电商行业应用领域的数据分析现状

1. 背景

电商行业产生了众多客户的网络操作数据，这些数据可以用来进行商业问题和行业趋势等分析研究。对于这些繁杂的网络数据信息，首先面临的挑战就是海量数据信息的分析问题。面对大量的数据，一方面需要更先进的技术设备来进行分析支持，再有就是数据质量难以保证，需要花费大量时间进行数据处理和确认；其次，数据信息的真假不容易辨别；最后，数据信息的形式不统一，难以分析。如何解析已有的历史数据并用于预测未来的行为，如何从这些海量数据中发现关键问题，成为了现代电商企业业绩增长的重要途径。

1. 目的

在大数据环境下利用数据挖掘技术来准确了解客户的真正需求；精准定位客户群体从而实施精细化的营销策略；利用自身已有的客户资源，在不增加新的资源成本的情况下实现自身利益和客户满意最大化的双赢局面。

1. 主要任务
   1. 顾客划分

消费者在各个电商平台上进行浏览商品信息时，会产生大量的数据信息，可以对消费者的这些信息进行分析，从而构建出各种用户画像，比如说消费者在首次登陆注册账号填写的各项信息、商品购买记录、历史浏览记录、方位信息、支付方式以及银行账号信息等，同时更加构建各种用户画像对消费者进行分类，为不同类别、不同层次的消费者提供差异化服务。

* 1. 顾客评估

利用大数据分析技术，结合构建的各种用户画像，可以将消费者按照特定原则划分为两个类别，分别是一般消费者群以及重要消费者群，对各个消费者建立正确的认知，并以此为基础为消费者提供高效、有用的产品与服务。比如说通过大数据分析技术，能够得知许多消费者在购买基金理财产品以后，还有很大概率会购买保险产品，为此，电商平台可以对这部分消费者的消费行为、习惯以及特征进行提取和分析，为其提供与之相匹配的差异化的产品方案。

* 1. 个性化推荐

依托消费者移动智能终端设备的数据以及在其他非电商网站的浏览数据，将这些数据信息导入到大数据平台上，实现数据信息的共享，进行消费者画像分析，为消费者推荐相应的产品，实现精准促销，加强消费者与企业之间的数据信息交互。例如根据百度搜索记录进行商品推荐，根据视频网站播放信息，推荐相关产品。

* 1. 优化销售平台

利用大数据挖掘还能够对消费者在消费习惯、访问习惯等方面有一定的了解,然后为网站和电商平台提供所需要的内容进行参考。电商网站通过用户访问以及消费上的实际情况来调整电商网站内容,例如将点击和交易量最高的商品放在网站首页,以此来吸引广大的消费者,并激发其内在的购买愿望。同时,还可以通过大数据挖掘对用户在各个电商平台上的数据进行分析,并根据不同客户的实际需求建立起其与网页之间所存在的关联性,根据用户自身的需求将需要的导航添加到界面中,然后再科学安排服务器缓存,最大程度降低服务器在响应上所消耗的时间,以此来有效提升用户对于电商服务的满意度。

* 1. 拓展增值业务

如果电商平台中的用户数据已经形成了一定的规模,此时则需要构建完善的数据库,然后对这些数据信息进行分析,通过得到各个用户的分析,商家可以为相应的用户提供所需要的商品,如果能够引起用户的兴趣,并购买了该商家的产品,那么则能够有效提高商家自身的收入。如果当企业或是商家没有利用大数据挖掘,那么将会在一定程度上大大增加业务开发难度,例如消费信贷,如果使用大数据挖掘,能够发现存在于大数据市场中的潜在价值,促使其更好的进行业务开发,如小额贷款等等。总体来说,数据信息的挖掘对于拓展电商企业和商家的增值业务能够产生积极作用,有利于电商市场的发展。

1. 常用模型
   1. RFM模型

RFM模型是衡量客户价值和客户创利能力的重要方法。通过一个客户的近期购买行为、购买的总体频率以及花费金额三项指标来描述该客户的价值状况。R(Recency)表示客户最近一次购买的时间，F(Frequency)表示客户在最近一段时间内购买的次数，M (Monetary)表示客户在最近一段时间内购买的金额。在电商数据分析中RFM模型被广泛运用，一般的CRM分析着重对于客户贡献度的分析，RFM则强调以客户的行为来进行客户细分。

给三个变量不同的权重或按一定的规则进行分组，然后组合使用，即可分出很多不同级别的会员。利用 RFM分析，我们可以建立会员金字塔，区分各个级别的会员，如高级会员、中级会员、低级会员，然后针对不同级别的会员施行不同的营销策略，制定不同的营销活动；发现流失及休眠会员，通过对流失及休眠会员的及时发现，采取营销活动，激活这些会员；在短信、EDM促销中，依据会员的精细化细分，更精准定位用户和挖掘用户需求；维系老客户，提高会员的忠诚度。

