

一种求解二部图最大匹配问题新算法及其应用^①

唐敏¹, 关健¹, 邓国强¹, 王海刚²

¹(桂林电子科技大学 数学与计算科学学院, 桂林 541004)

²(桂林电子科技大学 商学院, 桂林 541004)

摘要: 提出了解决二部图最大匹配问题的分层网络优化算法, 并应用新算法对排课问题进行求解。定义了分层网络的概念及匹配的规则, 结合广度优先搜索策略生成分层网络体系, 然后按网络逆序找出最大匹配。实验表明, 算法在解决大规模二部图最大匹配的理论问题和实际应用问题时均能获得准确的结果, 具备良好的性能。

关键词: 分层网络; 二部图; 最大匹配; 排课问题

A New Algorithm and Application of Solving Maximum Matching Problem of Bipartite Graph

TANG Min¹, GUAN Jiang¹, DENG Guo-Qiang¹, WANG Hai-Gang²

¹(School of Mathematics & Computing Science, Guilin University of Electronic Technology, Guilin 541004, China)

²(School of Business, Guilin University of Electronic Technology, Guilin 541004, China)

Abstract: This paper proposed an algorithm for maximum matching of Bipartite graph based on layered network model. Timetable Problems are solved by the new algorithm. A matching rule for layered networks is defined and proposed a concept of layered network first, and a layered network system is generated with the breadth first search strategy. Maximal matching is found according reversed network order. Experiments show that this algorithm can get accurate results and has a good performance with computational complexity in solving large-scale theoretical and practical maximum matching problems of bipartite graph.

Key words: layered network; bipartite graphs; maximal matching; timetable problems

二部图又称作二分图、偶图, 是图论中的一种特殊模型。若能将无向图 $G = \langle V, E \rangle$ 的顶点 V 分成两个子集 V_1 和 V_2 , V_1 与 V_2 的交集是空集, 使得 G 中任何一条边的两个端点一个属于 V_1 , 另一个属于 V_2 , 则 G 为二部图, 记为 $G = (V_1 \cup V_2, E)$ 。图 1(a)和(b)都是二部图, 其中图 1(a)的顶点集为 $V = \{a, b, c, d, e\}$, $V_1 = \{a, e\}$, $V_2 = \{b, c, d\}$, $V_1 \cap V_2 = \emptyset$, 图中有边 $(a, b), (a, c), (a, d), (b, e), (c, e), (d, e)$, 它们中的任意一条边的两个端点, 一个属于 V_1 , 另一个属于 V_2 。图 1(b)的顶点集为 V 可分成两个子集 $V_1 = \{a, d, e\}$, $V_2 = \{b, c, f\}$, 同样满足任意一条边的两个端点, 分属于集合 V_1 和 V_2 。图 1(c)则为非二部图。

二部图在“资源分配”、“工作安排”等问题的求解

中有重要应用。实际生活中常遇到这样的问题: 要给 n 个工作人员安排 m 项任务, n 个人用 $X(x_1, x_2, \dots, x_n)$ 表示, 而 m 项任务用 $Y(y_1, y_2, \dots, y_m)$ 表示, 并不是所有的工作人员都能从事每项工作, 比如 x_1 能做工作 y_1, y_2 , 而 x_2 能做工作 y_2, y_4 , 等等^[1]。应该如何安排工作才能做到最大限度地完成任务, 使任务有人做, 更多的人有工作做, 即 $y_1, y_2, \dots, y_m$ 都有人做, 而 $x_1, x_2, \dots, x_n$ 都有工作做。此问题可以转化为在一个二部图 $G = \langle V, E \rangle$ 中寻求一个最大匹配。在这个问题的求解过程中, 将 G 中顶点集 V 分为两个集合 V_1 和 V_2 , V_1 为 $X(x_1, x_2, \dots, x_n)$, V_2 为 $Y(y_1, y_2, \dots, y_m)$, 若 x_i 能做工作 y_j , 那么边 $(x_i, y_j) \in E$ 。匹配就是由 G 中的边集构成的集合 M , 满足 $M \subseteq E$ 且 M 中的任意两条边没有公共的端点, 若

