

以知识点为中心的智能教学系统的研究与构建

康健¹, 刘昌明², 罗楠楠³

(1. 东北师范大学理想信息技术研究院, 长春 130117; 2. 空军航空大学, 长春 130022;
3. 中国科学院长春光学精密机械与物理研究所, 长春 130033)

摘要: 在深入研究智能教学系统知识表示方法的基础上, 提出了一个以知识点为中心的智能教学系统构建方案, 采用知识点作为教学资源的组成单位, 通过知识点间的联系来构建知识体系, 并结合Web挖掘、人工智能、模糊理论等技术, 实现网络教学的智能教学策略。

关键词: 智能教学系统; 知识表示; Web数据挖掘; 人工智能; 模糊理论

Research and Construction of Intelligent Tutoring System based on knowledge-point as the center

KANG Jian¹, LIU Changming², LUO Nannan³

(1. Northeast Normal University, Changchun 130117; 2. Aviation University Of Air Force, Changchun 130022;
3. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033)

Abstract: On the basis of deeply research into the knowledge representation methods of the intelligent tutoring system, the paper puts forward a scheme of the intelligent tutoring system based on knowledge-point as the center, adopts the knowledge point as the composition unit of the educational resources and establishes knowledge system through the relationship between the knowledge points. The paper implement the intelligent tutoring tactics combining Web datamining, artificial intelligence, and fuzzy theory.

Key words: ITS; Knowledge Representation; Web Datamining; Artificial Intelligence; Fuzzy Theory

1 引言

网络教育是当今国际国内教育的主流发展方向, 其开放性、协同学习、共享资源、无时空限制等特点, 为学习者构建了一个随时随地自主学习的终身学习环境。本文提出的智能教学系统是以知识点为中心, 通过知识点间的联系来构建知识体系。实现网络教学的智能化。

2 访问记录的Web采集

智能教学系统的智能性主要体现为个别化教学。要实现个别化教学就必须对学习者的情况有足够的了解, 在网络教育过程中, 学习者通过浏览器访问服务器中教学资源的同时会在服务器中留下访问日志。这些访问日志为分析学习者的学习情况和实现个别化教学提供了宝贵的数据。我们可以通过Web日志挖掘的相关技术对服务器日志进行提取和分析。

目前Web日志挖掘的数据源主要分为两类, 一类是直接把Web服务器上的log日志作为Web日志挖掘的数据源。另一类是自定义用户访问记录作为Web日志挖掘的数据源^[6]。相对于服务器log日志, 自定义的用户访问记录并不是将用户访问的所有对象都记录下来, 而是有选择性地记录, 并且可以根据使用的需要剔除与挖掘无关的数据, 减轻数据预处理阶段的负担。所以本文选择自定义访问记录作为数据源。用户访问记录的格式定义如图1。

数据项	说明
UserID	用户标识
UserIP	代理
UserAgent	用户IP地址
DateTime	访问日期时间
Referrer	页面请求的来源地
TargetURL	对象URL
Key	查询关键字
TargetType	访问对象类型

图 1

本文通过在Web应用程序中定义一个servlet, 然后把对于该网站的请求转发给此servlet, 并在servlet中判断此访问页是否需要记录访问信息, 如果需要则根据访问记录的格式将相应的用户访问记录插入到数据库中。

3 知识体系的构建

在智能教学系统中, 知识的表示是教学系统智能化的基础, 本文主要以知识点为中心, 通过知识点间的联系构建智能教学系统的知识体系。并且引入模糊理论对知识点属性及其关联关系进行描述, 为智能化教学的实现打下坚实的基础。

3.1 知识点及知识点间的关系定义

知识点是知识的完整教学单元, 一个网络课程一般包括

很多章，每一章又包括若干节，每一小节又有多个知识点，而题库中的每道题也会和一个或者几个知识点相关联，来表示本题目考察的知识点。知识点有难度和重要性等属性。知识点之间存在的各种关联及其关联程度称为知识点间关系。它包括组成关系、前驱关系和相关关系。组成关系是指一个知识点可以由若干知识点聚合而成。前驱关系是指当前知识点和学习当前知识点之前应该掌握的知识点间的关系。相关关系是指在内容或逻辑上存在相关性的知识点间的关系。

知识点形式化定义如下：

知识点： $:= \langle \text{知识点编码}, \text{知识点名称}, \text{关键词}, \text{大纲教学目标}, \text{学习内容}, \text{难度}, \text{重要程度}, \text{学习时间}, \text{对应教学资源所在路径} \rangle$

知识点关系： $:= \langle \text{组成关系} \mid \text{前驱关系} \mid \text{相关关系} \rangle$

知识树 = $\{ \langle \text{知识点} \rangle, \langle \text{知识点关系} \rangle \}$ ---- 知识点及其关系的集合称为知识树。

