

# 选区划分模型及快速分类算法<sup>\*</sup>

何 朗<sup>1</sup>, 张好春<sup>1</sup>, 於 晓<sup>2</sup>

(1. 武汉理工大学理学院, 湖北武汉 430070)

(2. 武汉理工大学材料与工程学院, 湖北武汉 430070)

**摘要:** 本文主要研究选区重新划分问题. 利用公平性原则和广度优先搜索方法, 获得一个选区划分模型和一个求解该模型的定向目标算法. 理论分析和数值计算结果表明模型设计合理, 算法切实可行.

**关键词:** 选区划分; 选举; 定向目标算法; 席位选举曲线

**MR(2000)主题分类号:** 65K05; 68Q05

**中图分类号:** O229

**文献标识码:** A

**文章编号:** 0255-7797(2008)04-0469-04

## 1 引言

在美国国会议员选举历史中,“杰利蝾螈”(Gerrymander)<sup>[1]</sup>一词起源于 1812 年,时任美国马萨诸塞州州长的 Elbridge Gerry 为了胜选,蓄意将选区重划,把敌对党派的候选人票源集中在少数选区,以保护自己党派的候选人不成比例地当选. Gerrymander 表示为了政治目的,不公正地划分选区的方式. Gerrymander 的操作手段有两种:集中选票和分散选票. 集中选票指的是将某一党派的支持者划分到该党已经稳获选举的区域,造成该党的选票浪费;分散选票指的是将支持某一党派的选民分割到不同的选区,使选区无法集中足够的选票选出他们所支持的候选人. 因此,为了避免“杰利蝾螈”现象<sup>[2]</sup>,保证选举的公平性,美国通过立法明确要求每隔 10 年做一次人口普查,登记选民民意,重新划分选区,并且要求选区紧密相连. Lijphart<sup>[3]</sup>提出了 16 项标准用于规范选区划分原则. 本文的目的在于设计了一个简单可行的选区划分模型以保证选举公平性,并且通过算法自动实现选区的重划.

## 2 重新划分的公平选举模型

假设县是最小的行政单位,划分选区就是将某个或某些相邻的县划为一个选区. 选区划分最重要的是保证公平性,公平性对选民来说是一人一票,对党派来说是支持者多的党派获得选举席位多.

### 2.1 人口均等性

为了保证一人一票,票票等值,各小选区需要保证人口的均等性.

<sup>\*</sup> 收稿日期: 2007-12-24

接收日期: 2008-04-19

作者简介: 何 朗(1974-),男,湖北武汉,讲师,主要从事演化计算研究. E-mail: helang@whut.edu.cn

假设  $p_i$  表示  $i$  县人口数,  $P_l$  表示第  $l$  选区人口总数, 因此

$$P_l = \sum_{i=1}^N p_i \delta_{li}$$

这里,  $\delta_{li} = \begin{cases} 1 & s_i = l \\ 0 & s_i \neq l \end{cases}$ , 当  $s_i = l$  时, 表示  $i$  县被划入第  $l$  选区. 假设总共有  $m$  个选区, 所有选区的总人口数为

$$P = \sum_{l=1}^m P_l$$

各选区的平均人口数为

$$P_a = P/m$$

如果要绝对保证人口均等性, 需要做到对任意选区  $P_l = P_a$ , 但是考虑到现实问题, 差额难以避免, 实际操作中需要容许一定的偏差, 定义人口偏差函数如下

$$A_p = \sum_{l=1}^m \left| 1 - \frac{P_l}{P_a} \right| \quad (1)$$

并且取特定的常数  $A$ , 当  $A_p < A$  时, 认为人口均等性得到满足.

## 2.2 连通性

在选区的划分中, 需保证各个选区的连通性, 即每个选区应该由相连的县组合而成, 但是要避免复连通的情况, 即不能出现某个选区完全包围另外一个选区的情况.

