

基于融合AI模型的家庭宽带 满意度预测及应用研究

Research on Prediction and Application of Home Broadband Satisfaction Based on Fusion AI Model

张硕伟¹,刘贤松¹,陈 凌¹,任俊彦¹,杨治航²(1. 中国联通网络运营事业部,上海 200050;2. 中国联通天津分公司,天津 300000)

Zhang Shuowei¹,Liu Xiansong¹,Chen Ling¹,Ren Junyan¹,Yang Zhihang²(1. China Unicom Network Operations Division, Shanghai 200050, China;2. China Unicom Tianjin Branch, Tianjin 300000, China)

摘 要:

随着千兆家宽的普及和智慧家庭应用的快速发展,保障家庭宽带用户满意度的要求也有所提高。为了提升服务质量,增强用户黏性,构建一个基于融合AI模型的家庭宽带用户满意度预测方法。通过对工信部及本地满意度评分、宽带业务端到端数据的深入挖掘和分析,研究各个打分段用户的特征,结合人工智能构建融合型AI模型,实现家庭宽带满意度精准预测,对提升用户网络感知具有重要意义。

关键词:

宽带;满意度;人工智能

doi:10.12045/j.issn.1007-3043.2025.05.013

文章编号:1007-3043(2025)05-0074-04

中图分类号:TN915

文献标识码:A

开放科学(资源服务)标识码(OSID):



Abstract:

With the popularization of gigabit broadband and the rapid development of smart home applications, higher requirements are put forward to guarantee the satisfaction of home broadband users. In order to improve service quality and enhance user stickiness, it builds a home broadband user satisfaction prediction method based on the fusion AI model. Through in-depth mining and analysis of satisfaction scores from the Ministry of Industry, Information Technology and local areas, as well as end-to-end data of broadband services, it studies the characteristics of users in different score segments. Combined with artificial intelligence, a fusion AI model is built to achieve accurate prediction of home broadband satisfaction, which is of great significance for improving user network perception.

Keywords:

Broadband; Satisfaction; Artificial intelligence

引用格式:张硕伟,刘贤松,陈凌,等. 基于融合AI模型的家庭宽带满意度预测及应用研究[J]. 邮电设计技术,2025(5):74-77.

0 引言

近年来,随着家庭宽带业务的飞速发展和千兆家庭宽带的广泛普及,如何进一步提升用户黏性和满意度,已成为运营商的一个全新挑战^[1]。面对这一挑战,各大运营商已经提出并实施了多种预测用户满意度的方法。常见的方法有通过大数据分析家宽业务端到端数据并设定固定阈值,来对用户维度网络感知评

价进行分析、量化^[2];通过LightGBM算法进行家宽满意度评分预测^[3]。

本文提出了一种通过融合模型来预测家庭宽带满意度评分的方法,该方法基于内容侧、接入网侧、终端侧、服务质量等宽带端到端数据以及工信部和本地满意度评分数据,通过将分类模型和聚类模型融合,对用户满意度评分进行预测^[4]。该方法不仅预测准确性高,而且预测效率高,通过每月对用户进行满意度评分预测,并将预测评分结果提供给省分,以便其进行及时维系和排障,精准提升了家庭宽带用户的满意

收稿日期:2025-03-06

度。

1 宽带满意度影响因素分析

从家庭宽带用户端到端拓扑分段来看,宽带业务流程长、网元多、场景复杂,影响家宽用户满意度的因素众多^[5]。家庭宽带用户端到端拓扑流程如图1所示,本文从终端侧、接入网、数据网、内容侧以及服务侧5个方面对影响家庭宽带满意度的因素进行分析^[6]。

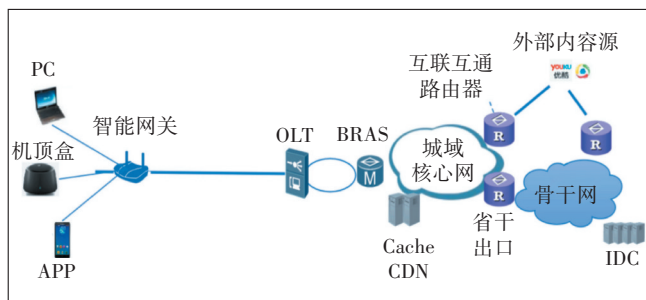


图1 家庭宽带用户端到端拓扑流程

a) 终端侧。主要指终端设备,如路由器、机顶盒等。据统计,家庭用户满意度差的原因有60%以上来自终端设备。

b) 接入网。PON网络是家庭宽带接入网络的关键节点,其各项性能指标都是影响家庭宽带用户满意度的关键指标,对分析用户满意度具有重要意义。

c) 数据网。上层数据网承载大客户等多种业务,具备很强的负载均衡和主备冗余能力,因此上层数据网很少发生故障。同时,数据网拓扑错综复杂,难以量化分析。因此本文不将数据网作为影响家庭宽带用户满意度的因素。