① 收稿时间:2011-06-20;收到修改稿时间:2011-09-04

一个匹配 M' 的元素数目大于或等于 G 中的任意一个匹配 M 的元素数目, 即 M' 中包含的边数最多, 那么 M' 就是一个最大匹配, 按照这样的匹配安排工作就能保证最大限度地完成任务。如图 2 所示, 为有 6 个工作人员和 6 项任务的工作安排问题, $\{(x_1, y_2), (x_2, y_4), (x_3, y_5), (x_4, y_3), (x_5, y_6)\}$ 和 $\{(x_1, y_1), (x_2, y_2), (x_3, y_5), (x_4, y_3), (x_5, y_6)\}$ 都是一个匹配, 这样的匹配有很多, 但最大匹配是 $\{(x_1, y_1), (x_2, y_2), (x_3, y_4), (x_4, y_3), (x_5, y_6), (x_6, y_5)\}$, 按照这样的对应关系安排工作就能保证每项任务都有人做, 每个人都有工作做。值得强调的是, 最大匹配不是唯一的, 只要包含的边数不小于图 G 中能找到的任意匹配, 它就是一个最大匹配。

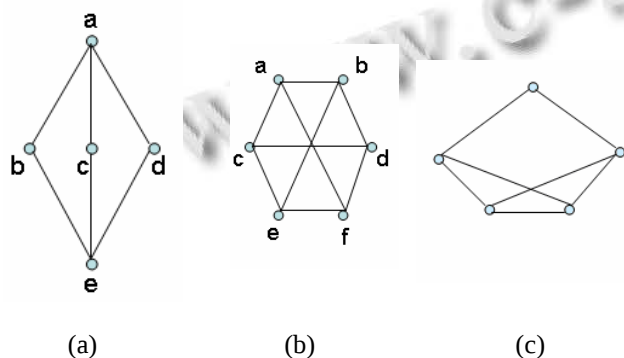


图 1 二部图和非二部图

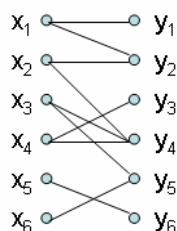


图 2 二部图实例: 工作安排问题

还有许多实际问题都可以转化为在一个二部图中寻求匹配。譬如排课系统设计、军事战略部署、体育竞赛对阵^[2,3]等等, 有学者将其应用在毕业论文的选题^[4]及镜头的检索^[5]中。目前, 此类问题的求解策略可以分为传统方法和现代优化方法两类: 传统方法是基于增广路径来得到最大匹配, 以匈牙利算法为代表^[6], 此外, 还有深度优先算法等^[7]。现代优化方法主要指生物算法, 如基于 DNA 表面积模型算法^[8]。

本文使用一种解决二部图最大匹配问题的新算法, 通过数值实验及实际应用, 验证算法的正

确性和有效性。

1 算法中的重要术语

1.1 二部图中的匹配和最大匹配

给定一个二部图 $G=(V_1 \cup V_2, E)$, M 是 E 的子集, 如果 M 中的任意两条边没有公共的顶点, 则称 M 是二部图 G 的一个匹配。如果 M 是一个匹配, 而不存在其他匹配 M' , 使得 $|M'| > |M|$, M 称为二部图 G 的最大匹配, 即 G 中边数最多的匹配。

图 3 二部图, $\{(a, f), (b, g), (c, h)\}$ 是 G 的一个匹配, $\{(a, f), (b, g), (c, h), (d, e)\}$ 是 G 的最大匹配。图 4 所示的排课问题是各大高校各个学院每学期都必须面临的问题, 假定一个学院有 n 名教师, 用 $T1 \sim Tn$ 表示, 每位教师可以讲授 1~3 门课程, 本学期一共有 n 门课程, 用 $C1 \sim Cn$ 表示, 那么如何安排工作才能做到最大限度的完成任务, 即每门课都有一个老师上, 而每个老师都有一门课上。此问题可以转化为在二部图中寻找最大匹配来求解。

