

知识点及其网络的特性分析^{*}

谢深泉

(湘潭大学计算机科学系 湘潭 411105)

摘要 优化 CAI(computer assisted instruction)课件设计中的超文本知识库,是提高课件质量的关键之一。首先给出知识点结构和知识点网络结构的数学描述;然后分析网络层次分划,给出先学备用和即学即用的并进网络遍历路线以及单进网络遍历路线;接着介绍知识点网络简化算法,以达到节省存储空间和便于操作的目的;最后综述利用网络特性优化 CAI 超文本知识库。

关键词 超文本, CAI(computer assisted instruction), 知识工程, 知识结构, 网络分层。

中图法分类号 TP18

知识库的组织和管理是 CAI(computer assisted instruction)课件设计的关键之一。良好的知识库应能反映知识的本来面目,并能便于操作。一本教科书的全部知识由若干个知识点组成,这些知识点分布在各个章节中。知识点间支持和被支持的联系是一种网络结构。因此采用网络形式的超文本结构来存储知识符合知识的本来面目。这样,CAI 课件设计采用超文本技术是很自然的事。为使这种超文本知识存储结构适应 CAI 课件设计要求并便于操作,必须从深入分析知识点网络的特性入手。本文在文献[1,2]的基础上进一步系统分析知识点网络的特性。文章首先给出知识点结构和知识点网络的数学描述,接着分析知识点网络的遍历路线,介绍知识点网络的简化算法,最后综述如何利用网络特性优化 CAI 超文本知识库。

1 知识点结构

知识点划分的基本原则是保证知识内容的局部完整性,而其大小可随需要而定,可能相差很悬殊。例如,一章可划为一个大的知识点,其中一节的内容又可细划为较小的知识点,一节中的定义、定理等还可以划为更小的知识点。

定义 1.1. 具有不能再分割的框架结构的知识点称为元知识点,而由相关的一组知识点组成的知识点称为复合知识点。仅由元知识点组成的知识点称为一级复合知识点;由不高于 k 级的复合知识点组成的知识点称为 $k+1$ 级复合知识点。元知识点为 0 级复合知识点。

定义 1.2. 元知识点一般为以下 9 个属性构成的框架。

主题 或称标题,是对本知识点的知识的一种高度概括的表述,用一个较短的字符串表示。例如,在微积分学中,常量、变量、函数、极限等等。

说明 或称注解,包括以下几方面的内容:提示学习本知识点前必须掌握的预备知识点;叙述本知识点的主要内容和特点;说明本知识点在本知识库中的重要程度等。

规则集 知识点的核心部分,知识点的知识,无论是叙述性知识、过程性知识或控制性知识。例如,在数学中的定义、定理、引理、推论,在物理学中的概念、定律、操作,都以规则形式表示出来。注意,这里的所谓规则形式是一种广义的可操作的形式,依据具体情况,规则可采用逻辑、产生式、语义网络、框架、概念依赖、对象等知识表达形式表示。

规则例示集 用例示说明本知识点的规则如何被引进,如何应用来解决实际问题。例示至少有以下几类:规则集中的规则的引进和证明例示;规则集中的规则的应用范围和应用方法例示;规则演变(操作)过程例示或模拟演示。

规则测试集 规则集中的规则应用实例集,该集合的大小依据本知识点内容及其重要程度来确定,用于测试对本知识点(及以前学习过的知识点)的知识理解与掌握的程度。要求每一个测试题都要集中到一个或几个测试目标(知识点或规则)。

测试评分标准 对测试集中的各测试题给出一个标准答案、评分标准和权值。评分标准是,答案全对给满分

* 本文研究得到湖南省教委科研项目基金资助。作者谢深泉,1939年生,教授,主要研究领域为人工智能,神经网络。

本文通讯联系人:谢深泉,湘潭 411105,湖南省湘潭大学计算机科学系

本文 1997-03-14 收到原稿,1997-09-08 收到修改稿

(1分),答案部分对或全错,给部分分(p 分, $0 \leq p_1 < 1$),并给出未达到的测试目标.每道题的权值依据本题的测试目标个数及其重要性来定.在评价对知识点掌握的情况时,可按某一原则抽取测试集的一子集来进行测试.每次测试满分为100,而实际得分为:对各题得分加权求和,除以该子集总权值,并乘以100.

