

基于AI的无线网络用户满意度分析

User Satisfaction Analysis of Wireless Network Based on Artificial Intelligence

李露, 李一喆 (中国联通网络技术研究院, 北京 100048)

Li Lu, Li Yizhe (China Unicom Network Technology Research Institute, Beijing 100048, China)

摘要:

利用人工智能算法,以用户侧数据和网络侧数据为基础,分析用户群体特征,找出影响用户网络感知评分的关键因素。基于决策树算法,建立用户网络感知关联模型,预测用户对通信网络的满意度,及时发现用户对于网络贬损的真正痛点,为运营商网络建设和运行维护部门制定提升用户的网络感知策略提供依据,从而提升用户体验。

Abstract:

Using artificial intelligence algorithm, based on user-side data and network measurement data, it analyzes user group characteristics and finds out the key factors affecting user network perception score. Based on the decision tree algorithm, the user network perception association model is established to predict the user's satisfaction, and the key points of the user for the network impairment are discovered in time. The algorithm model provides a basis for the operator network construction and maintenance department to formulate the network awareness strategy for the user. Moreover, the research helps improve the user experience.

Keywords:

User satisfaction; Artificial intelligence; Machine learning; Decision tree

引用格式: 李露, 李一喆. 基于AI的无线网络用户满意度分析[J]. 邮电设计技术, 2018(12): 64-67.

关键词:

用户满意度; 人工智能; 机器学习; 决策树

doi: 10.12045/j.issn.1007-3043.2018.12.014

中图分类号: TP181

文献标识码: A

文章编号: 1007-3043(2018)12-0064-04

0 前言

终端用户体验质量是决定未来5G网络快速成功发展的关键因素之一。根据某咨询公司的调查,因为对产品或服务不满意,82%的用户选择终止该业务,并且还会引起更严重的连锁反应,平均有另外13个人受到这种对服务不满评价的影响。因此,持续提高用户满意度对于通信网络运营商具有十分重要的意义。运营商通过提高服务质量、网络性能来提升用户满意度,有助于增加用户黏性,提高企业在同行业中的竞争地位。

1 用户满意度分析系统

移动通信网络的无线质量一直是导致用户满意

度降低的关键因素之一。网络中的覆盖质量、掉线率、下行用户速率等问题,会直接影响用户体验。基于网络的无线质量的满意度分析,不涉及用户主观信息,仅根据几种测量的网络质量参数给出用户满意度的网络关联模型,可以被认为是无参考客观模型的一个子类。然而,仅依靠网络的无线质量参数去分析用户满意度,是不全面的,往往会忽略用户侧的主观因素。用户侧的主观影响因素有年龄、使用服务过程中对网络各层级的主观感受、套餐类型、套餐费用等。结合网络侧的客观因素和用户侧的主观因素对用户满意度进行分析将更准确反映用户服务体验的关联模型。根据用户历史数据和网络质量数据,基于人工智能数据分析方法,可建立图1所示的用户满意度分析系统流程。

