



研究与开发

基于联合神经网络的投诉预测模型研究

马晓亮^{1,2,3}, 刘英^{2,3}, 高洁^{2,3}

- (1. 西安电子科技大学, 陕西 西安 710071;
2. 中国电信股份有限公司广州分公司, 广东 广州 510620;
3. 马晓亮劳模与工匠人才创新工作室, 广东 广州 510620)

摘要: 对影响电信运营商重复投诉的关键因素进行深入探讨, 旨在提高服务质量并构建风险预测模型。基于运营商客服数据, 研究采用了 Logistic 回归、BP 神经网络以及二者联合建模的方法。Logistic 回归模型确定了 5 个主要影响因素, 预测重复投诉发生的概率, 精度达到 80.0%。BP 神经网络则选取了 81 个影响因素, 预测精度为 90.6%。在此基础上, 构建了联合模型, 其精度高达 92.8%。实际应用于某省会电信运营商后, 重复投诉率下降了 3.2%, 成效显著, 为提高电信运营商服务质量、降低重复投诉率提供了有力支持, 对我国电信行业发展具有重要意义。

关键词: AI 客服; 联合建模; 重复投诉; Logistic 回归; 深度学习模型

中图分类号: TP393

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2024006

Research on a complaint prediction model utilizing joint neural networks

MA Xiaoliang^{1,2,3}, LIU Ying^{2,3}, GAO Jie^{2,3}

1. Xidian University, Xi'an 710126, China
2. Guangzhou Branch of China Telecom Co., Ltd., Guangzhou 510620, China
3. Ma Xiaoliang Innovation Studio for Model Workers and Creative Talents, Guangzhou 510620, China

Abstract: By conducting in-depth exploration on the key factors affecting repeat complaints of telecom operators, this study aimed to improve service quality and construct a risk prediction model. Based on the operator's customer service data, the study employed Logistic regression, BP neural network, and their combined modeling methods. The Logistic regression model identified five major influencing factors, predicting the probability of repeat complaints with an accuracy of 80.0%. The BP neural network selected 81 influencing factors, achieving a prediction accuracy of 90.6%. On this basis, a combined model was constructed with an accuracy rate of up to 92.8%. After practical application in a provincial telecom operator, the repeat complaint rate decreased by 3.2%, demonstrating a significant impact. Strong support is provided for improving the service quality of telecom operators and reducing repeat complaints, which is of great significance for the development of the telecom industry in China.

Key words: AI customer service, joint modeling, repeated complaint, Logistic regression, deep learning model

收稿日期: 2023-07-18; 修回日期: 2024-01-11

通信作者: 高洁, gaojie.gd@chinatelecom.cn

0 引言

随着电信行业竞争的日益激烈, 提高用户满意度和降低客户流失率成为电信运营商面临的重要挑战。作为服务质量的关键指标之一, 客户投诉引起了广泛的关注。本文旨在探索客户重复投诉的关键因素, 并建立可靠的预测模型, 以协助电信运营商提高其服务质量。客户重复投诉被定义为同一人对同一问题进行两次或以上的投诉行为^[1], 一直是研究者关注的焦点。

近年来, 研究者们广泛探讨了客服系统的智能化, 从而提高客服效率^[2]。同时, 研究者们采用了各种智能化的方法来分析重复投诉, 主要集中在定性分析上。例如, 雷霆等^[3]通过分析客户的网络情况等相关数据, 从规划、建设、维护、优化等多个维度找出重复投诉的原因, 并探讨了降低网络客户重复投诉率的策略。韩鹏^[4]强调要进行投诉根源分析, 并制定有效的纠正措施。然而, 这些研究并没有对重复投诉进行定量分析。

定量建模可分为两类, 即线性化的数学模型建模和非线性化的机器学习建模。其中, 当自变量为离散值时, Logistic 回归模型 (Logistic regression model) 有效预测分类结果^[5]。在金融领域, 该模型可以协助金融机构预测股票是否被操纵^[6], 也可以预测股票走势, 为投资决策提供依据^[7]。在医学领域, 该模型可以帮助医生更准确地诊断疾病^[8]和预测疾病趋势^[9]。在用户投诉领域, 一些研究探讨了 Logistic 回归模型在学术界和工业界的应用。例如, 张艳莉^[10]提出 Logistic 回归模型可以快速分类用户抱怨, 而庄贵军等^[11]则证明了该模型在构建用户投诉管理机制中的作用。

非线性化的机器学习建模中, 神经网络是一种主要的建模方式, 其中 BP 神经网络 (backpropagation neural network) 是一种常用的反向传播网络。在金融风险控制和医疗诊断领域, 神经网络具有广泛的应用前景。在金融风险控制和医疗诊

