

一类最优交通小区划分问题的一阶邻接约束建模方法

王霖青¹, 唐加福², 章宇¹, 吴影辉¹

(1. 东北大学 信息科学与工程学院, 沈阳 110004; 2. 东北财经大学 管理科学与工程学院, 辽宁 大连 116026)

摘要: 提出一种使用邻接矩阵保证最优交通小区划分一阶邻接约束的整数规划建模方法. 从求解复杂度和质量两个角度, 比较并分析了该邻接约束建模方法与其他3种方法对问题求解效率的影响. 设计了聚合式层次聚类启发算法以求解所提出的模型. 针对较大规模算例, 将所提出的建模方法与其他3种邻接约束建模方法的结果进行了对比与分析. 结果表明, 基于邻接矩阵表示的建模方法能在允许时间内求得满意解, 较其他3种方法更适合大规模问题.

关键词: 最优交通小区划分; 邻接约束; 整数规划; 启发式算法

中图分类号: U491

文献标志码: A

A model of first-order contiguity constraint on traffic analysis zone delineation problem

WANG Lin-qing¹, TANG Jia-fu², ZHANG Yu¹, WU Ying-hui¹

(1. College of Information Science and Engineering, Northeastern University, Shenyang 110004, China; 2. College of Management Science and Engineering, Dongbei University of Finance and Economics, Dalian 116026, China. Correspondent: WANG Lin-qing, E-mail: wanglinqing.chn@gmail.com)

Abstract: An explicit first-order contiguity constraint, adjacent matrix presentation, is proposed as a general integer programming model approach to the traffic analysis zone delineation problem. Model size and solution times are compared between the proposed contiguity constraint and another three. An agglomerative hierarchical clustering based heuristic algorithm is designed to solve the proposed model. For a large-sized case, the results of the proposed model and the other three ones are compared and analyzed. The results show that the proposed model is more suitable for solving a larger problem with approximate solutions in fairly reasonable time.

Keywords: traffic analysis zone delineation problem; contiguity constraint; integer programming; heuristic algorithm

0 引言

从最优化决策的角度来看, TAZ (Traffic analysis zone) 划分问题是指在满足某个目标最优的情况下, 如何将整体研究区域聚类划分为若干部分区域的问题. TAZ 问题属于一类 NP 难问题^[1], 已有研究大都集中于设计启发式算法^[2-3]. 这些启发式算法往往遵循一个通用的基本过程, 即初始化一个非最优的但满足空间连续的解, 在迭代的搜索过程中通过解空间单元的交流、重组得到新的可行解, 通过评估判定是否继续搜索. 由于缺乏对数学模型的精确求解分析, 一方面无法评价这些启发式算法的求解质量, 另一方面也造成了 TAZ 问题的算法研究方向的局限性. 这主要是因为邻接约束很难通过数学建模来表达.

目前, 只有很少文献讨论分析了 TAZ 问题以及

类似问题的整数规划建模, 如: Zoltners 等^[4]针对销售区域问题设计了“层次邻接树”(TRE)方案, 以保证其整数规划模型解的空间连续邻接; Cova 等^[5]通过有序路径方法 (ORP) 表示相邻关系, 给出了站点设计问题的混合整数规划模型; Shirabe^[1]基于嵌入网络流模型 (FLO) 提出了另一种保证邻接约束的整数规划建模方法; Duque 等^[6]综合比较并分析了上述3种邻接约束表示方法的整数规划模型的精确求解难度. 虽然上述几种邻接约束建模方案对于精确求解类似于 TAZ 问题的最优分区设计问题而言是一项重大的理论突破, 但在工程实现上存在一定局限: 其一, 在允许时间内仅能求解约 50 个聚类项的小规模问题; 其二, 需设计另外的约束以保证分区形状的饱满.

鉴于此, 本文采用邻接矩阵表示 (ADM) 提出一

收稿日期: 2013-10-28; 修回日期: 2014-03-08.

基金项目: 国家自然科学基金创新研究群体项目(71021061).

作者简介: 王霖青(1985-), 女, 博士生, 从事智能交通与物流优化的研究; 唐加福(1965-), 男, 教授, 博士生导师, 从事运作管理与决策、优化理论与技术、数据挖掘与商务智能等研究.

