

doi:10.3969/j.issn.1673-9833.2013.01.015

优化的Gale-Shapley算法在学生选课中的应用

邓蔚之, 刘 强, 任志虎, 黄盛达, 张 月

(湖南工业大学 计算机与通信学院, 湖南 株洲 412007)

摘 要: 学生选课是匹配问题, 提出了利用改进的 Gale-Shapley 算法来解决。根据学生对课程的喜爱程度, 以课程为单位对学生进行集合划分, 课程对该集合中喜爱度相同的学生发放随机种子, 再对选课的学生进行排序和录取, 对录取完的学生进行哈希标记, 并释放相应的存储空间, 从而完成学生和课程之间双向选择匹配。算法分析结果表明: 本优化算法降低了空间和时间复杂度, 且能完成学生与课程之间最大满意度的匹配。

关键词: Gale-Shapley 算法; 伪优先表; 哈希标记; 排序

中图分类号: TP301.6

文献标志码: A

文章编号: 1673-9833(2013)01-0067-04

The Application of Optimal Gale-Shapley Algorithm in Students Courses Selection

Deng Weizhi, Liu Qiang, Ren Zhihu, Huang Shengda, Zhang Yue

(School of Computer and Communication, Hunan University of Technology, Zhuzhou Hunan 412007, China)

Abstract: The courses selection is the matching problem, and an improved Gale-Shapley algorithm is put forward. According to the students' preference to a course, the course is taken as unit to set division of students, and the random seed is issued to the students with the same preference, then the students are sorted and admitted, the admission students are signed with the hash mark and the relative storage space is released, thus the two-way selection match between the students and the curriculum is completed. Algorithm analysis results show that the optimal algorithm reduces the space and time complexity and realizes the maximum satisfied match between the students and the curriculum.

Keywords: Gale-Shapley algorithm; pseudo-priority table; the hash mark; sort

0 引言

婚姻稳定匹配问题^[1-2]是 Gale 和 Shapley 于 1962 年提出的一种解决二部分图匹配问题的模型。典型的婚姻稳定问题可描述为: n 个男人构成集合 $N = \{p_1 | p_1, p_2, \dots, p_n\}$, n 个女人构成集合 $M = \{q_i | q_1, q_2, \dots, q_n\}$, 令 $N * M$ 得到的有序对表示为 (a, b) , 其中 $a \in N, b \in M$ 。用字母 A 来表示 $N * M$ 形成的笛卡尔积^[3], 且 A 具有下

述性质: N 的元素和 M 的元素随机组成的有序对出现在匹配 A 中一次且仅有一次。完美匹配 A' 具有下述性质: N 的元素和 M 的元素恰好出现在 A' 的有序对里一次且仅有一次。完美匹配^[1-2]是一对男女配对且互为理想对象, 不存在一个人有多个理想对象的情况。在完美匹配的背景下, 又引入优先级的概念, 男人 a 按自己的选择意愿对 n 个女人进行优先级排序, 但不允许出现并列的情况。如果男人 a 给女人 b 的排序

收稿日期: 2012-11-03

基金项目: 湖南工业大学教研委托基金资助项目(2011A03), 湖南省教育厅科研基金资助项目(11C0401), 湖南省自然科学基金资助项目(11JJ4050)

作者简介: 邓蔚之(1990-), 男, 湖南岳阳人, 湖南工业大学学生, 主要研究方向为 WEB 技术开发与应用,

E-mail: dengwz7788@gmail.com

高于女人 b' ，则表示 a 更喜欢 b 。同理，每个女人也对所有男人进行排序，最后将男女进行婚姻匹配。

定义一个完美匹配 A' ，若 A' 中存在 2 个配对 (a, b) 和 (a', b') ，其表示 a 的理想对象是 b 而不是 b' ，且 a' 的理想对象是 b' 而不是 b ，在这种情况下，称这个完美匹配 A' 是稳定的。但是，如果不符上述定义，则表明该完美匹配是不稳定的。

