融合AI模型在移网用户满意度预测中的应用研究

Research on Application of Fusion AI Model in Prediction of Mobile Users' Satisfaction

董莹莹，葛　阳，李坤树，沈　斌，黄双双（中国联通智网创新中心，江苏 南京 210000）
Dong Yingying，Ge Yang，Li Kunshu，Shen Bin，Huang Shuangshuang（Intelligent Network & Innovation Center of China Unicom，Nanjing 210000，China）

摘　要：
移动网络用户的满意度是运营商经营质量水平的一个重要评价标准，高水平的用户满意度评分是提升存量用户忠诚度的重要保障。基于O域信令数据、B域用户基础数据、用户感知数据、用户投诉数据及满意度打分数据等进行大数据分析，深入研究各个打分段用户的特征，结合人工智能算法搭建AI模型，精准预测全网用户的满意度。

关键词：
信令数据；投诉数据；打分数据；人工智能；满意度
doi：10.12045/j.issn.1007-3043.2022.08.014
文章编号：1007-3043（2022）08-0075-04
中图分类号：TN919
文献标识码：A
开放科学（资源服务）标识码（OSID）：

Abstract：
The mobile users' satisfaction is an important evaluation criterion for telecom operators' operational quality. A high score is an important guarantee for maintaining the loyalty of existing users. It conducts big data analysis based on O domain signaling data，B domain user basic data，user perception data，user complaint data，and satisfaction rating data，deeply studies the characteristics of each segmented users，builds an AI model combined with artificial intelligence algorithm to accurately predict the users' satisfaction of the network.

Keywords：
Signaling data；Complaint data；Scoring data；AI；Satisfaction level

引用格式：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.

0 引言

随着移网用户数量不断增长，提升用户满意度水平对运营商高质量发展具有重要意义。本文基于O域信令数据、B域用户出账数据、用户感知数据，再结合移网用户的投诉数据与打分数据，其中投诉数据中主要关注移网网络质量、套餐资费、服务质量等维度，打分数据是通过发放调查问卷的方式收集用户整体满意度评价打分，打分机制是10分制，从1分至10分，分别表示不满意至满意的程度，基于上述5类数据，分析各个分数段的用户数分布及特征，搭建分类加回归的融合AI模型，对全网用户进行满意度打分预测，将预测打分清单提供给网络部，支撑网络部进行差异化重保，高效提升移网用户满意度水平。

收稿日期：2022-06-28

1 背景、数据集和标准模型

1.1 背景

近年来，随着移动网络的迅猛发展尤其是5G网络的发展，以及中国联通、中国移动、中国电信3家运营商竞争格局的改变，移动网络用户对上网速度、网络感知、套餐资费等提出了更高要求，如何更好地进行客户服务工作对运营商来说是一个新的挑战。目前，

各运营商均提出了一些预测用户满意度的方法，常见的方法有通过客服部收集的历史用户评分，对低分用户进行回访维系；或者通过分析用户的基础信息（年龄、性别、行为偏好等）和用户的行为数据（用户的感知数据、历史记录等），对用户进行大数据画像，指导运营商对低分用户群进行重点保障。本文提出了一种基于融合机器学习模型来预测用户打分的方法，该方法基于用户的信令数据、感知数据、投诉数据、CES满意度打分数据等海量数据，利用大数据对用户满意度进行分析，再结合机器学习回归与分类融合模型对用户满意度进行预测，不仅预测准确率高，且实时性较强，通过每天或每周对用户进行预测，然后将预测打分数据清单输出给客服部进行及时干预维系，从而精准提升移网用户的满意度。

1.2 数据集

本文用户满意度预测模型使用的数据集是某省5G用户满意度的打分。打分为10分制，分数越高代表满意度越高。该数据集除了打分数据外，还包含了脱敏的用户基础数据（年龄、性别、常住地、入网时间等）、用户的投诉数据（投诉分类、渠道等）以及用户的感知数据（流量、时长、时延、下载业务次数、卡顿数、掉线次数等）。最终的数据集有172个特征。数据集中共有约37 000条打分数据，平均分为8.04分，方差为11.21，具体的分数分布如图1所示。

