

# 基于大数据的工商协同营销模式研究

冯波兴<sup>1,2</sup>,刘阳<sup>1</sup>,赵涛<sup>1</sup>,郭梁<sup>1</sup>,蒲雪松<sup>1</sup>,周家贤<sup>1</sup>

(1. 云南中烟有限责任公司营销中心,2. 山东省泰安市烟草专卖局(公司),650000)

**摘要:**随着信息化的发展,以及《国务院关于实施健康中国行动的意见》等控烟政策的发布,国家层面对烟草销售、购买的政策趋向逐步收紧,传统的烟草行业面临严峻挑战.从2015年至今,新零售发展迅猛,与此同时也产生了大量数据,烟草行业与大数据的结合势在必行.烟草行业信息化与行业中一致推行的工商营销协同已成为当前重要的课题.通过探索目前烟草行业的理念以及工商协同营销的案例,结合大数据以及烟草行业自身特征,对烟草行业的工商协同模式进行研究.新模式对于烟草行业的信息化改革发展具有重要指导意义.

**关键词:**工商深度协同;标签化;大数据;画像

**中图分类号:**TP-9 **文献标志码:**A **文章编号:**1672-8513(2024)02-0258-08

当前烟草行业工商协同的营销模式,一般是烟草工业公司与商业公司,或者商业与零售户之间的相互合作,很难覆盖到消费者环节.电子商务的高速发展,大数据、物联网等信息技术逐步实现,出现了新零售等新的销售模式,根据数据预测,到2022年中国新零售的市场规模将达到1.8万亿<sup>[1]</sup>.这使得协同营销中的消费者环节变得愈发重要.2021年在卫健委发布的调查报告中,中国的烟民数量已经超过3亿.大量中国烟民产生的大量结构化、非结构化和半结构化数据,而这些数据的采集、处理、存储常规的软件工具已无法高效处理,只能靠大数据技术提取、处理、分析这些大量数据集.烟草工商协同由于不存在专业的信息处理工具,并未将烟草工业、烟草商业、零售户、消费者串联起来.烟草工业掌握着烟草的基本信息,而烟草商业掌握着消费者的部分基础信息以及消费者的购烟行为信息.但是工业与商业数据不互通,工商协同的程度不够深,没有实现工商企业之间的深度协同.研究如何实现协同营销重要的一环是理解消费者的需求,如何把握消费者真实需求,是需要解决的一个问题.

## 1 基于大数据的工商深度协同精准营销的意义

### 1.1 以“工、商、零、消”为一体的工商协同营销体系

随着5G时代的到来,新零售模式自2015年发展至今,已与多个行业相结合发展出自己的特色,逐渐成熟,其中最典型的就是生鲜电商平台,已经落地多年.随着这种销售模式的快速发展,该模式产生的大量数据也可以作为协同营销的重要资源利用起来.大数据技术作为处理这些数据的有效手段,利用大数据技术将视线聚集在消费者、消费者市场身上,建立更加符合现代消费趋势的以消费者为中心营销体系也具有可行性.

整个流程中,从种植环节到设计、采购、生产、营销等环节的信息利用也不容忽视,将得到的信息经过处理以后,再作为各个环节的决策支撑,可以科学高效的发展产业.基于大数据的工商深度协同营销,将整个流程各个环节的信息作为基础,以大数据相关的技术作为支撑,构建工商深度协同营销体系,在产业的各个部分发挥重要作用,包括原料种植数据支撑、烟品设计支撑、商品配送支撑等等,都是体系的重要受益环节.

### 1.2 建立消费者数据采集的新模式

新零售的出现,使得工商企业能从新的消费者信息来源渠道获取消费者的基础信息以及行为信息.移动

收稿日期:2022-10-31.

基金项目:云南中烟工业有限责任公司科技计划项目(任务书编号2022XX01).

作者简介:冯波兴(1979-),经济师,工商管理硕士.主要从事市场营销和互联网研究.

通信作者:郭梁(1982-),学士.主要从事市场营销数字化研究.



