

基于大数据分析的 潜在固移融合用户挖掘研究

Research on Potential Fixed-Mobile Convergence User Mining
Based on Big Data Analysis

张晴晴,张 涛,韩玉辉,程新洲,王云云,高 洁(中国联通研究院,北京 100048)

Zhang Qingqing, Zhang Tao, Han Yuhui, Cheng Xinzhou, Wang Yunyun, Gao Jie (China Unicom Research Institute, Beijing 100048, China)

摘 要:

随着人口红利的消失和公众通信市场的饱和,发展固移融合用户成为当前运营商的一项重要战略。运营商通过融合业务创新,将移动业务与宽带业务进行捆绑,发展融合业务用户,增强用户黏性,提升业务营收。基于移动网络大数据和宽带网络大数据的联合分析,提出了基于模型的潜在固移融合目标用户挖掘方法体系,将该方法体系运用于市场单移转融营销场景中,可更加精准地发掘潜在目标用户,提升融合用户转化率。

关键词:

大数据分析;固移融合;精准营销;用户识别
doi:10.12045/j.issn.1007-3043.2021.11.007
文章编号:1007-3043(2021)11-0031-04
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 disappearance of the demographic dividend and the saturation of the public communication market, the development of fixed-mobile converged users has become an important strategy for operators. Through convergent service innovation, operators can bundle mobile services with broadband services and develop converged service users, which can enhance user stickiness and increase service revenue. Based on the joint analysis of mobile network big data and broadband network big data, it proposes a model-based method system for mining potential fixed-mobile converged target users, and applies this method system to the market for single mobile contract user transferring to fusion contract precision marketing scenario, which can more accurately explore potential target users and increase the conversion rate of convergence users.

Keywords:

Big data analysis; Fixed-mobile fusion; Precision marketing; User identification

引用格式:张晴晴,张涛,韩玉辉,等. 基于大数据分析的潜在固移融合用户挖掘研究[J]. 邮电设计技术,2021(11):31-34.

0 引言

在整体收入增速放缓、公众市场饱和的大背景下,各大运营商均在探索如何在公众市场实现存量客户价值运营,同时对增量客户进行有效挖掘。其中发展融合业务用户便成为运营商加固维稳存量客户,同

时有效挖掘增量客户的有效利器。所谓融合业务,是运营商根据其业务特点将多类产品进行捆绑销售,常见的捆绑策略为将固网产品、移动网产品和IPTV等产品进行融合,通过各类融合套餐产品的创新性设计可实现客户价值提升,同时提升客户满意度。融合产品将是势不可挡的大趋势,未来运营商的产品将随着业务边界不断扩展向深度融合和广度覆盖方向转变,故对于融合用户的发展和挖掘将是未来运营商客户运营的一项重要工作。

中国联通对于融合捆绑的策略相对于友商起步

基金项目:工业和信息化部大数据产业发展试点示范项目(5G大数据跨行业异构融合创新应用试点示范)

收稿日期:2021-09-22

较晚,虽然也有固移套餐、主副卡(亲情卡)等产品,但整体融合力度稍弱,所以对于融合用户发展的市场空间较大。尤其在携号转网服务全面开放后,大力发展融合业务将有助于中国联通加固维稳存量客户,同时也有助于实现异网拉新。本文基于移动网络大数据和宽带网络大数据的联合分析,针对单移用户的移动网络业务行为特征,构建机器学习模型识别其真实用宽带状态,并推送给市场前端进行精准营销,促进用户进行固移融合套餐产品的迁转。中国联通各省分公司可以根据自身资源禀赋和用户规模占比,分别采取不同的固移用户发展策略。

1 固移融合用户挖掘流程和方法体系架构

为了实现对单移用户中潜在宽带用户的精准识别,本文基于移动网络大数据和宽带网络大数据的联合分析,构建了潜在宽带用户识别的方法体系架构,如图1所示。

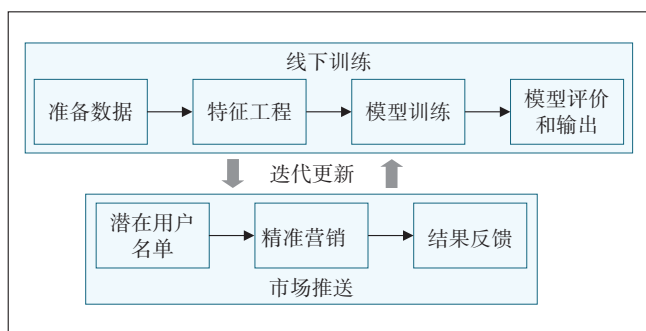


图1 潜在宽带用户识别流程

该框架由2部分组成:线下训练部分和市场推送部分。

a) 线下训练部分。主要目的是利用现有的历史数据得到理想的分类模型,并将训练好的稳定鲁棒模型用于现网数据识别,得到潜在宽带用户列表。

b) 市场推送部分。将模型得到的后台结果推送到市场前端进行精准营销,市场前端通过不同触点触达用户并进行融合产品营销。由于最终用户是否成功办理转化会涉及各类原因,故前端在接触客户时要分别记录模型预测准确率以及办理意向率,并将结果反馈给线下训练模型,从而实现模型的迭代训练。

