

基于DM的网络教学资源的反馈跟踪系统^①

于宝琴¹ 张悦¹ 王欢欢¹ 吴金晶¹ 叶娟娟²

(1. 天津财经大学商学院 天津 300222; 2. 河北工业大学 管理学院 天津 300401)

摘要: 网络教学资源的反馈跟踪系统是促进教学质量不断提高的重要手段,从以学生为主体的理念出发,研究设计了网络教学资源的反馈跟踪系统。该系统采用DM(Data Mining, 数据挖掘)技术中的粗糙集和关联规则对学生在线学习等行为特征数据进行数据收集、数据预处理、构造决策表、基于粗糙集的关联规则的提取,最终得出跟踪反馈的结果,对于推进教学现代化,提高教材质量有重要的意义。

关键词: 网络教学资源; 数据挖掘; 反馈跟踪; 粗糙集; 关联规则

Feedback Tracking System of Network-Aided Teaching Resources Based on DM

YU Bao-Qin¹, ZHANG Yue¹, WANG Huan-Huan¹, WU Jin-Jing¹, YE Juan-Juan² (1. Tianjin University of Finance and Economics, Tianjin 300222, China; 2. Hebei University of Technology, Tianjin 300401, China)

Abstract: The feedback tracking system of network-aided teaching resources is an important method to promote the quality of teaching, this article design the system from the student perspective. In this paper, we use the rough sets and association rules in the DM technique to data collection about students' online learning behavior, data preprocessing, construction decision-making table, association rules extraction based on rough set, finally, we get the results. The system has an important significance to promote the modernization of teaching and teaching quality.

Keywords: network-aided teaching resources; data mining; feedback tracking; rough sets; association rules

1 引言

随着 Internet 和 Web 技术的发展,网络教学资源已经成为网络研究和应用的热点之一。网络教学资源是通过计算机网络获得并能用于教学中的各种信息资源的总和。网络教学资源对于汇总大量平面教材,统一管理各种多媒体课件等电子辅助资料发挥了重要作用。立体化教材是以现代化信息技术为平台,以适应远程传输的数字化教育教学文档、软件为教材,以网络为学习和管理环境,适用于多层次教学对象,覆盖教学活动的各个步骤的教学资源体系。

但是现有的教学资源大多处于资源共享的状态,基本上都是静态的,即教学资源一放上去以后,就很少变更。设计者不知道资源的设计是否合理,是否符合

教学规律,页面之间的超文本链接是否符合用户的访问习惯,学生只能被动的接受完全相同的学习内容,从而没有真正体现出个性化教育的优势。因此,需要一种新的技术来充分利用这些有用信息,建立一个智能化、个性化的网络教学资源环境。数据挖掘技术从 Web 文档和访问数据中发现和抽取知识,将基于粗糙集的关联规则应用于网络教学资源,可以在海量的访问数据中发现学生的访问习惯、学习兴趣、学习倾向等,能够根据学生的访问规律调整网站结构,动态地为学生提供个性化的服务,能够发现学生的学习兴趣,并据此给学生推荐课程和学习资料。基于粗糙集的关联规则的采用为网络教学资源的智能化、个性化提供了重要的技术手段,对于推进教学现代化,提高教材质

^① 基金项目:天津市高等学校本科教学改革与质量建设研究计划(A06)

收稿时间:2010-04-18;收到修改稿时间:2010-06-21

量有重要的意义。

2 粗糙集和关联规则

2.1 粗糙集理论^[1,2]

粗糙集理论(Rough Set)是由波兰科学家 Z.Pawlak 教授提出来的对不完整数据进行分析、推理、学习、发现的新方法。目前已成为本领域中一个较新的学术热点,引起了越来越多的科研人员的关注。

一个知识表示系统 S 是一个四元组 $S=(U,A,V,f)$, 其中 U 为对象的非空有限集,称为论域; $A=CUD$ 是有限个属性的非空集合,子集 C 和 D 分别为条件属性和决策属性; V 是 C 和 D 的值域; $f:U \times A \rightarrow V$ 是一个信息函数。