3.2 结合模糊理论对知识点和知识点间关系的属性进行描述

(1) 知识点属性

知识点属性中知识点难度、重要度都属于不精确属性，可以结合模糊理论思想进行描述。其中知识点难度可以通过该知识点测试试题的通过率来衡量，其模糊隶属函数可定义为： $Diff(x) = \lfloor (1-x) * 10 \rfloor$ ($x \in [0, 1]$) 使得知识点难度值取为1至10之间的整数，并且随着知识点测试通过率x的增加而降低。

知识点重要程度与其后继知识点个数以及关于该知识点的试题数有关，即某知识点被其它知识点用到的次数越多，该知识点就越重要。题库中关于该知识点的试题越多，该知识点越重要。其模糊集的隶属函数可定义为：

$$Im p(x, y) = y \left(\frac{x}{10} \right) \quad \text{当 } x \in [0, 10]$$

$$y * 1 \quad \text{当 } x > 10$$

其中x为后继知识点的个数。y=当前知识点的试题数/所有知识点相关试题数中的最大值。知识点的重要程度取值在0至1之间，并且随着x, y的增加而增加。

(2) 知识点关系描述

知识点关系： $:= \langle \text{关系编码}, \text{知识点编号1}, \text{知识点编号2}, \text{关系类型}, \text{关联系数} \rangle$

关系类型： $:= \langle \text{组成关系} \mid \text{前驱关系} \mid \text{相关关系} \rangle$

关联系数的取值在0至1之间，在组成关系中，关联系数代表子知识点在构成父知识点中所占的比重；在前驱关系中，关联系数代表知识点2依赖知识点1的程度；在相关关系中，关联系数代表两知识点内容的相关程度。

3.3 以知识点为中心构建试题库

在教学过程中需要有系统地提供相应的训练习题来巩固学生对知识点的学习，同时需要通过习题来考察学生对于相应知识点的掌握情况，为了做到个别化教学，我们需要根据学生的学习情况、能力水平由系统自动生成相应难度的训练

试题，既给学生一定的成就感又会增加学生继续探究的兴趣。

习题： $:= \langle \text{习题编码}, \text{习题类型}, \text{难度系数}, \text{重要性系数}, \text{习题内容}, \text{习题答案}, \text{参考作答时间}, \text{考察知识点} \rangle$

本文主要以知识点为中心，通过知识点间的联系构建智能教学系统的知识体系，通过试题资源与知识点间的关联、学生学习情况与知识点间的关联，使得试题资源、知识点、学生学习情况成为一个有机的整体。此知识体系的具体数据表结构如图2所示。其中学生学习记录表是由图1的学习者访问记录表批量处理生成的。

4 模型构建

学生模型是智能导学系统中最重要的模型之一，很多教学理论的实现都需要以学生模型为基础。本文以知识点为中心，将学生的学习情况映射为对于相关知识点的掌握情况，学生对于相关知识点的掌握情况可以由一个七元组 {SID, KPID, T, D, I, E, C} 来表示。其中：

SID：学习者编号，是学习者的唯一标识。

KPID：知识点编号，是知识点的唯一标识。

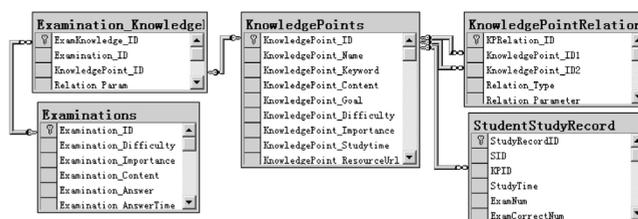


图 2

T：知识点学习时间，基于这样的假设，学习时间越长，该知识点的掌握情况越好。

D：知识点难度系数，基于这样的假设，难度系数越大的知识点越难于掌握。

I：知识点重要系数，基于这样的假设，越重要的知识点越应该掌握。

E：学生做过的和该知识点相关的练习题的总数（包括答对和答错的习题）。

C：E中答对的题目数。

学习者学习情况衡量函数定义为：

$$G(SID, KPID) = \left(\frac{T}{D} \right) * \left(\frac{C}{E} \right)$$

即与单位难度内的学习时间成正比，与答题正确率成正比。

5 教学策略的实现

本系统的智能教学过程主要是根据知识点间的关系和学生对于知识点的掌握情况来实现智能化教学策略。

5.1 学习路径引导

主要基于循序渐进和因材施教的原则，根据学生的学习情况对学习路径进行控制。具体解决方案如下：

(1) 在学生开始学习某知识点之前，搜索该知识点的重要前驱知识点，并且判断重要前驱知识点的掌握情况，若

(下转到24页)

多的样本时,存在标注瓶颈的问题;

(4)类别间的关系也更加复杂,需要有更好的类别组织方法

(5)Web文本是一种半结构化的数据,其结构信息(如链接关系、主题等)可能对分类提供某些帮助。

随着研究的逐渐深入,文本分类的应用越来越广泛,通过以上分析得出分类方法研究的主要趋势:

(1)对传统经典的分类方法进行改进;

(2)不断涌现新的分类方法;

(3)多种分类方法综合运用。另外,随着相关技术如:数据挖掘、人工智能和机器学习研究的深入,将很大程度上地促进文本分类的发展。

参考文献

[1] 于瑞萍. 中文文本自动分类相关算法的研究与实现[D]: 西安: 西北大学信息科学与技术学院, 2007.