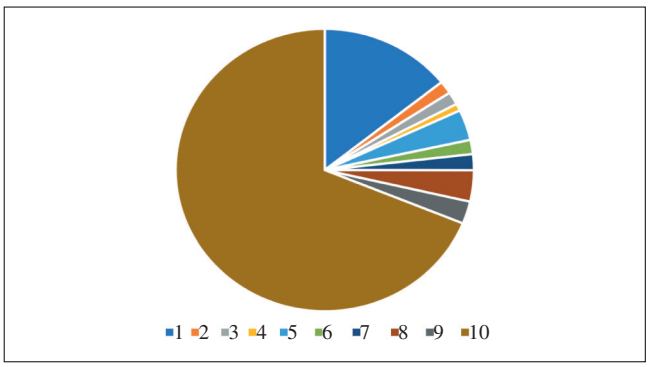

图1　打分分数用户分布

从图1可以发现，10分和1分的打分成绩占了数据总量的83%。其余分数（除了10分、8分、5分和1分）的数据量都非常少，占比均不到4%。由此可以看出这个数据集非常不均衡，这会降低模型的预测准确性。

1.3 标准模型及结果

本文建立了线性回归模型作为标准模型，对数据集进行诸如缺失值填充、对类别变量进行独热编码，降维等一系列操作后建模，最终得到的模型在测试集上的平均绝对误差为2.6。虽然平均绝对误差在可以接受的范围内，但观察预测出的数据本身还是暴露出了一些问题，如图2所示。

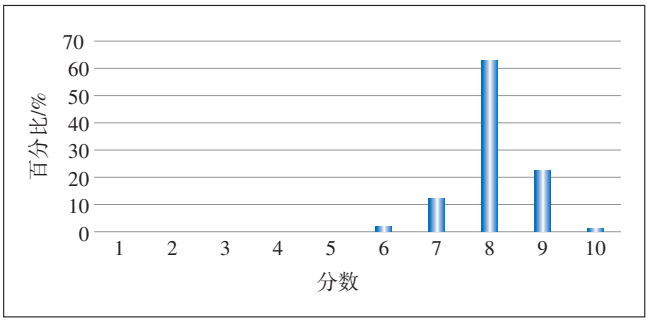

图2　预测打分用户占比

由图2可以看出，大部分的预测结果集中在6~9分，只有极少数的预测结果为1、3、4或10分，这与实际打分结果的分布完全不符。造成这种结果的原因是模型在尝试缩小平均误差时会将预测结果尽可能地靠近训练集的平均打分。因此预测结果中占比最多的自然是数据集打分的平均数。这样的预测结果会在实际的生产环境中给一线人员造成困惑，所以需要提出一些新的方法解决这个问题。

2 模型融合

为解决上述问题，本文采用模型融合的方法。其大致思路是，首先建立分类模型，从数据集中单独分出10分和1分用户，再通过回归模型预测剩余用户的分数，具体的过程如下。

2.1 建立分类模型

为解决数据分布不均衡问题，本文通过分类模型将以上打分为10分和1分的用户提取出来。本文针对10分和1分建立了模型A和模型B 2个分类模型。

模型A的任务是选出数据集中的10分用户，为训练模型A，首先将打分数据映射到0和1上。其中打分为10分的用户标记为1，其他用户标记为0，通过在修改后的数据集上训练分类模型，得到一个可以用来提取满意度打分为10分的分类器。与之相对的，模型B的任务是选出打分为1分的用户。同理，将数据集中打分为1分的用户标记为1，其余用户标记为0，从而训练能提取1分用户的分类器。

为了训练这2个分类模型，使用了LightGBM工具

库，一个基于梯度提升（Gradient Boosting）的机器学习库，它通过使用梯度提升决策树（Gradient Boosting Decision Tree）算法，实现了精准预测、快速训练、分布式和大规模部署等能力。LightGBM在分类任务中旨在最小化以下损失函数（Loss Function）。

$$\text{Loss} = \sum_{i=1}^{n} l\left(y_i, y_i^*\right) + \sum_{i=1}^{t} C(f_i) \tag{1}$$

式中：

n——数据集中的用户数量

y_i——用户的真实打分数据

y_i^*——模型预测的数据

t——决策树的个数

f_i——单个决策树

$C(f_i)$——衡量决策树复杂度的方程，该方程会随着决策树复杂度的增加而增加

$l\left(y_i, y_i^*\right)$——单个用户的模型评价函数

在分类模型中使用了交叉熵（Cross Entropy）函数作为评价，当预测结果和真实打分越接近，交叉熵会越小，具体公式如下。

$$l\left(y_i, y_i^*\right) = y_i \ln\left(y_i^*\right) + \left(1 - y_i\right)\ln\left(1 - y_i^*\right) \tag{2}$$