图1 大数据协同营销体系

化改变了零售商生存环境,客户关系从“失联”转变到“连接”,这是移动化时代零售商生存与发展的基本法则<sup>[2]</sup>。建立新的消费者数据采集模式必不可少,只有建立在线的平台,才能量化消费者信息,这是数字化的前提。再通过大数据,数据清洗等技术全方位采集、处理消费者信息。在在线的前提条件下,能更方便实现整体信息共享共用。将产品的生成、运输、销售,以及消费者的基础数据、行为数据共享共用。通过在线模式的数据采集,为多维主题的分析应用提供基础数据支撑。

### 1.3 建立面向消费者的多维分析模型

包含“消费者多渠道信息采集、消费者画像、品牌消费者画像、消费者市场画像、消费者标签化管理、零售消费者画像”多个角度,并研究以消费者需求为导向的工业企业生产需求模型、以消费者需求为导向的商业企业的货源投放模式、以消费需求为导向的面向零售户的经营指导,研究新零售模式下的品牌培育营销新方案,研究基于多异构大数据体系下的数智中台技术实现。

## 2 工商协同营销模式建设原则

构建的新的面向消费者的工商协同模式,也需要协同的基本原则以及行业规范:(1)方向正确性. 必须严格执行烟草行业相关政策,符合国家对烟草行业大方向的规划,协同营销工商企业之间注重沟通协调,明确各自责任。(2)整体性. 对于数据有必要公开透明,应将工商企业作为整体,将获得的消费者信息运用在整个营销流程中,在生成、工艺、供应、实施等方面实现信息共享,将采集到的信息作用于整体。(3)创新性. 与现有市场相结合,与新技术相结合,利用大数据、实时计算、用户画像、新型支付等新技术,更加精准高效的把握消费者真实需求,实现精准营销供应链优化的效果。(4)以消费者为中心. 营销的重点应该放在消费者身上,市场才是导向. 鼓励适度竞争,利用用户行为信息,产生市场画像,结合市场画像达到对于资源的更高效配置<sup>[3]</sup>。

## 3 工商深度融合面向消费者精准营销新模式

为了更加高效推动新零售相关的工商协同体系的建立,泰安烟草与云南中烟,共同开展了新零售模式下工商融合面向消费者营销模式研究工作,通过对新零售模式的特征分析,通过对比其他行业运用的研究,结合烟草行业自身的特点,使用大数据等相关技术,对消费者数据进行分析,从而构建新的工商协同体系。

### 3.1 总体框架

为了构建新的营销模式,模式中的主体是商业企业,同时借助大数据等技术来构建数据采集入口,来采集消费者的消费偏好信息,来形成一个标准的数据集,对卷烟消费者数据进行数据加工、清洗、建模、标签化,利用标签建立消费者、市场、品牌画像. 工商企业再参考画像进行营销、供货管理、品牌培育等方面的决策. 通过对品牌消费者进行画像分析,为商业企业货源精准投放提供数据支撑. 为工业企业品牌生产,提供辅助决

策数据.

### 3.2 消费者数据采集入口

烟草行业的现代化建设有部分成果,已经有了较为立体的数据采集体系.包括商盟系统客户管理系统、进销存管理系统等都可以产生所需数据<sup>[4]</sup>.但是对于消费者的信息来源.目前是缺少的.包括消费者的人口属性、消费水平、消费习惯、购买关联等信息是严重缺乏的.但是新零售模式下,烟草企业建立工、商、零、消共同参与的平 台,解决了消费者信息缺少的问题,使得消费者的信息也能够方便的采集到.通过以上信息的采集方式,形成消费者基本信息聚合,为后续消费者数据的加工、清洗、建模和应用提供基础数据支撑.

### 3.3 标签化管理

用户画像(persona)由学者 D. Travis<sup>[5]</sup>提出,是指基本性、移情性、真实性、独特性、目标性、数量和应用性.而生成用户画像一般需要按照以下步骤实现:(1)数据采集.(2)数据挖掘及过滤.(3)标签提取<sup>[6]</sup>.而标签提取是其中最具主观性的一步.根据不同的目的,会生成不同的标签,最后产生侧重点不同的用户画像.

标签的提取以及重组是生成用户画像的最后一个环节.对于标签的提取与重组方面,常用标签评分、TF-IDF 算法、关联规则、社会网络分析等来建立兴趣模型,使用数理统计,数据挖掘等方法来分析用户属性.根据新零售模式以及烟草行业的特点,我们再使用“专家意见法”完成.

