

文章编号: 1005-8451 (2018) 07-0054-04

铁路旅客用户画像系统设计与应用研究

张军锋

(中国铁道科学研究院集团有限公司 电子计算技术研究所, 北京 100081)

摘要: 结合铁路客票业务需求和系统现状, 利用大数据技术对铁路旅客用户画像系统进行了设计, 包括系统架构、标签体系及其实现。在此基础上以提高运营效率 and 提升服务质量为目标, 探索了画像系统在客运领域的应用, 提出了在旅客行为分析、精准营销、数据增值服务方面的应用方法和思路。

关键词: 用户画像; 精准营销; 大数据应用

中图分类号: U293.22 : TP39 **文献标识码:** A

Railway passenger user portrait system

ZHANG Junfeng

(Institute of Computing Technologies, China Academy of Railway Sciences Corporation Limited,
Beijing 100081, China)

Abstract: Combining with the service requirements in Railway Ticketing and Reserving System (TRS) and the system status, this paper applied the big data technology to design the railway passenger user portrait system including system architecture, label system and its implementation. Based on this, the paper explored the system application in passenger transport field, put forward the application methods and ideas in passenger behavior analysis, precision marketing and data value-added service.

Keywords: user portrait; precision marketing; big data application

国务院《关于促进大数据发展的行动纲要》强调以企业为主体、以市场为导向, 加大政策支持大数据产业发展力度, 深化大数据在各行业创新应用, 催生新业态、新模式。大数据技术和业态的成熟也为铁路部门实现数据驱动和业务创新提供了强有力的支撑。目前铁路部门已经形成了以 12306、95306 为代表的一系列互联网服务平台^[1], 在现有基础之上, 采用大数据技术, 开展用户画像、精准营销、延伸服务等“互联网+”产业升级已经具备条件, 也是产业发展的大势所趋。

用户画像系统是对现实生活中的用户行为进行数据建模, 以不同的数据维度对用户进行刻画。通过对用户的人口属性、行为偏好等主要信息进行建模分析, 从而抽象出能够让人理解的语义标签, 通过标签来形成一个用户的信息全貌, 为进一步分析和利用这些信息提供数据基础。目前在航空、电信、银行和互联网行业, 许多公司已经建立并应用了用

户画像系统, 在提升用户体验、精准营销、效益提升方面取得了很好的效果^[2-3]。

本文针对铁路行业大数据应用—旅客用户画像系统进行了探索: 对铁路用户的行为、交易等数据进行采集, 综合运用统计学模型、文本挖掘、机器学习等多种技术, 把海量数据转化成简洁、形象、人机可理解的画像标签, 通过与应用系统对接, 支撑客运产品优化、个性化服务、数据增值服务等业务的开展, 从而充分发挥数据的基础资源作用和创新引擎作用, 助力铁路部门提升客户服务能力和核心竞争力。

1 旅客用户画像系统架构设计

铁路旅客用户画像系统依托基础运行环境提供线性伸缩的计算存储资源, 采用基于 Hadoop 的大数据体系架构^[4-6], 通过采集、汇聚内部数据和外部数据, 为广告综合投放子系统、数据分析子系统、在线交易子系统以及客运相关服务提供实时的用户画像数据访问接口, 支撑精准广告投放、客户推荐服务、客运辅助决策支持等业务, 如图 1 所示。系统主要

收稿日期: 2018-05-10

基金项目: 中国铁路总公司科技研究开发计划项目 (2017X004-C, 2016X004-B)。

作者简介: 张军锋, 副研究员。

包括：

- (1) 实现系统大数据计算存储、数据汇聚、数据分析的基础运行环境；
- (2) 获取铁路客运业务系统数据、网站用户行为数据的采集系统；
- (3) 用于实现铁路用户画像标签定义、管理、计算、分析和呈现的业务平台；
- (4) 与外部系统交互的数据服务接口平台。

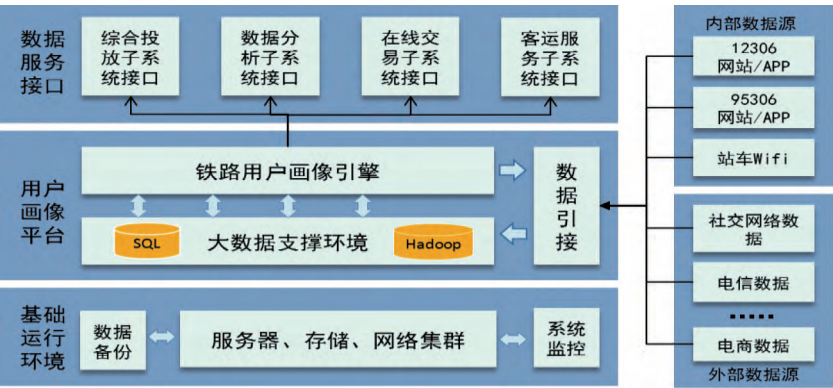


图1 用户画像系统总体架构

2 旅客用户画像系统实现

用户画像即用户信息标签化，就是从用户原始数据到用户标签的分析转换过程——通过对汇聚的海量用户数据的分析挖掘，形成每个用户的特征标签集合，并对外提供基于用户特征标签的数据服务的过程。它的核心工作是为用户打标签，打标签的重要目的之一是为了让人能够理解并且方便计算机处理。标签通常是人为规定的高度精炼的特征标识，既能够很方便地被人理解，也能够方便机器进行提取和聚合分析。

