第 19 卷 建模专辑 工 程 数 学 学 报 Vol. 19 Supp.
2002 年 02 月 JOURNAL OF ENGINEERING MATHEMATICS Feb. 2002

文章编号:1005-3085(2002)05-0107-06

建模需要思想,也需要数学训练和手上功夫
——B 题综合评述

刘宝光
(北京理工大学,北京 100081)

摘 要:本文对于 2001 年全国大学生数学建模竞赛 B 题的解答,从模型框架、模型建立和模型求解等三个方面给出评述。

关键词:公交车调度;数学规划模型;目标规划
分类号:AMS(2000) 90C08 中图分类号:Tb114.1 文献标识码:A

1 模型框架

正如题目的标题所示,这是一个"公交车调度问题"。题目给定了限制条件和目标,要求在一定统计数据资料的基础上,构建理论模型并据以实现调度。应当指出,构建"明确、完整的数学模型",是本题解案要求的一个本质性的方面。事实上,每一天,在全国各地,都有无数的车队调度在作这类问题。他们遵守各自的约束,追求各自的目标,用他们习惯了的方法调度车辆,指挥着全国各地公交车的运行。他们不见得都有多高深的数学修养,所给出的调度方案也不见得都是合理的优良的。本题的意义,并非要参赛者习作普通的车队调度,而是要求对这一问题用数学方法作更深一层次的探讨。这次有占相当比例的答卷没有完成明确、完整的模型的建立,只从数据出发,凭某种直观方法给出一个可行的调度方案。应当说这是不符合题目要求的,因而都不在得奖者之列。

考察题意最直接的想法会是建立数学规划模型。在给出了数学模型的答卷中,恐怕有七成以上是用的各种不同的数学规划模型。模型的变量,既然题目要求设计调度方案,最自然的就是发车时刻系列

$$T_1, T_2, \cdots, T_m$$

但这一变量,其维数 m 很高,而且是不定的。这会为使用某些成熟的优化数值方法求解带来麻烦。一种适当的简化方法是将全天分作若干时段,在每一类时段中等间距发车。这时,模型的变量可取作各类时段的发车时间间距,从而可建立有确定的低维数的数学规划模型。例如,将全天分为平峰时段和高峰时段,分别间距 t_1 和 t_2 分钟发一班车,则可得到以 (t_1, t_2) 为变量的 2 维模型。

题目称调度方案应满足四项要求。若记 $\bar{t}$ 为乘客候车时间,$\hat{t}$ 为早高峰时乘客候车时间,p 为车辆载客人数,则要求为

(a) $\bar{t} \leqslant 10$ 分钟; (b) $\hat{t} \leqslant 5$ 分钟;

(c) $p \geqslant 50$; (d) $p \leqslant 120$。

按照题目所列的乘车人数统计数据，严格满足这四项要求的调度方案是不存在的。而题目只是限定(d)为硬性约束，不应违反。其他三条使用"一般不要"措词，即是说可以违反，但应使违反程度尽可能低。这很自然导向使用目标规划(goal programming)模型，而且粗看起来，如下模型是合理的，

$$\min \quad (s_1^+, s_2^+, s_3^-) \tag{1}$$

$$s.t. \quad \bar{t} - s_1^+ \leqslant 10 \tag{2}$$

$$t - s_2^+ \leqslant 5 \tag{3}$$

$$p + s_3^- \geqslant 50 \tag{4}$$

$$p \leqslant 120 \tag{5}$$

$$t_1 \geqslant t_2 \tag{6}$$

$$t_1, t_2, s_1^+, s_2^+, s_3^- \geqslant 0 \tag{7}$$

然而，由于约束(2)和(3)要对每个乘客成立，约束(4)和(5)要对每辆车每个路段(指相邻二车站间的道路区段)成立，这一模型只有观赏意义，是难以处理的。将 $\bar{t}$ 和 t 换作所有乘客的最大候车时间，p 换作各车各路段的最小载客数，便真的是作成了三个约束。但此时 $\bar{t}$、t 和 p 作为(t_1, t_2)的函数难以给出表达。不仅如此，因为 $\bar{t}$ 和 t 是以分钟为单位的等候时间数，p 是车上的人数，二者是不可比较的，所以会给权因子的设置带来困难。这样的模型在答卷中有，为数极少。如上述，这一模型实际上是不能操作的。必须另外想变通的方法。

