Hans Journal of Data Mining 数据挖掘, 2019, 9(1), 1-7
Published Online January 2019 in Hans. http://www.hanspub.org/journal/hjdm
https://doi.org/10.12677/hjdm.2019.91001

Hans汉斯

Research Summary of Data Mining in University Information Management

Xi Zhang, Rui Fang

Chengdu University of Information Technology, Chengdu Sichuan
Email: 1024032707@qq.com, fangrui@cuit.edu.cn

Received: Dec. 18th, 2018; accepted: Jan. 2nd, 2019; published: Jan. 9th, 2019

Abstract

With the continuous development of data mining technology, significant achievements have been made in many fields. In recent years, the information management of colleges and universities has become a hot topic in the field of education. Data mining technology has been widely applied to the information management of colleges and universities. Beginning with the hotspot technology of data mining, this paper deeply studies the role of data mining technology in the transformation of university management from artificialization to informationization and related technology applications, so as to find more valuable research directions.

Keywords

Data Mining, University Information Management, Hot Technology, Artificialization, Informatization

数据挖掘在高校信息化管理中的应用研究综述

张　茜，方　睿

成都信息工程大学，四川　成都
Email: 1024032707@qq.com, fangrui@cuit.edu.cn

收稿日期：2018年12月18日；录用日期：2019年1月2日；发布日期：2019年1月9日

摘　要

随着数据挖掘技术的不断发展，在很多领域都取得了显著的成就。近几年，高校的信息化管理成为了教育领域研究的热点，数据挖掘技术被广泛的应用到高校信息化管理当中。通过数据挖掘这一热点技术入手，深入研究探讨数据挖掘技术在高校管理从人为化转变为信息化中起到的作用以及相关的技术应用，

文章引用: 张茜, 方睿. 数据挖掘在高校信息化管理中的应用研究综述[J]. 数据挖掘, 2019, 9(1): 1-7.
DOI: 10.12677/hjdm.2019.91001

从而发现更有价值的研究方向。

关键词

数据挖掘，高校信息化管理，热点技术，人为化，信息化

1. 引言

先进科学技术信息化的迅速发展，让各个研究领域数据都呈现出爆炸式的增长，包括淘宝交易数、银行交易数据、医疗数据、教育行业数据以及公司利润和业绩以及顾客反馈等[1]。在这些庞大的数据中隐藏着许多有价值的消费规则或者信息，而这些信息对各个领域的发展和管理具有重要的意义，因此如何挖掘出更有价值的信息则成了当今关注的热点话题[2]。数据挖掘技术的出现就是为了解决这些庞大的数据。数据挖掘是数据库研究、开发最活跃的分支之一[3]，可以说数据挖掘技术在国内的发展已经越来越成熟和强大。2017 年的 KDD 大会在加拿大新斯科舍省的首府哈利法克斯落下帷幕。纵观几十年的 KDD 发展，从参加会议的人数，论文发表影响力等各种情况来看，华人力量越来越强大，这说明我国现在也越来越重视数据挖掘的发展，相信这项技术能为科技的进步带来更加重大的发现。并且和我们在校学生密切相关的教育领域，数据挖掘技术也正在深入。近几年，随着进入高校学习人数的不断上升，摆脱传统的人为管理方式，使得管理信息化，成为了各个高校研究的热点。目前数据挖掘技术在一些高校管理中的应用很广泛并且也取得了一些成果，例如：在马来西亚以及印度的一些高校利用数据挖掘技术解决学生心理健康问题[4]；Brigitte Maier 利用数据统计方法分析了澳大利亚的 29 所大学的 2196 名学生在网络学习平台的经验和喜好[5]；Cristóbal Romero, Sebastián Ventura, Enrique García 以 Moodle 系统作为案例研究了数据挖掘技术在课程管理系统中的应用[6]；浙江大学使用关联规则发现挖掘技术对高校的人事信息库的作用，试图找到影响学科发展的因素[7]等。就这么多的案例来看，知道高校学生管理工作不仅需要经验指导，也需要科学引领[8]。如果能有效的将数据挖掘技术应用到高校信息化管理当中，不仅能够找到对学生学校发展最有利的数据资源，还能够帮助学校更好的了解学生的生活学习情况，实现更好的管理，提高学校管理质量。本文就将从几个方面入手，讨论数据挖掘技术在高校信息管理中的相关应用。

