

基于粗糙集理论的影响高校学生成绩因素研究

蔡兴雨¹, 徐 怡^{1,2}, 程智炜¹

(1. 安徽大学 计算机科学与技术学院, 安徽 合肥 230601;

2. 安徽大学 计算智能与信号处理教育部重点实验室, 安徽 合肥 230039)

摘 要:成绩是衡量教师教学质量以及学生学习效果的重要指标。由于影响学生学习成绩的因素众多,教师和学生不能清楚地认识影响成绩的关键因素。一方面,教师无法对教学方法做出有针对性的改进,以提高教学质量;另一方面,学生无法对学习方法做出有针对性的改进,以提高学习成绩。为了帮助高校教师及学生准确分析影响学习成绩的关键因素,设计了用于调查影响高校学生学习成绩因素的调查问卷,向该校大一至大四的学生分发调查问卷并收集数据,然后利用粗糙集理论的基于信息熵的启发式属性约简算法,找出影响高校学生学习成绩的关键因素,之后利用基于粗糙集理论的改进的基于分辨矩阵的属性值约简算法,挖掘出影响高校学生学习成绩的关键因素和成绩之间的依赖关系,导出规则集。通过实验验证了该规则集的有效性。研究成果可以帮助高校教师和学生了解影响成绩的关键因素,从而改进教师的教学方法和学生的学习方法。

关键词:高校学生;成绩影响因素;粗糙集;规则提取;属性约简

中图分类号:TP39

文献标识码:A

文章编号:1673-629X(2016)11-0200-05

doi:10.3969/j.issn.1673-629X.2016.11.043

Research on Factors Affecting College Achievement Based on Rough Set

CAI Xing-yu¹, XU Yi^{1,2}, CHENG Zhi-wei¹

(1. Department of Computer Science and Technology, Anhui University, Hefei 230601, China;

2. Key Lab of Intelligent Computing and Signal Processing of Ministry of Education, Anhui University,
Hefei 230039, China)

Abstract: Achievement is an important indicator of teaching quality and student learning. Because of many factors that affect student's achievement, teachers and students cannot clearly recognize the key factors affecting the results. Therefore, on the one hand, teachers cannot make an improvement to teaching methods to improve the quality of teaching. On the other hand, students are unable to make targeted improvements to the learning methods to improve study performance. To help college students and teachers for analysis of key factors influencing academic performance accurately, a questionnaire about factors affecting college student achievement is designed. Those data are collected from the school's freshman to senior, then using heuristic attribute reduction algorithm based on information entropy in rough set theory to identify the key factors affecting the performance of college students, and next applying improved property values reduction algorithm based on resolution matrix in rough set theory to mine key factors affecting student achievement and college students dependencies between the results derived the rule set. Finally through the experiment, the validity of the rule set is verified. Research can help university teachers and students to understand the key factors that affect performance, thereby improving the way teachers teaching and students learning.

Key words: college student; factors affecting achievement; rough set; rule extraction; attribute reduction

收稿日期:2016-01-21

修回日期:2016-04-26

网络出版时间:2016-10-24

基金项目:国家自然科学基金资助项目(61402005);安徽省自然科学基金项目(1308085QF114);安徽省高等学校省级自然科学基金项目(KJ2013A015, KJ2011Z020);安徽大学大学生科研训练计划项目(KYXL2014064);安徽大学计算智能与信号处理教育部重点实验室开放课题项目

作者简介:蔡兴雨(1994-),男,研究方向为机器学习、数据挖掘;徐 怡,副教授,博士,研究方向为智能信息处理和粗糙集理论。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20161024.1114.048.html>

