

第1章 跨境电子商务数据分析



学习目标

- 了解大数据的概念与特征
- 了解跨境电子商务数据分析的基础知识
- 熟悉跨境电子商务数据分析的原则
- 掌握跨境电子商务数据分析的流程
- 熟悉跨境电子商务数据分析的价值



技能目标

- 能够运用大数据处理工具解决实际问题
- 能够将大数据应用到实际工作中

1.1 大数据认知

随着物联网、云计算、移动互联网的迅猛发展，大数据（big data）吸引了越来越多的关注，正成为信息社会的重要财富，也给数据的处理与管理带来了巨大挑战。

1.1.1 大数据的概念与特征

1. 大数据的概念

很多人这样理解大数据：“大数据就是大规模的数据。”这个说法真的准确吗？

其实，“大规模”只是针对数据的量而言的，数据量大，但并不代表数据一定有可被数据挖掘、深度学习算法利用的价值。例如，在地球绕太阳运转的过程中，每秒记录一次地球相对太阳的运动速度、位置，可以得到大量数据，但如果只有这样的数据，其实并没有太多可以挖掘的价值。

关于大数据，这里参考马丁·希尔伯特（Martin Hilbert）的总结：今天我们常说的大数据其实是在 2000 年后，因为信息交换、信息存储、信息处理三个方面能力的大幅增长而产生的数据，如图 1-1 所示。

（1）信息交换。据估算，1986—2007 年，地球上每天可以通过既有信息通道交换的信息数量增长了约 217 倍，这些信息的数字化程度则从 1986 年的约 20% 增长到 2007 年的

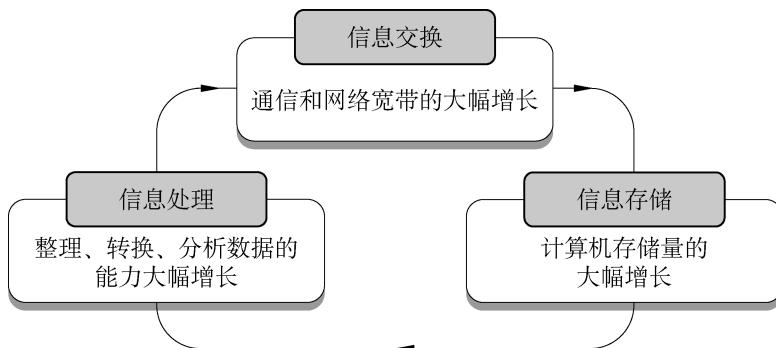


图 1-1 信息交换、信息存储、信息处理

约 99.9%。在数字化信息爆炸式增长的过程中，每个参与信息交换的节点都可以在短时间内接收并存储大量数据。

(2) 信息存储。全球信息存储能力大约每三年翻一番。1986—2007 年，全球信息存储能力增加了约 120 倍，所存储信息的数字化程度也从 1986 年的约 1% 增长到 2007 年的约 94%。1986 年，即便用上所有的信息载体、存储手段，我们也不过能存储全世界所交换信息的大约 1%，而 2007 年这个数字已经增长到大约 16%。信息存储能力的增加为我们利用大数据提供了近乎无限的想象空间。

(3) 信息处理。有了海量的信息获取能力和信息存储能力，我们也必须有对这些信息进行整理、加工和分析的能力。Google、Facebook 等公司在数据量逐渐增大的同时，也相应建立了灵活、强大的分布式数据处理集群。

2. 大数据的特征

大数据的 5V 特征如图 1-2 所示。

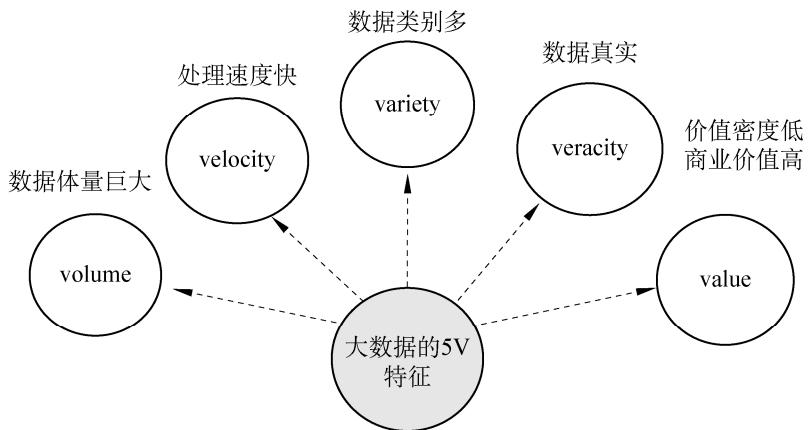


图 1-2 大数据的 5V 特征

(1) 数据体量巨大 (volume)。这是指收集和分析的数据量非常大，从 TB 级跃升到 PB 级。在实际应用中，很多企业用户把多个数据集放在一起，已经形成了 PB 级的数据量。