2. 数据挖掘

2.1. 数据挖掘的概念

随着大数据时代的蓬勃发展，数据挖掘技术在其中起着关键的作用，结合目前的发展形势，数据挖掘已经成为人工智能和数据库领域的研究热点，数据挖掘的意义是为了发现数据库中隐含的知识[9]，解决庞大数据集带来的问题，利用数据挖掘技术可以分析大量的数据，找到有利于社会发展的信息。

数据挖掘是一门结合了多个学科知识的产物，它包含了许多的学科知识，这些学科现在都是各部门的研究热点，主要包括人工智能，机器学习，模式识别，统计学，可视化技术等等。数据挖掘能够解释成为一个过程就是自动地帮助我们分析数据并从中得到数据中潜在知识，从而帮助决策者做出合理并且正确的决策[10]。

2.2. 数据挖掘的过程

数据挖掘的过程主要包括以下几个方面：数据集选取或构造、数据预处理、数据转换、数据建模以及对结果的分析与评价几个过程。

① 数据集选取或构造：大量数据都是由不同的数据源构成的，解决一个数据挖掘问题，首先要进行数据集的选取或者构造这一步骤，根据任务的目的和自己的需求选择合适的数据集或者是构造出需要的数据。

② 数据预处理：数据预处理是数据挖掘的关键步骤，我们收集到的数据类型、格式是多种多样的，并不是所有数据源都满足我们的实验需求，导致结果存在误差，不准确，所以就需要进行预处理。数据预处理包含数据清洗、数据集成、规范格式化数据，这个过程通俗来讲就是将“脏数据”转变为“干净的数据”从而为我们所用。

③ 数据转换：数据转换就是将上面处理后的数据转换为特征，这些特征要尽可能准备的去描述数据，并且可以使得数据挖掘算法达到最优。

④ 数据挖掘：数据挖掘的核心是模式发现[11]，也就是利用数据挖掘算法和数据挖掘工具对转换过后的数据进行分析，挖掘出我们预期想要的结果。

⑤ 结果分析与评价：当根据数据挖掘算法得到结果的时候，对得出的结果和理论的结果进行分析，得出相应的结论。

综上所述，将数据挖掘的过程表示出来如图 1：

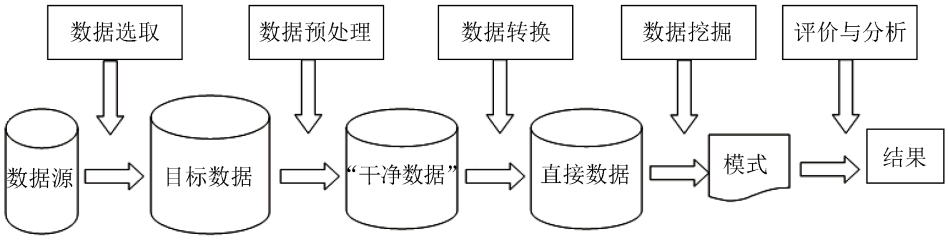

Figure 1. Data mining process
图 1. 数据挖掘过程

2.3. 数据挖掘的常用方法

在数据挖掘发展的进程中，积累下来的数据挖掘的方法有很多，每一个项目都有适合的数据挖掘方法，每种方法都有自己对应的优点和适合的领域，下面简单介绍几种数据挖掘算法。

① 分类方法：分类就是找到数据库中一组数据对象的相同特点并且按照分类模式将其划分成不同的类，其目的就是为了通过分类模式将数据库中的数据映射到某个给定的分类模式中。目前使用最为广泛的分类算法就是决策树算法。常用的领域有广告点击行为的预测。用户在上网的过程中会浏览许多的网页，根据点击网页的类型，网站可以向不同的用户投放不同的广告，产生更大的收益。

