

# 一种基于 Max-Flow 的协同学习活动调度算法

陈乙雄 吴中福 李 华 冯 永

(重庆大学计算机科学与工程学院 重庆 400044)

**摘 要** 远程教育中的在线学习过程并非简单的资源检索过程,而是由一系列学习活动组成的。现有的远程教育平台往往过于注重教学内容呈现的个性化,而忽视了对学习者活动的引导。从教育技术学的角度,不同学习目标和情景下的在线学习活动通常可以分为多种模式,针对其中的协同学习模式,提出了一种实现学习活动引导服务的方法。该方法将活动引导以调度生成的方式加以实现,提出了具体的调度模型和求解算法。该算法基于 Max-Flow 设计,具有较小的时间复杂度,能够有效地为学习活动引导服务提供支撑。

**关键词** 调度算法,远程教育,组合优化

**中图分类号** TP393 **文献标识码** A

## Collaborated E-learning Scheduling Algorithm Based on Max-Flow

CHEN Yi-xiong WU Zhong-fu LI Hua FENG Yong

(Department of Computer Science and Engineering, Chongqing University, Chongqing 400044, China)

**Abstract** To solve the problem of lacking activity guidance for learners in e-learning platform, a scheduling algorithm was proposed for the collaborated learning model which is classified by related e-learning projects. The basic idea of the proposed algorithm is to convert the problem into a Max-flow model, so that an effective learning schedule for learners can be generated in a reasonable time span.

**Keywords** Scheduling algorithm, E-learning, Combinatorial optimization

## 1 引言

实现个性化学习并不断提高学习效率和效果,是近年来远程教育研究的一个重要方向。在相关文献中,此类研究主要从学习内容组织方法、学习者个性特征建模、检索技术等方面展开。在学习内容组织方法方面,多数研究均以较为成熟的远程教育标准 SCORM 为基础,如文献[1]提出一种基于 SCORM 标准的个性化学习方法。在学习者个性特征建模方面,主要通过研究分析学习者对象在学习过程影响较大的显著性因素来完成学习者的特征建模,如一种集成学习者特征的远程教学环境<sup>[2]</sup>以及一种建立在学习者模型上的学习支持服务<sup>[3]</sup>。而相对于前两方面的研究,远程教育在检索技术方面的发展在相关 Web 技术的推动下取得了较大的发展。较为典型的研究包括通过本体<sup>[4-6]</sup>、语义<sup>[7]</sup>等技术来提高资源检索智能化程度的方法以及相应的一些学习效果评估方法<sup>[8]</sup>,以实现有效的个性化学习。

其中,笔者所在课题组多年来从事远程教育方面的研究,目前在数字化学习平台中,已先后研发了有关的个性化推荐技术<sup>[9,10]</sup>和资源组合技术<sup>[11,12]</sup>,这些技术能较好地帮助学习者在学习过程中获得符合其个性特征的教学资源。

通过对近期文献的综合分析可以发现,多数技术都是以教学内容为核心来实现的个性化学习服务。然而,从教育学

的角度来看:学习的过程并不等同于一个资源检索的过程,而应当是由按照先进教学思想,并结合学生个性特征设计的一系列教学活动组成的。为此,在分析网络学习环境中各类教学活动模式的基础上,本文对其中的协同学习活动模式进行了深入的研究,并利用调度理论建立了相应的模型,提出了一种基于最大流的调度算法。

## 2 问题模型

### 2.1 基本定义

根据调度理论的基本思想<sup>[13]</sup>,调度可以定义为:假设有  $n$  个作业  $J_i (i=1, \dots, n)$  需要在  $m$  台机器  $M_j (j=1, \dots, m)$  上处理,则调度就是给每个作业在一台或多台机器上分配一个或多个处理时间段;或者也可以等价地表述为给每台机器指定一个或多个作业的一个或多个处理时间段。

调度的结果一般由甘特图直观地表示,对于调度问题,则可以用一个三元组来刻画,即:

$$\alpha | \beta | \gamma (\alpha = \alpha_1 \alpha_2, \beta = \beta_1, \dots, \beta_k) \quad (1)$$