种TAZ问题的一阶邻接约束建模方法,建立了最小化分区系统差异的TAZ问题的整数规划模型,并与TRE表示、ORP表示以及FLO表示邻接约束的建模方法^[6],从模型规模以及求解效率等方面进行了比较分析.针对ADM模型无解的情形,设计了基于聚合式聚类技术的启发式方法以重构求解空间,并使用较大规模算例对4种模型的求解结果进行了分析,表明了本文方法在大规模问题上的有效性.

1 问题描述与数学模型

本文采用规则多边形网格来构建基本地理单元(BGU).TAZ问题考虑如下目标和约束:1)目标:最小化分区系统的差异;2)要决策的TAZ个数已知;3)每个BGU属于且仅属于一个小区;4)每个小区的BGU组成是连续且邻接的.小区 Z_k 的内部属性差异为

$\sum_{i,j \in Z_k | i < j} d_{ij}$.其中: i, j 为BGU的索引; d_{ij} 为度量任意两个BGU某属性值差异的变量.

引入如下符号: N 为BGU总数; K 为分区个数; U 为全集; $c_{ij} = 1$ 表示两个BGU相邻.基于 K 中位问题模型,引入3个决策变量: $x_{ij} = 1$ 表示 j 属于以 i 为中心的分区; $y_i = 1$ 表示 i 被选为某分区的中心; $d_{ij} = 1$ 表示 i 与 j 同属于一个分区.于是TAZ问题可描述为如下整数规划模型——ADM模型:

$$\min \sum_{i \in U} \sum_{j \in U | i < j} d_{ij} t_{ij}. \quad (1)$$

$$\text{s.t.} \sum_{i \in U} y_i = K; \quad (2)$$

$$\sum_{i \in U} x_{ij} = 1, \forall j \in U; \quad (3)$$

$$\sum_{i \in U} \sum_{j \in U} x_{ij} = N; \quad (4)$$

$$y_i - x_{ij} \geq 0, \forall i, j \in U, i \neq j; \quad (5)$$

$$y_i = x_{ij}, \forall i, j \in U, i = j; \quad (6)$$

$$x_{ij} \leq c_{ij}, \forall i, j \in U; \quad (7)$$

$$x_{ij} \leq t_{ij}, \forall i, j \in U; \quad (8)$$

$$t_{ij} + t_{im} - t_{jm} \leq 1, \quad (9)$$

$$\forall i, j, m \in U, i \neq j, m \neq j;$$

$$t_{ii} = 1, \forall i \in U; \quad (10)$$

$$t_{ij} = t_{ji}, \forall i, j \in U; \quad (11)$$

$$x_{ij} \in 0, 1, \forall i, j \in U; \quad (12)$$

$$y_i \in 0, 1, \forall i \in U; \quad (13)$$

$$t_{ij} \in 0, 1, \forall i, j \in U. \quad (14)$$

其中:目标(1)表示最小化分区系统的差异;式(2)表示中心BGU的个数等于分区个数;式(3)、(4)表示每个BGU对应且仅对应一个中心BGU,即属于且仅属

于一个分区;式(5)、(6)表示中心BGU只能属于自己所代表的分区;式(7)、(8)为ADM表示的邻接约束,即只有与中心BGU相邻的其他BGU才有可能属于该中心BGU对应的分区;式(9)表示如果 i 和 j 、 j 和 m 属于一个分区,则 i 和 m 也属于同一分区;式(10)表示对于每个BGU而言,总是与自身属于同一个分区;式(11)表示属于同一个分区是可逆的.

2 ADM模型与其他模型的计算复杂度比较

通常约束数量越多,模型规模越大,精确求解越复杂.表1给出了ADM模型(ADMM)、TRE模型(TREM)、ORP模型(ORPM)和FLO模型(FLOM)的约束数量理论计算表达式.

