

稳定婚姻匹配问题的一个快速枚举算法

宋旭东, 纪秀花

(山东经济学院计算机科学与技术学院, 山东 济南 250014)

摘 要: 稳定匹配问题是算法理论中的典型问题之一, 稳定婚姻匹配问题则是一种解决二部图匹配问题的模型。论文对稳定婚姻匹配问题进行了简单的阐述, 并介绍了求解典型稳定婚姻问题的 Gale-Shapley 算法的基本思想及其性质。为了快速求出所有的稳定匹配结果, 提出了基于先序遍历森林的快速枚举算法。由 Gale-Shapley 算法的性质得到一个定理及其推论, 利用得到的推论对算法做了进一步改进和优化。在满足推论的特定条件下, 提高了算法的执行效率。

关 键 词: 计算机应用; 算法理论; 稳定婚姻匹配; 先序遍历; 森林; 枚举

中图分类号: TP 391

文献标识码: A

文 章 编 号: 1003-0158(2010)03-0187-06

A Fast Enumeration Algorithm on Stable Marriage Matching Problem

SONG Xu-dong, JI Xiu-hua

(School of Computer Science and Technology, Shandong Economic University, Jinan Shandong 250014, China)

Abstract: The stable matching problem is one of the typical problems in the algorithm theory. The stable marriage matching problem is a model which can be used to solve the problem about bipartite graph matching. In this paper the stable marriage matching problem is introduced briefly, and the basic idea and property of the Gale-Shapley algorithm are described. And then a quick enumeration algorithm based on preorder traversal of forest is designed for quickly enumerating all the stable matching results. A theorem and its inference are derived from the property of the Gale-Shapley algorithm. The algorithm is further improved and optimized by utilizing the inference. The execution time of the original algorithm is improved in specified conditions where the inference is satisfied.

Key words: computer application; algorithm theory; stable marriage matching; preorder traversal; forest; enumeration

1962年两个数理经济学家 David Gale 和 Lloyd Shapley 在一篇名为 “College Admissions

and the Stability of Marriage”^[1] 的论文中首次引出并介绍了稳定婚姻问题。考虑 n 个男人的集合

收稿日期: 2008-09-26

基金项目: 国家自然科学基金资助项目 (60573114); 山东省自然科学基金资助项目 (Y2005G15); 山东省教育厅科技资助项目 (J07YJ10)

作者简介: 宋旭东 (1983-), 男, 山东烟台人, 硕士研究生, 主要研究方向为图像处理和算法。

$M=\{m_1, m_2, \dots, m_n\}$ 和 n 个女人的集合 $W=\{w_1, w_2, \dots, w_n\}$ 。令 $M \times W$ 为所有可能的形如 (m, w) 的有序对的集合。其中 $m \in M, w \in W$ 。一个匹配 S 是来自 $M \times W$ 的有序对的集合,并且具有下述性质:每个 M 的成员和每个 W 的成员至多出现在 S 的一个有序对中。一个完美匹配 S' 是具有下述性质的匹配: M 的每个成员和 W 的每个成员恰好出现在 S' 的一个对里。一个完美匹配仅仅对应于一个男女配对的方式,以这种方式,每个人最终与某个人结婚,且没有人与多个人结婚。在完美匹配的背景下又引入优先的概念,每个男人 $m \in M$ 对所有的女人排名,如果 m 给 w 的排名高于 w' ,我们就说 m 偏爱 w 超过 w' 。我们将把 m 的按顺序的排名作为他的优先表,但不允许排名出现并列的情况。类似的,每个女人也对所有男人排名。给定一个完美匹配 S ,在 S 中存在两个配对 (m, w) 和 (m', w') ,他们具有 m 更偏爱 w' 而不爱 w ,且 w' 更偏爱 m 而不爱 m' 的性质,在这种情况下,我们称这个完美匹配 S 是不稳定的,我们称像 (m, w') 这样的对是一个相对于 S 的不稳定因素。我们说一个匹配 S 是稳定的,那么要满足以下两点:① S 是完美匹配;② 不存在相对于 S 的不稳定因素^[2-9]。

这个问题被应用到学生的入学、工作招聘等处理过程中,最早是在美国国家医生匹配计划中一个类似的过程被用来将医生和医院进行匹配^[2]。

目前, Gale-shapley算法^[1-2],回溯法和回跳法^[10]能有效的让我们找到其中一个稳定匹配结果,其中最著名的算法是Gale-shapley算法。本文设计了一种基于先序遍历森林的快速枚举算法,以求得所有可能的稳定匹配结果,并且由Gale-shapley算法的性质推得一个定理及其推论,又根据推论进一步改进了这个枚举算法。

