

# 融合运营商网络行为数据的 权益营销应用

## Application of Equity Marketing Based on Network Behavior Data of Operators

郑正广<sup>1</sup>, 闫宇<sup>1</sup>, 蔡润昌<sup>1</sup>, 袁鹏<sup>2</sup>, 林立言<sup>1</sup>, 蔡惠坤<sup>1</sup> (1. 中移互联网有限公司, 广东广州 510640; 2. 中讯邮电咨询设计院有限公司广东分公司, 广东广州 510627)

Zheng Zhengguang<sup>1</sup>, Yan Yu<sup>1</sup>, Cai Runchang<sup>1</sup>, Yuan Peng<sup>2</sup>, Lin Liyan<sup>1</sup>, Cai Huikun<sup>1</sup> (1. China Mobile Internet Co., Ltd., Guangzhou 510640, China; 2. China Information Technology Designing & Consulting Institute Co., Ltd. Guangdong Branch, Guangzhou 510627, China)

### 摘要:

移动互联网时代下,如何挖掘海量用户网络行为数据,提升数据变现能力对于业务运营具有重要的意义。融合电信运营商计费订购数据、移动认证登录数据和上网流量DPI解析数据等数据源,按照统一的标准构建业务分类标签,并根据用户使用行为记录输出兴趣偏好标签,最后辅助视频类权益计费业务的营销推荐,实验结果表明该方法能够显著提高订购转化率。

### 关键词:

电信运营商;网络行为;兴趣偏好;权益营销  
doi:10.12045/j.issn.1007-3043.2022.07.016  
文章编号:1007-3043(2022)07-0083-05  
中图分类号:TN929.5  
文献标识码:A  
开放科学(资源服务)标识码(OSID):



### Abstract:

In the era of mobile Internet, how to mine massive user-network behavior data and improve data liquidity is of great significance for telecom operators. It integrates data sources such as telecom operator billing and ordering data, mobile authentication login data and Internet traffic DPI analysis data, and constructs service classification labels according to unified standards. According to the user behavior record, the interest preference tag is output, and finally it is used to assist the marketing recommendation of video equity billing service. The experimental results show that this method can significantly improve the ordering conversion rate.

### Keywords:

Telecom operator; Network behavior; Interest preference; Equity marketing

引用格式:郑正广,闫宇,蔡润昌,等.融合运营商网络行为数据的权益营销应用[J].邮电设计技术,2022(7):83-87.

## 0 引言

近十年移动互联网高速发展,智能终端已趋于饱和,增量市场已逐渐转变为存量市场博弈。在“提速降费”政策力度不减以及“携号转网”全面开启的新形势下,以语音和流量为代表的传统通信业务收入增长乏力,甚至出现负增长。因此如何高效融合现有数据资产,驱动增值业务发展成为电信运营商需要迫切解决的问题。

相较于众多APP应用提供商,电信运营商不仅拥有用户的基础信息,如性别、年龄、归属地、终端型号等静态属性标签,同时能够获取用户上传行为数据,从而解析出用户行为数据,如兴趣偏好等动态属性标签<sup>[1]</sup>。借助大数据用户级标签系统,对用户进行精准的个性化推荐服务,不仅可以提高用户感知体验,形成差异化竞争力,还可以为运营商实现流量变现,对于电信运营商具有重大意义和价值。

本文选取某电信运营商网络行为数据进行挖掘,侧重分析网络DPI(Deep Packet Inspection)数据、计费订购数据和移动认证数据。本文对不同来源行为数

收稿日期:2022-05-25

据进行整合,构建统一的规范化标签体系,进而对用户兴趣偏好标签进行刻画,然后进行个性化权益营销推荐服务,为运营商增值业务收入拓展提供参考。

## 1 电信网络行为数据概览

电信运营商拥有云、网、数三位一体的经营优势,掌握着海量的业务运营数据和网络行为数据,资源禀赋明显。这些数据包含了用户开卡资料、从智能终端到接入网、传输网到核心网等各环节数据,本文侧重分析网络行为数据。