## 2.3 紧凑性

选区紧凑性是防止“杰利蝾螈”的重要手段, 因为选区紧凑性就是要求选区形状满足一定的规则, 不能出现奇形怪状, 比如说选区出现细长形状. 本文的紧凑性函数定义如下: 假设  $r$  等于区域  $\Omega$  中某点到中心点的距离, 则其转动惯量为  $\iint_{\Omega} r^2 d\Omega$ , 这一值描述了周围的点向中心点靠拢的程度. 假设  $D$  表示区域的面积, 定义紧凑性函数

$$M = \frac{D}{\sqrt{4\pi \iint_{\Omega} r^2 d\Omega}} \quad (2)$$

这样, 如果  $\Omega$  是一个圆形区域的话,  $M$  等于 1, 这正好是理想的选区形状.

## 2.4 选区划分

在选区公平性的划分中, 以上已经通过选区的人口均等性、连通性和紧凑性给出了限制, 为了体现选区划分党派的公平性, 在各个选区中, 机制需要给定一个选票与当选席位的公平关系, 以保障不会出现选票有意浪费的情况. 假设  $V$  表示某个政党获得的选票率,  $f(V)$  表示该党获得的议员席位率, 在单一选区机制中, 最公平的关系是  $f(V) = V$ . 而在多选区中, 情况就复杂些.

首先, 公平的  $f(V)$  应该满足如下两个性质:

(1)  $f(V) \in [0, 1]$ , 且经过三个点  $(0, 0)$ ,  $(0.5, 0.5)$ ,  $(1, 1)$ , 意思很明确, 当某个党派获得 0 选票是理应不获得议员席位, 当获得全部选票的时候应该获得全部席位, 当他与其他党派势均力敌时, 席位数应当相等.

(2)  $f(V)$  是递增函数, 表示某个党派获得更多选票率的时候应该获得更多的议员席位.

其次, 我们采取这样的策略, 在某个选区中, 如果某个党派拥有绝大多数的选票率, 我们使得他能够获得的议员席位率比选票率稍低, 这样的做法可以增加小党派的竞争性, 因此,

这样做的目的就可以使得划分选区的行为趋向于使党派间的选票率均等,从而在一定程度上避免集中或分散选票. 如果不这样做容易造成选票集中或分散的恶化,而且还会导致一些其他的社会问题<sup>[4]</sup>.

因此,考虑到以上的分析,引进 bilogit 函数形式

$$f(V) = \{1 + e^{-\lambda - b(\frac{V}{1-V})}\}^{-1} \quad (3)$$

$\lambda$  表示党派偏向参数,  $\lambda=0$  表示无党派偏向, 即  $f(V)=V$ ;  $\lambda>0$  表示偏向大党派, 即选票率高的党派可以获得高于选票率的席位率;  $\lambda<0$  表示偏向小党派, 即选票率低的党派可以获得高于选票率的席位率.

### 3 定向目标算法

通过以上的模型,我们定义了公平划分选区的一些准则,下面的任务是如何在满足公平性模型的情况下,通过算法实现选区的划分,即不同县的组合来完成选区的划分. 我们的算法的策略是采取动态兼并,即从初始地区(比如某个县),在满足模型的条件搜索邻近县,当达到满足选区的条件时完成该选区的划分,依次类推直至所有地区划分为止.

考虑 2 个党派的竞选情况,首先假设  $V_A, V_B$  分别代表党派 A 和 B 的选票率,  $S_A, S_B$  分别代表他们获得的议员席位率,  $M$  代表紧凑性评价. 算法步骤如下:

**Step 1** 寻找州地图的最小凸多边形<sup>[5]</sup>, 由于各县各州的地图形状都不是规则的, 面积难以计算, 因此我们以他的最小凸多边形的面积代替;

**Step 2** 计算选区的人口平均值  $P_a$ , 选取最小凸多边形顶点处所在的县作为初始搜索点  $i$ , 并把该县作为第一个选区  $l$  所在县;

**Step 3** 标记初始选区的人口数  $P_l = p_i$ , 计算当前  $V_A, V_B, S_A, S_B$  及  $M$ , 同时搜索邻近县  $j$  将他加入该选区;

**Step 4** 判断县  $j$  是否满足模型条件, 如果满足模型方程(1)(2)(3), 将  $j$  加入选区  $l$ , 同时修正人口值  $P_l = p_i + p_j$ ; 如果不满足条件, 选取其他与  $i$  相邻的点;

