

# 数据挖掘技术下数字化校园系统的分析与研究

刘建花

(晋中师范高等专科学校数理科学系, 晋中 030600)

**摘要:** 随着科学技术的发展, 数据挖掘技术的应用越来越广泛。文章首先介绍了数据挖掘技术下数字化校园系统的重要性, 其次阐述了系统的构建, 其中包括总体技术架构、功能设计和关键技术, 最后重点分析了数字化校园系统中的成绩管理系统的评估模型。

**关键词:** 数据挖掘; 数字化校园; 决策树

doi: 10.3969/J.ISSN.1672-7274.2020.03.015

中图分类号: TP3

文献标识码: A

文章编号: 1672-7274 (2020) 03-0032-02

## 0 引言

数据挖掘技术在数字化校园系统中的作用主要是通过分析高校中各种类型的大量数据为高校制定各项决策提供数据依据, 以此来不断提高学校教学、管理、服务的各项水平。

## 1 数据挖掘技术下数字化校园系统中的重要性

### 1.1 实现智能化教学

在高校的数字化教学中, 学生可自主式学习, 与老师、同学互动留言, 提高学生对时间的利用率。智能化教学将每个学生的学习过程进行记录, 教师根据学生的学习情况进行合理的课堂设计, 对学生针对性的教学和辅导。

### 1.2 提高学校管理效率

在数字化校园系统中, 学校将网站中的一些留言、发表的稿子以及一些评论的数据信息分析处理后, 对教师的教学效果进行预测, 对教学活动进行更好地管理和服务。

### 1.3 实现智能化服务

系统将用户数据和资料进行管理和整合, 用户登录平台对身份识别验证后获取相应的使用权限, 享受其对应的功能服务。

## 2 数据挖掘技术下数字化校园系统构建

### 2.1 总体技术架构

在数字化校园中, 从上至下由四个方面进行建构: 一是业务系统应用层。有校教务管理系统、科研管理系统、网络教学和成绩管理系统以及档案管理系统等。二是平台管理层。利用数据挖掘技术, 将应用层的数据进行收集、筛选、备份、分析、整理, 让数据系统化。三是网络综合层。有校园网、移动网和物联网, 利用无线技术和 IP 技术, 将网络信息进行全面覆盖, 收集更多数据。四是感知层。通过 GPS、RFID 以及传感器等设备对校园中的信息进行采集、处理和筛选, 然后传输到信息平台中。

### 2.2 系统功能设计

借鉴其他学校数字化校园系统构建的经验, 将管理、科研、教学、安全、生活等方面结合, 为教师和学生构建数字化的校园环境。

(1) 校务和办公自动化。整个数字化校园的重要环节, 所有业务都在系统上完成, 包括教职工任务, 领导审批, 有效提高效率, 真正实现办公无纸。

(2) 智能教学。教师的教学计划、学生的考试、阅卷、成绩的录入等环节都在系统上进行。

(3) 教研查询。教师可以将科研成果, 包括科研论文、教案、教学经验的总结以及教学视频等资源上传到系统进行展示和交流, 相互分享。在课题管理系统中, 方便申报课题与中期检查, 教研人员节省了报告的时间。

(4) 智能平安校园。利用物联网技术, 植物灌溉和养护、照明、门禁、监控以及报警实现智能化, 既节省费用, 也提高效率。

(5) 智能校园生活。采用无限射频识别技术, 学生只需要一

部手机, 实现电子签到、超市购物、费用缴纳等功能。

### 2.3 关键技术

数字化校园构建中用到的技术和数据挖掘算法:

(1) 云计算: 主要对虚拟化的资源进行分配, 对数据进行整理和量化。

(2) 物联网技术: 将数据信息、互联网技术以及远程操作等技术结合起来, 实现智能化。

(3) 移动互联网技术: 提供安全的网络环境, 实现对人员的动态管理, 使得生活更加便捷。

(4) 决策树方法一般用于将数据分类, 由构造树和修剪树两个过程。先开始使用初始数据生成测试函数, 依据所得到的值生成树的分支, 重复步骤, 再各个分支上再生成下一层的分支和结点, 这样生成的树称之为决策树, 然后再对决策树进行修剪, 最后生成规则。该算法有 ID3、IBLE 等。

