

大数据分析

TIPDM

目录

Contents

实例引入：1 个性化用户
画像实现精准营销

实例引入：1 个性化用户
画像实现精准营销

2
大数据分析技术

大数据分析技术

3
主流的大数据分析处理框架

主流的大数据分析处理框架



实例引入：个性化用户画像实现精准营销

- 用户想购买一辆汽车，由于该用户经常通过某个网站浏览不同品牌和价格的汽车商品简介，因此，用户的浏览记录被存储在在网站后台数据库中，包含浏览的产品价位、汽车品牌、汽车的功能配置参数等。
- 汽车销售的技术人员也可以获取用户的基本信息和消费记录，通过大数据分析算法对该用户的个人喜好和购买能力进行分析，最终得到用户可能会购买的汽车品牌信息。
- 汽车销售的实例中，购车用户的特征可以通过用户的历史浏览数据进行描绘，形成该用户的用户画像，并依据特征对该用户未来的消费趋势进行预测，把用户可能会购买的商品推荐给用户，实现在大数据时代下的精准营销策略。

大数据实现精准营销

- 精准营销是在精准定位的基础上，依据现代信息技术特别是近些年发展快速的大数据技术，对企业的营销实施可衡量并且回报率高的精准策略，降低企业的营销成本，提升市场竞争力。
- 精准营销以用户为中心，通过现代化技术手段直接与用户沟通，使企业收集大量的用户数据，借助大数据分析技术，将用户数据加工为有用信息，然后企业利用加工后的信息，为用户推荐个性化产品，使用户享受到专业的客户服务。。
- 精准营销的关键在于如何精准地找到产品的目标人群，再让产品深入用户心坎里，让用户认识产品、了解产品、信任产品到最后依赖产品。

大数据实现精准营销

- 以选购汽车为例，为了满足用户的需求，汽车企业应从多个角度进行营销。一方面，将产品做好、做精、做强、生产出更多符合不同用户要求的产品。
- 另一方面，将汽车产品信息传达给目标用户，引领用户的选择，寻找吻合度高、对受众影响大的媒体进行宣传，在访问量较大的网站上进行汽车广告推送，增大用户点击感兴趣的商品的概率。
- 在网站上推送的汽车车型，由访问该网站的用户特征决定。通过用户画像进行精准营销。



什么是用户画像

- 阿兰·库珀（Alan Cooper）最早提出了用户画像（Persona）的概念，认为“用户画像是真实用户的虚拟代表，是建立在一系列真实数据之上的目标用户模型”。用户画像也称为用户的信息标签。
- 用户画像的主要用途是帮助商家了解用户，对用户了解得越深，刻画出的画像就越准确，用户画像被大量地应用在精准营销和智能推荐领域，是真实世界的用户在网络世界的映射



什么是用户画像

- 大数据时代的用户画像和传统的画像完全不同，传统的用户画像指的是画家利用画笔对用户的外貌进行描绘，体现出的是用户的轮廓和形态。互联网时代下的用户画像是根据用户社会属性、生活习惯和消费行为等信息抽象出的一个标签化的用户模型，即构建用户画像的核心工作是给用户贴“标签”，标签是通过对用户数据分析得到的高度精练的特征标识。
- 互联网时代的用户画像表现出来的信息更加丰富，信息种类也不局限于视觉特征，凡是能够对用户的特征进行描述的信息，都可以放到用户画像里面。大数据算法构建出来的用户画像具有更加丰富的属性，可以被更多的上层应用使用
- 通过对用户数据的分析，可以对用户进行画像，给出某个特定用户的相关信息，如年龄区间、从事的职业、婚姻状况、家庭成员、消费习惯、个人爱好、是否从事体育运动、消费习惯、经常购买哪类商品等

构建个性化用户画像

用于构建用户画像的数据，不仅需要数量多，而且还要和业务场景紧密结合，在本章介绍的汽车销售实例中，为了精准地推送给用户汽车的广告，首先要对用户进行用户画像，将用户的特征描述清楚，然后再根据画像的特征进行精准营销

- 如果用户画像勾勒出来的是一个年龄在20~30岁的年轻人，从事IT行业，平时喜欢选购电子产品，那么可以分析出该年轻人会比较钟情于经济型轿车
- 如果给出的用户画像是一个对美术比较感兴趣的人，热爱网络小说和文学，偏好人文社科书籍，也热爱音乐和舞蹈，那么可能会对外观设计和内饰风格比较注重，同时也会关注车辆的舒适性和安全性能，感兴趣的可能是运动型多用途汽车的车型获取信息的便利性增大