1 Gale-Shapley 算法

1.1 Gale-Shapley 算法基本思想

Gale-Shapley算法的基本思想如下^[2]:

(1) 初始,每个人都是未婚的。假设一个未婚的男人 m 选择了他的优先表上排名最高的女人 w ,并且向她求婚。我们不能立刻声明 (m, w) 将是最后稳定匹配中的一对,因为在将来的某个时候,女人 w 偏爱的男人 m' 可能向她求婚。另一

方面,对 w 来说,立刻拒绝 m 可能是危险的,她之后可能没有接收到来自她的优先表上排名高于 m 的某人的求婚。于是,一种自然的想法是使 (m, w) 这个对进入一种中间状态——约会。

(2) 假设我们现在处在某种状态,某些男人和女人是自由的(没有约会),某些是有约会的。任意一个自由的男人 m 选择他还没有求过婚的最高排名的女人 w ,并且向她求婚。如果 w 也是自由的,那么 m 和 w 就成为约会状态。否则, w 已经在与某个其他的男人 m' 约会,在这种情况下,根据 w 的优先表中 m 和 m' 的排名来选择,排名较高的男人变成与 w 约会而另一个人变成自由的。

(3) 最后,当没有一个人是自由的时候,算法将结束;此刻所有的约会将被宣告为最后的结果,且将返回最终的完美匹配。文献[2]中给出证明,该完美匹配是一个稳定匹配。

在上述算法中,由于总是未婚的男人向女人求婚,它常常被称为“man propose, women dispose”(“男性优先选择”,下文都用“男优先”表示)。当然该算法实现时也可以总是未婚的女人向男人求婚,这被称为“woman propose, men dispose”(“女性优先选择”,下文都用“女优先”表示)。

1.2 Gale-Shapley 算法的性质

根据上述Gale-Shapley算法的基本思想,可以得到该算法的几个性质^[1-3,9]。

在稳定婚姻问题中,设有 N 个男人和 N 个女人,男人 i 对女人 j 的偏好排名用一个整数 $k=MXH(i, j)$ 表示(MXH 为 $N \times N$ 矩阵,表示男人对女人的偏爱程度列表)。同理,女人 i 对男人 j 的偏好排名用一个整数 $t=WXH(i, j)$ 表示(WXH 为 $N \times N$ 矩阵,表示女人对男人的偏爱程度列表), $1 \leq k$ (或 t) $\leq N$ 。

按照 Gale-Shapley 算法,我们一定能得到一个稳定婚姻。也就是说,不管这 N 个男人和 N 个女人的优先表是如何分布的,至少存在一个稳定婚姻匹配。在“男优先”算法实现中,得到的一个稳定婚姻匹配具有如下三点性质:

(1) 男性能够获得尽可能好的伴侣,结果是:如果还存在其他的稳定匹配,那么里面任何一个男性的伴侣排名都不会比“男优先”得到的结果更好,我们说此种情况每个男性获得的是

“最好”的伴侣。

(2) 女性却只能被动地一步步接近她最爱的目标, 但最后往往达不到, 结果是: 如果还存在其他的稳定匹配, 那么里面任何一个女性的伴侣排名都不会比“男优先”得到的结果更差, 我们说此种情况每个女性获得是“最差”的伴侣。

(3) 无论男性们求婚的先后顺序如何, 最终得到的是同一个稳定婚姻匹配。

当然, 这个算法的实现也可是: “女优先”, 同样, 得到的此稳定婚姻匹配具有对应的三点性质:

