

网络学习中学习者个性挖掘方法的研究

刘均, 李人厚, 郑庆华

(西安交通大学电子与信息工程学院, 710049, 西安)

摘要: 提出了一种面向个性化网络学习的学习者个性挖掘方法, 以实现网络学习中学习者个性特征的自动获取. 该方法首先通过对样本学习者的个性属性归并, 去除冗余的个性属性; 进而通过对学习者行为模式与个性特征的定量关联规则分析, 得到两者间的关联规则; 最终据此规则对后续学习者的行为模式进行推理, 获取学习者的个性特征. 此方法不仅解决了个性化网络学习中学习者个性特征的自动获取问题, 而且还可表现出个性与行为间的关系. 该方法已在计算机教学管理示范系统中进行了试验, 结果表明该方法是有效的.

关键词: 个性挖掘; 个性特征; 网络学习

中图分类号: TP31 **文献标识码:** A **文章编号:** 0253 - 987X(2004)06 - 0575 - 04

Study on the Personality Mining Method for Learners in Network Learning

Liu Jun, Li Renhou, Zheng Qinghua

(School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: A personality mining method is proposed to obtain the personality characteristics of learners automatically in the network learning. In the method, the redundant personality attributes of learners are wiped off through attribute reduction, and then the association rules of the personality characteristics and learning behavior patterns are obtained through the analysis of the learning logs. Finally, the learners' personality characteristics are deduced from their learning behaviors by using the rules above. The method not only solves the issues of the automatic acquisition of the personality characteristics in the personalized network learning, but also indicates the relationship between the learners' personality characteristics and behaviors. The method has already been implemented and tested in our computer-managed instruction system, and the result manifests that the proposed method is effective.

Keywords: *personality mining; personality characteristics; network learning*

目前,网络教育的研究重点已从如何解决传统教学中的时空制约问题,逐步转移到如何基于现代教育学、心理学等理论,构建个性化的虚拟学习环境.此方面需要解决的核心问题是如何根据学习者的认知水平、性格、动机、偏好等个性化属性,制定相应的学习内容和学习策略,即如何实现学习内容的个性化导航以及教学策略的自适应生成.在内容的个性化导航方面,人们已开展了一些研究工作,并建立了学习内容的个性化导航机制^[1].然而,对于学

习策略的自适应生成,相应的机制在现有的网络学习系统中尚未得到体现,其根本原因在于难以准确、自动地获取学习者的个性特征.国内外在如何获取学习者学习行为与个性属性(研究过程中作为已知量)的关系上已进行了初步研究,但还未深入到对学习者的个性挖掘.有鉴于此,本文提出了一种面向个性化网络学习的学习者个性挖掘方法,其基本思路是:通过对样本学习者个性特征与学习行为模式的分析,挖掘出两者间关联规则;据此规则对后续学习

收稿日期: 2003 - 09 - 11. 作者简介: 刘均(1973~),男,博士生;李人厚(联系人),男,教授,博士生导师. 基金项目: 国家自然科学基金资助项目(60103022;60373105).

者的行为模式进行推理,动态地获取其个性特征.

1 个性挖掘方法

定义1 设 $C = \{c_1, c_2, \dots, c_n\}$ 是一个属性集合, V_i 是属性 $c_i \in C$ 的值域,则有:

(1) 对于任意 $v_{ij} \in V_i$, 二元组 (c_i, v_{ij}) 称为 $\{c_i\}$ 的一个属性模式;

(2) 若 $pattern_k$ 与 $pattern_l$ 分别为属性集 C_k 与 C_l 的属性模式, $C_k \cap C_l = \emptyset$ 且 $C_k, C_l \subset C$, 则 $pattern_k \cup pattern_l$ 为属性集 $C_k \cup C_l$ 的一个属性模式.

定义2 设 $P = \{p_1, p_2, \dots, p_m\}$ 是与学习者个性化学习相关的性格、动机、风格 3 方面个性属性的集合, 则个性模式为 P ($P \subseteq P$) 上的一个属性模式.

定义3 行为模式是学习行为特征的定量描述. 设 $B = \{b_1, b_2, \dots, b_l\}$ 是学习行为的特征集合, 则行为模式为 B ($B \subseteq B$) 上的一个属性模式.

定义4 个性挖掘是指通过对样本学习者的学习行为模式与个性模式的分析, 获取两者间的规则, 并基于此规则实现学习者个性模式的自动获取.