(2) 处理速度快 (velocity)。需要对数据进行近实时的分析，以视频为例，在连续不



间断的监控过程中，可能有用的数据持续的时间仅仅有一两秒。这一点和传统的数据挖掘技术有着本质的不同。

(3) 数据类别多 (variety)。大数据来自多种数据源，数据种类和格式日渐丰富，包括结构化、半结构化和非结构化等多种数据形式，如网络日志、视频、图片、地理位置信息等。

(4) 数据真实性 (veracity)。大数据的内容与真实世界息息相关，研究大数据就是从庞大的网络数据中提取出能够解释和预测现实事件的信息的过程。

(5) 价值密度低、商业价值高 (value)。通过分析数据可以得出如何抓住机遇及实现价值。

1.1.2 大数据的处理工具与应用

1. 大数据的处理工具

现有的大数据处理工具大多是对开源的 Hadoop 平台进行改进并将其应用于各种场景。在 Hadoop 完整生态系统中，各子系统都有相应大数据处理的改进产品。常用的大数据处理工具如表 1-1 所示，这些工具有的已经投入商业应用，有的是开源软件。在已经投入商业应用的产品中，绝大部分也是在开源 Hadoop 平台的基础上进行功能扩展，或者提供与 Hadoop 的数据接口。

表 1-1 常用的大数据处理工具

种 类		工 具 示 例
平台	Local	Hadoop、MapR、Cloudera、Hortonworks、BigInsights、HPCC
	Cloud	AWS、Google Compute Engine、Azure
数据库	SQL	MySQL (Oracle)、MariaDB、PostgreSQL、TokuDB、Aster Data、Vertica
	NoSQL	HBase、Cassandra、MongoDB、Redis
	NewSQL	Spanner、Megastore、F1
数据仓库		Hive、HadoopDB、Hadap
数据收集		ScraperWiKi、Needlebase、bazhuayu
数据清洗		DataWrangler、Google Refine、OpenRefine
数据处理	批处理	MapReduce、Dyrad
	流式计算	Storm、S4、Kafka
	内存计算	Drill、Dremel、Spark
查询语言		HiveQL、Pig Latin、DryadLINQ、MRQL、SCOPE
统计与机器学习		Mahout、Weka、R、RapidMiner
数据分析		Jaspersoft、Pentaho、Splunk、Loggly、Talend
可视化分析		Google Chart API、Flot、D3、Processing、Fusion Tables、Gephi、SPSS、SAS、R、Modest Maps、OpenLayers



2. 大数据的应用

(1) 商品零售大数据。在美国，有一位父亲怒气冲冲地跑到 Target 卖场，质问为何将带有婴儿用品优惠券的广告邮件寄给他正在念高中的女儿。然而后来证实，他的女儿果真怀孕了。这名女孩搜索商品的关键词，以及在社交网站所显露的行为轨迹，使 Target 捕捉到了她的怀孕信息。相关模型发现，许多孕妇在第二个妊娠期开始时会买许多大包装的无香味护手霜，在怀孕的最初二十周内会大量购买补充钙、镁、锌的善存片之类的保健品。最后，Target 选出了 25 种典型商品的消费数据构建了“怀孕预测指数”。通过这个指数，Target 能够在很小的误差范围内预测顾客的怀孕情况，因此 Target 就能早早地把孕妇优惠广告寄给顾客。

阿里巴巴公司根据淘宝网上中小企业的交易状况筛选出财务健康和讲究诚信的企业，对它们发放无须担保的贷款。零售企业会监控顾客在店内的走动情况及其与商品的互动，并将这些数据与交易记录相结合来展开分析，从而针对销售哪些商品、如何摆放货品及何时调整售价给出建议。此类方法已经帮助某领先零售企业减少了 17% 的存货，同时在保持市场份额的前提下，增加了高利润率自有品牌商品的比例。