随着互联网技术不断发展和完善,以微信、支付宝等为代表的应用软件正深刻地改变着人们的生活,同时也催生出一大批新的应用和业态。据工信部信息通信发展司统计,截至2018年底,我国移动APP应用规模排名全球第一,接近499万。用户对于APP的依赖程度越来越高,众多的行为习惯和兴趣正通过各种终端设备在网络数据中体现。大数据技术的发展也在改变着运营思路,从单纯的数据采集逐渐过渡到高质量运营和业务创新。

手机话费除了可以用于语音、流量等套餐消费外,还可以用于订购APP会员权益和游戏道具等增值业务,形成计费数据。部分APP应用接入了移动认证服务,可以获取到用户在什么时候通过手机号一键登录了什么APP。当用户中断操作APP,再次打开APP时一般不会触发一键登录,因此单纯依靠统一认证并不能完全掌握用户使用行为,且统一认证覆盖的APP数量相对有限。由于大多数网站和APP在网络传输过程中并非完全加密,可以对上网流量IP数据包进行DPI深度检测,根据网址、端口等特征识别出对应的业务和APP,从而获取用户浏览记录<sup>[2]</sup>。本文根据用户网络行为数据,可以刻画出用户兴趣偏好标签,从而为用户提供个性化营销推荐服务,具体如图1所示。

## 2 用户兴趣偏好标签构建

电信运营商通信网的用户网络行为面向的是所有APP,而DPI数据、计费数据和移动认证数据管理系统各不相同,对于APP业务的分类也有差异。此外各家软件应用商店对APP的分类和标签同样有着自己的设定规则,如表1所示,且覆盖度也不同。因此,电信运营商需要对采集的APP数据进行业务融合,并形成一套统一的规范化分类和标签体系,从而更加高效准确地依据APP使用行为来描绘用户兴趣偏好,为后

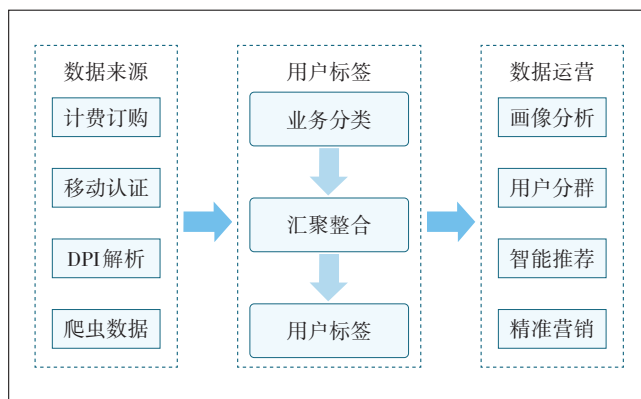


图1 精准营销构建框架

期精准推荐服务奠定良好的数据基础。

表1 “赶集网”在不同渠道上的分类标签

渠道	分类	标签
小米应用商店	居家生活	同城、兼职、求职招聘、租房、求职、二手
豌豆荚	生活休闲	用完就删、房产家居、求职
应用宝	生活	生活服务、招聘求职、全职招聘、综合服务
MM应用商场	生活助手	二手买卖、同城活动、求职、生活服务
华为应用商场	便捷生活	买房、本地生活、便民、二手车、生活服务、租房
易观	生活	房产家居、生活休闲、生活、便捷生活、居家生活、找工作、找房子、二手车、租房、求职、招聘

### 2.1 应用业务分类

针对各渠道来源数据不一致问题,本文首先对不同渠道来源的APP进行爬虫采集,获取应用名称、分类、标签和描述等信息,然后以某个渠道标准作为参考,进行APP关联匹配。对于无法匹配的应用,则采用机器学习的方法对分类进行预测,并提取关键词作为应用标签。

#### 2.1.1 应用采集

为了建立完备的APP基础信息库,本文针对主流应用商店设计了多线程数据采集系统,流程框架如图2所示。

由于采集数量较大,该采集系统需要在内存中使用布隆过滤器以高效识别重复网址,避免对数据进行重复采集。此外,配置IP代理池和User-Agent池可以突破基本的反爬策略,以确保系统稳定运转。

#### 2.1.2 应用分类

本文将采集得到的应用数据进行关联匹配,提取已识别的应用名称和描述作为训练数据,对无法匹配的应用进行机器学习分类预测。传统的文本分类方法主要有基于规则的关联模型和基于向量空间模型2