(上接第4页)

均已经达到教学大纲所要求的掌握水平,则可以进行当前知识点的学习;否则转为对未达到要求的重要前驱知识点的学习,这是一个递归的过程。

(2)当学生完成了当前知识点的学习并通过了测试,系统为学生提供当前知识点的相关知识点,由学生自行选则扩展学习内容。学生也可以退出此流程转为其它知识点的学习。

以上关键问题的解决方案:

(1)重要前驱知识点的发现:根据知识点联系表,得到该知识点的所有前驱知识点,前驱知识点的重要程度由两方面衡量,一是前驱知识点本身的重要程度,二是前驱知识点和当前知识点间的依赖系数。根据 $I=Imp(x) \times \text{依赖系数}$ 来判断其重要程度。当 $p>0.5$ 时,为重要前驱知识点。

(2)知识点掌握情况的衡量:根据学习者学习情况衡量函数 $G(SID, KPID)$ 与知识点要求掌握的程度值进行比较来判断前驱知识点的掌握是否符合要求。

(3)相关知识点的发现:根据知识点联系表中,相关关系类型的联系来发现与该知识点相关的知识点。并且按照相关系数进行排列返回给学习者。

5.2 薄弱知识点的发现和针对性训练

根据学习者学习情况衡量函数 $G(SID, KPID)$ 对该学生的所有已学知识点的掌握情况进行排序,找出薄弱知识点。通过知识点试题联系表,找出适合该学生的训练试题集,从而进行针对性训练。

6 结语

本文提出了一种以知识点为中心的智能教学系统解决方案,通过知识点间联系,知识点与试题、学习者学习情况间的联系,实现智能教学策略。本文的创新点是构建了一个以知识点为中心的知识体系,对知识点属性进行了模糊描述,并且提出了一种新的知识点掌握情况的衡量函数。

[2] 李荣陆. 文本分类及其相关技术研究[D]: 上海: 上海复旦大学, 2004.

[3] 蒲筱哥. 自动文本分类方法研究述评[J]. 情报科学, 2008, 26(3): 469-475

[4] 曾雪强. 一种基于潜在语义结构的文本分类模型[J]. 华南理工大学学报(自然科学版), 2004, (32): 99-102.

[5] 苏金树, 张博锋, 徐昕. 基于机器学习的文本分类技术研究进展[J]. 软件学报, 2006, 17(9): 1848-1859.

作者简介

陈思,女(1984-),吉林长春人,东北师范大学理想信息技术研究院计算机软件与理论。

钱铭宇,男(1983-),吉林长春人,东北师范大学理想信息技术学院,计算机软件与理论。

刘昌明,女(1982-),吉林长春人,在读硕士,助教,主要研究方向为智能导学、人工智能和信息检索。

参考文献

[1] G D Magoulas, K A Papanikolaou, M Grigoriadou. Towards a Computationally Intelligent Lesson Adaptation for a Distance Learning Course. University of Athens, 2002: 535-540.

[2] Barsky D B, Shafarenko A V. WWW and Java-based distributed examination system for distance learning applications. Dept of Electron&Electr Eng, Surrey Univ, Guildford, UK, 2001: 224-231.

[3] 周晓军, 黄河燕, 张普. 智能教学系统中的知识树增长模型. 计算机研究与发展, 2001, 10: 1217-1223.

[4] 武修琪. 基于SCORM的网络智能教学系统知识树模型研究. 哈尔滨理工大学, 2005.

[5] 陆汝钤. 世纪之交的知识工程与知识科学[M]. 北京: 清华大学出版社, 2001. 447.

[6] 何典, 梁英. 动态网页环境下的Web使用记录挖掘研究. 微计算机信息, 1008-0570(2006)08-1-0122-03

[7] 张燕姑, 谢胜利, 李虎雄. 智能教学系统知识表示的设计与实现. 温州师范学院学报, 2003, 24(5): 83-87.

[8] 王陆, 王美华. ITS系统中基于关系模型的知识表示. 北京大学学报(自然科学版), 2000(3): 53-56.

作者简介

康健,男(1984-),吉林长春人,在读硕士,主要研究智能导学,人工智能和数据挖掘。

刘昌明,女(1982-),吉林长春人,在读硕士,助教,主要研究智能导学,人工智能和信息检索。

罗楠楠,女(1983-),中国科学院长春光学精密机械与物理研究所,在读硕士,主要研究图像处理,人工智能。