d) 内容侧。内容侧多指内容供应商。内容侧业务种类众多,不同的用户偏好的业务不同,不同的业务用户体验要求也不一样,因此业务体验是影响用户满意度的直接因素。

e) 服务侧。主要是指装维服务,包括家宽故障是否频繁发生、故障修复是否及时以及智家工程师排障服务态度等,也是影响家庭宽带用户满意度的重要因素^[7]。

综上所述,终端侧、接入网侧、内容侧、服务侧数据对用户网络感知有重要影响;数据网侧分析难度较大,而且对用户网络感知影响较小。用户的网络感知直接影响了用户对家庭宽带满意度的评分^[8],因此本文主要从终端侧、接入网侧、内容侧、服务侧筛选重要

指标来构建家庭宽带满意度预测模型。

2 家庭宽带满意度预测方法

本文采用了模型融合的方法,其大致思路是:首先构建分类算法,从数据集中单独分出10分和1分用户,再通过聚类算法预测剩余用户的评分,具体建模流程如图2所示。

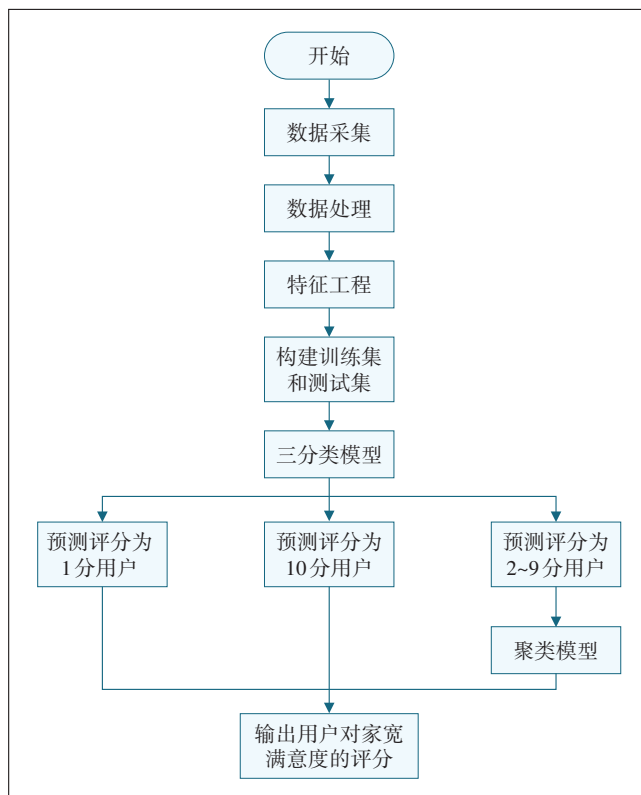


图2 融合模型整体流程

2.1 数据准备

本文家宽用户满意度预测使用的数据集是某省家宽满意度工信部评分以及省分本地评分,评分范围为1~10分,评分越高表示满意度越高。该数据集除了评分数据外,还包含了终端侧数据(路由器CPU占用率、路由器内存、Wi-Fi信号强度等)、接入网侧数据(OLT脱网次数、OLT脱网时长、ONU流入流出速率、ONU发射接收功率、用户访问连接时延、用户访问丢包数、用户访问传输时长等)、内容侧数据(某业务类占比、某时间段访问业务占比等)以及服务质量数据(前30天投诉单数量、30天内该用户修障单数量等)^[9]。通过筛选,最终数据集保留了129个特征,数据量约7000条,具体评分分布如图3所示。

从图3可以看出,1分和10分的评分占了整个数

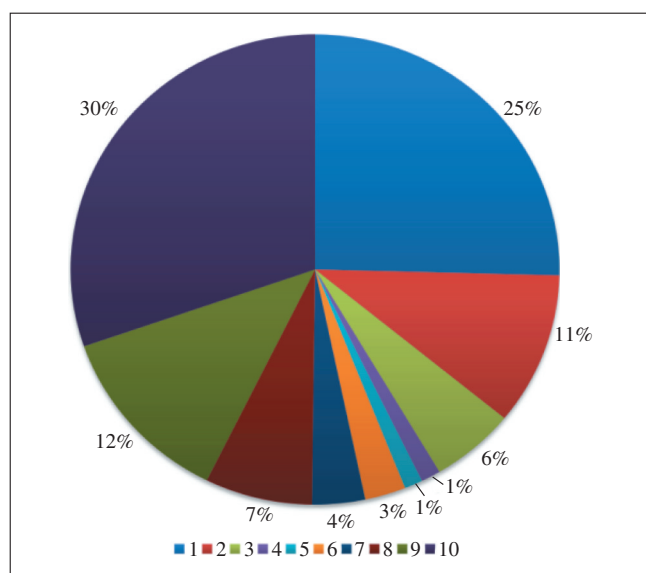


图3 用户真实评分分布

据集的55%,其中3~8分的数据量都非常少,每项占比均不超过10%。由此可以看出该数据集的评分分布极其不均衡,如果使用单一分类或聚类模型,预测结果会大幅集中在1分和10分,严重影响模型预测的准确性。