通常而言，复杂的模型往往预测精度更高，但同时也会损失掉一定的泛化能力。因此LightGBM通过最小化损失函数，可以使模型兼备较好的预测能力和泛化能力。

在通过LightGBM构建模型的过程中，往往需要手动选取一定的超参数（Hyperparameters），例如决策树的深度、树叶个数、惩罚项系数等。为实现最好效果，本文使用了sklearn提供的随机搜索交叉检验算法（Random Search Cross Validation）为2个不同目标的分类器寻找最优超参数。经过一系列的训练和验证后，便生成了2个拥有各自独立目标的分类器，其中一个用于筛选出满意度打分为1分的用户，而另一个用于筛选出满意度打分为10分用户。接下来的任务便是训练回归模型。

2.2 建立回归模型

建立回归模型需要对数据集做一定的修改。笔者将数据集中打分为1分和打分为10分的用户删除，仅留下打分为2~9分的用户作为训练样本。通过LightGBM中的回归模块建立回归模型。与分类模型相比，回归模型的评价函数变成了平均绝对误差（Mean Absolute Error）。

$$l'\left(y_i, y_i^*\right) = \left| y_i - y_i^* \right| \tag{3}$$

相较于常用的均方误差，平均绝对误差更适合本文满意度的使用场景。在经过与之前分类模型相同的训练、模型选择等过程后，便得到了适用于2~9分用户的回归模型。

2.3 模型融合

前文的2个过程建立了2个分类模型和1个回归模型，接下来在预测过程中需要融合上述3个模型的结果，具体融合方法如下所述。

a）将数据投入分类器A中，得到用户打分为10分的概率p_a，当p_a大于某个设定的阈值a时，便将此用户的预测分值设定为10。

b）然后将数据投入分类器B中，得到用户打分为1分的概率p_b，当p_b大于某个设定的阈值b时，便将此用户的预测分值设定为1。

c）将剩余的预测打分不为1分和不为10分的用户投入回归模型中，将回归模型预测结果四舍五入为整数作为这批用户的打分。

上述步骤将3个模型的结果融合到了一起。为了寻找到最佳阈值，本文使用网格化搜索（Grid Search）的方法，以最小化整体平均绝对误差为目标，经过寻找，将2个阈值设定为a=0.63，b=0.28。

2.4 模型融合流程

基于上述融合模型，具体的数据处理及建模流程如图3所示。

3 模型评价与应用

3.1 模型评价

本文在相同的测试集上进行了试验，结果发现融合模型的平均绝对误差仅为1.35，相较线性回归模型的平均绝对误差减少了约48%。本文模型预测结果的数据分布图4所示。

相较于线性回归的结果，融合模型预测打分的分布更接近真实用户打分的分布。因此，融合模型在减小了整体预测误差的同时，也使得预测结果更接近真实分布，使得一线业务人员可以更好地利用数据提升用户满意度打分。

