

# 贝叶斯网络在学生模型建模中的应用分析\*

陈丽花

(云南财经大学 信息学院, 云南 昆明 650221)

**摘要:** 将课程教学资源融合到学生模型构建中,描述了包括领域知识拓扑结构的建立、条件概率表学习算法的推理的详细过程,最终得到了学生模型中关于章节知识项的贝叶斯网络结构图,并通过一个实验系统对个性化教学系统中学生模型建构的整个框架的可行性进行了验证。

**关键词:** 个性化教学;贝叶斯网络;学生模型;建模

中图分类号: TP301

文献标识码: A

## Bayesian networks applied analysis in student model modeling

CHEN Li Hua

(Information School, Yunnan University of Finance and Economics, Kunming 650221, China)

**Abstract:** This article fuses the teaching resources in the student model, describes the model building detailed process, including the domain knowledge topology's establishment, the condition probability table learning algorithm's inference, obtains in finally the student model about the chapter knowledge item of Baye network structure drawing, and has carried on the confirmation through an experimental system to the personalized teaching system middle-school student model construction entire frame feasibility.

**Key words:** individualized teaching; Baye networks; student model; modeling

个性化教育是现代远程教育最重要的特点之一,其中计算机网络特有的信息、数据库管理技术和双向交互功能,为个性化教学提供了有效的实现途径与条件。在个性化教学系统中,学生模型是学生的认知状态表示。它的作用是为系统提供关于学生对知识的掌握情况的信息,以便能够对学生进行正确的评价和预测,从而提供个性化的学习内容。因此,如何建立一个准确、高效的学生模型是整个系统的核心。分析和表示学生模型的过程称为学生建模<sup>[1-2]</sup>。

### 1 有关贝叶斯网络理论知识

贝叶斯网络(Bayesian Networks)也称为信度网、因果网,主要用于表示变量之间的依赖关系,并为任何全联合概率分布提供一种简明的规范。贝叶斯网表达了各个节点间的条件独立关系,可以直观地从贝叶斯网中得出属性间的条件独立以及依赖关系;另外可以认为贝叶斯网以另一种形式表示出了事件的联合概率分布,根据贝叶斯网的网络结构以及条件概率表(CPT)可以快速得到每个基本事件(所有属性值的一个组合)的概率。贝叶斯

学习理论利用先验知识和样本数据来获得对未知样本的估计,而概率(包括联合概率和条件概率)是先验信息和样本数据信息在贝叶斯学习理论当中的表现形式。

### 2 学生模型中教学资源信息的建模

贝叶斯网络的建模目的是对所包含的定性知识和定量知识进行结构描述,定性部分由网络结构来描述,定量部分由条件概率分布函数来表示。从原始数据中构造贝叶斯网络模型,实质也是对原始数据进行数据挖掘:先找出最符合原始数据定性的网络图关系,然后根据网络图中的因果关系,计算节点间的条件概率。构造贝叶斯网络可以分为4个阶段:(1)定义域变量;(2)确定网络结构;(3)确定条件概率分布函数;(4)运用到实际系统中,并根据系统产生的数据优化贝叶斯网络。

在学生模型中教学资源信息的建模初期要考虑整个教学资源信息的网络结构,然后再考虑局部的知识结构:知识点的前驱、后继关系,知识点之间的相互影响程度,知识点的状态,以及与知识点相关的测试(包括识记、理解、应用、分析、综合等状态),只有经过测试,才能

\* 基金项目:校青年项目资助课题(80025090202)

## 技术与方法 Technique and Method

实现个性化教学。整个过程为将来个性化的评估做准备,根据评估结果实现贝叶斯网络的局部更新,也即学生知识状态的更新,贝叶斯网络根据更新过的学生状态,个性化地把内容呈现给学生<sup>[3]</sup>。

### 2.1 学生模型中整体教学资源的知识结构

本文把教学资源分为6个层次:课程知识、主知识项(按章划分)、一级知识项(按节划分)、二级知识项(也称复合知识项)、基本概念知识项、与上述知识相关的测试。层与层之间只有继承关系,同层内的知识可以是前驱和后继的关系,也可以是平行和关联等关系。可以认为前驱和后继关系是关联的一种特殊形式。