关键字集 与本知识点相关的关键字集合.关键字分三类两级.三类:本知识点、预备知识点和后继知识点.两级:知识点的级(一级)和规则级(二级).知识点的级关键字用知识点主题表示,规则级用规则编号表示.关键字的格式为(知识点的级), (规则级),省去前缀则对应本知识点规则,省去后缀则对应一知识点.

进入阈值 用于学习进程控制,只有当学习本知识点所要求的预备知识点的加权平均得分不小于进入阈值时才允许访问(学习)本知识点.

通过阈值 用于学习进程控制,只有当测试得分不小于这个阈值时才允许离开本知识点.

这些属性的值是多媒体信息,即数字、文字、图形、图象、声音、程序或它们的组合.

定义 1.3. 复合知识点是由以下6个属性构成的框架.

除主题、说明、关键字集、进入阈值、通过阈值5个属性(含义与定义1.2相同)外,还有组成属性,即指向低级知识点指针的集合.

2 知识点网络的数学描述

2.1 知识点网络的矩阵表示

定义 2.1. 若干相关的知识点按其内在联系构成的网络称为知识点网络.网络的节点表示知识点,节点间的链接表示知识点间的联系.

定义 2.2. 表示元知识点的节点称为元节点,表示 k 级复合知识点的节点称为 k 级复合节点.仅由元节点构成的网络称为元知识点网络.一个 $k(k > 0)$ 级复合节点可以扩展为知识点网络,扩展所得的网络称为 k 级子网.由不高于 k 级的复合子网构成的知识点网络称为 $k+1$ 级复合知识点网络.

定义 2.3. 节点间的链接是按支持者到被支持者的方向链接的.一个节点的入向链接是指端节点是本节点的链接,而出向链接是指始节点是本节点的链接.每个链接都有一个非负权值,它反映始节点对端节点支持的重要程度.只有出向链接的节点称为起始节点,只有入向链接的节点称为终结节点.从节点 i 到节点 j 的有向路径称为节点 i 对节点 j 的支持路径.

可见,知识点网络是一个带权的有向图 G .为便于描述,对节点给以自然数编号.在未作顺序处理前,编号是随机的.

通常两个知识点都不直接或间接相互支持.这样,知识点网络中任何两个节点都不会在一条有向环路上.因此,我们只讨论此类网络.

定义 2.4. 若有向网络图中不存在有向环路,则称该网络为前向网络.

有向图的邻接矩阵可描述节点间的联系.设 A 是有向图 G 的邻接矩阵,第 i 行第 j 列元素为1表示有一条从节点 i 指向节点 j 的有向边(链接),否则,表示无边(链接).

定义 2.5. 矩阵 $A_k(A^{(k)})$ 称为有向图 G 的 k 步(级)邻接矩阵,第 i 行第 j 列元素为 m ,当且仅当有 m 条长度为(长度不超过) k 的节点 i 对节点 j 的支持路径.

图 G 的闭包 G^* 的邻接矩阵是矩阵 A 的闭包矩阵 A^* ,它表示图 G 中节点间的所有支持.其第 i 行第 j 列的元素为 m 时,则表示有 m 条各种长度的节点 i 对节点 j 的支持路径,即节点 i 对节点 j 支持了 m 次,反映了节点 i 对节点 j 的支持程度, $m=0$ 则表示它们间无任何支持关系.

2.2 受支持广度的表示

受支持广度是指一个节点受多少节点支持,描述了学习该知识点要具备哪些预备知识.

定义 2.6. 以节点 j 为终止节点的所有长度不超过 k (长度非0)的支持路径构成的子图,称为对节点 j 的 k 级(最大)支持子图.节点 j 的 k 级(最大)支持子图的所有节点(包括节点 j 本身)的集合 $S_k(S)$,称为对节点 j 的 k 级(最大)支持节点集.

对节点 j 的 k 级(最大)支持节点集 S_k 的算法, k 级支持子图的邻接矩阵的算法、对节点 j 的最大支持节点集 S 的算法和支持子图的邻接矩阵的算法均已在文献[1]中给出.

2.3 支持深度的表示

支持深度刻画节点对被支持节点的基础深度,用其间的最长支持路径长度来描述.