用户满意度分析系统是基于多维数据的分析系统。系统从用户角度去感知和分析网络的业务信息,通过对大数据的挖掘和分析,从网络、用户、终端、业务

收稿日期: 2018-11-01

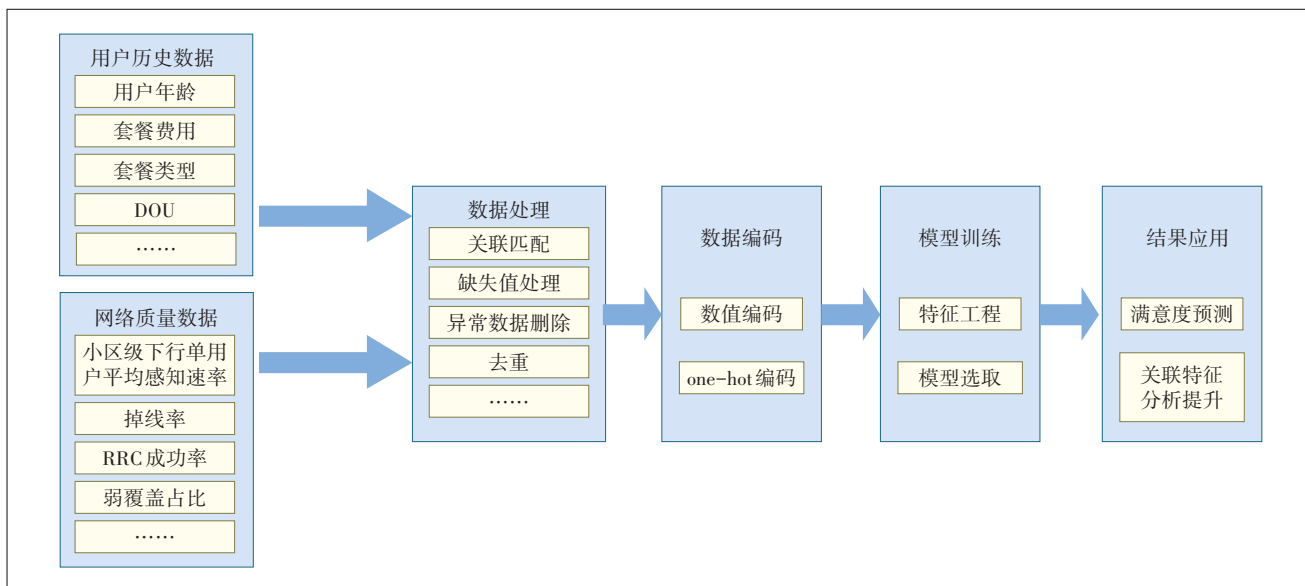


图1 用户满意度分析系统流程

等多个维度,对用户业务服务过程中的满意度的关联特征进行全方位的分析挖掘。

用户数据系统首先基于用户使用的网格ID将用户侧数据和网络质量数据关联匹配,并且对缺失数据进行标识处理,对异常数据进行删除(比如MOU、DOU为负值等违背业务逻辑或常理的脏数据)、去重。然后将数据中的中文属性进行数值编码或one-hot编码。由于数据维度太大,需要利用特征工程来选择对用户满意度作用明显的特征。利用已经确定的特征选取机器学习模型,如决策树、线性回归、神经网络、支持向量机等。特征工程和算法模型选择往往需要重复执行多次,才能确定。确定算法模型和特征之后,可以调整算法参数,提高结果准确度。得到机器学习模型之后,可以根据已有参数预测用户满意度,对不满意用户建立特征关联体系,及时调整涉及的重要服务参数(如网络质量、用户套餐类型等),提升用户满意度,避免用户离网。

2 用户满意度算法模型分析

2.1 特征选择

特征选择的过程其实就是用户满意度关联特征模型建立的过程。用于用户满意度分析的特征数量众多,直接输入到模型中会严重影响模型的预测效率和精度。特征选择可以解决上述问题,首先,可以避免属性过多造成的维数灾难问题,其次,可以去除不相关特

征,降低学习任务的难度。特征选择是为了减少特征数量,选取重要特征,从而提高模型的预测效率。所以,特征选择也是机器学习模型建立的一个过程,是用户满意度预测模型的前提工作。

将调查到用户的满意度作为标签,分为满意和不满意。用户的其他信息以及网络的参数作为特征输入。特征选择的过程其实就是用户满意度。

本文所用数据为常驻地/高业务量区域的网络质量和用户感知评分详细特征信息。由于特征维度众多,现根据专业知识将部分特征删除。用户使用网格有多个,但网络参数特征名称一致,为表示简略,将其写为“第一至第五名流量宫格”“第一至第五名语音宫格”,实际为多个属性。表格中的信息包括用户满意度信息、用户个人信息和网格级网络质量信息,具体如表1所示。