一种作法是计算两种时段中超时候车($\bar{t} > 10$, $t > 5$)的乘客数在各段乘客总数中所占的百分比(作为(t_1, t_2)的函数)，计算载客人数 $p < 50$ 的车路段数在全部车路段数中所占的百分比，即

$$g_1(t_1, t_2) = \frac{\text{平峰及晚高峰时段 } \bar{t} > 10 \text{ 的乘客数}}{\text{同时段乘客总数}},$$

$$g_2(t_1, t_2) = \frac{\text{早高峰时段 } t > 5 \text{ 的乘客数}}{\text{同时段乘客总数}},$$

$$g_3(t_1, t_2) = \frac{p < 50 \text{ 的车路段数}}{\text{车路段总数}},$$

并要求其取最小。这样直接导致多目标规划问题

$$\min \quad (g_1, g_2, g_3) \tag{8}$$

$$s.t. \quad p \leqslant 120 \tag{9}$$

$$t_1 \geqslant t_2 \geqslant 0 \tag{10}$$

三个百分率函数 g_1, g_2, g_3 的引入，既将涉及各个乘客和单个车路段的要求目标整合为一种集总的表述，又将三个函数的数量统一为无单位的百分率，使之有良好的可比性，易于用加权和形式化为单目标问题，从而使问题(8)-(10)成为一实际可操作的模型。

这次评定得奖的答卷，很大一部分都采用了各种不同的集总表示。如不少答卷定义了各种不同的乘客满意度指标和公交公司满意程度指标，也有的用总等待时间和总发车数，平均等待时间和平均载客量等。这些集总表示各有长短，所导致的调度方案也各不相同，但只要想法合理，表述和计算正确，都是合理的解答。但是有的答卷，设想了公交公司利益和乘客利益的一些不同的方面，作为要优化的目标函数，而将(a)-(d)四项要求作为约束条件，因无可行性，便放开某一条或某几条，仍将其余的作为硬性约束。更有个别答卷将(a)-(c)三条作为硬性约条处理，而完全放开表现载客数上限的条目(d)。这样做实际上已经改变了原来的题目。就

题目文字所表现的意义，公交公司利益和乘客利益的体现就是(a)－(d)这四项要求。

除了数学规划模型之外，部分答卷建立了排队论模型，有的用了计算机模拟。这样的答卷也有成绩很好的。

2　模型建立

上段谈及的模型，只是给出了一个框架，如(8)－(10)。要完成模型建立，还需利用统计数据，将 g_1, g_2, g_3 和 p 等各个函数的表达方法建立起来。

记早高峰时段为$[\hat{T}_1, \hat{T}_2]$，晚高峰时段为$[\hat{T}_3, \hat{T}_4]$。我们假设由发车间隔 t_1, t_2 所确定的全天共 m 班车的发车时刻序列为

$$T_1, T_2, \cdots, T_m$$

这些 T_i 值在各高峰和平峰时段内是均匀等间距的，在跨越 $\hat{T}_k, k = 1,2,3,4$ 时的确定 T_i 值的规则，要作一些补充规定。比如，设 $\hat{T}_1$ 左侧相邻的 T_i 值为 T_{i_1}，则可规定

$$T_{i_1+1} = \begin{cases} T_{i_1} + t_2 & \text{当 } \hat{T}_1 - T_{i_1} \leqslant t_2 \\ \hat{T}_1, & \text{当 } \hat{T}_1 - T_{i_1} > t_2 \end{cases}$$

$\hat{T}_2$ 左侧相邻 T_i 值为 T_{i_2}，则规定 $T_{i_2+1} = T_{i_2} + t_1$，等等。或按其他规则，此处不赘述。

设以 $j = 1,2,\cdots,n$ 标记车站。从 $j-1$ 站到 j 站的站间行车时间(包括 j 站的停车时间)记为 $\tau_j, j = 2,3,\cdots,n$。第 k 班车驶离 j 站的时刻记作 T_{kj}，则有

$$T_{k1} = T_k, \qquad T_{kj} = T_{k1} + \sum_{\ell=2}^{j} \tau_\ell, j = 2,3,\cdots,n-1 \tag{12}$$