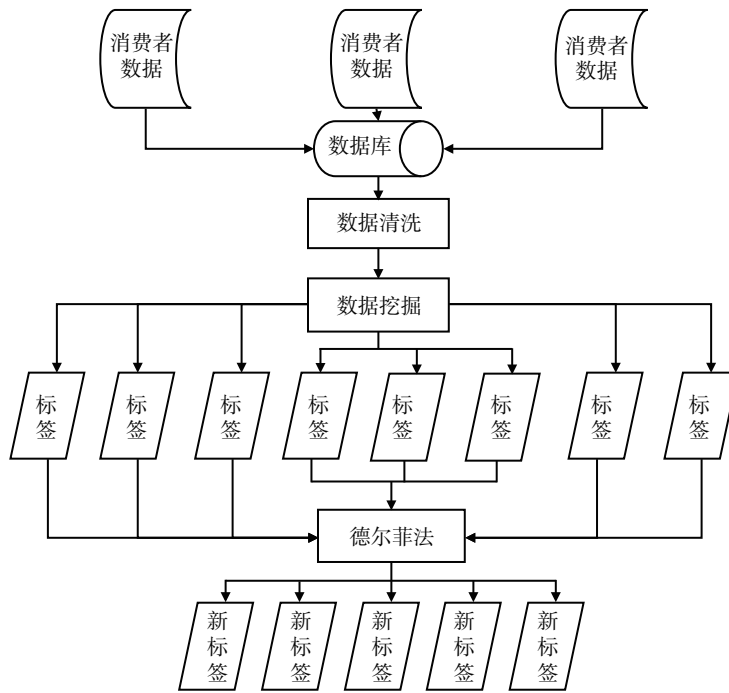


图2 标签生成

首先我们收集到以来源以平台为主的消费者数据,用户画像需要大量的数据才能得以实现,而新零售背景下提供的数据恰好能满足用户画像构建所需.将收集到的信息进行数据清洗,再通过聚类模型从中提取用户标签,比如长时间持续购买,用户会产生“老烟民”的标签.但是此时生成的标签针对性不够,生成的用户画像过于杂乱,还需要使用德尔菲法进行处理.德尔菲法,即“专家意见法”在聚类算法模型获得了自动生成标签后,将其中的高频便签找出来,结合具体的使用场景,通过专家问询,构建出让专家意见一致新标签.最后得到的新标签才会用于构建用户画像.

### 3.4 消费者画像

要建立立体的消费者画像一般会分别抽取消费者的动态标签和静态标签结合后再生成消费者画像标签.用户部分标签如下图.

对于不同的使用目的,如进行准营销和关联规则挖掘,就会使用不同的标签建立用户画像.就以常见的用户价值体现场景为例.

对于我们获取到消费者属性,一般会进行属性组合和规则定义.再针对特定的场景对消费者进行聚类.

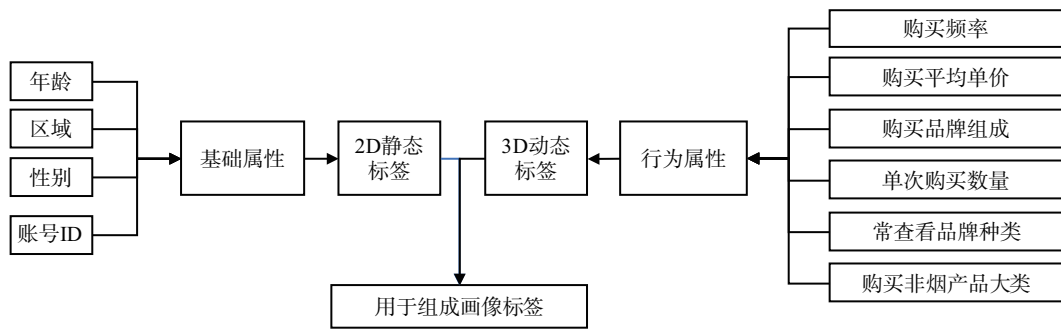


图 3 消费者属性

聚类算法多种多样, 根据不同的场合会使用不同的聚类算法.  $K$ -Means 算法最早是由 Mac<sup>[7]</sup> 于 1967 年提出的. 该算法由于原理简单、效率高被广泛应用于各种场景.

假定  $D = \{x_1, x_2, x_3 \dots x_n\}$  是拥有  $n$  个数据的集合, 而每个数据又拥有  $m$  条属性. 此时的  $m$  则根据具体场景下, 影响较大消费者的属性条数来决定.  $K$ -Means 以欧氏距离来判定集合中数据的相似性, 首先会将这些数据划分为不同的  $k$  个的类簇  $C_1, C_2 \dots C_k, 1 \leq k \leq n$ , 然后再根据欧式距离这个指标作为依据, 将  $D$  中的数据划分到  $k$  个类簇中. 具体的操作过程如下:

- (1) 从上述的数据集合  $D$  中, 随机去除  $k$  个数据作为初始的类簇中心.
- (2) 根据欧式距离的公式(1), 来计算每个类簇中数据到类簇中心的距离.

$$(3) d(x_i, C_j) = \sqrt{\sum_{t=1}^m (x_{it} - c_{jt})^2}$$

(4) 其中  $x_i$  表示第  $i$  个用户的数据,  $C_j$  表示第  $j$  个聚类中心.  $x_{it}$  表示第  $i$  个对象的第  $t$  个属性,  $c_{jt}$  表示第  $j$  个类簇中心的第  $t$  个属性.