(2) 消费大数据。亚马逊“预测式发货”的新专利，通过对用户数据的分析，在他们正式下单购物前，提前发出包裹。这项技术可以缩短发货时间，从而降低消费者前往实体店的冲动。从下单到收货之间的时间延迟可能会降低人们的购物意愿，导致他们放弃网上购物。所以，亚马逊会根据之前的订单和其他因素，预测用户的购物习惯，从而在他们实际下单前便将包裹发出。根据该专利文件，虽然包裹会提前从亚马逊发出，但在用户正式下单前，这些包裹仍会暂存在快递公司的转运中心或卡车上。

为了确定要运送哪些货物，亚马逊会参考之前的订单、商品搜索记录、愿望清单、购物车，甚至包括用户的鼠标在某件商品上悬停的时间。

(3) 中国证券监督管理委员会大数据。回顾“老鼠仓”的查处过程，在马乐案中，大数据首次介入。深圳证券交易所此前通过大数据查出的可疑账户多达三百个。实际上，早在 2009 年，上海证券交易所曾经有过利用大数据设置“捕鼠器”的设想。通过建立相关的模型，设定一定的预警指标，即相关指标达到某个预警点时，监控系统会自动报警。

而此次在马乐案中亮相的深圳证券交易所的大数据监测系统更是引起了广泛关注。深圳证券交易所设置了两百多个指标用于监测估计，一旦出现股价偏离大盘走势的情况，深圳证券交易所就会利用大数据查探异动背后有哪些人或机构在参与。

(4) 金融大数据。阿里“水文模型”会按小微企业类目、级别等统计商户的相关“水文数据”。例如，过往每到某个时点，某店铺的销售就会进入旺季，销售额就会增长，其对外投放的资金额度也会上升。结合这些“水文数据”，系统可以判断出该店铺的融资需求；结合该店铺以往资金支用数据及同类店铺资金支用数据，可以判断出该店铺的资金需求额度。

(5) 金融交易大数据。量化交易、程序化交易、高频交易是大数据应用比较多的领域。全球 2/3 的股票交易量是由高频交易所创造的，参与者总收益每年高达 80 亿美元。其中，大数据算法被用来做出交易决定。现在，大多数股权交易都是通过大数据算法进行



的，这些算法越来越多地开始考虑社交媒体网络和新闻网站的信息，从而在几秒内做出买入和卖出的决定。

当一种产品可以在多个交易所交易时，会形成不同的定价。谁能够最快地捕捉到同一种产品在不同交易所之间的显著价差，谁就能捕捉到瞬间套利机会。在这一过程中，大数据技术成了重要因素。

(6) 制造业大数据。在摩托车生产商哈雷·戴维森公司位于宾夕法尼亚州约克市新翻新的摩托车制造厂中，软件不停地记录着各种制造数据，如喷漆室风扇的速度，等等。当软件“察觉”风扇速度、温度、湿度或其他变量偏离规定数值时，它就会自动调节相应的机构。哈雷·戴维森公司还使用软件寻找制约公司每86秒完成一台摩托车制造工作的瓶颈。这家公司的管理者通过研究数据发现安装后挡泥板的时间过长。通过调整工厂配置，哈雷·戴维森公司提高了安装该配件的速度。

美国一些纺织及化工生产商根据从不同的百货公司POS机上收集的产品销售速度信息，将原来的18周送货周期缩短到3周。如此一来，百货公司分销商能以更快的速度拿到货物，减少仓储。对生产商来说，积攒的材料仓储也能减少很多。

(7) 医疗大数据。谷歌基于每天来自全球的三十多亿条搜索指令设立了一个系统，这个系统在2009年甲流暴发之前就开始对美国各地区进行“流感预报”，并推出了“谷歌流感趋势”服务。

谷歌在这项服务的产品介绍中写道：搜索流感相关主题的人数与实际患有流感的人数之间存在着密切的关系。虽然并非每个搜索“流感”的人都患有流感，但谷歌发现了一些检索词条的组合并用特定的数学模型对其进行分析，这些分析结果与传统流感监测系统监测结果的相关性高达97%。这就表示谷歌公司能做出与疾控部门同样准确的传染源位置判断，并且在时间上提前了1~2周。

继世界杯、高考、景点和城市预测之后，百度又推出了疾病预测产品。目前，百度可以就流感、肝炎、肺结核、性病四种疾病，对全国每个省份及大多数地级市和区县的活跃度、趋势图等情况进行全面的监控。未来，百度疾病预测监控的疾病种类将从目前的四种增加到三十多种，覆盖更多的常见病和流行病。用户可以根据当地的预测结果进行针对性预防。