2 基于模型的目标用户识别方法

2.1 数据准备

首先进行数据准备,构造模型训练所需的正负样

本集合。根据BSS端用户的签约信息,收集有宽带业务和无宽带业务用户的历史OSS数据和BSS数据。对于无宽带业务用户,仅通过套餐签约识别会出现不准确的情况,要结合家庭关系库以及OSS用户行为特征来进行准确识别,为模型训练打好数据基础。另外在数据准备过程中的另一个挑战是正负样本不均衡,会导致模型过拟合问题,所以在数据准备阶段要尽可能的收集正负样本集合。

2.2 特征工程

特征工程是提高机器学习模型表现和准确性的重要步骤。对于潜在宽带用户识别问题,基于对业务目标的理解以及手中掌握的数据,构造特征集合。本文在进行特征选择时,会通过可视化的方式,比较有宽带用户和无宽带用户在各个特征上的差异性,将有明显区分性的特征加入到特征集合中。图2展示了有宽带用户和非宽带用户的小时流量使用对比,可以看到有宽带用户的夜间流量使用有明显的下降特征。图3展示了有宽带用户和非宽带用户在白天(8:00-18:00)的流量与夜晚流量(19:00-24:00)的流量对比,可以发现无宽带用户无论在白天还是晚上的流量总体消耗都更多,尤其在夜晚会更加明显,无宽带用户与有宽带用户夜晚流量比值为1.6倍,要大于二者白天流量比值的1.2倍。其他特征的构造方法类似,这里不再赘述,最终形成的部分用户特征总结如表1所示,分别构造了O域特征与B域特征共约40个特征。

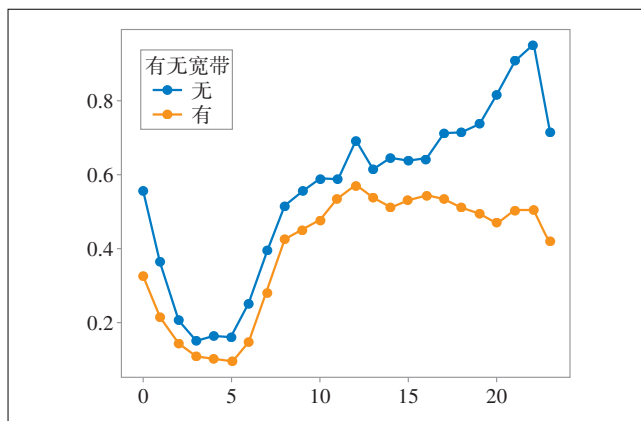


图2 有宽带用户和无宽带用户的小时流量对比

2.3 模型训练

对于模型训练部分,由于训练集合正负样本的不平衡特性,选择具有类权重参数的Class Weighted eXtreme Gradient Boosting (XGBoost) 作为模型来进行模型训练。对于分类中不同样本数量的类别,分别赋予

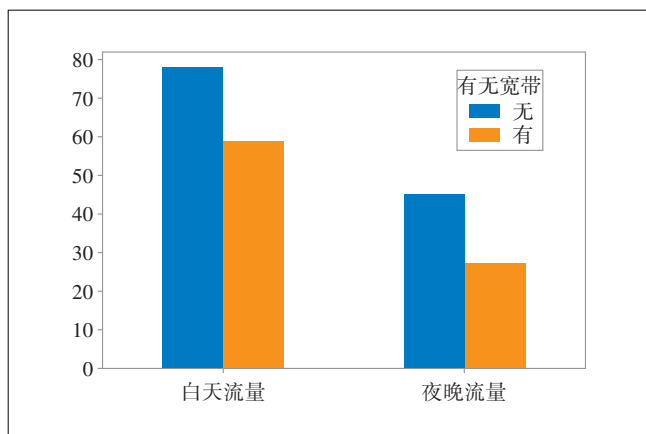


图3 有宽带用户和无宽带用户白天和夜晚流量对比

表1 O域与B域特征工程表

类别	特征名称
O域特征	按小时统计流量指标(共24h)
	白天总流量(8:00-18:00)
	夜间总流量(19:00-24:00)
	工作日白天总流量
	工作日夜间总流量
	工作日夜间流量占比
	节假日白天总流量
	节假日夜间总流量
	节假日夜间流量占比
B域特征	年龄
	性别
	在网时长
	套餐内总流量
	超套费用

