

客户流失管理研究现状及展望^①

张珠香, 骆念蓓

(福州大学 经济与管理学院, 福州 350108)

摘要: 梳理了客户流失和客户流失管理的定义, 客户流失问题的研究内容、应用场景, 客户流失预测算法及特征选择方法, 模型评估的常用技术与度量等方面的研究现状, 指出当前研究的不足, 并提出未来的研究方向。

关键词: 客户流失管理; 研究现状; 展望

引用格式: 张珠香, 骆念蓓. 客户流失管理研究现状及展望. 计算机系统应用, 2017, 26(12): 9-17. <http://www.c-s-a.org.cn/1003-3254/6086.html>

Current Situation and Prospect of Customer Churn Management

ZHANG Zhu-Xiang, LUO Nian-Bei

(School of Economics and Management, Fuzhou University, Fuzhou 350108, China)

Abstract: This paper summarizes the literature about the following aspects: the definitions of customer churn and customer churn management; research contents and application scenarios of customer churn issues; customer churn prediction algorithms and feature extraction methods; the evaluation technologies and measurements. In the end, we point out the shortcomings of the current research and put forward some future research directions.

Key words: customer churn management; current situation; prospect

随着信息技术的高速发展, 客户对产品或服务的信息获取越来越充分, 顾客的需求也因此更加多样化。尤其电子商务的兴起, 客户对产品有了更多的选择, 激烈的市场竞争, 使得客户流失已经成为许多企业不可避免的问题^[1]。而开发一个新客户的成本往往比保留一个老客户的成本高得多^[2], 例如银行客户流失率每降低 5%, 利润就会增长 85.5%^[3], 这一数据根据行业的不同而有所不同。为了应对客户流失, 很多企业从过去的以产品为导向向以客户保留、减少流失做战略转型^[4]。

虽然投放大量广告或不断优化产品是简单易行的保留措施, 但是, 这种毫无针对性的客户保留会使企业面临浪费资源的风险^[5], 有针对性地对潜在流失客户进行挽留已经成为企业客户关系管理的重要内容。大数据时代的到来使得客户交易数据更易被记录搜集, 利用数据建立预警模型来识别潜在的流失客户, 为企业

更有效率地理解客户消费行为与应对市场需求变化提供了新的路径。

本文从多个方面梳理了客户流失管理的现状研究, 并指出当前研究的不足, 提出了未来可研究的方向, 为深入研究客户流失理论、方法和应用提供参考。

1 客户流失和客户流失管理的定义

客户流失是指客户放弃使用某种产品或服务, 转而使用市场上另一竞争企业的产品或服务^[6,7]。Hadden 等^[1]将流失的客户分为两种, 一种是自愿流失, 一种是非自愿流失。非自愿流失是由于滥用服务或者未对服务付费等原因被企业撤销的用户, 非自愿流失的客户容易识别。而自愿流失是指客户主动决定和这个企业结束关系, 转向和另一个企业合作的流失行为, 自愿流失的客户是企业流失管理的重点对象。对于不同行业,

^① 收稿时间: 2017-03-10; 修改时间: 2017-03-27; 采用时间: 2017-03-29

流失客户的定义是不同的。Larivière等^[8]将客户关闭其金融账户时视为流失。Coussement等^[9]和Huang等^[10]是把电信客户销号作为离网流失客户。张玮等^[11]将电信客户报停、预销、强关、销号等八种状态确定为离网状态。Hadiji等^[12]将7天内没有玩该款网络游戏的玩家视为流失客户。Jahromi等^[5]把某电子商务网站半年内没有任何行为的客户视为流失客户。与基于消费频率的定义方法不同, Miguéis等^[13]对于零售公司的流失客户是根据消费金额定义的, 如果客户在第 t 个时段内的购买金额小于第 $t-1$ 个时段消费总额的40%, 视为流失客户。

客户流失管理是指通过客户的历史信息来对客户将来的流失行为进行预测, 通过计算客户流失概率, 将高流失概率的客户作为客户保留战的对象^[9]。Datta等^[14]于2000年提出客户流失管理的框架, 主要包括研究数据的选取、数据和业务的理解、特征选择、预测模型的建立和模型验证。Lima等^[15]将客户流失管理过程分解为6个步骤: 业务理解; 数据理解; 数据预处理; 建模; 评估; 部署。文献^[5,16,17]认为高流失概率的客户不一定是给企业带来最大收益的客户, 仅依据流失概率选择目标保留客户不一定会增加企业盈利, 因此, 流失概率预测只是客户流失管理的一部分。总体而言, 学者们对于客户流失管理的内涵有较为一致的认识: 客户流失管理是一个复杂的过程, 而核心工作就是对潜在的流失客户进行识别并采取保留措施。

2 客户流失问题的研究内容

客户流失问题是一个混沌、时滞、非线性、非对称的复杂系统^[18], 并具有以下特点^[6,19]: (1) 属于典型的二分类问题, 即流失与非流失。(2) 分类结果要有较高的准确性。(3) 客户数据集极端不平衡, 流失客户和非流失客户样本数目往往不是一个数量级。(4) 客户行为数据是海量的, 而且维数比较高, 数据类型涉及结构化、半结构化和非结构化, 并含有噪声。

在预测方面, 大多数学者通过使用统一数据集比较多个模型的预测效果, 以确定最好的模型。Nie等^[3]对中国某银行的信用卡客户数据建立Logit和决策树模型, 用以识别其潜在流失客户。Vafeiadis等^[20]采用UCI数据库中电信客户的公共数据集, 比较了多层人工神经网络、决策树、SVM、朴素贝叶斯、Logit、Adaboost等算法在客户流失预测方面的效果。Chen等^[21]

