

# 一种基于 AI 大数据的用户满意度预测算法

周文红 孙军亮 许妍青 杨嘉忱 孙凡晰

(中国电信股份有限公司 北京 100010)

**摘要** 随着大数据技术的发展和人工智能应用的广泛化,用户满意度预测已成为运营商解决信号差和网络质量问题,减少用户投诉并获取竞争优势的关键手段。文中提出了一种结合机器学习技术的用户满意度预测算法,通过分析用户行为数据和反馈,构建了一个高精度的预测模型。通过实验验证,该模型在多个数据集上表现出优越的预测准确性和良好的泛化能力。该算法的实现,对于理解用户需求和改进服务质量具有重要意义。

**关键词:** 大数据;用户满意度;预测算法;机器学习

**中图分类号** TN919

## An AI Big Data-Based Algorithm for Predicting User Satisfaction

ZHOU Wenhong, SUN Junliang, XU Yanqing, YANG Jiachen and SUN Fanxi

(China Telecom Corporation Limited, Beijing 100010, China)

**Abstract** With the development of big data technology and the widespread application of artificial intelligence, predicting user satisfaction has become a key method for operators to address poor signals and network quality issues, reduce customer complaints, and gain competitive advantages. This paper proposes a user satisfaction prediction algorithm that integrates machine learning techniques, building a high-accuracy prediction model through analyzing user behavior data and feedback. The model has demonstrated superior predictive accuracy and good generalizability across multiple datasets through experimental validation. The implementation of this algorithm is significant for understanding user needs and improving service quality.

**Key words** Big Data, User Satisfaction, Prediction Algorithm, Machine Learning

## 0 引言

随着技术的进步,尤其是数据分析与机器学习的不断发展,使得准确预测用户满意度成为可能。现有研究多侧重于通过传统统计方法进行满意度分析,而本文通过大规模用户数据,结合机器学习技术,不仅能提高预测的准确性,也为用户满意度的动态监测和实时优化提供了可能。该方法可帮助运营商在竞争激烈的市场环境中更好地理解客户需求,提高客户满意度和忠诚度。

## 1 AI 大数据技术概述

### 1.1 大数据技术的发展与应用

随着存储成本的降低和计算能力的提升,大数据技术已从初步的数据收集和存储发展到现在的复杂数据分析和实时数据处理。在通信行业,大数据技术被广泛应用于分析用户通信行为、优化网络流量管理。

### 1.2 人工智能与机器学习基础

人工智能(AI)通过模仿人类的思维过程,使计算机能

执行复杂任务,如语言理解、决策支持和图像识别等。机器学习作为人工智能的一个重要分支,它能让计算机基于经验自动改进性能<sup>[1]</sup>。机器学习的核心在于建立模型并使用算法来处理数据,从中学习并作出决策或预测。

### 1.3 AI 与大数据的结合趋势

AI 和大数据的结合,正推动技术创新向更深层次演进。AI 技术可以对大数据进行更深入的分析,发掘数据之间复杂的关联关系和模式,提供更精准的业务洞察。例如,通过机器学习模型可以对消费者数据进行分群,精准定位市场细分,以定制个性化营销策略。

## 2 用户满意度的评价指标

### 2.1 定义用户满意度

用户满意度通常被定义为消费者在购买或使用某种产品或服务后,其期望与实际感受之间的匹配程度。在通信领域,用户满意度则体现在用户对其生活区域内网络信号质量的感知,包括通话、上网、游戏、视频等业务的体验,以及服务人员态度的匹配等。这定义强调了满意度是主观评价,不仅受到产品或服务本身的影响,还受到个人期望、以

作者简介:周文红(1979—),硕士,高级工程师,研究方向为大数据。

往经验和情境因素的影响。在用户满意度高的情况下,消费者会认为服务或产品质量超过或符合他们的预期,相反则可能导致客户不满或转向竞争对手。

## 2.2 评价指标体系构建

建立一个有效的用户满意度评价指标体系是评估用户满意度的关键。这一体系应全面覆盖产品或服务的所有方面,并能反映出企业与消费者交互的各个环节。常见的评价指标包括产品质量、服务响应速度、服务人员的态度、价格公正性、解决问题的能力等<sup>[2]</sup>。在构建体系时,企业需考虑行业特性、目标市场和客户群体的特定需求。在通信行业中,用户满意度的评价可能重点关注网络覆盖的广度和质量、数据传输的速度,以及客服的响应效率和解决问题的能力等。

## 2.3 指标的量化与数据收集

将用户满意度的评价指标量化,是实现客观评估和持续改进的基础。在量化过程中,企业需确定每个指标的衡量方法和评价标准,这通常涉及到设计问卷、进行客户访谈、收集在线反馈等方法。例如,可以通过五点或十点满意度量表来收集用户对于各项服务的满意程度。此外,数据的收集不应仅限于事后的满意度调查,还可以包括交易数据、客户互动记录、社交媒体上的公开评价等多种来源,以构建更全面的满意度分析模型。