0 引言

学生成绩在衡量高校教师教学质量以及学生学习效果中意义重大,从成绩中可以看出教师是否完成教学目标,学生对知识的了解和掌握程度是否达到较满意的水平。所以无论是高校教师希望改善教学质量还是学生希望进一步掌握知识,提高成绩都是非常重要的,而影响高校学生学习成绩的因素众多,在面对这么多的因素时,学生和老师可能无法找出哪些是关键性的影响因素,进而无法有针对性地做出努力以有效提高学生成绩。目前对影响学生学习成绩关键因素的研究多是通过调研的方法获取数据,然后利用统计的方法进行分析,很少利用数据挖掘的方法对调研数据做深刻的分析,找出隐藏在数据中的规律性知识。因此,文中设计了影响高校学生学习成绩因素的调查问卷,向该校大一至大四的学生分发调查问卷并收集数据,然后利用粗糙集理论的属性约简算法和规则提取算法,挖掘出影响高校学生学习成绩的关键因素以及这些关键因素和成绩之间的依赖关系,导出规则集,并通过实验验证了规则集的有效性。文中的研究成果可以帮助高校教师和学生了解影响成绩的关键因素,从而改进教师的教学方法和学生的学习方法。

文中所采用的粗糙集(Rough Set,也称 Rough 集、粗集)理论^[1-2]是 Pawlak 教授在 1982 年提出的一种能够定量分析处理不精确、不一致、不完整知识与信息的数学工具。该理论最初的原型来源于比较简单的信息模型,它的基本思想是通过关系数据库分类归纳形成概念和规则,通过等价关系的分类以及分类对于目标的近似实现知识发现。粗糙集理论能处理模糊和不确定的知识,在保持分类能力不变的前提下,通过知识约简,导出决策和分类规则。所以文中利用粗糙集对收集的 300 份中有效的 279 份成绩样本先进行属性约简,然后进行规则提取,得出了影响高校学生成绩的关键因素。

1 基本概念

文中用到的相关概念介绍如下^[3]:

定义 1:信息系统(知识表达系统) $S = (U, A, V, f)$ 。其中: U 为对象的非空有限集合,称为论域; A 为属性的非空有限集合; $V = \bigcup_{a \in A} V_a$, V_a 是属性 a 的值域; $f: U \times A \rightarrow V$ 是一个信息函数,它为每个对象的每个属性赋予一个信息值,即 $\forall a \in A, x \in U, f(x, a) \in V_a$ 。

定义 2:信息系统 $S = (U, A, V, f)$, 若 $A = C \cup D, C \cap D = \emptyset$, C 为条件属性集, D 为决策属性集,这样的信息系统称为决策表,决策表是一类特殊而重要的信息系统。万方数据

定义 3: 对于一个决策表, $S = (U, C \cup D, V, f)$, $B \subseteq C$, 基于 B 的不可分辨关系定义为: $\text{ind}(B) = \{(x, y) \in U \times U\}$ 或 $\text{ind}(B) = B \mid \forall b \in B, f(x, b) = f(y, b)\}$ 。

如果 $(x, y) \in \text{ind}(B)$, 则称 x 和 y 是基于 B 不可分辨的。符号 $U/\text{ind}(B)$ 表示不可分辨关系 $\text{ind}(B)$ 在 U 上导出的所有等价类,可简记为 U/B 。通常, $\text{ind}(C)$ 称为条件等价类, $\text{ind}(D)$ 称为决策等价类。 $[X]_B$ 表示基于 B 得到的 x 的等价类。

定义 4: R 为一族等价关系 $P \in R$, 若 $\text{ind}(R) = \text{ind}(R - \{P\})$, 则称 P 为 R 中不必要的, 否则称 P 为 R 中必要的。如果每一个 $P \in R$ 都是必要的, 则称 P 为独立的, 否则称 P 为依赖的。

定义 5: 对于一个决策表 $S = (U, C \cup D, V, f)$, $B \subseteq C$, 若 B 是独立的, 且 $\text{ind}(B) = \text{ind}(C)$, 则称 B 是 C 的一个约简, 记为 $\text{red}(C)$ 。

定义 6: 核属性定义为 $\text{core}(P) = \bigcap \text{red}(P)$, 其中 $\text{red}(P)$ 表示 P 的所有约简。