2.2 建立分类模型

为了解决评分数据分布不均衡的问题,本文通过分类模型先将1分、10分、2~9分的用户提取出来。首先分别筛选出数据集中的1分、10分、2~9分用户,然后将评分数据映射到0、1和2上,其中评分为1分的用户标记为0,评分为10分的用户标记为1、评分为2~9分的用户标记为2。通过在重新打标后的数据集上训练分类模型,得到可以提取1分、10分以及2~9分用户的三分类分类器^[10]。

本文选取了LightGBM作为分类模型。LightGBM是一个基于GBDT算法的框架^[11],但LightGBM在很多方面比XGBoost表现的更为优秀^[12],具有训练效率更高、内存使用更低、支持并行化计算以及能够处理大规模数据的优势。在分类任务中,LightGBM的优化目标是 minimized 损失函数(Loss Function)^[13],如式(1)所示。

$$\text{Loss} = \sum_{i=1}^n l(y_i, y_i^*) + \sum_{i=1}^t \Omega(f_i) \quad (1)$$

其中, n 表示数据集中的用户数量, y_i 表示真实用户评分, y_i^* 表示模型预测评分, t 表示决策树的个数, $\Omega(f_i)$ 表示抑制决策树复杂度, $l(y_i, y_i^*)$ 表示模型的评

价函数。

在分类模型中通常使用交叉熵函数作为评价函数,预测值与真实值越接近,那么交叉熵会越小,具体如式(2)所示。

$$l(y_i, y_i^*) = y_i \ln(y_i^*) + (1 - y_i) \ln(1 - y_i^*) \quad (2)$$

通常来说,模型的复杂度越高,其在训练集上的精度越高,但同时模型的泛化能力也会有所损失。LightGBM通过正则项 Ω 来抑制模型复杂度,使模型在具有更好的泛化性的同时不损失模型精度^[14]。

在构建LightGBM模型的过程中,本文使用sklearn库提供的随机搜索交叉验证方法寻找最优超参数,如树的深度、叶子个数、惩罚项系数、学习率等^[15]。经过多轮的训练和验证后,得到三分类分类模型,该分类模型能够筛选出1分、10分、2~9分用户。

2.3 建立聚类模型

考虑到2~9分的评分受用户评分习惯以及主观因素的影响较多,相同的网络条件会存在1~2分的评分误差,因此对2~9分的用户采用聚类模型。将数据集中评分为1分和评分为10分的用户剔除,仅留下评分为2~9分的用户作为训练样本。本文选取K-means作为聚类模型, K 值设置为8,评价函数使用组内误差平方(SSE)^[16],如式(3)所示。

$$\text{SSE} = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2 \quad (3)$$

经过无监督训练后,将2~9分用户数据集聚类为8类,每类选取用户评分频数最高的作为该类的最终评分,最终得到适用于2~9分用户的聚类模型。

2.4 模型融合

前文分别建立了1个分类模型和1个聚类模型,接下来将这2个模型进行融合,融合过程如下。

a) 将数据投入到分类器中,得到用户打分为1分、10分及2~9分每类的概率 p_i ,选取概率 p_i 最大的一类作为分类器的分类结果,并将此用户的预测评分设置为1分、10分或2~9分。

b) 将预测评分为2~9分的用户投入到聚类模型中,用户距离哪一类类中心比较近,便将该类评分作为此用户的预测评分。

3 模型评价与应用

3.1 模型评价

本文将数据集按照8:2的比例拆分成训练集和测

试集,并在相同训练集和测试集上进行对比试验。实验结果表明融合模型的F1值为82%,评分的平均绝对误差为1.65,相较于单一分类模型或聚类模型,其F1值提高了20%,平均绝对误差减少了约40%。

相较于单一分类模型或聚类模型,融合模型的评分预测更接近真实用户评分。因此,融合模型在提升评分精度的同时,也使得预测结果分布很接近真实分布,一线业务人员可利用预测结果更好地维系客户,提高用户黏性,同时及时排查家庭宽带用户感知不好的问题。

