

一种网络导航学习路径生成方法

李孟, 曹晟, 秦志光, 邱丹

(电子科技大学计算机科学与工程学院 成都 611731)

【摘要】该文提出了一种基于反馈的、满足全局优化和充分必要条件的网络导航学习路径生成方法。将知识单元及其依赖关系作为知识地图的顶点和边,全部存储于广义表中。测试学习者已经学过的知识单元,根据知识单元中心度和难度量化计算方法,生成知识地图总路径集合,基于不同的学习基础,学习者在集合中“按需所取”知识单元。该方法产生的知识地图导航不再是海量知识单元的不完备近似集合,而具有全局精确性;解决了学习路径的充分性和必要性难以同时满足的缺陷;针对学习者个性化要求,有效提高学习效率,具有教育学和计算机科学的双重价值。

关键词 反馈机制; 知识地图; 学习路径; 网络导航学习

中图分类号 TP391.6

文献标志码 A

doi:10.3969/j.issn.1001-0548.2014.04.022

Network Navigation Learning Method Based on Feedback Mechanism

LI Meng, CAO Sheng, QIN Zhi-guang, and QIU Dan

(School of Computer Science and Engineering, University of Electronic Science and Technology of China Chengdu 611731)

Abstract A novel method on network learning is proposed. It can optimize the knowledge map in the global conditions and the necessary and sufficient requirements. All the knowledge units and their relations, stored in a generalized list, are viewed as the vertices and edges in the knowledge map. After testing the knowledge units which have been learned by users before, we can get the knowledge levels corresponding with each user. The sort sets of knowledge units can be deduced according to the algorithms of knowledge unit central and degree of difficulty, and then the complete sets of knowledge map paths are achieved. The approximation is substituted by the precision in the proposed method, and the defect that the necessary and sufficient conditions cannot be satisfied simultaneously is solved as well. The network navigation paths in our knowledge map can be used to avoid the problems of disorientation and information overloading, which can improve the efficiency of user learning on the network.

Key words feedback mechanism; knowledge map; learning path; navigation network learning

信息化时代,网络互联打破了传统教育模式。教学空间不再局限于课堂化和程式化,教学资源也不再受限于纸质化和离散化等因素,学习者通过网络在线学习已成为普遍现象,并在加速发展。网络培训、远程教育、移动学习、微型学习等模式各有特色,其共同的特点在于:在知识与信息日新月异、爆炸式增长的环境下,不论是以文档或是以多媒体的方式来组织,这些海量资源之间的耦合度越来越低、关联度越来越复杂,获得符合个性化要求资源的难度越来越大,学习者需要通过自身的知识积累和个人能力从搜索引擎给出的大量链接或数据库中找出需要的资源,严重降低了学习者的学习兴趣和效率,导致了“认知过载”和“学习迷航”等问题,

成为网络学习面临的重要挑战。

解决上述问题的关键在于如何自动生成高效的个性化学习路径,属于排课类NP问题。学习者通过网络导航学习路径可以获知3个要素:明确目标资源与其他关联资源的逻辑关系(关联状况);了解下一步要学习的资源,避免学习无用资源以防止认知过载;知道到达下一个资源位置的路径,避免迷航导致的时间和心力开销等。

1 相关工作

目前,导航学习路径的生成主要有3种方式:

1) 人工为主。通过教师等权威人士根据自身对知识结构的理解与认识,设定导航学习路径,这种

收稿日期: 2013-11-18, 修回日期: 2014-05-22

基金项目: 国家自然科学基金(61133016, 61103205, 61103206); 国家863项目(2011AA010706); 教育部中国移动科研基金(MCM20121041)

作者简介: 李孟(1981-), 女, 博士生, 主要从事计算机网络和知识工程方面的研究。

导航方式准确率较稳定。但学习路径千人一面,且路径一旦生成,不易更改。更重要的是,人工生成学习路径的速度远远落后于教育资源的海量增长,不能满足大规模、动态的应用要求。