表1 4种模型的约束数量

模型	约束数量计算公式
ADMM	$N^3 + 5N^2 - 4N + 2$
TREM	$N^3 - N^2 + 11N + 1$
ORPM	$NK(N - K + 1) + N + K(1 + 0.5 \times (N^2 - N))$
FLOM	$K(1 + 0.5 \times (N^2 - N)) + 2NK + 9N$

由表1可见:在约束数量方面,ADMM、TREM都是最庞大的模型,且两者的约束数量都与 K 无关;其次是ORPM,与 K 为非线性关系;FLOM是约束规模最小的模型,与前三者不同的是,FLOM的约束个数与 N 、 K 都线性相关.

对采用Cplex运行10个小规模实验算例的求解情况进行分析,可得如下结论:1)当 N 变化时,4种模型各有利弊;2)在求解时间消耗方面ADMM有明显优势;3)后3种模型的最优数值解优于ADMM,且当 $K < N/\hat{C}$ (见第3节定理1)时ADMM无解.

综上,使用ADMM建模邻接约束的最优数值解要劣于后3种模型,但求解速度快.下文将设计启发式算法以解决ADMM的无解情况,并使用较大规模算例对4种模型的求解效率进行分析.

3 基于聚合式层次聚类的启发式算法

假设 K 未知,根据ADMM约束(2)、(3)和(7),可得如下引理.

引理1 对于ADMM的任意可行解,有

$$\sum \sum c_{ij} y_i \geq N. \quad (15)$$

证明 记ADMM任意可行解为 $X^* = [x_{ij}]_{N \times N}$, $Y^* = [y_i]_{1 \times N}$, K^* ,假设 $\sum \sum c_{ij} y_i < N$,则根据 $x_{ij} \leq c_{ij}$,有 $\sum \sum x_{ij} y_i < N$;同理根据 $x_{ij} \leq y_i$,有

$$\sum \sum x_{ij} < N. \quad (16)$$

显然,不等式(16)与 $\sum \sum x_{ij} = N$ 矛盾,因而假设不成立.□

由引理1可得如下定理.

定理1 对于ADMM的任意可行解,有

$$K \geq N/\hat{C}. \quad (17)$$

证明 同样,记ADMM的任意可行解为 X^* 、 Y^* 、 K^* ,根据引理1有 $\sum \sum c_{ij} y_i \geq N$;记 Y^* 中非零元为 $y_{m_1}, y_{m_2}, \dots, y_{m_K}$,则对于 $\forall i \in U$,因为 $\sum c_{ij} \leq \hat{C}$,所以 $\hat{C}(y_{m_1} + y_{m_2} + \dots + y_{m_K}) \geq N$,即 $K \geq N/\hat{C}$. \square

根据定理1,针对ADMM,设计基于聚合式层次聚类的启发式算法(AHC算法),其步骤如下.

Step 1: 判定 $K \geq N/\hat{C}$ 是否满足,若不满足则转Step 3;

Step 2: 求解ADMM,若无解则转Step 3,否则停止;

Step 3: 计算 d_{ij} ,根据计算结果选取 $\min[d_{ij}]$,并记为 i', j' ,转Step 4;

Step 4: 合并BGU i' 和BGU j' ,并重新计算 N' , d'_{ij}, c'_{ij} ,返回Step 1.

4 实验与结果分析

表2给出了在AHC算法框架下,结合Cplex运行100个较大规模算例的求解情况.其中后3者求解时间上限设置为600s.第2列中 N' 为AHC算法处理后 N 的变化,当该列值不为空时,4种模型的计算结果都是在 N' 条件下求解得到的.列Gap为ADMM最优解与最好解的差异比.一般而言,ADMM能在10min内求解约76%的问题,其中48%为最好解.求解质量方面,根据表2分别得到了ADMM与TREM(Gap-A.T.)、ORPM(Gap-A.O.)、FLOM(Gap-A.F.)中最好解的差异比,如图1所示.

当 $N > 50$ 时,TREM和FLOM基本无法给出解;ORPM虽然可以给出解,但质量远不如ADMM.虽然当 N 较小时,ADMM的最优解不及后3者,但随着 K/N 变大,ADMM最优解与后3者的差异也在逐渐减小.计算时间消耗方面,25个算例中ADMM最短用时为0.23s,最长用时约为2000s,平均用时少于400s.后3种模型均出现因内存不足而导致无法求解的情况.其中:TREM在10min内仅求解了8个算例;ORPM求解了17个算例,且只有3个算例的最优解优于ADM-OP;而FLOM也求解了较少的7个算例,仅有5个优于ADMM.

