

后,把在副应用端运行的应用,返回给主应用端。

5.2 手动恢复系统

手动完成将主应用端的数据切换到远程的副应用端的操作。

5.3 数据备份系统

在这种系统中,系统将主应用端的数据实时地备份到远地的存储器中。此方法与上述两种相比,恢复系统正常应用所时间最长。

参考文献:

- [1] 吴勇, 杨琛. 浅论医院数据中心 [J]. 医疗设备, 1002 - 2376 (2008) 04 - 0016 - 03:16-18.
- [2] 王志奇,徐兴良,张红. 浅析医院中心机房安全稳定运行的必备要素[J]. 医疗设备信息, 1007 - 7510 (2007) 01 - 0033 - 02:33-34
- [3] 王经雨,甘哲. 数据中心机房的设计与建设[J]. 首都医科大学学报社会科学版增刊(2006 年),06-150:118-121
- [4] 徐鹏, 薛建锋. 数据中心容灾系统研究 [J]. 计算机工程与设计,

1000-7024 (2007) 22-5556-03:5556-5558

- [5] 日立数据存储系统有限公司上海 [J]. 中国数字医学, 3 (10):69-70.
- [6] 楚宏硕, 张楠. 医院计算机机房建设 [J]. 医疗设备信息, 1007-7510(2007)12- 0054- 03:54-56
- [7] 程钦安,杨保卫. 医院数据中心建设探讨[J]. 中国数字医学, 2(3):45-47.
- [8] 罗维,罗自力,李明,马睿. 医院网络中心机房建设探讨[J]. 西南军医,2007,19(6):140-141.
- [9] 李 婕. 医院信息化促进数据存储中心的建立[J]. 医学信息,2006 (9):1520-1522.
- [10] 田玉兔, 骆晔. 医院信息化建设中的网络机房建设[J]. 医院数字化,1003- 8868(2008) 04- 0049- 03:49-51.
- [11] 郭燕南. 医院信息中心机房建设工程架构探讨 [J]. 医疗卫生装备,1003- 8868(2006)09- 0066- 02:66-67.
- [12] 张黎玉, 黄□. 医院中心机房的规划设想 [J]. 井冈山医学学报, 1008- 2735 (2008) 01- 0055- 01:55/84.

编辑/任鸿兰

试题分析中智能疑难知识点诊断的设计和应用

周 珂

(广东医学院计算机科学教研室, 广东 湛江 524023)

摘要:目的 目前我们已经通过无纸化方式获取了某课程实验考核的所有数据,期望对现有的数据进行分析统计,及时发现实验中学生存在的疑难点并解决。**方法** 常规的分析方法还缺少了对学生成绩变化趋势和能力等学生信息的进一步分析。本文把数据挖掘技术用于某课程的试题分析,创新地提出了基于知识点的试卷分析方法,并给出疑难知识点的诊断方法。**结果** 用该方法对某年级学生实验数据进行分析并提取疑难知识点。**结论** 通过实验数据与传统的方法比较,证明该方法是有效可行的。

关键词: 数据挖掘 试题分析 疑难知识点诊断

中图分类号: TP398.1 **文献标识码:** B

Design and Application of Diagnosed in Difficult Knowledge to Analysis Test Paper

ZHOU Ke

(Dean of computer science, Guangdong Medical College,Zhanjiang 524023,China)

Abstract: Objective We have already obtained all data that examine of some course experiment through the paperless way at present, expect to analyze and count the existing data, find the difficulty that the student of the experiment exists is clicked and solved in time. **Methods** The regular analytical method has also lacked and trended towards the further analysis in student's information such as knowing and ability to student's achievement change. This article excavates technology of the data mining for Analysis on test paper of some courses, have proposed the paper analytical method based on knowledge, provide the diagnosis method of difficult knowledge. **Results** Analyze to student's experimental data of one grade and the difficult knowledge ahead of time is clicked by this method. **Conclusion** Compared with traditional method through the experimental data, prove this method is effective and feasible.