2) 机器为主。通过各种启发式机器学习方法自动生成学习路径^[1-4],随着教育资源的增长而同步匹配。该方法忽略了知识单元之间的关联,产生的知识地图往往缺乏甚至违背知识之间的逻辑关系,反而导致更严重的学习迷航和认知过载。

文献[1]将知识地图的生成看作是动态约束优化问题,通过设计合理的优化算法自动生成学习路径。该方法得到的知识地图是局部解而不是全局解,导致知识单元之间的关联关系不完善,有可能导致学习者遗漏重要的知识单元,与认知过载相对应,属于认知不完整。

文献[2]借用进化计算的思想生成知识地图,采用模量因子优化算法,可以得到较为理想的典型学习路径。由于模量因子算法的效率依赖于因子的选择,而不同学习者的学习因子难以预先知道并区分,若模量因子选择不当,将直接影响学习路径的准确性,存在认知过载的隐患。

文献[3]引入近似计算思想构建知识地图,通过测试等预处理发现学习者的知识不足之处,建立学习者适应函数,继而生成学习路径。这种方法严重依赖于对学习者的知识结构的测试,若测试给出的结果与学习者真实的知识结构偏差较大,知识地图中的路径也不能给出有效的指引。

文献[4]采用贝叶斯网络的概率特性描述学习能力与学习目标之间的关系,以大多数认可的学习过程生成最佳的导航学习路径,在学习效率和学习群之间形成了良性的平衡,代价是丢弃了学习群体中少数“与众不同”的学习者。

3) 人工与机器两者结合,即基于知识地图的导航学习路径生成方法^[5-6]。该方法通过对知识之间依赖关系的划分,保证了方法2)的逻辑合理性,从而成为近年来的研究热点。

文献[5]将知识地图的生成与排序约束问题联系起来,通过描述本体中的概念及概念之间的关系,辅以专家知识,对知识单元进行排序并给出学习路径,但不构成掌握某个知识点所需知识单元之间依赖关系的充要条件,使得学习者学习的知识单元不全面。

文献[6]也将知识地图的路径生成看成排序约束问题,给出了一种满足全局优化的学习路径生成方

法,可以满足学习者个性化的学习要求。该方法未能重视学习者在学习过程中自我反馈的效用,随着学习能力的不断变化,对知识单元的掌握程度引起知识单元中心度及难度的变化,这种非收敛的自由选择方式会导致学习者选择知识单元的准确性逐步降低。

通过对以上各种方法的优缺点分析,本文采用方法3)并将知识地图的导航路径生成看成是排序约束问题,提出一种基于反馈机制的知识地图全局优化生成方法,且同时满足学习知识单元的必要条件(提供基本知识单元)和充分条件(避免遗漏知识单元),保证学习效率的可达性。

2 知识地图全体路径生成

知识地图(knowledge map, KM)是一个有向无环图 (V, E) ,顶点 $(V_1 \cup V_2) \in V$ 表示某教育学科领域内全部知识单元的集合,即学科领域中具有完整表述能力的基本知识单位,比如定义、定理、算法等。边 E 是知识单元之间的依赖关系集合,可定义如下等价关系:

$$\begin{aligned} \exists(R_i, R_j) \in E &\Leftrightarrow \\ (\exists R_i \in V)(\exists R_j \in V)(S(R_j) = 1) &\Leftrightarrow \\ \text{Level}(R_i) = 1 & \end{aligned}$$

在知识地图中, R_i 是 R_j 的前驱节点, R_j 是 R_i 的后继节点,并定义 C 是前驱节点集合, F 是后继节点集合。 $\text{Level}(R_i)$ 是 V 上知识单元已被学习理解的状态函数,如果学习理解了 R_i ,则 $\text{Level}(R_i) = 1$;反之 $\text{Level}(R_i) = 0$ 。 $S(R_j)$ 是 V 上知识单元可以被理解的状态函数,如果可以学习理解 R_j ,则 $S(R_j) = 1$;反之 $S(R_j) = 0$ 。

将 (V, E) 存储于广义表中,成为学习者提取所需知识单元的总表,保证学习的充分性要求。再依据学习者的学习目标,给出满足学习必要性的知识单元,即可实现所有学习者的全体学习路径。