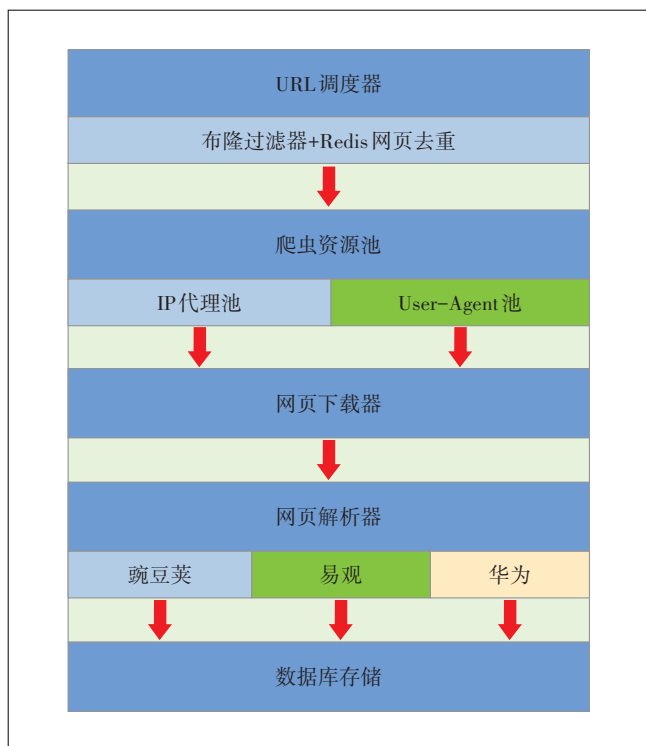


图2 爬虫数据采集流程框架

种<sup>[3]</sup>。基于规则的文本分类需要依据专家知识人工标注关键词词典,但随着应用规模的扩充和新业态的不断涌现,需要耗费大量的人力对词典频繁更新,且效率低下,难以适应运营发展。向量空间模型则把文档的语义看成诸多词语的表达,通过深度学习的方法进行向量化表示,且在长文本分类中取得了较好的效果<sup>[4]</sup>。

### 2.1.2.1 Word2Vec 模型

Word2Vec 基于神经网络对大量文本语料的词汇和上下文关系进行训练,将其中的词汇映射到一个较低维度的向量空间。常用的训练模型有 CBOW 和 Skip-Gram 2 种,在 CBOW 模型中,需依据上下文信息  $S_{w_t} = (w_{t-k}, \dots, w_{t-1}, w_{t+1}, \dots, w_{t+k})$  对当前词汇  $w_t$  进行预测,包含输入层、隐藏层和输出层,其目标优化函数如下。

$$L_{CBOW} = \sum_{w_i \in C} \log p(w_i | S_{w_t}) \quad (1)$$

其中  $C$  为语料库中所有词汇的集合,  $k$  为上下文信息窗口长度。

与 CBOW 相反, Skip-Gram 模型则是依据当前词汇对上下文信息进行预测,优化目标函数如下。

$$L_{Skip-Gram} = \sum_{w_i \in C} \sum_{k_j \in C, k_j \neq 0} \log p(w_{t+j} | w_t) \quad (2)$$

在 Word2Vec 模型输入层,每个单词均进行 one-hot 编码,而隐藏层神经元的数量通常设置为 100~300,训练后得到隐藏层的训练参数矩阵,以此构建词向量。

### 2.1.2.2 TF-IDF 算法

TF-IDF 是一种在文本信息检索领域常用的加权技术,其结合了文档中单词出现的频次 TF 和逆文档频率 IDF 综合计算,公式如下。

$$TF_{ij} = \frac{N_{ij}}{\sum_m N_{mj}} \quad (3)$$

$$IDF_i = \log \frac{N_d}{D_i + 1} \quad (4)$$

$$TF-IDF_{ij} = TF_{ij} \times IDF_i \quad (5)$$

其中  $TF_{ij}$  表示单词  $w_i$  在文档  $j$  中的出现频率,  $N_{ij}$  为单词  $w_i$  在文档  $j$  中的出现次数,  $N_d$  表示语料库中总文档数目,  $D_i$  为包含单词  $w_i$  的文档总数,分母加 1 是为了防止测试集中出现新词,导致逆文档频率无法计算。