断领域都有重要的应用。在金融业, 神经网络可以通过分析历史金融数据和预测未来市场趋势, 协助投资者做出明智的投资决策。例如, 深度神经网络已经被用来预测沪深 300 指数的涨跌, 为投资决策提供参考^[12]。在医学领域, 神经网络模型可以帮助医生实现更准确的诊断, 为病人设计更好的治疗方案, 并提高治疗效果^[13]。

针对运营商的客户投诉, 神经网络得到了越来越多的关注。周文杰等^[14]提出了一种利用神经网络自动学习非线性复合特征的方法, 从而提高传统分类器的分类精度, 并证明了深度学习在电信大数据领域的实用性。在 2018 年的一项研究中, 穆晓玛等^[15]比较了 3 种算法——差分自回归移动平均模型 (autoregressive integrated moving average model, ARIMA)、多元线性回归分析 (multiple linear regression analysis) 和 BP 神经网络对客户投诉的预测效果, 经过预测值与实际值的对比计算, 发现单独采用 BP 神经网络模型来预测客户投诉是最为适宜的方案之一。在所得结果中, BP 神经网络模型相对误差小于 10% 的占 31.4%, 相对误差小于 20% 的占 58.5%。尽管 BP 神经网络模型在客户投诉预测方面较为优异, 但存在过拟合和效果不稳定等问题。然而, 文献[16]显示, 联合模型能够降低过拟合, 提高模型泛化性。

基于以上情况, 本文提出了一种用于投诉预测的联合模型, 该模型基于联合神经网络方法, 将 Logistic 回归模型和 BP 神经网络模型组合使用。该模型综合考虑了线性和非线性因素对重复投诉概率的影响。实验结果表明, 在运营商投诉预测的场景下, 该联合模型的预测准确率达到 92.8%, 超过了单独的 Logistic 回归模型或 BP 神经网络模型。这表明, 在运营商投诉预测的应用场景下, 两种不同建模方法的结合可以增强模型学习能力, 缓解单一模型的局限性, 从而更好地适应实际应用场景。因此, 本文提出的联合模型在预测客户重复投诉方面具有较高的实用价值。



1 研究路线及模型构建

基于某电信运营商的“客服工单数据(connect center order, CCO)”及“业务支撑系统(business support system, BSS)”数据,“客服工单数据”包含客户向电信运营商发起的投诉咨询诉求的数据,以及工单处理、满意度回访等闭环过程的数据。“业务支撑系统”包含客户基础信息、装机信息、移动/固话/宽带信息及费用信息等。

1.1 重复投诉特征项选择

(1) “客服工单数据”的特征项

在重复投诉的预测模型中,“客服工单数据”来源于电信运营商的客户服务系统,涵盖了客户的投诉记录、提交的工单属性等,可反映客户当前的需求,共计 21 个与投诉单相关的特征项。

(2) “业务支撑系统”的特征项

“业务支撑系统”的特征项包括客户在电信运营商里面的基础及消费行为信息(如用户套餐信息、每月欠费信息)等。这类信息可以从运营商的数据库中获取,根据关键质量指标筛选出 60 个特征项。

重复投诉的特征项共计 81 个,数据特征项见表 1。

1.2 数据清洗

确定重复投诉的相关数据特征后,需要对数据清洗处理,从而降低对数据模型输出结果的影响。数据清理包括几个步骤,具体如下。

(1) 删除手动输入到工单数据中的无意义字符

“客服工单数据”特征项中“受理内容”“处理内容”等为人员手工录入内容,容易出现无空格信息而影响数据分析结果,例如,“用户致电反映对前单处理结果不满,要求局方继续跟进处理,否则进行媒体曝光!”,在内容里面“前单”“媒体曝光”都出现了空格,需要删除这类空格。

(2) 删除数据缺失率高的特征项

“客服工单数据”特征项除已选中的 21 个特

征项外,还有一些特征因为其数据缺失率较高而被舍弃处理,例如,95%以上的“战略分群”“客户感知级别”“责任定性”等特征项数据都缺失,因此在数据清洗处理中,应删除较高缺失率的特征项。

(3) 缩小特征的非离散文本描述的范围

对于“回访结果”这一特征,工单数据可能有“未确认”“不需要回复”“理解”“满意”“无法确认”“无法联系客户”等值。在这种情况下,将数据列成“未确认”“满意”和“其他”等值。对于“投诉客户情绪”这一特征,工单数据可能有“满意”“很满意”“十分满意”“很不满”“不满”“有些不满”“不知道”等值,在这种情况下,则将该数据特征项的数据列成“满意”“不满”和“其他”等值。

1.3 构建 Logistic 回归函数

Logistic 回归模型常常用来解决二分类问题,即将数据集中的数据分为两个类别,Logistic 回归模型还可以用于预测某个事件的概率。

本文采集某省会城市电信“客服中心”工单系统 2022 年 7 月 1 日到 2022 年 12 月 31 日的投诉咨询工单数据,数据共计 358 665 个。采用 sklearn 库中的 train_test_split 函数来划分,得到训练集共计 178 141 个,验证集为 108 790 个,测试集为 71 734 个(详见第 2.3 节的数据集)。