Seton Healthcare是采用IBM最新沃森技术医疗保健内容分析预测的首个客户。该技术允许企业找到大量与病人相关的临床医疗信息，通过大数据处理，更好地分析病人的信息。在加拿大多伦多的一家医院，针对早产儿，每秒钟有超过三千次的数据读取。通过数据分析，医院能够提前知道哪些早产儿可能出现问题，并且有针对性地采取措施，避免早产儿夭折。

大数据让更多的创业者更方便地开发产品，如通过社交网络来收集数据的健康类App，也许在数年后，它们收集的数据能让医生的诊断变得更为精确。社交网络为许多慢性病患者提供了临床症状交流和诊治经验分享平台，医生借此可获得在医院通常得不到的临床效果统计数据。基于对人体基因的大数据分析，可以实现对症下药的个性化治疗。公共卫生部门可以通过全国联网的患者电子病历库快速检测传染病，进行全面疫情监测，并通过集成的疾病监测和响应程序快速做出响应。



(8) 交通大数据。UPS 最新的大数据来源是安装在公司 4.6 万多辆卡车上的远程通信传感器，这些传感器能够传回车速、方向、刹车和动力性能等方面的数据。收集到的数据流不仅能反映车辆的日常性能，还能帮助公司重新设计物流路线。大量的在线地图数据和优化算法最终能帮助 UPS 实时地调整驾驶员的收货和配送路线。该系统为 UPS 减少了 8500 万英里^①的物流里程，由此节省了 840 万加仑^②的汽油。

可基于用户和车辆的 LBS（基于位置服务）定位数据，分析人车出行的个体和群体特征，进行交通行为的预测。交通部门可预测不同时点不同道路的车流量，进行智能的车辆调度或应用潮汐车道。用户则可以根据预测结果选择拥堵概率更低的道路。百度基于地图应用的 LBS 预测涵盖范围更广。在春运期间预测人们的迁徙趋势，指导火车线路和航线的设置；在节假日预测景点的人流量，指导人们进行景区选择；在平时通过百度热力图告诉用户城市商圈、动物园等地点的人流情况，指导用户进行出行选择和商家选点选址。

(9) 公安大数据。大数据挖掘技术的底层技术最早是英国军情六处研发用来追踪恐怖分子的技术。利用大数据技术可筛选犯罪团伙，如与锁定的罪犯乘坐同一班列车、住同一酒店的人可能是其同伙。过去，刑侦人员要证明这一点，需要通过把不同线索拼凑起来排查疑犯。

通过对越来越多数据的挖掘分析，可显示某一区域的犯罪率及犯罪模式。大数据可以帮助警方定位最易受到不法分子侵扰的区域，创建一张犯罪高发地区热点图和时间表。这不但有利于警方精准分配警力、预防打击犯罪，也能帮助市民了解情况、提高警惕。

(10) 文化传媒大数据。与传统电视剧有别，《纸牌屋》是一部根据“大数据”制作的作品。制作方 Netflix 是美国最具影响力的影视网站之一，在美国本土有约 2900 万名订阅用户。Netflix 的成功之处在于其强大的推荐系统 CineMatch，该系统将用户视频点播的基础数据，如评分、播放、快进、时间、地点、终端等存储在数据库中，然后通过数据分析，推断出用户可能喜爱的影片，并为他们提供定制化的推荐。

Netflix 发布的数据显示，用户在 Netflix 上每天产生 3000 多万个行为，如暂停、回放或快进；同时，用户每天还会给出 400 万个评分，发出 300 万次搜索请求。Netflix 遂决定用这些数据来制作一部电视剧，投资过亿美元制作出《纸牌屋》。

Netflix 发现，其用户中有很多人仍在点播 1991 年的 BBC 经典老片《纸牌屋》，这些观众中许多人喜欢大卫·芬奇，而且观众大多爱看奥斯卡奖得主凯文·史派西的电影。由此 Netflix 邀请大卫·芬奇作为导演，凯文·史派西作为主演，翻拍了《纸牌屋》这一政治题材剧。2013 年 2 月《纸牌屋》上线后，用户数增加了 300 万，达到了 2920 万。

(11) 航空大数据。Farecast 已经拥有惊人的约 2000 亿条飞行数据记录，用来推测当前网页上的机票价格是否合理。作为一种商品，同一架飞机上每个座位的价格本来不应该有差别，但实际上，价格却千差万别，其中缘由只有航空公司自己清楚。