用过滤式特征选择方法对数据集进行特征选择。对每一维的特征“打分”,即给每一维的特征赋予权重,权重代表着该维特征的重要性,然后依据权重排序。论文利用随机森林进行特征选择。随机森林训练决策树的时候,可以计算出每个特征平均减少了多少树的不纯度,并把它平均减少的不纯度作为特征选择的值。

从图2中可知,基于随机森林算法,与用户满意度相关性较强的特征分别为提及B室内型号标签、提及C语音型号稳定性标签、提及B上网标签、提及B室外信号标签、提及A网络标签、客户年龄、第一名流量宫

表1 用户满意度模型输入特征属性名称

提及A网络标签	提及B语音标签	提及B上网标签	提及C语音信号覆盖标签	提及C语音信号稳定性标签	提及C上网信号覆盖标签	提及C上网信号稳定性标签	提及C上网速度标签	提及B室内信号标签	提及B室外信号标签
业务类型	VIP级别	付费模式	产品类别	已入网年限	是否上网卡	客户年龄	客户性别	DOU	非漫游的总流量数
MOU	主叫的总分钟数	被叫的总分钟数	非漫游的总分钟数	月度月小区总数	ARPU	套餐费用	超套月流量费用	超套月语音费用	超套月增值费用
融合类型	老人机标签	第一至第五名流量宫格流量	第一至第五名语音宫格语音	住宅宫格弱覆盖占比110	住宅宫格日均下行流量	住宅宫格小区级上行单用户平均感知速率	住宅宫格小区级下行单用户平均感知速率	住宅宫格倒流比	住宅宫格ERAB建立成功率
住宅宫格掉线率	住宅宫格RRC成功率	住宅宫格CQI质差占比	工作宫格弱覆盖占比110	工作宫格日均下行流量	工作宫格小区级上行单用户平均感知速率	工作宫格小区级下行单用户平均感知速率	工作宫格倒流比	工作宫格ERAB建立成功率	工作宫格掉线率
工作宫格RRC成功率	工作宫格CQI质差占比	第一至第五名语音宫格弱覆盖占比110	第一至第五名语音宫格日均下行流量	第一至第五名语音宫格小区级上行单用户平均感知速率	第一至第五名语音宫格小区级下行单用户平均感知速率	第一至第五名语音宫格倒流比	第一至第五名语音宫格ERAB建立成功率	第一至第五名语音宫格掉线率	第一至第五名语音宫格RRC成功率
第一至第五名语音宫格CQI质差占比	第一至第五名流量宫格弱覆盖占比110	第一至第五名流量宫格日均下行流量	第一至第五名流量宫格小区级上行单用户平均感知速率	第一至第五名流量宫格小区级下行单用户平均感知速率	第一至第五名流量宫格倒流比	第一至第五名流量宫格ERAB建立成功率	第一至第五名流量宫格掉线率	第一至第五名流量宫格RRC成功率	第一至第五名流量宫格CQI质差占比
调研时间	用户归属省份	用户归属地(市)							