核的概念的用处: (1) 可以作为所有约简的基础; (2) 在知识约简中是不能消去的知识特征集合。

定义 7: 在信息系统 S 中, 若知识 $P \in A$, P 在 U 上导出的等价划分为: $\text{IND}(P) = \{X_1, X_2, \dots, X_n\}$ 。则 P 的信息熵记作 $H(P)$, 定义为: $H(P) = - \sum_{i=1}^n p(X_i) \log p(X_i)$ 。其中, $p(X_i)$ 表示等价类 X_i 在 U 中的概率, $p(X_i) = \frac{|X_i|}{|U|}, i = 1, 2, \dots, n$ ^[4]。

定义 8: 在信息系统 $S = (U, A, V, f)$ 中, $C \in A$, $\forall a \in A - C$ 关于属性集 C 的重要性定义为:

$$\text{sgf}_c(a) = H(C \cup \{a\}) - H(C)$$

由定义可知, 属性 $a \in A - C$ 关于属性集 C 的重要性由 C 中添加 a 后所引起的信息熵的变化大小决定, 此值越大, a 关于 C 越重要。

定义 9: 令 $S = (U, C \cup D, V, f)$ 是一个决策表, $U = \{x_1, x_2, \dots, x_n\}, c \in C$ 。 $c(x)$ 是对象 x 在属性 C 上的值, $D(x)$ 是对象 x 在决策 D 上的值, 则分辨矩阵记为 $M(S) = [c_{ij}]_{n \times n}$, 其 i 行 j 列处元素为:

$$[c_{ij}]_{n \times n} = \begin{cases} c \in C: c(x_i) \neq c(x_j), D(x_i) \neq D(x_j) \\ 0, D(x_i) = D(x_j) \\ -1, \forall c: c(x_i) = c(x_j), D(x_i) \neq D(x_j) \end{cases}$$

因此, 分辨矩阵中元素 c_{ij} 是能够区别对象 x_i 和 x_j 的所有属性的集合; 但若 x_i 和 x_j 属于同一决策类时, 则分辨矩阵中元素 c_{ij} 的取值为空集 \emptyset 。显然, 分辨矩阵是一个依主对角线对称的 n 阶方阵, 在进行分辨矩阵运算时, 只需考虑其上三角(或下三角)部分^[5]。

定义 10: 对于一个决策表, $S = (U, C \cup D, V, f)$, $x \in U$, $[x]_C$ 表示 x 根据条件属性 C 定义的等价类, $[x]_D$ 表示 x 根据决策属性 D 定义的等价类, 对象 x 对应的决策规则 R_x 定义为: $\text{Des}([x]_C) \rightarrow \text{Des}([x]_D)$ 。其中, $\text{Des}([x]_C)$ 表示基于属性集 C 对 x 的特征描述; $\text{Des}([x]_D)$ 表示基于属性集 D 对 x 的特征描述。

对于规则的质量, 有三个衡量标准, 分别是置信度、覆盖度和支持度^[6]:

定义 11: 规则的置信度。在决策信息系统中, 定义 R_x 的置信度为 $\text{Cer}(R_x) = \frac{[x]_C \cap [x]_D}{[x]_C}$ 。

确定性程度 $\text{Cer}(R_x)$ 反映了决策规则的可信性, 或者设置置信度, 是衡量规则 r 中条件类分配到决策类的精度。

定义 12: 规则的覆盖度。在决策信息系统中, 定义 R_x 的覆盖度为 $\text{Cov}(R_x) = \frac{[x]_C \cap [x]_D}{[x]_D}$ 。

覆盖度 $\text{Cov}(R_x)$ 用来评估决策规则的质量, 反映了决策规则条件类对决策类的覆盖程度。

定义 13: 规则的支持度。在决策信息系统中, 定义 R_x 的支持度为 $\text{Sup}(R_x) = \frac{[x]_C \cap [x]_D}{U}$ 。

支持度 $\text{Sup}(R_x)$ 用来衡量决策规则的强度, 反映了论域中支持此决策规则的对象占全体对象的比例。支持度越大, 说明支持该决策规则的对象越多, 规则的强度就越大, 规则的泛化能力和鲁棒性就越好。

