A communication user behavior analysis model based on big data

Hailong Deng Haibo Dou Yinan Yang Gang Wang Honggan Li

China United Network Communications Co., Ltd. Shenzhen Branch, Shenzhen, Guangdong, 518000, China

Abstract

This paper discusses the construction and application of communication user behavior analysis model based on big data. This paper first introduces the basic concepts and application value of big data technology in the analysis of communication user behavior, and elaborates the data acquisition, preprocessing and quality evaluation standards. Then, this paper focuses on the construction of clustering analysis model, including the extraction of user behavior features, the selection and application of K-means clustering algorithm, and the correlation analysis between clustering results and user behavior characteristics. Through the K-means algorithm, the users were successfully divided into different groups, and the behavioral characteristics of each group were analyzed in depth. Finally, this paper discusses the application value of clustering results in user portrait construction, precision marketing strategy formulation, and product function optimization suggestions, which provides strong data support for operators.

Keywords

big data technology; communication user conduct; analytical models; Data preprocessing

基于大数据的通信用户行为分析模型

邓海龙 窦海波 杨宜楠 王刚 李弘甘

中国联合网络通信有限公司深圳市分公司, 中国·广东 深圳 518000

摘要

本文探讨了基于大数据的通信用户行为分析模型的构建与应用。文章首先介绍了大数据技术在通信用户行为分析中的基本概念和应用价值,详细阐述了数据的获取、预处理以及质量评估标准。接着,文章重点讨论了聚类分析模型的构建,包括用户行为特征的提取、K-means聚类算法的选择与应用,以及聚类结果与用户行为特征的关联分析。通过K-means算法,成功将用户划分为不同群体,并深入分析了各群体的行为特征。最后,文章探讨了聚类结果在用户画像构建、精准营销策略制定以及产品功能优化建议等方面的应用价值,为运营商提供了有力的数据支持。

关键词

大数据技术; 通信用户行为; 分析模型; 数据预处理

1 引言

随着信息技术的飞速发展,大数据已成为各行各业不可或缺的宝贵资源。在通信领域,用户行为数据海量且复杂,蕴含着丰富的价值信息。本文旨在探索基于大数据的通信用户行为分析模型,通过深入挖掘用户通信数据,揭示用户行为特征与偏好,为运营商提供精准的用户画像和决策支持。通过构建高效的数据处理与分析框架,为提升用户体验、优化网络资源分配及促进通信业务创新提供有力支撑,进一步推动通信行业的智能化发展。

2 大数据技术与通信用户行为分析

2.1 大数据技术的基本概念

大数据技术的基本概念涉及对海量、高增长率和多样化数据的处理与分析。这种技术旨在从各种类型的数据集合中快速挖掘出有价值的信息和知识。大数据通常指那些规模巨大、类型复杂多样,超出了传统数据库软件工具能力范围的数据集合。其核心特征可以概括为“4V”: Volume(大量)、Velocity(高速)、Variety(多样)和 Veracity(真实性)。大数据技术不仅仅关注数据的收集,更注重数据的存储、处理、分析和可视化。它利用一系列先进的技术和方法,如分布式技术、流处理技术等,来处理和分析这些庞大的数据集。通过这些技术,大数据技术能够极大地增加数据处理速度,确保系统能够实时响应,从而洞察数据背后的规律和趋势。此外,大数据技术与云计算、人工智能等先进技术存在着紧密的联系。云平台提供了强大的计算和存储能力,为大数据

【作者简介】邓海龙(1978-),男,中国广东廉江人,本科,工程师,从事通信互联网研究。

处理提供了坚实的基础。而人工智能技术的发展,尤其是机器学习和深度学习,为大数据分析提供了强大的工具,进一步推动了数据挖掘和利用的效率与精准度^[1]。

2.2 大数据在通信用户行为分析中的应用

大数据在通信用户行为分析中扮演着至关重要的角色。随着移动通信技术的飞速发展,通信运营商能够收集并储存海量的用户数据,这些数据涵盖了用户的呼叫记录、短信记录、网络流量记录等,具有极高的信息价值。大数据技术的运用,使得运营商能够深入分析用户的历史记录和行为习惯,从而揭示用户的特点和习惯。例如,通过大数据分析,可以描绘出用户的社交圈子和社交活跃度,了解用户所处的社交群体特征。同时,还能分析用户的上网记录,挖掘用户的兴趣爱好,进而提供更为个性化的服务和内容推荐。此外,大数据在预测用户行为方面也展现出强大的能力。通过分析用户的呼叫、短信和上网行为等数据,可以预测用户的流量使用情况、是否可能流失等,为运营商提供决策支持。运营商可以据此及时调整流量套餐、推出挽留措施等,以提高用户满意度和黏性^[2]。

