

# 基于二分图最优匹配算法的毕业论文选题系统<sup>①</sup>

## Choosing Topics System of Undergraduates' Theses Based on Bipartite Graph Optimal - matching Algorithm

杨胜超 张瑞军 (武汉科技大学管理学院 湖北 武汉 430081)

**摘要:** 本文在介绍毕业论文选题系统的系统用例、功能模块和流程图的基础上,针对学生选题不均衡这一突出问题,引入了二分图最优匹配的经典算法—KM 算法,该算法能够根据学生的题目预选、自命题、未定题等多种情况,完成题目与学生的智能匹配,使最终题目的整体满意度最高,从而提高学生的毕业论文选题质量。该系统在武汉科技大学管理学院 04 级毕业论文选题中实施效果良好。

**关键词:** 二分图 毕业论文选题 最优匹配 KM 算法

### 1 引言

在毕业论文选题系统中,一个学生只能选择一个题目作为自己的最终论文题目;同样,一个题目也只能分配给一个学生。如果最终题目由学生自己确定,那么就会出现这样的情况:先选的学生具有更大的选择余地,后选的学生由于不能再选已经选定的题目,所以其可选择的题目会越来越少,这对很多学生来说是很不公平的。如果学生选择自己的志愿,而最终题目由老师来定,这不但加大了老师的工作量,而且还是不能保证每位同学的公平性。如果采用计算机智能辅助选题,设计最优匹配算法实现学生与题目的整体最优匹配,无疑将大大提高选题的效率。

一些学者曾对题目的智能化匹配作过比较深入的研究,如汤颖采用模糊匹配技术进行学生-题目的自动匹配<sup>[1]</sup>;潘志方将题目与学生的匹配抽象为二分图的匹配,并采用改进的 Ford - Fulkenson 算法实现了题目与学生的自动匹配<sup>[2]</sup>。但是,他们只是考虑了题目与学生的最大匹配数,并没有同时考虑学生整体满意度最优的情况。

本文将题目与学生的匹配抽象为带权二分图的匹配(如图 1 所示)。学生可以预选多个题目,并根据自己对题目的满意度由高到低排序,这些满意度就成为

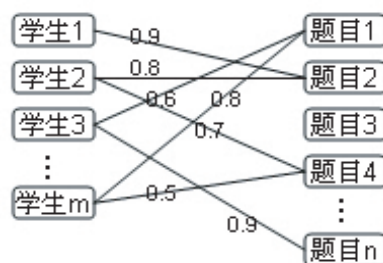


图1 题目与学生的带权二分图匹配算法

二分图中的权值,然后通过最优匹配算法来确定一种匹配方案,使得学生的整体满意度最高。

## 2 系统功能模块

### 2.1 系统用例图

为了使系统能快速准确地开发完成,保证系统功能的完整性和可行性。经过比较深入的系统分析后,确定了整个系统可分为四个大的管理模块:个人管理、题目信息、用户信息和留言管理,但在不同权限的用户角色下,提供不同的管理功能。并使用 UML 统一建模语言给出整个系统如图 2 所示的用例图,图中的每一个叶子用例分别对应于系统控制菜单里的一个功能选项。

<sup>①</sup> 基金项目:武汉科技大学博士科研启动基金项目(060105),武汉科技大学教学研究项目(2006036)

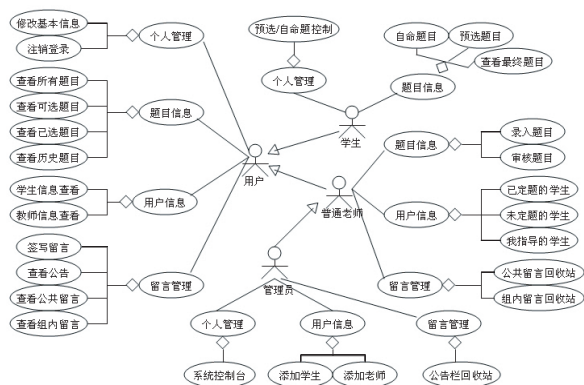


图 2 选题系统用例图

从用户的角度来看,整个系统的用户角色分为三类:管理员、普通老师和学生。首先,对于三种角色所共有的功能,抽象出一个共同的用户角色,学生和普通老师都由该共同角色泛化而来,然后再根据不同的角色提供各自特有的功能。其次,管理员角色是由普通老师角色泛化而来,其功能是在普通老师的基础之上加上自己更大权限的管理功能。下面将分别介绍各用户角色所具有的功能。