3.2 模型应用效果

由以上建模过程和结果分析可知，本文提出的融合模型非常适合客户满意度打分预测的场景。在实际满意度打分场景下，用户的打分往往集中在最满意

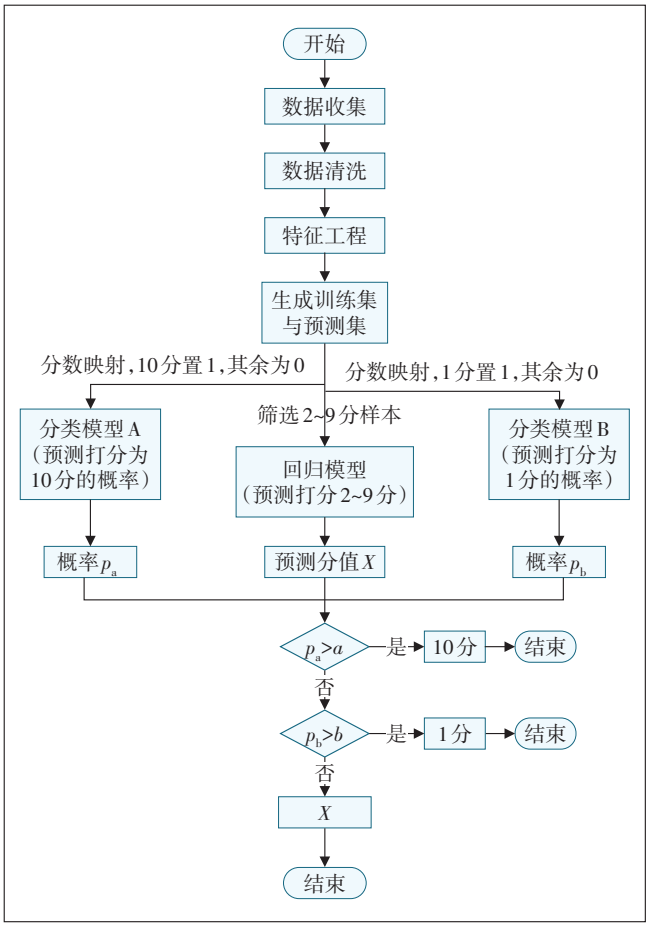

图3　数据处理集建模流程

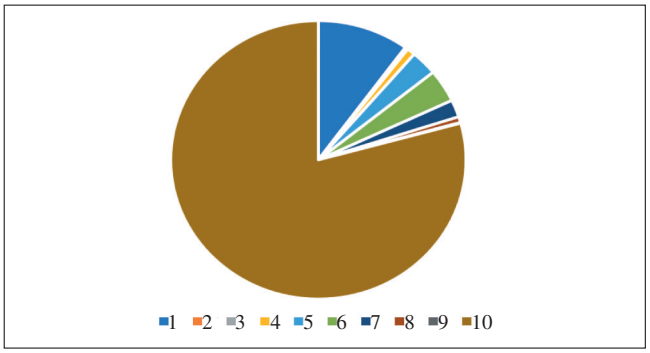

图4　预测打分结果占比

和最不满意两端，直接进行回归，会使预测分数在平均分数附近而忽略了数据的分布特征。本文首先使用分类器筛选出这2种占比最多的打分，再将剩余的分数通过回归模型拟合，这样可以在保留数据分布特征的同时尽可能提高预测准确度。

本文对某省分公司A全月打分用户进行预测及结果验证，共计2.5万个用户，实际打分均值为8.8分，预测打分均值为9.2分，较实际均值高0.4；其中预测打分为10分的用户中，75.6%的用户实际打分大于8分，模型预测效果大大高于预期。

另外，针对某省分公司B提供的工信部回访的5 000个用户进行效果验证，并将预测打分与用户实际打分结果进行对比分析，预测打分平均分为7.61，实际打分平均分为7.89，较实际均分低0.28分；其中预测打分为10分的用户中，实际打分大于8分的用户占比82.8%，实际打分小于4分的用户占比仅3.6%。预测效果得到了省分B的认可，准确率较高，大大提高了重保效率。

4 结束语

本文通过对移动网络用户满意度数据的分析，提出了一种融合机器学习模型的方法对用户满意度打分进行预测，详细介绍了数据采集、数据处理、分类模型建立、回归模型建立、模型融合、模型评价以及模型应用效果等方面的内容。本文在选定融合AI模型之前也对单一的回归模型和分类模型进行了建模测试，由于用户满意度打分数据分布极其不均衡，单一的回归与分类模型的预测效果均不尽如人意，预测的打分分布与实际打分分布偏差较大。为了解决上述问题，本文提出了一种基于多模型动态融合的方法，在提升精度的同时，模型的预测结果分布也更符合实际打分分布。在模型的实际应用与推广中，该方法也得到了客户的一致认可。

参考文献：

[1] TANG G H，ZENG H. Evaluation of tourism E-commerce user satisfaction[J]. Journal of Organizational and End User Computing，2021，33(5)：25-41.

[2] 章凯兵. 基于神经网络的用户满意度预测研究——以yelp数据集为例[D]. 赣州：江西理工大学，2020.

[3] 刘凡. 电信用户满意度预测研究——以浙江省为例[D]. 杭州：浙江工商大学，2020.

[4] 关守姝，孟岩，韩伟，等. 基于电力大数据的用户用电感知研究[J]. 电力大数据，2019，22(10)：58-64.

[5] 陆赟，杨晓康，李翔明，等. 移动通信运营商客户满意度提升策略研究[J]. 通信世界，2021(9)：19-20.

作者简介：

董莹莹，毕业于西安工程大学，工程师，硕士，主要从事面向市场的精准营销与用户识别算法研发工作；葛阳，毕业于南京邮电大学，工程师，硕士，主要从事面向市场支撑相关需求推广工作；李坤树，毕业于加州大学尔湾分校，工程师，硕士，主要从事用户识别算法工作；沈斌，毕业于河海大学，学士，主要从事智能运维类系统产品相关研发工作；黄双双，毕业于东南大学，硕士，主要从事面向运维支撑的OSS系统的产品研发设计工作。