

# BP 神经网络电信行业客户流失风险预测研究

王素娟

(安徽财经大学 管理科学与工程学院, 安徽 蚌埠 233030)

**摘要:**随着移动互联网的蓬勃发展,电信服务企业处于日趋激烈的竞争漩涡中,对于电信企业来说,客户的多少决定着企业的自身竞争力与效益,精准预测客户流失风险对企业的战略决策来说至关重要。因此,本文结合电信行业的背景,利用 BP 神经网络的机器学习理论,融合 CRM 系统中的数据,建立了电信行业客户风险预测模型,并通过 Matlab 实验平台验证模型预测的可行性、准确性与有效性。

**关键词:**BP 神经网络;电信行业;风险预测;客户流失

**中图分类号:**F416 **文献标识码:**A

**文章编号:**1005-913X(2020)10-0050-03

近些年来,我国的互联网产业进入了高速发展的轨道,各种互联网产品和移动终端相继出现,其中移动互联网行业更是连续几年保持飞速增长,这种飞速增长的态势相应地也加快了我国电信行业的发展。根据有关数据表明,我国的移动互联网产业产值已跃居全球第一,移动智能手机产量位居世界第一,4G 用户也在保持持续增长。截至 2019 年,中国电信拥有 3.258 亿 4G 用户,中国联通拥有 3.2462 亿用户,中国移动用户数高达 9.38641 亿,占全国用户 60.16%。电信行业面临着日益严峻的竞争压力,激烈的竞争也就迫使各电信企业采取措施发展新客户和留住老客户,以此提高自身在行业的竞争力,而从运营成本来说,挽留老客户的成本相对较少。因此,如何精确的预测客户流失的风险,对运营商的决策起着举足轻重的作用。

行业客户流失风险的预测是当前学术界研究的热点,从研究对象和研究方法上来看,相关研究可分为两类:第一类大多从客户层面出发,利用数据挖掘技术,结合决策树、支持向量机等分类算法深入研究客户性质和消费行为,构建客户流失的风险预测模型。例如 Ahmad AK<sup>[5]</sup>利用 SyriaTel 电信公司的大量原始数据,结合 XGBOOST 算法构建电信

公司的客户流失风险预测模型,并应用于客户流失的分类。第二类从客户关系管理角度出发结合其他的算法例如逻辑回归、灰色预测等对客户流失行为预测并采取挽救措施。例如,Vijaya J<sup>[6]</sup>结合粗糙集理论本提出了一种识别电信客户流失预测有效特征的方法。

这些研究为电信行业决策者理解客户流失风险预测提供了很好的视角,但也存在一些不足:一是,大多学者的研究对象在客户层面,研究领域主要集中在金融行业、电子商务、银行业务领域,缺乏电信行业的针对性的客户流失风险预测;二是,从大量学者的研究反馈中得出,众多研究者的研究仿真精准度不高,须要对模型展开改进与完备。本文认为要克服上述的第一个问题就必须充分结合我国电信行业的客户关系管理的相关理念,从企业的客户管理角度构建评价指标体系,克服第二个问题应挖掘相关电信企业的数据,借助于 BP 神经网络算法,构建电信行业客户风险预测模型,并通过 MATLAB 仿真实验验证模型的科学性、精确性。

## 一、电信行业客户流失风险预测指标体系的构建

在电信行业中,影响着客户的流失存在诸多的原因,可能是客户自身原因,也有可能是运营商的原因。对于运营商来说,影响客户流失主要有套餐资费优惠程度、工作人员服务态度等,而对于客户本身来说,则影响原因更为复杂,有可能是工作人员的服务质量问题,也有可能是客户自身的原因。构建合理、高效又完善的客户流失风险预测指标体系对提高风险预测的准确性极其重要。综合考虑了国内外关于电信行业客户流失风险预测的相关资料,用频数统计分析挑选出 4 个一级指标,财务、客户、运营商和服务,16 个二级指标。其中财务方面包括销售收入  $X_1$ 、欠费率  $X_2$ 、利润率  $X_3$ 、业务价格  $X_4$ 、