### 2.2 管理员角色的功能模块

(1) 个人管理模块 在该模块里,最重要的是为管理员集成了系统的核心功能——系统控制台。在这里,管理员可以控制选题年级、学生自命题数、老师的可带人数和可审核题目数、修改满意度权值、启动智能匹配算法以及整理数据库等。

(2) 题目信息模块 管理员除了具有公共的题目查看权限以外,还具有录入题目、管理自己的题目和审核学生自命题的功能。

(3) 用户信息模块 添加老师和学生、删除老师和学生、初始化所有用户的密码(为了保密起见,所有密码都是经过不可逆算法 MD5 函数加密后存储的)、查看已经定题、还未定题和自己指导的学生。

(4) 留言管理模块 查看和管理所有的留言:包括公告栏、公共留言和组内留言。

### 2.3 普通老师角色的功能模块

(1) 个人管理模块 修改自己的密码、联系方式和研究方向。

(2) 题目信息模块 和管理员的权限一样。

(3) 用户信息模块 除了不能管理老师和学生以外,其它功能和管理员一样。

(4) 留言管理模块 签署留言、查看留言、管理公共留言和组内留言。

### 2.4 学生角色的功能模块

(1) 个人管理模块 修改自己的登录密码和联系方式,还可以选择自己是自命题还是预选老师录入的题目。如果选择自命题,那么预选功能将被系统锁定;如果选择预选题目,同样自命题功能将被锁定。

(2) 题目信息模块 录入自己的题目或者预选老师录入的题目、查看自己的最终题目,此外还具有公共用户角色的所有功能:查看所有题目、可选题目、已选题目和历史题目。

(3) 用户信息模块 查看系里所有老师和学生的基本信息。

(4) 留言管理模块 签署公共留言和组内留言,查看公告栏、公共留言和组内留言,并能管理自己的留言内容。

## 3 系统流程图

基于二分图最优匹配算法的选题系统,充分考虑了学生确定自己论文题目的自由性:学生可以自主命题由老师来审核,如果审核通过则可作为自己的最终论文题目,如果未通过审核还可以反过来参加预选或者再次自主命题(有最大自主命题数限制)。同时还优化了题目预选的匹配:通过管理员启动智能匹配算法,确定出学生与题目的最优匹配方案,这样便大大减轻了老师的工作量,提高了选题的效率。最后,如果通过以上两个步骤还有学生没有定题,就只有通过老师手动确定学生的最终题目。整个系统的流程图如图 3 所示:

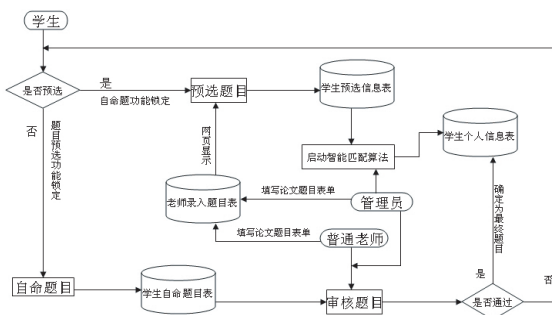


图 3 选题系统的工作流程图

系统运行的示例性界面如图 4 所示:



图 4 系统运行的界面

## 4 题目与学生智能匹配的关键算法

### 4.1 二分图及相关概念

系统在设计时,为了提高学生对题目的整体满意度,采用了一种基于二分图的问题与学生智能匹配算法,以下先对该算法所涉及的概念进行介绍。

**二分图:**如果图  $G = (V, E)$  的顶点集合  $V$  可分为两个集合  $X, Y$ , 且满足  $X \cup Y = V, X \cap Y = \Phi$ , 则  $G$  称为二分图(或二部图或偶图);图  $G$  的边集用  $E(G)$  表示,点集用  $V(G)$  表示。

**增广路:**若  $M$  是二分图  $G = (V, E)$  的一个匹配。设从图  $G$  中的一个顶点到另一个顶点存在一条道路,这条道路是由属于  $M$  的边和不属于  $M$  的边交替出现组成的,则称这条道路为增广路(或称交互树或交错树)。如果增广路的头尾两个顶点不属于  $M$ ,则称这条增广路为可增广道路。

**最优匹配:**有二分图  $G = (X, Y, E)$ , 其中  $|X| = |Y| =$  匹配数,  $E$  中每条边  $(X_i, Y_j)$  有权  $W_{ij} > = 0$ , 若能找到一个匹配  $M(|M| =$  匹配数), 满足所有匹配的边权和最大(或最小), 则称  $M$  为  $G$  的一个最优匹配(或称最大权匹配)<sup>[3]</sup>。