在 TF-IDF 算法中,单词在文档中出现的频率越高,说明其权重越大,同时,若该单词出现的文档数目越多,则权重也将随之降低。比如,中文汉字“的”尽管词频较高,但由于逆文档频率很低,其综合权重反而较小。

### 2.1.2.3 神经网络分类模型

本文通过对短文本进行处理得到 APP 应用的向量化表达,相比传统基于规则的关联方法,采用神经网络模型进行训练预测,具备较高的可扩展性和推广性。

在文本分类任务中,神经网络模型可视为输入层、隐藏层和输出层的组合,相邻层的神经元进行连接,且具有对应的权重<sup>[5]</sup>。输入神经元的数量等于文本特征数量,而输出层神经元数量与应用分类数量相等。在初始训练阶段,可通过正向传播算法和反向修正算法进行多次迭代,以获取神经元之间的连接权重。

本文依据 APP 应用描述简介等短文本数据,首先进行分词操作等预处理,然后借助 Word2Vec 训练词向量,通过 TF-IDF 算法加权得到短文本向量化表达,并作为神经网络分类器的输入信号,如图 3 所示。分类完成后,针对单个应用,本文汇聚主流应用商店的标签,并选择 TF-IDF 权重较高的进行表征。



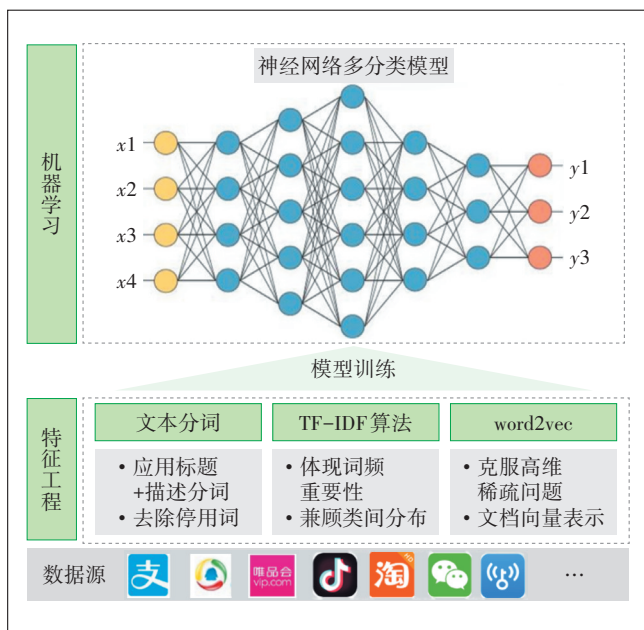


图3 APP应用分类框架

#### 2.1.2.4 分类效果评估

本文的试验选择影音、学习、旅游等11个类别共6万个APP数据进行训练预测。文本分词后,设定上下文窗口长度为10,词向量维度为100,采用CBOW模型进行训练得到词向量库。数据按照80%训练、20%测试的原则进行分割。

本文通过对测试集进行APP分类预测可得不同类别的准确率,结果如图4所示,总体准确率为76.3%,其中旅游类别的准确率最高,达到84.7%。

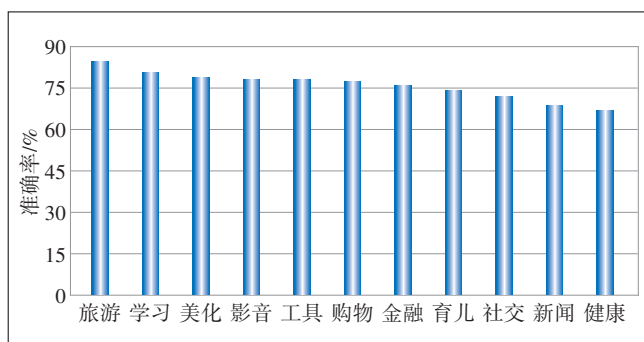


图4 应用分类准确率测试结果

#### 2.2 用户兴趣偏好计算

当用户对某类应用持续产生使用行为,如登录记录、浏览记录和订购记录等,使用行为次数越多,则说明该用户对这类APP的兴趣热度越高,且该热度一般只与近期内行为相关性较高。根据艾宾浩斯遗忘曲

