

基于 Web 和数据挖掘的 ITS 系统的设计与实现

The Design and Realization of Intelligent Tutoring System Based on Web and Data - mining

乔向杰 王万森 (首都师范大学信息工程学院 100037)

摘要:网络化和智能化是计算机辅助教学发展的两大趋势,基于网络环境的智能教学系统(ITS)的研究已经成为当前和今后相当长时期的一个重要课题。本文结合“首都师范大学智能教学系统”的开发过程,主要讨论了系统模型的构建、数据挖掘技术的应用以及知识库、教师模型、认知型学生模型的建构、基于模糊理论的知识表示和推理等关键技术和实现方法。目的在于借助通用的、一般的技术手段,为学习者提供一个自主的、智能化、个性化的、有利于学生意义建构的学习环境。

关键词:ITS 数据挖掘 认知型学生模型 模糊知识表示

本文结合“首都师范大学智能教学系统”的研发过程,在典型的 ITS 系统模型的基础上,利用超媒体技术、人工智能技术、信息集成技术、数据挖掘技术和组件技术等,建立起一个具有开放式结构、智能化、个性化、通用的网络教学系统。

1 系统的总体设计

1.1 功能模型(如图 1)

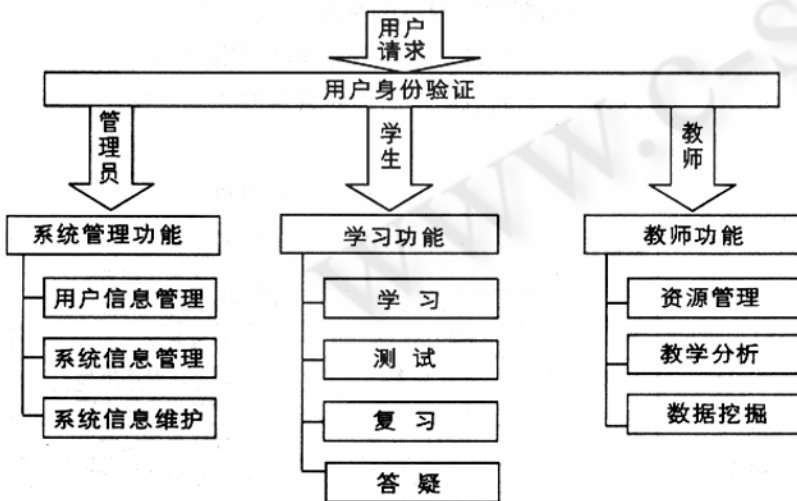


图 1 功能模型图

本系统主要由三类用户组成:管理员、教师和学生。其中管理员主要负责对系统信息的管理和维护及用户信息和权限的管理等;教师主要对课件、试题库进

行管理,对学生的学习行为进行分析和评价,依据数据挖掘的结果及时进行教学策略和教学进度的调整和改进等;学生针对系统提供的学习界面展开自主化、个性化的学习、测试和复习、答疑等。

1.2 结构模型

基于 Web 和数据挖掘的智能教学系统模型主要采用数据挖掘技术通过对站点上积累的大量信息进行分析,以发现用户感兴趣的模式和规则,为课程的设计者和管理者提供有关改进课程设计和重构站点的信息,为老师提供有关学生学习课件的情况,根据学生的学习情况,为其提供不同难度的学习内容,以建立起一个智能化、个性化的远程教育环境。该模型的结构如图 2 所示。

1.3 系统的实现

(1) 系统架构——浏览器/服务器模式(B/S)

由于本系统面向的对象范围广,如果采用 C/S 模式,在每一个客户端都必需安装一个应用程序,就会加大人力、物力、财力的损耗,并且一些微小的改动或版本升级都会带来维护困难。而采用 B/S 模式,用户只需在客户端安装浏览器软件,就可上网学习;统一简洁的浏览器界面减少了用户熟悉各种界面所需的大量培训以及客户端的维护工作。