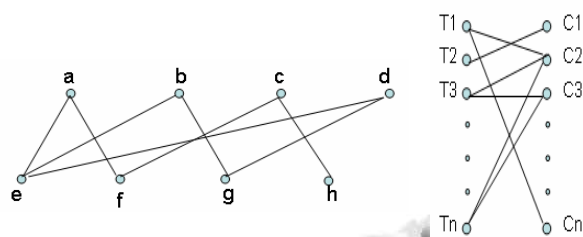


图 3 二部图最大匹配实例 图 4 排课问题

本文使用分层网络优化算法对二部图最大匹配问题进行求解, 并将之用于实际的排课问题中进行测试。

1.2 分层网络

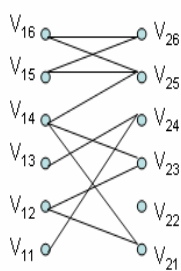
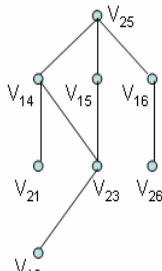
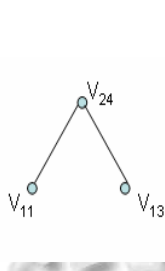
一个分层网络 Net 是二部图 G 的子图, 在新算法中, 一个二部图 G 可以导出若干个分层网络, 为 Net_1 , Net_2 , Net_3 , ..., 所有的分层网络构成求解二部图最大匹配的网络体系 $Nets$ 。

在导出各个分层网络时, 结点和边的产生是有顺序的, 如图 6 所示的 Net_1 , 首先生成结点 v_{25} , 它是第一层, 然后生成结点 v_{14}, v_{15}, v_{16} 及与上一层 (即第一层) 的边 $(v_{25}, v_{14}), (v_{25}, v_{15}), (v_{25}, v_{16})$, 此为第二层, 接着生成结点 v_{21}, v_{23}, v_{26} 及与上一层 (即第二层) 的边 $(v_{14}, v_{21}), (v_{14}, v_{23}), (v_{15}, v_{23}), (v_{16}, v_{26})$, 此为第三层, 最后生成的结点是 v_{12} 及与上一层 (即第三层) 的边 (v_{23}, v_{12}) , 此为第四层。对于图而言, 原本是无层次的, 由于生成 G 的

子图时,有一定的产生顺序,在进行匹配时也和层次有关,故而将之命名为分层网络。

将各层网络中的顶点 $P \in (U \cup V)$ 称为网点,网点之间的连线(简称网线) $Line \in E$,即为 G 的边。

例如,图 5 所示的二部图 G ,图 6、图 7、图 8、图 9 是图 G 导出的分层网络 Net_1 , Net_2 , Net_3 , Net_4 ,它们构成的集合就是图 G 的分层网络体系。

图 5 二部图 G 图 6 Net_1 图 7 Net_2

1.3 网络逆序

在一个分层网络 Net 中,从最下层、最右端的网点开始,按自右向左、自下而上的顺序来实现网点配对的一种匹配顺序。

如图 6 所示的分层网络,按网络逆序得到的匹配是 $\{(V_{12}, V_{23}), (V_{16}, V_{26}), (V_{21}, V_{14}), (V_{15}, V_{25})\}$ 。

2 分层网络优化算法

2.1 算法思想及步骤

求解二部图的分层网络优化算法,思想是通过二部图 G 导出有序分层网络体系,按照一定的规则,找出网络体系中的各个网络中的匹配,它们共同构成了二部图的最大匹配。具体做法如下:

第一步:生成二部图 G

二部图可随机生成,也可按要求进行设定,应用 Matlab,可随机产生 $n \times n$ 的 0-1 矩阵 A_{ij} ($1 \leq i, j \leq n$), $A_{ij} = 1$ 表示 V_{1i} 与 V_{2j} 有边相连,若无边相连,则 $A_{ij} = 0$,二部图 $G = (V_1 \cup V_2, E)$ 可由此矩阵表示,此时 $|V_1| = |V_2| = n$, $M = \emptyset$ 。

第二步:计算二部图中各顶点度数

计算二部图 G 中所有顶点的度数,记作 $d(V_{11}), d(V_{12}), \dots, d(V_{1n}), d(V_{21}), d(V_{22}), \dots, d(V_{2n})$ 。将它们进行排序,找出全部顶点中度数最大的那个顶点做为 Net_1 的起始网点,若有多个,任选其中一个。置 $m = 1$ 。

