

# Research on Cross-selling Forecast of Auxiliary Products of Low-cost Airlines

Tao Li Zhiyu Zhu Jing Liu Zhiqiang Ren

Civil Aviation Flight University of China, Guanghan, Sichuan, 618307, China

## Abstract

This paper analyzes the problem of cross selling of assistive products of low-cost airlines, and proposes ways to build a combination forecasting model based on NPTB model to predict the probability of passengers purchasing assistive products at the next flight. On this basis, passengers with high probability of purchasing assistive products at the next flight are selected to promote auxiliary products, so as to achieve the purpose of increasing revenue. The actual data show that the combination forecasting model constructed in this paper is better than the single model.

## Keywords

assistive products; cross-selling; NPTB; portfolio forecasting

## 低成本航空公司辅助产品交叉销售预测研究

李涛 朱志愚 刘静 任志强

中国民用航空飞行学院, 中国·四川广汉 618307

## 摘要

论文分析低成本航空公司辅助产品交叉销售的问题, 提出在 NPTB 模型的基础上构建组合预测模型来预测旅客在下次乘机时购买辅助产品的概率。在此基础上选择下次购买辅助产品概率高的旅客进行辅助产品推广, 以达到增加收益的目的。通过实际数据证明, 论文构建的组合预测模型优于单项模型。

## 关键词

辅助产品; 交叉销售; NPTB; 组合预测

## 1 引言

根据咨询公司 Idea Works Company 与在线租车分销系统供应商 Car Trawler 联合发布的最新报告预测, 2018 年, 全球航空公司的辅助收入预计将达到 930 亿美元, 相比于 2017 年的 822 亿美元, 增长了 13%; 相比于 2016 年的 674 亿美元, 增长率更加显著<sup>[1]</sup>。所以, 如何增加辅助收入是中国航空公司一直思考的问题。而低成本航空公司与全服务航空公司之间有一个显著的区别, 就是低成本航空公司的辅助收入占其总收入的比重大, 重视辅助产品的开发, 且远远大于传统航

空公司。目前, 航空运输市场面临竞争加剧, 油价上涨, 机队成本增加等挑战, 使低成本航空公司的利润空间进一步缩小。由于中国低成本航空公司起步晚, 辅助收入以及辅助产品的开发与国外航空公司存在巨大的差距。

论文主要研究与旅客服务相关的辅助产品即菜单式自选附加服务产品的交叉销售预测。利用某低成本航空公司的辅助产品订单数据和会员旅客信息数据进行实证分析, 对相关数据进行处理, 选取变量, 在 NPTB 模型的基础上构建组合预测模型进行预测, 用 Clementine 软件进行模型计算, 预测旅客下次购买某种辅助产品的概率, 能够使航空公司精准定位目标客户, 旅客的需求也得到满足。

## 2 辅助产品分类

航空公司的主营业务就是提供旅客、货物的位移服务,

**【基金项目】**中国民用航空飞行学院大学生创新创业训练计划项目(项目编号: S201910624182)

**【作者简介】**李涛(1995-), 中国民用航空飞行学院研究生, 从事交通运输工程研究。

表 1 辅助产品按旅客流程分类

流程	购票	值机	登机	机上服务	下机	其他
辅助产品	保险					
	机票退改签	超额行李费		机上餐食	接机服务	景点门票
	送机服务	提前选座	贵宾休息室	机上 Wi-Fi	空铁、空陆联运	报销凭证
	行李费预付	优先选座	优先登机	机上零售	酒店	里程销售
	金融服务	优先值机		机上娱乐	租车	免税商店

而辅助产品是指航空公司在提供位移服务的基础上围绕旅客所开展的其他一系列服务，如机上餐食、超额行李、贵宾休息室、机上 Wi-Fi、保险、提前选座等。不同的研究者将其命名为附加服务、增值业务、增值服务，非航空产品，其含义是相同的。根据旅客的乘机流程，论文将常见的辅助产品归纳为表 1 所示的内容。

### 3 基于 NPTB 模型的组合预测

2002 年，Knott 等人提出 NPTB (Next-Product-To-Buy) 模型，该模型能够使交叉销售的有效性增强<sup>[2]</sup>。其模型的结构框架如图 1 所示。

