

DOI:10.12361/2661-3263-05-08-115620

# 大数据建模在保险电销系统中的运用

宋磊

对外经济贸易大学统计学院, 中国·北京 100105

**【摘要】**构建可运用于保险电销系统中的客户倾向性大数据模型, 并对模型效果进行评估分析, 确认模型提升效果较为显著后, 采用对照组方法检验客户倾向性模型的运用成效, 最终发现模型运用效果较好, 可作为保险电销系统的客户倾向性分析模型。

**【关键词】**大数据; 模型构建; 保险电销系统; 运用成效

## Application of Big Data Modeling in Insurance Electrical Marketing System

Song Lei

School of Statistics, University of International Business and Economics, Beijing 100105

**[Abstract]** A big data model of customer orientation that can be applied to the insurance e-business system is constructed, and the model effect is evaluated and analyzed. After confirming that the model has a significant improvement effect, the control group method is used to test the application effect of the customer orientation model. Finally, it is found that the application effect of the model is good and can be used as the customer orientation analysis model of the insurance e-business system.

**[Keywords]** Big data; Model construction; Insurance electrical marketing system; Application effect

对于保险企业来说, 大数据技术既是机遇, 也是挑战, 企业必须要结合自身实际, 利用大数据技术不断提高自身数据使用和分析能力, 并以此把握机遇, 应对挑战。在此背景下, 本文以保险电销系统为研究视角, 介绍一种客户倾向性大数据模型构建方法, 以期能够有效提高保险电销工作效率及效果。

### 1 大数据模型构建

#### 1.1 模型构建思路

客户倾向性大数据模型以现有客户群体为出发点, 结合客户倾向数据特点, 构建客户倾向性模型, 并配套完善的电销流程, 保障客户黏性, 提高客户续保率。设计初步续保率提升目标为10%。根据保险公司各类保险的经营发展特点, 选择客户刚需的车险作为研究对象, 并将车险细化为交强险和商业险, 深入挖掘客户商业险最佳险别组合, 以此来提升客户服务满意度, 保障保险续保率。

#### 1.2 变量界定及样本提取

在模型构建前, 需先实施模型目标变量选择和数据样本提取。

第一, 目标变量选择: 客户倾向性大数据模型以挖掘客户需求为出发点, 目标在于最大化提高客户产能。因此, 在参考现有研究成果综合分析后, 确认模型构建的目标变量分别为已有客户是否续保、已有客户预测件均、已有客户预测百产。

第二, 数据样本提取: 基于保险电销系统的保险订单信息, 选择2022年1月~11月到期客户的基本客户信息、车辆信息、电销任务信息、活动批次信息、订单信息、保险信息、险别组合信息、上年购买险别组合信息、出险信息、理赔信息、万里通注册信息、万里通积分及交易信息等。所有采集信息等比例分为训练样本和测试样本两部分样本数据, 其中训练样本占比80%, 总样本数量为1286404条, 所对应的续保成功率约为71.36%; 测试样本占比20%, 总样本数量为321601条, 所对应的续保成功率约为71.39%。训练样本和测试样本对应的续保成功率差异较小。

#### 1.3 变量分组

模型采用变量分箱法进行连续变量分组, 具体变量分组过程如下:

第一, 粗分箱: 根据样本数据中每个变量的分位数进行平均分组, 使分组后每个分组中的变量均等同。一般粗分箱会将样本

数据平均分为20个分组。

第二, 排序: 连续变量采用由小到大的顺序进行排序; 类别变量直接采用WOE排序。

第三, 细分箱: 连续变量比较相邻两组的WOE或者续保概率, 将WOE相近的合并成一组; 类别变量直接将WOE相近的合并成一组。

以某市客户年龄变量分组为例, 所有测试及数据中某市客户数据共有120310条, 年龄作为连续变量, 其需要由小到大进行排序, 并安装名单量平均划分为10组, 统计每组客户的产品组合购买概率、续保概率, 并以此为基础将相近两组进行合并, 重新统计续保概率, 最终形成4组变量数据。

#### 1.4 预测变量筛选

采用IV值分析法实施预测变量筛选, 具体筛选中应先对WOE(证据权重)进行说明。作为原始自变量的一种编码形式, WOE需要对变量进行分箱处理, 并在分组后, 采用以下公式对获取第*i*组的WOE。

$$woe = \ln\left(\frac{py_i}{pn_i}\right) = \ln\left(\frac{ny_i/ny_r}{n_i/n_r}\right) \quad (1)$$