# 3 预测模型的构建

## 3.1 数据预处理与特征选择

在构建任何预测模型之前,首先需彻底清洗数据,确保输入数据的质量和准确性。数据清洗的过程包括识别并处理缺失值、异常值和重复数据等。例如,缺失值可以通过删除缺失数据的记录、使用平均值或中位数填充,或应用更复杂的插补技术如K-最近邻(K-NN)插补进行处理。对于异常值,可以通过Z分数或箱型图(IQR)方法识别并处理,以防止它们对模型预测造成不良影响。特征工程是机器学习中的关键环节,它涉及到从原始数据中提取、创建和选择那些对预测任务有用的特征,这一过程对模型的性能有着直接影响。首先,可以通过探索性数据分析(EDA)来理解数据的基本特征,识别可能的预测因子<sup>[3]</sup>。然后,基于业务理解和统计分析,可以创建新的特征,这些特征可能是原始数据的函数,如日志转换、多项式特征或交互项特征。特征选择是特征工程的另一重要部分,旨在减少特征空间的维度,提高模型的泛化能力。常用的特征选择方法包括过滤方法(Filter)、包裹方法(Wrapper)和嵌入方法(Embedded)。过滤方法如使用相关系数或卡方检验来评估各个特征与目标变量的相关性,选择相关性强的特征。包裹方法如递归特征消除(RFE)通过实际模型的性能来选择特征。嵌入方法则结合了特征选择和模型训练过程,如使用Lasso回归,该方法通过对回归系数施加L1正则化来实现特征的自动选择。

## 3.2 算法选择与模型训练

选择合适的机器学习算法是构建有效预测模型的关键步骤。对于用户满意度预测,可选的算法有很多,包括逻辑回归、决策树、支持向量机(SVM)、随机森林和梯度提升机(GBM)等。逻辑回归适用于结果为二分类的预测问题,它是一种简单且高效的算法,通过估计概率来进行分类,主要优点是模型输出具有很好的可解释性。决策树通过构造一个树形结构来做决策,能自动捕捉特征间的非线性关系,易于理解和实现,但容易过拟合。支持向量机(SVM)在高维空间中寻找最佳超平面来区分不同的类别,特别适合于特征维数较高的数据集。SVM试图最大化不同类别之间的间隔,使得模型具有较好的泛化能力。随机森林是一种集成学习方法,其能构建多个决策树,并将它们的预测结果进行汇总。随机森林对于数据的异常值和非平衡数据集具有很强的鲁棒性。选定算法后,需进行模型训练和参数调整。模型训练的目标是使用训练数据来让机器学习算法学习数据模式,参数调整则是为了最大化模型的预测性能。

在训练阶段,机器学习算法会尝试找出最佳的模型参数来拟合数据。例如,在决策树中,算法会学习如何在每一个节点处分割数据,以优化目标函数。在SVM中,则是学习如何放置决策边界以最大化分类间隔。大多数机器学习模型有多个参数需要调整,这些参数的设定对模型的性能影响巨大<sup>[4]</sup>。例如,随机森林中的树的数量、决策树的最大深度、SVM的C参数(错误项的惩罚系数)和核函数类型。参数调整通常使用网格搜索(Grid Search)和随机搜索(Random Search)等方法,结合交叉验证来确定最优参数组合。

## 3.3 模型评估与优化

交叉验证是一种评估模型泛化能力的重要技术,特别适用于防止机器学习模型的过拟合。这一方法通过将数据集分割成多个较小的子集来进行,常见的方法包括K折交叉验证和留一交叉验证。在K折交叉验证中,原始数据被分为K个大小相等的子集。其中,每个子集轮流作为验证数据集,其余的K-1个子集作为训练数据。如此,模型会被训练K次,每次都会得到一个模型性能的评估,最后的性能评估结果是这K次评估结果的平均值。这种方法的优点是每个数据点都既被用作训练又被用作测试,从而可以更全面地利用数据。通过交叉验证,可以减少模型因训练数据选择不同而产生的性能变异,提高模型评估的稳定性和可靠性。模型的性能评估是通过具体的指标来实现的,对于用户满意度预测模型,常用的性能评估指标包括以下几个方面。(1)准确率。其是最直观的性能度量,通过正确预测的数量除以总样本数来计算。(2)精确度和召回率。精确度指预测为正的样本中实际为正的比例,而召回率指实际为正的样本中被预测为正的比例。这两个指标通常一起使用,尤其是在数据集类别不平衡的情况下。(3)F1分数。这是精确度和召回率的调和平均,是一个综合考虑了精确度和召回率的性能指标。F1分数对于那些对精确度和召回率同等重视的场景非常有用<sup>[5]</sup>。