式中,  $\alpha$  表示处理作业的机器环境,  $\alpha_1 \alpha_2$  具体表示各个属性,如是否同构,处理速度是否相等等;  $\beta$  表示作业的性质,同样地,作业是否具有各自的释放时间、到期时间,作业之间是否有前驱后继关系等具体属性可以由  $\beta_1 \beta_2 \beta_3 \beta_4 \beta_5 \beta_6$  等表示;  $\gamma$  用于定义目标函数,调度常见的目标有最小时间跨度、最少延迟

到稿日期:2011-12-04 返修日期:2012-03-16 本文受国家自然科学基金(ACA07004-08),重庆市高等教育教学改革重点项目(112023)资助。

陈乙雄(1977-),男,博士,讲师,主要研究方向为调度理论、云计算、远程教育, E-mail: chenyx@cqu.edu.cn; 吴中福(1938-),男,教授,主要研究方向为远程教育、云计算。

作业数、最小代价等。

## 2.2 协同学习活动模型

### 2.2.1 处理机环境

在网络学习环境中,学习者可以通过参加不同的学习活动来实现其学习目的。其中,协同学习模式是指若干个学习者  $L_j (j=1, \dots, m)$  通过分工合作的方式去共同完成学习任务  $T_i (i=1, \dots, n)$ , 从而达到期望的学习效果。

为了让每个学习者发挥自己的特长,在这种学习模式下每个学习者所扮演的角色都是不同的。如果根据参与者角色的不同,将每个学习任务  $T_i$  都分解为  $O_{i1}, O_{i2}, \dots, O_{im}$  个操作,则表示  $O_{i1}$  只能由学习者  $L_1$  进行,  $O_{i2}$  只能由学习者  $L_2$  进行, ... 以此类推。

根据上述处理机环境的特性,结合文献[13]中对调度三元组各个变量取值的说明:对于各处理机为专用、作业由多个操作组成的情况,可以得出以下定义:

$$\alpha = \alpha_1 \alpha_2 = O(\alpha_1 = O, \alpha_2 = o) \quad (2)$$

### 2.2.2 活动的性质

这里的活动对应于调度基本模型的作业,下面分别对有关作业的6个属性值  $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$  进行考察(仅针对协同学习活动模式):

1)  $\beta_1$ : 作业是否允许中断

学习者在进行学习(或者说完成学习任务)过程中应该允许中断。即学习者可以暂时停止执行当前的学习活动,在之后的某个时刻继续该活动,故  $\beta_1 = \text{pmtn}$ 。

2)  $\beta_2$ : 作业的前后关系

在协同学习模式下,每个学习者都可以独立进行各自能够完成的作业,如果协同学习的作业集合元素与知识点之间具有关联关系,即渐进式学习,则  $\beta_2 = \text{outtree}$  或  $\text{chain}$ ; 如果没有上述关联关系,即自由式学习,则  $\beta_2$  可去掉。

3)  $\beta_3$ : 作业的释放时间

在网络学习环境中,如无特殊限制一般作业释放的时间即学习活动开始的时间,本文中可以定义作业的释放时间  $\beta_3 = r_i = 0 (i=1, \dots, n)$ 。

4)  $\beta_4$ : 描述作业中每个操作所需的处理时间

由于协同学习模式下作业的每个操作分别由不同的学习者来进行,因此其处理时间  $p_{ij}$  通常不同,但根据学习任务性质的不同,可以限制其处理时间的范围(可由用户自己定义)。这里定义每个操作的处理时间最短为  $t_u$  分钟,最长为  $t_v$  分钟,则  $\beta_4 = p_i \in \{t_u, t_{u+1}, \dots, t_v\}$ 。

5)  $\beta_5$ : 作业的到期时间

绝大多数学习型活动的设计目的都在于让学习者掌握理解知识,一般不会事先规定到期时间,因此本文的研究中对  $\beta_5$  不作限制。

6)  $\beta_6$ : 定义批处理作业

协同学习模式下不存在可以由单个学习者完成全部操作的作业,也不存在可以成批处理的作业,即  $\beta_6$  在  $\beta$  中也不出现。

### 2.2.3 活动模型

综上所述,并结合最短完成时间  $C_{\max}$  作为调度目标,我们可以得出最后的调度模型三元组:

$$\alpha | \beta | \gamma = O | \text{pmtn}; p_i \in \{t_u, t_{u+1}, \dots, t_v\} | C_{\max} \quad (3)$$