这便给出了每一班车的行车时刻表。

为了求 $g_3(t_1,t_2)$ 和 $p(t_1,t_2)$，需要计算每一班车驶离各站时车上的乘客数 $p_k(T_{kj}), k = 1,2,\cdots,m, j = 1,2,\cdots,n-1$。而为了计算 $g_1(t_1,t_2)$ 和 $g_2(t_1,t_2)$，需要计算每个乘客的候车时间。这些都要从乘客流的统计数据出发。

首先对每一车站 j 引入乘客来站的时间密度，即单位时间来站乘客数，记以函数 $u_j(t), j = 1,2,\cdots,n-1$。统计数据恰好给出每隔 60 分钟一个点上的平均密度值。据此或用插值或用拟合，总可以得出 $u_j(t)$ 的一个近似表示。这里最简单的是用阶梯函数；作分段线性插值得一折线函数也很简单好用。答卷上对此一细节的处理花样繁多，有的甚至作了长篇幅的讨论。其实，这在整个解案中不过是一个细节，下功夫过大则势必舍本逐末。类似可建立乘客离站的时间密度函数 $d_j(t), j = 2,3,\cdots,n$。虽说乘客来站是分散的，离站是成批的，但统计数据中并未反映这一区别，因而在计算中也就作完全一样的处理。

乘客等候时间的计算较为曲折。这里使用将乘客按等候时间长短分类统计的方法。为此，设第 k 班车到达 j 站时，站上已等候过 h 辆车而仍未能上车的乘客数为 $w_{kj}(h)$。以下将 $p_k(T_{kj})$ 和 $w_{kj}(h), j = 1,2,\cdots,n-1, k = 1,2,\cdots,m$，用对 k 和 j 二重递推的方式表述出来。

注意对任意的 k, j，当 $h = 0$，因为 $w\int kj(0)$ 为时间段$[T_{k-1\,j}, T_{kj}]$的来客数，所以有

$$w_{kj}(0) = \int_{T_{k-1\,j}}^{T_{kj}} u_j(t)\,dt \tag{13}$$

其中的 T_{0j} 理解作题目所讨论的全天的开始时间，即

$$T_{0j} = T_0 = 5 \times 60,\ j = 1,2,\cdots,n \tag{14}$$

对于 $k=1$，根据(13)式，对任意的 j，$w_{1j}(0)$ 为已知。

以下证明：对于任意 k，若对任意 $h \leqslant k-1$，任意 j，$w_{kj}(h)$ 为已知，则可以导出 $p_k(T_{kj})$，$j=1,2,\cdots,n-1$ 和 $w_{k+1,j}(h)$，$j=1,2,\cdots,n-1$，$h \leqslant k$。我们假设乘客按先到先上车的队列原则乘车。

记 h_{kj} 为 j 站上等待最久的乘客的候车趟数，即

$$h_{kj}=\max\{h \mid w_{kj}(h)>0\}$$

我们规定 $T_{k0}=0$，$P_k(T_{k0})=0$。依次对于 $j=1,2,\cdots,n-1$，记

$$a_{kj}=\max\{p_k(T_{kj-1})-\int_{T_{k-1j}}^{T_{kj}} d_j(t)dt, 0\} \tag{15}$$

即第 k 班车到达 j 站，乘客下车后，车上仍留下的乘客数。这里的 T_{0j} 仍如(14)式。此时，可容纳的乘客上车数上界

$$b_{kj}=120-a_{kj} \tag{16}$$

这时有

$$p_k(T_{kj})=\min\left\{120,\ a_{kj}+\sum_{h=0}^{h_{kj}} w_{kj}(h)\right\} \tag{17}$$

$$h_{k+1j}=\max\left\{h \mid h=0 \text{ 或使 } \sum_{r=h}^{h_{kj}} w_{kj}(r)>b_{kj}\right\} \tag{18}$$