构建个性化用户画像

用户画像的主要步骤可以分为如下3步

- 首先需要明确研究的目标，即对哪些用户进行画像。例如，为了研究电商平台用户流失的情况，就要将那些购物体验较差的用户设定为目标用户；如果要研究潜在客户是否能成为正式客户，就要将那些目前还未接触过本产品但采购了同类型其他品牌商品的用户设定为目标用户
- 对目标用户的所有的相关数据进行收集，如用户的性别、职业、年龄、地域、消费层次等基本信息；也可以是用户的行为信息，如浏览记录、搜索过的关键词、发表过的评论等
- 通过大数据分析技术，包括描述性统计分析、数据挖掘算法等，为用户贴上相应的标签，标示出用户的兴趣、偏好和需求等。

通过用户画像实现精准营销

构建好用户画像后，即可对用户需求、基本特征、用户价值进行分析，实现精准营销

- 用户需求分析：了解用户需要什么，才能精准地提供需要的服务和商品。通过大数据分析实现对用户画像，可以得到准确的用户需求。在移动互联网时代，用户的消费数据不断积累，利用用户消费数据可勾画出用户可能需要哪类商品，用户的需求隐含在其浏览和选购过程中，是更深层次需求的挖掘，需要对用户的消费习惯进行分析。

通过用户画像实现精准营销

- 用户基本特征分析：用户画像是对一个用户全方位的展示，为了让用户的画像内容丰富，标签要尽量多。用户画像的目的之一是企业找到目标用户，目标用户是可能要购买企业产品的，并且是有能力购买的。例如，用户购买产品的类型、采购的频率、采购商品的价格、用户所在的区域等基本属性信息就非常重要，了解了基本属性信息，企业可以和用户进行沟通，将产品推荐给用户通过大数据分析技术，包括描述性统计分析、数据挖掘算法等，为用户贴上相应的标签，标示出用户的兴趣、偏好和需求等。
- 用户价值分析：在对人物画像时，可以根据大数据分析给出人物的价值特征。用户价值可以理解为用户在系统中的商业变现能力，包括广告价值、付费价值。

目录

Contents

实例引入：1 个性化用户
画像实现精准营销

回顾营销流程搭建基础

2 大数据分析技术

大数据分析技术

3 主流的大数据分析处理框架

主流的大数据分析处理框架



大数据分析技术

1. 数据分析与数据挖掘

大数据时代的战略意义不仅在于掌握庞大的数据信息，还在于发现和理解信息内容及信息与信息之间的关系，而大数据分析就是大数据研究领域的核心内容之一。大数据分析是决策过程中的决定性因素，也是大数据时代发挥数据价值的关键环节。

大数据分析核心即为挖掘。

- 数据分析的定义是用适当的统计分析方法对收集来的大量数据进行分析，将数据加以汇总和理解并消化，以求最大化地开发数据的功能、发挥数据的作用
- 数据分析是为了提取有用信息和形成结论而对数据加以详细研究和概括总结的过程

大数据分析技术

1. 数据分析与数据挖掘

- 数据分析的数学基础在20世纪早期就已确立，但直到计算机的出现才使得实际操作成为可能，而随着计算机的不断发展，数据分析也得以推广。数据分析是数学与计算机科学相结合的产物
- 数据分析的目的是将隐藏在一大批看来杂乱无章的数据中的信息集中和提炼出来，从而找出所研究对象的内在规律

大数据分析技术

1. 数据分析与数据挖掘

- 数据挖掘是指通过人工智能、机器学习等方法，从大量的数据中挖掘出未知的且有价值的信息和知识的过程
- 数据挖掘主要侧重解决4类问题，即分类、聚类、关联和预测
- 数据挖掘的重点在于寻找未知的模式与规律，寻找那些事先未知的但又非常有价值的信息，主要采用统计学、人工智能、机器学习等方法进行挖掘
- 数据分析是将数据变成信息的方法，数据挖掘是将信息变成认知的方法，如果想要从数据中提取一定的规律往往，需要数据分析和数据挖掘结合使用

大数据分析技术

2. 数据认知

随着大数据技术和体系的发展，越来越多的人使用大数据技术。大数据技术是以数据为核心的，人们对大数据的认知和传统数据有着很大区别。数据的质量也有区别，数据质量分析的主要任务是检测原始数据中是否存在脏数据，脏数据一般是指不符合要求的数据

- 数据完整性
- 数据准确性
- 数据重复性
- 数据的一致性

大数据分析技术

2. 数据认知

科学地分析数据特征是数据分析的基础，对数据分析而言，对数据特征的准确把握是至关重要的。数据特征分析常用的方法有分布分析、对比分析、统计分析、相关性分析

➤ 分布分析

分布分析是指根据数据在坐标图里分布的特点来对数据进行分析的方法，在生产工作正常的情况下，产品的质量不可能完全相同，但也不会相差太大，而是围绕着一一定的平均值，在一定的范围内变动和分布。分布分析是通过对质量的变动分布状态的分析发现问题的一种重要方法。