**收稿日期:**2020-06-23

**基金项目:**安徽财经大学科研基金项目(XSKY2091)

**作者简介:**王素娟(1998-),女,安徽芜湖人,本科学生,研究方向:信息管理与信息系统。

合同签约率  $X_5$ , 客户方面包括客户满意度  $X_6$ 、新客户增长率  $X_7$ 、投诉率  $X_8$ 、忠诚度  $X_9$ 、合约到期时间  $X_{10}$ ; 运营商方面包括技术开发费  $X_{11}$ 、市场开拓力  $X_{12}$ 、营销人员态度  $X_{13}$ ; 服务方面包括网络功能实现度  $X_{14}$ 、业务多样性  $X_{15}$ 、服务人员流动率  $X_{16}$ 。经过分析众多专家学者的观念并结合数据的特征, 去除了代表性差、随机性大、实用性低的指标, 以确保各指标的科学与独立性, 使得指标体系更加科学与完善。

## 二、神经网络相关理论

### (一) BP 神经网络概述

BP(back propagation)神经网络, 是一种应用广泛、误差方向传播的多层前馈神经网络, 可以对输入的数据进行学习, 并输出相应的结果, 其优势在于不需要去建立数学方程式, 就可以进行训练和学习, 其核心思想是梯度下降法, 运用一些数学手段使得 BP 神经网络的实际输出和预期输出的误差方差达到最小。BP 神经网络的结构大致分为三层: 输入层、隐含层、输出层。

### (二) BP 神经网络训练过程

BP 的训练过程存在两个子过程, 分别为正向传递与误差反向传递过程。在正向传递过程中, 数据集通过输入层传输至隐含层, 在隐含层经过计算, 传输至输出层。输出层的输出结果与真实值进行比较, 若误差大或未符合预期值, 则进行反向传递, 传递途径与正向传递过程相反, 但需要逐层修改权重与阈值, 不断循环此过程至与预期输出值的误差值达到最小。

## 三、电信行业客户流失风险预测模型结构的确定

一是输入层的确定。本文把影响客户流失风险的 16 个二维指标作为模型的输入。

二是输出层的确定。模型的输出层为 4 个, 本文用客户流失风险的等级来表示, 共划分为 4 个等级, 目标输出格式为(0001)、(0010)、(0100)和(1000), 分

别对应风险较小、风险一般、风险较大和风险极大。

三是隐含层层数和节点数的确定。本文所建立的电信行业客户流失风险预测模型结构相对较简单, 故模型的隐含层的层数可以取 1。据相关实验表明, 当构建的风险预测模型时, 若隐含层的节点数少会导致神经网络处理信息能力差, 但若节点数过多又会导致训练速度慢。因此, 节点数应合理选取, 既要保证训练速度又要保证模型的处理信息能力。目前 BP 神经网络的隐含层的节点数并没有明确规, 一般由经验确定。众多学者总结出相关经验公式如下: ( $m, n$  和  $h$  分别表示输入层、输出层和隐含层的节点数)

$$(1) h_1 = \log_2^m; (2) h_2 = \sqrt{mn}; (3) h_3 = 2m + 1$$

本文根据上述经验公式得出最优节点数, 然后继续利用最优节点数的附近值对模型的最优隐含层节点数进行实验验证。

四是传递函数的确定。S 型函数是 BP 神经网络中传递函数的使用主流。本文使用“Tansig”函数作为隐含层的传递函数, 表达式为(4); 输出层的传递函数使用“Logsig”函数, 表达式为(5); 而对于训练函数, 本文使用“Trainlm”函数。

$$(4) f(x) = \frac{2}{e^{-2x} + 1} - 1; (5) f(x) = \frac{2}{e^{-x} + 1}$$

## 四、电信行业客户流失风险预测模型的构建

### (一) 数据采集与处理

本文采取了某移动公司近五年的统计数据, 由于指标数据之间量纲不同, 本文采用公式:

$$(6) x' = \frac{x_{ij} - \min_j x_{ij}}{\max_j x_{ij} - \min_j x_{ij}}$$

对数据进行标准化, 标准化的数据均落在[0,1]之间, 并对数据进行平滑处理, 得到1000个标准化数据。处理后的部分数据如下表。

表 样本数据

	信用评估指标															
	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$	$X_{11}$	$X_{12}$	$X_{13}$	$X_{14}$	$X_{15}$	$X_{16}$
1	0.52	0.59	0.62	0.58	0.61	0.60	0.56	0.61	0.58	0.55	0.64	0.54	0.55	0.51	0.49	0.53
2	0.57	0.50	0.64	0.62	0.58	0.61	0.52	0.57	0.58	0.50	0.64	0.48	0.52	0.55	0.57	0.60
3	0.49	0.61	0.76	0.65	0.69	0.62	0.57	0.77	0.65	0.68	0.57	0.63	0.56	0.51	0.63	0.59
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1000	0.61	0.61	0.70	0.60	0.52	0.53	0.62	0.66	0.53	0.60	0.53	0.56	0.59	0.56	0.54	0.60