在SVM、MK-SVM基础上提出分层多核支持向量机(H-MK-SVM), 并使用多个零售和电信的真实数据集进行实证比较三种模型的效果。De等^[22]使用银行、电信、零售、邮购等真实数据集比较RotBoost、Rotation Forest、Bagging、随机森林、随机子空间、CART、C4.5的预测效果。Moeyersoms等^[23]着重研究如何处理有穷多个不同值的离散变量(例如邮政编码), 使之纳入预测模型中, 结果显示, 纳入这些高基数属性的C4.5、Logit、SVM模型精度均得到提高。Coussement等^[9]比较了不同数据预处理方式对Logit模型效果的影响, 从而选出最佳的数据处理方式。Amin等^[24]比较了不同的过取样方式对解决正负样本不平衡问题的效果。

在流失原因方面, 目前的文献多采用定性和实证的方法进行研究。Jamil等^[25]认为预测模型的准确度可能并不是企业最关注的, 企业更想知道顾客是因为什么原因流失。Keaveney^[26]较早地对客户流失问题进行研究, 通过对服务行业500个客户进行调查, 得出重要结论: 导致客户流失的最主要原因是不好的服务质量, 价格反而不是最主要的因素。此后, 便出现了很多客户流失原因的研究, 考虑因素包括客户的统计学特征、价格、市场竞争程度、产品升级、客户的社会地位以及客户满意度和服务质量等^[27]。盛昭瀚等^[28]认为产生客户流失危机的内部因素包括企业创新能力下降、经营不善、观念滞后、战略决策失误等, 外部因素包括政治、经济政策、科技发展、市场需求和竞争条件等外部经营条件的突变或恶化。内部因素可控性较强, 是企业自身矛盾的结果; 外部因素几乎难以控制, 是经营环境矛盾的结果, 但客户流失危机往往又是二者综合的结果和体现。Hadden等^[1]分析了导致客户流失的内因和外因, 例如, 竞争对手的产品升级了或是价格上更优惠了; 自己的服务不到位导致顾客的体验差等等。于小兵等^[29]根据电子商务的交易过程, 提出客户流失指标体系, 包括技术、网页、销售产品、物流、服务5个一级指标和10个二级指标, 并用模糊直觉法评估各流失原因的重要性。李婷婷^[30]采用实证的方法分析技术、物流、产品、售后服务、网络安全5个因素对客户流失的影响程度。

在挽救策略方面, 学者们多采用数学建模的方法进行研究。在流失预警系统报告高危流失群后, 决策者并不会对所有的潜在流失客户进行挽留, 也不会简单

的依据流失概率高低选择高概率的客户进行挽留,因为企业的挽留利润不仅与流失概率有关,也与客户的价值、客户被成功挽留的概率、挽救人数、企业的成本等息息相关^[17]. Neslin 等^[16]基于流失概率、流失成本等参数构建了挽救利润函数. Glady 等^[31]基于客户生命周期构建了挽救利润函数. 胡理增等^[32]研究了用于挽救的总经费有限的条件下,如何确定各组客户的挽救比例、费用和次序,实现全体流失客户价值总和的最大化. 之后,胡理增等^[33]又研究了无资源约束下的最优挽救方案数学模型. 罗彬等^[34,35]建立了客户挽留周期计算模型、客户挽留价值计算模型、客户流失挽留评估模型以及客户挽留顺序选择模型. 与上述文章思路不同的是,Chen 等^[21]和 Tamaddoni 等^[36]在预测模型建立时,就使用了挽留最大利润 (MP) 作为对模型的评估标准.

总的来说,客户流失问题涵盖了流失客户的预测识别、流失原因分析以及挽救策略三方面内容,但是,目前对于客户流失问题的研究重点主要放在提高预测模型准确度上,对流失原因及挽救策略的研究较少.

3 客户流失问题的应用场景

3.1 电信业

Coussement 等^[9]采用欧洲电信运营商的 30104 个客户数据,其中,流失客户占 4.52%,数据包含 956 个变量,涉及客户通话行为、客户和运营商的交互行为、套餐订阅以及人口统计学特征,通过 Logit 模型进行客户流失预测. Huang 等^[10]使用 104199 个客户的电信数据集,其中流失客户占比 5.8%,数据包含 121 个变量,涉及客户的人口统计学信息、账号信息、通话信息,通过基于 K-means 的混合模型进行预测. 张玮等^[11]对某电信公司 184761 个客户数据,其中流失客户占比 7.3%,数据包含 15 个变量,通过 CART 与 Boosting 集成后的模型进行客户流失预测. Masand 等^[37]以 GTE 公司 20 个最大的手机通信市场的客户为研究对象,使用简单回归、最近邻分类、决策树和神经网络进行客户流失预测. 丁君美等^[38]收集了某电信公司 2013 年 9 月至 2014 年 2 月的 7913 条客户数据,其中,流失客户占比 3.3%,通过改进的随机森林算法对客户流失进行预测. 罗彬等^[18]通过将 BP 神经网络、RBF 神经网络、ELMAN 神经网络和广义回归神经网络 4 个子分类器进行线性集成,并采用人工蜂群算法优化线性组合的权重,以此

对某电信企业 20000 个客户数据进行分析.

3.2 金融业

Nie 等^[3]抽取了某银行信用卡 5456 条,其中,非流失客户占 8.1%,流失客户占 91.1%,包含 135 个变量,涉及信用卡持有人信息、信用卡信息、交易信息、异常使用信息等,使用 Logit 和决策树进行预测. Larivière 等^[8]使用生存分析研究比利时金融服务机构的客户流失问题. 应维云等^[6]使用改进的支持向量机对深圳市某银行个人信贷部的客户信贷数据进行研究,客户数据共计 12 万条,每条记录有 16 个变量,且包括文本型变量.