定义 2.7. 称节点 i 对节点 j 的最长支持路径的长度为节点 i 对节点 j 的支持深度. 矩阵 D 称为有向图 G 的支持深度矩阵, 其第 i 行第 j 列的元素的值为节点 i 对节点 j 的支持深度.

利用 k 步邻接矩阵, 文献[1]给出了支持深度和支持矩阵的算法.

2.4 支持程度的表示

支持程度刻画节点对被支持节点的重要程度, 用链接权来描述.

定义 2.8. 矩阵 W 称为有向图 G 的权矩阵, 其第 i 行第 j 列的元素的值为从节点 i 到节点 j 的链接的权值.

定义 2.9. 矩阵 $W_k(W^{(k)})$ 称为有向图 G 的 k 步(级)权矩阵, 其第 i 行第 j 列的元素值为所有长度为(长度不超过) k 的节点 i 对节点 j 的支持路径上的链接权值之和. 它反映节点 i 通过所有长度为(长度不超过) k 的路径对节点 j 的支持程度.

定义 2.10. 矩阵 W^* 称为有向图 G 的全权矩阵, 其第 i 行第 j 列的元素值为节点 i 对节点 j 的所有支持路径上的链接权值之和. 它反映节点 i 通过所有支持路径对节点 j 的支持程度.

文献[1]给出了 k 步权矩阵的算法.

3 知识点网络的遍历路线

学习就是以较好的成绩遍历知识点网络的各个节点. 这里存在两个问题, ①如何保证学习质量; ②如何遍历网络各节点. 对于第 1 个问题, 简单的回答就是: 学好了再用. 具体作法是, 用知识点结构中的进入阈值和通过阈值属性来控制进入要学习的后继知识点和离开正在学习的本知识点; 对于第 2 个问题, 则必须先学后用, 即要遵循学习新知识必须在掌握好预备知识后才能进行的规则. 下面来讨论如何求出先学后用的遍历路线.

先学后用给出学习进程的基本路线, 而学好了再用则允许在学习进程的基本路线上回溯, 二者结合为智能控制学习进程提供了基础.

3.1 前向网络的性质

学习过程, 即先学后用的网络遍历, 总是从起始节点集节点开始, 终结节点集节点结束. 在有限前向网络上, 这总是可行的, 因为它具有以下性质(见文献[1]).

(1) 有限前向网络一定存在起始节点集和终结节点集.

(2) 对任何有限前向网络, 总存在一种节点编号, 使得其邻接矩阵和相应的权矩阵均为具有零对角块的上三角分块矩阵, 且其非零元素分布相同.

(3) 上述分块邻接矩阵的第 1 个对角块的行号正好是网络图的起始节点号, 最后一个对角块的行号正好是网络图的终结节点号.

3.2 网络层次的划分

定义 3.1. 网络的起始节点集称为第 1 层节点. 从网络中删去第 1 到 $k-1$ 层节点及其所有的出向链接后, 剩下的网络中的无入向链接的节点集称为第 k 层节点. 对网络节点的这种分层称为网络的层次划分.

网络层次划分的原则实际上就是第 k 层的节点只受到前 $k-1$ 各层的节点的支持. 网络层次划分就是知识点间联系层次的划分, 即学习第 k 层知识点的预备知识全含在前 $k-1$ 各层中. 因此, 网络层次划分可以用来指导学习的进程, 它对于 CAI 课件设计是有重要意义的.

定义 3.2. 对任意 $k(k>1)$, 第 $k-1$ 层的节点中有的可能不直接支持第 k 层的节点, 而直接支持第 $k+1$ 层和更高层次的节点, 这种层次划分称为先学备用层次划分.

有限前向网络性质 2 给出了一种网络层次划分, 即分块邻接矩阵的第 k 个零对角块的行号正好是网络图的第 k 层节点号. 这种划分是一种先学备用层次划分.

定义 3.3. 对任意 $k(k>1)$, 第 $k-1$ 层的任意节点至少直接支持第 k 层中的 1 个节点的层次划分称为即学即用层次划分.

求即学即用层次划分先要删除冗余边(见 4.1 节). 文献[2]给出了即学即用层次划分及相应邻接矩阵的算法.

