

文章编号:1005-3085(2005)07-0113-04

# DVD 在线租赁业务中的数学模型

谢金星

(清华大学数学科学系, 北京 100084)

**摘 要:** 2005全国大学生数学建模竞赛 B、D 题背景是 DVD 在线租赁业务。本文对竞赛论文中出现的主要模型和方法进行简要介绍和分析, 并指出以此为背景还有很多问题可以继续研究。

**关键词:** DVD 在线租赁; 数学建模竞赛

**分类号:** AMS(2000) 62J05

**中图分类号:** O212.1

**文献标识码:** A

## 1 引言

2005全国大学生数学建模竞赛 B、D 题背景是 DVD 在线租赁业务, 由亚马逊公司全球供应链运营副总裁于刚先生提供。于刚先生同时还是美国德州大学奥斯汀分校管理学院 Jack G. Taylor 讲席教授, 他的理论研究课题主要从应用中产生并直接返回到应用中, 已成功申请到多项美国专利。于刚教授于1995年创建美国科莱科技公司 (CALEB Technologies Corp.) 并任董事长和总裁, 他十多年来开创了实时运作管理的新领域并成功地将其应用于航空管理, 2001年为美国大陆航空公司所创造的价值超过6000万美元, 荣获2002年度运筹与管理科学应用 Franz Edelman 奖, 这一奖项被认为是运筹与管理科学应用的“世界杯”[1]。

于刚先生目前所在的亚马逊英国公司 (amazon.co.uk) 正在开展网上 DVD 在线租赁业务, 市场潜力很大, 但竞争十分激烈, 例如在美国已经有 netflix.com 和 blockbuster.com 等、在欧洲已经有 lovefilm.com 等著名公司。众所周知, 以尽可能低的成本向用户提供尽可能优质的服务, 对赢得竞争至关重要。在企业的实际运营中, 对降低成本和优质服务的内容、手段可以有很多不同的理解, 而且与企业的几乎所有业务流程相关联, 是高度综合、非常复杂的问题。这次赛题中的问题只能看成是实际运营中几个关键问题的非常粗略和理想化的简化版本, 例如问题中的数据规模似乎已经不算小了, 但与实际相比较仍然是小很多(实际中很多公司租赁的 DVD 多达几万种, 用户多达几十万甚至几百万, 有的公司还包括多个配送中心)。

题目中主要要求解决以下几个问题: 题中(1)、(2)问分别考虑购买和分发子问题, 第(3)问则同时考虑购买和分发, 第(4)问要求参赛队自己提出新问题和尝试建模和求解。问题初看起来似乎很容易理解而且并不复杂, 但考虑到 DVD 在线租赁业务中存在的各种不确定性和多阶段特征, 建立好的数学模型并不容易。对题目思考的深度和广度不同, 也就可能理解不同、假设不同, 得到的模型和结果可能很不相同, 因此一篇好的论文应特别注意假设的合理性及所建立的模型与假设之间的一致性。

## 2 网站购买 DVD 的最优数量

这里仅针对一个月的情况对参赛论文中出现的几种主要模型和方法进行介绍, 三个月的情况可相应考虑, 只是由于需要考虑的时间阶段数增加, 算法更复杂一些。解决本小问时遇到的

第一个假设可能就是每种 DVD 应该独立考虑, 因为题中仅给出了其中5种 DVD 的需求预测, 显然不是网站用于出租的所有 DVD, 而且也没有说明他们之间的关联性。

题中要求保证希望看到该 DVD 的会员中至少50%在一个月内看到该 DVD, 故首先应弄清希望看到该 DVD 的会员数量。对问卷调查数据普遍的一种理解方式是由此得到该 DVD 被选中的概率(记为  $p$ ), 设网站的会员总数为  $n$ , 则该 DVD 的需求可看成二项分布  $B(n, p)$ 。题中要求的虽然是“保证”其中至少50%在一个月内能看到该 DVD, 但由于抽样调查的随机性, 这种保证可理解为在一定置信水平下的保证, 即在一定置信水平下计算希望看到该 DVD 的会员人数的上限  $M$ 。很多论文将 DVD 的需求看成是常数而不是随机变量, 相当于认为上限  $M$  等于二项分布的均值, 不太合理。在  $n$  较大时, 该 DVD 的总需求也可用正态分布  $N(np, npq)$  近似 ( $q = 1 - p$ ), 不过是否进行这一近似对本题并不是本质的。

确定该 DVD 的购买数量时, 一种简单的想法是先确定每张 DVD 在一个月内可被重复使用的次数。这显然也是随机变量, 可以对它进行不同的估计, 不同论文之间差异较大。题中只说明了60%的会员每月租赁 DVD 两次, 而另外40%只租一次, 一种简单近似方法是认为1个月该 DVD 可重用次数是  $0.6 \times 2 + 0.4 \times 1 = 1.6$  次, 这相当于对每月租赁两次的会员把一个月简单地分成前半个月和后半个月两个阶段进行处理。如果假设租出去的 DVD 的返还率在总体上相对稳定, 即40%的会员每天有  $1/30$  的概率归还, 60%的会员每天有  $1/15$  的概率归还, 结果是一致的。有的论文通过对网站事先是否知道哪些会员每月租赁两次和一次作出不同的假设, 从而对可重用次数作出保守和乐观等多种估计, 也有一定道理。还有的论文将用户归还 DVD 的时间周期假设成随机变量(如服从均匀分布或正态分布等), 这样计算得到的可重用次数可能较大, 但由于往往需要假设归还回来的 DVD 马上能再借出去, 在一般情况下是否合理值得商榷。