婚姻稳定匹配问题就是对匹配双方的理想对象进行最大满意度的描述。由于婚姻稳定匹配问题是双向选择，所以只要涉及双向选择的匹配问题都可以用该算法解决。学生选课也是匹配问题，但课程可选择多个学生，是一对多的匹配问题，因此，本文对 Gale-Shapley 算法进行了优化，以解决学生选课问题。

1 优化 Gale-Shapley 算法

1.1 算法设计

基于 Gale-Shapley 算法^[4-5]的基本思想和匹配方式，本文对其进行了优化和改进，以解决学生选课匹配问题。学生根据个人的理想对象匹配到其最喜欢的课程，而课程要根据其选课人数的限制来录取学生。由于婚姻匹配问题中，不存在一对多的情况，而在课程匹配问题中存在一对多的情况，因此，课程对学生就没有了“优先表”。针对以上问题，若多个学生对一门课程的喜爱度相同，则给其发放随机种子，即将随机种子作为学生对课程的喜爱度，最后课程对所有的学生排序，得到“伪优先表”，从而解决喜爱度并列的问题。学生对课程的喜爱度用自然数表示，如 1, 2 等，而随机种子是以学生对课程的喜爱度和该喜爱度相同的学生人数为依据发放，假设学生对课程的喜爱度为 i ，则随机种子的表达式为： $i+l$ 个上标“'”，如 i' ， i'' 等，其中，“'”的个数表示随机种子是该并列情况中发放的第 l 个。喜爱度的排序定义如下。

定义 喜爱度为 $1 >$ 喜爱度为 $1' >$... > 喜爱度为 $2 >$ 喜爱度为 $2' >$... > 喜爱度为 $3 \dots$ ，其它依次类推。

全部学生 \$student 和全部课程 \$task，通过函数 SXT 的处理，得到二维优先表 \$yxb，其代码如下：

```
function SXT(array $student, array $task){
    foreach($student as $stu_value){
        $stuid = $stu_value["id"];
        foreach($task as $task_val){
            $taskid = $task_val["id"];
            $yxb["$taskid"][$stuid] = $stu_value["like"];
        }
    }
}
```

```
}
}
}。
```

本文算法流程如图 1 所示。

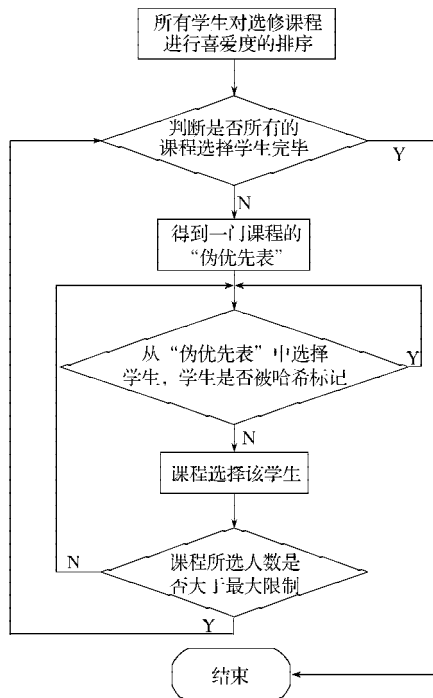


图 1 算法流程图

Fig. 1 Algorithm flowchart

优化的 Gale-Shapley 算法的具体步骤如下：

1) 学生先对所有选修课程进行喜爱度的排序，得到一张二维表。假设有 7 个学生，3 门选修课程，学生对课程喜爱度见表 1。

表 1 学生对课程喜爱度表

Table 1 The students' preference list of courses

学生	课程 1	课程 2	课程 3
学生 1	3	1	2
学生 2	3	2	1
学生 3	1	2	3
学生 4	1	2	3
学生 5	2	3	1
学生 6	2	1	3
学生 7	2	3	1

2) 根据所得二维表，当一门课程开始录取学生前，先要对该门课程喜爱度相同的学生发放随机种子，得到“伪优先表”。例如：课程 1 所建立的“伪优先表”为{学生 1: 3'，学生 2: 3，学生 3: 1，学生 4: 1'，学生 5: 2，学生 6: 2'，学生 7: 2''}，根据文中对喜爱度的排序定义，可得课程被学生选择的优先级别为：学生 3 > 学生 4 > 学生 5 > 学生 6 > 学生 7 > 学生 2 > 学生 1。