(1) 特征工程

经过数据清洗后,Logistic 回归模型变量见表 2。

根据上述测量指标的分类,当前工单是否会产生后续重复投诉工单可以划分为“是”“否”两种结果,是一个二向性的问题,将工单是否重复投诉设为因变量,将“否”定义为 $Y=0$,“是”定义为 $Y=1$,故可选择 Logistic 回归模型对工单“是否重复投诉”的影响因素进行计量分析。

在 Logistic 回归模型中,使用 sklearn 库中的 preprocessing 模块进行特征变换,使用 pandas 库

表 1 数据特征项

“客服工单数据”特征项	“业务支撑系统”特征项	
服务类别	语音套餐名称	关联副卡套餐外使用流量（后付费）
是否有前单	无宽套餐名称	本月注册终端系统型号
工单总历时	近 6 个月内欠费停机次数	本月注册终端品牌
工单流转次数	用户信用额度	最近一次更换终端时间
回访结果	套餐内总可用流量	上上个月出账金额
挂起次数	套餐内已使用流量	上个月出账金额
是否为升级工单	是否超流量用户（后付费）	12 个月内最近一次充值渠道
客户星级	套餐外超流量值（后付费）	12 个月内最近一次充值操作系统
是否小额快速退费	当月关联副卡上网流量	12 个月内最近一次充值方式
受理时长	当月出账金额	是否开通 5G 功能
关键信息	是否欠费用户	是否 5G 终端
投诉风险	当月累计欠费金额	实际速率
投诉客户情绪	是否 5G 套餐用户	套餐速率
受理内容，如出现曝光、越级、媒体标识	当月上网流量（出账）	是否高竞争网格
处理内容	欠缴月份	是否城中村网格
是否认可	是否敏感客户（VIP）	是否商业网格
退费与否	是否黑名单用户	是否敏感客户（非 VIP、VPN）
是否受理	停机类型	落地局向 ID
支持内容，如出现追、催标识	停机原因	落地局向名称
是否有协查单	半年内投诉次数	落地营服 ID
服务类别	累计投诉次数	落地营服名称
	客户编码	用户装机地址 ID
	客户名称	装机地址
	客户星级	接入方式
	揽装人工号	是否 FTTH+HUB
	网点名称	揽装工号 ID
	套餐内总可用通用流量	缴费方式名称
	套餐内已使用通用流量	缴费银行账号
	套餐外使用流量（后付费）	缴费账户类型
	3 个月内投诉次数	缴费账户名称

中的 DataFrame.corr 函数进行特征选择。通过 Pearson 算法分析特征变量的相关性分布情况，如果 Pearson 相关系数低于 0.2，那么其相关性极低或无相关性，选出所需要的特征变量。过滤

掉相关性低于 0.2 的特征变量后，自变量为“是否有前单”“工单总历时时长”“工单流转次数”“挂起次数”“受理时长”，Pearson 相关系数分析见表 3。



表 2 Logistic 回归模型变量

变量	测量特征	赋值
自变量	是否有前单	0=否; 1=是
	工单总历时	连续值, 例如: 13.48、9.22
	工单流转次数	连续值, 例如: 6、4
	回访结果	0=不参与; 1=基本满意; 2=满意; 3=不满意
	挂起次数	连续值, 例如: 0、1、2
	是否为升级工单	0=否; 1=是
	客户星级	0=无星; 1=1 星; 2=2 星; ...; 7=7 星
	是否小额快速退费	0=否; 1=1
	受理时长	连续值, 例如: 103、156
因变量	是否重复投诉	0=否; 1=是

表 3 Pearson 相关系数分析

变量	Pearson 相关系数
是否有前单	0.342
工单总历时时长	0.265
工单流转次数	0.235
挂起次数	0.225

(2) 获取参数

指定模型的参数, 正则化范数 L1, 迭代次数 50。模型训练得到的回归系数, 其 p 值均小于 0.05 检验有效, Logistic 回归模型变量分析见表 4。

表 4 Logistic 回归模型变量分析

变量	回归系数	p 值
是否有前单	0.133	0.005
工单总历时时长	0.054	0.025
工单流转次数	0.051	0.023
挂起次数	0.093	0.016
受理时长	0.043	0.039

得出 Logistic 函数为:

$$\ln(p/(1-p)) = 0.133 \times \text{是否有前单} + 0.054 \times \text{工单总历时时长} + 0.051 \times \text{工单流转次数} + 0.093 \times \text{挂起次数} + 0.043 \times \text{受理时长} \quad (1)$$

其中, p 为事件成功的概率, $1-p$ 为事件失败的概率, 即 $\ln(p/(1-p))$ 是一种因变量与事件发生的概率之间的一种映射关系。

1.4 构建 BP 神经网络模型

BP 神经网络作为一个非线性系统, 可用于逼近非线性映射关系, 而深度学习是 BP 神经网络的一种加强版, 用于解决特征值海量的场景。考虑本文所涉及的特征量为 81 个, 并非海量特征量, 采用 BP 神经网络的定性描述、定量计算、精确逻辑分析和非确定性推理能力, 同时也兼顾了其运行性能。BP 模型示意图如图 1 所示。