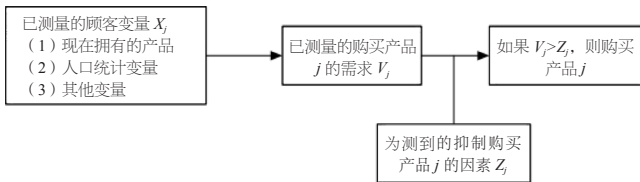


图 1 NPTB 模型结构

图 1 中， $X_j$  代表了模型的变量向量，表示对产品  $j$  测得顾客的相关数据，例如现在顾客拥有的产品，人口统计变量以及其他变量等 ( $X$  表示对某产品产生需求的客户变量，当前产品拥有权，人口统计变量等， $j$  表示产品)； $V_j$  表示基于预测变量的顾客对产品  $j$  的需求 ( $V$  表示顾客对某种产品的需求， $j$  表示产品)； $Z_j$  表示购买  $j$  产品的抑制因素 ( $Z$  表示购买某产品的抑制因素， $j$  表示产品)，假如顾客对产品  $j$  的需求大于抑制因素，即  $V_j > Z_j$ ，则顾客就会购买产品  $j$ 。其中，NPTB 模型的构建分为四个步骤。

#### 3.1 数据准备

需要确定  $t$  时期和  $t + 1$  时期的数据， $t$  时期的数据作为自变量，包括： $t$  时期顾客购买的某产品数量， $t$  时期的人口统计变量（年龄，收入等），以及  $t$  时期公司针对每个顾客

所作的营销努力等； $t + 1$  时期的数据作为因变量，即顾客是否在  $t + 1$  时期购买该产品。

#### 3.2 选择统计模型

预测下一个顾客最有可能购买产品的统计模型包括：逻辑回归、多元逻辑回归、判别式分析和神经网络。其中，逻辑回归模型的普适性比较广，效率也较好，容易使用和解释；多元逻辑回归对于销售的预测适用于在企业全部的产品或者服务中，当顾客每次只能选择一种产品或者服务进行购买时，才会利用该模型进行预测；判别分析式算法适用于二元或者多元的预测，在研究中的应用也相当普遍；神经网络具有很强非线性的映射能力，能够捕捉非线性变量之间的关系。Knott 等人的文章实验结果证明神经网络模型的预测准确性要高于其他三种模型。

#### 3.3 预测模型

Knott 在文中以逻辑回归模型为例，顾客选择下一个购买产品的概率为：

$$P = 1 / (1 + e^{-V_{ij}}) \tag{式 1}$$

$$V_{ij} = \sum_{k=1}^k \beta_{ijk} X_{ijk} \tag{式 2}$$

其中， $X_{ijk}$  表示对顾客  $i$  的  $K$  预测变量的度量，预测了顾客  $i$  下一次是否将购买产品  $j$ ； $\beta_{ij}$  表示根据最大似然方法对数据估计的系数。如果有  $J$  种产品，就有个单独的逻辑回归模型，每个模型都有对应的产品。对每个模型的变量我们应用 0-1 变量，即表示顾客在下次消费中买或者不买产品  $j$ ， $X_j$  为自变量。

#### 3.4 确定目标顾客

根据统计模型得到的预测结果，依据预测的结果的高低对顾客重新排序，选择前 20% 的顾客作为银行实施交叉销售的目标客户<sup>[3]</sup>。

## 4 实例分析

论文的数据来源于某低成本航空公司的辅助产品订单数据和会员旅客信息, 根据研究的需要在信息比较完整的条件下, 随机抽取 10000 名会员旅客, 这些会员旅客的注册日期必须早于 2015 年, 且登录账号的次数大于 10 次, 确保有足够的订单信息。数据可以分为订单数据和会员信息数据, 会员 ID 可以对这两部分数据进行连接。

### 4.1 预测对象选取及变量定义

每个航空公司的辅助产品都有同质化现象的存在, 并且航空公司的定位不同其辅助产品的重心也不同, 由于数据的保密性, 本研究针对某低成本航空公司的八种辅助产品进行交叉销售的研究, 即超额行李、机上餐食、提前选座、机上 Wi-Fi、保险、接送机、贵宾休息室、景点门票。旅客在出行的过程中都有可能购买这八种辅助产品。

本研究所需的变量根据会员信息数据和订单数据进行如下提取:

#### 4.1.1 旅客最后一次乘机是否购买某项辅助产品

利用旅客的订单信息数据获取。如果购买过某项辅助产品即为 1, 如果没有购买过某项辅助产品即为 0, 作为因变量, 即为  $t+1$  时期的数据。下列各项为  $t$  时期的数据。