3) 每门课程根据“伪优先表”和课程选课人数

限制来选择学生。依据课程得到的“伪优先表”对学生进行排序, 并将排序结果存放于数组 \$Stu 中; 再对学生节点进行哈希标记检查, 如果学生没有被标记, 则表示符合条件, 课程可选择该学生, 并对其进行哈希标记^[6](MD5 加密), 且计数器数值增加 1; 重复哈希标记检查操作, 直到计数器的数值达到选课人数的最大限制; 课程录取学生完毕, 对未标记的学生节点进行释放, 从而进入下一门课程的选择, 且对已选课的学生不再进行排序。如课程 1 已选择学生完毕, 所选学生有学生 3、学生 4 和学生 5, 则这 3 个学生都要被哈希标记, 不能进入课程 2 的选择。部分代码如下:

```

For($i=0;$i<=count($Stu);$i++){
    $j = 0; //设置计算器
    If($j==$num){
        Return 1;
    }
    Else{
        Foreach($Stu as $value){
            If($value["hask"]!=md5("ok")){
                $value["hask"]=md5("ok");
                $j++;
            }
        }
    }
}

```

4) 重复步骤 2~3, 直到所有的课程录取学生完毕, 即完成学生与课程之间的最大满意度匹配。

1.2 空间回收机制

课程录取学生前, 第一次申请的学生节点空间是最大的, 当一门课程录取学生完毕, 已录取的学生节点会退出下一门课程的录取, 因此, 学生节点的存储空间会逐步缩小。

假设一门课程一次录取 X 个学生, 一共有 N 个学生和 M 门课程, 第一次课程录取时, 需要的存储空间为 N , 则第二次课程录取的存储空间为 $N-X$, 依此类推, 第三次存储空间为 $N-2X$, 直到 M 门课程录取完毕时, 存储空间为 $N-(M-1)*X$ 。因此, 当课程循环录取学生时, 动态申请的空间会逐渐减小, 这样可降低算法的空间复杂度, 提高算法效率。部分代码如下:

```

for($i=1;$i<=$tasknum;$i++){
    $space = $studentnum-($i*$num);
    Get_student($space,$student,$task)
} //调用学生匹配课程的函数。

```

// \$tasknum 表示课程数, \$studentnum, 表示学生人数, \$space 表示运行空间, \$num 表示课程需求人数。

1.3 哈希标记

哈希标记^[7-10]的目的是为了对处理后的节点进行标记, 以免重复处理, 造成错误数据进入队列。课程录取学生前, 本算法会对每个学生节点进行哈希标记检查, 即只有所有的学生节点都没有被哈希标记, 课程才能开始录取学生。录取后的学生节点的哈希值属性将被加密, 用 MD5 加密的 40 位密钥。

2 算法复杂度分析

2.1 空间复杂度

假设有 N 个学生和 M 门课程, 每门课程录取人数都为 X , 且 $X*M=N$, 形成序列对 (n, m) 需开辟一个 $N*M$ 的存储空间, 课程对学生的筛选, 还需开辟一个长度为 N 的存储空间, 每完成一门课程的录取工作, 筛选空间就会减小, 即 $N-X*N'$, 其中 N' 表示已经完成的课程数。因此, 本算法的空间复杂度^[6]为

$$S = N*M + N*X - [(M-1)*(M-2)]*X/2。$$

2.2 时间复杂度

由于数据都是离线处理^[8-9], 所以对于 N 个学生和 M 门课程的获取是在课程选择之前得到的, 不加入算法耗时中。因此, 算法的时间复杂度为

$$O = N+M+M*N。$$

3 结语

稳定婚姻问题是一个 NP 问题, 其思想被应用到许多实际问题中。为解决学生选课中一对多的问题, 提出了基于 Gale-Shapley 的优化算法, 使每个学生都选择其最满意的课程, 并分析了该算法的空间复杂度和时间复杂度。这说明了该算法能完成学生与课程之间最大满意度的匹配, 通过动态空间降低了空间和时间复杂度, 提高了算法的效率。