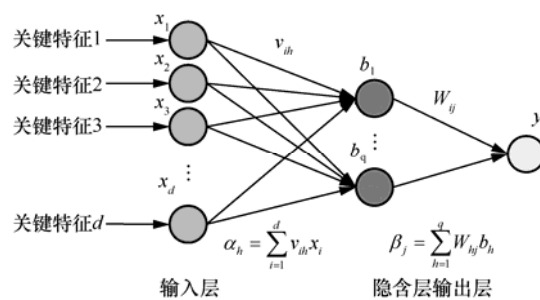


图 1 BP 模型示意图

BP 神经网络算法的主要思路是: 将输入信号从输入层输入, 通过隐含层运算后, 将输入输出值与记录值进行比较, 如果发现误差, 则将误差反向传播到输入层, 并通过梯度降低算法调节神

经元的权重,以达到最优的模型性能,BP 算法流程如图 2 所示。

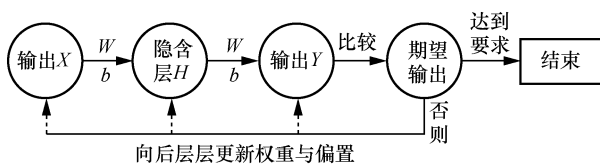


图2 BP 算法流程

每个工单的特征作为输入向量,包括投诉类型、受理时间、客户信息等,输出层为重复投诉概率。输出值被限定在 0 到 1 之间,期望值越大,表明未来一段时间客户重复投诉的可能性越大。

(1) 特征工程

本文共采用 81 个全量特征项为 BP 神经网络模型输入层的节点,其中包括 21 个“客服工单数据”的特征项,如服务类别、是否有前单、工单总历时、工单流转次数、回访结果等;60 个“业务支撑系统数据”的特征项,如语言套餐名称、无宽套餐名称、近 6 个月内欠费停机次数、用户信用额度、套餐内可用流量等。

(2) 模型训练

步骤 1 计算隐含层神经元数。BP 神经网络的隐含层神经元数太多将导致过拟合的问题,而太少又会导致欠拟合的问题,因此根据式(2):

$$p = \sqrt{n+m} + \alpha \quad (2)$$

其中, p 为所求隐含层神经元数; n 为输入层节点数; m 为输出层节点数; α 为修正常量, α 的取值范围一般为 1~100。本文选择的节点神经元数 p 为 100 个。

步骤 2 训练参数。采用与 Logistic 回归模型完全相同的 178 141 个训练集数据,使用网络模型进行训练,根据前面初始参数的设定开始进行迭代训练,得到模型的隐含层层次为 20 层,隐含层每层神经元数为 100 个,迭代次数 1 000 次。训练迭代平均绝对百分比误差 (mean absolute percentage error, MAPE) 收敛图如图 3 所示。

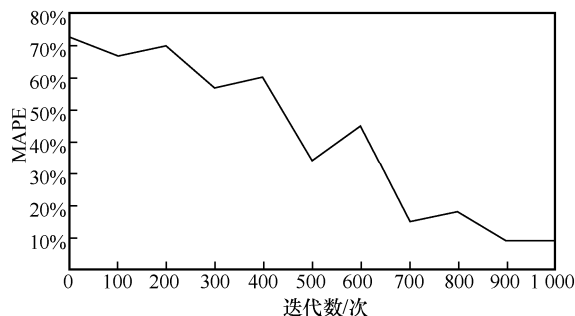


图3 训练迭代 MAPE 收敛图

训练得到一个较优的模型训练参数,该 BP 模型参数见表 5。该模型的 MAPE 较理想,误差率为 9.529% (详见第 2.5 节的 BP 神经网络模型实验结果)。

表 5 BP 模型参数

参数名	参数值
激活函数	ReLU
L2 正则项	0.000 1
迭代次数	1 000
隐含层神经元数	100

步骤 3 最终使用 sklearn 库中的 MLPRegressor 模块进行程序开发工程实践,该程序主要参数设定如下。

```
model_mlp = MLPRegressor(\#得出的 BP 模型对象
```

```
hidden_layer_sizes=(100,100,...,100), \#隐含层层次 20 层及每层 100 个神经元
```

```
activation=relu, \#激活函数为 relu
```

```
solver = lbfgs, \#优化算法为 lbfgs, 即拟牛顿法
```

```
learning_rate_init=0.01, \#学习率初始为 0.01
```

```
max_iter=1000) \#最大迭代次数为 1 000
```