(18)式的意思是按候车趟数从多到少依次累加，$w_{kj}(h_{kj})+w_{kj}(h_{kj}-1)+\cdots$ 直到首次出现大于 b_{kj} 则停止累加，以累加到的最后的(最小的)一个 h 作为 h_{k+1j}；如果直加至 $w_{kj}(0)$ 仍不大于 b_{kj}，则 $h_{k+1j}=0$。然后，若 $h_{k+1j}=0$，则 $w_{k+1j}(1)=0$；否则

$$w_{-k+1j}(h)=w_{kj}(h-1),\ h=1,2,\cdots,h_{k+1j}-1 \tag{19}$$

$$w_{k+1j}(h_{k+1j})=w_{kj}(h_{k+1j}-1)-b_{kj}-\sum_{h=h_{k+1j}}^{h_{kj}} w_{kj}(h) \tag{20}$$

$w_{k+1j}(0)$ 则如(13)式。

现在可以建立模型(8)－(10)中的各个函数了。利用 $P_k(T_{kj})$，有

$$g_3(t_1,t_2)=\frac{\sum_{k,j}\{1 \mid p_k(T_{kj})>50\}}{m(n-1)} \tag{21}$$

注意 $p_k(T_{kj})$ 的构造过程已保证了约束条件(9)式的成立，因而无须再求函数 $p(t_1,t_2)$，该约束在建成的模型中也不再出现。

为求 g_1 和 g_2，注意到至 T_{kj} 时刻，j 站上已候 h 辆车而仍未能上车的乘客的等候时间满足

$$T_{kj}-T_{k-hj} \leqslant t \text{ 或 } t \leqslant T_{kj}-T_{k-h-1j}$$

以保守的原则取其下限，则有

$$g_1(t_1,t_2)=\frac{\sum_{T_{kj}\in[T_1,T_2]}\{w_{kj}(h) \mid T_{kj}-T_{k-hj}>10, h=1,2,\cdots,h_{kj}\}}{\sum_j \int_{t\in[T_1,T_2]} u_j(t)dt} \tag{22}$$

$$g_2(t_1,t_2)=\frac{\sum_{T_{kj}\in[T_1,T_2]}\{w_{kj}(h) \mid T_{kj}-T_{k-hj}>5, h=1,2,\cdots,h_{kj}\}}{\sum_j \int_{t\in[T_1,T_2]} u_j(t)dt} \tag{23}$$

三个函数都是百分率,数值上有良好的可比性,可以简单地取其加权和作为单一目标函数。题目中没有关于优先权及权重的规定,可以认为公交公司利益与乘客利益等权,并且表现乘客利益的二函数 g_1 和 g_2 等权。这样便有

$$\min \quad \frac{1}{4}g_1(t_1,t_2)+\frac{1}{4}g_2(t_1,t_2)+\frac{1}{2}g_3(t_1,t_2) \tag{24}$$

$$s.t. \quad 0 \leqslant t_2 \leqslant t_1 \tag{25}$$

可以以此作为求解 t_1、t_2 的规划模型。

三个函数的表述,是该题目的最具难度之点。构想一个如上一段所述的模型框架并不很难,要表述出这些函数从而完成模型的建立,却是对参赛者的数学训练和手上功夫的考验。这次的答卷中有表现极好的,但是不多。从这一现象是否可以提出一个问题。近一些年来,在教学中很注意思想性的方面,学生在这一方面也确实有很大提高。但对于实际的手上功夫的训练却渐趋弱化。固然技术的进步代替了许多手上的操作,但同时却派生出许多新的需要人工操作的事情。即如本题,不是计算技术如今天的进步,也就不会提出这样的题目。在教学上应当如何适应情况的这种变化呢?