3.3 电子商务

Gordini 等^[39]研究了意大利某公司在 2013 年 1 月至 2014 年 1 月期间的 80000 个客户数据,其中,流失客户占比 10%,数据变量涉及注册信息、交易信息、访问信息,使用改进的支持向量机进行预测. Yu 等^[40]利用中国某电子商务网站 50000 个客户的注册信息、登陆信息、交易信息、网站日志信息建模,通过改进的支持向量机进行预测. Jahromi 等^[5]使用 Logit、决策树、Boosting 等多个模型研究了澳大利亚某 B2B 电子商务平台的客户流失问题. 由于 B2B 数据的获取渠道比 B2C 有限,因此在客户流失方面的研究集中在 B2C 方面^[5].

3.4 其他

还有一些例如人才流失研究^[41]、网络游戏方面的客户流失研究^[12,42]、保险业的客户流失^[43]等.

总体上看,目前客户流失的研究多集中于电信、金融、电子商务,尤其以前两者居多,而最新的一些研究开始将客户流失拓宽至网络游戏等新领域. 此外,客户流失问题涉及的客户数据规模大,维数高,预测复杂程度高,因此,通常需要使用多个模型进行预测.

4 客户流失预测算法及特征选择方法

4.1 客户流失预测算法

由于客户流失问题的复杂性,传统的经验时间阈值法、RFM 法等均难以奏效, Schmittlein 等^[44]于 1987 年提出了预测客户交易行为的概率模型组 Pareto/NBD 模型,是首个考虑到客户流失现象的客户重复购买预测模型^[45]. 之后, Fader 等^[46]对其进行改进,提出了 BG/NBD 模型. 客户流失预测研究开始于 20 世纪 90 年代,我国学者在 21 世纪初开始研究客户流失预测方法^[47],目前用于客户流失预测的算法主要包括

以下几种:

(1) 基于传统统计学的方法. 传统统计学分类器包括 Logit 模型、线性/二次判别器. 前者是通过假设服从 Logistic 分布, 后者是通过假设服从高斯分布, 二者均是通过极大似然估计求出参变量, 估计条件概率. 朴素贝叶斯也是通过求条件概率进行分类. 基于传统统计学的方法都是较为基础的算法.

(2) 基于递归分割方法. 递归分割的典型代表是决策树, 属于机器学习领域的方法. 主要包括 ID3、J48、C4.5、C5.0 等算法.

(3) 基于统计学习理论. 典型代表是支持向量机. SVM 从数学的角度看与人工神经网络密切相关, 但其起源于统计学习理论.

(4) 基于人工智能的预测. 典型代表是仿照人脑学习的人工神经网络, 其中, 多层感知器是最流行的人工神经网络.

(5) 基于集成分类器的预测. 集成分类器是利用各个子分类器之间的互补性, 按照一定的规则将各子分类器的输出结果进行融合, 以提高其性能^[22]. Bagging, Boosting 和随机森林是典型的集成方法. 不同集成模型的共同点在于都是由一些基本的模型构成, 通过集成实现将弱分类器转变为强分类器, 不同之处就在于使用的基本模型的不同以及集成规则的不同, 因此有线性判别法的集成方法^[48,49], 决策树的集成方法^[50], 支持向量机的集成方法^[20], 神经网络的集成方法^[18], Hu 等^[51]将决策树、贝叶斯网络和神经网络等多个分类器共同集成, Kim^[52]将 Logit 和 ANN 共同集成等. 另外还有使用旋转集成的算法进行客户流失预测, 如旋转森林和 RotBoost^[22].

(6) 其他方法. 除了上述分类, 还有很多其他方法, 如 Amin 等^[7]用模糊粗糙集做客户流失预测的研究. Huang 等^[10]提出基于 K-means 的有监督和无监督混合模型. Lu^[53]和 Larivière 等^[8]使用生存分析估计客户即将流失的时间. Adebisi 等^[54]使用马尔科夫链对尼日利亚的电信客户流失进行预测. 袁旭梅等^[55]根据 RFM 标准划分顾客群体, 利用马尔科夫链构建动态 CRM 模型. 文献^[56-58]使用社会网络分析方法, 根据客户与客户之间的通话记录构建电信客户关系网络, 研究客户流失. 文献^[59-61]使用控制图法研究电信客户流失. Chen^[62]提出控制图法与贝叶斯分层模型结合的混合模型, 动态监测电子商务客户流失.

哪一种预测模型精度更高? 这个问题目前还没有定论. 客户流失的模型精度通常是基于特定的样本数据而言, 同样的模型在不同的研究中会有不一样的表现. 因此, 学者们对于哪一种预测模型精度更高没有一致的看法. Nie 等^[3]认为 Logit 回归优于 J48 决策树. Mozer 等^[63]认为 Adaboost 和 boostd neural network 比其他模型好. 尽管支持向量机是处理高度非线性分类问题的高级方法^[64], 但是它是一种黑箱模型, 只能给出分类结果, 而无法给出模型所学习到的知识^[25]. Masand 等^[37]使用简单回归、最近邻分类、决策树和神经网络进行客户流失预测, 结果显示神经网络的效果最好. Kirui 等^[65]认为朴素贝叶斯和贝叶斯网络在召回率上比 C4.5 决策树好. Rehman 等^[58]的结果显示社会网络分析比多种决策树算法精度高了 8%-10%. Amin 等^[7]认为粗糙集处理未知分布的数据比机器学习更有效. Gordini 等^[39]认为其改进的支持向量机在处理有噪声、不平衡、非线性的电子商务数据时比 Logit、神经网络、经典的支持向量机效果更好. Neslin 等^[16]认为 Logit 和决策树在流失预测的实际应用中最为广泛, 在流失预测中应作为基础模型先行试验. Chen^[62]认为现有的很多模型在建模时是基于某段时间内的客户信息, 相对静态, 因此需要定期建模, 而他提出的控制图法与贝叶斯分层混合模型可以实现客户流失的动态监测.