## 3 数字化校园系统中成绩管理系统的评估模型分析

本校学生最终考试成绩由平时成绩和期末成绩决定, 平时成绩占 40%, 期末成绩占 60%, 考试形式有笔试、机试、论文、表演等, 考试方式有考查和考试两种。考完后由代课老师登录成绩管理系统提交成绩生成最终考试成绩。

在成绩管理系统中, 为了分析考试成绩数据, 第一项将成绩数据进行预处理, 第二项将连续成绩数据转换为离散的数据, 第三项利用决策树中 ID3 算法进行属性归纳出成绩规则, 第四项教师和学生进行相应的分析评估。

数据预处理过程有对空值的处理、对无用信息进行消减等步骤。系统中采用忽略的方式解决因为缺考、作弊产生的空值。将开课单位、班级、学期、科目类型、代课老师等无关字段删去, 只保留学号、姓名、性别、科目、考核方式、平时成绩、期末成绩和最后的成绩等字段。

数据的离散型处理分两过程组成: 一先将成绩按是否大于等于 60 分来判断成绩是否合格划分, 二将上一步 60 分以上的成绩执行下列步骤: 一是将成绩数据进行排序; 二是去除重复数据; 三是根据相应的比例划分等级; 四是根据确立好的等级标准离散化数据。

比如期末考试成绩 (67, 87, 69, 79, 77, 88, 79, 86, 79, 60, 70, 71, 76, 98, 89, ...) 将分数从高到低排序, 并去除重复的值, (98, 89, 88, 87, 86, 79, 77, 76, 71, 70, 69, 67, 60, ...)。可将成绩的 25%, 50%, 25% 划分 A、B、C 三个等级, 把产生的等级作为成绩表中新的字段值。最后需在原成绩表中的字段 XH (学号)、XM (姓名)、KM (科目)、PC (平时成绩)、QC (期末成绩)、ZC (综合成绩) 中增加字段 PCDJ (平时成绩等级)、QCDJ (期末成绩等级)、YX (优秀否)、JG (及格否)。

利用 ID3 算法构造决策树。该算法是递归算法(下转第 30 页)

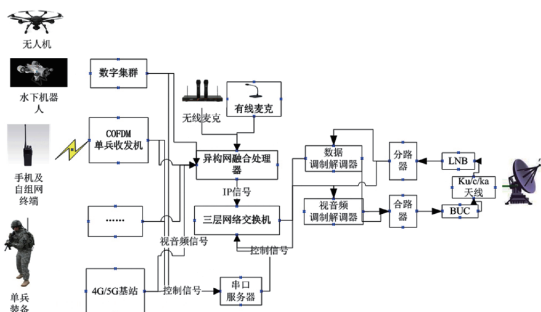


图2 岛礁远程医学地面通信异构网融合硬件方案

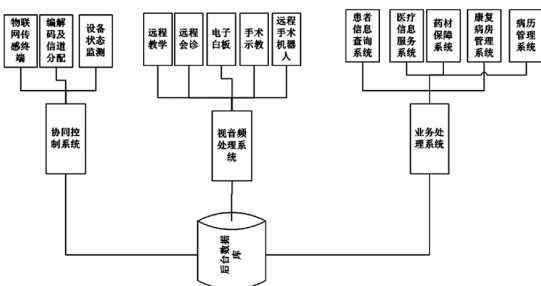


图3 岛礁远程医学地面通信异构网融合软件方案

协同控制系统主要用于获取物联网传感信息、设置信道带宽、调整编解码方式，以及监测设备的工作状态等，确保地面与卫星网络协同工作。

视音频会议处理系统主要用于传输、交互视频、音频信号，集成电子白板，开展远程教学和手术示教，并操控远程手术机器人进行手术等。

业务处理系统主要用于医学业务开展工作的相关信息处理。如患者信息查询、医疗信息服务、药材保障、康复病房和病历管理等。制定医疗保障方案计划、筹措保障物资，提供自动化的手段；实现卫勤信息快速上报和下传双向反馈、预警分析、指挥控制和可视功能，实现疫情、病情、伤情和管理等各种信息快速传