构建用户画像标签体系包含以下步骤。

2.1 数据源分析

构建用户画像是用户对用户信息的还原，涉及的数据源尽可能包括用户相关的所有数据。通过对铁路客票系统相关数据源的分析，根据标签的特性将数据分成静态数据和动态数据。静态数据包括人口属性、社会属性、账户属性等；动态数据包括乘车、购票、支付、查询、咨询行为等方面数据。

2.2 设计标签

标签和标签规则需要人为定义，通过对目标的确定和数据的分析，定义出对应的标签，最终为用户打上标签。根据用户画像标签实现方式和应用场景的不同，铁路旅客用户画像系统的标签可以分成事实标签、业务标签、模型标签等类型，如图2所示。

2.3 标签实现

标签实现是数据建模和流程化的过程，即根据定义好的标签，确定采用什么样的方式为用户打上标签的过程。铁路旅客用户画像系统在技术上使用了离线的MapReduce/Pig/Hive、准实时的spark/impala、数据挖掘的R/mahout/spark-mllib等数据处理技术，结果存储到Hbase和Elasticsearch快速检索引擎上，以供数据服务接口实现快速访问。铁路旅客用户画像系统建模的方式包括直接取值、统计分析、业务规则、预测模型等几种类型，流程示意图，如图3所示。

3 旅客用户画像系统应用

目前，铁路旅客用户画像系统设计并建成了包括事实标签、业务标签和模型标签在内的一整套标签体系，数量超过1500个，涵盖了旅客的基础属性（姓名、性别、年龄、常驻地址等）、出行行为（出行次数、出发城市、到达城市、席位等级等）、交易行为（线上线下选择、支付方式、退改签情况等）、出行规律、消费习惯、潜在消费等多个方面。

建立用户画像系统的目的是为了帮助业务部门了解自己的用户，便于在产品设计和旅客服务过程中



图2 旅客用户画像标签体系



图3 标签建立流程示意图

做到目标明确、高效和精准。结合目前铁路业务现状，旅客用户画像系统在下述几个方面可以进行业务对接和应用。

3.1 旅客行为分析

通过对旅客成分结构、产品选择行为、购票行为、旅行行为等方面进行分析，可以发现群体旅客特征、个体旅客特征与客运产品的关系。

3.1.1 旅客成分结构

通过分析得到铁路旅客的组成情况，以做好铁路产品设计和提供数据支撑。例如：

- (1) 性别比例；(2) 旅客区域（城市）比例；(3) 各年龄段比例；(4) 各票种的比例。

3.1.2 产品选择行为分析

旅客在进行产品选择时，会受多种因素的影响。通过下列分析，有助于分析判定旅客选择客运产品的影响因素，为产品设计提供依据。

- (1) 按旅客选择的列车类型分析；(2) 按旅客选择的铺别和席别分析；(3) 按旅客选择的票价区间分析；(4) 按旅客乘车里程分析；(5) 按旅客选择的列车开点分析；(6) 按旅客选择的列车到点分析；(7) 按旅客选择列车担当企业分析。

3.1.3 交易行为分析

通过交易记录的分析，发现旅客在购票交易过程中的行为特点，以针对性的做好售票组织和服务工作。

- (1) 购票渠道分析；(2) 互联网换票分析；(3) 电订取票分析；(4) 自助售票分析；(5) 支付方式分析；(6) 预售规律；(7) 通票、联程、往返票分析；(8) 退票时间规律；(9) 改签时间规律；(10) 改签分类分析。

3.1.4 旅客乘车行为分析

通过对旅客出行行为的分析，可以发现需要重点关注的旅客，从而进行产品推荐或者个性化服务，提升旅客出行体验。

- (1) 按旅客出行目的分析：通过设计合理的模型进行评判，将出行目的分为旅游、商务、公务、探亲、民工、学生、通勤职工等；(2) 按旅客出行频度分析：按出行次数对旅客进行分群分析，得到潜在的常旅客用户；(3) 按旅客出行范围分析：按行政区划进行出行范围分析，如省内客流、省间客流；(4) 按旅客出行行程分析：按旅客出行行程进行分析，由聚类模型分析得到旅客行程的合理划分。

3.2 精准营销

通过构建的铁路旅客用户画像，为旅客建立了人口属性、出行行为、交易行为等多维度的标签，通过对多维度标签的筛选、聚类，找出符合要求的特定人群，然后在推荐平台为人群制定对应的营销策略，并利用广告平台实现对用户群体的精准营销。

基于铁路旅客用户画像数据的推荐系统利用Elasticsearch强大的组合和聚合搜索能力，提供快速的标签筛选能力，在此基础上利用机器学习算法——基于用户的协同过滤算法，根据用户的偏好实现用户聚类，即将相似度较高的用户分为一个群体，在用户访问时进行信息的推送，从而实现产品的精准推荐。