以 DVD1 为例, 则  $p=0.2$ ,  $n=100000$ , 如取置信水平为95%, 可得  $M=20208$ ; 若重用次数取1.6次, 得到的结果是购买  $0.5M/1.6 = 6315$  张。很多论文把  $M$  简单地取二项分布的均值, 得到的结果为6250张。采用数值模拟(仿真)直接求解或先进行理论分析再用模拟进行验证, 也是一种可取的方法, 多数论文模拟的结果比上述数值还要低一些, 但往往没有很规范地阐明模拟的详细过程及其合理性, 说明参赛同学在这方面的训练还需要进一步加强。

值得指出的是, 对问卷调查结果可以存在其他理解方式, 例如考虑到该 DVD 租赁业务的生命周期, 可以认为表中给出的某 DVD 的需求率只是初始时段(如一个月或半个月)的需求率, 并进一步假设以后时段的需求率持续不变或按某种规律变化, 这时的系统类似于随机服务系统或随机库存系统, 从而可以利用随机过程、排队论或随机决策的一些模型、理论和方法。不过由于会员每月最多只能租赁两次, 如何在现有模型中考虑这个约束是比较棘手的问题。有一些论文显示出这方面的思想萌芽, 但由于需要的数学知识和技巧性比较专业, 往往并不成功。本小问提出的问题看似简单, 但还有很多工作可以继续研究下去。

### 3 网站分发 DVD 的数学模型

本小问大概是本题中最容易数学化的部分。由于已知手头 DVD 的数量和会员的订单, 因此无需考虑会员是否会归还等多阶段问题, 只需考虑当前进行一次分发的情形。下面用  $n, m$  分别表示当前需要分发的会员数量和 DVD 种类数, 用  $c_j$  表示第  $j$  种 DVD 的现有数量, 用  $a_{ij}$  表示表格文件中给出的订单矩阵。用变量  $x_{ij}$  表示是否选择第  $j$  种 DVD 分配给第  $i$  位用户, 用变量  $y_i$  表示第  $i$  位用户是否得到 DVD。顾客的满意度可以自行定义, 如满意

度定义为  $s_{ij} = 1/a_{ij}$  或者  $27 - a_{ij}$  等 (当  $a_{ij} > 0$  时, 否则  $s_{ij} = 0$ ), 一种可能的模型为

$$\begin{aligned} \text{Max} \quad & \sum_{ij} s_{ij} x_{ij} \\ \text{s.t.} \quad & x_{ij} \leq a_{ij}, \sum_i x_{ij} \leq c_j, \sum_j x_{ij} = 3y_i, x_{ij}, y_i = 0, 1. \end{aligned}$$

如果容许为会员分发没有定购的 DVD, 则第一个约束可以去掉; 如果容许会员得到的 DVD 数量少于3张, 上面约束中的等号需要换成小于等于号。

本题 (尤其 B 题) 数据量较大, 但容易用 LINGO 软件算出最优结果[2]。根据本人计算经验, D 题规模较小, 也可用 MATLAB7.0 以上版本的纯 0-1 规划程序 Bintprog 求解, 但速度较慢; B 题规模较大, 用 Bintprog 求解几乎不可能。如将模型松弛成线性规划求解再舍入到 0-1 解, 应该验证解的可行性和说明解的误差。有些参赛同学没有找到合适的软件进行求解, 所以设计启发式算法搜索求解, 也是可以的, 甚至对更大规模问题可能是真正实用的方法。需要注意的是, 论文应该详细说明搜索算法的设计思想和步骤, 最好还能对算法的计算效率和效果进行一定的分析。针对模型中参数变化的情况, 对结果进行敏感性分析也是值得鼓励的。也就是说, 验证和说明你所提出的算法的优点是你的责任, 应该是论文的重要组成部分, 也只有这样才能指导实践, 否则实际决策者就会提出质疑, 而不会轻易相信和采纳你的算法。

值得指出的是, 有些论文建立了相应的网络优化模型, 在所有论文中显得比较有特色。例如, 可设计一个包括  $n + m + 2$  个节点的网络, 其中  $n$  个节点对应于会员,  $m$  个节点对应于 DVD, 另外两个虚拟的节点分别是源点和汇点。从源点到每个会员节点有一条弧, 对应的容量为3 (每个会员最多得到3张 DVD), 费用为0; 从每种 DVD 节点到汇点有一条弧, 对应的容量为手头 DVD 的现有数量, 费用为0。会员节点与 DVD 节点之间有弧连接, 当且仅当该会员的订单中包含该 DVD, 对应的容量为1 (每个会员最多得到1 张该 DVD), 费用可以定义成题中给出的偏好排序数字。此时, 原问题就转化成了在这个流网络上寻找最小费用流, 已有有效算法。可参见本期发表的上海交通大学参赛的论文, 网络优化的更多知识可参见[3]。