Key words: Data mining ;Analysis of test paper ;Diagnosis of Difficult knowledge

收稿日期:2009-06-02

0 引言

目前,我教研室的《VFP 程序设计》已经实现基于局域网的无纸化的实验考核,也就是说,通过网络考试,我们可以在计算机上获得学生实验考核的所有数据。有了这些原始数据以后,我们可以利用数据挖掘的技术,对每个学生的答题情况进行试题分析,对考试、学生掌握情况、教师教学情况进行全面的了解,从而达到促进教学的作用。

数据挖掘,又称数据库中的知识发现,近几年来已被数据库界所广泛研究。它是在数据仓库或大型数据库的基础上,从大量的、模糊的、随机的数据中提取出数据间重要的但容易被人工分析忽略的知识和信息,自动地发现隐藏在数据间的模式,作出预测性分析。数据挖掘技术涉及数据库、人工智能、神经网络、预测理论、机器学习和统计学等多种相关技术。数据库中的知识发现(KDD)是从大量数据中提取出可信的、新颖的、有效的并能被人们理解的模式的高级处理过程。模式可以看作是我们所说的知识,它给出了数据的特性或数据之间的关系,是对数据包含的信息更抽象的描述^[1,2]。

本文把数据挖掘技术和试题分析有效结合在一起,使考试分数能真实、客观地反映所要测量的属性状态、水平,要对试题进行筛选,对试卷进行分析、鉴定。试卷分析是智能化题库建设中的一重要工作,试卷的质量分析对于提高命题水平、组卷质量和考试质量,完善题库建设具有十分重要的意义。

1 常用的试题分析方法

一般试卷分析是利用试卷的应试结果从难度、区分度、信度和效度等方面进行分析,以便考察试卷结构是否合理。通常我们筛选的题目指标是难度和区分度;鉴定试卷的指标是信度和效度,只有效度合格,才能保证考试在较好程度上测量到所要测量的属性;只有信度合格,才能保证考试有较好的稳定性。

下面给出上述几个指标的计算方法。

1.1 难度

主要是对试卷中试题的难易程度进行分析,这关系到学生得分是否正态分布。一般用难度系数 P 表示,用公式 $P=X/W$ 来计算。其中 X 是全体考生某试题得分平均值, W 试题的满分值。从公式中不难看出 P 值越大, 试题越简单; P 值越小, 试题越难。一般认为 P 值在 0.3-0.9 之间较为适中。

1.2 区分度

区分度是用于判定试题能否把学习上优秀与较差的学生区别开来的指标。计算的方法是先将学生按总成绩排序, 前 27% 的学生为高分组, 后 27% 的学生为低分组, 按下列公式求出区别指数:

$$r=(H_j-L_j)/Q_j$$

其中 H_j 是高分组第 j 题平均分; L_j 是低分组第 j 题平均分; Q_j 是 j 题的满分值。 r 值越高, 表示试题的区分度越好, r 越接近 0, 表示试题的区分度越差。一般认为 $r < 0.15$ 表示区分度较差, 表示区分度较好。

1.3 信度

信度是指考试结果的可信程度, 为考试的实测值与被测真值相差的程度, 也可以理解为同一考生进行 2 次等价的考试, 比较 2 次考试分数所得到的一致性程度。信度是反映考试结果稳定性、可靠性的指标。信度的高低主要反映考试过程中存在的随机误差的大小, 信度高的试卷受随机因素影响较小, 对任何学生多次测定都会产生稳定的结果。以往我们是利用克伦巴赫系数- α 系数法求得可靠性系数:

$$r_{xy} = \frac{n \sum xy \cdot \sum x \cdot \sum y}{\sqrt{n \sum x^2 \cdot (\sum x)^2} \cdot \sqrt{n \sum y^2 - (\sum y)^2}}$$

式中 n 为样本数, x 和 y 分别为两个分布数列, r_{xy} 为分半信度。由于折半信度法是对同一试卷的两部分进行比较, 一方面节省了时间和精力, 简化试卷信度的分析工作, 提高了组卷效率; 另一方面是对完全等价的同一试卷相关因数进行计算, 所以对组卷质量分析的结果更真实可靠。