(5) 分别计算每个数据到每个类簇中心的距离, 将数据划分到距离最短的类簇中. 得到第一次划分好的  $k$  个类簇  $\{s_1, s_2 \dots s_k\}$ . 此时可以根据类簇内现存的数据计算新的类簇中心.

$$(6) C_l = \frac{\sum_{x_i \in s_l} x_i}{|s_l|}$$

(7) 其中  $C_l$  表示第  $l$  个类簇中心,  $S_l$  则表示的是  $l$  对应的类簇中心的数据数量.

(8) 重复过程(2)(3)直至到类簇中心固定下来, 过着达到指定的重复迭代次数为<sup>[8]</sup>.

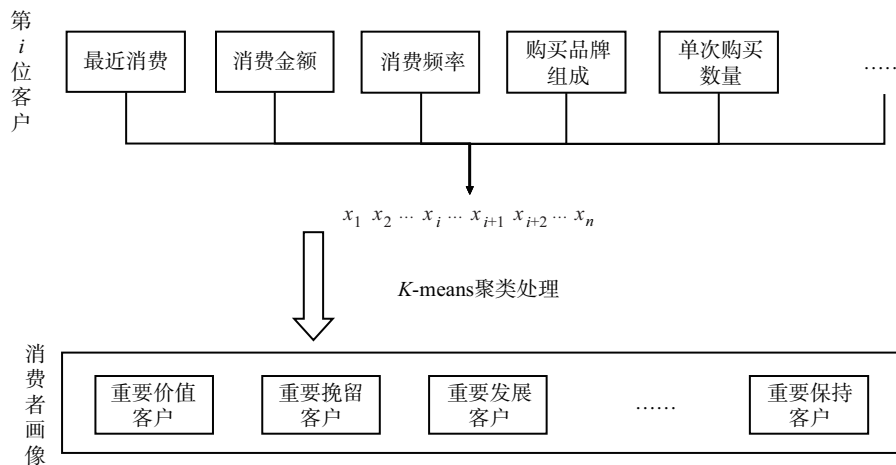


图 4 消费者画像

商业模式中有常见的一种情况是少数的客户, 贡献大多数的利润. 在这种情况下就会使用到一个经典的划分体系: RFM(recency frequency magnitude)模型, 是行数据挖掘重要途径<sup>[9]</sup>. 因此我们也参考 RFM, 将最近消费, 消费频率, 消费金额将这 3 个消费者属性加入, 同时还将用户购买品牌一致性, 和用户单次购买数量等

加入组成消费者的数据当中. 每个用户信息就是一个  $m$  维的向量, 将收集都的用户信息变成数据以后. 在使用  $K$ -Means 聚类算法, 将代表用户的向量分成以不同的类簇. 这些类簇就是不同的用户画像: 如重要价值客户、重要挽留客户、重要发展客户、重要保持客户等等. 以此确保不会流失重要客户.

### 3.5 市场画像

对于市场来说, 分析市场一般是分析市场的需求和供应两个方面, 对于烟草行业, 最直观的表现就是销售数据和生产数据. 通过收集某一市场上的某一段时间的销售与供应数据, 以及市场其他方面如市场内品牌、市场本身经济发展程度等为市场构建画像. 通过画像和标签, 能预测市场的发展趋势, 及时把握消费人群的定位, 同时对市场格局有明确的理解, 便于销售策略的制定.

市场画像的部分标签如图所示, 我们根据烟草市场的部分特点从两个大的维度建立标签.