用户定量标签的相似度计算方法^[7-8]：

$$sim(u_i, u_j) = \sum_k w_k sim(profile_k(u_i), profile_k(u_j))$$
 (1)

式(1)中， w_k 表示第 k 个定量标签的权重； $sim(profile_k(u_i), profile_k(u_j))$ ，表示用户 u_i 和用户 u_j 在第 k 个定量标签的相似度。

3.3 数据核验

央行发布的《非银行支付机构网络支付业务管理办法》于2016年7月1日起正式实施。其要求支付机构为客户开立支付账户的，应当对客户实行实名制管理。在推行和实施的过程中，第三方支付平台需要简化用户核验流程，在提高用户体验的情况下完成用户的核验，故需要借助外部系统提供的核验服务。目前铁路互联网售票系统注册用户突破2.5亿，乘车旅客信息超过5亿，旅客信息基本都是通过实名制核验的，所有的数据真实可靠，所以铁路互联网数据服务平台可以对第三方支付平台提供身份核验服务，助力第三方支付平台实名制的实施。

因此，数据核验是铁路部门实现数据共享、数据共创和数据增值的重要业务。系统基于旅客用户画像系统，为第三方系统提供不同查询条件的服务接口，包括姓名、证件号和手机号的比对。

3.4 互联网售票系统风险防控

铁路12306互联网售票系统为旅客提供便捷购票的同时，也成为部分机构和个人囤票、倒卖的途径。如何对互联网售票系统存在的恶意访问进行有效防控，构造公平、和谐的购票环境是永恒的主题^[9]。基于画像系统对用户行为进行深入分析挖掘之后，评估存在异常的用户行为，针对性地研究相关风险控制手段形成风险控制系统，实现应用防火墙（WAF）成为行之有效的方法。

风险防控系统覆盖了风险识别、风险控制、风险告警等环节，主要功能包括：

（1）用户风险等级管理。结合业务系统的用户特点，综合考虑和分析用户的各类特征因素并进行分类，通过用户评级列表可以展现基于风险评估的评级结果。（2）规则管理平台。规则引擎基于多维度的风险特征、灵活的风控模型，可以准确地分析识别恶意访问行为，经过规则判断的每一条数据，综合各个维度的分数，形成最终的评分，最终建议该条请求拒绝或者通过。（3）统计分析。按照风险性质和类型、业务类型、风险趋势、风险核查等不同角度进行统计，通过可视化辅助监控人员全面了解当前系统的运行状况，并将相关统计分析结果运用在优化风险规则上。（4）报警监控。对系统分析出

的异常行为，依据规则达到风险告警阈值时进行告警，以便管理人员及时人工介入处理。（5）风险控制。实现通用安全防御、基于IP的流量控制、IP黑白名单、统一资源定位符（URL）黑白名单、用户名黑白名单、用户/URL/IP高频控制、业务流程控制、监控、管理、多样化的安全管控措施等业务或管理功能。

4 结束语

通过运用大数据技术，设计了铁路行业的旅客用户画像系统，并基于该系统进行了业务应用的探索。目前，旅客行为分析、数据核验和互联网售票系统风险控制应用已经部署上线、进入试运行阶段，精准营销相关功能正在开发过程中，后续将在完善现有应用的基础上，积极探索新的业务模式和方法，通过不断的完善、优化、丰富，充分发挥铁路旅客用户画像系统在客运产品设计、旅客服务水平提升方面的支撑作用。

参考文献：

- [1] 朱建生，王明哲，杨立鹏，等. 12306互联网售票系统的架构优化及演进[J]. 铁路计算机应用，2015，24（11）：1-4.
- [2] 王冬羽. 基于移动互联网行为分析的用户画像系统设计[D]. 成都：成都理工大学，2017.
- [3] 李璋. 基于Hadoop的互联网数据营销系统的设计与实现[D]. 北京：中国科学院大学，2017.
- [4] 单杏花，王富章，朱建生，等. 铁路客运大数据平台架构及技术应用研究[J]. 铁路计算机应用，2016，25（9）：14-17.
- [5] 阎志远，翁涅元，戴琳琳，等. 铁路客运大数据平台的数据采集技术研究[J]. 铁路计算机应用，2016，25（9）：17-21.
- [6] L Deng, J Gao, C Vuppapapati. Building a Big Data Analytics Service Framework for Mobile Advertising and Marketing [C]// IEEE First International Conference on Big Data Computing Service & Applications, 2015: 256-266.
- [7] 阮利男. 大数据时代精准营销在京东的应用研究[D]. 成都：电子科技大学，2016.
- [8] 张慷. 手机用户画像在大数据平台的实现方案[J]. 信息通信，2014（2）：266-267.
- [9] 王雪峰，王明哲，朱建军，等. 如何防范针对12306网站的恶意抢票插件[J]. 铁路计算机应用，2015，24（3）：26-28.

责任编辑 徐侃春