② 关联规则方法：关联规则是数据挖掘算法中运用最广泛也是最重要的方法之一。关联规则是形如下 $X \rightarrow Y$ 的一种蕴含条件式，也就是发现变量之间存在的关系和规律。例如顾客和购买东西这两者之间就体现了关联规则，可以通过顾客购买东西的记录找到不同东西之间存在的联系。

③ 聚类方法：聚类的意思就是把一组数据按照他们之间存在的相似和差异来分成几种类别，能够让同一类别的数据之间相似性大，让不同类别数据之间的相似性小。例如在市场细分领域中，根据同一商品，不同顾客有不同的消费特点，研究这些特点根据聚类算法制定出不同的销售组合应用在市场中。

④ 回归分析方法：回归分析是一个统计预测模型，用来描述和评估一个或多个自变量之间的关系，反映的是事务数据库中属性值在时间上的特征，产生一个将数据项映射到一个实值预测变量的函数，发现变量或属性之间的相互依赖关系。比如研究产品质量和用户满意度之间的关系，在两者之间建立线性回归关系，体现出质量的提升对用户满意度带来的影响。

3. 数据挖掘技术在高校信息化管理中的研究

如今随着数据挖掘技术的快速发展，将数据挖掘技术应用到高校信息化管理中已经成为了各大高校的趋势，所谓的高校信息化就是指将计算机的各种技术与学校教务教学管理融合在一起，从而提高了学校管理人员的管理效率。下面主要从数据挖掘在学生选课、学生就业、贫困生认定、学生成绩预警四个方面进行应用分析。

3.1. 数据挖掘技术在学生选课方面的研究

学生的选课问题一直是被忽视的问题。在多数高校普遍存在课程设置不合理、选课方式不完善、选课指导体系不健全等问题。加上学生经常盲目的去询问高年级学生的选课情况，往往忽视自己的兴趣以及优势，这样不利于高校对人才的培养。数据挖掘技术能够更好的解决这个问题，利用数据挖掘技术中的关联规则算法，关联规则就是发现变量之间存在的某种联系，大家最熟悉的关联规则体现就尿布与啤酒的关系，那么这种表现同样可以应用在学生选课当中，发现学校课程之间的相关性。比如一个频繁项集是{计算机科学基础，计算机网络，网络编程}，那么由此可见，如果该同学计算机基础学的很好很有兴趣，那么他的计算机网络和网络编程也是没有问题，而且可以为学校的排课提供一个参考，在学习有争对性的专业课的时候，需要学习相关的基础课程，更有利于学生更好的学校专业课知识。这样就可以为学生选课提出一个更有科学性的参考，避免盲目性选课为学生以后毕业或找工作带来困扰。

3.2. 数据挖掘技术在学生就业情况方面的研究

学生就业问题一直以来是学生也是家长和学校最为关注的问题，目前我国部分高校已经对学生就业系统进行开发并且也投入使用。例如上海交通大学毕业生就业办公室的开发和应用[12]，功能很强大，速度很快，为学校的就业管理提供了帮助。其实每个学校的信息数据库中，收集了学生的各种信息，也收藏了许多招聘单位的条件信息等。那么对高校来说最重要的一点就是如何从这些海量的数据中发觉有用的信息和数据，能够对人才培养方案的制定，课程的设置，实验实训的内容设计，以及如何开展毕业生的就业指导都有着显著的意义。利用数据挖掘技术中的关联规则算法，通过设置一些对就业有影响的关键词集，首先通过计算找到能够匹配最小支持度值的频繁词项，再通过匹配最小置信度来生成关联规则得出强关联规则就能够分析出各种对就业有影响的因素，找到这些因素之后，比如学生的学位课成绩，学生的城市经济状况等，学校可以根据这些因素进行一个参考，为学生找工作提供宝贵的意见，提高学校的就业率。