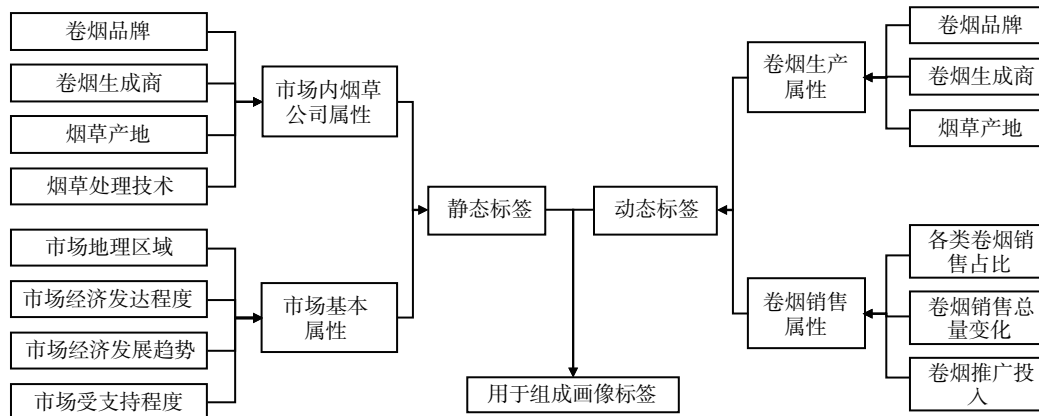


图 5 市场属性

我们通过收集到的信息, 将根据不同区域的划分, 对应一个个大小不一的市场. 在不同场景的应用中, 选取各种合适的市场属性, 将其变为市场所对应的数据, 然后再运用合适的聚类方法, 建立市场化下. 其中 Meanshift 聚类也是适合的聚类方法之一.

Meanshift 是一种基于滑动窗口的算法, 他的作用是可以找到数据密集的区域. Meanshift 算法不需要自己去确定要分的类簇数量. 同时它是基于质心的算法, 受到均值的影响相对于  $K$ -Means 算法要小. 我们同样假定  $D = \{x_1, x_2, x_3 \dots x_n\}$  是拥有  $n$  个数据的集合, 而每个数据又拥有  $m$  条属性. 其中每个  $x$  都代表一个被划分的市场属性, 需要做的是将这  $n$  个市场根据不同目的分成合适的具有相似特征的  $K$  个类, 其中的类簇数  $K$  由 Meanshift 算法自己生成, 每个类簇中的数据代表一种市场的集合<sup>[10]</sup>. 对基本向量的影响, 一般与采样点与参考点的距离成反比. 具体的步骤如下:

- (1) 随机选择一个未被分类的数据  $x$  作为质心
- (2) 将以  $x$  为质心, 半径为  $r$  的球状区域中的点为类簇  $C$  的元素, 并且将这些点在该区域出现的频次加 1
- (3) 计算漂移量  $M_h(x)$
- (4) 
$$M_h(x) = \frac{1}{h} \sum_{x_i \in s_r} (x_i - x) i = 1, 2, 3, \dots, m.$$
- (5) 其中  $x_i$  代表的是市场在空间中被表示的点,  $x$  代表质心,  $r$  则表示半径长度,  $s_r$  则表示这个空间中, 质心为  $x$  的球状空间,  $m$  表示在这个区域内点的个数.
- (6) 令  $x = x + M_h(x)$ , 即将质心偏移  $\|M_h(x)\|$ , 偏移的方向是  $M_h(x)$ .
- (7) 重复步骤(2)(3)(4)直到  $\|M_h(x)\|$  为零或者小于某值, 则收敛.
- (8) 重复步骤(1)(2)(3)(4)(5)直到所有的点都被分类.

统计每个点在各个类簇出现的频次, 将其中累计频次最大的点划分到对应的类簇中, 如果这个过程中有两个类簇的质心距离小于  $r$ , 则合并两个类簇为一个<sup>[11]</sup>.

至此我们将不同的市场划分出来, 整个过程如下图. 将市场内烟草属性、市场内烟草产地、市场的经济发展情况、市场内政策支持程度、卷烟推广成本等作为市场的属性, 经过 Meanshift 聚类以后, 生成不同的市场

画像, 如高端烟市场、高端烟潜力市场等等, 让商业企业对某区域的卷烟投放和发展规划都有一定调整.

