

基于多元回归模型和 Topsis 模型对货运线路价格的分析

张子璇¹ 刘 鹏² 杨洪滔¹ 王会帆¹ 崔思雯¹

(1、河北农业大学理学院,河北 保定 071000 2、河北农业大学机电工程学院,河北 保定 071000)

摘 要:在“互联网+物流”的无车时代,无车承运人作为一种新型物流运作模式正日渐兴起。为实现高效运作、服务规范、合作共赢,如何对承运路线进行科学定价将尤为重要。文中借助 MATLAB 和 SPSS 应用软件,利用因子分析确定了影响定价的主要因素,提出了基于多元回归的货运线路的定价模型,并利用 Topsis 模型对定价区间进行了综合评分。通过对某无车承运人平台的历史交易数据进行试验,得到了货运线路定价的回归模型为: $y=18.27+0.88x_1+0.16x_2-15.86x_3-443.81x_4+15.01x_5+108.57x_6$ 。线路指导价格最优区间为:(210,284],最劣区间为:(9433,12226],中间区间为:(1460,1800]和(780,1120]。

关键词:货运线路定价;因子分析;Topsis 模型;多元回归模型

中图分类号:F252,U492.3

文献标识码:A

文章编号:2096-4390(2021)07-0065-02

当今我国运输市场开放,无车承运行业不断发展。无车承运主要利用网络平台让每个运输任务被承运端司机接收,这需要提前将任务以一定的价格发布到互联网上,司机浏览完后决定是否接收该运输任务,如果任务未被接收,则需要调整价格来确保任务最终被接收,因此承运线路的科学定价问题也成为了无车承运人平台亟待解决的问题。

1 实验数据与处理方法

1.1 数据来源

本文所得数据来自我国无车承运人货运线路历史交易平台。

1.2 所用软件

SPSS、MATLAB。

1.3 数据处理

1.3.1 指标的筛选

为了选择适当的模型,需要对数据进行预处理,首先确定研究的指标,我们查阅了影响货运线路定价因素的相关资料后,确定了要研究的指标为:总里程数、线路编号、调价比例、始发网点、目的网点、交易成功时常、车辆长度、车辆吨位、线路总成本、业务类型、需求类型 1、需求类型 2、运输等级、调价紧急程度、需求紧急程度。

1.3.2 数据分类

由于数据量太大,不易直接对其价格进行评价,所以我们

对数据进行了分类,将线路指导价格分区间组合,共分为 12 类,将其可视化如图 1 所示。

1.3.3 数据指标降维

为了便利实验研究,本文利用因子分析降维。经计算得 KMO 值为 0.718>0.5,可知原有因子适合进行因子分析。又经计算得累积方差贡献率达到 82.765%,故降维后的数据能够代表绝大多数的初始数据。降维后的指标为线路价格、交易成功时长、调价紧急程度、线路成本、调价比例、车辆吨位、运输等级。其中线路价格用 y 表示,其他指标依次用 $X=\{x_1, x_2, x_3, \dots, x_6\}$ 表示。

2 模型的建立与求解

2.1 基于多元回归模型的线路定价

2.1.1 计算步骤

Step1 线性模型假设

设 X 与 y 满足如下线性关系:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \mu$$

Step2 联合显著性检验

对于多元线性回归模型,在对每个回归系数进行显著性检验之前,应该对回归模型的整体做显著性检验。这就是联合显著性检验——F 检验^[1],假设:

$$H_0: \beta_i = 0 \text{ vs } H_1: \exists \beta_i \neq 0$$

选取 F 统计量: $F \sim F(k-1, T-k)$, 我们得到联合显著性检验 $F(2, 15) = 16.008, P = 0.0001 < 0.01$, 所以在 99% 的水平下,拒绝原假设,即通过了联合显著性检验。