知识表达系统 $S=(U,A,V,f)$ 对应一张二维表,表中每一列描述对象的一种属性 $A=CUD, C \cap D=?$, A 为非空有限集称为属性集合, C 称为条件属性集, D 称为决策属性集。每一行表示一个对象的一条由观察或测量得到的信息即该对象的各个属性值,这样得到的表称为决策表,决策表对应一个信息库。

给定一个论域 U 和 U 上的一簇等价关系 S , 若 $P \subseteq S$, 且 $P \neq ?$, 则 $\cap P$ (P 中所有等价关系的交集)仍然是论域 U 上的一个等价关系,称为 P 上的不可分辨关系,记为 $IND(P)$, 也常简记为 P 。而且,

$$\forall x \in U, [x]_{IND(P)} = [x]_P = \cap [x]_R$$

这样, $U/IND(P) = \{[x]_{IND(P)} | x \in U\}$ 表示与等价关系 $IND(P)$ 相关的知识,称为知识库 $K=(U,S)$ 中关于论域 U 的 P -基本知识(P -基本集)。

2.2 关联规则^[3,4]

设 $I=\{i_1, i_2, \dots, i_n\}$ 是项的集合。包含 K 个项的项集称作 K -项集。设 D 是数据库记录的集合,其中每个事务 T 是项的集合,且 $T \subseteq I$ 。设 X 是一个项集,事务 T 包含 X 当且仅当 $X \subseteq T$ 。

关联规则是形如 $X \Rightarrow Y$ 的蕴涵式,这里 $X \subseteq I, Y \subseteq I$, 且 $X \cap Y = ?$ 。 X 称为规则的左部或规则的前提(简记 LHS), Y 称为规则的右部或结论(简记 RHS)。

度量规则的参数是支持度(Support)与置信度(Confidence):

$$Support(X \Rightarrow Y) = P(X \cup Y)$$

$$Confidence(X \Rightarrow Y) = P(Y | X)$$

支持度揭示了规则的重要性,置信度揭示了规则

的可信度。

经典的粗糙集理论中的规则是一个特例,即 $support=0$ 且 $confidence=100\%$ 。

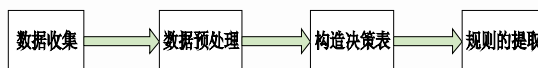
3 基于粗糙集的关联规则的跟踪系统设计

3.1 系统设计原理

由于网络教学资源平台的跟踪反馈系统具有学员在线学习行为的复杂性、动态性、多因素性和模糊性等特点,这种带有大量不确定因素的半结构化问题或非结构化问题难以采用传统的统计学评价模型进行科学地计算处理。而从信息论的角度来看,网络教学资源系统的所有状态特征是通过相关数据表现出来的;从控制论角度来看,对教学资源进行跟踪,实际上是依据系统所挖掘出来的数据进行处理并做出决策。因此,我们可以利用基于粗糙集的关联规则技术来研究这个问题。

3.2 系统设计

表 1 基于粗糙集的关联规则的跟踪体系设计框架



3.2.1 数据收集

网络教学资源平台的跟踪反馈系统主要是对学生的学习行为进行实时跟踪,需要收集的学习行为特征数据主要来源于 Web 日志。

3.2.2 数据预处理

通过对 Web 日志中的初始数据进行数据清理,滤掉多余记录,合并相关数据并将不恰当的或冗余的数据项从数据集中清除,识别会话,获得初始 Web 访问序列表,如下图所示:

表 2 Web 访问序列表

TID	Web访问序列
1	A,B,C,D
2	B,C,E
3	A,B,C,E
4	A,B,C,D,E

3.2.3 构造决策表^[6]

决策表是粗糙集理论中一类特殊而重要的知识表达系统,多数决策问题都可以用决策表形式来表达,要想把粗糙集理论应用于网络教学资源、立体化教材数据资源的反馈跟踪系统中,就要把 Web 日志中的访