第三步:迭代生成分层网络 Net_m

按广度优先搜索(BFS)策略得到 Net_m 的其余网点,并作出相应的网线。

第四步:将得到的 Net_m 进行匹配

每次迭代均可获得一个分层网络 Net_m ,按网络逆序进行匹配,此次匹配记为 M_m 。 $M = M \cup M_m$ 。

第五步:判断迭代是否结束

若 M 中包含的顶点为 $V_1 \cup V_2$ 时,最大匹配完成,跳转到第六步。否则在 G 中选出关于 M 非饱和顶点(即在 $V_1 \cup V_2$ 中出现的顶点,而不包含在 M 中的顶点)以及相应的边构成新的二部图,选取度数最大的一个顶点作为新的分层网络的起始网点, $m = m + 1$,跳转到第三步。

第六步:输出最大匹配 M 。

2.2 实例分析

为了更好的说明算法的每一个步骤的进程,本节以一个顶点数为 12 的二部图为例,即 $|V_1| = |V_2| = 6$,如图 5 所示,说明通过分层网络优化算法如何得到此图的最大匹配。

随机生成如图 5 所示的二部图 $G = (V_1 \cup V_2, E)$,其中 $V_1 = \{V_{11}, V_{12}, \dots, V_{16}\}$, $V_2 = \{V_{21}, V_{22}, \dots, V_{26}\}$ 。

计算可得 $(d(V_{11}), d(V_{12}), \dots, d(V_{16})) = (1, 2, 1, 3, 2, 2)$, $(d(V_{21}), d(V_{22}), \dots, d(V_{26})) = (2, 0, 2, 2, 3, 2)$, $M = \emptyset$ 。

根据各顶点度数计算的结果,把 V_{14} 作为 Net_1 的起始网点。

按广度优先搜索策略(BFS)依次寻找分层网络 Net_1 的网点,并作出相应的网线,如图 6 所示。然后按网络逆序进行匹配,记录该匹配 M_1 。依次得到图 G 的各个分层网络 Net_2 , Net_3 , Net_4 ,如图 7~图 9 所示。由图 G 生成的网络体系 $Nets$ 由 Net_1 , Net_2 , Net_3 , Net_4 4 个分层网络组成。

最大匹配 $M = M_1 \cup M_2 \cup M_3 \cup M_4$,其中 $M_1 = \{(V_{12}, V_{23}), (V_{16}, V_{26}), (V_{21}, V_{14}), (V_{15}, V_{25})\}$, $M_2 = \{(V_{13}, V_{24})\}$, $M_3 = M_4 = \emptyset$ 。故 M 内共有 5 个匹配,如图 10 所示。

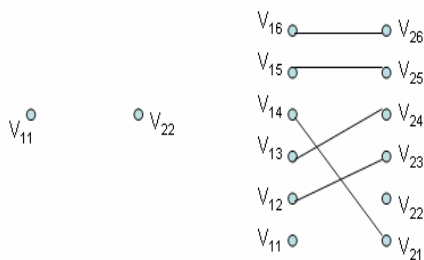


图 8 图 9 图 10 最大匹配

2.3 数值实验

为了说明该算法可以很好地适应大规模问题的求解, 在 Matlab 环境下编程, 执行了多个不同规模的测试用例, 列举顶点数为 40 个顶点的二部图, 即 $|V_1| = |V_2| = 20$ 。

如图 11 所示的二部图 (随机生成), 应用分层网络优化算法求得最大匹配, 如图 12 所示, 经验证 (采用匈牙利算法检验), 图 12 所得到的解是图 11 的最大匹配图。