TableGen是KM的结构存储表,知识单元 R_i 的存储结构要素包括 Num_{id_i} 、 $\text{Level}_{\text{id}_i}$ 、 d_{id_i} 、 Pre_{id_i} 和 Suf_{id_i} 。其中, ID_i 是 R_i 的标示符, Num_{id_i} 是 R_i 在全体路径中的序号; $\text{Level}_{\text{id}_i}$ 是 R_i 的知识层次,表示在KM上学习 R_i 的难度级别; d_{id_i} 是 R_i 的学习中心度,表示掌握 R_i 对学习KM上其他 R_j 的重要程度; Pre_{id_i} 是以 R_i 的前驱结点为元素的广义表; Suf_{id_i} 是以 R_i 的后继结点为元素的广义表。

知识地图中全体路径totalpath是由KM中所有

知识单元及其依赖关系构成的一个序列:

$\text{totalpath}(1), \text{totalpath}(2), \dots, \text{totalpath}(|V|)$

一般地, 当序列满足以下条件:

1) $\forall R_i \in V \Rightarrow \exists i(1 \leq i \leq |V|) \text{ totalpath}(i) = R_i$;

2) $(|C_{\text{totalpath}(1)}| = 0) \text{ and } (|F_{\text{totalpath}(|V|)}| = 0)$;

3) $\forall i(1 \leq i \leq |V|) L(\text{totalpath}(i)) = 1$

$\Rightarrow S(\text{totalpath}(i+1)) = 1$;

4) $L(\text{totalpath}(|V|)) = 1 \Rightarrow S(KM) = 1$;

5) $\forall i \forall k(1 \leq i < k \leq |V|) \Rightarrow \text{Level}_{\text{ID}_i} = \text{Level}_{\text{ID}_k}$

$\Rightarrow d_{\text{ID}_i} \leq d_{\text{ID}_k}$ 。

基于以上约束条件, totalpath 分别按照 $\text{Level}_{\text{ID}_i}$ 的升序和 d_{ID_i} 的降序对 V 中所有 R_i 进行排序, 所得路径集合即为知识地图路径, 该用户导航学习路径生成算法如图1所示。

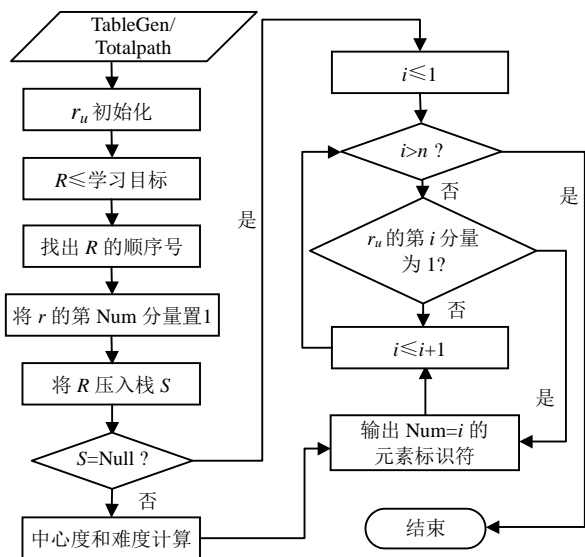


图1 用户导航学习路径生成算法

3 知识单元中心度量及难度量化

3.1 知识单元中心度量

学习中心度 d_{ID_i} 是对知识单元 R_i 在知识地图上重要性的度量。在所有的知识单元集合里, 有些知识单元与其他知识单元之间的关联较大, 被认为是重要的知识单元。知识单元在集合中是否处于中心地位, 可以通过启发式方法进行自动认定, 这也是文献[5]的思路。根据教育学规律, 知识单元是否重要, 除了启发式方法外, 学习者随着知识能力的提高, 还需要通过测试学习者对某个知识单元的掌握程度来度量^[7]。因此, 明确知识单元与学习目标、试题之间的关系, 对知识单元中心度的认定也具有决定性意义。实际上也是满足学习者真正达到个性化学习要求的标志^[8]。