Farecast 预测当前的机票价格在未来一段时间内会上涨还是下降。这个系统需要分析所有特定航线机票的销售价格，并确定票价与提前购买天数的关系。

① 1 英里=1.6039 千米。

② 1 加仑=3.785 升。



Forecast 票价预测的准确度已经高达 75%。使用 Forecast 票价预测工具购买机票的旅客，平均每张机票可节省 50 美元。

(12) 人体健康大数据。中医可以通过望闻问切发现人体内隐藏的一些慢性病，甚至看体质便可知晓一个人将来可能会出现什么症状。人体体征变化有一定规律，而慢性病发生前人体会有一些持续性异常。从理论上来说，如果大数据掌握了这样的异常情况，便可以进行慢性病预测。

结合智能硬件，慢性病的大数据预测变为可能。可穿戴设备和智能健康设备可帮助网络收集人体健康数据，如心率、体重、血脂、血糖、运动量、睡眠量等。如果这些数据足够精确且全面，并且有可以形成算法的慢性病预测模式，或许未来的设备就会提醒你的身体有罹患某种慢性病的风险。Kickstarter 上的 My Spiroo 便可收集哮喘病人的吐气数据来指导医生诊断其未来的病情趋势。

(13) 体育赛事大数据。在世界杯期间，谷歌、百度、微软和高盛等公司都推出了比赛结果预测平台。百度预测结果最为亮眼，预测全程 64 场比赛，准确率为 67%，进入淘汰赛后准确率为 94%。现在互联网公司取代章鱼保罗试水赛事预测，也意味着未来的体育赛事会被大数据预测所掌控。

谷歌世界杯预测基于 Opta Sports 的海量赛事数据来构建其最终的预测模型。百度则是搜索过去 5 年内全世界 987 支球队（含国家队和俱乐部队）的 3.7 万场比赛数据，同时与中国彩票网站乐彩网、欧洲必发指数数据供应商 SPdex 进行数据合作，导入博彩市场的预测数据，建立了一个囊括 199 972 名球员和 1.12 亿条数据的预测模型，并在此基础上进行结果预测。

从互联网公司的成功经验来看，只要有体育赛事历史数据，并且与指数公司进行合作，便可以进行其他赛事的预测，如欧洲冠军联赛、NBA 等赛事。

(14) 灾害大数据。气象预测是最典型的灾害预测。地震、洪涝、高温、暴雨这些自然灾害如果可以利用大数据进行预测，便有助于减灾、防灾、救灾、赈灾。过去的数据收集方式存在着死角、成本高等问题，物联网时代可以借助廉价的传感器、摄像头和无线通信网络进行实时的数据监控收集，再利用大数据预测分析，做到更精准的自然灾害预测。

以气象卫星数据为例，虽然气象卫星是用来获取与气象要素相关的各类信息的，然而在森林草场火灾、船舶航道浮冰分布等方面，气象卫星也能发挥出跨行业的实时监测服务价值。气象卫星、天气雷达等非常规遥感遥测数据中包含的信息十分丰富，有可能挖掘出新的应用价值，从而拓展气象行业新的业务领域和服务范围。例如，可以利用气象大数据为农业生产服务。美国硅谷有家专门从事气候数据分析处理的公司，它从美国气象局等数据库中获得数十年来的天气数据，然后将各地降雨、气温、土壤状况与历年农作物产量的相关度做成精密图表，可预测各地农场来年产量和适宜种植品种，同时向农户提供个性化保险服务。气象大数据应用还可在林业、海洋、气象灾害等方面拓展新的业务领域。

(15) 环境变迁大数据。大数据除可以进行短时间微观的天气、灾害预测，还可以进行长期和宏观的环境与生态变迁预测。森林和农田面积缩小、野生动植物濒危、海岸线上升、温室效应等问题是地球面临的“慢性问题”。如果人类知道越多地球生态系统及天气形态变化数据，就越容易模拟未来环境的变迁，进而阻止不好的转变发生。大数据能帮



助人类收集、存储和挖掘更多的地球数据，并且提供预测的工具。

除了上面列举的十五个领域，大数据还可被应用于房地产预测、就业情况预测、高考分数线预测、选举结果预测、奥斯卡大奖预测、保险投保者风险评估、金融借贷者还款能力评估等方面，让人类具备可量化、有说服力、可验证的洞察未来的能力。

美国的维克托在《大数据时代》一书中提到：“未来，数据将会像土地、石油和资本一样，成为经济运行中的根本性资源。”