智能组卷的试题库系统采用折半信度法, 即将同一试卷试题按难度分为两部分, 然后计算这两部分试题得分的相关因数, 即试卷信度 r 计算公式为:

$$r = \frac{2r_{xy}}{1+r_{xy}}$$

按规定标准信度值以 为好。

1.4 试卷的效度

效度是指一次考试能测量到的知识能力的程度。它包括内容效度、结构效度和准则效度。内容效度主要是用来比较试卷考核的内容与教学大纲的一致性。结构效度主要是用来反映试题是否考虑到教学目标、教育层次等内容。

2 基于疑难知识点的试卷分析

以上介绍的四个指标确保了试卷分析中的试卷信息分析及指标反馈这两部分工作的完成。但还缺少了对学生成绩变化趋势和能力等学生信息的进一步分析。因此一个实用性更强, 智高的试题库组卷系统需要有更完善的分析方法。

在这里, 智能组卷的试题库系统可采取基于知识点的试卷即通过对学生试卷的测试结果进行分析, 可以获得对学生知识价, 即对学生的测试结果进行评价, 并诊断出学生在学习过程的疑难知识点, 从而进行相应的教材反馈。通过这种试卷分析, 可以针对学生学习过程中遇到的具体情况, 开展

因材施教的教学。

通过对学生答案进行批改,可以获得描述学生对相应知识点缺乏理解程度的学生知识点缺乏度矩阵 QFM,其定义如下^[34]:

$QFM=[QF(Ki)]$

其中 QF(Ki)表示学生对知识点 Ki 理解的缺乏度, $QF(Ki) \in [0,1]$ 且 $1 \leq i \leq m$ 生成学生知识点缺乏度矩阵学生知识点缺乏度矩阵的算法(QFM_Algorithm)如下:

输入:XSM 学生答题状况矩阵

SJM 试卷知识点矩阵;

输出:QFM 学生知识点缺乏度矩阵

** 初始化 QFM 使其为 0

Initial(QFM)

** 遍历 XSM 和 SJM

For i=1 TO m &&m 为试卷中的知识点数

sum=0

For J=1 TO n &&n 为试卷中的试题数

**SJ(ki,tj)来自于 SJM,XS (tj)来自于 XSM

sum=sum+ SJ(ki,tj)

QF(ki)=QF(ki)+ SJ(ki,tj)*XS(tj)

ENDFOR

QF(ki)=QF(ki)/sum

ENDFOR

如果 $QF(ki) > K_LIMIT(ki)$ (知识点 ki 的缺乏度闭值),即表示学生对知识点 ki 的学习效果不好,则将知识点 ki 视为疑难知识点要加强学习。 $K_LIMIT(ki)$ 可以通过计算试卷得分排名后 50%的学生对知识点 ki 的平均错误率而确定。

综上所述,疑难知识点诊断的算法(KDD_Algorithm)如下:

Step 1:生成试卷的试卷知识点矩阵 SJM;

Step 2:根据学生答题情况生成学生答题状况矩阵 XSM;

Step 3: 根据 SJM 和 XSM 生成学生知识点缺乏度矩阵 QFM;

Step 4:判断是否遍历完 QFM,若遍历完则转 Step 6;

Step 5:如果 $QF(ki) > K_LIMIT(ki)$,则将该知识点加入到疑难知识点集合 DM 中,并转 Step 4;

Step6:输出疑难知识点集合 DM。

3 应用实例

利用我们自行设计的《VFP 程序设计》无纸化考试系统对学校 2005 级全体学生(483 人)进行测试,根据上面的公式,得到对应的难度、区分度、信度如表 1 所示。并以班为单位对考试结果进行了试卷分析,其中 1 班的试卷分析见表 2,所有班级的统计见表 3:

从试题分析可以看出,题目 06028 难度较高,S3048 难度较低,因此两题的区分度也较底,所有题目的信度均大于 0.5,符合标准要求。

表 1 试题分析统计

题号	难度	区分度	信度
06028	0.4	0.13	0.65
Q3034	0.6	0.16	0.64
R3012	0.6	0.17	0.67
S3048	0.8	0.14	0.8

从表 1 可以看出,1 班有 24%的学生对题号为 06028 的知识点不理解,12%的学生对题号为 Q3034 和 R3012 的知识点不理解,9%的学生对题号为 S3048 的知识点不理解,从表 2 可以看出每个班的疑难知识点理解程度,并经过有经验的教师对学生答题情况的分析以及学生的反馈意见,说明利用疑难知识点诊断的算法(KDD_Algorithm)是有效的。同时,根据此数据,还可以对不同班级进行针对性讲解。