## 4 应用案例分析

### 4.1 案例背景

在通信行业,一家领先的运营商面临提高用户满意度和减少客户流失率的挑战。该运营商收集了大量用户交互数据,包括通话记录、数据使用行为、用户反馈以及客服互动记录。初步分析显示,尽管网络用户数稳步增长,但用户满意度的波动对客户忠诚度和品牌声誉有直接影响。运营商需要一种方法来准确预测哪些因素最影响用户满意度,并识别潜在不满意的客户,以便提前采取措施改善其网络体验。此外,运营商希望能通过优化网络服务和提供个性化的客户支持来提升用户满意度。

### 4.2 AI大数据解决方案

运营商决定采用基于机器学习的用户满意度预测模型来解决这一问题。其选择了随机森林算法,因为它在处理大型数据集和高维特征时表现良好,且能有效管理特征间的非线性关系。模型训练使用的数据来源包括以下几方面。(1)用户行为数据。包括用户上网时长、上网业务类型(如浏览、流媒体播放、在线游戏等)、断传率以及信号质量测量等。(2)用户反馈。来自在线调查和网络服务评价的文本数据。(3)客服互动记录。记录了用户投诉数据、服务请求以及解决问题的过程和效率等相关互动。模型训练首先进行数据预处理,包括将清洗数据、处理缺失值、转换文本数据作为数值特征(如TF-IDF权重)<sup>[6]</sup>。基于业务理解,从原始数据中提取有价值的特征,如用户满意度得分(通过分析产品评价计算得出)。使用交叉验证来优化随机森林参数,并进行模型训练。

### 4.3 训练结果

模型在不同参数设置下的交叉验证结果如表1所列。基于训练结果,模型在树的数量为200,最大深度为20时表现最佳。进一步调整模型,增加特征选择步骤,减少过拟合现象,并重新调整树的数量和深度。通过细化特征和调整模型参数,进一步提高模型的精确度和召回率。

表1 模型交叉验证结果

参数设置	准确率/%	精确度/%	召回率/%	F1分数/%
树的数量:100, 最大深度:10	85.2	86.1	84.3	85.2
树的数量:150, 最大深度:15	87.5	88.0	87.0	87.5
树的数量:200, 最大深度:20	89.3	89.8	88.7	89.2

### 4.4 模型应用效果

在优化后的用户满意度预测模型中可以发现,网络覆盖质量和客服响应效率是对用户满意度影响最大的因素。该发现基于模型中这两个特征的重要性评分,以及它们对

用户行为的统计影响。基于上述关键因素的分析,运营商采取了以下业务策略。(1)优化网络覆盖。通过升级基站、优化网络配置、增加小区覆盖来提高网络的覆盖质量。(2)提升客服团队的响应效率。引入更高效的客户服务管理系统,采用自动化工具来预先筛选和分类客户咨询,以及培训客服团队提高响应速度和服务质量。实施上述策略后,在一个季度内对优化效果进行了评估。评估标准主要基于以下几个方面。(1)用户满意度调查。通过在线调查收集用户对网络覆盖和客服满意度的反馈。(2)网络性能监控。使用统计工具监控网络的平均信号强度和用户的数据使用行为。(3)客服效率分析。分析客服响应时间和解决问题的效率,以及用户对服务的评价。结果如表2所列。

表2 模型应用结果分析

评估指标	改进前	改进后	变化率
平均页面加载时间/s	4.5	2.2	-51%
客服平均响应时间/min	10	4	-60%
用户满意度评分(满分5分)	3.2	4.5	+40.6%

从表2可以看出,页面加载速度和客服响应时间的改进直接提高了用户满意度评分。页面加载时间的提升减少了用户的等待时间,提高了用户的访问体验;客服响应时间的减少则直接提高了用户对问题解决的满意度。这些结果验证了模型预测的准确性和策略的有效性,表明通过精确的数据分析和策略实施能显著提升用户满意度,增强客户忠诚度和企业的市场竞争力。

## 5 结语

本文利用大数据和AI技术预测了用户满意度,且指导了实际业务决策。本文提出的模型通过优化关键性能指标,提高了用户满意度,验证了数据驱动策略在客户关系管理中的实用价值。未来,这种预测模型的应用范围将更加广泛,对业务的推动力也将更加显著。

### 参考文献

- [1] 付晓东,陈秋琳,冯艳.基于Double DQN的用户群体满意度最大化在线服务信誉度量[J].昆明理工大学学报(自然科学版),2024,49(4):105-117.
- [2] 侯宁.基于用户满意度和形状文法的青铜器气质特征提取及应用研究[J].洛阳理工学院学报(社会科学版),2024,39(4):80-85.
- [3] 程江洲,胡敏,刘闯.考虑需求响应和用户满意度的主动配电网优化调度[J].全球能源互联网,2024,7(4):443-453.
- [4] 李莹.基于用户体验的H5新闻用户满意度与持续使用意愿研究[D].武汉:武汉纺织大学,2024.
- [5] 张龙昌,白静.基于用户满意度的大数据服务可信评价与优化[J].计算机技术与发展,2023,33(8):1-8.
- [6] 董莹莹,葛阳,李坤树,等.融合AI模型在移网用户满意度预测中的应用研究[J].邮电设计技术,2022(8):75-78.