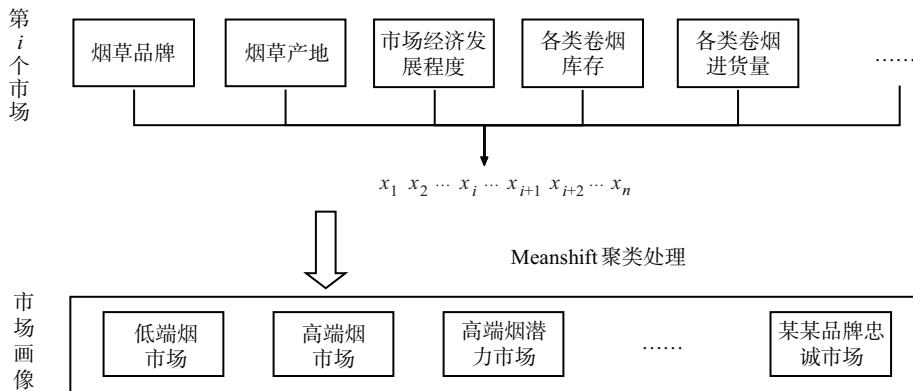


图 6 市场画像

### 3.6 品牌画像

品牌是名称、术语、标记、符号、设计或者其组合的产物, 起区分其他产品的作用. 通过品牌管控, 公司可以监督控制消费者与品牌的关系, 以达到形成品牌特有的竞争优势的目的.

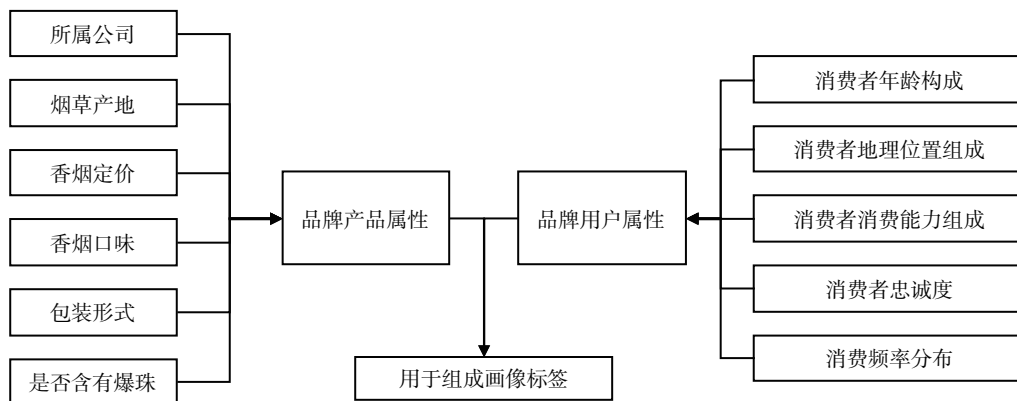


图 7 品牌属性

品牌作为香烟消费者群体里中的主要区分方式. 对品牌有科学、全面的认知会在推广、培训品牌的过程中起到重要的作用. 而新零售模式下, 借助大数据技术构建品牌画像也成为可能. 部分品牌画像标签如图, 我们用以下几个维度来建立了标签体系.

由于影响品牌的因素很多, 这个聚类算法是一个降维算法, 使得聚类的过程更加顺畅简单. t-SNE (t 分布随机邻域嵌入)<sup>[9]</sup>, t-SNE 算法是对称随机近邻嵌入 (SNE) 算法的改进<sup>[12-13]</sup>. 是一个简单而且便于可视化的方法.

以品牌画像举例, 假设品牌的数据集合为  $X$ , 其中有  $n$  个品牌. 每个品牌  $x_i$  都是  $n$  维的数据. 一般情况下选取为 2 维的低维平面. 就是将高维的品牌属性数据, 全部用二维数据表示出来. 原始数据之间的相似性一般通过下列欧氏距离转化的概率来表示:

$$P_{ji} = \frac{\exp(-\|x_j - x_i\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\sigma_i^2)}, \tag{1}$$

其中  $x_i$  为第  $i$  个品牌,  $x_j$  为第  $j$  个品牌,  $\sigma_i$  为预先设置的方差. 其结果就是说, 在低维平面上  $i$  会选择在高维空间中距离其近的点作为其邻居的可能性更高, 高维空间中距离其远的点的可能性很低. 经过 t-SNE 的降维以后, 再结合 K-Means 算法, 将不同的品牌分类将会更加容易.

在提取了上述标签之后, 根据不同的目的, 利用德尔菲法, 从不同维度去构建标签, 再建立品牌所对应的品牌画像. 品牌如果想要做好, 可以通过宣传、自主产权、和差异化的方法来实现<sup>[14]</sup>. 但是由于烟草行业的特殊性, 差异化是培育品牌的重要手段. 通过品牌画像, 来挖掘品牌特质, 再针对性的发展品牌, 将是未来品牌

培育的一条可行路线.

### 3.7 分析应用

新的消费者用户标签体系紧密结合了新零售模式,充分挖掘利用了客户信息,为后续的工商协同营销活动提供数据支撑.