## 3 调度算法

为求解以上调度问题模型,将该问题转换为最大流问题来求解。其具体思路是:首先构造一个所有节点(除起始和终结节点)出入流量均相等的有向图;划分若干个时间槽,针对每个时间槽在图中的作业节点和学习者节点之间寻求完全匹配,即对作业和学习者之间进行了一次分配;将所有时间槽的作业与学习者匹配后即得到一个调度结果。

### 3.1 变量定义

首先,定义  $m$  个学习者  $L_1, L_2, \dots, L_m$ ,  $n$  个学习活动  $A_1, A_2, \dots, A_n$ 。每个学习活动  $A_i$  由  $m$  个操作  $O_{i1}, O_{i2}, \dots, O_{im}$  组成,每个操作  $O_{ij}$  分别由对应的  $L_j$  完成。

任一学习者  $L_j$  完成所有学习活动操作所需要的时间跨度(即该学习者不间断地进行操作)为:

$$TS_j = \sum_{i=1}^n p_{ij}, p_{ij} \in \{t_u, t_{u+1}, \dots, t_v\} \quad (4)$$

学习活动  $A_i$  能够被完成的最短时间跨度(即该学习活动不间断地在推进)为:

$$TA_i = \sum_{j=1}^m p_{ij}, p_{ij} \in \{t_u, t_{u+1}, \dots, t_v\} \quad (5)$$

### 3.2 构造流量图

分别设置起始节点  $S$  和终结节点  $T$ ,同时定义  $m$  个伪学习活动节点  $A_{n+1}, A_{n+2}, \dots, A_{n+m}$  和  $n$  个伪学习者节点  $L_{m+1}, L_{m+2}, \dots, L_{m+n}$ ,使得活动节点数和学习者节点数相等,均为  $m+n$ 。

构造有向图  $G=(V, E)$ ,其中顶点集为  $V=\{S, T, A_1, A_2, \dots, A_{n+m}, L_1, L_2, \dots, L_{m+n}\}$ 。从  $S$  到  $T$  的所有节点都添加有向边,形成完全图。最后再为各个边赋予流量值,具体方法如下:

1)从  $S$  到  $A_1, A_2, \dots, A_{n+m}$  节点和  $L_1, L_2, \dots, L_{m+n}$  节点再到  $T$  的所有边权值均为:

$$I = \max \left\{ \max_{j=1}^m TS_j, \max_{i=1}^n TA_i \right\} \quad (6)$$

2)从所有真实学习活动节点  $A_1, A_2, \dots, A_n$  到所有真实学习者节点  $L_1, L_2, \dots, L_m$  之间的边权值为活动  $A_i$  由学习者  $L_j$  完成操作  $O_{ij}$  的处理时间为  $p_{ij}$ 。

3)从所有真实学习活动节点  $A_1, A_2, \dots, A_n$  到所有伪学习者节点  $L_{m+1}, L_{m+2}, \dots, L_{m+n}$  之间的边权值根据流量平衡的原则来确定。由于进入任一活动节点  $A_i$  的流量为  $I$ ,从  $A_i$  流出到所有真实学习者节点  $L_1, L_2, \dots, L_m$  的流量之和为  $\sum_{j=1}^m p_{ij}$ ,即  $TA_i$ ,则可定义  $A_i$  到  $L_{m+i}$  的流量为  $I - TA_i$ ,从而使所有  $A_1, A_2, \dots, A_n$  两边的流量达到平衡。

4)同理,由于任一学习者节点  $L_j$  的流出量为  $I$ ,从所有真实活动节点流入到  $L_j$  的流量为  $TS_j$ ,因此可以定义伪活动节点  $A_{n+j}$  到  $L_j$  的流量为  $I - TS_j$ ,使其两端流量达到平衡。

5)最后,再根据流量平衡的原则在伪活动和伪学习者节点之间添加流量  $f_{n+j, m+i}$ ,同时去掉图中所有流量为0的边,则构造出图  $G$ ,如图1所示。