**完备匹配:**对于二分图  $G = (X, Y, E)$ ,  $M$  是  $G$  中的一个匹配, 如果  $G$  中每个顶点都是  $M$  中的一个匹配顶点, 则称  $M$  为  $G$  的完备匹配(或称完全匹配、完美匹配)<sup>[4]</sup>。

### 4.2 KM 算法

KM 算法是 Kuhn 和 Munkras 分别于 1955 年和 1957 年独立提出来的, 这是求解二分图最优匹配的经

典算法<sup>[5]</sup>。该算法是通过给每个顶点一个顶标  $(A[i], B[j])$  来把求最大权匹配的问题转化为不断地寻找增广道路以使二分图的匹配数达到最大的完备匹配问题。KM 算法的正确性基于以下定理:

**定理** 若由二分图中所有满足  $A[i] + B[j] = W_{ij}$  的边  $(X_i, Y_j)$  构成的子图(称作相等子图)有完备匹配, 那么这个完备匹配就是二分图的最优匹配<sup>[5]</sup>。

**KM 算法的初始化:**

令  $G(X, Y, E)$  是一个二分图, 其中  $X = \{x_1, x_2, \dots, x_n\}, Y = \{y_1, y_2, \dots, y_n\}$ 。设顶点  $X_i$  的顶标为  $A[i]$ , 顶点  $Y_j$  的顶标为  $B[j]$ , 顶点  $X_i$  与  $Y_j$  之间的边权为  $W_{ij}$ 。要求在算法执行过程中的任一时刻, 对于任何一条边  $(X_i, Y_j)$ , 使得  $A[i] + B[j] > = W_{ij}$  始终成立, 可以令  $A[i]$  为所有与顶点  $X_i$  关联的边的最大权, 而  $B[j] = 0$ 。数组  $match[n]$  表示  $Y_i$  与  $X_{match[i]}$  匹配。

KM 算法的关键在于, 不断地寻找二分图中的可增广道路。如果找到了一条可增广道路, 就可以将属于和不属于相等子图的边取反, 从而相等子图里就会增加一条边, 一直到所有的顶点都进入相等子图为止。本文采用的是调用递归函数的方式, 来寻找二分图的可增广道路。KM 算法的伪代码描述如下:

```

For i = 0 to n - 1
    将 X 和 Y 所有的顶点初始化为未扫描
    If path (i) 返回值为真 then
        进行下一值的循环
    Else
        令 dx = min{ A[i] + B[j] - Wij | Xi 在交错树中, Yj 不在交错树中 }
        For k = 0 to n - 1
            If Xk 节点已被扫描过 then
                A[k] 减去 dx
            End if
            If Yk 节点已被扫描过 then
                B[k] 加上 dx
            End if
        End for
    End if
End for

//////// 以下是递归函数 path(int i) //////////
Function Boolean path (int i)
    
```

```

标记  $X_i$  节点为扫描过的
For  $j = 0$  to  $n - 1$ 
  If  $Y_j$  节点未扫描且  $A[i] + B[j]$  等于  $W_{ij}$  then
    标记  $Y_j$  节点为扫描过的
    If 节点  $Y_j$  还未被匹配 或者调用函数 path
      ( $match[j]$ ) 的返回值为真 then
        将  $i$  赋值给  $match[j]$ 
        返回真
      End if
    End if
  End for
返回假
End function

```

在算法运行过程中,如果对于某个节点  $X_i$  找不到一条从它出发的可增广道路,但其实已经形成了一条交错树,我们就把交错树中  $X$  顶点的顶标全都减小某个值  $dx$ ,  $Y$  顶点的顶标全都增加同一个值  $dx$ 。当  $dx = \min\{A[i] + B[j] - W_{ij} \mid X_i \text{ 在交错树中}, Y_j \text{ 不在交错树中}\}$  时,将至少有一边会进入相等子图,从而说明了算法必定在有限步内结束。该算法的时间复杂度为:  $O(n^3)$ 。

#### 4.3 KM 算法在选题系统中的应用

KM 算法可以很好地解决在选题系统中,题目与学生最优匹配的问题。在匹配过程中,令学生集合为  $X = \{X_1, X_2, \dots, X_m\}$ , 题目集合为  $Y = \{Y_1, Y_2, \dots, Y_n\}$ , 学生对自己预选题目的满意度为二维矩阵  $W[m][n]$ , 其它的题目规定其权值为 0。系统在初始化时要求: 学生最多只能预选 5 个题目, 并按照自己对题目的满意度由高到低排序, 其满意度分别为 0.9、0.8、0.7、0.6 和 0.5, 该满意度可以由管理员以后根据需要进行修改, 其取值范围是 0 到 1 之间。