精准客户维护:通过消费者用户画像,将消费者归为几类,对于不同的消费者类型进行不同的维护方式,例如给潜在客户推送其经常浏览品牌香烟活动信息,对于高价值的活跃用户推荐对应品牌的精品香烟等,以实现客户发掘,客户忠诚度的维护等效果.

精准品牌培育:通过了解品牌的销量、库存、品牌发展信息、主要消费群体等信息来为品牌建立画像.通过品牌画像的形成,对于香烟品牌的投放能够更加精准,以达到更好的实现销售目标,同时为消费者带来便利的目的.

精准的市场划分:通过对整个卷烟营销中的信息收集,建立市场画像.掌握市场上各品牌培育状况,对品牌培育提供帮助.根据市场特点对各品牌香烟有计划投放,规划出合理的品牌培育路线.

### 3.8 基于大数据的工商协同营销平台技术架构

营销平台实现框架示意图如下:

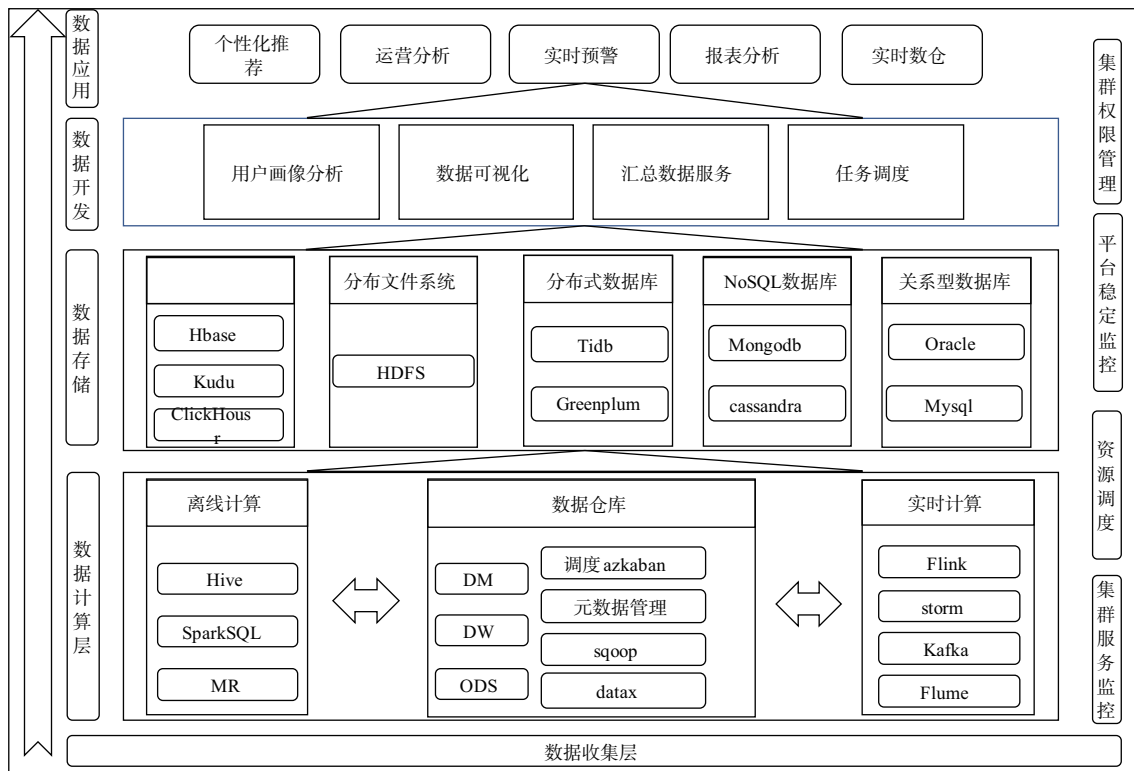


图 8 平台架构图

该方案的实时处理部分是由 flink 对接 kafka 消息流的处理方式,保证实时处理的同时有较高的稳定性,离线处理框架是基于 sparkSQL 的离线 ETL 框架对离线数据进行清洗压缩.任务调度则采用了稳定准确和易于管理的 azkaban.平台监控通过 flume、Filebeat 来收集日志,再将收集到的信息传输到 kafka 消息队列中,借助实时处理框架分析处理.

数据存储目前选用的是 ES 作为实时存储方案,后期再根据实时数据增长以及需求来调整,同时借助离线处理框架,对 HDFS 文件进行压缩合并,减轻服务器的负载.后续则将这些数据应用于不同的方案,用户画像分析等等.