(1) 在所有可能的稳定婚姻匹配中, “女优先”得到的结果中每个女性获得的是“最好”的伴侣。

(2) 在所有可能的稳定婚姻匹配中, “女优先”得到的结果中每个男性获得的是“最差”的伴侣。

(3) 无论女性们求婚的先后顺序如何, “女优先”最终得到的是同一个稳定婚姻匹配。

2 基于先序遍历森林的算法

Gale-Shapley算法只是能找到一个稳定匹配

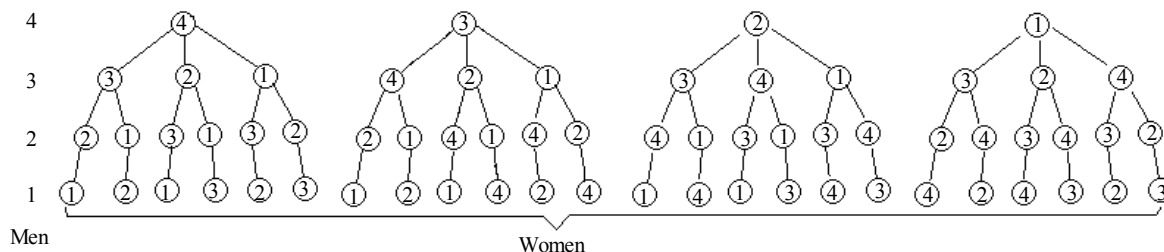


图1 婚姻匹配森林 ($N=4$)

2.1 新算法基本思想

要检查每一种婚配方式是否存在不稳定因素, 可按先序遍历的顺序访问整个森林, 当每到一结点时, 需要将此结点对应的配对分别与其祖先的某一些结点(即满足一定条件的结点, 后面将做详细说明)对应的配对按照优先表, 检查是否存在不稳定因素:

(1) 假设不存在不稳定因素, 若访问的结点为叶子结点, 则输出此稳定婚姻匹配; 若不是叶子节点则继续向下层遍历访问。

(2) 假设存在不稳定因素, 则跳过此结点

结果。如果我们想找到所有的稳定匹配结果, 若用穷举法解决这个问题, 对有 N 个男人和 N 个女人的情形, 需要考虑 $N!$ 种可能的婚姻匹配结果, 并且需要对其中每一种婚配方式检查其是否存在不稳定因素。显然, 该算法的复杂度太大。

当有 $k(>1)$ 种稳定匹配结果, 下面将设计算法找出这 k 种稳定匹配。可将这 $N!$ 种可能的婚姻匹配结果构成一森林(见图1, $N=4$ 情形)。树中各结点的数字对应女性序号, 树的每一层对应同一男性序号(见图1中最左侧数字)。从森林中每棵树的根到叶子结点路径上, 所有结点中的数字组成一个数字串表示一种完美婚姻匹配。

例1 第一棵树最左边路径④③②①对应的婚姻匹配为:

man 4 – woman ④、man 3 – woman ③、man 2 – woman ②、man 1 – woman ①

例2 第二棵树最右边路径③①②④对应的婚姻匹配为:

man 4 – woman ③、man 3 – woman ①、man 2 – woman ②、man 1 – woman ④

及其所有子孙结点, 继续向其他分支遍历访问。直到遍历结束。

这样做的优点是消除了许多重复的操作, 从而大大节省了时间。

2.2 新算法的具体描述

整个过程按照先序遍历的顺序访问森林的每个结点, 每到一个结点按照优先表, 检查其配对情况与其祖先的配对情况, 看是否存在不稳定因素。下面给出求解所有稳定婚姻匹配结果的算法实现描述。该算法实现主要包括两个过程: 一个是主过程GSM(), 一个是子过程SM()。

```

GSM( $N$ ,  $MXH$ ,  $WXH$ )
// 有 $N$ 个男人,  $N$ 个女人;
//  $MXH$ 为 $N \times N$ 矩阵, 表示男人对女人的偏爱
程度列表(优先表);
//  $WXH$ 为 $N \times N$ 矩阵, 表示女人对男人的偏爱
程度列表(优先表);
{
  If  $N > 1$  then
    FOR  $m=N$  to 1           //对应各棵
    树的根结点。
       $X()=[1,2,\dots, N]$     //  $X()$ 数组表
      示配对情况, 即( $n, X(n)$ )对应其中一个配对。首
      先初始为 $1 \sim N$ 
      //自然数,
      即: 男性 $n$ 与女性 $n$ 配对。
      Swap  $X(N), X(m)$       //通过交换,
      让男性 $N$ 与女性 $m$ 配对;
      //即顺序遍
      历森林的每棵树。
      SM( $N-1, MXH, WXH, X$ ) //向下
      遍历孩子结点。
      Next
    END If
  END
}
SM( $n, MXH1, WXH1, Y$ )
// 有 $n$ 个男人,  $N$ 个女人;
//  $MXH1$ 为 $N \times N$ 矩阵, 表示男人对女人的喜
好程度列表;
//  $WXH1$ 为 $N \times N$ 矩阵, 表示女人对男人的喜
好程度列表;
//  $Y$ 为一维数组, 表示当前婚姻匹配情况;
{
  If  $n=1$  then             //是否到叶子结点
    将 $Y(1)$ 对应的配对分别与其祖先结点 $Y(2)$ 、
     $Y(3)$ 、 $\dots$ 、 $Y(N)$ 对应的配对进行比较看是否有不
    稳定因素(具体检测方案见下说明);
    If 没有不稳定因素, 则输出 $Y()$ 当前的
    稳定婚姻匹配。
  Else
    FOR  $mm=n$  to 1
      Swap  $Y(n), Y(mm)$  //让男性 $n$ 与女