大数据分析技术

2. 数据认知

科学地分析数据特征是数据分析的基础，对数据分析而言，对数据特征的准确把握是至关重要的。数据特征分析常用的方法有分布分析、对比分析、统计分析、相关性分析

➤ 对比分析

对比分析主要是分析两个相互联系的指标，从数量上展示和说明研究对象的各种关系（规模的大小、水平的高低、速度的快慢等）是否协调，分析其中的差异，从而揭示事物代表的发展变化情况和变化规律。对比分析分为绝对数比较和相对数比较

大数据分析技术

2. 数据认知

科学地分析数据特征是数据分析的基础，对数据分析而言，对数据特征的准确把握是至关重要的。数据特征分析常用的方法有分布分析、对比分析、统计分析、相关性分析

➤ 统计分析

统计分析是对定量数据进行统计描述，常从集中趋势度量和离中趋势度量两个方面分析。

- 集中趋势度量。集中趋势度量是指数据向某一中心靠拢的倾向，核心是寻找数据的代表值或中心值，通过算数平均数、中位数和众数来度量。
- 离中趋势度量是指一组数据中各数据以不同程度的距离偏离中心的趋势。衡量离中趋势的4个度量值分别为极差、分位距、标一组准差和方差，其中分位距最常用的是四分位距。

大数据分析技术

2. 数据认知

科学地分析数据特征是数据分析的基础，对数据分析而言，对数据特征的准确把握是至关重要的。数据特征分析常用的方法有分布分析、对比分析、统计分析、相关性分析

➤ 相关性分析

数据相关性是指数据之间存在某种关系，该关系一般通过相关系数来体现，而相关系数就是用于反映变量之间相关关系密切程度的统计指标。

相关系数是研究变量之间线性相关程度的量，是按积差方法计算，以两个变量与各自平均值的离差为基础，通过两个离差相乘来反映两个变量之间的相关程度。常见的相关系数有两类，分别是皮尔逊积矩相关系数（Pearson相关系数）和斯皮尔曼秩相关系数（Spearman等级相关系数）。

大数据分析技术

3. 数据处理

对海量的数据进行处理时，处理的方式包括**数据清洗**、**数据规约**、**数据变换**。

数据清洗包括4个方面：

- 缺失值分析处理
- 异常值分析处理
- 重复值分析处理
- 数据一致性分析处理

大数据分析技术

3. 数据处理

数据规约技术能在很大程度上移除数据中错误的实例或样本属性，不但能提升数据挖掘的速度，还会提升数据挖掘的准确度。

数据规约算法可分为5类：

- 特征选取（Feature Selection）
- 实例选取（Instance Selection）
- 离散化（Discretization）
- 特征提取（Feature Extraction）
- 实例生成（Instance Generation）

大数据分析技术

3. 数据处理

数据规约算法说明

算法	说明
特征选取	用于减少数据的维度，从数据维度的角度出发，目的在于移除数据集合中的不相关或冗余属性，最终选出一个能代表或接近原始集合数据分布的属性子集
实例选取	用于减少数据集合中实例样本数据样本的数量，目的在于选出能代表集合特征的实例子集，其随机选取的方式被称为取样，常用在大体量数据集合中，防止数据的过拟合
离散化	又称作特征简化，用于简化样本属性的描述，是将一种定量化的数据转换为另一种定量化数据的过程，该过程会将数据集合中的数值属性进行离散化处理，转化为在一定区间内的有限数值。在后续的挖掘过程中，可将数据属性当成固定区间内的可计算数值进行处理
特征提取	用于生成新的属性或样本，主要分为线性和非线性提取两种方式。
实例生成	实例生成算法除了移除数据集合中的数据，在规约的过程中还会对原始集合中的样本进行改动，抽取多个样本特征，生成更能代表数据特征的新样本

大数据分析技术

3. 数据变换

数据变换是将数据进行转换或归并，通过平滑处理、数据泛化、规格化等方法将数据转换成适用于数据挖掘的形式

方法	方法说明
平滑处理	帮助去除数据中的噪声
合计处理	对数据进行总结或合计操作
数据泛化处理	用更抽象的概念取代低层次或数据层的数据对象
规格化处理	将有关属性数据按比例投射到特定的小范围之中
属性构造	根据已有属性集构造新的属性，以在数据处理过程中起帮助作用

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：
<https://d.book118.com/557031120153006156>