1.5 构建 Logistic 回归函数和 BP 神经网络联合模型

Logistic 回归和 BP 神经网络两个模型分别具有不同的优点,Logistic 回归模型在处理线性分类问题时表现良好,而 BP 神经网络能够处理非线性关系。因此,构建 Logistic 回归和 BP 神



经网络的联合模型,可以全面地考虑线性和非线性因素对重复投诉风险概率的影响,从而提高预测的准确性。

(1) 模型训练

步骤1 采用训练集数据分别构建两个个体学习器模型 Logistic 回归模型及 BP 神经网络模型。

步骤2 再将验证集样本数据输入 Logistic 回归模型及 BP 神经网络模型,得到两个模型的输出结果集。

步骤3 对两个结果集进行分类,分出结果一致的结果集及结果不一致的结果集,使用式(3)和式(4)判定结果是否一致。

$$p_i = 1 - (1 - \text{Pred}_{\text{logistic}(i)}) \times (1 - \text{Pred}_{\text{BP}(i)}) + \delta(\|\text{Pred}_{\text{logistic}(i)} - \text{Pred}_{\text{BP}(i)}\|)^2 \quad (3)$$

$$p = \frac{1}{n} \times \sum_{i=1}^n p_i \quad (4)$$

其中, $\text{Pred}_{\text{Logistic}}$ 和 Pred_{BP} 分别表示 Logistic 回归模型和 BP 神经网络模型的预测结果概率值,两个模型的预测概率值差值的范数的平方代表输出概率分布之间的差异程度。 δ 函数可以用于强调模型输出概率分布之间的相似性。当范数的平方较小时, δ 函数将其映射为接近 1 的值,表示两个模型的输出概率分布非常接近;反之,当范数的平方较大时, δ 函数将其映射为接近 0 的值,表示两个模型的输出概率分布之间差异较大。因此当两个模型的输出概率非常相似时, δ 函数的值将接近 1;而当它们的输出概率差异较大时, δ 函数的值将为 0。 P_i 为

各样本的结果相近度,而 P 为训练样本的整体结果相近度。对于结果不一致的数据则进行人工打标处理,并精调 BP 神经网络模型的隐含层层数、隐含层神经元数等参数,最后将打标后的数据重新作为 Logistic 回归模型及 BP 神经网络模型的输入。

步骤4 多次训练后最终得到最优模型参数为隐含层 20 层,隐含层神经元数 100 个,激活函数为 ReLU (rectified linear unit),迭代次数为 1 000 次。

步骤5 构建所得联合模型,即 BP 模型示意图如图 4 所示。

2 实验与分析

本文使用了 Logistic 回归建模、BP 神经网络建模和 Logistic 回归与 BP 神经网络联合建模 3 种方法,通过实验比较它们在客户重复投诉预测方面的表现,3 种方法均使用相同的数据集。在比较后,选择了预测精度更优的模型来服务运营。

2.1 实验环境

实验硬件配置为 16 核 CPU, 64 GB 内存, RTX 3090 显卡, 操作系统为 CentOS7.6, 编程环境为 Python3.7.0, 深度学习框架为 sklearn 0.22.1。