总之，未来的信息世界是三分技术、七分数据，得数据者得天下。

1.2 跨境电子商务数据分析概述

数据分析是一种思维方式和技能，是指通过数据化的工具、技术和适当的统计分析方法对收集来的大量资料进行分析，以最大化地发挥数据的作用，提取有用信息、形成结论的过程，即对数据进行详细研究和概括总结的过程。

1.2.1 跨境电子商务数据分析的基础知识

跨境电子商务数据分析的基础知识主要包括跨境电子商务数据分析的概念、意义与作用。

1. 跨境电子商务数据分析的概念

数据分析是指用适当的统计分析方法对收集的大量第一手资料进行分析，以最大化地利用数据资料，发挥数据资料的作用，提取有用的信息并形成结论，从而对数据加以详细研究和概括总结的过程。跨境电子商务数据分析就是跨境电子商务通过数据分析得到有助于跨境电子商务发展的相关资料。例如，全球速卖通的卖家通过数据分析，能将整个店铺的运营建立在科学分析的基础之上，对各种指标进行定性、定量分析，从而为决策者提供准确、科学的参考依据。

2. 跨境电子商务数据分析的意义

在实践中，数据分析可以帮助人们做出判断，以便采取适当的行动。数据分析的意义在于发现问题，并且找到产生问题的根源，最终通过切实可行的办法解决存在的问题；基于以往的数据分析，总结发展趋势，为营销决策提供数据支持。

3. 跨境电子商务数据分析的作用

跨境电子商务数据分析的作用有分享线上活动成效，考核相关人员绩效（key performance indicator, KPI），监控推广的投入产出（return on investment, ROI），发现客服、营销等方面的问题，预测市场未来趋势，帮助改进网站用户体验设计（user experience design, UED）。数据分析贯穿于产品的整个生命周期，从市场调研到售后服务



的各个过程，都需要适当运用数据分析，以提升产品策划的有效性。

1.2.2 跨境电子商务数据分析的原则

跨境电子商务数据分析具有科学性、系统性、针对性、实用性和趋势性五个原则，如图 1-3 所示。

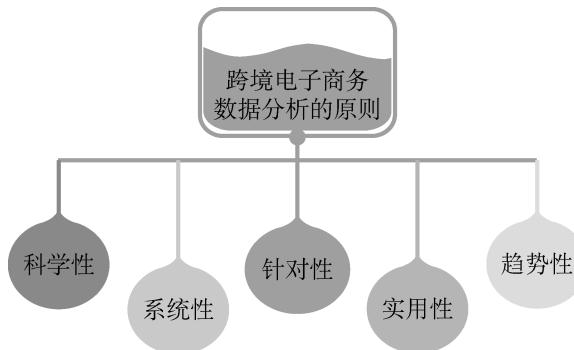


图 1-3 跨境电子商务数据分析的原则

1. 科学性

科学方法的显著特征是数据收集、分析和解释的客观性。数据统计分析要具有同其他科学方法一样的客观标准。

2. 系统性

数据分析不是单个资料的记录、整理或分析活动，而是一个周密策划、精心组织、科学实施，并由一系列工作环节、步骤、活动和成果组成的过程。一次完整的数据分析应包括分析目的与框架、数据收集、数据处理、数据分析、数据展现和报告撰写六个环节。

3. 针对性

不同的数据分析方法，无论是基础的分析方法，还是高级的分析方法，都会有它的适用领域和局限性。例如，行业宏观分析采用 PEST 模型（P——politics, E——economy, S——society, T——technology），用户行为分析采用 5W2H 模型（5W——what, why, who, when, where, 2H——how, how much），客户价值分析采用 RFM 模型（R——recency, F——frequency, M——monetary），销售推广分析常采用多维指标监测，等等。只有根据数据分析的目标选择合适的分析方法与模型，才能得到科学有效的结果。

4. 实用性

数据分析是为企业决策服务的，因此在保证其专业性和科学性的同时，也不能忽略其现实意义。在进行数据分析时，还应考虑分析指标的可解释性、报告的可读性、结论的指导意义与实用价值等。