网络学习环境下学习者的学习行为模式异构、多样^[2], 个性模式维数较多且属性值多为数值型, 简单地采用常规数据挖掘算法, 将导致挖掘过程复杂度高、精度低下. 针对此问题, 提出了如图 1 所示的个性挖掘工作机理: 通过对样本学习者个性调查获得个性模式, 并由个性属性归并降低个性模式的维数; 通过对学习日志的统计分析获得学习行为模式; 对个性特征与行为模式进行关联规则分析, 挖掘出行为-个性规则; 据此规则, 对当前学习者的行为模式进行推理, 从而获得学习者的个性特征.

1.1 个性属性归并

个性属性归并是指根据个性属性间的关系去除

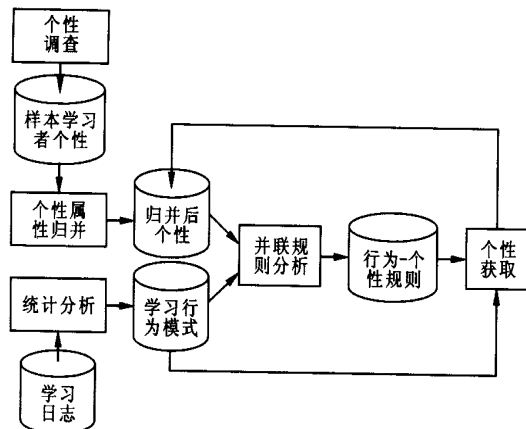


图1 个性挖掘的工作机理示意图

冗余属性,其目的是降低个性特征与行为模式的关联规则分析的复杂度.为了描述个性属性归并的机理,提出了“可归并”的概念.

定义5 若存在某个个性属性 $p_i \in P$, 对于 $\{p_i\}$ 的任意个性模式 $pattern_{p_i}$, 都存在对应某个属性集 P_j ($p_i \notin P_j$) 的 $pattern_{P_j}$, 使得关联规则 $pattern_{p_i} \Rightarrow pattern_{P_j}$ 与 $pattern_{P_j} \Rightarrow pattern_{p_i}$ 都成立, 则称属性 p_i 是可归并的.

个性属性归并就是发现并去除可归并的个性属性的过程, 属性之间的关系主要是通过关联规则分析获得的. 由于个性属性采用整型数值描述, 单纯地基于布尔关联规则 (Boolean Association Rules, BAR) 的分析方法还无法实现个性属性间关系的发现. 对于定量关联规则 (Quantitative Association Rules) 分析问题, 一般是通过将其转化为 BAR 问题解决的. 目前, 较常用的方法是将属性值域划分为多个区间, 每个区间映射为一个布尔属性^[3]. 该方法对于属性值域区间较大或精度要求较高的应用, 存在两方面的问题: 数据源的模糊化过程使得挖掘结果的精度较低; 划分区间的数目与属性值域的大小成正比, 当值域较大时, 将指数级地增加频繁项集挖掘算法的时间复杂度^[4].

有鉴于此, 提出了一种基于进制转换的定量关联规则分析方法, 即将二进位属性值的每一位作为两个布尔属性, 分别代表 0 与 1, 然后再采用 BAR 分析方法. 它的优点是在源数据类型转化过程中不会失真, 且布尔型属性的数目与值域大小的对数成正比, 即后序挖掘算法的空间与时间复杂度和值域大小成幂函数关系.

设某个属性 p 的值域长度为 L ($L > 1$), 若采用区间划分的映射方法, 设划分的最大区间为 s , 则至少需要 $\lceil (L + s - 1) / s \rceil$ 个布尔属性; 若采用所提方法, 则需要 $2 \lceil \lg(L - 1) + 1 \rceil$ 个布尔属性. 显然, 当 L 较大或者 s 较小时, $\lceil (L + s - 1) / s \rceil > 2 \lceil \lg(L - 1) + 1 \rceil$, 即采用划分区间的方法导致 BAR 过程具有更大的时间复杂度.

基于进制转换的定量关联规则分析方法的正确性等价于: 对于任意一条满足一定支持度与置信度的定量关联规则, 通过本文所提方法必生成一条满足同样支持度与置信度, 且前件与后件均为完整模式的布尔型关联规则; 反之亦然.

定义6 设 $p_i \in P$ 是某个个性属性, $P_i = \{p_{i1}, p_{i2}, \dots, p_{in}\}$ 是属性 p_i 对应的二进制属性集, 属性

$p_{i 2j-1}$ 与 $p_{i 2j}$ 分别表示某一位为 0 或 1, $2 \leq j \leq n$, 则有:

(1) 对于属性集 $P_i \subseteq P_j$, 若任意属性对 $p_{i 2j-1}, p_{i 2j} \in P$ 中有且仅有一个属性属于 P_i , 则称 P_i 上的属性模式为完整模式;

(2) 若 $pattern_k$ 与 $pattern_l$ 分别为二进位属性集 P_k 与 P_l 的完整模式, $P_k \cap P_l = \emptyset$, 则 $pattern_k \cup pattern_l$ 为属性集 $P_k \cup P_l$ 的一个完整模式.