Step3 确定回归方程

我们对回归系数 $\beta_1, \beta_2, \dots, \beta_6$ 进行显著性检验,假设:

$$H_0: \beta_i = 0 \text{ vs } H_1: \beta_i \neq 0$$

构造 t 统计量: $t_i = \frac{\hat{\beta}_i}{\sqrt{c_{ii} \hat{\sigma}^2}} \sim t(n-1)$, 取显著性水平 $\alpha = 0.05$,

通过计算得知 $\beta_i < 0.05$, 所以 β 显著异于 0, 回归系数有意义。故回归模型为:

$$y = 18.27 + 0.88x_1 + 0.16x_2 - 15.86x_3 - 443.81x_4 + 15.01x_5 + 108.57x_6$$

2.1.2 实验结果分析

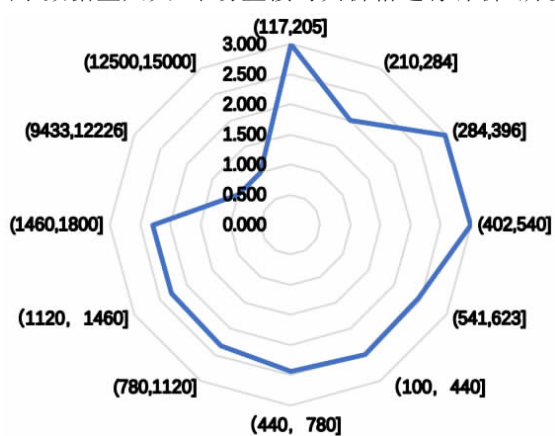


图 1 线路指导价格运输能力图

作者简介:张子璇(1999,10-),女,汉族,籍贯:河北石家庄,河北农业大学理学院本科在读,研究方向:数学与应用数学。

关注【公众号：数模加油站】获取更多

利用上述回归模型对未知的线路价格进行预测并初次定价,选用部分任务编号下的指标数据例如:

$$\begin{cases} 17281: x_1 = 826.3, x_2 = 117.6, x_3 = 0.9, x_4 = 1.3, x_5 = 6.7, x_6 = 3.2 \\ 15629: x_1 = 920.9, x_2 = 101.5, x_3 = 2.6, x_4 = 1.3, x_5 = 6.7, x_6 = 2.4 \\ \dots \end{cases}$$

现将自变量取值分别带入上文求得的回归模型,我们可以预测得到每个任务编号对应的第一次报价,将其结果和已知实验数据对比的预测效果比较好。

2.1.3 第二、三次调价策略

无车承运平台市场复杂多变,假设每条线路最多只能发布三次价格,为了最大程度降低承运人的成本,建立运输企业与货主之间的长期供应链合作关系,需要我们对调价报价制定一个合理的策略。

(1) 将第一次报价分为 5 等级(级数越高,定价越高),其中总里程数小($\leq 100\text{km}$)、业务类型为重货、需求类型分别为普通和计划、续签状态为已签时,我们定义第一次报价等级为 0 等级,有一个因素不满足时,定价水平变为 1 等级。

(2) 对于第一次报价中未成交的任务,需要进行第一次调价处理,此次调价则以报价合理性为目标,在第一次报价策略的基础上,每一等级相应的报价都是处于该等级的合理报价。

(3) 若两次报价都未成交,则以让司机接受任务为目标,进行调价处理,此时每一等级相应的报价都是满足司机要求的较高报价。

2.2 Topsis 模型对线路价格的评价

2.2.1 模型介绍

为了对线路指导价格进行评价,我们利用 Topsis 优劣解距离法^[2]对各个类别的价格进行评分,得分越高,则价格区间类别越好,从而确定最优类别价格、最劣类别价格和中间类别价格,并对其进行评价,分析其合理性。

2.2.2 计算步骤

Step1 指标分类

根据实际应用的可操作性,需对文中的七个指标分类,其中,线路价格、交易成功时长、调价紧急程度、线路成本为极小型指标,其他指标分类见表 1:

表 1 指标分类表

名称	指标分析	指标类型
调价比例	越接近 1 越好	中间型指标
车辆吨位	[7, 20]	区间型指标
运输等级	越大越好	极大型指标

Step2 指标正向化

为了计算方便,需要将上述四种类型的指标正向化处理。

正向化处理的依据:

(1) 极小型指标正向化公式:

$$\bar{x}_i = \max\{x_j\} - x_i$$

(2) 中间型指标正向化公式:

$$M = \max\{|x_i - x_{best}|\}, \bar{x}_i = 1 - \frac{|x_i - x_{best}|}{M}, \text{其中本题的 } x_{best} = 1$$

(3) 区间型指标正向化的公式(本文中 $a=7, b=20$):

$$\bar{x}_i = 1 - \frac{a-x}{M}, x < a; \bar{x}_i = 1, a \leq x \leq b; \bar{x}_i = 1 - \frac{x-b}{M}, x > b$$

Step3 标准化处理

为了避免不同指标量纲的影响,需要对正向化的指标进行标准化处理。本题有 12 个要评价的价格区间类别,7 个评价指标构成的正向化矩阵,正向化矩阵中的每一个元素用 x_{ij} 表示,记标准化的矩阵为 Z , Z 中每一个元素的计算公式为:

$$z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^n x_{ij}^2}}$$

Step4 计算得分并归一化处理

在标准化后的指标中找到最优方案和最劣方案,然后分别计算评价对象与最优方案和最劣方案间的距离,获得评价对象与最优方案的相对接近程度,作为评价优劣的依据,计算出价格的各个区间类别的最终得分,并对得分做归一化处理:

$$S_i = \frac{D^-}{D_i^+ + D_i^-}$$

2.2.3 实验结果分析

将归一化后得到的得分排序后可可视化得到图 2, 我们看到得分最低的价格类别是 (9433, 12226], 得分最高的价格类别是 (210, 284], 得分处于中间的价格类别是 (1460, 1800] 和 (780, 1120], 并且我们得到除最后两类价格, 其他价格得分比较均匀, 没有太大的差异结论。基于实验结果可以看出各分数段评分大体相同, 由此我们认为: 平台定价较为合理, 故其有一定的参考价值, 在后续分析中我们可以以历史数据的定价为基础进行分析。

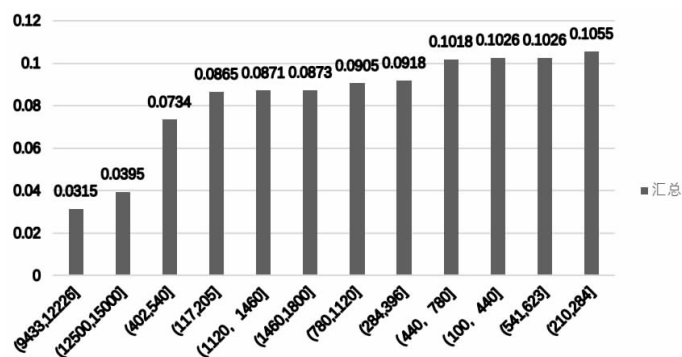


图 2 各段价格类别分数图

3 结论

通过研究分析,我们发现无车承运人平台的线路定价非常重要,对整个竞争市场的格局都产生着非常大的影响,所以,定价调价是具有挑战性的决策。基于本文的回归模型与实验结果,结合我们制定的调价政策,能较为准确的预测出未完成订单的报价。利用 Topsis 模型对预测报价进行评价,发现不同区间之间评分较为均匀,合理性具有一定保证。由此评价结果也可以反应出我们的调价策略与预测结果较为合理,具有一定可行性。

参考文献

- [1]xuxiatian, 多元线性回归模型的 F 检验, <https://blog.csdn.net/xuxiatian/article/details/55047998> 2020-05-02, 2020-05-24.
- [2]黄发明, 李友铎. 基于改进 Topsis 法的湖泊水质评价[J]. 水电能源科学, 2020(4): 51-54.