5. 趋势性

市场所处的环境是不断变化的，在进行商务数据分析时，要用发展的眼光看待问题，不能局限于当前现状与滞后指标，要充分考虑社会宏观环境、市场变化与先行指标。

1.2.3 跨境电子商务数据分析的流程

跨境电子商务数据分析的流程主要包括收集数据、量化分析、提出方案和优化改进，如图 1-4 所示。

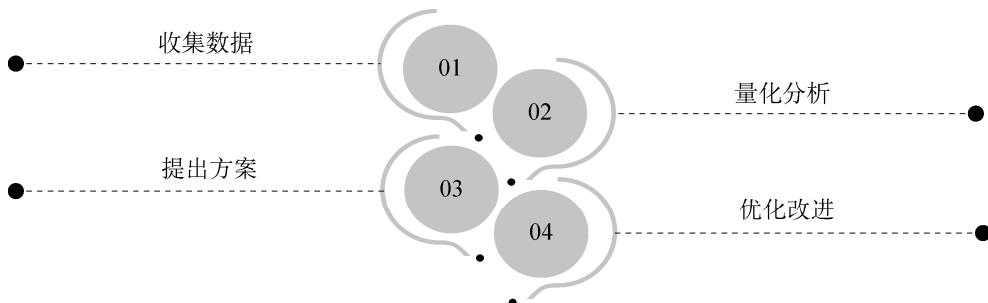


图 1-4 跨境电子商务数据分析的流程

1. 收集数据

在对网站数据进行分析之前，企业首先需要收集和获取数据，尽量获得完整、真实、准确的数据，做好数据的预处理工作，以便量化分析工作的开展。收集数据的类型及具体内容如表 1-2 所示。

表 1-2 收集数据的类型及具体内容

收集数据的类型	具体内容
网站后台的数据	网站用户数据（注册时间、用户性别、所属地域、来访次数、停留时间等） 订单数据（下单时间、订单数量、商品品类、订单金额、订购频次等） 反馈数据（客户评价、退换货、客户投诉等） 网站在各个搜索引擎中的收录量（site）
搜索引擎的数据	网站在搜索引擎中的更新频率 关键词在搜索引擎中的竞价排名情况 网站取得的搜索引擎信任的权重（谷歌的 PR 值、搜狗的 SR 值）等
统计工具的数据	网站的统计工具很多，这些工具基本上都会提供访客来自哪些地域、来自哪些网站、来自哪些搜索词、浏览了哪些页面等数据信息，并且网站还会根据需要进行广告跟踪

2. 量化分析

分析不是只对数据的简单统计和描述，还会在数据中发现问题的本质，然后针对确定



的主题进行归纳和总结。常用的分析方法有以下四种，如表 1-3 所示。

表 1-3 量化分析的方法及要点说明

量化分析的方法	要点说明
趋势分析	趋势分析是指将实际分析的结果与不同时期的报表中的同类指标的历史数据进行比较，从而确定变化趋势和变化规律。具体的分析方法包括定比和环比两种。定比是以某一时期的数据为基数，对其他各期的数据均与该基数进行比较；而环比是以上一时期的数据为基数，对下一时期的数据与上一时期的数据进行比较
对比分析	对比分析是指对两个相互联系的指标数据进行比较，从数量上展示和说明研究对象规模的大小、水平的高低、速度的快慢，以及各种关系是否协调。在对比分析中，选择合适的对比标准是十分关键的，标准合适，才能做出客观的评价，反之，可能得出错误的结论
关联分析	关联分析意为如果两个或多个事物之间存在一定的关联，那么其中一个事物就能够通过其他事物进行预测。它的目的是挖掘隐藏在数据间的相互关系
因果分析	因果分析是为了确定引起某一现象变化的原因的分析，主要解决“为什么”的问题。因果分析就是在研究对象的先行情况中，把引起它发生变化的原因的现象与其他非原因的现象区分开来，或者是在研究对象的后行情况中，把因它而产生的结果的现象与其他的现象区分开来

3. 提出方案

对数据量化分析的结果进行汇总、诊断，并提出最后的解决方案，如表 1-4 所示。

表 1-4 提出解决方案及要点说明

解决方案	要点说明
评估描述	对评估情况进行客观描述，用数据支撑自己的观点
编制统计图表	运用柱形图和条形图对基本情况进行更清晰的描述，运用散点图和折线图表现数据间的因果关系
提出观点	根据实际情况进行数据分析，提出自己的观点，预判发展趋势，给出具体的建议性的改进措施
演示文档	基于以上三点进行归纳总结，列出条目，制作一份详细的演示文档，将具体方案讲解给部门领导

4. 优化改进

根据改进措施的实施，及时了解数据相应的变化，不断进行优化和改进，不仅要治标，而且要治本，使同类的问题不再出现。运营人员要持续监控和反馈，不断寻找能从根本上解决问题的最优方案。