### 2.2 学生模型中局部教学资源的知识结构

#### 2.2.1 定义域变量和确定网络结构

学生模型的建模就是要针对学生所学课程掌握的知识结构进行建模,从而根据学生的知识状况给出学生学习状况的评价。

以《Delphi 程序设计教程》为例,按照课程的章节安排列出课程的知识结构,为了简化,只选取部分知识内容做出说明,整体结构可依次类推。比如在基本程序控制结构这一章,一级知识项(章)为:基本程序控制结构;二级知识项(节)有:顺序结构、循环结构、选择结构;复合概念有:While 循环、For 循环、Repeat 循环;基本概念有:关系运算符、赋值语句、变量、数据类型。需要在贝叶斯网络中加入测试项作为输入节点,也可称之为观测节点。每个节点代表一个测试该知识点的测试项。

节点的状态表示学生对该知识项的掌握程度,例如可以用4个等级来表示学生对知识项的掌握程度,A:表示学生对特定知识项已完全熟练掌握;B:表示学生对知识项基本掌握,但是仍有一些问题;C:表示学生处在初级阶段,有许多问题;D:表示学生对该知识项完全不了解。每一等级对应一个条件概率值(值的加权和为1),该值是基于对用户的观测值得到的用户对知识项的掌握程度。在系统设计过程中,为了简化,设置节点的状态为 True(掌握)、False(未掌握)。

#### 2.2.2 条件概率估计

经过上面的分析,接下来将讨论如何获得知识项之间的影响程度的条件概率表。

在贝叶斯网络构建中所有节点都要给定条件概率,对于没有父节点的节点,需要给定其先验概率,条件概率可以通过样本学习和通过专家估计获得<sup>[4]</sup>。

##### (1) 设计分析

在由数据类型(A)、关系运算符(B)、变量(C)、While 循环(D)教学资源构成的网络拓扑结构中,每个节点有两个离散的状态:True 和 False,如果需要,节点也可以包括更多的状态。设计网络时,必须指定每个节点的条件概率表,需要获得  $P(A)$ 、 $P(C)$ 、 $P(B)$ 、 $P(D|C)$ 、 $P(D|B)$ 。

##### (2) 样本收集

经验数据来自于某校某班学生学习“基本程序控制结构”这一章节后,经过测试,对其测试结果进行分析并获得学习知识项之间的相互影响程度,这样就确定了利用贝叶斯网络所建学生模型中网络拓扑知识项之间的条件概率函数。

将学生的得分情况分为 100~90、90~80、80~70、70~60、60 以下 5 个分数段。认为如果学生答对涉及到该知识项的题目,那么他对该知识项的知识状态为 True(掌握);反之,如果答错,则认为他对该知识项的知识状态为 False(未掌握)。

##### (3) 学习条件概率表<sup>[5-6]</sup>

算法的设计思想是:首先按照网络构建中节点的拓扑顺序,根据每个分数段随机抽取  $n$  份考试的试卷用于对网络中各节点的状态进行赋值,当所有节点都被赋值一遍后就得到该网络中的一次采样样本,依此重复,得到  $10n$  个样本,即一个样本序列;然后对该样本序列按相关列进行联合统计可得到网络近似推理结果。由于随机采样时考虑了节点的先验信息,所以当样本数  $n$  取很大时,这些采样样本的边缘统计量和条件统计量就可以趋近于事件的真实发生概率和条件发生概率。

对所有节点都采样了  $m$  遍,得到  $m$  组样本,即样本序列  $S$ ,然后对样本序列  $S$  按列进行单独统计,得到节点的全概率;对多列进行联合统计,得到节点间的联合概率和条件概率。

所谓联合统计,就是把证据节点和推理目标节点的采样样本联合起来进行统计,就得到了给定证据后目标节点发生的条件概率和联合概率。对网络中的所有节点按拓扑顺序采样  $m$  次,得到样本序列  $S$ ,对样本序列进行适当的统计,可以得到网络的近似推理结果。

节点  $N_i(i=1, \dots, n)$  发生的全概率就是单独对该列进行统计的结果。例如,要知道学生对于“变量”(C)已经 True 的情况的概率  $P(C=True)$ ,只要把对是 T 的那些样本挑选出来,然后统计一下 T 的个数,除以总的采样次数  $m$ ,就可以近似为 C 这一知识项学生已经 True(掌握)的概率,即:

$$P(C=True) = \frac{\text{正确答案数}}{\text{抽取试卷总数}}$$

然后得到

$$P(C=False) = 1 - P(C=True)$$

节点  $N_i(i=1, \dots, n)$  发生的条件概率就是把那些条件节点和该节点所在的相关列联合起来进行统计。例如,计算节点“D”在节点“B”已经 True(掌握)情况下发生 T 状态(D 也为 True)的概率,也即求  $P(D=True|B=True)$ ,可以这样进行:首先,把在“B”中是 T 的那些样本挑选出来,形成新的样本序列  $S^0$ ,其组数为  $n_0$ ;其次,把样本序列  $S^0$  中在“D”中是 T 的那些样本挑选出来,然后统计 T