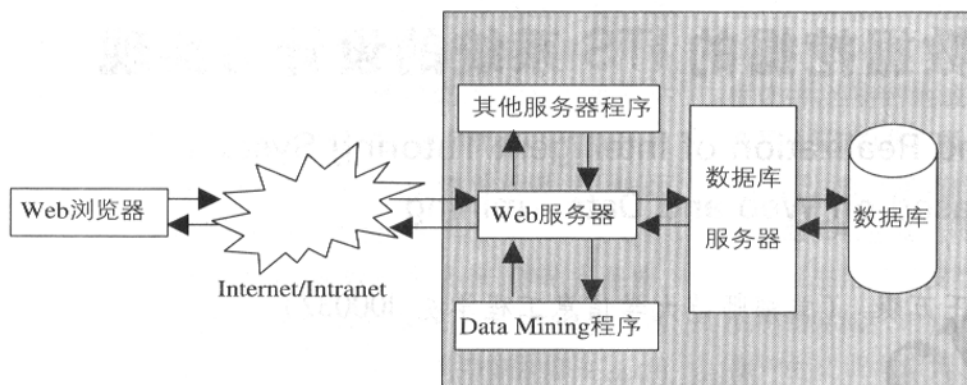


图 2 基于 Web 和数据挖掘的智能教学系统模型

- 最临近分类、基于案例的推理、遗传算法、粗糙集方法、模糊集方法等^[4]。在本系统中，由于选用的数据为 300 条学生记录，数据量相对不大；所分析的学生属性大致分为两类，静态属性，包括年龄层次、性别、学历等，动态属性，包括在线学习时间、学习频度等，离散化程度较好，同时各个属性之间没有很大的相互关系；在个性化前台学习环境的构建要求下，

分类规则应该易于描述，以及规范化，以便于后续的灵活处理。基于以上考虑，该系统主要采用判定树归纳分类的方法 ID3 版本。

该系统中，选定的属性集为 { 性别，最高学历，每周学习时间，自学能力 }，类标号为不同的难度等级 { 初级、中级、高级 }，通过对于学生动态属性和静态属性的分类，提供给学生不同难度等级的内容。经过挖掘算法分析处理后，得到了如图 3 所示的决策树。

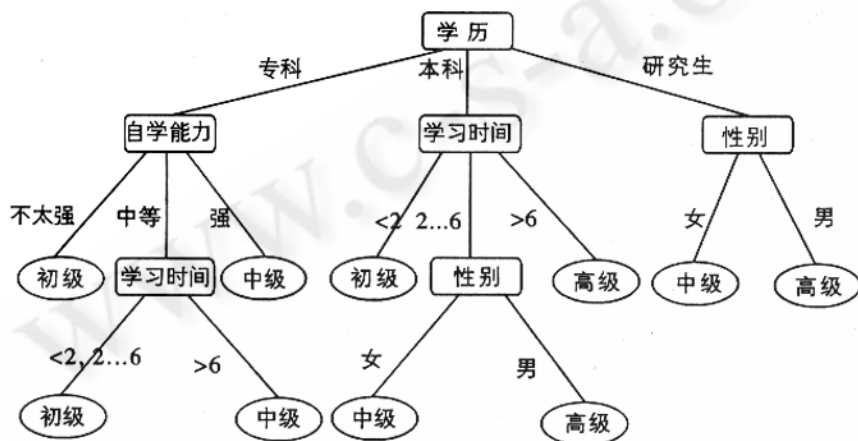


图 3 经过 ID3 算法形成的判定树

(2) 系统开发环境和开发工具

- 操作系统: Windows2000 Server
- Web 服务器: Apache + Tomcat
- 数据库服务器: SQL Server2000
- 数据库访问: JDBC 技术
- 开发工具: JAVA、JSP + JavaBeans + Applet

2 关键技术

2.1 数据挖掘技术

通过学生填写注册信息，将每个学生的情况都存入到学生档案库中，这样就可以根据这些信息，将学生进行分类，从而采取相应的教学策略，呈现不同的教学内容。

目前，对于不同的数据类型，不同的数据规模存在着多种的分类方法，如：判定树归纳分类、贝叶斯分类、后向传播分类、基于源自关联规则挖掘概念的分类、k

2.2 知识库的构建(如表 1)