大数据在通信用户行为分析中的应用,不仅有助于运营商更好地了解用户需求,提供个性化服务,还能优化资源配置,降低运营成本。随着技术的不断进步和应用场景的持续拓展,大数据在通信用户行为分析中的价值将得到进一步凸显^[3]。

3 通信用户行为数据获取与预处理

3.1 数据来源及获取方式

数据来源及获取方式是关键内容。通信用户行为数据主要来源于多个渠道,包括运营商自身的业务数据、计费数据、客户数据以及设备数据。业务数据涵盖了用户的通话、短信、流量消耗等通信行为,计费数据则提供了用户套餐订购、消费明细等财务信息。客户数据主要来源于用户注册、实名认证等环节,包括姓名、性别、年龄等个人信息。设备数据则涉及用户手机型号、操作系统、应用安装情况等设备相关信息。此外,运营商还与第三方合作伙伴共享数据,获取更为丰富的外部数据资源,如社交媒体、论坛、博客等公开网络数据。在数据获取方式上,运营商采用多种技术手段,如日志采集、数据抓取、API接口调用和数据同步等。日志采集通过收集运营商系统的日志文件,获取用户行为、业务数据等相关信息。数据抓取则利用网络爬虫技术,从公开网络中抓取数据。API接口调用与第三方合作伙伴建立数据共享机制,实现数据的高效获取。数据同步则通过数据库同步、数据导入导出等方式,整合内部各个系统的数据。

3.2 数据清洗与整合方法

数据清洗是提升数据质量的关键步骤,旨在识别并去除数据集中的错误、重复及异常值,确保数据的一致性和准确性。这一过程包括对数据的全面检查,识别缺损数据、

错误数据以及噪声数据,随后运用填充技术、SQL语句及噪声平滑方法等手段进行处理,将“脏数据”转化为“干净数据”。整合方法则侧重于将来自不同来源、格式各异的数据集合并为一个统一的数据集。在整合前,需标准化数据格式,统一日期、单位及文本字段等,以提高合并效率。通过键值连接、时间戳合并等方式,将多个数据源无缝对接,同时解决数据冲突和重复问题。在这一过程中,需定义明确的优先级规则,以确保数据的一致性和准确性。通过数据清洗与整合,能够显著提升数据的可用性和可靠性,为后续的用户行为分析模型提供坚实的基础。这一环节不仅关乎数据的质量,更直接影响到最终分析结果的准确性和有效性^[4]。

3.3 数据质量评估标准

数据质量的评估主要从完整性、一致性、准确性和及时性四个方面进行深入考量。完整性是指数据信息是否存在缺失的情况,这包括整个数据记录或数据中某个字段信息的缺失。不完整的数据会显著降低其参考价值,是数据质量评估的基础。评估完整性时,可通过数据统计中的记录值和值域进行,例如对比日常访问量与异常访问量,或检查地域分布统计值是否全面。一致性则关注数据是否遵循统一规范,保持格式统一。这体现在数据记录的规范性和逻辑一致性上,如手机号码的位数、IP地址的格式,以及PV和UV间的逻辑关系等。一致性检验通常依据标准编码规则进行。准确性强调数据记录信息的正确性,包括检查乱码、异常数据等。这可通过统计最大值和最小值,或使用数据分析工具进行复杂统计分析来实现。及时性则指数据从产生到可供查看的时间间隔,即数据延时长。确保数据的及时性对于捕捉用户行为变化的最新趋势至关重要。

数据质量评估标准是多维度、综合性的考量,旨在为后续的用户行为分析奠定坚实基础。

4 基于聚类分析模型构建

4.1 用户行为特征提取

在构建基于大数据的通信用户行为分析模型时,用户行为特征提取是至关重要的一环。用户的在线行为数据包含了丰富的信息,可以用于提取用户的特征,从而深入了解用户的兴趣、行为习惯以及潜在需求。用户的在线行为多种多样,包括搜索、点击、购买等。针对每一类行为,可以通过提取不同的特征进行分析。例如,在用户搜索行为中,可以提取搜索的关键词、搜索词的搜索次数、搜索结果的点击率等特征,这些特征能够反映用户的兴趣偏好和需求。而在用户点击行为中,可以提取点击的广告类型、点击时间、所在的网页位置等特征,从而了解用户的广告喜好和点击行为规律。对于用户的购买行为,可以提取购买的商品种类、购买时间、购买金额等特征,以此来挖掘用户的购买习惯和消费特点。此外,为了更全面地理解用户行为,还需要考虑维度和指标的应用。维度描述了一个事物身上所具备的特征或属

性，而指标则是具体的数值。通过结合维度和指标，可以更深入地分析用户行为数据，为模型的构建提供有力的支持^[5]。

用户行为特征提取是构建基于大数据的通信用户行为分析模型的基础，通过深入分析用户的在线行为数据，可以提取出丰富的用户特征，为后续的模式构建和应用提供有力的数据支撑。