问序列用决策表的形式表示出来。由于用户访问资源的序列不等长，因此就要构造维数不同的决策表，然后运用粗糙集理论提取规则。

设用户的访问序列 $T=\{t_1, t_2, t_3, \dots, t_n\}$ ，其中第 1 至 t_{n-1} 个属性为条件属性，第 t_n 属性为决策属性。

决策表 $DT=?DT_i$ ，其中 i 为用户访问序列的维数， $1 \leq i \leq n$ ， DT_i 表示对应维数为 i 的决策表。

3.2.4 基于粗糙集的关联规则的提取^[7]

针对上述方法进行处理后得到的决策表，提取网络教学资源的推荐规则，应用基于粗糙集的关联规则方法，具体算法的伪代码如下所示：

输入 Database: 事务数据库

min_sup:最小支持度计数阈值

min_conf:最小置信度计数阈值

令 C 为条件属性集，D 为决策属性集

(1) $k=1$

$LC1=\{i \mid i \in C \wedge \sigma(\{i\}) \geq N \times \min_sup\}$

{发现所有条件属性的频繁 1-项集}

(2) $k=1$

$LD1=\{i \mid i \in D \wedge \sigma(\{i\}) \geq N \times \min_sup\}$

{发现所有决策属性的频繁 1-项集}

(3) $k=2$

$L2=\{i \mid i \in LC1 \cup LD1 \wedge \sigma(\{i\}) \geq N \times \min_sup\}$

{连接 LC1 与 LD1，并生成频繁 2-项集}

for L2 中的每条记录 $W2, W2=\{W21, W22, \dots,$

$W2n\}$

if $W2$ 满足决策属性 D 相同

then $L2=L2 \cup W2$ {将 L2 分成具有共

同决策属性值的若干簇}

end if

end for

(4) for ($k=3; Lk-1 \neq \emptyset; k++$) {

$Ck=apriori_gen(Lk-1)$ {应用

apriori 算法产生候选项集}

for each 事务 $t \in T$ do

$Ct=subset(Ck, t)$ {识别属于 t 的所有候选}

for each 候选 $c \in Ct$ do

$\sigma(c)=\sigma(c)+1$ {支持度计数增值}

end for

end for

$Fk=\{c \mid c \in Ck \wedge \sigma(\{c\}) \geq N \times \min_sup\}$

{提取频繁 k-项集}

(5) for each Lwi do

$conf=\sigma(i)/\sigma(i-l)$

if $conf \geq \min_conf$

then output: rule $(i-l) \rightarrow (l)$ {输出规则}

end if

end for

4 网络教学资源的个性化推荐

网络教学资源、立体化教材数据资源站点上积累了大量有用的信息，这些信息对于学生有一定的指导作用，可以发现学生的学习兴趣，并且据此推荐合适的网络教学资源。

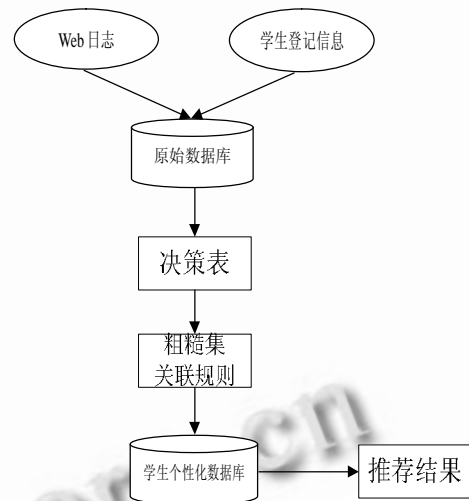


图 1 网络教学资源的个性化推荐系统流程图

5 实验结果及分析

实验数据采用立体化教材实例网站的学生学习行为日志数据，该网站是由网络教学资源共享、立体化教材课题组开发完成的，学生通过该网站访问网络课程，浏览教学内容、检索相关知识、进行模拟仿真实验和自我测试，可以在学习本门课程的同时有效地培养在信息化社会中学习能力与素质。

为了测得本文中所提方法的有效性，首先对学生行为日志进行预处理，然后从中选择 674 条访问数据作为试验数据集，该数据集中包括 674 个用户和 2839 个项目。本文方法采用 MATLAB 进行处理和实现。MATLAB 实现关联规则的界面和结果分别如图 2 和图 3 所示：