#### 4 购买和分发同时考虑

总体来说这一问需要考虑购买和分配 DVD 两个方面, 需要兼顾减少购买成本和提高满意度并满足一定的服务水平。同时由于会员一个月可能租赁两次, 因此需要考虑多阶段决策的因素, 很难建立精确的数学模型。大致有以下一些思路:

1) 在一定的假设下, 把问题近似分解成前面考虑过的购买和分发两个子问题。例如, 有的论文先根据会员订单统计 DVD 的需求情况, 确定 DVD 购买量, 然后用前一问中建立的模型进行第一次分发, 再对网站是否知道哪些会员租赁两次作出一定假设, 进行第二次分发。

2) 有的论文对前一问中建立的模型进行一定修改, 建立购买和分发统一的多目标数学规划模型, 且同时考虑两次分发和服务水平约束, 不过往往在二次分配和服务水平约束方面考虑有些缺陷。

3) 考虑到一个月内可能一个会员要发货两次, 这又是一个多阶段的决策问题, 建立随机决策模型并寻找最优决策是可能的 (例如采用马氏决策方法), 但由于后一阶段决策时需要考虑前一阶段哪些会员归还了哪些 DVD, 因此这样建立模型的难度较大, 评委们在评阅中几乎没有见到非常成功的论文。

4) 采用数值模拟 (仿真) 建模和求解, 或检验其他模型。与求解第一小问类似, 这时同样需要很规范地阐明模拟的详细过程及其合理性, 参赛论文在这方面往往注意不够。

## 5 其他问题

DVD 在线租赁网站在需求预测、采购和分发方面所面对的问题是很多的。本小问要求同学自己思考、提出问题,并解决问题。这也是本题的重要组成部分,可供同学充分发挥其创造性,特别是提出问题的能力,提出的问题应具有实用价值或理论价值,并给出建模思路。例如,有的论文提出网站应充分利用会员过去租赁的历史数据从各方面认真进行统计分析(数据挖掘),并对会员进行分类处理;有的论文提出网站对 DVD 应分多次购买而不是一次性购买;有的论文提出网站的短期利益与长期利益应该权衡;有的论文讨论如果不限会员每月最多只租赁2次,对网站运营会有什么影响;等等。

## 6 关于论文的写作

好的写作技巧可为论文增色,这也是竞赛培养的能力之一。一般来说,论文写作没有固定形式,关键是简洁、清晰地阐述思想、方法和结果。按照竞赛要求,摘要限定为1页,本人认为这可以理解为详细摘要,应该包括对问题的基本理解、建立模型的关键假设、主要的模型和算法及其特点、主要结论等,让人阅读完摘要后对论文的概貌能有基本全面的了解。有些论文摘要与全篇论文几乎没有直接关系,过于空泛,作为其他人的论文摘要可能也适用,这样的摘要有价值的信息太少,不是好摘要。论文正文中有些同学总是模仿“八股文”似地重复“问题重述”、“问题分析”,而实际上只是把题目重抄一遍,这样的重述还不如不要。有些论文专用一节列出文中的所有假设和数学符号,这当然可以,但本人觉得即使这样,在以后第一次用到它们时还是应该说明其含义以增加论文的可读性。论文正文写作应该有一条主线,思路清晰,一些关键性的计算结果应该在正文中给出,其他有用的信息(包括源程序和详细结果)应该很有条理地整理好,放到附录中并在正文中给出引用。

总的来说,本题背景非常实际,牵涉面广,可以考虑的因素很多,不大可能有惟一、精确、最优的解决方案。因此,本人觉得本题是一道比较开放性的题目,基本上符合“宽无边、深无底”的原则,还有很多问题可以继续研究下去。

### 参考文献:

- [1] 中国运筹学会. 热烈祝贺于刚教授荣获Franz Edelman奖[J]. 运筹通讯, 2002,12(2):51-53
- [2] 谢金星, 薛毅. 优化建模与LINDO/LINGO软件[M]. 北京: 清华大学出版社, 2005
- [3] 谢金星, 邢文训. 网络优化[M]. 北京: 清华大学出版社, 2000

## Mathematical Models for Online DVD Rent Business

XIE Jin-xing

(Department of Mathematical Sciences, Tsinghua University, Beijing 100084)

**Abstract:** The background for problems B and D in 2005 CUMCM (China Undergraduates Mathematical Contest in Modeling) is Online DVD rent business. This paper briefly introduces and analyzes the mathematical models and methodologies used in the contest answer papers, and points out that they are many issues which can be regarded as future research directions in this field.

**Keywords:** online DVD rent; mathematical contest in modeling