3.3. 数据挖掘技术在贫困生认定方面的研究

高校教育的普及给许多农村的孩子带来了方便，当然，也存在着读书困难的问题。我国也针对贫困生进行了帮助，我国已经初步形成贫困生资助体系，建立了“奖、助、补、贷、减”相结合，以国家奖学金、国家励志奖学金、国家助学金、国家助学贷款、生源地助学贷款为主的家庭经济困难学生资助体系[8]。我们学校也是积极的在为贫困生提供帮助。但是怎样将资金帮助真正贫困的同学，成为了高校现在需要解决的问题。目前大多是通过同学之间相互评定和老师的讨论进行资助，这里面存在很多人为因素，可能导致需要帮助的同学得不到帮助。那么贫困生认定系统就可以很好的解决这个问题，利用数据

挖掘技术能够排除其他因素干扰，使得调查的数据更加合理，结果更可信。之前也有学者做过相关的研究，例如：赵炳起[13]等提出建立包括五类指标的评定指标体系；杨晴[14]在硕士论文中提出了评价学生家庭经济困难程度的五类指标；刘善槐、邬志辉[15]使用线性回归方法构建了困难生认定二分类模型等。这些都是利用各种合理的数据挖掘算法，然后根据采集到的有用的数据，对其进行分析，挖掘出其中隐含联系[16]，那么根据这些研究我们可以从多方面入手，比如收集关于校园一卡通的消费数据。在校园一卡通中包含大量的消费记录，这是最好反映学生经济情况的数据；其次还有学生填写的家庭情况表，包含学生的家庭情况信息，提供一个辅助作用，将这些数据进行收集整理后，可以利用数据挖掘中的关联规则算法来设置最小支持度和最小置信度，通过计算找出超过最小置信度的所以数据，来和研究的数据进行对比，将学校的贫困生挖掘出来。也可以利用决策树算法，通过计算每一个属性的信息增益值，找到具有最大信息增益的值，作为根节点，依次构造成一棵决策树，可以直观的看出对贫困生认定最有用的属性。由此可知，利用数据挖掘算法能够为贫困生认定提供更加直观有效的结果，帮助学校更好的解决这一问题。

3.4. 数据挖掘技术在学生学业预警方面的研究

由于高校教育的普及，学生人数持续的增长，带来的数据也越来越多。目前大多数学校针对学生产生的庞大的数据仅仅采用简单的查询，统计等手段，很少挖掘数据隐藏的含义，因此我们可以利用数据挖掘技术挖掘出学生各科成绩之间隐含的关系，来起到提醒和警示学生学业的作用，能够更好的帮助学生为自己的学业做好准备。其实部分学校已经做出了一些措施，比如：上海海事大学教务处建立了严格的预警规则[17]，对学生发放预警数据，对那些存在问题的同学提出严肃的正式的提醒，督促学生能够更好的完成下面的学习任务，顺利毕业；复旦大学为了督促学生主动学习，在校园里实施了“博学计划”，构建了完善的学习预警体系，对于绩点低的学生进行提前干预[18]。我们可以利用数据挖掘技术中的分类分析的方法，对大一大二学生的成绩进行分析，找到容易引起挂科的科目，和存在学分低无法毕业的学生，进行预警，对这部分学生做出一些成绩提示，能够帮助他们更好的重视自己的成绩，顺利毕业。

4. 实例——贫困生认定系统的设计

4.1. 数据挖掘技术应用于贫困生认定系统的重要性

由于目前高校人数越来越多，那么随之增加的贫困生的数量也越来越多，如何公平公正的为学生做出贫困生等级判断成为目前高校需要解决的一个问题。但是由于目前大多数高校的认定贫困生方法存在一定的偏差，多数是采用人为评定的方法，那么学校可以采取一定的措施，通过深层次的挖掘大量学生数据中与贫困生评定相关的数据，提取出相关的规则来为学校管理人员提供贫困生认定的帮助。