上述情况(1)的完整模式也称平凡完整模式.

由定义 5, 进而得到如下定理.

定理 1 设 $p_i \in P$ 是某个个性属性, $P_i = \{ p_{i1}, p_{i2}, \dots, p_{in} \}$ 是属性 p_i 对应的二进制属性集, 若 p_i 是可归并的, 当且仅当满足条件: 对于所有 P_i 的平凡完整模式 $pattern_i$ 均存在某个完整模式 $pattern_j$, 使得 $pattern_i \Rightarrow pattern_j$ 和 $pattern_j \Rightarrow pattern_i$ 均成立. 其中, $P_i \subseteq P_j$, 且任意属性对 $p_{i 2j-1}, p_{i 2j} \in P$ 中有且仅有一个属性属于 P_i .

下面是个性属性归并的过程描述.

STEP1: 将个性属性由十进制整型转化为布尔型, 转化式为

$$((p_i, v_i))_{10} = ((p_{i1}, v_{i1}) \dots (p_{ij}, v_{ij}))_2 \quad (1)$$

式中: (p_i, v_i) 为 $\{ p_i \}$ 的某个属性模式, v_i 的范围为 $[\min_i \dots \max_i]$; (p_{ik}, v_{ik}) 为布尔型属性集合 $\{ p_{ik} \}$ 的属性模式, 每一位仅对应一个二进制属性模式, $1 \leq k \leq j, j = \lceil \lg(\max_i - \min_i) + 1 \rceil, v_{ik}$ 恒为 1.

STEP2: 对布尔型属性之间的关联规则进行分析. 采用 Apriori 算法获得满足一定支持度与置信度的规则. 由于挖掘出来的关联规则必须是由完整模式构成的, 因而可基于此对 Apriori 算法进行改进. 根据定义 6, 模式中不能同时出现对应于同一数值型某一位的两个布尔型属性, 若发现此类型的模式, 则可去除此模式以及包含它的其他模式.

STEP3: 去除结果中非法的关联规则, 非法规则是指那些非完整模式构成的规则.

STEP4: 在规则结果集中寻找规则组, 该组规则的后件是某个数值型个性属性对应的所有平凡模式, 且规则的前件与后件倒置后仍属于规则结果集, 则由定理 1 可知此数值型个性属性是可归并的. 设规则组有 m 个规则, 每个规则的置信度为 c_i , 则定义规则组的置信度为 $\sum_{i=1}^m c_i / m$. 重复该步骤, 直到找不到符合条件的规则组.

STEP5: 对得到的规则组按置信度由高到低进行排序, 并根据此次序逐个去除可归并的数值型个性属性.

1.2 个性特征与学习行为的关联规则分析

此部分主要用于获取学习者行为与个性之间的关系. 数据源包括样本学习者的学习行为模式与归并后的个性属性. 现有网络学习系统中, 学习行为主要包括课件学习、BBS 讨论以及师生多模式交互等, 可分别采用以下元组进行描述.

(1) 课件学习: (学习者 ID, 可分配单元 ID, 入口, 停留时间).

(2) BBS 讨论: (学习者 ID, 帖子 ID, 发帖类型, 字数, 发帖质量等级).

(3) 师生多模式交互: (学习者 ID, 实时课堂 ID, 交互类型, 交互时间).

行为模式中的各个特征一般也采用数值进行描述, 因而对其与个性特征之间的关联规则也应采用数值型分析方法, 分析获得以下形式的规则, 即

$$b \in B, B \subseteq B \quad ((b, S), S \subseteq S) \quad ((p, T), T \subseteq T) \quad (2)$$

式中: B 为所有行为模式的属性集合, b 为 B 的一个非空子集; S 为 B 中的某个属性 b 的所有可能取值的集合, S 为 S 的一个非空子集, 表示属性 b 在此规则中取值的集合; T 为某个个性特征属性 p 所有可能取值的集合, T 为 T 的一个非空子集. 上述规则的前件是各个行为模式特征的“合取”, 后件则是某个个性模式.

根据该规则的置信度, 需要对规则中个性属性的权重进行调整. 若某个个性属性仅在一条规则的后件中出现, 则直接采用该规则的置信度作为权重; 若个性属性存在于多个规则的后件中, 则采用各个规则置信度的加权平均作为个性属性权重

$$w = \frac{\sum_{r \in R} \text{sup}(r) \text{conf}(r)}{\sum_{r \in R} \text{sup}(r)} \quad (3)$$

式中: R 为包含某个个性属性的规则集合; $\text{sup}(r)$ 与 $\text{conf}(r)$ 分别为集合中某个规则的支持度与置信度. 采用支持度作为权重, 有效避免了在个性聚类中高置信度、低支持度(或相反)规则所导致的属性权重的偏差.