## (二) 模型结构的确定

根据上文可知,模型的输入层、隐含层和输出层的层数均为1,其中输入层的节点数确定为16,输出层的节点数确定为4,根据目前的隐含层个数确定的经验公式,分别为4、8、33,为了使模型的结构拟合程度更高,采用五种隐含层个数,添加节点数为31、32和34三种方式,其他参数值保持一致。则BP神经网络的结构可以大致确定为:16-4-4、16-8-4、16-31-4、16-32-4、16-33-4、16-34-4三种。

将标准化数据中前900组数据作为模型的训练样本,神经网络的训练误差设置为0.001,显示中间结果的周期设置为50,最大迭代次数设置为1000,学习率设为0.1,运用建立好的模型对数据进行训练,通过不断的迭代学习,直至训练结束。经过训练实验,根据所得误差分析结果知,隐含层的节点数分别为4、8和33时,模型的误差与预先设定的误差均存在差距,相差分别为0.058、0.061、0.002、0.012、0.041、0.027、0.027。由回归分析的结果可知,模型的回归系数分别 $R_1=0.98921$ 、 $R_2=0.98074$ 、 $R_3=0.97436$ 、 $R_4=0.9717$ 、 $R_5=0.98037$ 、 $R_6=0.9579$ 。当神经网络层的误差值设为0.001时,一般认为当模型的回归系数达到80%时即为高精度模型。以上隐含层节点数下,模型均满足高精度模型的要求,通过对比模型的误差值,当隐含层数为33时,误差值最小。因此,当模型的隐含层数设为33时,模型的拟合程度最好,可用于电信行业客户风险的预测。

## (三) 模型的训练

经过上述分析,电信行业客户流失风险预测模型的结构确定为16-33-4,为了进一步验证模型的预测准确率,采用余下的100组测试数据对模型进行验证,在MATLAB上对测试数据进行仿真误差分析,其测试样本最小误差为0.00612,其训练结果迭代曲线如图所示。

横坐标表示算法的迭代次数,纵坐标表示误差的结果,由图可知,BP神经网络算法的经过12次迭代后,网络停止了训练,此时训练样本的误差趋近于 $10^{-3}$ 。误差处在可控范围之内,说明该模型预测电信行业客户流失的风险出错率较低,可以用于客户流失风险的预测。通过与实际值的比较,发现所建

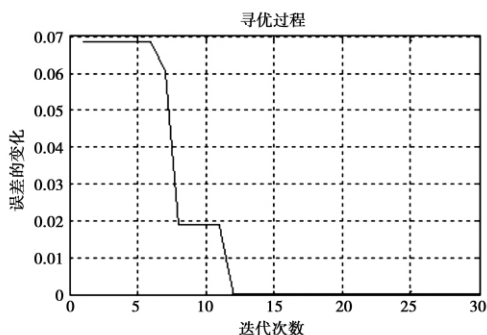


图 优化结果

立的电信行业客户流失的预测模型成功反映了输入和输出之间的关系,也成功预测了客户流失的风险,因此,该模型适用于电信行业客户流失风险的预测。

我国三大电信运营商正处于深化“提速降费”和“携号转网”的激烈竞争环境中,运营商们均有“存量经营”战略转型的考量。从CRM角度来说,存量客户的价值远大于新增用户,即需采取相应措施挽留老客户进而提高企业自身竞争力。因此,如何精准预测客户流失风险并制定相应的挽留与维系对策是运营商降低运营成本,提高核心竞争优势的关键。在此背景下,本文在企业客户关系管理角度上,利用BP神经网络学习能力强、网络稳定的特点,以某移动运营商近五年客户数据为研究对象,构建了电信行业客户流失预警模型,研究结果表明基于BP神经网络算法的电信客户流失预测模型准确率高达到98.8%,验证了模型的精确性、科学性,帮助电信行业运营商对潜在客户流失风险的预测提供新思路,从而采取相应个性化措施以便降低客户的流失率。

## 参考文献:

- [1] Ahmad AK, Jafar A, Aljoumaa K. Customer Churn Prediction in Telecom Using Machine Learning in Big Data Platform[J]. Journal of Big Data, 2019(1).
- [2] Vijaya J, Sivasankar E. Computing Efficient Features Using Rough Set Theory Combined with Ensemble Classification Techniques to Improve the Customer Churn Prediction in Telecommunication Sector [J]. Computing, 2018(8).

[责任编辑:谭志远]