总之, 用于客户流失预测的算法很多, 既包括统计学的方法, 也包括机器学习的方法, 还有一些其他方法. 每一种方法都具有自己的特点从而适用于不同的条件, 表 1 对最常用的几种预测方法进行了总结.

4.2 特征选择方法

由于客户流失数据集里包含了很多无关的多余的变量, 过多的变量会带来多重共线性、过拟合、过参数化甚至欺骗性的结果等问题^[62], 所以在建立预测模型之前, 需要进行变量选择. De 等^[22]使用主成分分析、独立成分分析、稀疏随机投影三种特征提取方式研究模型. Nie 等^[3]先对变量做相关性分析, 将相关性程度很强的变量做删除, 再用逐步回归筛选变量. Masand 等^[37]先使用决策树进行属性选择, 再使用遗传算法降维. Amin 等^[24]使用 mRMR 进行特征选择. Fathian 等^[66]认为使用主成分分析进行预处理后的数据更适合建模. 文献^[67-69]通过粗糙集方法约简模型解释变量. 张玮等^[11]利用直方图检验和卡方检验相结合的方法对模型变量进行筛选.

表1 客户流失常用预测方法

方法	特点	适用范围
LR	逻辑回归(LR)是实际应用最为广泛的模型之一,易于理解和解释,可用于概率预测,也可用于分类,但仅能用于线性问题,对多重共线性反应较为敏感,需要事先对离散型变量设置虚拟变量,分类准确度也一般。	适用于线性问题的概率预测或分类,且对于准确度要求不高的情形。
NB	朴素贝叶斯(NB)是基于传统统计学的条件概率公式,易于理解和解释,但其假定各个变量之间相互独立,而这一点在现实中很难满足,因此,在属性个数比较多或者属性之间相关性较大时,分类效果不好。	适用于属性个数较少或属性之间相关性较小的情形。
DT	决策树(DT)是实际应用最为广泛的模型之一,容易提取规则,可同时处理离散型和连续型变量,但对于缺失数据处理起来比较困难,且对于大规模数据运行效率较低,易产生过拟合问题。	适用于样本量不大且数据完整的情形。
SVM	支持向量机(SVM)的分类准确度高,擅长处理高度非线性分类问题,但它是一种黑箱模型,难于解释,且内存消耗大。	适用于小样本、非线性、数据高维的情形。
ANN	人工神经网络(ANN)的分类准确度高,具有很强的非线性拟合能力和记忆能力,对噪声数据具有较强的鲁棒性和容错能力,但它需要训练大量的参数,且训练过程无法观察,也是一种黑箱模型,输出结果难以解释。	适用于非线性、数据高维的情形。
Bagging/Boosting/RF	集成分类器的准确度高,其中, Bagging和随机森林(RF)属于并行式集成, Boosting属于串行式集成。一般来说,集成分类器的的预测能力比单个分类器好。RF对于不平衡的数据集可以平衡误差,但是在数据噪声比较大的情况下会产生过拟合。Boosting对噪声数据和异常数据敏感,运行速度较慢。	RF可用于处理分类不平衡、高维数据集。Boosting适用于对准确率要求高且样本量不大的数据。
Rough Set	粗糙集具有较强的不精确信息处理能力,用来处理未知分布的数据比机器学习更有效,但多用来筛选模型属性。	适用于不完全的、有噪声的、模糊的数据。
SNA	使用社会网络分析(SNA)预测客户流失视角较新,它考虑了客户之间的社会关系,认为客户的行为会受到身边人的影响,比将客户看作是独立个体预测更为准确,但通常在电信业使用较多,因为关系网络很难建立,而电信业可以通过通话记录建立客户之间的关系网络。	适用于客户关系网络有必要建立并且有条件建立的情形。
马尔可夫链	将马尔可夫链应用于客户流失预测是较新的做法,既可以用来预测客户流失转移,又可以预测市场占有率,且预测效果良好。	可用于客户群的转移预测。

5 模型评估的常用技术与度量

5.1 模型评估的常用技术

评估分类器性能的常用技术有3种。保持方法是最常用的方法,在这种方法中,给定的数据集随机地划分成两个独立的集合:训练集和测试集。使用训练集建立模型,使用测试集检验模型。例如 Vafeiadis 等^[20]使用 2/3 的数据作为训练集, 1/3 的数据作测试集。Coussement 等^[9]将原始数据分为学习集(50%)、选择集(20%)、验证集(30%),其中,学习集用于建立模型,选择集用于预处理和建模过程的参数选择,验证集用于检验模型性能。

k 折交叉验证法是将初始数据随机地划分成 k 个互不相交的子集或“折”,每个折的大小大致相等。 $k-1$ 份作为训练集,1份作为验证集,依次轮换训练集和验证集迭代 k 次,总准确率取每次迭代准确率的平均值。例如 Vafeiadis 等^[20]采用 100 折交叉验证; Moeyersoms 等^[23]采用 10 折交叉验证; Huang 等^[10]采用 5 折交叉验证; De 等^[22]采用 2 折交叉验证。

与上面两种方法不同,自助法是用小样本估计总

体值的一种非参数方法,它是从给定训练集中有放回地均匀抽样。最常用的一种是 632 自助法:假定给定的数据集有 d 个元组,那么从该数据集中有放回地抽样 d 次,产生的自助样本集作为训练集,没有进入训练集的元组作为测试集,计算该次迭代的准确率,继续这样迭代 k 次,模型的总体准确率用下式估计:

$$Acc(M) = \sum_{i=1}^k (0.632 \times Acc(Mi)_1 + 0.368 \times Acc(Mi)_2)$$

其中, $Acc(Mi)_1$ 是自助样本 i 得到的模型用于测试集 i 的准确率, $Acc(Mi)_2$ 是自助样本用于原数据元组集的准确率。Holtrop 等^[70]和丁君美等^[38]即是采用了 Bootstrap 抽样。