2.3 认知型学生模型的构建

学生模型就是用以表示学生实际认知状况，并通过解释学生的活动得出他对领域知识和技能的掌握情

表 1 知识库的组成

领域知识库	教学内容库	存放教学知识点
	试题库	存储测试题目
教师模型	教学策略库	由一定顺序的知识点组成
	教学规则库	存放教学规则
学生模型	学生档案库	存放学生基本信息
	学习历史库	存储学生已学内容
	测试历史库	存储学生测试情况
	学生认知库	存放学生认知水平
	学习/测试断点库	存放学生上次学习和测试断点

况，是为了反映学生在学习过程中的思维活动、学习方法、对新知识的应用等方面而建立的数学模型。根据学生知识表示方法的不同，学生模型可分为三类^[5]：覆

盖模型、偏差模型和认知模型。前两种模型都只强调了知识而忽视了认知能力在学习过程的作用,而学生的认知能力在学习过程中将会起很大的作用,甚至是决定性的作用,因此,建立认知型学生模型成为一个新研究领域。

2.4 教师模型的构建

2.4.1 教学策略库

教学策略库是由一系列的教学知识点所构成,该库结构只有一个字段即知识点编号,是由教师根据学科或教学特点动态生成的。同时对于知识点序列是有顺序要求的,并由它来决定教学序列,如果学生学习能力较强,还可以跳过某些知识点。如:教学策略库 1 的教学序列为:0100 - 0101 - 0102 - 0103 - 0104 - 0105 - 0106,而教学策略库 2 的教学序列为:0100 - 0103 - 0104 - 0105 - 0106,等等。

2.4.2 教学规则库

由于本系统中对于教师的经验知识采用的是模糊知识表示,因此在规则库中也为每一规则增加了可信度,以优化推理过程,具体包括以下规则:

规则 1: IF (学习成绩好) AND (认知能力较高) THEN (继续下一章的学习, 0.9);

规则 2: IF (学习成绩好) AND (认知能力一般) THEN (继续下一章的学习, 0.8);

规则 3: IF (学习成绩一般) AND (认知能力一般) THEN (复习当前章, 0.8);

规则 4: IF (学习成绩一般) AND (认知能力较低) THEN (重复当前章的学习, 0.8);

规则 5: IF (学习成绩较差) AND (认知能力较低) THEN (重复当前章的学习, 0.9);

2.4.3 模糊知识表示和推理

在 ITS 系统中,按知识的性质,一般可将知识分为领域知识(事实性知识)和经验知识(过程性知识)两种。在本系统中,领域知识是用利用关系数据库按照知识树的结构组织的;经验知识由于是教师在教学过程中积累的教学经验,受人类自然语言表达方式以及非确定性思维方式的限制,这类知识常常带有模糊性、不确定性,因此,需要引入模糊思想,运用模糊知识表示方法表示这类非确定性知识,并在系统的推理过程中运用模糊推理机制,使系统的推理客观、公正,做出的判断也更为合理、有效^[6]。

(1) 相关定义

设 U 是给定论域(即问题所限定的范围), μ_f 是把任意 $u \in U$ 映射为 $[0, 1]$ 上某个值的函数,即 $\mu_f: U \rightarrow [0, 1], u \rightarrow \mu_f(u)$, 则称 μ_f 为定义在 U 上的一个隶属函数,由 $\mu_f(u)$ (对所有 $u \in U$) 所构成的集合 F 称为 U 上的一个模糊集, $\mu_f(u)$ 称为 u 对 F 的隶属度^[7]。

(2) 模糊集的运算

在系统中,我们常用到的是两个模糊集的交集和并集,它们的隶属函数分别为:

$$\mu_{F \cap G}(u) = \min\{\mu_f(u), \mu_g(u)\} \text{ 或 } \mu_{F \cap G}(u) = \mu_f(u) \wedge \mu_g(u)$$

$$\mu_{F \cup G}(u) = \max_{u \in U} \{\mu_f(u), \mu_g(u)\} \text{ 或 } \mu_{F \cup G}(u) = \mu_f(u) \vee \mu_g(u)$$