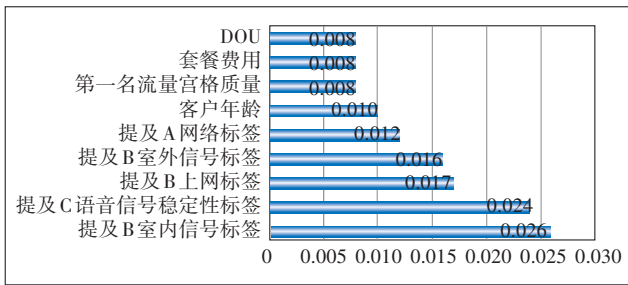


图2 特征与用户满意度的相关性

格网络质量、套餐费用和DOU。这表明上述的特征是模型建立不可或缺的信息,但是其他特征对用户的满意度也有组合影响。

为了排除与用户满意度完全不相关的特征,利用贪心算法找出不相关特征。设输入特征集合 X 为 $\{X_1, X_2, X_3, \dots, X_n\}$,分别将 X_1 至 X_n 抽取后,计算训练模型的准确性。通过分析得出,可将调研时间、用户归属省份、是否在网、是否本期出账、是否上网卡、是否合约计划、合约时效期、终端类型、终端厂家、终端型号、是否锁屏、互联网套餐等特征删除。综上可知,用户终端情况和是否有合约业务对用户满意度基本无影响,要提高不满意用户的满意度,可以从室内信号覆盖、语音信号稳定性、常用流量宫格质量等多个方面提升改善。

2.2 用户满意度预测模型

基于2.1节中的特征分析,可以建立用户满意度预测模型。用户满意度预测模型有利于运营商提前感知

用户对网络服务质量的体验情况。根据预测,运营商可通过提前调整网络参数、推荐用户更换套餐类型、调整套餐内容等方式,及时提升用户对网络服务的满意度,避免用户离网。用户满意度的预测分析可以映射为一个二分类问题。常用的机器学习二分类方法有线性回归、决策树、SVM、Adaboost等。

本文基于机器学习理论中的梯度提升决策树GBDT算法建立用户满意度预测模型。算法中的Boosting是一种模型组合的思想,对于分类和回归问题,单一模型往往非常复杂而且容易出现过拟合现象,解决办法是训练多个简单快速的弱分类器,通过加权投票等方式组合到一起成为预测准确的强分类器。

假设输入 N 个训练样本集合表示为 $\{X_i, y_i\}^N$,其中向量 X_i 包含 k 个参数, $X_i = \{x_{i1}, \dots, x_{ik}\}$ 。模型目标是在样本空间上,找到最优的预测函数 $F^*(X)$,使 X 映射到 y 的损失函数 $L(y, F(X))$ 最小,即:

$$F^*(X) = \arg \min_{F(X)} E_{y,x} [L(y, F(X))]$$

损失函数的形式为:

$$L(y, F(X)) = (y - F(X))^2$$

对上述过程进行求解,具体算法如表2所示。

2.3 模型结果分析

本文介绍的用户满意度预测模型是一个二分类问题,可以将样例根据其真实类别与学习器预测类别的组合划分为真正例、假正例、真反例、假反例4种情形,

表2 预测模型的具体算法

算法1 GBDT-用户满意度预测模型
输入: 训练集 $\{X_i, y_i\}^n$, 迭代次数 M
初始化: $F_0(x)=0$
循环迭代: For $m=1$ to M do:
对每个样本的函数估计值做logistic变换
遍历所有样本的每个类别的概率, 求每个样本在第 k 类上概率梯度
第 i 个样本属于第 k 个类别的残差=真实的概率-估计的概率
沿着梯度方法学习到 J 个叶子结点的回归树
求每个叶子节点的增益
更新所有样本在第 k 类下的估计值
End

另TP、FP、TN、FN分别表示其对应的样例数,则显然有TP+FP+TN+FN=样例总数。定义查准率为P,查全率为R。具体计算公式为: $P=TP/(TP+FP)$, $R=TP/(TP+FN)$ 。因为本文主要关心不满意用户的群体特征,所以将不满意用户设为正例。由于F1值是精确率和召回率的调和均值,即 $F1=2 \times P \times R / (P+R)$,相当于精确率和召回率的综合评价指标,更能反映算法模型的优劣,因此,用F1值来衡量机器学习模型的性能。

基于用户满意度分析问题,本文对多项式回归算法、SVM算法、决策树回归算法(DTR)、Adaboost算法和梯度增强决策树算法5种算法进行了仿真分析,具体结果如图3所示。从图3中可知,决策树算法明显优于其他算法,F1值可达80%以上。其中,GBDT决策树算法最优,F1值可达到83%。