3.2 模型应用效果

经过上述建模和结果分析可知,本文提出的融合模型非常适合这种满意度评分分布不均匀的场景。在实际满意度评分调研中,发现用户的评分通常集中在最满意和最不满意这2个极端,如果直接进行分类或聚类,会使模型忽略中间评分,无法拟合真实的数据分布。本文首先将1分和10分用户识别出来,然后通过聚类模型将剩余分布比较均匀的2~9分用户数据进行聚类,这样可以将相同网络感知质量的用户聚在一起,尽可能提高预测准确度。

本文对某省分公司某月家庭宽带用户满意度进行了预测及结果验证。实验共计400万个用户,预测评分均值为8.8分,其中预测评分为10分的用户占比最高。该省分从预测评分为1~10分的用户中各抽取20名用户,通过电话回访调研其实际用户评分,其中真实评分和预测评分相同的用户有158人,预测评分与真实评分误差在2分以内的占88%。预测效果得到了该省分公司的认可。

4 结束语

本文通过对家庭宽带用户满意度相关数据的研究和分析,基于家宽业务端到端数据,提出一种融合AI模型的方法对用户满意度评分进行预测,经过数据采集、数据处理、模型构建、模型训练以及模型融合,最终输出满意度预测。单一的分类模型或聚类模型在评分数据分布极度不均衡的情况下预测精度不高,而融合模型很好地解决了这个问题。在模型的实际推广应用中,该方法也得到了客户的认可。

参考文献:

[1] 赖孙芳. 如何提升光宽带用户的网络质量感知[J]. 长江信息通信, 2021, 34(2): 202-204.

- [2] 陈雪芬,吴凯洲,刘雯,等. 家庭宽带用户网络感知评价的建模与综合分析[J]. 邮电设计技术, 2022(1): 48-53.
- [3] 何均均,王江舟. 基于LightGBM算法的家庭宽带满意度分类研究和实现[J]. 江苏通信, 2023, 39(2): 92-96.
- [4] 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.
- [5] 冯笑杰,梁童,方波,等. 基于用户感知的家宽端到端质量分析方案研究[J]. 电信工程技术与标准化, 2023, 36(4): 35-39.
- [6] 商冶,修志超. 构建面向用户和业务的宽带数字化运营体系[J]. 邮电设计技术, 2023(6): 1-7.
- [7] 林浩,张克刚,梁军君,等. 家庭宽带业务质量保障分析[J]. 广东通信技术, 2023, 43(6): 15-19.
- [8] 岳斌,胡慧敏. 基于业务感知的家庭宽带上网满意度影响因素研究[C]//天津市电子工业协会2023年年会论文集. 天津:天津市电子工业协会, 2023: 198-203.
- [9] 张延盛,徐银. 基于用户访问感知的家宽业务质差识别研究与实践[J]. 电信工程技术与标准化, 2020, 33(8): 63-68.
- [10] 祝喆,周奕昕,于洋,等. 精准识别关键问题、快速提升移网用户满意度[J]. 江苏通信, 2021, 37(4): 56-58.
- [11] SONG R W, CHEN S D, DENG B L, et al. eXtreme gradient boosting for identifying individual users across different digital devices [C]// Web-Age Information Management. Cham: Springer, 2016: 43-54.
- [12] FRIEDMAN J. Greedy function approximation; a gradient boosting machine[J]. Annals of Statistics, 2001, 29(5): 1189-1232.
- [13] 高洁,张涛,程新洲,等. 一种基于LightGBM机器学习算法的用户年龄及性别预测方法[J]. 邮电设计技术, 2019(9): 36-39.
- [14] WANG D H, ZHANG Y, ZHAO Y. An effective miRNA classification method in breast cancer patients [C]// ICCBB '17: Proceedings of the 2017 International Conference on Computational Biology and Bioinformatics. Newark: Association for Computing Machinery, 2017: 7-11.
- [15] MACARINI L, MERRY W J, PATERNAIN G P. On the growth rate of leaf-wise intersections[J]. Journal of Symplectic Geometry, 2012, 10(4): 601-653.
- [16] 张瑜,周丽永. 基于k-means聚类算法的西安市游客消费体验评价[J]. 现代商业, 2023(8): 73-78.

作者简介:

张硕伟,毕业于上海师范大学,工程师,硕士,主要从事机器学习、深度学习算法研发及固网和云网智能化的相关工作;刘贤松,毕业于武汉水利电力大学(武汉大学),中国联通网创新中心网络AI产品部总监,硕士,主要从事网络服务智能化产品研发和网络服务智能化产品推广管理工作;陈凌,毕业于复旦大学,硕士,主要从事算法技术的研究工作;任俊彦,毕业于英国伦敦大学学院,硕士,主要从事接入网与数据网解决方案、智能化规划及行业大模型解决方案相关工作;杨治航,毕业于英国伦敦国王学院,工程师,硕士,主要从事宽带接入网络质量分析、客户感知提升、数智化运营以及网络产品创新相关的工作。