线可知,用户使用APP的时间与兴趣偏好的相关性为<sup>[2]</sup>:

$$\rho(t) = 1 - \alpha(T - t)^\beta \quad (6)$$

其中 $t$ 为使用时间, $T(T \geq t)$ 为当前时间, $\alpha$ 和 $\beta$ 为兴趣热度衰减参数。

由于不同的APP所体现的可信度不同,不同网络行为其权重亦不同,因此在对用户兴趣偏好进行定量化表达时应尽量体现如下原则。

a) 用户在同一APP上的使用行为次数越多,代表兴趣越浓厚。

b) 用户行为记录发生时间(以天为颗粒度)越近,其兴趣偏好分值越高。

c) 同一APP内部,经济或时间成本越高的行为,所代表的分值越高,如订购行为权重大于登录,而登录则大于浏览。

d) 为体现不同APP之间的差异,可对APP赋予不同的权重。

用户兴趣偏好计算公式如下

$$s(u,p,t) = \sum_v \sum_t \rho(t) \times w(p) \times \gamma(v,p) \times f(u,v,p,t) \quad (7)$$

其中 $u,v,p,t$ 分别表示用户、行为(浏览、登录、订购)、APP应用和时间, $\rho(t)$ 表示时间兴趣衰减程度, $w(p)$ 表示APP应用 $p$ 的权重,可根据运营需要灵活设置, $\gamma(v,p)$ 表示应用 $p$ 中行为 $v$ 的权重, $f(u,v,p,t)$ 表示时间 $t$ 内用户 $u$ 在应用 $p$ 中产生行为 $v$ 的频次。

同一APP内不同网络行为的权重可借鉴文本IDF思想进行计算,例如某应用当月产生行为记录的总人数为1万人,而涉及订购行为的有100人,登录行为有8000人,那么订购行为的权重则为 $\log(10000/100)=4.61$ ,同理登录行为权重为0.22。

将相同类别的所有APP兴趣偏好值进行累加,即可得到用户所在类别的兴趣偏好标签值,表2为某运营商某用户排名前10的兴趣偏好标签。

### 3 权益计费业务营销推荐

为了持续改善计费业务收入结构,促进包月业务快速发展,选取爱奇艺25元包月视频会员权益等优质业务进行精准营销。依托大数据技术,分析用户网络行为数据,挖掘视频兴趣偏好较高且历史订购过爱奇艺的用户进行营销推送。为了验证本文兴趣偏好标签的有效性,同时选择通信消费高ARPU高增值费业务人群,以及历史订购过且视频记录频次较高的用户

表2 某用户Top 10兴趣标签偏好(采用10分制)

兴趣标签	偏好值	兴趣标签	偏好值
视频	9.95	微博	3.24
游戏	5.01	微店	2.88
头条	4.72	漫画	2.84
Wi-Fi	3.77	K歌	2.75
直播	3.31	生活	2.72

人群作为测试组,同时剔除近1个月有订购爱奇艺会员的用户,每组包含约30万用户。在相同时间进行同样的广告投放,并在3天后评估订购转化率效果,结果如图5所示。

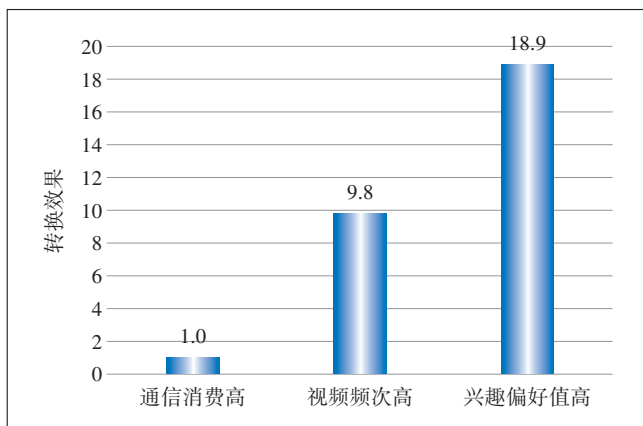


图5 不同投放人群归一化订购转化率提升效果对比

从图5可以看出,以通信消费高人群的订购转化率作为基准值,视频频次较高的测试组转化率提升了9.8倍,而本文提出的兴趣标签偏好值较高的用户则效果显著,提升到18.9倍,充分说明该标签体系的有效性。