3　模型求解

此次的答卷中,模型建立与其后的求解计算两相脱节者所在多有。这是不好的。建立了模型,而又不用它去寻解案,那么建模又是为了什么呢?所建的模型难解,不会解,那就应当再进一步抽象,再引入新的假设以简化模型,直到你能够处理它为止(包括引出理论结果或数值求解)。前面两段中,我们先是作等间隔发车的假设,以便将一个有不确定的高维数的问题变为一个 2 维问题;然后又将本应是面向各单个乘客和单个车路段的目标要求,集总化为三个百分率函数。这些都是为了使得做成的模型能够作求解处理。

模型(24)-(25)是容易求解的。作为一个非线性规划问题,只有2维,且约束极为简单,除非目标函数的求值不能实现,总可以用某种直接方法(不涉及求导数)求得近似解。上段中函数 $P_k(T_{kj})$ 和 $w_{kj}(h)$ 的二重递推式的表述,实际上就给出了对给定的 (t_1,t_2) 值求这些函数值的算法。(21)、(22)和(23)式则表示如何用这些函数值完成目标函数求值运算。而实际上,(21)式的分子是合条件的 $P_k(T_{kj})$ 的个数的计数器,(22)和(23)式的分子是合条件 $w_{kj}(h)$ 值的累加。因此,g_1、g_2 和 g_3 的计算可以放人到为产生 $P_k(T_{kj})$ 和 $w_{kj}(h)$ 的对 k 和 j 的 2 重循环中。随产生随累加,以免去对这些不再有用的 $P_k(T_{kj})$ 和 $w_{kj}(h)$ 的存储。使用参赛者熟知的 Matlab 语言,容易作成给定 (t_1,t_2) 值后产生目标函数值的程序段,再使用 Matlab 的优化工具箱的程序即可完成求解。这当中或许要为 t_1、t_2 设一个合理的上界,或许要通过随机搜索选一个好一点的初始点,都是容易做到的事情。有的答卷中选用遗传算法求解此模型,单从其目标函数构成复杂,可以求值而难于求导或作其他处理等特点,这一选择就是正确的。

另外要指出一点。如果将函数 g_3 和 g_1 进一步分拆成两项,即分别单独计算平峰时段和高峰时段的载客数不到 50 的车路段百分率和候车超时乘客的百分率,并且对两种时段的交接点处的情况作适当处理,目标函数就可以化为可分离的形式,即

$$f(t_1,t_2)=f_1(t_1)+f_2(t_2)$$

这时就可以分别求解两个一维问题

$$\begin{array}{llcll} \min & f_1(t_1) & \qquad & \min & f_2(t_2) \\ s.t. & t_1 \geqslant 0 & & s.t. & t_2 \geqslant 0 \end{array}$$

比如分别为 t_1 和 t_2 设定合理的上界,就可以用0.618方法求解。(其关键的运算还是目标函数 f_1 和 f_2 的求值。)设得解为 t_1^* 和 t_2^*。如果 $t_2^* \leqslant t_1^*$,则(t_1^*,t_2^*)就是所求之解。

4 结束语

作者认为,这个题目的难点在于目标函数的具体表述(如本文第2段)。如果判定建立此表述的难度是适当的,那么这个题目是选得很好的。事实上,尽管漂亮的答卷不多,但众多答卷上对于这样一个活生生的相当复杂的实际问题,表现出分析的思路是合理的,也很有深度。为解决问题,表现的思路也很广。不足之处,从总体上说,相对于完成表述,显得数学训练和手上功夫欠缺些。或难度大了些(相对于三天集中时间)?或的确是训练差了些?或二者兼有?以供讨论。建模是重要的。在各类信息系统平台日趋完善的今天,如何从中发现知识,如何优化决策,人们祈求建模。而建模不光要有思想,也要靠人工完成表述,这要求数学训练和手上功夫。

致 谢:本文参考了大量的参赛者的答卷,和全国及北京赛区两级阅卷组的评论意见,并特别引用了供题者所提供的参考答案。谨在此一并致谢。

Ideas, Well-training and Skills are Needed in Mathematical Modeling

——A Summary of Answers of the Problem B

LIU Bao-guang

(Beijing Institute of Technology, Beijing, 100081)

Abstract: This paper presents a summary of answers of the problem B of China Undergraduate Mathematical Contest in Modeling in 2001, from three aspects: framework of models, model building and model solving.

Key words: mathematical modeling; model building; bus scheduling