2 属性约简方法

现有的主流属性约简方法有基于正域的属性约简方法、基于属性依赖度的属性约简方法、基于信息熵的属性约简方法和基于差别矩阵的属性约简方法等^[7-10]。基于信息熵的属性约简方法利用信息熵来度量属性重要度, 算法运行效率和准确性都相对较高。因此, 文中采用基于信息熵的属性约简方法, 算法描述如下^[11]:

输入: 信息系统 $S = (U, A, V, f)$;

输出: 信息系统的核与最小约简。

步骤 1: 计算 $H(A)$ 。

步骤 2: 求 $\text{Core}(A)$ 。

(1) $\text{Core}(A) = \emptyset$;

(2) 对每个 $a \in A$, 计算 $\text{sgf}(a)$;

(3) 对每个 $a \in A$, IF $\text{sgf}(a) > 0$ THEN $\text{Core}(A)$:
 $= \text{Core}(A) \cup \{a\}$;

(4) 输出 $\text{Core}(A)$ 。

步骤 3: 求最小约简。

(1) 令 $C = \text{Core}(A)$, 执行以下过程;

(2) 若 $H(C) = H(A)$, 则 C 为 A 的一个约简, 转

(3), 否则转(2);

(3) 对每个属性 $a \in A - C$, 计算属性重要性 $\text{sgf}(a)$, 选择属性 a 使满足: $\text{sgf}_C(a) = \max_{a \in A-C} \text{sgf}_C(a)$, 令 $C := C \cup \{a\}$, 转(1);

(4) 输出约简 C 。

3 规则提取方法

规则提取即为属性值约简, 是在保持决策能力不变的前提下, 在属性约简的基础上, 进一步删除冗余的属性值以提高生成规则的泛化能力。属性值约简有很多种算法, 常见的有一般值约简算法、基于分辨矩阵的值约简算法、基于 LEM2 的规则提取算法等^[12-13]。由于基于分辨矩阵的规则提取方法简单易操作, 但效率较低, 所以为了提高效率, 文中采用一种改进的基于分辨矩阵的属性值约简算法, 描述如下^[14]:

输入: 一个决策表 $S = (U, C \cup D, V, f)$;

输出: 该决策表的一个决策规则集 R 。

步骤 1: 构造 $M(S)$ 。

步骤 2: 计算第 i 行核属性 C_0 。

(1) $C_0 = \emptyset$;

(2) $\text{num}(j)$ 为 $M(S)$ 第 i 行第 j 个元素所含的属性个数, 对 $M(S)$ 每一行 IF $\text{num}(j) := 1$ THEN $C_0 := C_0 \cup c_{ij}$ 。

步骤 3: 修改 $M(S)$ 。

(1) $B \leftarrow C_0$;

(2) IF $C_{ij} \cap B \neq \emptyset$

THEN $M(S) = M(S) - \{C_{ij}\}$

其中 i 为定值, 即将分辨矩阵的第 i 行中所有与 B 相交不空的属性集赋空;

(3) IF $M(S)$ 第 i 行为空, THEN 执行步骤 4, ELSE 转(4);

(4) 计算每个 c 属于 $C - B$ 的重要性, 扩充 B , 计算 $c \in C - B$ 在 $M(S)$ 的第 i 行中出现的次数, 将出现次数最多的属性加入 B 转(2)。

步骤 4: 得出规则。

(1) 对于第 i 行 IF $c \cap B := \emptyset$ THEN $\text{Des}([x]_B) \rightarrow \text{Des}([x]_D)$;

(2) 用上述算法对每一行对象进行处理, 最终得到决策规则集 R 。

4 实验分析

4.1 数据收集

为了得到大量相关数据来研究影响高校学生学习成绩的关键因素, 设计了一份影响高校学生学习成绩

的调查问卷。问卷内容涵盖了学生基本信息、学习态度、学习方法三大方面。其中,基本信息包括学生的个人信息和家庭信息,学习态度包括学习目的与学习动力,学习方法包括学习的时间地点与方式手段。问卷共包括 21 个问项,前 20 个构成条件属性,后一个为决策属性。为了便于数据处理与提取,所有问项皆为单选,每个答案独立不重叠。具体的调查问卷见表 1。