4.2. 贫困生认定系统的设计方案

在贫困生认定的过程当中采用数据挖掘技术，将数据挖掘技术应用于一卡通消费记录的关联以及家庭信息和学习情况的判断的过程中，为参与评定的人员提供帮助。

4.2.1. 设计流程

① 数据的选取与收集：可以向学校相关负责部门收集数据，并自己通过对学生做出一些调查来收集更多数据。将收集的数据进行整合选取，选择更有用，更规范化的数据。

② 数据处理：根据选择好的规范化的数据，根据所需要的数据挖掘的算法将其进行离散化化的处理。

③ 数据挖掘：选择合适的数据挖掘算法以及数据挖掘工具，对之前处理的数据进行挖掘，找到对方案有用的信息和规则。

④ 设计系统：选择合适的程序开发语言，设计一个贫困生认定系统，便于更直观的展示结果。

⑤ 应用：将通过数据挖掘技术产生的结果应用到贫困生系统当中，并通过系统展示出来。

综上所述，方案设计流程如下图 2：

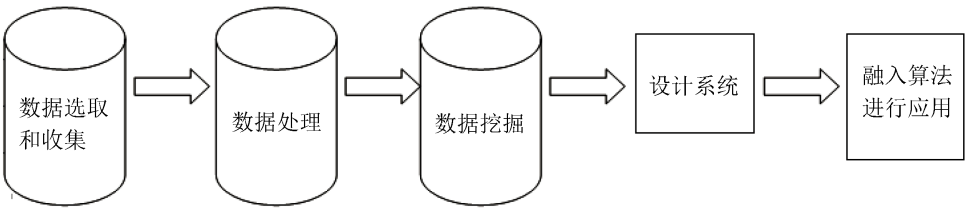

Figure 2. Scheme design flow chart
图 2. 方案设计流程图

4.2.2. 数据来源

① 首先是一卡通数据：根据第三部分的几个研究来看，发现就高校管理问题来说，一卡通已经成为了研究的热点和必不可少的一个部分。原因是校园一卡通实现人工辨别与电子认证结合的功能[19]；可以作为校内识别证件，准确认证身份，提高学校行政管理水平[20]，因此校园卡中包含着大量有价值的数据，比如：校园一卡通卡面上有个人的姓名、性别、学号、院系、照片等信息；校园卡还包括门禁系统的管理数据；校园卡包括校园消费数据；还有校园网上网的数据，这在之前提过的分析学生的学习行为起到了一定作用；有些学校校园卡也代替了水卡电卡的功能。

② 除了一卡通数据，对于贫困生这一问题来说还有其他可用的数据，包括学生填写的个人以及家庭情况表、学生的综合成绩、学生申请贷款的情况，以及学生的德育分情况等。

4.2.3. 解决方法

① Apriori 算法来分析。该算法通过很多次的扫描数据库来得到项集从而找到我们需要的规则，那么利用 Apriori 算法应用到贫困生系统中[21]的主要原理是从众多影响因素中选出几个对贫困生认定最有价值的项，然后将里面的数据进行离散化处理，设置置信度之后，再利用算法找出这些项之间的关联性进行分析，最终得到我们需要的结果；

② 支持向量机算法来分析。贫困生问题也是一个分类问题，那么支持向量机算法能够很好的将贫困生和非贫困生分割开来，达到最大的分割操作，在收集到的所有数据中提取训练集，然后对其进行支持向量机的测试，解决问题。

③ 遗传算法来分析：学生数据库，以及校园一卡通中的信息每天都在不断的变化，如果需要进行长期分析观察结果，那么就应该采用优化之后的算法，利用遗传算法的聚类，分类分析可以解决这个问题。

4.2.4. 系统设计

在设计系统的时候，需要选择一个合适的框架，B/S 框架[22]比 C/S 框架更适合，B/S 因为是直接访问网页的，所以更容易理解和操作，设计也没有那么困难。然后需要前台后台的页面，需要 web 服务器和数据库服务器相结合。需要选择合适的开发语言和数据库环境来进行开发，最后完成合适的系统设计，为贫困生的认定工作带来更大的方便。

5. 展望

综上所述，数据挖掘技术在各个领域的发展都越来越快，我们作为学生，应该更加注重高校信息化管理这一块的应用。利用数据挖掘中的各种方法，设计出合理的系统，帮助学校管理人员摆脱传统管理

方式，合理利用一些看似不起眼的数据，挖掘出数据背后潜在的含义，能够使得管理更加科技化，信息化[23] [24]。这样不仅给学校的教学管理带来方便，有利于更好的培养学生，对于学生来说也能帮助学生更加顺利的毕业，真正的实现高校校园信息一体化工作，实现人才培养的宗旨。

参考文献

[1] 张红蕾. 数据挖掘在校园卡消费中的研究与应用[D]: [硕士学位论文]. 兰州: 兰州交通大学, 2016.