若某个知识单元与其他知识单元之间存在关联, 知识地图上则一定有边相连, 用知识地图上经过 R_i 总路径条数的期望值来表示。由于地图在逻辑结构上是一个有向无环图, 可将基于知识地图上的学习过程看作吸收马尔科夫(Markov), 状态迁移概率矩阵表示学习掌握知识单元 R_i 后, 进而掌握 R_j 的可能性:

$$P = \begin{bmatrix} RB \\ \mathbf{0} \mathbf{I} \end{bmatrix}$$

其中, R 是 $|V_1| \times |V_1|$ 阶矩阵; B 是 $|V_1| \times |V_2|$ 阶矩阵; I 是 $|V_2| \times |V_2|$ 阶单位阵; $\mathbf{0}$ 是 $|V_2| \times |V_1|$ 的零阶矩阵。依据Markov链基本特性可知, 知识单元 R_i 的中心度为:

$$d_{\text{ID}_i} = \begin{cases} \sum_{k=1}^{|V_1|} q_k & i \leq |V_1| \\ \sum_{k=1}^{|V_1|} \sum_{l=1}^{|V_1|} b_{li} & |V_1| < i \leq |V| \end{cases} \quad (1)$$

式中, q_k 是矩阵 $Q = (I - R)^{-1}$ 的第 k 行第 i 列元素; b_{li} 是矩阵 B 的第 l 行第 i 列元素。通过式(1)得到各知识单元中心度的值, 为各个知识单元在知识地图中的定位给出程式化途径。

在自动计算的基础上, 因为学习者基础互不相同, 导致不同的知识单元 R_i 对于不同的学习者来说, 其难度各不相同, 所以还需要结合学习者的实际情况进行个性化配置。式(1)表示对于任一个学习者, 其学习难度均一致。因此, 需要在上述公式中引入变量来体现这种差别, 称为“反馈因子”, 记为 θ_i , 表示学习者答错某试题与知识单元 R_i 的关系, 即学习者未能掌握某知识点, 知识单元 R_i 所占因素百分比:

$$d_{\text{ID}_i} = \begin{cases} \sum_{k=1}^{|V_1|} q_k \theta_i & i \leq |V_1| \\ \sum_{k=1}^{|V_1|} \sum_{l=1}^{|V_1|} b_{li} \theta_i & |V_1| < i \leq |V| \end{cases} \quad (2)$$

经过校正后, 可以进一步提高知识单元中心度的准确性。获得 θ_i 的方式很多, 依据不同教育学科领域特点和不同知识单元学习目标要求, 可以采取不同的测验类型, 诸如考试、问答、测验和调查等。

《数据结构与算法》吴跃等编写教材中, 以“树”这一章及课后习题为资源, 经过Matlab计算, 得到相应知识单元中心度, 并与权威专家(任课教师)所理解的知识单元中心度相对比。本文采用测试的方法获得反馈因子, 并假定权威专家对于经典知识的理

解和认识基本趋于一致, 且专家给出的中心度量值已经过等比例转换, 如表1所示。

表1 知识单元中心度量对比

知识单元	式(1)	式(2)	权威专家
二叉树	0.92	0.66	0.70
平衡二叉树	0.34	0.42	0.50
霍夫曼树	0.66	0.73	0.70
四叉树	0.11	0.19	0.20
B树	0.31	0.57	0.50

从表中可以看出, 式(1)的知识单元中心度与专家给出的值偏差较大, 且计算出的多个值之间的变化幅度较大; 式(2)给出的值更接近专家认识, 且变化幅度平滑, 在教育资源达到海量规模时, 具有更细的粒度, 使得知识单元中心度的划分更加有层次。

知识单元难度量化的准确度还与设定的试题难度有正相关关系, 当资源规模较大时, 参照《教育信息化技术标准CELTS41.1》中试题分类标准^[10], 对试题资源库的规范性的要求变高。

3.2 知识单元难度量化

知识单元难度度量参数 $Level_{iD_i}$ 是一个相对量, 用于定量描述知识单元的学习难度。现有知识单元难度量化均以知识单元难度相加总和来评判^[6], 只能近似估计知识单元理论上的难度, 不能真实反映知识单元的实际应用难度。