式中:  $\frac{py_i}{pn_i}$  和分别为组中响应客户/未响应客户占有所有数据样本中响应客户/未响应客户的比例;  $\frac{ny_i/ny_r}{n_i/n_r}$  和分别为组中响应客户和所有样本中响应客户数量, 和分别为组中未响应客户和所有样本中未响应客户数量。

对公式(1)进行变化拆除后, 可获取到以下公式。

$$woe = \ln\left(\frac{py_i}{pn_i}\right) = \ln\left(\frac{ny_i/ny_r}{n_i/n_r}\right) = \ln\left(\frac{ny_i/n_i}{ny_r/n_r}\right) \quad (2)$$

同时, 分组*i*也具备对应的IV值, 具体计算公式如下。

$$IV_i = (py_i - pn_i) \times woe = (py_i - pn_i) \times \ln\left(\frac{py_i}{pn_i}\right) \\ = (ny_i/ny_r - n_i/n_r) \times \ln\left(\frac{ny_i/ny_r}{n_i/n_r}\right) \quad (3)$$

获取分组 IV 值后, 可对整个变量的 IV 值进行计算, 具体计算公式如下。

$$IV = \sum_i^n IV_i \quad (4)$$

式中: n 为变量分组数量。

通常情况下,  $IV < 0.02$  为无相关;  $0.02 \leq IV < 0.1$  为弱相关;  $0.1 \leq IV < 0.3$  为中等相关;  $IV \geq 0.3$  为强相关。除 IV 值分析法以外, 研究中还采用 VIF 法 (方差膨胀因子), 该方法可以判断主要变量之间的多重贡献性, VIF 值越大, 说明变量之间的共线性越严重。通常情况下,  $0 < VIF < 4$  为不存在共线性;  $4 \leq VIF < 10$  为可能存在共线性;  $10 \leq VIF < 100$  为存在较强共线性;  $VIF \geq 100$  为存在严重共线性。

### 1.5 模型构建

完成变量的 IV 值和 VIF 检验后, 取  $IV > 0.1$ ,  $VIF < 10$  的变量进入模型进行建模筛选, 并通过逐步回归法构建逻辑回归模型, 根据模型分析自变量与因变量之间的关系。此过程中所涉及的模型变量分组包括车龄分组、年龄分组、年收入分组、座位数分组、排量分组、客户性别、盗抢险购买标签分组、座位险购买标签分组、主险购买个数分组、附件险构建个数分组、商业险出险次数分组、交强险出险次数分组、万里通90天内活跃情况分组、客户赔付金额分组、客户标识分组、销售任务时间标识分组等。

