

电子商务大数据分析综合实践平台建设

Construction of E-commerce Big Data Analysis and Comprehensive Practice Platform

周勇

Yong Zhou

中移(苏州)软件技术有限公司 SRE 部 / 安全部 中国·江苏 苏州 210000

SRE Department/Security Department of China Mobile (Suzhou) Software Technology Co., Ltd., Suzhou, Jiangsu, 210000, China

摘要: 现代社会发展态势下, 中国高度关注并推动“互联网+”计划落实, 有效利用大数据、云计算等新兴技术, 更好地满足电子商务事业发展需求。为此, 提出了电子商务大数据分析综合实践平台的构建计划, 这也是电子商务行业重要的发展方向, 借此平台可以培育更多的电子商务专业人才。论文基于当前电子商务实践平台存在的问题, 侧重分析了电子商务大数据分析综合实践平台建设具体方法, 希望可以更好地为电子商务行业发展提供借鉴和参考。

Abstract: Under the development trend of modern society, China pays great attention to and promotes the implementation of the “Internet +” plan, and makes effective use of emerging technologies such as big data and cloud computing to better meet the development needs of e-commerce. To this end, the construction plan of the comprehensive practice platform of e-commerce big data analysis is put forward, which is also an important development direction of the e-commerce industry, and this platform can cultivate more e-commerce professionals. Based on the problems existing in the current e-commerce practice platform, this paper focuses on analyzing the specific construction methods of e-commerce big data analysis and comprehensive practice platform, hoping to provide better reference for the development of e-commerce industry.

关键词: 大数据分析; 电子商务; 专业人才; 实践平台构建

Keywords: big data analysis; e-commerce; professional personnel; practice platform construction

DOI: 10.12346/etr.v3i11.4659

1 引言

电子商务行业的发展, 是信息化时代催生出的新兴行业, 尤其是云计算、大数据、物联网的出现和应用, 极大地推动了电子商务领域发展。而此背景下, 行业所需专业人才日益增加, 但是从当前电子商务行业实际情况看, 并没有优质的电子商务实践平台提供技术支撑, 严重影响了电子商务事业发展, 对此, 论文提出了基于电子商务数据的大数据分析综合实践平台构建方案, 从而为电子商务优质数据分析人才的培养提供实践平台保障。

2 当前电子商务实践平台存在的问题

电子商务作为一种新型活动方式, 很好的促使企业生产、采购、销售等环节组建成全新的信息化流动模式, 在互联网平台作用下, 让交易变得更加高效, 实现了实时交互功能, 搜集、分析数据, 对企业把控市场发展方向起到重要作用。但是从目前多数电子商务市场发展看, 很多行业内的企业缺乏对数据分析技术的应用, 在实践平台建设方面存在严重不足, 究其原因, 多半是由于资金不充足, 影响了实践平台

建设所需设备等硬件。除此之外, 很多企业即使买了硬件, 也不愿再花钱买软件, 在实践层面, 缺乏足够的实践性, 此种情况直接导致电子商务实践缺乏实质性, 严重降低了电子商务实践平台使用效果, 也阻碍了行业工作人员使用需求, 极大地抑制了电子商务行业的进一步发展^[1]。此外, 当前电子商务大数据分析平台中的内容, 只能满足部分工作需求, 无法做到深入分析、处理数据, 导致数据本身的价值没有被挖掘出来, 所以, 能够为电子商务企业提供的有价值信息数据非常有限。不仅如此, 电子商务由于其较强实践性的特性, 决定了人员需要具备一定的实践能力, 能够熟练掌握电子商务基本技能并可以应用于实践中。但从当前实际情况看, 虽然大部分从业者具备较高的学历, 但由于缺乏相应的实践经验, 很多时候无法灵活应对电子商务实践平台进行高效利用, 难以处理好数据, 直接影响了数据价值和作用的发挥。

3 电子商务大数据分析综合实践平台具体构建方法

3.1 选择大数据分析平台

考虑到大部分互联网初创公司资金问题, 在选择大数

【作者简介】周勇(1980-), 男, 中国江苏靖江人, 本科, 高级架构师, 从事软件架构设计研究。

据分析平台上,硬件方面主要应用 Hadoop YARN。由于 Hadoop YARN 框架属于分布式平台系统,所以,能够整合目前各类的硬件设备,这在平台构建中,就可以不用另外再购买新硬件。而在软件方面,因为 Hadoop 属于 Apache 社区开发的开源软件,这也就意味着,以其他软件在使用中,如果是依托 Hadoop 为框架,则可以免费使用。所以,在该平台下,所使用的软件无需再投入资金。