在广告投放过程中所采用的文案并没有体现用户之间的差异性,即所有用户看到的订购页海报内容均一致,难以满足用户个性化需求。后续笔者将进一步细化用户内容偏好标签,并进行客户分群,同时制定不同的广告文案,以便进行更为精准的营销推送,拉动收入增长。

#### 4 结束语

本文通过对某电信运营商的权益计费订购数据、移动认证登录数据和上网流量DPI解析数据等数据源进行融合,按照统一的标准规范建立业务分类标签体系,并量化表达用户级兴趣偏好,辅助权益计费业务精准营销推荐。该方法不但可以提升数据运营能

力,还可以实现运营商流量的有效变现,具有较大的意义和价值。

#### 参考文献:

- [1] 宋春涛,张帆,王勇,等. 电信运营商的数据资产综述:数据、内联及外延[J]. 邮电设计技术,2019(9):20-24.
- [2] 金石. 基于运营商管道大数据的智能电商推荐系统[D]. 南京:南京邮电大学,2018.
- [3] 牛雪莹,赵恩莹. 基于Word2Vec的微博文本分类研究[J]. 计算机系统应用,2019,28(8):256-261.
- [4] 王根生,黄学坚. 基于Word2vec和改进型TF-IDF的卷积神经网络文本分类模型[J]. 小型微型计算机系统,2019,40(5):1120-1126.
- [5] 冯兴杰,张志伟,史金钊. 基于卷积神经网络和注意力模型的文本情感分析[J]. 计算机应用研究,2018,35(5):1434-1436.
- [6] 缪超,张辉,仇琦. 基于大数据的电信运营商业务模式研究[J]. 电脑迷,2017(12):75-75.
- [7] 杜娟. 电信运营商基于大数据的业务模式分析[J]. 移动通信,2015(13):77-81.
- [8] 李嫚,姚文胜,龚至晖. 运营商基于大数据的BI架构重构及数据应用研究[J]. 电信科学,2013(S1):5-5.
- [9] 左超,耿庆鹏,刘旭峰. 基于大数据的电信业务发展策略研究[J]. 邮电设计技术,2013(10):1-4.
- [10] 陈志恒,张瑜. 运营商大数据变现创新模式研究[J]. 现代电信科技,2016(4):1-4.
- [11] 黄仁,张卫. 基于word2vec的互联网商品评论情感倾向研究[J]. 计算机科学,2016,43(z1):387-389.
- [12] 苏秀芝,左国才,张珏. 基于深度学习框架的短文本情感分析方法研究[J]. 数字技术与应用,2019,37(2):80,82.
- [13] 张苗. 基于深度学习的中文评论文本情感分析研究[D]. 武汉:华中科技大学.
- [14] 杨政,易绵竹. 基于词向量的中文新词情感倾向性分析[J]. 计算机系统应用,2019,28(10):245-250.
- [15] 相若晨,孙美凤. 基于词向量与句法树的中文句子情感分析[J]. 计算机与现代化,2016(8):27-31.
- [16] 杨阳,刘龙飞,魏现辉,等. 基于词向量的情感新词发现方法[J]. 山东大学学报(理学版),2014(11):51-58.
- [17] 周萌. 基于LSTM和Word2Vec的情感分析在商品评论中的应用[J]. 统计与管理,2019(12):81-84.
- [18] 王勤勤,张玉红,李培培,等. 基于word2vec的跨领域情感分类方法[J]. 计算机应用研究,2018,35(10):2924-2927.

#### 作者简介:

郑正广,工程师,硕士,主要从事数据分析挖掘、机器学习等工作;闫宇,工程师,学士,主要从事智能推荐、机器学习等工作;蔡润昌,工程师,学士,主要从事应用分发、数据分析等工作;袁鹏,高级工程师,学士,主要从事网络规划与设计等工作;林立言,经济师,硕士,主要从事数据分析、经营分析等工作;蔡惠坤,中级经济师,学士,主要从事通信与互联网产品运营工作。