## 2 大数据模型效果评估

根据训练集数据构建 LIFT 图, 具体结果如下。

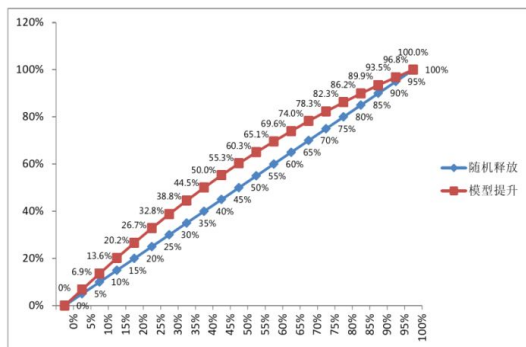


图2 训练集数据 LIFT 图

如图2所示, 前50%的客户已经覆盖60%以上的续保客户, 而训练集数据所对应的客户续保率为71.36%, 总体来说模型提升效果较为显著。

再根据测试集数据构建 LIFT 图, 具体结果如下。

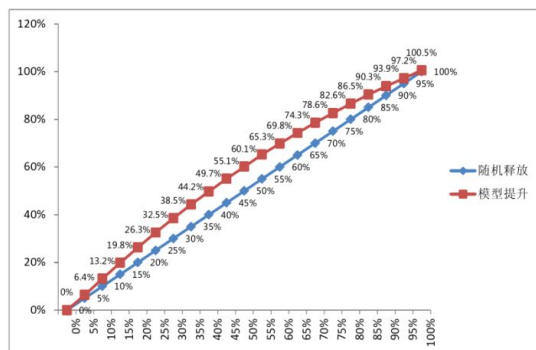


图3 测试集数据 LIFT 图

如图3所示, 测试集数据中前50%的客户也覆盖了超过60%的续保客户, 说明模型整体稳定性较强。

## 3 大数据建模在保险电销系统中的运用成效

### 3.1 大数据建模的运用

车险电销业务基本要素包括资源、团队、产品及影响方案。现有的客户购买概率预测模型多以资源为基础进行预测分析, 但仅考虑资源要素所获取的预测结果精准性较差, 所以需

要在预测模型中引入其他基本要求, 以提升预测精度。过往预测模型主要采用逻辑回归分析法进行资源筛选, 重点考虑优质资源 (LIFT 图中前80%资源), 舍弃劣质资源 (后20%资源), 以提高资源利用成功率, 降低保险电销成本。

但已有客户均是保险公司的重要客户, 是保险公司多年发展积累下的重要资源, 对其进行舍弃势必会导致保险公司现有资源被浪费, 所以具体预测分析中不应资源进行筛选使用, 应尽可能提高资源利用率。据此, 具体预测中以原有概率预测模型为基础, 结合件均预测模型、百产预估模型、推荐方案博弈模型, 根据客户续保百产目标, 保障预测结果与推荐方案的客户续保百产最大化。

具体应用中, 概率预测模型以现有概率预测模型为基础, 结合客户续保主要影响辨认成熟, 根据客户上年度商业险别组合、车龄、收入、出险次数等分组对客户进行微分组处理。以某市为例, 上年度商业险别组合主要为三者车损, 实际收入为02-A, 车龄为1年, 出险次数为1次的客户均为1个微分组, 针对微分组内客户所购买的商业险别组合实施客户购买概率预测。

以概念预测模型为基础, 根据微分组中客户购买商业险别组合预测险别组合件均, 构建件均预测模型。最后以每个微分组客户购买商业险别组合预测件均为基础, 预测分析每个微分组客户的百产, 优先选取百产最高商业险别组合, 形成推荐方案博弈模型。以上模型预测过程均为客户倾向性预测模型。

### 3.2 模型测试及结果

具体模型测试主要包括资源优化、系统改造、销售策略三部分内容。

第一, 资源优化: 所采用的资源包括车龄、车型、车架、上年度购买商业险别组合等, 对所有资源进行分组, 按照等比例分层抽样法将客户分为测试组、对照组两部分, 并确保两组客户之间无差异, 避免因客户差异所导致的模型测试精准性不足问题。

第二, 系统改造: 根据模型预测结果向公司技术部门、管理层人员提出系统优化需求, 并在条件允许情况下对现有保险电销系统的客户倾向性模型进行优化调整, 获取最优商业险别组合, 将其提供给电销人员用于客户电销服务。

第三, 销售策略: 根据预测结果对过往使用的销售推荐策略进行优化调整, 并以此为基础对所有电销人员进行针对性培训, 要求电销人员对客户开展电话销售时优先按照客户倾向性模型预测结果进行产品推荐。

具体模型测试过程为2022年4月和5月两个月, 4月和5月两个月测试中百产政府均在9%以上, 并且相较于对照组, 测试组平均拨打次数更少, 说明测试组可以在更少拨打次数情况下, 有效提高成功率, 证明客户倾向性模型测试效果较为优秀。

## 4 结束语

综上所述, 本文介绍一种可用于保险电销系统的客户倾向性模型构建方法, 此模型采用大数据建模技术, 可实现客户最佳商业险别组合预测及推荐, 并结合具体运用测试确认此模型可有效提高预测精度, 并提升保险电销百产效果, 具有较强应用价值, 可在后续保险电销系统优化调整中进行参考应用。

### 参考文献:

- [1]周菲, 吴美珍. 大数据时代电话营销在保险业中的运用研究——以杭州地区人寿保险为例[J]. 现代商业, 2019(1): 52-53.
- [2]陈红玲. 大数据时代人寿保险公司营销策略优化转型研究[J]. 现代商业, 2021(30): 24-26.
- [3]焦丽杰. 大数据视角下保险领域精准营销解读[J]. 商业2.0(经济管理), 2021(14): P. 1-1.

### 作者简介:

宋磊 (1986-) 男, 汉, 湖北武汉人, 大学本科, 研究方向: 统计学。对外经济贸易大学统计学院在职人员高级课程研修班学员。