综合求解质量和时间消耗两个方面评价表2所示的算例结果可知:1)对于 $N < 50$ 的问题,10min内TREM能给出最好解,其次是FLOM和ADMM,ORPM的求解质量最差,同时其他3种模型的求解时间远大于ADMM;2)对于 $50 \leq N \leq 100$ 的问题,

表2 AHC算法求解4种模型结果

p	N-K-N'	ADM-OP	TRE-OP	ORP-OP	FLO-OP	Gap
1	36-6-25	1755	516.24	967.7	516.24	2.40
2	36-9-35	1192	325.03	1787.65	412.23	2.67
3	36-10	549	259.46	896.5	364.68	1.12
4	36-16	109	87.45	1276.16	123.13	0.25
5	36-18	77	63.97	119.99	79.69	0.20
6	49-7-31	2408	1069.11	1507.04	1374.14	1.25
7	49-13	1624	748.69	2865.33	∞	1.17
8	49-17	482	364.59	2900.42	∞	0.32
9	49-20	283	219.39	4037.74	∞	0.29
10	49-25	124	92.07	1307.35	∞	0.35
11	64-8-34	4520	2254.44	4356.89	2278.65	1.00
12	64-16	3423	∞	67761.94	∞	0.00
13	64-18	2050	∞	63661.83	∞	0.00
14	64-20	1365	∞	3773.12	∞	0.00
15	64-32	195	163.27	30117.18	∞	0.19
16	81-9-44	6936	6062.64	10737.66	∞	0.14
17	81-19-80	6406	∞	119549.8	∞	0.00
18	81-21	3990.97	∞	134443.4	∞	0.00
19	81-30	908	∞	∞	∞	0.00
20	81-41	268	∞	∞	∞	0.00
21	100-10-53	13306	∞	∞	∞	0.00
22	100-20-89	10879	∞	∞	∞	0.00
23	100-25	7400.58	∞	∞	∞	0.00
24	100-30	4381.56	∞	∞	∞	0.00
25	100-50	338.25	∞	∞	∞	0.00

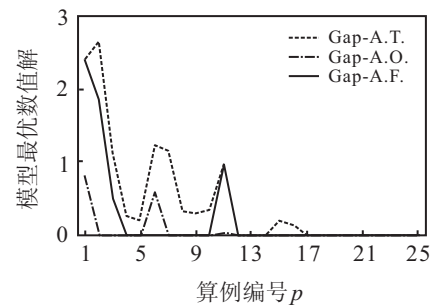


图1 ADM模型与其他3种模型的最优值Gap比较

ADMM能在短时间内求解全部算例,其他3种模型可求解的算例很少.

根据表2,当 $K/N = 0.5$,即平均两个BGU组成一个分区时,ADM-OP与TRE-OP的差异约为20%.参考上文对求解时间消耗的比较,结合模型分析可知,ADM-OP与其他3种模型差异是由相邻空间单元的搜索方式导致的.对于任意BGU,ADMM仅搜索与其一阶相邻的单元,而其他3种模型均采取多级相邻关系的方式传递搜索空间单元,因此后3种模型的最优解更优但时间消耗较大.为了分析不同邻接搜索方式对分区解形状的影响,图2给出了表2第 p_1 算例中TREM与ADMM的最优解图形表示.

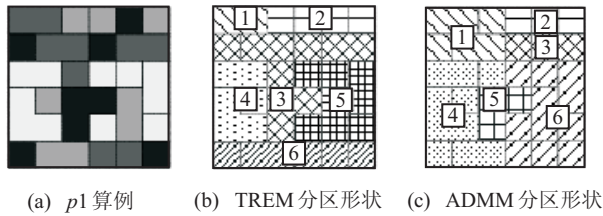


图2 TREM与ADMM关于算例 $p1$ 的最优解分区形状

图2(a)为算例 $p1$ 中经过AHC算法重构之后的25个BGU集合,网格颜色的深浅表示某度量属性值(如分区可达性、分区出行数量等TAZ问题指标)的大小.观察TREM(图2(b))以及ADMM(图2(c))分区最优解的形状可知:TREM中的第3个分区与第6个分区形状过于狭长;而ADMM产生的分区解形状较饱满,与实际交通规划中对交通小区的要求更相符.