3.3 学习进程控制网络遍历路线

学习进程控制有两种方式, 并进遍历和单进遍历.

3.3.1 并进遍历

这是指当学习有多个目标要深入时, 不是先深入某个目标, 而是兼顾各目标, 最后几乎同时达到各目标. 执行时又有两种方式:

(1) 先学备用并进遍历

它是按先学备用层次分划得到的节点编号顺序来作为学习进程的基本路线,当前学习的知识点并不一定在学习下一层知识点时立即用到,这种学习方法是可行的,但不一定最好。

(2) 即学即用并进遍历

它是按即学即用层次分划得到的节点编号顺序来作为学习进程的基本路线,当前学习的知识点一定在学习下一层知识点时用到,这种学习方法也可行,是比较好的学习方法。

3.3.2 单进遍历

这是指虽然学习有多个目标,但选定某个目标深入学习,学完后再选另一目标学习。

把对学习目标节点的最大支持子图的邻接矩阵,应用即学即用层次分划算法所得的节点编号顺序作为学习进程的基本路线,是单一目标的即学即用式学习。

4 知识点网络的简化算法

4.1 冗余边的删除

定义 4.1. 支持路径长度为 1 称为直接支持,支持路径长度大于 1 称为间接支持. 两节点间仅有一条长度为 1 的支持路径称为仅直接支持。

定义 4.2. 若一个节点 i 对另一个节点 j 的支持路径不只一条,并且有一条长度为 1 的路径,即不仅间接支持,而且直接支持,这种直接支持的边称为冗余边。

由于存在冗余边,使得在学习中不能通过正在学习的知识点时,可能发生对间接支持的知识点直接回访,而在大多数情况下,造成对本知识点不能很好掌握的原因是未学习好仅直接支持的知识点,而不是那些既间接支持又直接支持的知识点,因此,删除冗余边是合理的。

在删除了冗余边后,前向网络图中任何两个节点之间若存在一条长度大于 1 的路径,即间接支持,则两个节点之间不会再有边,即不会再有直接支持;反之,若两个节点之间有边,即直接支持,则不会有长度大于 1 的支持路径,即间接支持。或者说,此时的直接支持总是仅直接支持。

在即学即用分划的基础上容易删除冗余边,文献[2]给出了两个算法。

4.2 网络子图的分划

将高级的知识点网络图转换为低级的知识点网络图,只需将复合节点扩展为其相应的子图。反过来,将低级的知识点网络图转换为高级的知识点网络图却不很容易。后者对节省存储空间和提高操作效率有重要的作用。

注意,在子图分划中,如果所得的子图有 1 个以上的终结节点,则应引进一个虚节点,使所有终结节点都指向它,而它继承所有终结节点的出向链接,于是它就成了唯一的终结节点,这样所得的子图称为扩展子图,可以在网络图中作为一个复合知识点节点出现。

网络子图的分划算法参见文献[2]。

4.3 可替换分层网络的生成

经过即学即用的层次分划算法处理后,一个节点除了直接支持下层的节点外,还可能直接支持更下层的节点,由于上述情况的出现,使得邻接矩阵的第 1 上对角线块以上还可能有非零元素,因此,存储邻接矩阵要占用多的空间,如果我们保持层次不变、支持关系不变,用另一个网络来替换,又不会对 CAI 课件设计带来影响,这样就可以节省存储空间。

定义 4.3. 对任意 $k(k > 1)$,第 $k-1$ 层的任意节点只直接支持第 k 层中节点的网络称为分层前向网络。

根据定义容易证明下述定理。

定理 4.1. 一个有限前向网络是分层的,当且仅当网络中任何支持路径上的相邻两节点的层次差为 1。

下面是上述定理的另一种表述。

定理 4.2. 一个有限前向网络是分层的,当且仅当网络中任何两节间的不同支持路径的长度是相等的。

由即学即用层次分划算法得出以下定理。

定理 4.3. 对任何分层有限前向网络,总存在一种节点编号,使得其邻接矩阵是仅具有非零的第 1 上对角线分块矩阵。

对于一般的有限前向网络,上述定理不成立,因为当一个节点 i 到另一个节点 j 时,可能有两条长度不同的支持路径,但是,只要对网络中的支持关系作适当处理,可以将任何有限前向网络转换为分层网络。