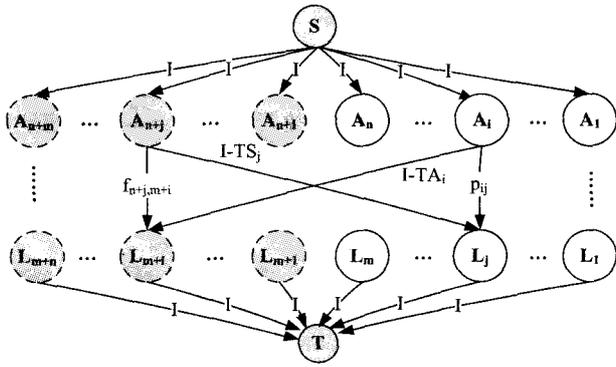


图1 活动与学习者节点最大流图

### 3.3 生成调度结果

在图  $G$  中,若从  $S$  到  $T$  发送流量  $I$ ,则可使图中的流量达到最大,处于饱和状态。从学习者节点集合中任取一个子集,并设  $L$  中节点  $L \subseteq \{L_1, L_2, \dots, L_{m+n}\}$  的所有前驱节点组成的集合为  $P(L)$ ,则根据图中  $L$  集合中节点的流量进出关系,有:

$$|L|I = \sum_{L_j \in L} \sum_{A_i \in P(L_j)} f_{ij} \leq |P(L)|I \quad (7)$$

即  $P(L)$  中的所有节点都贡献可能的最大流量  $I$  给  $L$  中节点时,流量为  $L$  中流量的上限。除去  $S$  和  $T$  外,所有学习者节点和活动节点及其它它们之间的边实际上构成了一个二部图。根据有关文献[14],关于二部图的最大匹配存在以下定理。

**定理 1** 设二部图  $G=(V_1 \cup V_2, E)$ , 其中,

$$|V_1| \geq |V_2| = m$$

式中,  $M$  为  $V_2$  的任一子集,则图  $G$  存在最大匹配  $m$  的充要条件为:

$$\forall M \subseteq V_2, \left| \bigcup_{v_i \in M} P(v_i) \right| \geq |M| \quad (8)$$

通过前面公式不难发现,对任一学习者节点集合中的子集  $L$ :

$$|L|I \leq |P(L)|I \Rightarrow |L| \leq |P(L)| \quad (9)$$

由于添加伪节点后学习者节点和活动节点数目相等,结合定理可以证明,在学习者节点和活动节点之间存在  $m+n$  的完全匹配,可用  $W$  表示。设  $T_k$  为当前活动节点到学习者节点的最小流量边的流量值,则最后生成调度方案的算法伪代码如下:

```

K=1;
T_0=0;
I=max{max_{j=1}^m TS_j, max_{i=1}^n TA_i};
For I>=0
{

```

```

  将图  $G$  中的最大流更新为  $I$ ;
  删除图中流量 flow=0 的边;
  建立活动节点到学习者节点之间的匹配  $W (A \rightarrow L)$ ;
   $T_k = \min\{f_{ij} | (A_i, L_j) \in W\}$ ;
  设定时间槽  $[T_{k-1}, T_k]$ ;
  If  $(A_i, L_j) \in W$  &  $i \in [1, n]$  &  $j \in [1, m]$  then
  将  $A_i$  在时间槽  $[T_{k-1}, T_k]$  分配给  $L_j$ ;
  End if
  I=I-T_k;
}

```

从上述算法中可以看出,当  $I$  减小到 0 后,所有边的流量都变成了 0,则算法结束,同时得到一个时间跨度为  $I$  的调度。由于  $I$  已经是该调度模型可能的时间跨度的下限,因此得到的调度结果也是最优的。

## 4 算法的评估

下面对上述算法从理论和仿真实验两个方面进行评估。

### 4.1 算法复杂度分析

由于在算法执行过程中,每次设定的时间跨度为  $T_k = \min\{f_{ij} | (A_i, L_j) \in W\}$ ,因此在循环过程中,每次循环结束后至少有一条边在图  $G$  中被删除。若设  $x$  为处理时间不为 0 的学习活动操作  $O_{ij}$  的总数,则需要匹配的次数最多为  $x$ 。同时根据文献[15]关于二部图最大匹配算法的复杂度研究,可以得出完成一次完全匹配(即在时间槽  $[T_{k-1}, T_k]$  完成活动任务的分配)所需的时间复杂度为  $O(x(m+n)^{0.5})$ ,故最终完成所有活动操作分配的算法总时间复杂度为  $O(x^2(m+n)^{0.5})$ 。

### 4.2 数据论证与实验分析

另一方面,利用 Matlab 对上述算法的时间复杂度分析进行了验证。实验在处理器为 Intel i5(2.27GHz)、内存为 4G 的单机环境下,采用 Matlab2010 进行。