2.2 数据来源

采集某省会城市电信“客服中心”工作单系统 2022 年 7 月 1 日到 2022 年 12 月 31 日的投诉咨询工单数据,数据共计 358 665 个。

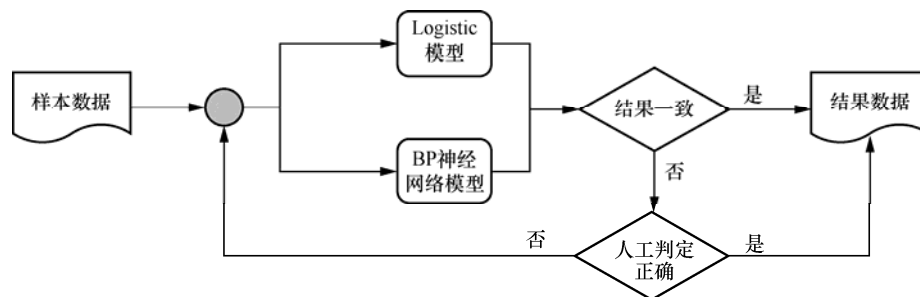


图4 BP 模型示意图

2.3 数据集

使用 sklearn 库中的 train_test_split 函数对 358 665 个投诉咨询工单分成训练集、验证集及测试集。

(1) 训练集

训练集涉及 205 种工作单产品类型、284 种服务类别，训练集数据共有 178 141 个。

(2) 验证集

验证集为 108 790 个真实投诉咨询工单：事实重复投诉工单有 11 342 个，事实非重复投诉工单有 97 448 个。

(3) 测试集

测试集为 71 734 个真实投诉咨询工单：事实重复投诉工单有 7 479 个，事实非重复投诉工单有 64 255 个。

2.4 Logistic 回归模型实验结果

(1) 信度检验

工单样本数据 α 系数为 0.853，说明工单样本数据信度较高，内部一致性较好。

(2) 似然比检验

本模型的似然比结果见表 6。 p 值小于 0.05，所选择的 Logistic 回归模型是有效的。

表 6 Logistic 回归模型似然比结果

模型	p 值	AIC 值	BIC 值
最终模型	0.017	75.277	88.302

(3) 实验结果

训练后得到的模型函数如式 (1) 所示。其中，是否有前单、工单总历时时长、工单流转次数、挂起次数和受理时长是影响重复投诉发生概率的主要因素。

Logistic 回归模型预测准确率为 80.09%，预测能力一般，预测结果见表 7。模型所用的训练数据信度较高、一致性较好，模型本身的 p 值小于 0.05，说明从模型合理，但预测结果并不理想，整体准确率为 80.09%。

表 7 Logistic 回归模型预测结果

	实际重复投诉/宗	实际非重复投诉/宗
预测重复投诉/宗	496	7 298
预测非重复投诉/宗	6 983	56 957
整体准确率	80.09%	—

2.5 BP 神经网络模型实验结果

(1) 模型评估

通过均方误差 (mean squared error, MSE)、MAPE 指标来衡量 BP 神经网络的精度，其中，MSE 表示预测值与实际值之差的平方的期望值，该值越小，模型的精度越高，如式 (5)；MAPE 的值越小，模型的精度越高，如式 (6)。

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (5)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\% \quad (6)$$

其中， n 为样本数量， $\hat{y}_i$ 为 i 样本在模型中的预测值， y_i 为实际值。

BP 模型评估结果见表 8，可见 BP 神经网络的各项评价指标相差不大，MAPE 误差率仅为 9% 左右，模型精度较好。

表 8 BP 模型评估结果

MSE	MAPE
55.07	9.529%

(2) 实验结果

BP 神经网络模型总体准确率为 90.61%，实验结果见表 9，可见 BP 神经网络模型合理，且预测整体准确率较为理想。

表 9 BP 神经网络模型实验结果

	实际重复投诉/宗	实际非重复投诉/宗
预测重复投诉/宗	1 149	406
预测非重复投诉/宗	6 330	63 849
整体准确率	90.61%	—



2.6 Logistic 回归和 BP 神经网络联合模型实验结果

(1) 模型评估

同样使用 MSE、MAPE 指标来衡量联合模型的精度，联合模型评估见表 10，可见联合模型的各项评价指标均优于未优化前的结果，MAPE 误差率仅为 8% 左右，模型精度较好。

表 10 联合模型评估

模型	MSE	MSE 下降	MAPE	MAPE 下降
BP 神经网络模型	55.07	—	9.529%	—
联合模型	50.83	-4.24	8.124%	1.405%

(2) 实验结果

Logistic 回归和 BP 神经网络联合模型总体准确率为 92.78%，联合模型实验结果见表 11，可见联合模型的预测准确率更为理想。

表 11 联合模型实验结果

	实际重复投诉/宗	实际非重复投诉/宗
预测重复投诉/宗	2 613	313
预测非重复投诉/宗	4 866	63 942
整体准确率	92.78%	—

2.7 3 个模型预测能力对比

在完全相同的 71 734 个测试集数据环境下，BP 神经网络模型预测重复投诉的准确率比 Logistic 回归模型提升了 10.52%。联合模型预测重复投诉的准确率比 Logistic 回归模型约提升了 12.69%，这说明，联合模型可以更全面地考虑线性和非线性因素对重复投诉风险的影响，从而获得更高的预测精度。3 个模型预测能力对比见表 12。

表 12 3 个模型预测能力对比

重复投诉预测模型	预测准确率	比 Logistic 回归模型提高
Logistic 回归模型	80.09%	—
BP 神经网络模型	90.61%	+10.52%
联合模型	92.78%	+12.69%

3 案例应用

3.1 全流程的重复投诉管理

某省会电信运营商的客服部门部署本文所提联合模型，涵盖了重复投诉的事前预防、事中预警和事后学习全流程。

在事前预防阶段，加强管控联合模型所发现的预测特征值，降低重复投诉的发生风险，具体如下。

- 加强对“工单总历时时长”“工单流转次数”“挂起次数”的管控，这些预测特征值反映服务效率和问题解决能力的水平。通过对这些指标针对性地改进，提高服务质量和效率。
- 加强对员工的培训，提高其服务意识和情商水平，以便更好地与客户沟通、了解客户需求并及时解决问题。
- 建立完善的客户反馈机制，及时收集、分析、处理和回复客户反馈信息，以便更好地了解客户需求和满意度，并对不足之处进行改进。

在事中预警阶段，基于联合模型的预警方法识别潜在的投诉风险，以便及时采取措施、避免事态扩大。该方法结合了 Logistic 模型和 BP 神经网络模型两种算法，能够快速准确地预测出可能存在的风险因素，进而提供相应的预警信息。