[2] 陈卓民. 数据挖掘技术在国内外的研究和发展现状[J]. 青年文学家, 2009(16): 122-123.

[3] 胡春红. 数据挖掘技术在高校信息化系统中的应用[J]. 长江大学学报(自科版), 2010, 7(2): 274-276.

[4] Paechter, M. and Maier, B. (2010) Online or Face-to-Face? Students' Experiences and Preferences in E-Learning. *Internet & Higher Education*, **13**, 292-297. https://doi.org/10.1016/j.iheduc.2010.09.004

[5] 孙云帆, 齐美玲. 数据挖掘在教育应用中的浅析[J]. 商场现代化, 2012(24): 167-168.

[6] 陈丽, 陈根才. 基于高校人事信息库的数据挖掘研究[J]. 计算机工程, 2000, 26(11): 117-119.

[7] 陈新中. 大数据时代高校学生管理工作信息化建设分析[J]. 科技风, 2016(7): 79-79.

[8] 陈新宇, 林长英. 某高校校园一卡通应用分析[J]. 吉林医药学院学报, 2014, 35(1): 48-49.

[9] 廉文武. 数据挖掘下贫困生认定辅助系统设计研究[J]. 当代教育实践与教学研究: 电子版, 2016(7): 47-48.

[10] 王梦雪. 数据挖掘综述[J]. 软件导刊, 2013, 12(10): 135-137.

[11] 陶双红. 决策树关联规则算法在高校贫困生评定管理中的应用[D]: [硕士学位论文]. 长沙: 湖南大学, 2013.

[12] Balbi, S., Misuraca, M. and Spano, M. (2016) A Cosine-Based Validation Measure for Document Clustering. JADT 2016: 13ème Journées internationales d'Analyse statistique des Données Textuelles.

[13] 李永宁, 赵炳起. 高校贫困生经济资助绩效的模糊综合评价模型[J]. 统计与决策, 2007(11): 44-45.

[14] 杨晴. 中国高校贫困生贷款资格判定[D]: [硕士学位论文]. 武汉: 华中科技大学, 2005.

[15] 刘善槐, 邬志辉. 高校贫困生评价体系与界定模型研究[J]. 高教探索, 2010(5): 115-117.

[16] 王鑫家. 大数据思维在高校学生信息化管理中的支撑作用[J]. 黑龙江高教研究, 2016(7): 47-50.

[17] 柏美屹, 罗颖. 中部地区高校选课制实施现状研究——以三所"211 工程"大学为例[J]. 科技致富向导, 2013(11): 55, 104.

[18] 郭佳. 数据挖掘技术在高校学生就业信息管理系统中的应用研究[J]. 桂林师范高等专科学校学报, 2015, 29(3): 148-150.

[19] Chang, C.I. (2016) Recursive Band Processing of Fast Iterative Pixel Purity Index. *SPIE Commercial + Scientific Sensing and Imaging*, 98740G.

[20] 罗华群, 易国平. 校园一卡通数据的挖掘与应用[J]. 科技信息, 2010(1): 41-41.

[21] 郑丹. 数据挖掘技术在高职院校贫困生认定中的应用[D]: [硕士学位论文]. 合肥: 安徽大学, 2016.

[22] Hao, S., Qiao, Y., Hu, Q., *et al.* (2016) Handbill Release System Using B/S and C/S Hybrid Framework. *Control and Decision Conference*, IEEE, 6789-6793.

[23] 黄文静. Apriori 算法在高校毕业生就业数据挖掘中的应用研究[J]. 电子技术与软件工程, 2015(4): 207-208.

[24] 王凯成. 基于数据挖掘的大学生学业预警研究[D]: [硕士学位论文]. 上海: 上海师范大学, 2012.