* 1. 关联分析

关联分析最经典的案例来自于沃尔玛的“啤酒与尿布”。它的基本思想就是买了A商品的很多用户又买了B商品，那么我们就可以认为A、B两个商品的关联性比较高，通过这种分析进行捆绑销售或相关陈列。

* 1. 聚类分析

电商行业的聚类分析主要是指将具有相似购物行为的用户进行群体的细分，以支持精细化的营销活动，带来更大的营销效果，从而节省成本。聚类分析是进行会员精细化管理，精细化营销的基础，可以构建用户画像实现精准化营销。

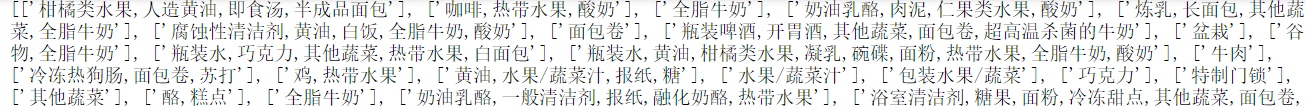
1. 模型分析

以关联分析为例，它是是一种无监督的学习算法，用于寻找交易事务中不同商品之间存在的关联关系。频繁项集是指在数据集中频繁出现的项集，该项集是满足最小支持度、置信度阈值的所有项的集合。支持度是指关联的数据在数据集中出现的次数在总数据集中所占的比例。例如购买了A商品，还购买了B商品，支持度计算公式为Support(A→B)=P(A∪B），该度量值表示A商品和B商品同时出现的概率。置信度用来表示一个数据出现后，另一个数据出现的概率，即数据的条件概率。置信度的计算公式为Confidence (A→B)=Support(A∪B）/Support(A）该度量值表示商品A和商品B同时出现的概率占商品A出现概率的比值。

其中应用比较广泛的是Apriori算法，它包含两个步骤，连接和剪枝。Apriori算法是将最小支持度和数据集作为输入参数，算法首先扫描所有项集表，生成所有单个元素的项集，接着通过比较项集大于或等于最小支持度和置信度，不满足的项集将会被删除，接着通过连接生成包含两个元素的项集，重复上述步骤，在通过算法中的剪枝，将不满足先验性原理的项集删除，直到找到满足要求的频繁项集。

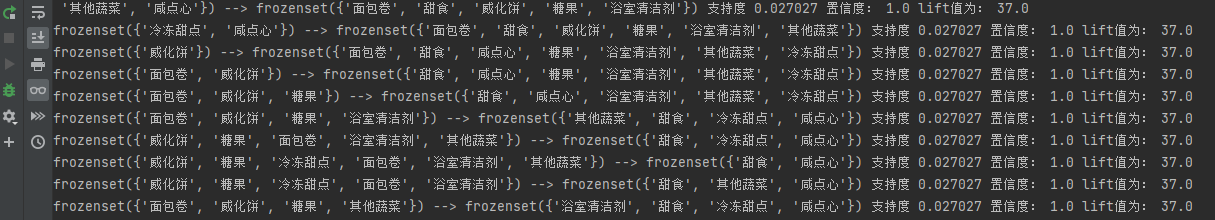
下面以某商品的零售企业购物车数据为例，进行Apriori关联分析。

首先对原始数据进行数据的预处理，转换数据形式，使其满足Apriori关联分析的要求。



然后构建Apriori模型，调整超参数。首先设置Apriori的最小支持度、最小置信度、输入建模样本数据，采用Apriori关联分析算法对建模样本数据进行分析，以模型设置的参数作为条件和目标，目前如何设置最小支持度与最小置信度设置并没有统一的大小的标准，大部分都是通过设置初始值，然后通过不断的调整，获取与业务相符的关联结果。本文输入的模型参数为：最小支持度0.02，最小置信度为0.35。

最后对模型结果进行分析，部分结果如下。



置信度为同时购买两件商品的概率，支持度为这种情况发生的可能性。

1. 参考文献：

[1]赵雅菲.数据挖掘对电商平台的影响[J].现代营销(下旬刊),2017,(06):140.

[2]高昀. 基于社交电商平台的用户行为分析与研究[D].北京邮电大学,2021.DOI:10.26969/d.cnki.gbydu.2021.001850.

[3]罗燕君.大数据背景下电子商务精准营销策略[J].鄂州大学学报,2021,28(04):51-54.DOI:10.16732/j.cnki.jeu.2021.04.016.