需要说明的是,目前的远程教育个性化学习服务主要以学习内容的各类组合或推荐方法为主,并有相关数据集支持实验仿真,而本文提出的支撑个性化协同学习服务的学习活动调度方法目前暂无标准化的测试数据集可用,因此本文的实验仿真建立在自拟的一组数据集上。该数据集主要结合笔者所在课题组多年来在远程教育领域的项目开发和实际应用的经验而设计,能较好地满足实验仿真的需要,并验证算法的有效性。

鉴于篇幅限制,这里以表 1 所列的一个典型测试用例来阐述实验过程。首先,定义了在某协同学习模式下参与的学习者分别为  $L_1, \dots, L_m, m=5$ ,学习活动  $A_1, \dots, A_n, n=30$ ,即 5 名学习者合作完成 30 项学习活动(任务)。各活动持续进行所需的时间长度如表 1 所列。

表 1 测试数据

学习活动	时间长度	学习活动	时间长度	学习活动	时间长度
A1	21	A2	4	A3	12
A4	42	A5	7	A6	32
A7	12	A8	19	A9	15
A10	45	A11	27	A12	1
A13	45	A14	8	A15	16
A16	49	A17	50	A18	11
A19	17	A20	22	A21	31
A22	2	A23	12	A24	8
A25	16	A26	24	A27	37
A28	41	A29	6	A30	12

这里,假定实际网络学习环境中的活动时间长度的最小值为 1,最大值为 50(即  $t_u=1, t_v=50$ ),具体 30 个学习活动的长度是随机给出的。同时,在调度过程中,活动允许被中断并在另一时刻继续进行。通过 MATLAB 运行得到的调度结果如图 2 所示。

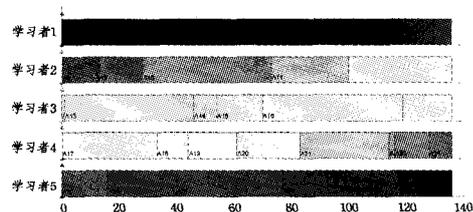


图 2 调度结果图

从图2可以看出,由于活动允许中断并在另一时刻开始,因此在调度结果中没有空隙产生。该调度方案也能让5个学习者能够同时完成学习活动,符合协同学习的约束要求且总的时间跨度为最优值。

根据该调度结果,可得到各个学习者在整个学习过程中的任务参与情况为: $L_1(t_1 \sim t_7)$ ,  $L_2(t_8 \sim t_{11})$ ,  $L_3(t_{12} \sim t_{16})$ ,  $L_4(t_{17} \sim t_{24})$ ,  $L_5(t_{25} \sim t_{30})$ 。另外,若将每个学习者参与完成的任务数定义为贡献度(设每个任务的权重都相同),则可以用图3来描述每个学习者在整个学习过程的参与程度。

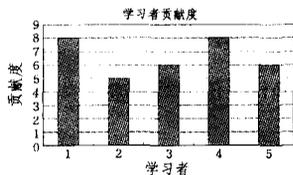


图3 学习者贡献度图

当然,如果各个任务的重要程度并不相同,即权重值不同时,图3中学习者的贡献度可以结合任务权重重新定义。在实际应用中,通过学习者参与任务的情况和贡献度,结合具体任务的属性,可以对学习者的能力水平和学习效果进行评估,并将相关结果作为学习者的历史学习数据加以保存。系统在生成下一次的学习活动调度时,可以利用这些数据使求解的结果更加符合学习者的个体特征。

此外,为了考察算法求解耗费的时间长度与问题规模大小之间的关系,分别以  $m=5, 10, \dots, 30$ ,  $n=10, 15, \dots, 50$  对算法的执行时间进行了评估:

(1) 在学习者数目固定为5,任务数分别为10,15,20, ..., 50时,得到的算法执行时间如图4所示。

(2) 在任务数目固定为50,学习者人数分别为5,10,15, ..., 30时,得到的算法执行时间如图5所示。

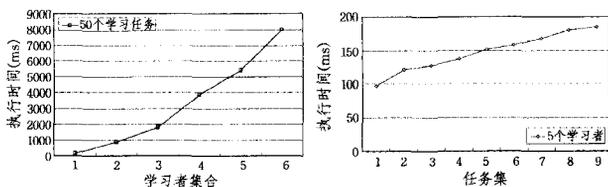


图4 不同任务集的执行时间图

(3) 从(1)、(2)两项测试可以看出,算法执行的时间长度与学习者数目  $n$  和学习任务数目  $m$  之间的关系与上一节中的时间复杂度分析是基本吻合的。为了进一步确认以上关系,对  $m=5, 10, \dots, 30$  和  $n=10, 15, \dots, 50$  进行了综合测试,得到如图6所示的测试结果,该结果也验证了时间复杂度分析的正确性。