参考文献:

- [1] 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.
Guo Dongliang, Zhang Lichen. Solving Stable Marriage Problem by Backjumping Method[J]. Application Research of Computers, 2005, 22(1): 59-60.
- [2] 宋旭东, 纪秀花. 稳定婚姻问题的研究[C]//全国第19届计算机技术与应用学术会议(CACIS 2008). 合肥: 中

- 国科学技术大学出版社, 2008: 968-972.
- Song Xudong, Ji Xiuhua. Study on Stable Marriage Problem [C]/The 19th National Conference on Computer Technology and Application (CACIS 2008). Hefei: The Press of University of Science and Technology of China, 2008: 968-972.
- [3] Alsuwaiyel M H. 算法设计技巧与分析[M]. 北京: 电子工业出版社, 2004: 16-39.
- Alsuwaiyel M H. Algorithms Design Techniques and Analysis[M]. Beijing: Publishing House of Electronics Industry, 2004: 16-39.
- [4] 张楠, 齐俊玲. 稳定完备婚姻问题的算法及推广[J]. 软件, 2012, 33(9): 112-114.
- Zhang Nan, Qi Junling. The Algorithm Theory and Extension on the Stable Matching Problem[J]. Software, 2012, 33(9): 112-114.
- [5] 范华, 秦茂玲, 张新法. 典型“稳定婚姻问题”的简明矩阵算法实现[J]. 山东师范大学学报: 自然科学版, 2007, 22(1): 17-21.
- Fan Hua, Qin Maoling, Zhang Xinfu. Brief Matrix Algorithm Realization of Typical "Stable Marriage Problem" [J]. Journal of Shandong Normal University: Natural Science, 2007, 22(1): 17-21.
- [6] 严蔚敏, 吴伟民. 数据结构: C语言版[M]. 北京: 清华大学出版社, 1997: 4-36.
- Yan Weimin, Wu Weimin. Data Structure: C Language Version[M]. Beijing: Tsinghua University Press, 1997: 4-36.
- [7] 魏晓玲. MD5 加密算法的研究及应用[J]. 信息技术, 2010(7): 145-147.
- Wei Xiaoling. Research and Application of MD5 Encryption Algorithm [J]. Information Technology, 2010(7): 145-147.
- [8] 张浩华, 齐维毅, 赵子夫, 等. 基于 JAVA 技术的 MD5 加密算法的设计与实现[J]. 沈阳师范大学学报: 自然科学版, 2009, 27(1): 75-77.
- Zhang Haohua, Qi Weiyi, Zhao Zifu, et al. Design and Implementation of MD5 Algorithm Based on JAVA Technology[J]. Journal of Shenyang Normal University: Natural Science, 2009, 27(1): 75-77.
- [9] 么丽颖. MD5 算法的分析和改进[J]. 哈尔滨师范大学自然科学学报, 2011, 27(5): 34-37.
- Yao Liying. Analysis and Improvement of MD5 Algorithm [J]. Natural Science Journal of Harbin Normal University, 2011, 27(5): 34-37.
- [10] 孙维国, 李浩然. MD5 算法在数据安全中的应用及安全性分析[J]. 微计算机应用, 2010, 31(10): 66-69.
- Sun Weiguo, Li Haoran. MD5 Algorithm in Data Security and Safety of the Application[J]. Microcomputer Applications, 2010, 31(10): 66-69.
- [11] 李学武, 陈丹. 数据库访问离线处理分析设计[J]. 网络与信息, 2010, 24(4): 34.
- Li Xuewu, Chen Dan. A Database Access Off-Line Processing Analysis and Design[J]. Network and Information, 2010, 24(4): 34.
- [12] 杨华. 智能客户端离线处理的分析与实现[J]. 计算机与信息技术, 2006(12): 10-12.
- Yang Hua. Smart Client Off-Line Processing Analysis and Realization[J]. Computer and Information Technology, 2006(12): 10-12.

(责任编辑: 邓彬)