对 n 个学习者在学习某个知识单元 R_i 后, 依据做题的反馈结果对 R_i 进行校正, 才是实际的难度系数。定义 ε 为知识单元所对应的题目 Q_i 的实际得分, x 为学习者学习某知识单元所对应的题目 Q_i 的得分, 难度系数记为 $l_{iD_i}(x)$ 。

概率 $p(\varepsilon \geq x)$ 越大, 表示得分能达到 x 的学习者人数越多, 则该题难度较小; 反之成立。定义 $F(x) = p(\varepsilon < x)$, g 的分布函数 $l_{iD_i}(x) = g(F(x))$ 。

$\forall x \in R_i$, 当 $\Delta x > 0$ 时, 定义:

$$p(\varepsilon \geq x) - p(\varepsilon \geq x + \Delta x) = p(x \leq \varepsilon < x + \Delta x) \quad (3)$$

式中, Δx 表示难度值在 x 处的变化程度。

若 Δx 不变,

$$P(x_1 \leq \varepsilon < x_1 + \Delta x) > P(x_2 \leq$$

$$\varepsilon < x_2 + \Delta x) \Leftrightarrow \Delta l_{iD_i}(x_1) < \Delta l_{iD_i}(x_2)$$

$$\Delta F(x_1) > \Delta F(x_2) \Leftrightarrow \Delta l_{iD_i}(x_1) < \Delta l_{iD_i}(x_2)$$

表明 $l_{iD_i}(x) = g(F(x))$ 随 $F(x)$ 单调递增外, 其增速与 $F(x)$ 相反, 即难度系数 $l_{iD_i}(x)$ 具有如下特性:

1) $l_{iD_i}(x) = g(F(x))$ 是 $F(x)$ 的单调增函数, 也是 x

的单调增函数;

2) $l_{iD_i}(x)$ 在 x 点的增加速度的快慢与 $F(x)$ 相反,

即 $\forall x_1 \in R_i, x_2 \in R_i, \Delta x > 0$, 若 $\frac{\Delta F(x_1)}{\Delta x} > \frac{\Delta F(x_2)}{\Delta x}$, 则:

$$\frac{\Delta l_{iD_i}(x_1)}{\Delta x} < \frac{\Delta l_{iD_i}(x_2)}{\Delta x} \quad (4)$$

根据点积变换, 取:

$$l_{iD_i}(x) = g(F(x)) = cF^{-1}(x) + d \quad (5)$$

式中, $F^{-1}(x)$ 是 $F(x)$ 的反函数; c 、 d 表示待定参数且 $c > 0$ 。当 $x \in [0, 1]$ 时:

$$x = F(F^{-1}(x)) = p(\varepsilon < F^{-1}(x)) = \phi\left[\frac{F^{-1}(x) - \mu}{\sigma}\right] \quad (6)$$

式中, $\phi(x)$ 表示 $(0, 1)$ 随机变量的分布函数。对于

$$\phi^{-1}(x) = \phi^{-1}\left[\phi\left[\frac{F^{-1}(x) - \mu}{\sigma}\right]\right] = \frac{F^{-1}(x) - \mu}{\sigma}$$

$$\Leftrightarrow F^{-1}(x) = \sigma\phi^{-1}(x) + \mu$$

$\forall x \in R$, 由于学习者群体 $\varepsilon \sim N(\mu, \sigma^2)$ 满足正态分布, 所以:

$$p(\varepsilon - 3\sigma < 3\sigma) = p(\mu - 3\sigma < \varepsilon < \mu + 3\sigma) = 0.9974 \approx 1 \quad (7)$$

对 x 进行标准化, $x^* = \frac{x - \mu + 3\sigma}{6\sigma} = \frac{x - \mu}{6\sigma} + 0.5$, 因