数据分析是一项长期工作，同时也是一个循序渐进的过程，需要运营人员实时监测网站运行情况，及时发现问题、分析问题并解决问题，这样才能使跨境电子商务健康、持续地发展。



1.2.4 跨境电子商务数据分析的价值

跨境电子商务数据分析的价值主要体现在以下三个方面，如图 1-5 所示。

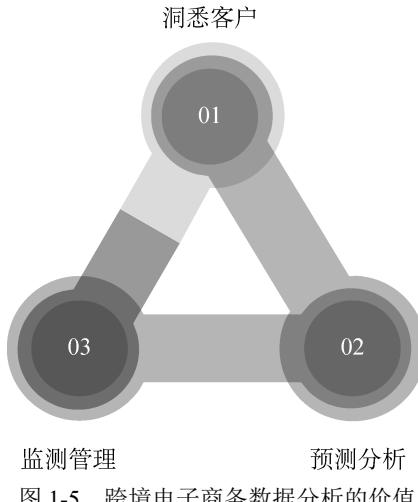


图 1-5 跨境电子商务数据分析的价值

1. 洞悉客户

跨境电子商务数据分析可以帮助跨境电子商务卖家更好地了解客户的需求和行为偏好；客户的来源渠道是什么？是新客户还是老客户？客户关注哪个产品？以此决定自己的营销投放渠道、策略和方向。例如，在分析客户的来源渠道时，通过分析客户是从 Instagram、Facebook 等外部平台来，还是从平台网站的搜索、导购频道来，这可以帮助跨境电子商务卖家调整营销投放策略，发现对客户更有吸引力和价值的渠道，调整渠道经营策略。

2. 预测分析

预测是通过一段时期的历史数据拟合出数据变化模型，据此测算未来的数据拟合值，这是数据分析的重要目标。对于跨境电子商务卖家而言，一方面，通过预测可以合理地制定企业的经营目标和发展规划；另一方面，通过预测能够及时发现运营过程中可能出现的问题，从而优化原有业务流程、提升和改进客户体验、制定预警和防范策略。此外，预测可以更合理地优化资源配置，达到实现效益最大化的目的。

3. 监测管理

数据分析可以帮助跨境电子商务卖家实时监控整个运营流程，方便快捷地把握各个关键运营节点的表现，适时制定更准确、更高效的执行策略。此外，通过社交平台的舆情监测，根据客户的评价数据，及时地进行产品改进，改善客户体验，打造良好的口碑，完善社交平台运营，帮助跨境电子商务卖家始终朝着正确的方向和目标前进。

数据分析对于跨境电子商务全业务链具有重要的价值。随着近年来跨境电子商务的迅



猛发展和竞争日趋白热化，消费者对于线上消费体验和产品服务的要求日益多元化和个性化，跨境电子商务卖家越来越意识到通过数据挖掘有效信息，能够支持精细化的客户管理和产品运营。



复习思考题

一、填空题

1. 大数据除了进行短时间微观的天气、灾害预测，还可以进行长期和宏观的环境和_____预测。
2. 数据分析的意义在于发现问题，并且找到产生问题的根源，最终通过切实可行的办法解决存在的问题；基于以往的数据分析，总结发展趋势，为营销决策提供_____。
3. 从数据化管理的流程来看，_____是数据化管理的核心。
4. 一次完整的数据分析应包括分析目的与框架、数据收集、数据处理、数据分析、数据展现和_____六个环节。
5. 预测是通过一段时期的历史数据拟合出数据变化模型，据此测算未来的数据拟合值，这是数据分析的_____。

二、判断题

1. “大规模”只是针对数据的量而言的，数据量大，代表数据一定有可被数据挖掘、深度学习算法利用的价值。（ ）
2. 全球信息存储能力大约每5年翻一番。（ ）
3. 现有的大数据处理工具大多是对开源的Hadoop平台进行改进并将其应用于各种场景。（ ）
4. 对比分析是指将实际分析的结果与不同时期的报表中的同类指标的历史数据进行比较，从而确定变化趋势和变化规律。（ ）
5. 数据分析可以帮助跨境电子商务卖家实时监控整个运营流程，方便快捷地把握各个关键运营节点的表现，适时制定更准确、更高效的执行策略。（ ）

三、简答题

1. 简述大数据的特征。
2. 简述跨境电子商务数据分析的原则。
3. 简述跨境电子商务数据分析的流程。
4. 简述量化分析的常用方法。
5. 简述跨境电子商务数据分析的价值。