3.2 构建综合实践平台

3.2.1 基本模块构成

①对于数据采集模块的实现,平台主要就是利用网络爬虫等工具,直接获取相应的分析对象^[2]。②现阶段,社会发展已经呈现出了信息大爆炸趋势,这对数据存储提出了挑战,以往关系型数据库很难满足海量数据。因此,在数据存储模块中,主要就是利用分布式数据库,储存相应的数据,为日后使用提供便利。③采集后的数据并不全是完整的数据,非常容易存在缺失、异常等情况,所以,需要对数据实施清洗处理,主要的目的就是保证数据质量,从而为后续数据分析做相应的准备。④做好上述相关准备工作后,就需要对数据进行分析。只有做好数据分析和挖掘,才可以更好地为决策提供支持。在数据分析模块中,主要就是运用 Hadoop 框架下的机器学习算法,满足数据处理需求,探寻数据和数据之间潜在的关系。⑤分析后的数据,用户很难直接利用其进行分析、研究,所以应当以可视化方式呈现,如图片、表格等,只要可以满足用户使用需求的形式都可以,进而提升用户体验感,帮助用户提高决策速度。

综合分析这五个模块,不难发现,想要促进大数据与电子商务专业实现融合,构建大数据分析综合实践平台是非常关键的一步。整体看,现有实验室的数据还无法达到大数据的 5V 特征,但却已经具备了应用大数据的基础,所以,能够很好地与电商企业实际需求相符,将这一个平台应用到人才培养中,有助于拉近与企业实际工作的距离,提升学生实践能力。

3.2.2 模块具体分析

①采集数据。

网络爬虫技术在数据采集过程中的应用,主要包含了两种运行模式,即静态网页、动态网页,其中静态网页,主要就是利用网络爬虫工具采集数据,由 Python 编写而成,而针对动态网页,则主要使用 JavaScript 编写的工具实现。在这两种运行模式下,可以很好地保障爬取信息全面性。

②存储数据。

考虑到以往数据存储模式与当前数据需求不相适应,所以,在此次构建方案中,主要引入了分布式数据存储方式,将 NoSQL 数据库中的 HBase 作为载体,在这一模式下,可以很好地为大数据场景做铺垫。除此之外, NoSQL 数据库本身具有较强的优势,如内存数据库、列式数据库、横向扩展等,有助于进一步优化数据存储场景^[3]。

③清洗数据。

在数据清洗模块中,涉及前期准备、数据检测、质量评估、数据修正、数据输出几个部分。

第一,在前期准备过程中,会先初步分析对应的数据,明确具体的清洗目标,实现措施,而后整合成相应的数据清洗方案。第二,进入到数据检测阶段,这一环节主要就是预处理、检测数据,主要的作用就是消除当中的空数据、无效数据。并掌握冗余、缺失以及异常的数据。在此操作下,最终会得到高质量的数据信息。第三,对数据检测后所得结果实施相应的质量评估和定位,通过参照之前的清洗方案进行改进和调整,最后得出新的数据修正方案。第四,数据修正,主要就是利用不同方法修正检测后的数据。例如,运用排序、融合、基于规则等方法消除、删除不一致数据、空数据、重复数据^[4]。第五,在提交数据前,需要进行验证其合理性、真实性,主要依据就是之前的清洗方案、质量评估,如果符合要求,可以直接则输送给数据挖掘模块,若是不满足,则需要再次清洗数据,直到数据质量满足要求。

④分析数据。

在决策中,需要依靠数据分析为其提供基础,在此阶段,多个组件在 YARN 框架作用下,相互紧密结合,而后利用机器学习算法完成综合大数据分析流程,而后实施合理归纳、推理,找出数据关联性。其中,对于离线数据的处理,采用了 Mahout 中的协同过滤、关联分析、聚类算法以及分类算法等,分析相应的数据。而在实时数据处理过程中,为了更好地分析数据,主要采用 Spark 框架下 Streaming 与 MLlib 融合方式,有效提高了数据分析结果。

⑤可视化处理数据。

对于数据的可视化处理,在构建此平台过程中,应用了百度开源项 Fcharts 可视化工具,在 Fcharts 作用下,以图形的形式,绘制出了数据挖掘的结果,直观地为用户提供了数据间的关联性。

4 结语

总而言之,建立电子商务大数据分析综合实践平台,全面考虑到了电商企业需求,同时也促使大数据在现代社会发展中发挥了巨大应用价值,有助于进一步推动了电子商务事业发展。

参考文献

- [1] 李文强.基于Hadoop的电子商务大数据分析综合实训平台建设[J].现代工业经济和信息化,2018,8(11):46-47.
- [2] 齐伟东.基于传统BI系统的大数据分析平台建设[J].中国管理信息化,2017,11(21):61-63.
- [3] 何迪.第三方电子商务大数据分析平台的构建与应用研究[J].饮食科学,2017,14(45):194.
- [4] 张禹,金个,周志拓.基于大数据的电子商务专业网店运营实训指导平台建设[J].科学与财富,2020,12(9):293+290.