表 1 调查问卷

编号	问题	选项
1	您的性别	A. 男 B. 女
2	您是否有男(女)朋友	A. 有 B. 无
3	您每周上自习的时间	A. 12 小时以上(平均每天 2 小时以上)
		B. 6~12 小时(平均每天 1 小时左右)
		C. 6 小时以下(平均每天不到 1 小时)
		D. 偶尔
4	您每周上网的时间	A. 28 小时以上(平均每天 4 小时以上)
		B. 14~28 小时(平均每天 2~4 小时)
		C. 14 小时以下(每天小于 2 小时)
		D. 偶尔
5	您是否玩游戏	A. 是 B. 否
6	您上课时一般坐在第几排	A. 1~3 B. 4~6 C. 7~8
7	您父亲的职业	A. 其他 B. 企业 C. 公务员或事业单位
8	您母亲的职业	A. 其他 B. 企业 C. 公务员或事业单位
9	您父母的平均文化程度	A. 小学 B. 初中 C. 高中 D. 大学
10	您经常去图书馆借书吗	A. 是 B. 否
11	您对考试的态度	A. 能过就行 B. 很在乎 C. 无所谓
12	您毕业后的打算	A. 就业 B. 考研 C. 出国
13	您的政治面貌	A. 党员 B. 团员 C. 群众
14	您偏向哪些爱好	A. 运动类(如:体育运动、旅游等)
		B. 安静类(如:看电影、听音乐、看书等)
15	您每周锻炼身体的时间	A. 12 小时以上(平均每天 2 小时以上)
		B. 6~12 小时(平均每天 1 小时左右)
		C. 6 小时以下(平均每天不到 1 小时)
		D. 偶尔
16	您会购买学习辅导书吗	A. 会 B. 不会
17	学习上遇到问题您倾向以什么方式解决	A. 询问老师
		B. 独立思考
		C. 查阅书籍材料
		D. 查阅网络资料
18	您的作业是否独立完成	A. 是 B. 否
19	您翘课的频率	A. 总是 B. 经常 C. 偶尔 D. 从不
20	您努力学习的原因	A. 为未来工作打下基础
		B. 为继续深造打下基础
		C. 对专业知识感兴趣
		D. 父母的期望
21	您的平均绩点	A. 4.0~3.1 B. 3.0~2.5C. 2.4~1.6

为了使调研对象的分布较均匀,调研结果更加可靠,调研对象为在校的大一到大四的学生,调研地点有宿舍、图书馆、自习室、考研教室、食堂、校园街道。总共收了 300 份问卷,剔除其中不符合要求以及明显随意填写无参考价值的 21 份问卷,剩余 279 份为真实有用数据。万方数据

4.2 数据处理

使用第 2 节中基于信息熵的属性约简算法对收集的 279 份数据进行属性约简,约简结果为 14 个属性:性别*,父母的平均文化程度*,每周上网的时间,上课时一般坐在第几排,母亲职业,图书馆借书频率,爱好,锻炼身体的时间,购买学习辅导书,学习上遇到问题时的解决方式,作业是否独立完成,翘课的频率,努力学习的原因。其中,打*的 2 个为核属性。

在上述属性约简的基础上利用第 3 节中改进的基于分辨矩阵的属性值约简算法对约简后的决策表进行处理,得出了 33 条有效规则,鉴于篇幅有限,在此就不列出了。其中,覆盖度 0.5 以上,支持度 0.3,置信度 0.9 以上的规则共 16 条,如表 2 所示。