#### 4.1.2 旅客购买某项辅助产品的数量

利用旅客订单信息数据获取。将有关的数据导入 SPSS 软件利用求和功能, 求出旅客购买某项辅助产品的次数。

#### 4.1.3 旅客拥有辅助产品的种类数

根据旅客购买某项辅助产品的数量表进行计算。如果某一旅客购买过超额行李、机上餐食以及提前选座, 则该旅客辅助产品的种类数为 3。

#### 4.1.4 旅客性别、年龄、职业、会员龄

在旅客的会员信息表中包含旅客的性别、年龄、职业等变量, 其中性别用 1 (代表女性), 2 (代表男性); 旅客的年龄通过出生日期与统计截止日期之差计算而得; 职业用 1-10 表示, 1 代表医护工作者, 2 代表律师, 3 代表教育工作者, 4 代表金融, 5 代表商务人士, 6 代表公司职员, 7 代表个体, 8 代表事业单位人员, 9 代表学生, 10 代表离退休人员; 会员龄可以根据会员注册日期与统计截止日期之差计算而得。

#### 4.1.5 RFM 值

利用旅客订单数据和会员信息数据进行计算而得。

R: 旅客最后一次乘机距离统计日期截止间隔的天数;

F: 旅客在统计日期截止前乘机的次数;

M: 旅客购买增值服务消费的平均金额, 增值服务消费总金额除以数量。

### 4.2 预测过程

论文使用的软件为 Clementine, 该软件简单易操作而且具备许多建模技术, 如神经网络、决策树、聚类、关联规则等数据挖掘技术, 便于分析预测数据中的关系。在预测旅客购买和不购买的二元预测问题中, 利用 Clementine 二元分类器功能选择八种针对二元预测的统计模型, 其中包括逻辑回归模型、神经网络模型、贝叶斯网络模型、SVM 模型、C&R 树、决策列表模型、C5.0、QUEST 算法 [4-6], 利用这些模型预测旅客下次购买某种辅助产品的概率, 选取 3 种预测准确性较高的模型构建组合预测模型。以超额行李为例, 执行二元分类器节点, 如图 2 所示。根据八种模型预测准确性的高低, 选取了神经网络模型、C5.0、贝叶斯网络模型进行组合, 构建适用于超额行李的组合预测模型对旅客下次是否购买进行预测。模型构建过程如图 3 所示。

生成	图形	模型	构建时间 (分钟)	最大利润	最大利润发生比率 (%)	提升 (Top 30%)	总体精确性 (%)	使用的字段编号	曲线下方区域
<input type="checkbox"/>		贝叶斯网络 1	< 1	7,870	51	1.4	64.92	9	0.701
<input type="checkbox"/>		C5.1	< 1	7,751,534	66	1,364	64.9	8	0.895
<input type="checkbox"/>		神经网络 1	< 1	7,390	51	1,388	63.99	9	0.894
<input type="checkbox"/>		SVM 1	< 1	7,340	60	1.4	63.83	9	0.899
<input type="checkbox"/>		Logistic 回归 1	< 1	7,260	52	1,383	63.93	9	0.894
<input type="checkbox"/>		C&R 树 1	< 1	7,066,033	64	1,207	63.54	2	0.834
<input type="checkbox"/>		QUEST 1	< 1	7,066,033	64	1,207	63.54	2	0.834
<input type="checkbox"/>		决策列表 1	< 1	4,728,413	21	1,268	58.9	4	0.596