## 4 结语

对新零售模式下面向消费者的工商协同营销模式进行了研究,介绍了山东泰安烟草有限公司和云南中

烟工业有限责任公司在工商协同营销方面的探索工作.结合新零售能提供的海量数据,使得通过大数据技术建立消费者画像,市场画像,品牌画像成为可能.形成了以消费者为核心,以市场为导向的现代营销体系.符合新零售的发展趋势与特点.提高了工商零对消费者的认知能力,优化了消费者的消费体验,为品牌培育方向和市场战略安排提供了重要的参考.该模式为烟草行业与大数据的结合,面向消费者的工商协同模式,为后续实践提供了重要的参考意义.

#### 参考文献:

- [1] 董闰.新零售发展趋势研究[J].经济研究导刊,2022(12):51-53.
- [2] 周勇,池丽华.新零售从1.0走向3.0[J].上海商学院学报,2022,23(3):83-95. Wit, E. and McClure, J. (2004) Statistics for Microarrays: Design, Analysis, and Inference. 5th Edition, John Wiley & Sons Ltd., Chichester, 5-18.
- [3] 高林,陈浩.互联网时代面向消费者的新型工商协同营销模式研究[J].现代商贸工业,2020,41(5):76-77.
- [4] 乔湛青.卷烟企业大数据营销策略研究[J].现代商贸工业,2016,37(23):50-51.
- [5] TRAVIS D. E-commerce usability: tools and techniques to perfect the on-line experience[M]. CRC Press, 2002.
- [6] 徐芳,应洁茹.国内外用户画像研究综述[J].图书馆学研究,2020(12):7-16.
- [7] 郭阳敏,唐云,杜宇,等. Cluster analysis of polymers using laser-induced breakdown spectroscopy with K-Means[J]. Plasma Science and Technology, 2018, 20(6): 103-107.
- [8] 王森,刘琛,邢帅杰. K-Means 聚类算法研究综述[J]. 华东交通大学学报, 2022, 39(5): 119-126.
- [9] 苏腾飞,张圣微,李洪玉.基于可变尺度 Mean-Shift 的农田高分遥感影像分割算法[J]. 国土资源遥感, 2017, 29(3): 41-50.
- [10] 王震飞.基于RFM模型的科学网博客博主群体画像研究——以图书馆学、情报学、档案学三个学科领域为例[J]. 情报探索, 2020(11): 26-33.
- [11] 王著鑫,耿秀丽,王龙羽,等.考虑 MeanShift 用户聚类的云服务推荐[J]. 软件导刊, 2022, 21(4): 169-176.
- [12] CHAN D M, RAO R, HUANG F, et al. GPU accelerated t-distributed stochastic neighbor embedding[J]. Journal of Parallel and Distribute Computing, 2019, 131(1): 1-13.
- [13] 崔文泉,黄禹侨.高维数据情形下的一种基于随机投影的集成分类方法[J].中国科学技术大学学报, 2019, 49(12): 974-984.
- [14] 陈泽勋,俞蘅君,乔添,等.生活用品销售市场分析——以卫生用品为例[J].中国商论, 2022(9): 59-61.

## Research on the industrial – commercial synergy marketing model based on big data

FENG Bo-xing<sup>1,2</sup>, LIU Yang<sup>1</sup>, ZHAO Tao<sup>1</sup>, Guo Liang<sup>1</sup>, PU Xue-song<sup>1</sup>, ZHOU Jia-xian<sup>1</sup>

(1. China Tobacco Yunnan Industrial Co., Ltd, Kunming 650000, China;

2. Shandong Tai'an Tobacco Monopoly Bureau, Tai'an 271000, China)

**Abstract:** With the development of informatization and the issuance of anti-smoking policies such as the "Opinions on Implementing the Healthy China Action" by the State Council, the policy orientation at the national level towards tobacco sales and purchases is gradually tightening. This poses severe challenges to the traditional tobacco industry. Since 2015, the development of new retail has been rapid, generating a substantial amount of data. Concurrently, the integration of the tobacco industry with big data is becoming imperative. The informatization of the tobacco industry and the consistent promotion of industrial-commercial marketing synergy have become crucial topics. The article explores the current concepts in the tobacco industry and cases of industrial-commercial synergy marketing. Combining big data with the unique characteristics of the tobacco industry, the study focuses on researching the industrial-commercial synergy model in the tobacco industry. The new model holds significant guiding significance for the informatization reform and development of the tobacco industry.

**Key words:** industrial-commercial deep synergy; tagging; big data; profiling