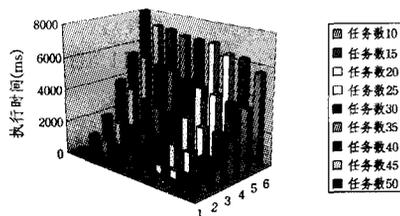


图6 综合测试结果图

通过以上实验可以看出,针对本文定义的协同学习活动调度模型,相应求解算法能在有效的时间内生成时间跨度最短的学习活动调度方案。

**结束语** 本文为远程教育中的协同学习活动设计了一种调度模型并提出了以最优时间跨度为目标的调度算法。根据理论和实验评估的结果来看,该模型和算法能在较短的时间内为学习者生成所需的调度方案,使学习者在该种教学模式下能以最短的时间达到学习目的,从而为进一步提高现有网络教学平台的教学效果,解决学习者缺少活动引导的问题提供了一种有效的解决方法。

本文目前只讨论了协同学习活动的调度方法和问题,下一步工作将结合相关研究项目的成果,对其它网络教学模式及其调度模型展开研究,并以算法池的形式建立全面的解决方案。

## 参考文献

- [1] Srimathi H, Srivatsa S K. SCORM-compliant personalized e-learning using instructional design principle[C]// International Conference on Signal Processing Systems, 2009;738-742
- [2] Pavlis K M, Thanassis K, et al. Integrating adults' characteristics and the requirements for their effective learning in an e-Learning environment[J]. Communications in Computer and Information Science, 2009(49):570-584
- [3] Wang Yan-li, Cheng Yun. The design and implementation of e-learning support service system based on learner model[C]// WRI World Congress on Computer Science and Information Engineering, 2009;411-414
- [4] Jia Hai-yang, Wang Min-hong, et al. Design of a performance-oriented workplace e-learning system using ontology[J]. Expert Systems with Applications, 2011, 38;3372-3382
- [5] Tatyana I. Adaptive open corpus e-learning and authoring, using collaborative ontology learning[C]// IEEE International Conference on Emerging eLearning Technologies and Applications, 2011;83-87
- [6] 徐和祥,张世明. 基于元数据语义模型的数字资源 Top-N 检索[J]. 计算机工程, 2010, 36(22):272-273
- [7] Pan Jian-guo, Zhang Bo-feng, Wang Shu-feng, et al. A personalized semantic search method for intelligent e-learning[C]// International Conference on Intelligent Pervasive Computing, 2007; 11-14
- [8] Granda J C. Quality assessment of speech codecs in synchronous e-learning environments[C]// IEEE International Conference on Multimedia and Expo, 2011
- [9] Gao Min, Wu Zhong-fu, Du Jian-jun. Incorporating Pragmatic Information in Personalized Recommendation Systems[C]// Proceedings of the 11th International Conference on Informatics and Semiotics in Organizations, Beijing, China, 2009;156-164
- [10] Gao Min, Wu Zhong-fu. EPN-based method for web service composition[J]. Lecture Notes in Computer Science, 2009, 5854;345-354
- [11] Gao Min, Wu Zhong-fu. Personalized Context-Aware Collaborative Filtering Based on Neural Network and Slope One[J]. Lecture Notes in Computer Science, 2009, 5738;109-116
- [12] 朱郑州,吴开贵,吴中福,等. 基于知识点本体的个性化课程组织方法[J]. 计算机科学, 2009, 36(12):124-128
- [13] Pinedo M. Scheduling: Theory, Algorithms, and Systems(2nd Edition)[M]. USA: Prentice Hall/Pearson, 2005;16-38
- [14] 卢开澄,卢华明. 图论及其应用(第2版)[M]. 北京:清华大学出版社, 2002;189-190
- [15] West D B. Introduction to Graph Theory (2nd Edition) [M]. USA: Prentice Hall/Pearson, 2004;124-128