## 技术与方法 Technique and Method

的个数,除以  $n_0$ , 就可以近似为“D”在“B”取状态为 True 的情况下发生 False 状态的概率:

$$P(D=True|B=True)=\frac{P(D=True, B=True)}{P(B=True)}$$

$$P(D=False|B=True)=1-P(D=True|B=True)$$

$$P(D=True|B=False)=\frac{P(D=True, B=False)}{P(B=False)}$$

$$P(D=False|B=False)=1-P(D=True|B=False)$$

类似地,几个节点的联合概率则是把同时满足这些节点状态的样本挑出来,得到的样本组数除以总的样本组数  $m$  即可。

单独的节点状态概率,可以使用条件概率规则,并沿着边进行传播。

网络设计完成后,条件概率表进行了指定,网络进行了传播。

通过上面的分析,从定性方面完成了学生所学知识项的逻辑关系,并且通过经验数据获得了关于这些知识项之间存在影响程度的定量数据,最终构建成的个性化教学中关于《Delphi 程序设计教程》课程中章节知识项的学生模型的框架如图 1 所示。

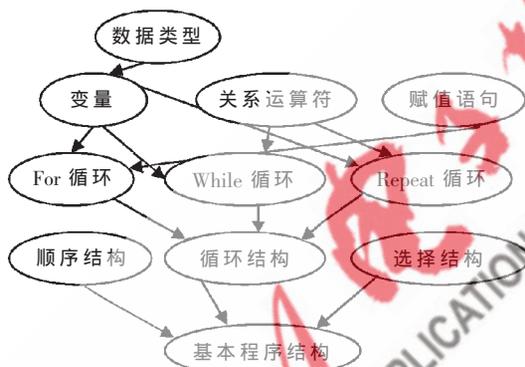


图 1 学生模型中关于章节知识项的贝叶斯网络结构图

### 2.2.3 学生模型原型系统的实现

根据上面理论方面的分析,已经构建了个性化教学系统中学生模型的贝叶斯网络结构图,为了证明整个系统框架的可行性,需要把前面的理论构建与具体的技术相结合以实现可用的系统。

该系统设计中运用不同的颜色来标记知识项掌握的 4 种情况:绿色表示已掌握;红色表示已学习但未掌握;黄色表示未学习,适合学习;灰色表示不适合学习。当贝叶斯网络获得证据信息后标记的颜色状态会发生变化。基于贝叶斯网络先验概率,知识项全部标记为黄色,如图 2 所示,新注册用户登录系统时可在黄色节点中点击选择最初学习节点,学习后对相应知识项进行测试,根据系统导航,若测试合格则按引导进行下一知识项的学习,未通过系统会提示返回上一个节点重新学习。



图 2 系统初始界面

从这个系统中可以看到:基于贝叶斯网络的个性化教学系统既有交互性,又体现了个性化,通过系统的导航功能(不同颜色的应用)可以帮助进行学习的指导,当学生想学某一概念时系统会把它的前驱知识点以序列形式呈现给学生,不同学生要求不一样,系统提供的导航也不一样,真正体现了个性化教学。

随着网络教学平台的不断涌现,个性化教学系统的研究和开发成为远程教育中的关键问题和热点。而在个性化教学系统中的学生模型是其中的重要组成部分,在远程教育领域起着重要的作用。因此,研究如何设计和利用学生模型来实现远程教育中的个性化教学有着重要的意义。

### 参考文献

- [1] 刘桓,曹红菊.自适应学习系统的研究与设计[J].教育技术导刊,2006,(3):39-40.
- [2] 余胜泉.适应性学习——远程教育的发展趋势.开放教育研究,2000,6(3):70-76.
- [3] 冀俊忠,刘椿年.贝叶斯网及概率推理在智能教学中的应用[J].北京工业大学学报,2002,9(28):353-358.
- [4] 冀俊忠,刘椿年,沙志强.贝叶斯模型的学习、推理和应用[J].计算机工程与应用,2003,7(5):145-147.
- [5] 胡兆勇,屈梁生.贝叶斯网络推理的一种仿真算法[J].系统仿真学报,2004,26(2):286-288.
- [6] 刘伟娜,霍利民,张立国.贝叶斯网络精确推理算法的研究[J].网络与通信,2006,22(9):92-94.

(收稿日期:2009-12-21)

### 作者简介:

陈丽花,女,1974年生,讲师,工学硕士,主要研究方向:人工智能,数据挖掘,计算机基础教学。