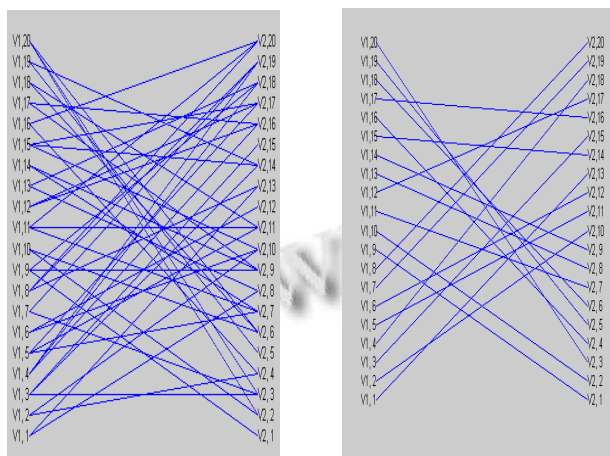


图 11 40 个顶点的二部图 图 12 图 11 的最大匹配

多次数值实验的结果, 表明对于高阶的二部图, 新算法均能获得准确解, 在一定程度上解决了二部图最大匹配问题大规模求解的困难。

二部图 G 的结点为 n 时, 求解二部图最大匹配的分层网络优化算法的时间复杂度为 $O(n \log_2 n)$ 。

3 分层网络优化算法的一个应用

分层网络优化算法的一个应用是解决简化的排课问题。

排课问题可以转化为图论中的二部图最大匹配问题, 如图 13 所示, 顶点集 $T = \{T_1, T_2, T_3, \dots, T_{30}\}$ 为 30 名教师, 顶点集 $C = \{C_1, C_2, C_3, \dots, C_{30}\}$ 为 30 门课程, 各个老师可以讲授的课程与老师之间的关系用边表示, 应用分层网络优化算法, 得到一个最大匹配, 如图 14 所示, 即该问题的一个解决方案。

排课问题可以转化为图论中的二部图最大匹配问题, 如图 13 所示, 顶点集为 30 名教师, 顶点集为 30 门课程, 各个老师可以讲授的课程与老师之间的关系用边表示, 应用分层网络优化算法, 得到一个最大匹

配, 如图 14 所示, 即该问题的一个解决方案。

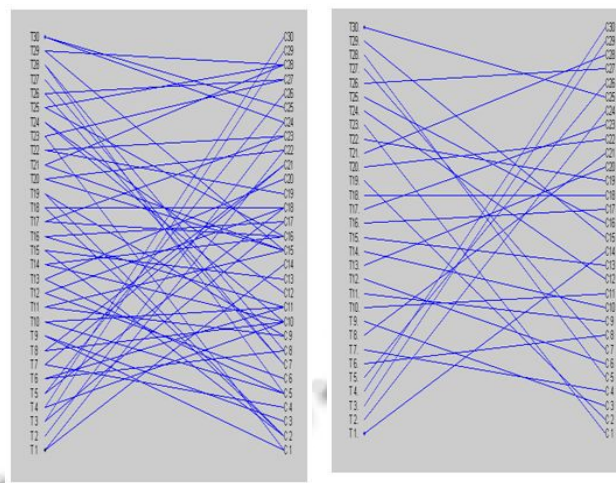


图 13 排课问题的图形转化 图 14 图 13 的一个最大匹配

算法的结果显示, 每个老师都有工作做, 而每门课程都有老师讲授, 即保证最大限度的完成任务。

值得一提的是, 算法并不局限于 30 个老师和 30 门课程, 对于任意人数的老师、任意门数的课程、任意的初始匹配关系, 都可以通过分层网络优化算法得到一个最大匹配, 做为一个参考的排课方案。

4 结语

本文使用分层网络优化算法来求解二部图最大匹配问题, 在建立分层网络的优先匹配规则的基础上, 结合广度优先搜索策略生成分层网络体系, 然后按网络逆序找出最大匹配。在算法的效率上, 时间复杂度为。数值实验表明: 该算法在保证求得二部图最大匹配的前提下, 时间复杂度较低; 易于理解和实现; 适合大规模运算。

对于新算法的应用, 采取了实际的排课问题进行说明, 此类问题均可转化为二部图的最大匹配问题进行求解, 而对于更多的教师和课程, 算法均能求得最大匹配, 得到解决问题的一个有效方案。

二部图最大匹配问题既是一个重要的图论问题, 又是许多实际应用问题的理论模型, 它们都可以转化为在一个二部图中寻求匹配或最大匹配, 比如残缺棋盘问题、拉丁方阵问题等, 只要涉及到两个集合, 而这两个集合的元素之间有着某种联系, 都可以尝试用二部图来解决。在算法的设计上, 也会出现越来越多的求

(下转第 28 页)