定义 4.4. 如果将一个前向网络通过修改节点间的链接产生的另一前向网络, 虽然在后者中可能增加了一些新的支持关系, 但仍保持了原网络中所有的支持关系(不区分直接与间接支持), 称后者为原网络的一个可替换网络。

上面定义中的可替换的含义是指当用知识网络来组织教学时是可允许的, 因为它仍保持了原有的支持关系, 而增加的新关系对教学的进行无实质影响。

算法 4.1. 转换一般有向网络为可替换的分层有向网络的算法。

① 调用即学即用层次分划算法将网络节点重新编号;

② 依次对第 $k(k=1, 2, \dots)$ 层进行操作, 若该层节点 i 直接支持第 $m+1(m>k)$ 层的某节点 j , 则在第 m 层中找一节点 p , 它受节点 i 支持, 删除从节点 i 到节点 j 的链接, 添加从节点 p 到节点 j 的链接。

下面的算法 4.1A 是算法 4.1 的另一种表述。

算法 4.1A. 转换一般有向网络为可替换的分层有向网络的邻接矩阵的算法。

① 调用即学即用层次分划算法求出新的编号下的邻接矩阵;

② 依次对第 $k(k=1, 2, \dots)$ 个第 1 上对角线块进行操作, 若该块所在的行 i 中不属于该块的元素 a_{ij} 非零, 则在第 j 列所处的第 m 个第 1 上对角线块中找一行号 p , 使得闭包矩阵元素 a_{ip} 非零, 并置 a_{ij} 为 0, a_{pj} 为 1。

定理 4.4. 算法 4.1(或 4.1A)将一般有向网络转换所得的网络是可替换的分层有向网络。

由定理 4.3 可知, 可替换的分层有向网络的邻接矩阵是一个仅具有非零的第 1 上对角线分块矩阵。

5 超文本知识库

定义 5.1. 以知识点为节点的超文本结构称为超文本知识库。

通过知识点结构中的后继知识点类关键字集, 就可以生成知识网络的邻接矩阵, 即建立超文本知识库的初始结构。然后利用本文讨论的网络特性, 对初始结构进行优化预处理。预处理的顺序是: (1) 求先学备用邻接矩阵, (2) 删除冗余边, (3) 求即学即用邻接矩阵, (4) 求可替换分层网络的邻接矩阵。经过预处理后的邻接矩阵是仅具有非零的第 1 上对角线分块邻接矩阵。如果是大型知识点网络, 则通过子图的分划建立多级网络, 即多级知识库。

经过预处理的超文本知识库, 具有两个特点: (1) 如果用邻接矩阵来存储节点间的联系, 则存储空间可以大大节省; (2) 适合于即学即用并发的学习进程控制, 如果要指定目标单进, 则在此基础上可再通过求支持子图来实现。

参考文献

- 1 谢深泉. 超文本知识库及其在 CAI 中的应用. 湘潭大学自然科学学报, 1995, 17(3): 101~108, 120
(Xie Shen-quan. Hypertext knowledge base and its application in CAI courseware. Natural Science Journal of Xiangtan University, 1995, 17(3), 101~108, 120)
- 2 谢深泉. 应用超文本知识库于 CAI 课件中的若干问题(I). 湘潭大学自然科学学报, 1996, 18(1): 124~129
(Xie Shen-quan. Some problems in design of CAI courseware applying hypertext knowledge base (I). Natural Science Journal of Xiangtan University, 1996, 18(1): 124~129)

Analysis of the Properties of Knowledge Points and Their Networks

XIE Shen-quan

(Department of Computer Science Xiangtan University Xiangtan 411105)

Abstract Optimizing the CAI (computer assisted instruction) hypertext knowledge base is one of the keys for improving quality of CAI courseware. First, the structure of knowledge points and the mathematical description of their networks are given in the paper. Then the parallel network traveling in "learn first use later" or "learn right use right" manner and the single destination network traveling are given by some layer partitions of network nodes. Some algorithms of reducing the network structure to save storage space and ease operating are introduced. Finally, optimizing the CAI hypertext knowledge base by means of the network properties is summarized.

Key words Hypertext, CAI (computer assisted instruction), knowledge engineering, knowledge structures, network management layer.