(3) 模糊知识表示

① 对前提、结论中涉及到的教学特性进行分类,如:“成绩”特性可分为“好”、“中”、“差”三个级别。

② 提取出教学特性中涉及到的模糊修饰语,如“很”、“非常”、“稍微”等。

③ 确定隶属函数,对每一种教学特性进行模糊处理。即对教学特性的每一种模糊状态使用隶属函数来表示。

④ 用隶属函数描述各种模糊修饰语。

· 表示否定,如“不”、“非”等,其隶属函数的表示为: $\mu_{\text{非}f}(u) = 1 - \mu_f(u) \quad u \in [0, 1]$

· 表示“很”、“非常”等,其效果是减少隶属函数的值: $\mu_{\text{非常}f}(u) = \mu_f^2(u) \quad u \in [0, 1]$

· 表示“有些”、“稍微”等,其效果是增加隶属函数的值: $\mu_{\text{有些}f}(u) = \mu_f^{1/2}(u) \quad u \in [0, 1]$

⑤ 设置规则的可信度为 $CF(0 < CF < 1)$ 。

(4) 模糊推理的基本过程

① 初始化事实数据库,把欲解决问题的已知事实送入事实库中。

② 检查规则库中是否存在尚未使用过的规则,若有则执行③;否则就终止;

③ 检查规则库的未使用规则并与已知事实进行模糊匹配并得到相应的结论;否则转⑥。

④ 执行当前选中规则,并对该规则作上标记,把执行该规则后所得到的结论作为新的事实放入事实库;如果该规则的结论是一些操作,则执行这些操作。

⑤ 检查事实库中是否包含了该问题的解,若已包

(下转第 60 页)

(上接第15页)

含,则说明已求出解,在结论集中选择具有最大可信度的结论作为推理的最终结论,问题求解过程结束;否则,转②。

⑥ 当规则库中还有未使用的规则,但均不能与事实库中的已有事实相匹配时,要求用户进一步提供关于该问题的已知事实,若能提供,则转②;否则,说明该问题无解,终止问题求解过程。

若一学生在测试中测得其学习成绩为92,认知能力为0.81,则可分别求得其学习成绩对于“好”、“一般”、“差”的隶属度为0.7、0.3、0;认知能力对于“高”、“一般”、“低”的隶属度为0.1、0.9、0。然后对于每条规则进行模糊集的交通算,得出相对于每条规则的可信度。对于规则1: $0.7 \wedge 0.1^2 \wedge 0.9 = 0.01$;对于规则2: $0.7 \wedge 0.9 \wedge 0.8 = 0.7$;对于规则3: $0.3 \wedge 0.9 \wedge 0.8 = 0.3$;对于规则4: $0.3 \wedge 0^2 \wedge 0.8 = 0$;对于规则5: $0^2 \wedge 0^2 \wedge 0.9 = 0$,因此,选取具有最大可信度0.7相应的规则2,即要求该生可以继续下一章的学习。

3 结束语

本文从实现的角度提出了一完整的ITS系统的实

施策略和方法,将数据挖掘技术应用到系统中,使得系统具有一定的自适应性;教师及教学专家可实时监督学生的学习情况,实时更新知识点库、试题库、教学策略库、教学规则库等,从而为不同学生提供不同的教学进度和教学内容。

参考文献

- 1 王陆、王美华,ITS系统中基于关系模型的知识表示[J],北京大学学报(自然科学版),2000,36(5):660-663,
- 2 焦加麟、徐良贤、戴克昌,人工智能在智能教学系统中的应用[J],计算机仿真,2003,20(8):49-51。
- 3 顾纪鑫、丁煜,教学新概念——网上远程教学,中国远程教育[J],2000(1):44-46。
- 4 (加) Jiawei Han Micheline Kamber 著,范明、孟小峰等译,数据挖掘概念与技术[M],机械工业出版社,2001。
- 5 王萃寒、赵晨等,智能计算机辅助教学系统的实现研究[J],计算机工程与科学,2003,25(3):83-85。