定义如图3所示的形状压缩率 ρ 来定量描述分区形状的饱满程度.

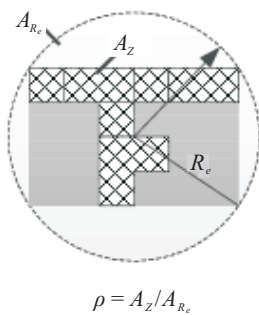


图3 形状压缩率示意

如图3所示, ρ 为分区面积 A_z 与分区当量半径 R_e 定义的圆面积 A_{R_e} 之比. ρ 越大,越接近于 $1/\pi$,表示分区越饱满.比较表2算例的各模型最优解的平均形状压缩率如图4所示.

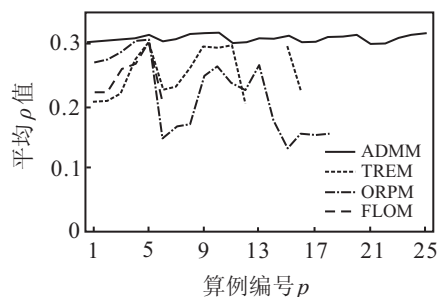


图4 4种模型最优解平均形状压缩率比较

如图4所示,ADMM分区解的形状具有较平稳的压缩率,而其他3种模型的波动较大,表明ADMM分区解的饱满程度较为平均.此外,根据 ρ 范围,4种模型中,ADMM的分区解的形状最为饱满,其他3种模型由于考虑多级的邻接关系容易产生极其狭窄的分区.

综上所述,从4种模型求解的数值质量、图形质量以及求解时间消耗3个方面看,ADMM可以在较短时间内产生形状较为饱满的分区解,因此当最优交通小区划分问题对优化目标的数值质量要求不是很严格但期望分区形状均匀且饱满的情况下,ADMM是最为合适的建模方法.

5 结论

本文提出了一类最优交通小区划分问题的一阶邻接约束,基于K中位模型,给出了最小化分区系统总差异的整数规划模型.通过比较邻接矩阵表示的一阶邻接约束与其他3种邻接约束,得到如下结论:

1) 4种模型在问题求解空间上有所不同,ADM模型的最优解是其他3种模型最优解的一个下界;

2) 所提出的ADM模型具有能在短时间内给出较好解的特点,而其他3种模型的解虽然质量更好但耗时巨大,当 $N > 50$ 时,在实际可允许时间范围内可能无法给出比ADM模型更好的解;

3) 由于仅搜索一阶相邻的单位空间构造交通小区,ADM模型给出的分区解具有形状均匀且饱满的特点.

综上所述,本文提出的基于邻接矩阵表示的一阶邻接约束建模方法适用于一类对优化目标的数值质量要求不是很严格但期望分区形状均匀且饱满的最优交通小区划分情况,能够实现高效求解的目的.

参考文献(References)

- [1] Shirabe T. A model of contiguity for spatial unit allocation[J]. *Geographical Analysis*, 2005, 37(1): 2-16.
- [2] Martin D, Cockings S, Harfoot A. Development of a geographical framework for census workplace data[J]. *J of the Royal Statistical Society: Series A*, 2013, 176(2): 585-602.
- [3] Huang K N, Liu X P, Li X, et al. An improved artificial immune system for seeking the Pareto front of land-use allocation problem in large areas[J]. *Int J of Geographical Information Science*, 2012, 27(5): 922-946.
- [4] Zoltners A A, Sinha P. Sales territory alignment: A review and model[J]. *Management Science*, 1983, 29(11): 1237-1256.
- [5] Cova T J, Church R L. Contiguity constraints for singlesite search problems[J]. *Geographical Analysis*, 2000, 32(4): 306-329.
- [6] Duque J C, Church R L, Middleton R S. The p problem[J]. *Geographical Analysis*, 2011, 43(1): 104-126.