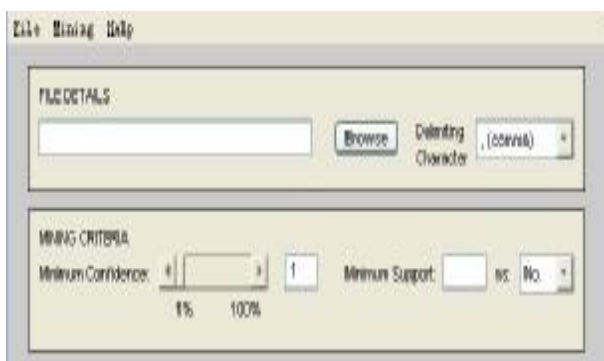


图 2 matlab 实现关联规则界面

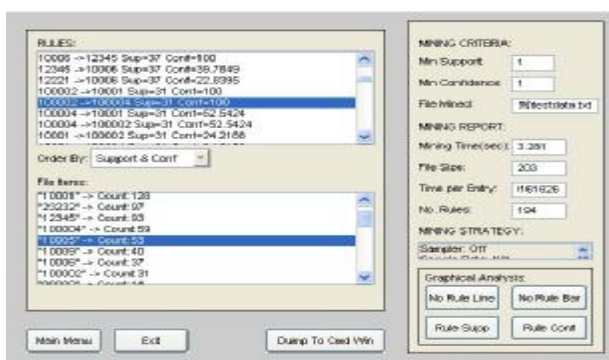


图 3 用 matlab 实现关联规则的结果

为了说明本方法的有效性，本文还对基于粗糙集的关联规则与没有应用粗糙集的关联规则的推荐准确率进行了数据分析，并做了对比。如图 4 所示：

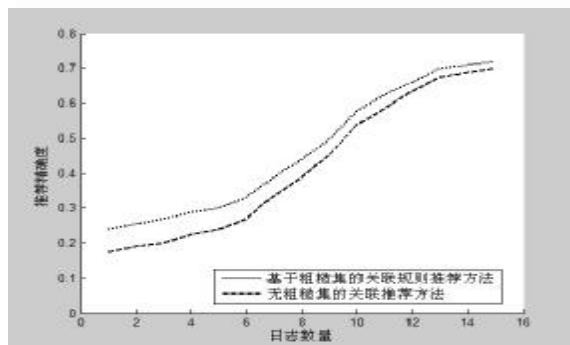


图 4 推荐精确度的对比

6 结论

教育信息化带来的是信息量的急剧增长和对信息提取的更高要求，必须借助数据挖掘技术去发现数据中隐藏的规律和模式，系统将基于粗糙集的关联规则应用于网络教学资源、立体化教材数据资源中，充分利用站点上积累的各种信息，提取并分析不同学生的学习习惯，将分析的结果及时反馈给网站管理员，以达到合理配置资源，充分利用立体化教材的目的，提高教材质量，推动教学的现代化。

参考文献

- 1 Pawlak Z. Rough sets: theoretical aspects of reasoning about data. Dordrecht, The Netherlands:Kluwer Academic Publishers. 1991.
- 2 苗夺谦,李道国.粗糙集理论、算法与应用.北京:清华大学出版社, 2008.
- 3 Han JW, Kamber M.范明, 孟小峰译.数据挖掘概念与技术.机械工业出版社, 2007.
- 4 张云涛,龚珍.数据挖掘原理与技术.北京:电子工业出版社, 2004.
- 5 白秀玲,崔林,王向阳.一种基于关联规则挖掘的粗糙集简约算法.计算机工程与应用, 2003,39(10):185 - 186.
- 6 Pandey A, Pardasani KR. Rough Set Model for Discovering Multidimensional Association Rules. International Journal of Computer Science and Network Security, 2009,9(6):159 - 164.
- 7 Tomasz Str, Kowski Henryk, Rybiski. A distributed version of apriori rule generation based on rough set theory. Workshop Concurrency, Specification and Programming, 2004.