表 2 1 班试卷分析情况

学号	知识点 1	知识点 2	知识点 3	知识点 4
05101010001		Q3034	R3021	
05101010002	06028			S3048
05101010003	06028			
05101010004				
05101010005				
05101010006				
05101010007				
05101010008				
05101010009				
05101010010	06028	Q3034		
05101010011				
05101010012				
05101010013				
05101010014	06028			
05101010015				
05101010016				
05101010017	06028		R3021	S3048
05101010018				
05101010019				
05101010020				
05101010021	06028	Q3034	R3021	
05101010022	06028			
05101010023				
05101010024				
05101010025				
05101010026				
05101010027				
05101010028				
05101010029	06028		R3021	S3048
05101010030				

表 3 所有班级统计

班级	06028	Q3034	R3012	S3048
1 班	24%	12%	12%	9%
2 班	23%	10%	12%	10%
3 班	15%	8%	11%	5%
4 班	17%	15%	12%	6%
5 班	27%	13%	15%	8%
6 班	13%	16%	13%	11%
7 班	20%	11%	9%	8%
8 班	24%	11%	9%	14%
9 班	26%	10%	8%	6%
10 班	25%	11%	11%	9%
11 班	29%	12%	11%	10%
12 班	27%	15%	13%	8%
13 班	20%	13%	12%	7%
14 班	22%	11%	12%	8%
15 班	25%	10%	11%	9%
16 班	27%	14%	10%	6%
17 班	26%	11%	13%	10%

4 小结

在诊断整个班级共同疑难知识点的基础上,教师可以调整教学方向针对这些疑难知识点重点教学。同时通过对每个

学生疑难知识点的诊断,我们可以利用知识点的支撑关系进行分析,以获得疑难知识点相应的前驱知识点的掌握状况以及学习后继知识点的可行性,从而可以相应的指导学生有针对性的学习。

智能试卷分析是对试题和试卷使用后的数据进行分析来找出其中的不足和缺陷,并以此为依据来实现试题和试卷的完善,从而客观的为老师的教和学生的学提供了评价及参考依据。

参考文献:

- [1]Jiawei Han Micheline Kamber.Data Mining Concepts and Techniques[M],China Machine Press,2001.8.
- [2] 王斌. 浅析数据挖掘的主要方法和研究方向 [J]. 计算机仿真. 2005.10 1-3.
- [3]C.S.Hsu,S.F.TU and G.J.HWang,Aconcept inheritance method for learning diagnosis of a network-based testing and evaluation system[C]. The 7th International Conference on Computer-Assisted Instructions, 1998.602-609.
- [4]Vasandani,V.&Govindaraj,T.Knowledge organization intelligencet tutoring system for diagnostie problem solving in complex dynamic domains [J].IEEE Transaction on Systems,Man and Cybernetics,1995,25 (7):1076-1096.

编辑/樊延南

数据挖掘技术在医院资金分析管理中的应用

熊友生,徐祖铭

(解放军第 455 医院,上海 200052)

摘要:经过多年“医院信息系统”的应用,很多医院都积累了大量的原始资金数据信息,但目前对这些信息的利用,大多仅限于简单的查询和统计,停留在低层次应用上,缺乏对数据的有效集成、自动钻取,无法形成医院管理者用以决策的信息。本文主要研究数据挖掘技术在医院资金分析管理中的应用,对应用系统中采用的数据挖掘策略和方法进行了重点阐述。

关键词:数据挖掘;资金分析管理;决策

Application of Data Mining in Analysis and Management of Hospital-Funds

XIONG You-sheng,XU Zu-ming

(The No. 455 Hospital of PLA, Shanghai 200052,China)

Abstract:After many years of application of the HIS(hospital information system), many hospitals have gathered large amounts of data information of original capital. But recently the use of this information is limited via simply inquiry and statistics nowadays, and the information application remains on the low level, lacking for effective integration and automatic acquisition, therefore these

收稿日期:2009-06-10