表 2 规则集

序号	规则的条件
1	if $C_3=1$ and $C_1=1$ and $C_8=1$ then $D=1$
2	if $C_1=2$ and $C_4=4$ then $D=1$
3	if $C_8=2$ and $C_1=1$ and $C_3=2$ and $C_9=3$ then $D=1$
4	if $C_4=4$ and $C_6=2$ and $C_3=4$ and $C_8=1$ then $D=3$
5	if $C_{10}=1$ and $C_1=1$ and $C_3=4$ then $D=3$
6	if $C_3=4$ and $C_6=2$ and $C_1=1$ and $C_4=1$ and $C_9=2$ then $D=3$
7	if $C_8=1$ and $C_1=1$ and $C_3=3$ and $C_4=2$ and $C_{16}=2$ then $D=2$
8	if $C_1=2$ and $C_3=1$ and $C_6=1$ then $D=1$
9	if $C_3=1$ and $C_4=2$ and $C_8=1$ then $D=1$
10	if $C_6=1$ and $C_4=2$ and $C_9=2$ then $D=1$
11	if $C_1=1$ and $C_6=3$ and $C_3=4$ and $C_9=4$ then $D=2$
12	if $C_8=1$ and $C_1=1$ and $C_3=3$ and $C_4=2$ and $C_{16}=2$ then $D=2$
13	if $C_1=1$ and $C_3=2$ and $C_6=3$ and $C_{18}=1$ and $C_4=2$ then $D=2$
14	if $C_4=1$ and $C_3=4$ and $C_6=1$ and $C_8=2$ then $D=1$
15	if $C_{15}=2$ and $C_4=1$ and $C_3=4$ then $D=1$
16	if $C_9=4$ and $C_1=1$ and $C_3=2$ and $C_4=3$ and $C_{14}=1$ then $D=2$

其中, $C_i=j$ 中的 C_i 代表表 1 中的第 i 个问题也是第 i 个条件属性, C_i 的值 j 代表表 1 中第 i 个问题的第 j 个选项也是条件属性的值, D 为决策属性, D 的值为表 1 中第二十一问题的选项也是决策属性值。

为了验证得出规则的质量,从 279 份数据中分别随机抽出 30%、60%、90% 的数据作为测试用数据,以验证上述 16 条规则的分类精度,也就是规则的可靠程度。为了保证结果的准确性,对每种比例的测试数据都抽取了 100 次,如 30% 的测试数据,抽取 100 次,计算每次抽取组的分类精度,取 100 次分类精度的平均值作为最终结果。不同比例测试数据的分类精度如表 3 所示。

从表 3 中可以看出,通过运用基于信息熵的属性约简算法以及改进的基于分辨矩阵的属性值约简算法后,得出的 16 条规则的分类精度均在 0.7 以上,由此可见从数据集中得出的 16 条规则是较为可靠的。

表 3 16 条规则的分类精度

组数	测试用数据占 总体数据比例/%	测试次数	分类精度
一	30	100	0.714
二	60	100	0.725
三	90	100	0.731

为了进一步验证文中方法的正确性,从 279 份数据里随机抽取 50%、60%、70%、80%、90% 的数据作为训练数据,剩余数据作为测试数据,进行交叉测试,每组交叉测试进行 100 次。计算结果如表 4 所示。

表 4 交叉测试结果

组数	训练集 比例/%	训练集	测试集	测试次数	分类精度
一	50	140	139	100	0.632
二	60	167	112	100	0.651
三	70	195	84	100	0.674
四	80	223	56	100	0.718
五	90	251	28	100	0.696

从表 4 的测试结果中不难得出,当训练数据量和测试数据量各占总数据量的 50% 时,分类精度依然可以达到 0.6 以上,说明文中采用的算法较为恰当,得到的规则的准确度和泛化能力也比较好,从而验证了文中方法的有效性,证明了文中结论的精确性。而且当训练数据增多时,分类精度也逐渐提高。这是因为训练数据越多,所提取规则的质量也越好,这个结果也符合人的直观认识。

