 39: 786—792.
- [12] WOO JI-YOUNG, BAE SUNG MIN, et al. Visualisation method for customer targeting using customer map[J]. Expert System with Applications, 2005, 28: 763—772.
- [13] KUMAR A N V, UMA G V. Improving academic performance of students by applying data mining techniques[J]. European Journal of Scientific Research, 2009, 4: 526—534.

(下转第 139 页)

- 心理科学进展,2012,20(3):344—353.
- [28] 徐芬,李欢,马凤玲. 大学生信任倾向与人格特质的关系[J]. 应用心理学,2011,17(3):202—211.
- [29] 李晔,刘华山. 中小学生对教师的信任及问卷编制[J]. 心理发展与教育,2007(4):88—94.
- [30] 井维华,张庆强. 青少年学生的人际信任度与家庭因素的相关研究[J]. 心理学探新,2006(2):61—64.
- [31] 于少勇,潘峰. 篮球运动员的互依性与合作绩效的关系:团队信任的中介作用[J]. 中国体育科技,2010,46(6):8—14.
- [32] 叶鹏,潘峰,于少勇. 篮球运动员对教练员的信任与训练比赛满意度的关系[J]. 沈阳体育学院学报,2010,29(5):34—41.
- [33] 于少勇,韩践,谢卫平. 篮球运动员对队友信任的测量及与相关变量的关系[J]. 体育科学,2009,29(2):30—38.
- [34] VATHSALA WICKRAMASINGHE V, WIDYARATNE R. Effects of Interpersonal Trust, Team Leader Support, Rewards, and Knowledge Sharing Mechanisms on Knowl-

edge Sharing in Project Teams[J]. The Journal of Information and Knowledge Management Systems,2012,42(2):214—236.

- [35] HONG I B, CHA H S. The Mediating Role of Consumer Trust in An Online Merchant in Predicting Purchase Intention[J]. International Journal of Information Management, 2013,33(6):927—939.
- [36] JOANNE D, LECK PENNY M WOOD. Forming Rust in E-Mentoring: A Research Agenda[J]. American Journal of Industrial and Business Management,2013,3(1):101—109.
- [37] OVERBYE MARIE BIRCH, WAGNER ULRIK. Experiences, Attitudes and Trust: An Inquiry into Elite Athletes' perception of The Whereabouts Reporting System[J]. International Journal of Sport Policy and Politics,2013(ahead-of-print):1—22.

Progress in the Field of Interpersonal Trust in China

——Based on Econometric Analysis for CSSCI Journals from 2001 to 2012

CHEN Zhong-wei, WU Li-huan, DAI Yun-yun

(Center of Entrepreneurship and Business Growth, Anhui University of Finance and Economics, Bengbu Anhui 233030, China)

Abstract: Using the 232 papers on the interpersonal trust which published in CSSCI journals during the period 2001—2012 as samples, then we analysis the annual, discipline, periodical and region and find that: the research of interpersonal trust in China gradually becomes the hotspot, the literature number on this research are increasing more and more. Although the time of this research is not very long, we are continuing to narrow the gap with the abroad. Besides, based on the different subjects of trust, we try to analysis and look forward four kinds of relationship which are bosses and subordinates, teachers and students, network consumers and businesses and athletes and coaches.

Key words: interpersonal trust; econometric analysis; progress

(上接第 124 页)

- [14] MASSA S, PULIAFITO P P. An application of data mining to the problem of the university students' dropout using Markov chains[C]//Zytknow J M, Rauch J (Eds.), Principles of Data mining and Knowledge Discovery, third European conference, PKDD 99, Prague, Czech Republic, September 15—18, 1999, proceedings, Springer, vol. 1784, 51—60.
- [15] 毛国君,等. 数据挖掘原理与算法[M]. 北京:清华大学出版社,2005.
- [16] 徐保民,杨铨玖,等. 数据库系统原理与应用[M]. 北京:清华大学出版社,2005.

- [17] 周勇. 数据挖掘技术发展综述[J]. 中国科技信息,2005(16):35.
- [18] 刘雨露. 数据挖掘在高校学生管理决策中的应用模式分析[J]. 成都信息工程学院学报,2006,21(3):373—377.
- [19] 丁元明. 数据挖掘技术在高校教学质量评估中的应用研究[D]. 上海:华东师范大学,2005.
- [20] 白雪,段富. 决策树分类算法的研究及其在教学评估中的应用[J]. 电脑开发与应用,2007(2):24—26.
- [21] 常桐善. 数据挖掘技术在美国院校研究中的应用[J]. 复旦教育论坛,2009,7(2):72—79.

The Application of Data Mining Technology in the Field of Education

GUO Ru

(School of Statistics, Xi'an University of Finance and Economics, Xi'an 710061, China)

Abstract: Each year thousands of students enrolled in major institutions, learning different courses. Gathering information to the students on admissions and stored in the computer. Understand the benefits of using the data is very important for the colleges. Data can be used for classification and prediction the future behavior of the students, the possibility to drop out of school, etc. for teachers, too. Therefore, this article discussed that the role of data mining technology in the field of education. In addition, pay attention to the application of the data mining technology will help colleges to provide more competitive courses for the students, in order to improve their abilities.

Key words: major institutions; the field of education; data mining; the application of data mining technology