在事后学习阶段，本文采取了完整的风险管理体系，包括调查定责、监督学习等环节。调查定责环节对重复投诉事件进行核实，查明事实及责任；监督学习环节及时提炼重复投诉案例信息，并对联合模型的预测结果打标，通过有监督学习改进模型，提高预测精度。联合模型应用全流程如图 5 所示。

3.2 应用效果

联合模型的预测重复投诉性能优越，指导服务人员及时发现并解决客户的问题，提高了服务质量。通过对风险工单提前干预，重复投诉率下降约 3.2%，提高了服务质量。模型对生产的指导效果见表 13。

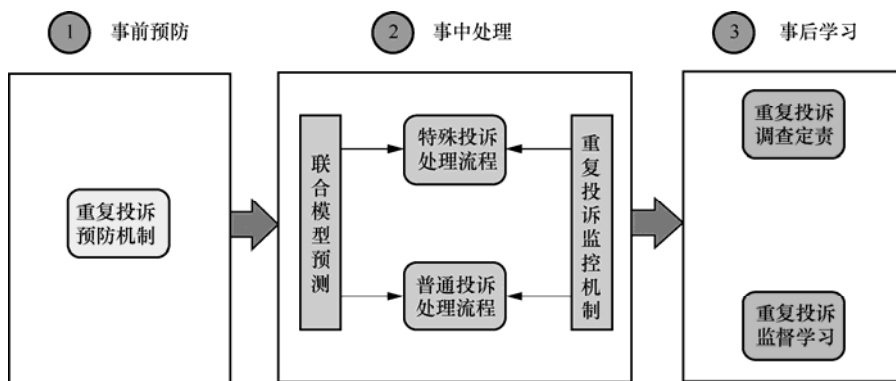


图 5 联合模型应用全流程

表 13 模型对生产的指导效果

干预前的重复投诉率	干预后的重复投诉率	下降
10.47%	7.27%	-3.2%

3.3 模型应用的创新性和与传统投诉管理方法的差异

该联合模型在客户重复投诉预测方面具有很强的创新性。相较于传统的投诉管理方法，例如，雷霆等^[3]和韩鹏^[4]对投诉管理的研究，该模型不仅可以实现事前的风险控制和主动干预，还能预测客户是否会重复投诉，解决了重复投诉率高的问题。

同时，该模型采用了机器学习算法，结合了大量数据进行建模，具有较高的精度和有效性，相比传统管理方式缺乏数据驱动的特点，更具优越性。而且该模型还通过联合建模的方式进一步提高了预测精度，使得实际应用效果显著，重复投诉率下降约 3.2%。

从预测精度和应用效果的角度来看，本文提出的联合模型，为运营商客服管理提供了新的思路和方法，实践表明其效果优秀。该模型集成了多种算法和模型，能够提供全面、科学的解决方案，从而提高客服管理效率和服务质量。

4 结束语

本文提出了基于联合神经网络的投诉预测模型，通过统计建模和机器学习建模，构建了

Logistic 回归模型，挖掘了影响重复投诉的关键因素，为提升服务质量提供了新视角。同时，本文比较了 BP 神经网络的重复投诉风险预测模型与 Logistic 回归模型的预测性能，发现 BP 神经网络模型的预测精度提高了 10.5%，这表明在大规模数据和高算力的支持下，BP 神经网络模型具有更高的预测性能。进一步地，本文将 Logistic 回归模型与 BP 神经网络模型相结合，构建了联合模型，实验结果显示，联合模型的预测精度提高了 12.7%，证实两种算法结合使用能够互补优势，提升模型整体性能。

此联合模型已被成功应用于某省会电信运营商的服务运营，显著改善了服务质量。展望未来，一方面，可以对模型进行进一步优化以提高预测精度；另一方面，可以研究更多适用于不同场景的风险预测模型，为提升服务质量提供更多有力支持。同时，随着数据规模和算力的不断提升，可以探索更多先进的人工智能技术在风险领域的应用，以期为我国电信运营商提供更高效、更优质的服务。

参考文献：

- [1] 中国政府网. 中华人民共和国消费者权益保护法[Z]. 2013-10-25.
GOV. Consumer rights protection law of the National People's Congress of the People's Republic of China[Z]. 2013-10-25.
- [2] 马晓亮, 刘英, 杜德泉, 等. 电信运营商 AI 客服平台技术研究与应用分析[J]. 电信科学, 2023, 39(9): 141-152.
MA X L, LIU Y, DU D Q, et al. Research and application anal-