图 2 超额行李二元分类器结果

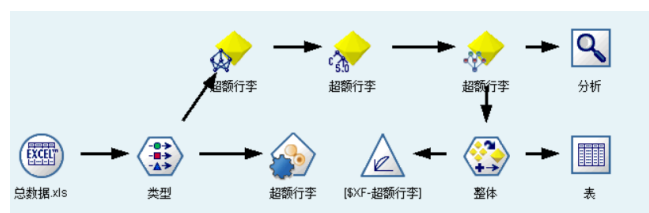


图 3 组合预测模型构建过程

### 4.3 预测结果及性能比较

#### 4.3.1 预测结果

按照超额行李预测的原理, 构建适用于其他七种辅助产品的组合预测模型分别对旅客下次购买的可能性进行预测, 预测结果如表 2 所示。

表2 组合模型预测结果

编号	下次购买超额行李的概率	下次购买机上餐食的概率	下次购买提前选座的概率	下次购买机上Wi-Fi的概率	下次购买保险的概率	下次购买贵宾休息室的概率	下次购买接送机的概率	下次购买景点门票的概率
1	53.90%	45.74%	67.97%	70.95%	65.85%	63.19%	62.03%	64.41%
2	60.35%	63.76%	64.38%	68.27%	61.89%	65.70%	66.35%	64.59%
3	52.03%	67.12%	65.91%	66.10%	62.53%	65.10%	64.20%	66.38%
4	51.05%	65.78%	65.63%	71.29%	65.87%	66.39%	63.11%	64.58%
5	60.40%	70.82%	76.15%	69.26%	70.49%	76.82%	71.72%	77.33%
6	61.85%	72.70%	70.99%	72.11%	70.95%	78.73%	71.41%	72.90%
7	18.30%	55.37%	54.42%	35.39%	52.52%	36.89%	35.95%	40.07%
8	65.04%	73.02%	72.73%	71.40%	71.23%	73.64%	70.93%	73.97%
9	54.48%	63.18%	67.85%	66.36%	65.31%	67.71%	66.42%	63.28%
10	54.76%	73.03%	70.26%	64.56%	76.68%	76.29%	73.62%	74.97%
...	...	...	...	...	...	...	...	...
7956	83.07%	84.63%	84.58%	80.55%	82.90%	85.61%	79.98%	80.75%
...	...	...	...	...	...	...	...	...
9999	58.68%	67.18%	46.33%	69.09%	65.02%	64.74%	64.69%	63.19%
10000	67.75%	76.07%	71.09%	74.72%	70.74%	74.65%	74.12%	71.65%

根据表2可以清楚地看到每位旅客下次购买某项辅助产品的概率，可以根据概率的高低将辅助产品选择合适的旅客进行推广。例如第7956名旅客接受超额行李的概率为83.07%，可以作为超额行李最佳的推广对象，或者选择概率大于50%的旅客进行推广，针对高概率的旅客可以在价格、服务方面进行创新，吸引旅客的眼光，促进消费。

#### 4.3.2 不同模型准确性比较

预测模型的准确率代表着预测结果的好坏，评估预测准确性的方法有两种：第一种是比较两个不同模型的预测准确率；第二种是与水平基准比较。论文选择第一种方法评估模型预测的准确性，根据t时期统计数据预测的结果，与t+1时期数据进行对比，从而确定模型预测的准确性。在NPTB模型原有的模型中选择神经网络模型、逻辑回归模型两种模型进行准确性对比，因为神经网络模型的预测结果相比较于其他模型拥有较高的准确性；逻辑回归模型简单、方便，应用的比较广泛。针对预测的结果为二元分布，而SVM模型擅长对二元结果进行预测。所以选择上述三种模型与论文构建的组合预测模型进行准确性对比。

在构建模型的过程中添加分析节点，执行分析节点就可以得到模型预测的准确性。为验证八种辅助产品组合模型的优越性，对处理后的数据分别选择神经网络模型、逻辑回归模型、SVM模型对旅客下次购买某种辅助产品的可能性进行预测，进行准确性对比。预测结果的准确性如表3所示。

表3 模型预测准确性对比

辅助产品种类	组合预测模型	神经网络模型	逻辑回归模型	SVM模型
超额行李	68.03%	63.71%	63.93%	63.83%
机上餐食	68.06%	64.92%	64.72%	64.92%
提前选座	69.22%	65.2%	64.92%	65.52%
机上WiFi	67.11%	64.24%	64.11%	64.46%
保险	68.34%	63.91%	64.15%	65.01%
贵宾休息室	68.15%	64.51%	64.68%	64.77%
接送机	68.17%	63.44%	63.25%	63.26%
景点门票	68.1%	63.8%	63.77%	64.35%

从上表中我们可以看出八种组合预测模型的准确性明显高于其他三种单项统计模型,同时也高于图2中二元分类器的预测准确性,这说明组合模型预测的结果能够帮助航空公司选择合适的旅客作为辅助产品的推广对象,增加航空的收益。

## 5 结语

论文在传统 NPTB 模型的基础上,扩展了统计模型的数量,并选取预测准确率较高的三种模型构建组合预测模型进行实证分析,便于航空公司精准定位旅客的需求。实验结果表明,论文构建的组合预测模型准确性优于其他单项模型。

## 参考文献

[1] Lizzy.2018 年全球航司辅助收入预计可达 930 亿美元 [EB/OL].

<http://news.carnoc.com/>,2018-11-30.

[2] Knott A, Hayes A, Neslin S A, Next-product-to-buy models for cross-selling applications[J],*Journal of Interactive Marketing*,2002,16(3):59-75.

[3] 杨帅,马少辉.基于 NPTB 模型的影视网站影片推广组合预测[J].*计算机与现代化*,2014(1):164-167.

[4] 顾静秋,吴华瑞,朱华吉.数据集浓缩研究综述[J].*计算机应用与软件*,2012,29(10):211-215.

[5] 谢邦昌.数据挖掘 Clementine 应用实务[M].北京:机械工业出版社,2008.