```

性 $Y(mm)$ 配对,

//即让男性 n 依

次与其他女性配对。

将 $Y(n)$ 对应的配对分别与其祖先结点 $Y(n+1)$ 、 \dots 、 $Y(N)$ 对应的配对进行比较看是否有不稳定因素(具体检测方案见下说明);

If 没有不稳定因素, 则SM($n-1, MXH, WXH, Y$) //递归向下一孩子结点访问。

Next

END IF

END

}

说明: 当每到一结点时, 要检测该结点对应的配对($manX, womanY$)加入后是否会出现不稳定因素, 只需考虑两种情况^[10]: ① 男性 $manX$ 是否会与其他已婚女性产生不稳定因素, 在这里并不需要考虑所有已婚女子(即其祖先的所有结点对应的配对女性), 而只需考虑男性优先表中 $manX$ 对女性排名的前 $k-1$ ($k=MXH(manX, womanY)$)个女性当中的已婚女性, 因为排名在 k 以后的已婚女性一定不会与男性 $manX$ 产生不稳定因素; ② 女性 $womanY$ 是否会与其他已婚男性产生不稳定因素, 同样在这里也不需要考虑其所有祖先结点对应的所有已婚男子, 而只需考虑女性优先表中 $womanY$ 对男性排名的前 $L-1$ ($L=WXH(womanY, manX)$)个男性中的已婚男性。只有上述两种情况都为否定, 才能继续向下遍历孩子结点。

此算法是以递归的方式实现。从算法的流程易看出, 每种婚姻匹配都已考虑在内, 说明此算法是正确的。此算法的优点: 对于众多不同的婚姻匹配, 它们包含相同的配对部分是不会重复判断的, 这样大大节省了时间。为了进一步提高速度, 还可对上述算法进一步改进。

3 改进算法

通过对按照“男优先”或“女优先”实现的Gale-Shapley算法性质的研究, 可推得如下定理:

定理 对某优先表分布情况, 分别按照“男优先”和“女优先”来求得两个稳定婚姻匹配, 若这两个稳定婚姻匹配是相同的, 则此种情况就

只存在一个稳定婚姻匹配;若这两个稳定婚姻匹配是不相同的,则此种情况就存在多于1个的稳定婚姻匹配。

证明 定理的后半部分显然正确。下面对定理的前半部分给出证明。

假设有 N 个男人为: $A, B, C, D, \dots, X, \dots, Y, \dots$, 有 N 个女人为: $a, b, c, d, \dots, x, \dots, y, \dots$ 。“男优先”和“女优先”两个稳定婚姻匹配结果相同都是: $A-a, B-b, C-c, D-d, \dots$ (记为 α 匹配);

如果还存在另一个与上不同的稳定婚姻匹配,对其中不相同的一部分,不妨可表示为: $\dots, B-x, \dots, Y-b, \dots$ (记为 β 匹配)。

因为 α 匹配是按照“男优先”或“女优先”来求得的,所以根据前面提到的 Gale-Shapley 算法的性质可知:在所有可能的稳定婚姻匹配中, b 是男性 B 获得的最好的女性伴侣,同样, B 是女性 b 获得的最好的男性伴侣,这样,针对 β 匹配中,男性 B 更喜欢女性 b 胜过 x ,同样女性 b 也更喜欢男性 B 胜过 Y ,所以推得 β 匹配不是稳定婚姻匹配,与假设不符,所以不会存在其他的稳定婚姻匹配。□

由上述定理,可得如下推论:

推论 当分别按照“男优先”和“女优先”求得两个不同的稳定婚姻匹配时,若这两个不同的稳定婚姻匹配中有一部分配对是相同的,则:若还存在其他稳定婚姻匹配,仍会包含这部分相同的配对。

证明 假设分别按照“男优先”和“女优先”求得两个不同的稳定婚姻匹配时,若这两个不同的稳定婚姻匹配中有 m 个配对是相同的,令:

这 m 个配对中男人为: $A, B, C, D, \dots, X, \dots, Y, \dots$

女人为: $a, b, c, d, \dots, x, \dots, y, \dots$

设按照“男优先”和“女优先”得到的两个稳定婚姻匹配中都共同含有如下 m 个配对:

$A-a, B-b, C-c, D-d, \dots$ (记为 α 匹配)

假设 还存在另一个稳定匹配其中不完全包括这个公共部分,有不相同处,对其中不相同的一部分,不妨可表示为:

$\dots, B-x, \dots, Y-b, \dots$ (记为 β 匹配)

剩余部分证明与上述定理的证明相似,在此略。□

根据此推论,可得到改进算法的基本思想:首先分别按照“男优先”和“女优先”求得两个不同的稳定婚姻匹配,根据推论可知,若这两个不同的稳定婚姻匹配中有 m 个配对是相同的,则所有稳定婚姻匹配中的这一部分是已知的。这样,在求解之前,可将这些已配对的 m 个男性和 m 个女性从问题中删去。这样问题就降为求 $N-m$ 个男人和 $N-m$ 个女人的稳定婚姻匹配了。

改进算法具体描述:

G_GSM(N, MXH, WXH)

{

按“男优先”求得一个稳定婚姻匹配;

按“女优先”求得另一个稳定婚姻匹配;

若二者相同,则输出这唯一的稳定婚姻匹配,退出。

若二者不相同:

找出二者相同的配对部分;

若无相同的配对部分,则执行GSM();

若有 $m(m \neq 0)$ 个相同的配对部分,则:

输出这 $m(m \neq 0)$ 个相同的配对部分;

将 MXH, WXH 中的这些男、女对

应的行列删除;

// MXH, WXH 变为 $(N-m) \times (N-m)$ 矩阵

G_GSM($N-m, MXH, WXH$);

END

}

此改进算法的正确性说明:此思想是依据推论,所以已证明。

此改进算法的时间复杂性分析:

(1) 若求得两个稳定婚姻匹配中有 m 个配对是相同的,问题可就转化为求 $(N-m)$ 对男女的婚配问题,则图1中森林的规模将减小 $N(N-1)\dots(N-m+1)$ 倍,所以算法的速度比原算法GSM()的速度提高 $N(N-1)\dots(N-m+1)$ 倍。

(2) 若求得两个稳定婚姻匹配中无配对是相同时,也只是多了2次运行Gale-Shapley算法的时间,相比之下,不会影响原算法GSM()的时间复杂度。

4 结 束 语

稳定婚姻问题被应用到许多实际问题的处理过程中,例如学生的入学,工作招聘,医生和医院进行匹配等。为快速找出所有可能的稳定匹配结果,我们设计了基于先序遍历森林的快速枚举算法。利用此算法,对于众多不同的婚姻匹配,不会重复判断它们包含相同的配对子部分,这样大大节省了时间。为了进一步提高速度,由 Gale-Shapley 算法的性质证明得到了一个定理及其推论,并利用推论对算法做了进一步改进。在满足推论的特定状况下,提高了原算法的执行时间。

参 考 文 献

- [1] Gale D, Shapley L S. College admissions and the stability of marriage [J]. American Mathematical Monthly, 1962, 69: 9-15.
- [2] Jon Kleinberg, Éva Tardos. Algorithm design [M]. Addition Wesley, 2005. 1-12.
- [3] Hopfield J J, Tank D W. Neural computation of decisions in optimization problems [J]. Biological Cybernetics, 1985, (52): 141-152.
- [4] McVitie D G, Wilson L B. The stable marriage problem [J]. Communications of the ACM, 1971, 14(7): 486-492.
- [5] Irving R W, Leather P. The complexity of counting stable marriages [J]. SIAM Journal on Computing, 1986, 15(3): 655-667.
- [6] Knuth D E. Marriage stable [J]. Les Presses de L'universite de Montreal, Montreal, 1976, (8): 66-68.
- [7] Irving R. An efficient algorithm for the "stable roommates" problem [J]. J. Algorithms, 1985, (6): 577-595.
- [8] Wilson L B. An analysis of the stable marriage assignment problem [J]. BIT, 1972, (12): 569-575.
- [9] Gusfield D, Irving R W. The stable marriage problem, structure and algorithms [M]. MIT Press, 1989. 1-3, 5-20.
- [10] 郭东亮, 张立臣. 用回跳法求解稳定婚姻问题[J]. 计算机应用研究, 2005, 22(1): 59-60, 63.