5.2 模型评估度量

最常用的评估度量包括精度(Precision)、召回率(Recall)或称敏感度(Sensitivity)、特异性(Specificity)、准确率(Accuracy)、F值(F-measure); ROC曲线和AUC值;提升度(Lift)和Gini系数。

虽然 Accuracy 在模式分类中是最主要的衡量指标,但是在客户流失方面,AUC和Lift是常用的指标^[22]。

Masand 等^[37]指出虽然 Accuracy 常用来评价模型的性能,但是并不适合市场环境,因此他对于模型的评估采用了 Lift 而不是 Accuracy. Accuracy 是需要指定固定阈值计算而得,而 AUC 是不受阈值变化影响的,相比 Accuracy, AUC 具有较强的鲁棒性,因此能较为客观衡量分类模型的性能^[71,72]. Chen 等^[21]也认为 Accuracy 不适合用来评估正负样本不平衡的数据. 尽管如此, Hand^[73]提出了 AUC 存在最大的缺点是对于不同的分类器没有采用统一的误判成本分布,对于同一个问题,正样本误判成本与负样本误判成本的倍数关系是固定的,而 AUC 却在不同的分类器上采用了不同的倍数关系.

在大部分研究中,学者们通常使用上述多种指标共同评估分类器对所用样本的分类能力. Vafeiadis 等^[20]采用 Precision、Recall、Accuracy、F-measure 比较 SVM、BPN、决策树、朴素贝叶斯、Logit 及 SVM 集成、BPN 集成、决策树集成的模型效果. Coussement 等^[9]使用 AUC 和 Top-decile Lift 评估 Logit 模型效果. De 等^[22]使用 Accuracy, AUC 和 Top-decile Lift 三个指标衡量模型效果. Neslin 等^[16]和 Holtrop 等^[70]使用 Top-decile Lift 和 Gini 系数作为衡量指标. Lima 等^[15]使用 Accuracy、Sensitivity、Specificity、AUC 作为评价标准. Jahromi 等^[5]采用 ROC 和 Lift 对 Logit、决策树、Boosting 进行评价. 夏国恩等^[74]使用 Accuracy、Precision、Recall、Lift 评估 SVM、人工神经网络、C4.5 决策树、Logit 回归、贝叶斯分类器的分类效果.

除此之外,还有一些其他的评估标准,如 Hand^[73]于 2009 年提出了 H-measure 的评估指标. 丁君美等^[38]认为在不平衡分类中,分类精度并不能作为衡量分类性能的评价指标,因此提出了生命价值比 (LVR) 的评估指标. Chen 等^[21]除了使用 Accuracy、Sensitivity、Specificity、AUC、Lift 指标,同时考虑 H-measure、MP(最大利润)对 H-MK-SVM、MK-SVM、SVM 模型进行评价. Tamaddoni 等^[36]认为现有的评估标准都关注模型准确率,却忽视了企业利润,因此,他除了使用 Lift,也使用了 MP(最大利润). Lessmann 等^[75]除了使用 Accuracy, AUC, Gini 系数,还将 H-measure, Brier Score, KS 统计量作为评估指标.

6 不足和未来的研究方向

(1) 流失客户的定义研究. 流失客户的定义常见有基于消费频率和基于消费金额两种方法. 对于不同行

业,流失客户的定义不同,电信业的客户属于契约型客户,其流失定义较为明确,终止契约即可视为流失,但是对于诸如电子商务、零售业等非契约型应用场景,无法准确判断客户是否真的流失,流失客户的定义全凭研究者决定,导致研究结果带有一定主观性. 现有文献中对如何合理定义流失客户还未总结出系统的经验指导. 因此,探索如何对流失客户进行合理定义有其必要性.

(2) 客户流失原因及挽救方案的研究. 目前对于客户流失问题的研究主要集中在提高预测模型准确度上,对流失原因及挽救策略的研究较少. 但是,一个好的流失管理系统不应只预测潜在流失客户,也应包括对流失原因的分析以及挽救措施的实施,因此,客户流失原因及挽救是客户流失管理的重要研究方向.

(3) 客户流失问题的方法研究. 从研究方法上看,目前用于流失客户的预测模型很多,但是大多数预测模型都只是对单个客户的流失行为进行预测,对特征客户群的预测研究较少. 而流失原因分析多以定性和实证分析为主,缺乏量化研究. 挽救策略多采用数学建模的方法,但由于挽救行动的利润计算涉及的变量较多,包括客户生存时间、流失速率、流失概率、潜在流失数量、客户的价值、客户被成功挽救的概率、挽救人数、企业的成本等等,很容易遗漏变量导致研究结果失真.

(4) 客户流失预测模型的评估度量研究. 对于客户流失预测模型的评估度量有很多,Accuracy、AUC 和 Lift 是常用的指标,但这些传统的指标均聚焦于准确性上,忽视了企业利润,而企业进行客户流失管理就是为了实现利润最大化,因此,探索一些新的评估度量,例如将企业利润因素考虑到度量中,使得模型效果的评判更加有意义.

(5) 客户流失应用场景研究. 从应用场景上看,目前客户流失问题研究多集中于电信、金融、电子商务,尤其以前两者居多. 但事实上,“客户”的概念是广义的,客户流失问题广泛存在于各行各业,可以考虑将现有的客户流失研究场景拓宽至其他行业,例如网络游戏玩家的流失、人才流失等等.