因此,平台对功能进行了分解,细化为三级,并采用表格存储客户的订阅信息,按订阅信息进行功能的开放,对应不同的应用界面并进行使用量的度量。

5.5 可伸缩技术

随着平台的不断深化,应用平台的租户(保险公司和第三方)会不断增加,应用系统的直接用户数会呈几何级的速度不断增长,对数据库容量的要求不断增加,对运行性能也提出了很高的要求。平台建立了可伸缩的体系架构,即随着用户数的增加,不必对系统架构进行调整。采用 F5 或其他软件方式实现应用服务器层的负载均衡,实现水平扩展;采用垂直切分、读/写分离、水平切分等方法实现数据库的可伸缩性。

6 总结

中国机动车辆保险信息共享平台的建设,以云计算技术为基础,服务于不同应用和用户。本文分析了中国车险信息平台的现状和面临的问题,介绍了平台的技术架构、服务和关键技术,在基础设施利用方面,通过网络为车险信息共享平台提供 IT 基础设施服务,包括数据存储和网络资源、负载均衡、网络加速等服务;在技术平台方面,通过网络向车险信息共享平台的开发团队提供统一的、可定制、可开发的平台服务,包括可复用的组件和一系列工具;在服务功能方面,通过网络向保险公司和用户车险业务生产云、数据精算云、行业电子商务云、清算支付云服务,以便降低企业信息化成本,提高企业信息化水平,为用户提供一站式服务。

到目前为止,中国机动车辆保险信息共享平台已经上线运行,全国范围内车险的投保都实时通过平台进行,实现了对保费、手续费等的有效控制,对抑制车险恶性竞争、防止欺诈等作用逐步显现。实践表明,基于云计算技术的车险信息共享平台更能适应新形势下新业务特点的要求,达到对保险企业及最终用户的监管与服务的双重目标。

参考文献

- 1 孙祁祥.中国保险业:矛盾、挑战与对策.北京:中国金融出版社出版,2000.
- 2 邓兴玲.汽车保险中无赔款优待系统的改进.西南财经大学,2010.
- 3 黄丽娟.论保险代位权制度的建构.西南财经大学,2010.
- 4 张慧.汽车保险精算模型与奖惩系统研究.新疆大学,2009.
- 5 Armbrust M, Fox A, Griffith R. A view of cloud computing. Communications of the AC M, 2010,53(4):50-53.
- 6 Bhardwaj S, Jain L, Jain S. Cloud Computing: a study of infrastructure as a service. International Journal of Engineering and Information Technology, 2009,2(1): 60-61.
- 7 Marinos A, Briscoe G. Community Cloud Computing. Computer Science, 2009,5931:472-484.
- 8 刘鹏.云计算.北京:电子工业出版社,2010.
- 9 王庆波,何乐.虚拟化与云计算.北京:电子工业出版社,2009.
- 10 彭荣.SaaS 模式下多租户系统架构及关键技术研究.大连海事大学,2010.

(上接第 75 页)

解方法。若求解的问题为 NP 难问题,那么在求得精确解的算法的基础上(这类算法必定花费大量的时间)还可加入现代优化计算方法的某些策略,得到一个折中解。这些研究任务在以后的工作中必定有所体现。

参考文献

- 1 陈光喜,丁宣浩,古天龙.离散数学.北京:电子工业出版社,2008.
- 2 Bruce L. Clarke. Stability analysis of a model reaction network using graph theory. Chemical Physics, 1974,60(4): 1493-1501.
- 3 Woess W. Random Walk on Infinite Graphs and Groups. Cambridge Tracts in Math: Cambridge University Press,

2000: 35-36.

- 4 杨胜超,张瑞军.基于二分图最优匹配算法的毕业论文选题系统.计算机系统应用,2008,17(7):14-17.
- 5 彭宇新, Ngo CW,肖建国.一种基于二分图最优匹配的镜头检索方法.电子学报,2004,32(7):1135-1139.
- 6 Golumbic MC. Algorithmic Graph Theory and the Perfect Graph. New York: Academic Press,1980.57-58.
- 7 李洪波,翟金刚.二部图最大匹配的快速动态优化算法.鲁东大学学报,2006,22(3):168-170.
- 8 刘文斌,高琳,王淑栋,刘向荣,许进.最大匹配问题的 DNA 表面计算模型.电子学报,2003,31(10):1496-1499.