不同权重的方法,具体操作是设置类样本权重反比于类样本数量。XGBoost的最小化目标函数公式如下:

$$L(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (1)$$

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \quad (2)$$

式(1)和(2)分为2个部分,第1部分为损失函数,第2部分为正则化参数。对于XGBoost,在模型训练时,可通过调节参数'scale_pos_weight'值来平衡正负权重。

2.4 模型评价和输出

在模型评估时,应选择与业务问题相匹配的评估方法。本文中的潜在用户挖掘问题是二分类问题,对于二分类模型,可采用多种不同的评估方式,如AUC(Area Under Roc Curve)、F1值、查准率(Precision)、查

全率(Recall)等。为了体现模型预测的准确性,将实例分为正类(Positive/+)或负类(Negative/-),对于模型是否预测正确,可形成混淆矩阵,基于得到的混淆矩阵结果可计算F1值、查准率和查全率。三者的计算公式如下:

$$\text{Precision} = \frac{m}{m+n} \quad (3)$$

$$\text{Recall} = \frac{m}{m+p} \quad (4)$$

$$F1 = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (5)$$

根据交叉验证模型在测试集上的表现,得到模型评价指标结果:Precision为51%,Recall为36.8%,F1值为42.4%。图4显示了Roc曲线结果,AUC值为0.648。同时为了验证模型的泛化能力,绘制了学习曲线,学习曲线是将训练集误差和交叉验证集误差在不同样本点数量下的误差进行对比,从图5的学习曲线结果来看,模型具备较好的泛化能力。

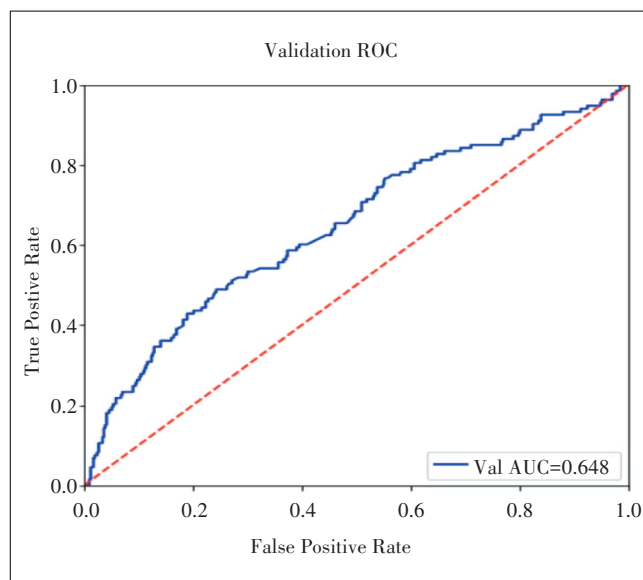


图4 潜在宽带用户识别Roc曲线

3 现网验证效果

将形成的潜在用户列表推送到市场前端开展各种形式的触达及融合产品推广活动,在触达过程中记录用户真实宽带状态,以此作为2种方法的直接验证指标。图6为验证效果。由结果可以看出,命中到的真实无宽带用户及有宽带用户比例均高于随机组,说明方法是有效的,能够对现网中用户的真实状态进行