参考文献

- 1 Hadden J, Tiwari A, Roy R, *et al.* Computer assisted customer churn management: State-of-the-art and future

- trends. *Computers & Operations Research*, 2007, 34(10): 2902–2917.
- 2 Roberts JH. Developing new rules for new markets. *Journal of the Academy of Marketing Science*, 2000, 28(1): 31–44. [doi: [10.1177/0092070300281004](https://doi.org/10.1177/0092070300281004)]
- 3 Nie GL, Rowe W, Zhang LL, *et al.* Credit card churn forecasting by logistic regression and decision tree. *Expert Systems with Applications*, 2011, 38(12): 15273–15285. [doi: [10.1016/j.eswa.2011.06.028](https://doi.org/10.1016/j.eswa.2011.06.028)]
- 4 Blattberg RC, Kim BD, Neslin SA. *Database Marketing: Analyzing and Managing Customers*. New York: Springer, 2008.
- 5 Jahromi AT, Stakhovych S, Ewing M. Managing B2B customer churn, retention and profitability. *Industrial Marketing Management*, 2014, 43(7): 1258–1268. [doi: [10.1016/j.indmarman.2014.06.016](https://doi.org/10.1016/j.indmarman.2014.06.016)]
- 6 应维云, 覃正, 赵宇, 等. SVM 方法及其在客户流失预测中的应用研究. *系统工程理论与实践*, 2007, 27(7): 105–110.
- 7 Amin A, Anwar S, Adnan A, *et al.* Customer churn prediction in the telecommunication sector using a rough set approach. *Neurocomputing*, 2017, (237): 242–254.
- 8 Larivière B, Van Den Poel D. Investigating the role of product features in preventing customer churn, by using survival analysis and choice modeling: The case of financial services. *Expert Systems with Applications*, 2004, 27(2): 277–285. [doi: [10.1016/j.eswa.2004.02.002](https://doi.org/10.1016/j.eswa.2004.02.002)]
- 9 Coussement K, Lessmann S, Verstraeten G. A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. *Decision Support Systems*, 2017, 95: 27–36. [doi: [10.1016/j.dss.2016.11.007](https://doi.org/10.1016/j.dss.2016.11.007)]
- 10 Huang Y, Kechadi T. An effective hybrid learning system for telecommunication churn prediction. *Expert Systems with Applications*, 2013, 40(14): 5635–5647. [doi: [10.1016/j.eswa.2013.04.020](https://doi.org/10.1016/j.eswa.2013.04.020)]
- 11 张玮, 杨善林, 刘婷婷. 基于 CART 和自适应 Boosting 算法的移动通信企业客户流失预测模型. *中国管理科学*, 2014, 22(10): 90–96.
- 12 Hadiji F, Sifa R, Drachen A, *et al.* Predicting player churn in the wild. *IEEE Conference on Computational Intelligence and Games*. Dortmund, Germany, 2014. 1–8.
- 13 Miguéis VL, Van Den Poel D, Camanho AS, *et al.* Modeling partial customer churn: On the value of first product-category purchase sequences. *Expert Systems with Applications*, 2012, 39(12): 11250–11256. [doi: [10.1016/j.eswa.2012.03.073](https://doi.org/10.1016/j.eswa.2012.03.073)]
- 14 Datta P, Masand B, Mani DR, *et al.* Automated cellular modeling and prediction on a large scale. *Artificial Intelligence Review*, 2000, 14(6): 485–502. [doi: [10.1023/A:1006643109702](https://doi.org/10.1023/A:1006643109702)]
- 15 Lima E, Mues C, Baesens B. Monitoring and backtesting churn models. *Expert Systems with Applications*, 2011, 38(1): 975–982. [doi: [10.1016/j.eswa.2010.07.091](https://doi.org/10.1016/j.eswa.2010.07.091)]
- 16 Neslin SA, Gupta S, Kamakura W, *et al.* Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research*, 2006, 43(2): 204–211. [doi: [10.1509/jmkr.43.2.204](https://doi.org/10.1509/jmkr.43.2.204)]
- 17 Lemmens A, Gupta S. *Managing churn to maximize profits*. Boston, MA: Harvard Business School, 2013.
- 18 罗彬, 邵培基, 罗尽尧, 等. 基于粗糙集理论—神经网络—蜂群算法集成的客户流失研究. *管理学报*, 2011, 8(2): 265–272.
- 19 于小兵, 曹杰, 巩在武. 客户流失问题研究综述. *计算机集成制造系统*, 2012, 18(10): 2253–2263.
- 20 Vafeiadis T, Diamantaras KI, Sarigiannidis G, *et al.* A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, 2015, 55: 1–9. [doi: [10.1016/j.simpat.2015.03.003](https://doi.org/10.1016/j.simpat.2015.03.003)]
- 21 Chen ZY, Fan ZP, Sun MH. A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data. *European Journal of Operational Research*, 2012, 223(2): 461–472. [doi: [10.1016/j.ejor.2012.06.040](https://doi.org/10.1016/j.ejor.2012.06.040)]
- 22 De Bock KW, Van Den Poel D. An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. *Expert Systems with Applications*, 2011, 38(10): 12293–12301. [doi: [10.1016/j.eswa.2011.04.007](https://doi.org/10.1016/j.eswa.2011.04.007)]
- 23 Moeyersoms J, Martens D. Including high-cardinality attributes in predictive models: A case study in churn prediction in the energy sector. *Decision Support Systems*, 2015, 72: 72–81. [doi: [10.1016/j.dss.2015.02.007](https://doi.org/10.1016/j.dss.2015.02.007)]
- 24 Amin A, Anwar S, Adnan A, *et al.* Comparing oversampling techniques to handle the class imbalance problem: A customer churn prediction case study. *IEEE Access*, 2016, (4): 7940–7957. [doi: [10.1109/ACCESS.2016.2619719](https://doi.org/10.1109/ACCESS.2016.2619719)]
- 25 Jamil S, Khan A. Churn comprehension analysis for telecommunication industry using ALBA. *2016 International Conference on Emerging Technologies (ICET)* Islamabad, Pakistan. 2016. 1–5.
- 26 Keaveney SM. Customer switching behavior in service industries: An exploratory study. *Journal of Marketing*, 1995, 59(2): 71–82. [doi: [10.2307/1252074](https://doi.org/10.2307/1252074)]