- ysis of artificial intelligence-based customer service platform technologies for telecom operators[J]. Telecommunications Science, 2023, 39(9): 141-152.
- [3] 雷霆, 陈小鹏. 降低网络客户重复投诉率策略研究[J]. 信息与电脑(理论版), 2022, 34(1): 208-210, 215.
- LEI T, CHEN X P. Research on strategies of Xi'an Mobile to reduce repeated complaints rate of network customers [J]. Information & Computer, 2022, 34(1): 208-210, 215.
- [4] 韩鹏. 强化故障全程管控有效压降重复投诉[J]. 通信管理与技术, 2021, 175(5): 50-51, 62.
- HAN P. Strengthening full-process management of faults to effectively reduce repetitive complaints[J]. Technical & Operation Exchanges, 2021, 175(5): 50-51, 62.
- [5] 李航. 统计学习方法[M]. 北京: 清华大学出版社, 2012.
- LI H. Statistical learning methods[M]. Beijing: Tsinghua University Press, 2012.
- [6] 张建锋, 付强, 杜金柱. 基于 Logistic 模型的上市公司股票价格操纵预判研究[J]. 西安理工大学学报, 2018, 34(2): 240-245.
- ZHANG J F, FU Q, DU J Z. Study on the manipulation anticipation of listed companies' stock prices based on Logistic model[J]. Journal of Xi'an University of Technology, 2018, 34(2): 240-245.
- [7] 邵志高. 基于二元 logistic 回归模型的收入异动影响未来股票收益研究[J]. 湖南科技大学学报(社会科学版), 2017, 20(5): 103-111.
- SHAO Z G. On the impact of revenue abnormal variation on future stock returns based on the binary logistic model[J]. Journal of Hunan University of Science & Technology (Social Science Edition), 2017, 20(5): 103-111.
- [8] 陈蒙蒙, 方振红, 涂文怡, 等. 基于 Logistic 回归模型的心脏病预测模型构建及效果分析[J]. 医院管理论坛, 2022, 39(2): 32-35.
- CHEN M M, FANG Z H, TU W Y, et al. Regression model effect analysis of heart disease prediction model based on logistic[J]. Medical Management, 2022, 39(2): 32-35.
- [9] 孙敬怡. 基于智能算法的中青年居民健康状况与疾病预警模型研究[D]. 天津: 天津中医药大学, 2021.
- SUN J Y. Research on health status and early warning model of disease of young and middle-aged residents based on intelligent algorithm[D]. Tianjin: Tianjin University of Traditional Chinese Medicine, 2021.
- [10] 张艳莉. 基于 Logistic 回归算法的企业顾客抱怨管理机制的构建[J]. 科技通报, 2014, 30(5): 156-159.
- ZHANG Y L. Enterprise management mechanism for build customer complaints based on logistic regression algorithm[J]. Bulletin of Science and Technology, 2014, 30(5): 156-159.
- [11] 庄贵军, 朱美艳. 顾客抱怨行为与重购意愿的 logistic 回归分析[J]. 商业经济与管理, 2009, 211(5): 90-96.
- ZHUANG G J, ZHU M Y. Logistic regression analysis of customer complaint behavior and repurchase intention[J]. Journal of Business Economics, 2009, 211(5): 90-96.
- [12] 罗鑫, 张金林. 基于多时间尺度复合深度神经网络的股票价格预测[J]. 武汉金融, 2020, 249(9): 32-40.
- LUO X, ZHANG J L. Stock price prediction based on multi-time scale composite deep neural network [J]. Wuhan Finance, 2020, 249(9): 32-40.
- [13] 唐思源, 杨敏, 白金牛. 基于深度卷积神经网络的肺结节检测与识别[J]. 科学技术与工程, 2019, 19(22): 241-248.
- TANG S Y, YANG M, BAI J N. Lung nodules detection and recognition based on deep convolutional neural networks [J]. Science Technology and Engineering, 2019, 19(22): 241-248.
- [14] 周文杰, 严建峰, 杨璐. 基于深度学习的用户投诉预测模型研究[J]. 计算机应用研究, 2017, 34(5): 1428-1432.
- ZHOU W J, YAN J F, YANG L. Research on prediction model of complaint based on deep learning [J]. Application Research of Computers, 2017, 34(5): 1428-1432.
- [15] 穆晓玛. 主流客户投诉预测模型的对比与研究[J]. 智能计算机与应用, 2018, 8(5): 100-102, 105.
- MU X M. Comparison and research of several mainstream customer complaint prediction models [J]. Intelligent Computer and Applications, 2018, 8(5): 100-102, 105.
- [16] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
- ZHOU Z H. Machine learning [M]. Beijing: Tsinghua University Press, 2016.

[作者简介]



马晓亮 (1973-), 男, 西安电子科技大学博士生, 华南理工大学工商管理学院讲席教授, 中国电信股份有限公司广州分公司副总经理、高级工程师, 从事数据通信、互联网运营、数据挖掘、商业呼叫中心和运营商客服等工作, 主要研究方向为人工智能、自然语言处理和数据安全保护等。



刘英 (1979-), 男, 现就职于中国电信股份有限公司广州分公司, 主要研究方向为智能客服平台组件、自然语言处理等。



高洁 (1985-), 女, 现就职于中国电信股份有限公司广州分公司, 主要研究方向为机器学习、算法运用等。