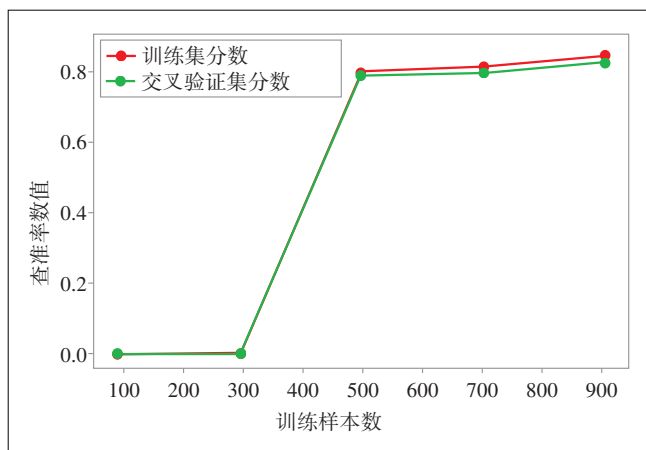


图5 潜在宽带用户模型学习曲线

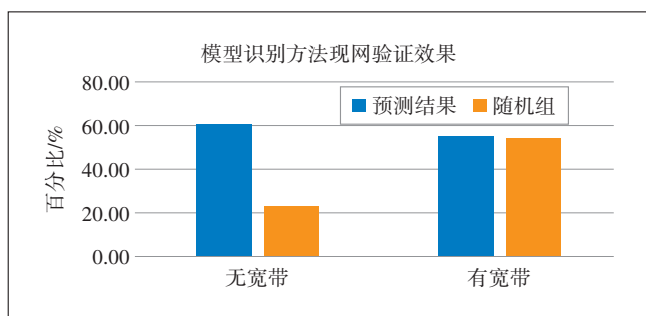


图6 模型识别方法现网验证效果

更有效的判断。

4 结论

运营商发展融合业务用户是势不可挡的大趋势。本文基于移动网络大数据和宽带网络大数据的联合分析,提出基于模型的潜在固移融合目标用户挖掘方法体系可以显著提高目标用户的识别率,同时将结果推送到市场前端进行精准营销,促进用户进行固移融合套餐产品的迁转。现网实际验证的结果表明,运用本文提出的方法发展固移融合用户,不仅可以提升单用户 ARPU 值,也有助于提高用户体验和用户忠诚度。在后续的研究中,根据市场前端的反馈,将对算法进行不断迭代,进一步提升整体模型的精准性。

参考文献:

[1] 曹蕊,胡万里. 电信运营商大融合业务及数字化转型[J]. 信息技术与网络安全, 2021, 40(7): 6-11.
[2] 周晓东. Z电信公司家庭宽带业务营销策略研究[D]. 长春: 吉林大学, 2019.
[3] 董润莎,徐争莉,袁明强,等. 基于机器学习的用户离网预测研究[J]. 邮电设计技术, 2018(10): 1-5.

[4] 艾达,罗爱平. 移动通信重入网用户识别算法分析研究[J]. 西安邮电学院学报, 2012, 17(3): 30-33.
[5] 张岳,车维崧,罗远源,等. 运营商宽带数字化运营方法探析[J]. 邮电设计技术, 2021(6): 41-45.
[6] 刘健峰. 佛山市电信家庭宽带业务营销策略研究[D]. 广州: 中山大学, 2020.
[7] XU L X, SHAO G L, CAO Y, et al. Research on telecom big data platform of LTE/5G mobile networks [C]//2019 IEEE International Conferences on Ubiquitous Computing & Communications (IUCC) and Data Science and Computational Intelligence (DSCI) and Smart Computing, Networking and Services (SmartCNS). IEEE, 2019.
[8] 石晓利. 大数据在宽带精准营销业务中的应用[J]. 中外企业家, 2019(16): 235.
[9] 柯晓燕,刘万里. “单宽”转融合精准营销运营商保存激增之道[J]. 通信世界, 2020(20): 13-15.
[10] 淡培彦. 基于大数据分析的家庭宽带业务精准营销研究[D]. 北京: 北京化工大学, 2019.
[11] SUN S L. Signal and Information Processing, Networking and Computers [M]//ZHU C B, CHENG X Z, CHENG C, et al. A novel architecture and machine learning algorithm for the prediction of user equipment replacing. Singapore: Springer, 2019.
[12] 左政. 基于数据挖掘的固网及手机融合套餐用户流失分析[D]. 长沙: 湖南大学, 2019.
[13] CHENG X Z. A novel Big Data based Telecom Operation architecture [C]//Proceedings of the 1st International Congress on Signal and Information Processing, Networking and Computers (ICSINC), 2015.
[14] 郑舒方. K-means 聚类算法在通信运营商精准营销中的应用研究[D]. 长春: 吉林大学, 2019.
[15] ZHANG T, LI H B, XU L X, et al. Comprehensive IoT SIM card anomaly detection algorithm based on big data [C]//2019 IEEE International Conferences on Ubiquitous Computing & Communications (IUCC) and Data Science and Computational Intelligence (DSCI) and Smart Computing, Networking and Services (SmartCNS). IEEE, 2019.
[16] 袁石勇,周兵. 一种基于固网大数据挖掘用户联系电话的方法及系统: CN107332694A [P]. 2020-12-29.

作者简介:

张晴晴, 毕业于中国科学院大学, 助理工程师, 硕士, 主要从事 DPI 解析及数据分析, 大数据行业应用等工作; 张涛, 毕业于北京邮电大学, 工程师, 硕士, 主要从事运营商大数据分析及应用行业创新应用产品的研究工作; 韩玉辉, 毕业于北京邮电大学, 网络智能运营研究中心 DPI 解析团队负责人, 高级工程师, 硕士, 主要从事 DPI 及通信大数据行业应用等工作; 程新洲, 毕业于北京邮电大学, 中国联通研究院网络智能运营研究中心总监, 教授级高级工程师, 硕士, 主要从事通信大数据分析及架构等研究工作; 王云云, 毕业于北京邮电大学, 助理工程师, 硕士, 主要从事网络态势感知分析及研究, 网络大数据行业应用产品研发等工作; 高洁, 毕业于北京邮电大学, 高级工程师, 硕士, 主要从事大数据 IT 工具研发及数据分析工作。