**Step 5** 当所有县都搜索完成之后, 仍然无法完成该选区的划分工作, 则转入 Step 2 重新寻找初始点进行, 直到该选区满足条件为止;

**Step 6** 将选区划分好的县排除在外, 转 Step 1, 继续下一个选区的划分;

**Step 7** 若所有县都已划分, 算法停止.

本算法能够反映模式的效果. 算法在搜索过程中始终搜索的是临近点, 因此选区的连通性得到保证; 当人口数达到要求时, 选区划分终止, 保证了人口的均等性, 而  $V_A, V_B, S_A, S_B$  及  $M$  的计算是最重要的, 它们体现选区公平性. 当一个县加入该选区时, 是否破坏了公平原则, 就通过这些值给予评价. 算法同时存在着一定的风险性, 当选区划分完成时, 出现孤立县, 当考虑初始搜索点为多个时, 可能会出现争夺一个县的情况, 即当前的选区都要求纳入某个县进入才能完成选区划分, 当出现这两种情况时, 算法需重新进行. 算法的核心步骤是搜索临近点, 它是宽度优

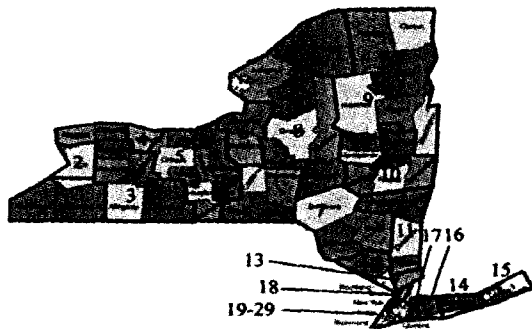


图 1 划分结果图

先搜索,因此,复杂性为  $O(n^2)$ .

## 4 数值实验

以 2004 年美国纽约州的选举结果作为初步数据. 根据美国宪法,美国纽约州的议员席位为 29 席. 在 2004 年,划分的选区有 29 个. 根据定向目标算法,结果如图 1 所示. 其他一些信息,如地区人口,面积,紧凑性见表 1(只显示前 3 个区). 从表 1 中,可以发现,形状各区各有不同,但它们的形状都是可行的,紧凑性评价值与理想图形较接近. 因此,本文的紧凑性评价值可信.

表 1 选区选民数、面积、紧凑度信息表

选 区	选民数	面 积	紧密度
1	417348	2785	1.0209
2	41000	631	0.9885
3	392752	5056	1.1989

## 5 结 论

本文针对民主选举问题,设立了一个涵盖广泛的定义和标准,将政治问题转变为数学模型,并设计了一个多项式时间算法——定向目标算法来求解该模型. 通过对美国纽约州选区划分问题的研究表明,本文所设计的模型是合理的,算法有效、适用.

### 参考文献:

- [1] Alan G.. The American Political Process[M]. New York: Ashgate Publishing Company, 1997.
- [2] Gary K.. Representation Through Legislative Redistricting—A Stochastic Model [J]. American Journal of Political Science, 1989, 33(4):787-824.
- [3] Lijphart A.. Democracies: Patterns of Majoritarian and Consensus Government in Thirty-One Countries[M]. New Haven: Yale University Press, 1984.
- [4] William D N. The Political Business Cycle [J]. Review of Economic Studies, 1975, 65(2): 169-190.
- [5] Bondy J A., Murty S R. Graph Theory with Applications[M]. London: The Macmillan Press, 1976.

## A FAST SORTING ALGORITHM TO REDISTRICTING ELECTION MODEL

HE Lang<sup>1</sup>, ZHANG Hao-chun<sup>1</sup>, YU Xiao<sup>2</sup>

(1. School of Sciences, Wuhan University of Technology, Wuhan 430070, China)

(2. School of Materials Science and Engineering, Wuhan University of Technology, Wuhan 430070, China)

**Abstract:** In This paper, we study the problem of redistricting election. Then a redistricting election model is proposed and an aim directed algorithm is established to cope with the model. Theoretical analysis and numerical experiment indicate that the model is reasonable and the algorithm is quite feasible.

**Keywords:** Gerrymander; Election; aim directed algorithm; voters-seats curve

**2000 MR Subject Classification:** 65K05; 68Q05