- 27 Padmanabhan B, Hevner A, Cuenco M, *et al.* From information to operations: Service quality and customer retention. *ACM Trans. on Management Information Systems*, 2011, 2(4): 1–21.
- 28 盛昭瀚, 柳炳祥. 客户流失危机分析的决策树方法. *管理科学学报*, 2005, 8(2): 20–25.
- 29 于小兵, 曹杰, 张梦男. B2C 电子商务客户流失原因评估研究. *模糊系统与数学*, 2012, 26(6): 166–172.
- 30 李婷婷. 影响 B2C 电子商务企业客户流失因素的实证分析. *对外经贸*, 2014, (1): 136–137.
- 31 Glady N, Baesens B, Croux C. Modeling churn using customer lifetime value. *European Journal of Operational Research*, 2009, 197(1): 402–411. [doi: [10.1016/j.ejor.2008.06.027](https://doi.org/10.1016/j.ejor.2008.06.027)]
- 32 胡理增, 于信阳, 张长赋, 等. 基于经费约束和广义客户终身价值最大化的多客户流失挽救模型. *系统工程理论与实践*, 2009, 29(2): 63–69.
- 33 胡理增, 陈建军. 无约束条件下多客户流失挽救最优化决策. *中国管理科学*, 2009, 17(6): 39–43.
- 34 罗彬, 邵培基, 罗尽尧, 等. 基于竞争对手反击的电信客户流失挽留研究. *管理科学学报*, 2011, 14(8): 17–33.
- 35 罗彬, 邵培基, 罗尽尧, 等. 基于预算限制和客户挽留价值最大化的电信客户流失挽留研究. *管理学报*, 2012, 9(2): 280–288.
- 36 Tamaddoni A, Stakhovych S, Ewing M. Comparing churn prediction techniques and assessing their performance: A contingent perspective. *Journal of Service Research*, 2015, 19(2): 123–141.
- 37 Masand B, Datta P, Mani DR, *et al.* CHAMP: A prototype for automated cellular churn prediction. *Data Mining and Knowledge Discovery*, 1999, 3(2): 219–225. [doi: [10.1023/A:1009873905876](https://doi.org/10.1023/A:1009873905876)]
- 38 丁君美, 刘贵全, 李慧. 改进随机森林算法在电信业客户流失预测中的应用. *模式识别与人工智能*, 2015, 28(11): 1041–1049.
- 39 Gordini N, Veglio V. Customers churn prediction and marketing retention Strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry. *Industrial Marketing Management*, 2016, 62: 100–107.
- 40 Yu XB, Guo SS, Guo J, *et al.* An extended support vector machine forecasting framework for customer churn in e-commerce. *Expert Systems with Applications*, 2011, 38(3): 1425–1430. [doi: [10.1016/j.eswa.2010.07.049](https://doi.org/10.1016/j.eswa.2010.07.049)]
- 41 Saradhi VV, Palshikar GK. Employee churn prediction. *Expert Systems with Applications*, 2011, 38(3): 1999–2006. [doi: [10.1016/j.eswa.2010.07.134](https://doi.org/10.1016/j.eswa.2010.07.134)]
- 42 Runge J, Gao P, Garcin F, *et al.* Churn prediction for high-value players in casual social games. *IEEE Conference on Computational Intelligence and Games (CIG)*. Dortmund, Germany. 2014. 1–8.
- 43 Günther CC, Tvette IF, Aas K, *et al.* Modelling and predicting customer churn from an insurance company. *Scandinavian Actuarial Journal*, 2014, 2014(1): 58–71. [doi: [10.1080/03461238.2011.636502](https://doi.org/10.1080/03461238.2011.636502)]
- 44 Schmittlein DC, Morrison DG, Colombo R. Counting your customers: Who are they and what will they do next? *Management Science*, 1987, 33(1): 1–24. [doi: [10.1287/mnsc.33.1.1](https://doi.org/10.1287/mnsc.33.1.1)]
- 45 马少辉, 刘金兰. Pareto/NBD 模型实证与应用研究. *管理科学*, 2006, 19(5): 45–49.
- 46 Fader PS, Hardie BGS, Lee KL. “Counting your customers” the easy way: An alternative to the Pareto/NBD model. *Marketing Science*, 2005, 24(2): 275–284. [doi: [10.1287/mksc.1040.0098](https://doi.org/10.1287/mksc.1040.0098)]
- 47 夏国恩. 客户流失预测的现状与发展研究. *计算机应用研究*, 2010, 27(2): 413–416.
- 48 Xie YY, Li X. Churn prediction with linear discriminant boosting algorithm. *2008 International Conference on Machine Learning and Cybernetics*. Kunming, China. 2008. 228–233.
- 49 应维云, 蔺楠, 谢雅雅, 等. 用 LDA Boosting 算法进行客户流失预测. *数理统计与管理*, 2010, 29(3): 400–408.
- 50 Abbasimehr H, Setak M, Tarokh MJ. A comparative assessment of the performance of ensemble learning in customer churn prediction. *The International Arab Journal of Information Technology*, 2014, 11(6): 599–606.
- 51 Hu XH. A data mining approach for retailing bank customer attrition analysis. *Applied Intelligence*, 2005, 22(1): 47–60. [doi: [10.1023/B:APIN.0000047383.53680.b6](https://doi.org/10.1023/B:APIN.0000047383.53680.b6)]
- 52 Kim YS. Toward a successful CRM: Variable selection, sampling, and ensemble. *Decision Support Systems*, 2006, 41(2): 542–553. [doi: [10.1016/j.dss.2004.09.008](https://doi.org/10.1016/j.dss.2004.09.008)]
- 53 Lu JX. Predicting customer churn in the telecommunications industry—An application of survival analysis modeling using SAS. *SAS User Group International (SUGI27) Online Proceedings*. <http://www2.sas.com/proceedings/sugi27/p114-27.pdf>. [2012/2017-02-28].
- 54 Adebisi SO, Oyatoye EO, Mojekwu JN. Predicting customer churn and retention rates in Nigeria’s mobile telecommunication industry using markov chain modelling. *Acta Universitatis Sapientiae Economics and Business*, 2015, 3(1): 16