从上述实验分析结果可以看出,影响高校学生成绩的因素可以分为两个方面:一是主观态度,包括:上自习时间、上网时间、锻炼身体的时间、上课时坐在第几排、图书馆借书频率、作业是否独立完成以及翘课频率等;二是客观环境,包括:性别、父母文化程度、母亲职业等。再由得出的规则综合来看,一是态度积极向上的学生成绩多优于态度一般的学生,像自习时间、图书馆借书频率、作业独立完成情况、锻炼身体的时间等与学校成绩呈正相关,而上网时间、翘课频率与学生成绩呈负相关;二是女生成绩普遍好于男生,因为女生比较细心,在学习考试当中更容易取得优良的成绩;三是父母文化程度越高孩子成绩越好,因为文化程度高的父母可以为子女提供更好的学习指引,更擅长培养子女良好的学习习惯。从提取的规则中发现有趣的是母亲的职业对孩子成绩的影响要大于父亲职业的影响,也许是受中国相夫教子的传统影响,母亲在子女成长的过程中陪伴的更多,母亲的很多行为习惯比父亲对子女造成的影响更大,所以母亲在孩子的学习生活中扮演着非常重要的角色。

以上研究成果可以帮助高校教师和学生了解在众多的影响因素中哪些是影响学生学习成绩的最关键

因素,以及这些关键因素和成绩之间的依赖关系,从而帮助教师改进教学方法,并帮助学生改进学习方法,以更好地提高成绩。

5 结束语

为了帮助高校教师及学生准确分析影响学习成绩的关键因素,以有效提高学生成绩,文中设计了影响高校学生学习成绩因素的调查问卷,构成决策表,然后利用粗糙集理论的属性约简和规则提取算法,从决策表中提取影响高校学生学习成绩的关键因素以及这些关键因素和成绩之间的依赖关系,导出规则集,通过实验验证了规则集的有效性。研究成果可以帮助高校教师改进教学方法,帮助高校学生改进学习方法。

参考文献:

[1] Pawlak Z, Grzymala-Busse J W, Slowinski R, et al. Rough sets [J]. Communications of the ACM, 1995, 38(11): 88-95.

[2] Pawlak Z, Skowron A. Rough sets: some extensions [J]. Information Sciences, 2007, 177(1): 28-40.

[3] 张文修, 吴伟志, 梁吉业, 等. 粗糙集理论与方法 [M]. 北京: 科学出版社, 2001.

[4] 吕林霞, 赵锡英, 唐占红. 一种基于信息熵的信息系统属性约简算法 [J]. 自动化与仪器仪表, 2013(5): 197-199.

[5] Pawlak Z. Rough set approach to knowledge-based decision support [J]. European Journal of Operational Research, 1997, 99(1): 48-57.

[6] Pawlak Z. Rough sets, decision algorithms and Bayes' theorem [J]. European Journal of Operational Research, 2002, 136(1): 181-189.

[7] 陈娟, 王国胤, 胡军. 优势关系下不协调信息系统的正域约简 [J]. 计算机科学, 2008, 35(3): 216-218.

[8] 路松峰, 刘芳, 胡波. 一种基于属性依赖的属性约简算法 [J]. 华中科技大学学报: 自然科学版, 2008, 36(2): 39-41.

[9] Faustino Agreira C I, Machado Ferreira C M, Maciel Barbosa F P. Roughset theory: data mining technique applied to the electrical power system [M]. Netherlands: Springer, 2013.

[10] 杨萍, 李济生, 黄永宣. 一种基于二进制区分矩阵的属性约简算法 [J]. 信息与控制, 2009, 38(1): 70-74.

[11] 吴尚智, 苟平章. 粗糙集和信息熵的属性约简算法及其应用 [J]. 计算机工程, 2011, 37(7): 56-58.

[12] 徐怡, 李龙澍, 李学俊. 改进的 LEM2 规则提取算法 [J]. 系统工程理论与实践, 2010, 30(10): 1841-1849.

[13] Grzymala-Busse J W. A new version of the rule induction system LERS [J]. Fundamental Information, 1997, 31(1): 27-39.

[14] 饶泓, 夏叶娟, 李姆竹. 基于分辨矩阵和属性重要度的规则提取算法 [J]. 计算机工程与应用, 2009, 44(23): 163-165.