4.2 聚类算法的选择与应用

在聚类分析模型中，选择 K-means 算法作为主要的聚类方法。K-means 算法是一种基于划分的聚类算法，其核心思想是通过迭代优化，使得同一聚类内的数据点尽可能相似，而不同聚类间的数据点差异尽可能大。

K-means 算法的基本公式包括目标函数和更新公式。目标函数是求解各个聚类内数据点到聚类中心的距离平方和的最小值，即：

$$J = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$$

其中， k 是聚类数目， C_i 是第 i 个聚类， μ_i 是第 i 个聚类的中心， x 是聚类内的数据点。

更新公式则是在每次迭代中，根据当前聚类中心重新分配数据点，并更新聚类中心，即：

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$$

通过不断迭代，直到聚类中心不再发生变化或达到预设的迭代次数，最终得到稳定的聚类结果。K-means 算法因其简单高效、易于实现，在通信用户行为分析中具有广泛应用。

4.3 聚类结果与用户行为特征的关联分析

经过 K-means 聚类算法的处理，得到了通信用户的多个聚类群体。为了深入理解这些群体，将聚类结果与用户行为特征进行了关联分析。以下是用户行为特征与聚类群体的关联表格：

聚类群体	平均通话时长 (分钟)	平均上网流量 (MB)	短信发送频率 (条/周)
群体 1	12.5	850.3	7.2
群体 2	45.8	2100.7	2.1
群体 3	8.0	400.1	15.3

从表格中可以看出，不同聚类群体的用户行为特征存在明显差异。例如，群体 2 的平均通话时长和上网流量都较高，而短信发送频率较低；群体 3 则相反，短信发送频率较高，但通话时长和上网流量相对较低。这些差异为后续的用户画像构建和精准营销策略制定提供了有力依据。

5 聚类结果的应用价值探讨

5.1 用户画像构建

聚类结果为我们构建细致入微的用户画像提供了基础。以表格中的数据为例，我们可以清晰地描绘出不同群体的用户特征。群体 1 的用户通话时长适中，上网流量和短信发送

频率也处于平均水平，他们可能是普通的通信用户，对通信服务的需求较为均衡。群体 2 的用户则表现出高通话时长和高上网流量的特点，他们可能是商务人士或重度手机用户，对通信服务的质量和速度有较高要求。而群体 3 的用户短信发送频率极高，他们可能是社交活跃分子或年轻人群体，对短信服务有特别的需求。这些用户画像的构建，有助于我们更深入地了解用户需求，为后续的服务提供指导。

5.2 精准营销策略制定

基于聚类结果，我们可以制定更加精准的营销策略。对于群体 1 的普通用户，我们可以推出性价比高的套餐服务，满足他们日常通信的需求。对于群体 2 的商务人士或重度用户，我们可以提供高品质的通信服务，如 5G 网络、高速上网等，以吸引他们选择我们的产品。而对于群体 3 的社交活跃分子或年轻人，我们可以推出包含丰富短信服务的套餐，或者与他们喜爱的社交应用合作，提供定制化的服务。这样的精准营销策略，有助于提高市场营销的效果和用户满意度。

5.3 产品功能优化建议

聚类结果还为我们的产品功能优化提供了方向。例如，针对群体 2 用户对高速上网的需求，我们可以进一步优化网络架构，提高网络速度和稳定性。针对群体 3 用户对短信服务的需求，我们可以开发更加便捷、有趣的短信功能，如短信表情包、语音转文字等，以提升用户体验。同时，我们还可以根据不同群体的用户反馈，持续优化产品的其他功能，如通话质量、用户界面等，以满足用户多样化的需求。

聚类结果在用户画像构建、精准营销策略制定以及产品功能优化建议等方面都具有重要的应用价值。我们将继续深入挖掘聚类结果的潜力，为业务运营提供更有力的数据支持。

6 总结

本文提出的基于大数据的通信用户行为分析模型，通过聚类分析成功揭示了用户行为的多样性。聚类结果在用户画像构建、营销策略制定和产品功能优化等方面展现出重要价值，有助于运营商更深入地了解用户需求，提高市场营销效果和用户满意度。该模型为通信行业的智能化发展提供了有力支撑。

参考文献

- 谷红勋,杨珂.基于大数据的移动用户行为分析系统与应用案例[J].电信科学, 2016, 32(3):8.
- 吕洋.基于用户行为数据分析的移动互联网业务推荐模型[D].华中科技大学[2024-12-27].DOI:10.7666/d.d190567.
- 李嘉彬,施勇,薛质.基于大数据平台的用户行为分析研究[J].信息安全与通信保密, 2016(4):5.
- 孙加萌,宋文凯.基于大数据的企业内部用户实体行为分析研究[J].中国新通信, 2023, 25(2):53-55.
- 周志远.基于电力大数据的用户用电行为分析模型探究[J].电子器件与信息技术, 2023(11):126-129.