- 67–80.
- 55 袁旭梅, 康健, 张昕. 动态 CRM 模型在电子商务中的应用. 中国管理科学, 2003, 11(S1): 343–347.
- 56 Richter Y, Yom-Tov E, Slonim N. Predicting customer churn in mobile networks through analysis of social groups. Proc. of the 2010 SIAM International Conference on Data Mining. Columbus. 2010. 732–741.
- 57 Verbeke W, Martens D, Baesens B. Social network analysis for customer churn prediction. Applied Soft Computing, 2014, 14: 431–446. [doi: [10.1016/j.asoc.2013.09.017](https://doi.org/10.1016/j.asoc.2013.09.017)]
- 58 Rehman A, Ali AR. Customer churn prediction, segmentation and fraud detection in telecommunication industry. ASE Science Division Conferences. Cambridge, MA, USA. 2014. 1–9.
- 59 Pettersson M. SPC with applications to churn management. Quality and Reliability Engineering International, 2004, 20(5): 397–406. [doi: [10.1002/\(ISSN\)1099-1638](https://doi.org/10.1002/(ISSN)1099-1638)]
- 60 Jiang W, Au T, Tsui KL. A statistical process control approach to business activity monitoring. IIE Trans., 2007, 39(3): 235–249. [doi: [10.1080/07408170600743912](https://doi.org/10.1080/07408170600743912)]
- 61 Samimi Y, Aghaie A. Monitoring usage behavior in subscription-based services using control charts for multivariate attribute characteristics. IEEE International Conference on Industrial Engineering and Engineering Management. Singapore, Singapore. 2008. 1469–1474.
- 62 Chen SH. The gamma CUSUM chart method for online customer churn prediction. Electronic Commerce Research and Applications, 2016, 17: 99–111. [doi: [10.1016/j.elerap.2016.04.003](https://doi.org/10.1016/j.elerap.2016.04.003)]
- 63 Mozer MC, Wolniewicz R, Grimes DB, *et al.* Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. IEEE Trans. on Neural Networks, 2000, 11(3): 690–696. [doi: [10.1109/72.846740](https://doi.org/10.1109/72.846740)]
- 64 Brandusoiu I, Todorean G. Churn prediction in the telecommunications sector using support vector machines. Acta Technica Napocensis, 2016, 57(1): 27–30.
- 65 Kirui C, Hong L, Wilson C, *et al.* Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining. IJCSI International Journal of Computer Science Issues, 2013, 10(2): 165–172.
- 66 Fathian M, Hoseinpoor Y, Minaei-Bidgoli B. Offering a hybrid approach of data mining to predict the customer churn based on bagging and boosting methods. Kybernetes, 2016, 45(5): 732–743. [doi: [10.1108/K-07-2015-0172](https://doi.org/10.1108/K-07-2015-0172)]
- 67 柳炳祥, 盛昭翰. 一种基于 Rough 集的客户流失风险分析方法. 中国管理科学, 2002, 10(S1): 130–133.
- 68 朱帮助. 基于 SMC-RS-LSSVM 的电子商务客户流失预测模型. 系统工程理论与实践, 2010, 30(11): 1960–1967. [doi: [10.12011/1000-6788\(2010\)11-1960](https://doi.org/10.12011/1000-6788(2010)11-1960)]
- 69 琚春华, 卢琦蓓, 郭飞鹏. 融入个体活跃度的电子商务客户流失预测模型. 系统工程理论与实践, 2013, 33(1): 141–150. [doi: [10.12011/1000-6788\(2013\)1-141](https://doi.org/10.12011/1000-6788(2013)1-141)]
- 70 Holtrop N, Wieringa JE, Gijzenberg MJ, *et al.* No future without the past? Predicting churn in the face of customer privacy. International Journal of Research in Marketing, 2016, 34(1): 154–172.
- 71 Provost FJ, Fawcett T, Kohavi R. The case against accuracy estimation for comparing induction algorithms. Proc. of the 15th International Conference on Machine Learning. Morgan Kaufmann, San Francisco, USA. 1998. 98. 445–453.
- 72 Langley P. Crafting papers on machine learning. Proc. of the 17th International Conference on Machine Learning. San Francisco, CA, USA. 2000. 343–354.
- 73 Hand DJ. Measuring classifier performance: A coherent alternative to the area under the ROC curve. Machine Learning, 2009, 77(1): 103–123. [doi: [10.1007/s10994-009-5119-5](https://doi.org/10.1007/s10994-009-5119-5)]
- 74 夏国恩, 金炜东. 基于支持向量机的客户流失预测模型. 系统工程理论与实践, 2008, 28(1): 71–77.
- 75 Lessmann S, Baesens B, Seow HV, *et al.* Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research, 2015, 247(1): 124–136. [doi: [10.1016/j.ejor.2015.05.030](https://doi.org/10.1016/j.ejor.2015.05.030)]