输与分类共享，从而用得上、看得见、判得准。依托岛礁远程医学软件系统，构建一个综合数据库、具备三种能力，实现监控信息实时可知，资源实时可视，处置过程实时可控。

### 3.3 岛礁远程医学地面通信异构网融合接口方案

在接口处理上，需要考虑硬件接口方案和软件接口方案。如图4所示。

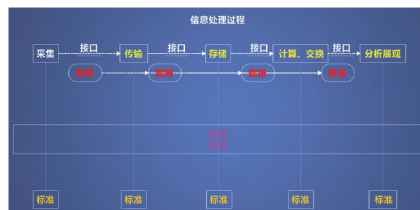


图4 接口方案

针对硬件接口，采用统一标准的传输协议、编解码格式和调制解调方式，可以有效保证通信的一致性。

在软件接口上，需要考虑公共的数据标准、公共编码体系和对应的数据库接口标准，同时确保不同系统之间按照信息资源规划的方式，实现信息交互和信息传输。

特别强调的是，岛礁软硬件接口方案需采用共同的标准实现，否则容易导致系统之间互不相通，信息无法交互。

## 4 结束语

通过岛礁远程医学地面通信异构网融合方案设计，并按照对应方案进行试点试验，可以有效指导岛礁远程医学站点建设，确保系统发挥实际功效，为岛礁人员服务。

### 参考文献

- [1] 吴豪，刘运成，郑重，薛艳明，等. 军民融合远程医疗物联网的构建[J]. 中国数字医学，2017，38-40.
- [2] 刘运成，吴豪，郑重，等. 基于宽带无线网络的远程医疗异构网融合技术架构研究[J]. 中国数字医学，2017，12(2): 19-21.

(上接第32页)的一种,其算法思想:(1)首先生成一个结点。(2)根据判断是否属于同一类的结果来确定是否是叶结点,如果是,该结点是叶结点。(3)如果否,依据信息增益的基于熵的度量启发信息,重新选择属性,将数据样本进行分类,成为该结点测试和判定属性。(4)再根据测试属性的每一个值,创建一个分支,以此划分样本数据,所有的属性都进行离散化转化。该算法符合以下三种情况递归结束:一是结点的数据属于同一类;二是数据全部属性都测试完,没有可利用的属性来划分数据;三是分支没有数据样本了,遇到此情况下创建一个叶结点。

算法步骤:

输入:由离散值属性表示的样本值 samp, 候选属性集合 att\_list

- (1)初始化一个结点 M。
- (2)如果结点中的所有样本都属于同一类别 B, 该结点属于叶结点。
- (3)返回决策树 tree, B 为结点的类标记。
- (4)判断 att\_list 是否为空。
- (5)如果空, 该决策树为单结点树, B 记作类别个数最多的类别。
- (6)如果非空, 在 att\_list 里选择一个最大的信息增益的属性特征 m\_att。
- (7)m\_att 的信息增益和阈值相比, 如果 m\_att 大, 则决策树为单结点树, B 作为实例数最大的类的类标记。
- (8)如果 m\_att 小, 对 m\_att 的取值 mm, 根据取值划分样

本值子集 sampi, 将子集中实例数最大的类作为标记, 创建子结点, 由结点和其子树构成决策树, 返回。

(9)对下个结点, 以 sampi 为新的样本值数据集, 以 att\_list-{m\_att} 为候选属性递归重复以上步骤。

创建好决策树后, 使用后剪枝方法中代价复杂性剪枝对树进行修剪。此算法需要计算树中每个非叶结点删去该结点后子树所产生的期望错误率, 低的期望错误率则修剪该子树, 否则保留该子树。

决策树剪枝完成后, 将决策树分析出的信息进行提取, 生产成绩系统的评估模型, 也就是生成分类规则, 产生学生的考核等级, 可预测学生的成绩是否及格或优秀, 分析规则, 分析影响学生成绩的原因。

## 4 结束语

目前, 数据挖掘技术处于不断的探索和研究过程中, 在完善数字化校园系统的过程中利用数据挖掘技术让数据变得更加有用, 使数字化校园系统功能更加完善。

### 参考文献

- [1] 李爱凤, 刘葵, 唐连章, 葛志游. 数据挖掘技术在数字化校园共享数据中心的应用[J]. 实验室研究与探索, 2013, 32(11): 232-236.
- [2] 黎未然. 数据挖掘技术在数字化校园建设中的应用[J]. 软件, 2012, 33(10): 61-63.
- [3] 邹晶晶. 数据挖掘技术在数字化校园中的应用[D]. 中南大学, 2009.