$\sigma > 0$, 则标准化过程不会改变难度系数值, 即 $\forall x \in [\mu - 3\sigma, \mu + 3\sigma]$ 有 $F^{-1}(x^*) = \sigma\phi^{-1}(x^*) + \mu$ 。则有:

$$l_{iD_i}(x) = cF^{-1}(x^*) + d = c\sigma\phi^{-1}(x^*) + (c\mu + d) =$$

$$a\phi^{-1}(x^*) + b = a\phi^{-1}\left[\frac{x - \mu}{6\sigma} + 0.5\right] + b \quad (8)$$

式中, $a = c\sigma \in (0, 1)$; $b = c\mu + d \in (0, 1)$ 。 a 和 b 均为相对量, 在同一类知识单元集合中取值相同。

$\phi^{-1}(x^*)$ 值可通过计算 x^* 后查询方差表获得, 为资源处于海量状态时, 计算机自动、批量地计算不同类别知识单元的难度系数提供新的参考途径。

4 结束语

与现有技术相比, 本文提出的方法优点在于可生成知识地图上基于反馈的具有全局优化特点的总路径, 为知识单元之间提供满足学习充分性、必要性和优化的学习顺序参照表。知识单元中心度和难度量化在教育学和认知规律的指导下, 采用与学习者互动的反馈机制, 并给出了具体的计算公式。在此基础上构建的知识地图, 能够有效显式地展现出知识单元及其依赖关系, 使学习者不仅能轻松看出知识结构的变化, 也能选取最有效的学习路径及其到达方式, 为解决学习迷航和认知过载问题起到推

动作用。

参 考 文 献

- [1] ZHAO Cheng-ling, WAN Li-yong. A shortest learning path selection algorithm in e-learning[C]//The Sixth International Conference on Advanced Learning Technologies-ALT'06. Kerkraide: IEEE Press, 2006.
- [2] ACAMPORA G, GAETA M, LOIA V, et al. Optimizing learning path selection through memetic algorithms[C]// IEEE International Joint Conference on Neural Networks. Hongkong: IEEE Press, 2008.
- [3] CHEN Chih-ming, PENG Chi-jui, SHIUE Jer-yeu. Ontology-based concept map for planning a personalised learning path[J]. British Journal of Education Technology, 2009, 40(6): 1028-1058.
- [4] Jose Manuel Marquez Vazquez, Juan Antonio Ortega Ramirez, Luis Gonzalez-Abril, et al. Designing adaptive learning itineraries using features modelling and swarm intelligence[J]. Neural Computing&Applications, 2011, 20(5): 623-639.
- [5] COLACE F, DE Santo M. Ontology for e-learning: A bayesian approach[J]. IEEE Journal of Education, 2010, 53(52): 223-233.
- [6] 郑庆华, 黄香君, 刘均, 等. 一种知识地图上导航学习路径的生成方法: 中国, CN 102508874A[P]. 2012-06-20.
- ZHENG Qing-hua, HUANG Xiang-jun, LIU Jun, et al. A method for navigation learning path on knowledge map: China, CN 102508874A[P]. 2012-06-20.
- [7] 陶乃航, 赵星, 彭晓东, 等. 中心度指标在期刊引文网络分析中的运用及改进[J]. 图书情报工作, 2009, 3(4): 143-146.
- TAO Nai-hang, ZHAO Xing, PENG Xiao-dong, et al. Application and improvement of centrality indexes in journal citation networks analysis[J]. Library and Information Service, 2009, 3(4): 143-146.
- [8] 雷茵. 基于概念地图的网络化学习路径控制研究[D]. 重庆: 西南大学, 2006.
- LEI Han. A study of concept map based learning path control for web learning[D]. Chongqing: Southwest University, 2006.
- [9] 李瑾坤. 知识作业过程表示方法的研究[J]. 电子科技大学学报, 2008, 37(5): 769-772.
- LI Jin-kun. Knowledge work process representation method research[J]. Journal of the University of Electronic Science and Technology of China, 2008, 37(5): 769-772.
- [10] 教育部. 教育信息化技术标准CELTS41.1[M]. 北京: 人民教育出版社, 2002.
- Ministry of Education (The People's Republic of China). Education informatization technology standard CELTS41.1 [M]. Beijing: People's Education Press, 2002.

编辑 叶芳