数组  $match[n]$  表示题目  $Y_i$  被分配给学生  $X_{match[i]}$ 。

在使用 KM 算法时, 有一个条件: 要求集合  $X$  的顶点个数必须等于集合  $Y$  的个数(实际上, 经分析后发现, 对于初始化顶标为 0 的集合的顶点个数大于另一个集合的顶点个数时, 该算法同样适用)。但当应用在选题系统中时, 由于无法预先确定是学生人数多还是题目的个数多, 所以在使用 KM 算法进行匹配运算时规定: 如果参加预选题目的学生人数  $m$  大于被选题目

数  $n$ , 就添加一些空题目节点使得  $m = n$ , 与其相关的权值都为 0; 同样, 如果  $m < n$ , 就添加一些空学生节点使得  $m = n$ , 与其相关的权值也都为 0。

在经过智能算法匹配后, 去掉那些添加的空节点的匹配和权值为 0 的匹配(如果有的话), 从而数组  $match[n]$  中剩下的就是学生与题目的最优匹配方案。

表 1 给出的测试数据来自武汉科技大学管理学院 2004 级管理科学与工程、工商管理系和市场营销系的真实选题情况:(第一志愿满意度为:0.9, 第二志愿满意度为:0.8; 为了简单起见, 每个学生最多只有两个选题志愿)

表 1 测试数据

	管理科学与工程系(学生: 111 人, 题目: 116)		工商管理系(学生: 256 人, 题目: 244)		市场营销系(学生: 99 人, 题目: 69)	
	匹配数	整体满意度(%)	匹配数	整体满意度(%)	匹配数	整体满意度(%)
最优匹配方案	108	86.757	227	78.242	64	56.869
最大匹配方案	108	85.675	227	76.992	64	55.757
手工分配方案	98	79.459	173	60.820	41	37.273

下面以工商管理系为例对上表数据进行说明。刚开始时, 采用手工分配的方式, 使得 256 个学生中只有 173 个学生分得了题目, 其总体满意度仅为 60.820%; 如果采用最大匹配算法进行分配, 可以使匹配数达到最大, 即 227 个学生分得题目, 而其总体满意度上升为 76.992%; 最后采用最优匹配算法进行分配, 使总体满意度达到了最高水平 78.242%, 但同时匹配数也达到了最大, 即 227。虽然在上表所列出的三个实例中, 都是在满意度最高的时候匹配数也达到最大, 这当然是最理想的情况。但这只是针对该系统而言, 同时达到最优的概率比较大, 因为这主要受权值等因素的影响。

有一点需要说明的是, 本系统采用 KM 算法只是找到了整体最优匹配而不一定是最大数匹配。如果在

(下转第 34 页)

(上接第 17 页)

整体最优的情况下匹配数与最大匹配数相差得太大的话,那么整体最优的方案就显得不太可取。所以,最好的情况就是同时考虑整体最优匹配和最大匹配,但这要求设计一个能对满意度或匹配数进行排序的算法,来同时控制两者的大小。

## 5 结束语

按照以上描述的设计思路和算法,采用 Asp 技术 + Access 后台数据库实现了毕业论文选题系统。在学生预选题目的智能匹配方面采用了 KM 最优匹配算法,保证了学生的整体满意度最大,同时还大大减轻了老师手工分配题目的工作量和难度。系统除了提供学生预选题目的功能以外还提供了学生自命题目由老师进行审核的功能,在预选中没有被分配题目的学生,可以由老师再次组织预选题目,或者由老师手工分配题目。该系统在武汉科技大学管理学院 2004 级学生的毕业论文选题中进行了应用,取得了良好的效果。

## 参考文献

- 1 汤颖. 毕业设计立项与选题管理及其支持系统. 合肥工业大学学报(自然科学版),2006,29(5):613-616.
- 2 潘志方. 一种改进的 Ford - Fulkenson 算法在选题系统中的应用研究. 计算机应用与软件,2007,24(9):120-121.
- 3 彭宇新, Ngo Chong - Wah, 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,2004,7:1135-1139.
- 4 俞万禧. 基于完全二分图矩阵的  $\Delta(G)$  - 边着色求解完全图  $K_{4n}$  的完备匹配. 井冈山学院学报(自然科学版),2007,28(4):50-52.
- 5 孙惠泉. 图论及其应用. 北京:科学出版社,2004:89.