1.3 个性特征的获取

个性特征获取是指通过将当前学习者行为模式与所获规则的前件进行匹配, 获得其个性特征, 过程如下.



STEP1:将学习者行为模式逐条匹配规则的前件,获得一组个性属性值.

STEP2:若存在某个属性具有不一致的属性值,则取置信度最高的规则生成的属性值.

STEP3:采用个性属性值与相应权重的乘积构建个性特征向量.

STEP4:根据所获个性特征向量,对学习者的行为模式按照个性聚类的结果进行归类.

2 实验

上述研究内容已在教育部“教育振兴行动计划”项目——计算机教学管理(CMI)示范系统中得以实现,并结合该系统进行了实验.首先,对我校网络教育学院260名学生进行了个性测量,并取220名学习者的数据作为挖掘的数据源,其余学习者的数据作为检验数据.第二,根据个性属性的取值范围,将数据转化为布尔型属性,共得到116个属性.第三,采用Apriori算法,以 $s_{\min} = 0.15$, $c_{\min} = 0.8$ 进行关联规则分析,共得到1007209个关联规则,其中的前、后件均为完整模式的规则,共有101个.由此分析出可归并的属性是所有的动机属性,以及风格中的场独立依存性.第四,取 $s_{\min} = 0.15$, $c_{\min} = 0.6$ 进行个性特征与学习行为之间的关联规则分析.由于应用条件限制,在个性属性与学习行为关联规则分析方面,只作了BBS发帖以及课件学习两方面学习行为的分析,共得到33195个规则,其中的前、后件均为完整模式的规则,共有16个.第五,将后续学习者的行为模式去匹配上述规则,得到学习者的个性特征.最后,对40名学习者的挖掘结果数据与测量

数据进行对比,对比结果验证了个性挖掘方法的正确性.

3 结论

为了实现网络学习中个性化策略的生成,本文提出了一种面向网络学习的个性挖掘方法.该方法通过对样本学习者个性特征以及学习行为模式的分析,挖掘出两者间的关联规则;根据此规则以及后续学习者的学习行为,可自动获取学习者的个性特征.该方法已在CMI示范系统中进行了实现与验证.相对于个性测量方法,个性挖掘方法实现了学习者个性的自动获取,为个性化学习策略的自适应生成奠定了基础.

参考文献:

- [1] Chen Pinde, Li Kedong. Design of student model centered Web based adaptive learning system [A]. Society for Information Technology in Teacher Education Conference, Nashville, USA, 2002.
- [2] Cooley R, Mobasher B, Srivastava J. Web mining: information and pattern discovery on the world wide web [A]. The 9th IEEE International Conference on Tools with Artificial Intelligence, Newport Beach, USA, 1997.
- [3] 陈富赞,寇纪淞,李敏强.基于网络的数值关联规则挖掘方法[J].系统工程理论与实践,2002,4(4):1~10.
- [4] Agrawal R, Srikant R. Fast algorithm for mining association rules [A]. International Conference on Very Large Databases, Santiago, Chile, 1994.

(编辑 苗凌)

[文摘预登]

一种侧链含偶氮的聚噻吩甲烯的合成以及三阶非线性研究

孟令杰, 吴洪才, 易文辉, 高潮, 张远

(西安交通大学电子与信息工程学院, 710049, 西安)

为了提高聚噻吩甲烯的非线性光学性能,通过4-醛基-4'-二甲氨基偶氮苯和噻吩的缩聚反应,合成了聚噻吩对二甲氨基偶氮苯甲烯(PTDMAAB),PTDMAAB与四氯苯醌发生脱氢反应合成了侧链含偶氮的聚噻吩甲烯(PTDMAABQ),并用红外光谱、核磁共振氢谱、紫外-可见吸收光谱对其进行了结构表征.与PTDMAAB相比,PTDMAABQ的红外谱图在 1650 cm^{-1} 附近出现肩峰,紫外-可见吸收光谱上 λ_{max} 吸收峰红移35 nm,都说明了醌化反应的发生和共轭结构的形成.X衍射表明PTDMAABQ主要呈非晶态,但存有微小晶区.热重分析显示PTDMAABQ的热分解温度在275℃以上.用Z-scan系统研究表明PTDMAABQ薄膜具有大的非线性折射率和三阶非线性系数,分别为 4.926×10^{-6} 和 $3.644 \times 10^{-9}\text{ esu}$.