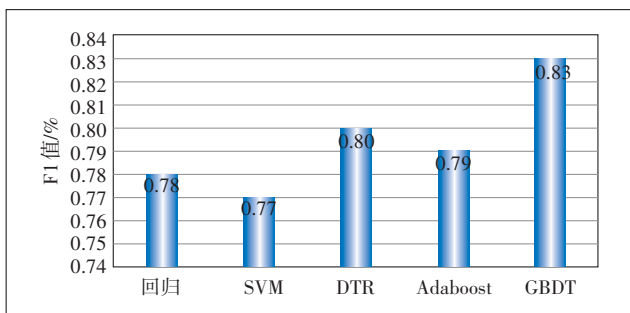


图3 算法模型F1值对比图

综上所述,可以基于机器学习算法模型对用户满意度进行预测分析,从客户角度去感知、分析网络和业务信息,提供面向用户的端到端的运营分析能力,对业务使用过程中的质量和特征进行全方位的分析挖掘。基于AI的用户满意度分析,一方面可以通过特征选择的分析过程,建立用户满意度与多维特征间的关联,发现用户对于网络贬损的痛点;另一方面可以通过预测用户满意度,及时帮助运营商网络建设和运行维护部门制定提升用户的网络感知策略。

3 结束语

本文提出了一种基于AI的用户满意度分析方法,得出了用户满意度的关联特征和用户满意度预测模型,对运营商网络建设和运维部制定提升用户的网络感知策略具有一定指导意义。

未来5G网络需要借助人工智能算法演进发展,为用户提供智能的网络服务。本文的分析还可以扩展到用户多维智能价值体系分析,通过全方位数据采集,并基于大数据和人工智能的全方位挖掘,为网络提供端到端、全业务生命周期的运营支撑,从而实现运营智慧化,如潜在离网用户挖掘分析、主动关怀、提升客户维挽率;主动式投诉处理方式,做前瞻客服推送;深度解析,多维度用户标签,立体用户画像;深度学习,用户关联信息分析等。

参考文献:

- [1] 邹蕾,张先锋. 人工智能及其发展应用[J]. 信息安全, 2012(2):11-13.
- [2] 毛毅. 人工智能研究热点及其发展方向[J]. 技术与市场, 2008(3):4.
- [3] 董慧,盛凌志. 基于用户感知的移动网络优化体系及关键技术[J]. 电信工程技术与标准化, 2011(11):1-4.
- [4] 斯坦尔. UMTS蜂窝网络系统的QoS与QoE管理[M]. 吕召彪,王慧,李沛,译.北京:机械工业出版社,2009:5-6.
- [5] 陈光. 移动用户感知数据监控分析系统的设计与实现[D]. 北京:北京邮电大学,2008.
- [6] 孙志军,薛磊,许阳明,等. 深度学习研究综述[J]. 计算机应用研究, 2012,29(8):2806-2810.
- [7] HAN J, KAMBER M. Data Mining: Concepts and Techniques[M]. 北京:机械工业出版社,2012.
- [8] ITU-T Recommendation P. 10/G. 100, Vocabulary for performance and quality of service[S]. 2006.
- [9] 孙志军,薛磊,许阳明,等. 深度学习研究综述[J]. 计算机应用研究, 2012,29(8):2806-2810.
- [10] WANG Q, DE CHUN B A, WANG X D, et al. Diagnosis of RH-KTB Vacuum System Based on Decision Tree ID3 Theory[J]. Journal of Iron & Steel Research, 2006, 18(4):59-62.
- [11] PETR P. DATA Mining[J]. Tools and Techniques, 1994, 29(1):47.

作者简介:

李露,毕业于北京邮电大学,工程师,硕士,主要研究方向为5G通信、人工智能等;李一喆,毕业于北京邮电大学,高级工程师,博士,AIIA电信项目组边缘计算工作组副组长,主要研究方向为AI在移动通信中的应用。

