

# 基于大数据技术的个性化推荐系统分析

张云天 陈娜

(兰州工业学院 计算机与人工智能学院, 甘肃 兰州 730050)

**摘要:** 随着信息技术的快速发展, 互联网为人们的生产生活提供了诸多便利, 但也带来了一些问题。如何从庞大、繁杂的数据中获取有价值的信息, 是当前研究的重要。基于此, 本文首先阐述了大数据、个性化推荐系统的内涵及特征, 然后对基于大数据技术的个性化推荐系统进行分析, 以为相关研究提供一些帮助。

**关键词:** 大数据技术; 个性化; 推荐系统

**中图分类号:** F713.36; TP183 **文献标识码:** A **文章编号:** 1003-9767(2021)15-098-03

## Analysis of Personalized Recommendation System Based on Big Data Technology

ZHANG Yuntian, CHEN Na

(College of Computer Science & Artificial Intelligence, Lanzhou Institute of Technology, Lanzhou Gansu 730050, China)

**Abstract:** With the rapid development of information technology, the Internet provides many conveniences for people's production and life, but it also brings some problems. How to obtain valuable information from huge and complex data is an important topic of current research. Based on this, this article first expounds the connotation characteristics of big data and personalized recommendation system, and then conducts research and analysis on the personalized recommendation system under big data technology, hoping to provide some help for related research.

**Keywords:** big data technology; personalization; recommendation system

### 0 引言

随着信息技术的不断进步, 推陈出新的先进技术为人们的生产生活提供了诸多便利。推荐系统在此过程中发挥了十分重要的作用, 已广泛应用在电子商务、社交网络、音乐、电影等领域。推荐系统作为一种软件工具, 可为人们推荐各种感兴趣的商品、服务。但即便如此, 如今的推荐系统仍面临着预测准确率不足、冷启动、稀疏性等一系列问题<sup>[1]</sup>。传统的推荐系统已难以满足人们不断增长的需求。随着大数据技术的不断发展, 借助大数据技术的高准确性、强灵活性等优势, 能够有效解决推荐系统面临的诸多问题。因此, 本文将对基于大数据技术的个性化推荐系统进行研究。

### 1 大数据概述

大数据指无法在一定时间范围内用常规软件工具进行捕捉、管理和处理的数据集合, 是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率

和多样化的信息资产。一般来说, 大数据可以分为传感器数据、移动设备数据、网站点击流数据等类型。其中, 传感器数据分布在不同地理方位的传感器中, 可通过感知所处环境生成即时数据; 移动设备数据是指相关部门通过移动设备、导航等设施采集与用户群体行为、人员地理方位等相关的数据; 网站点击流数据是指用户群体在互联网开展有效操作被系统记录的数据<sup>[2]</sup>。

大数据主要有以下几个特征。第一, 数据规模庞大。随着大数据时代的到来, 海量的信息数据被存储在数据库中, 在需要时可对它们进行提取应用。数据是大数据时代的基础, 并来自不同领域, 无论是社交网络中的数据, 还是日常生活中的数据, 均可被存储在数据库中。相关调查结果表明, 2018年全球数据总量达到19.4 ZB, 并呈现不断上升趋势。第二, 数据类型多样。在大数据时代, 信息数据不再仅仅是文字类内容, 诸如图形、音视频等都属于大数据的范畴。第三, 处理速度快。在先进信息技术的支持下, 大数据可在短

**基金项目:** 2021年甘肃省高等学校创新基金项目“基于知识表示的个性化网络学习资源推荐算法研究”(项目编号: 2021B-311)。

**作者简介:** 张云天(1982—), 男, 甘肃兰州人, 本科, 助教。研究方向: 数据挖掘。

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

时间内完成对数据的处理。目前,网络信号存在于世界各地,基于网络信息的支持可实现对各类信息数据的实时传输,进而可显著提升数据传输效率。

## 2 个性化推荐系统概述

以电子商务个性化推荐系统为例,个性化推荐系统针对对不同消费群体的多元需求为他们提供针对性服务,或者电商企业依据消费者的上网习惯,为他们提供专业的服务,为广大消费者打造一个优质的购物平台<sup>[3]</sup>。个性化推荐系统具有以下几个特征。第一,推荐内容针对性。电子商务个性化推荐系统不仅可以满足广大消费者的多元需求,还可以开展信息交换。消费者可以自主向系统提出需求,也可以对服务进行评价。第二,推荐内容多样化。消费者可以结合自身需求设置信息内容获取的形式、时间,电商企业则可以基于消费者的实际需求,为其提供丰富的信息内容服务。用户可以通过电子邮件、音视频等途径接收信息内容。此外,用户基于个性化推荐系统即可摆脱时空束缚<sup>[4]</sup>。第三,推荐内容智能化。个性化推荐系统具有突出的智能化特征,可以结合用户的浏览轨迹进行实时刷新,为用户提供实时的个性化信息内容。

## 3 基于大数据技术的个性化推荐系统

### 3.1 个性化推荐系统的关键技术

个性化推荐系统的重点在于对数据的采集、处理以及分析,可以划分为数据源、数据采集、数据清洗、数据建模、个性化推荐算法、用户交互等模块。个性化推荐系统的运行流程如图1所示。

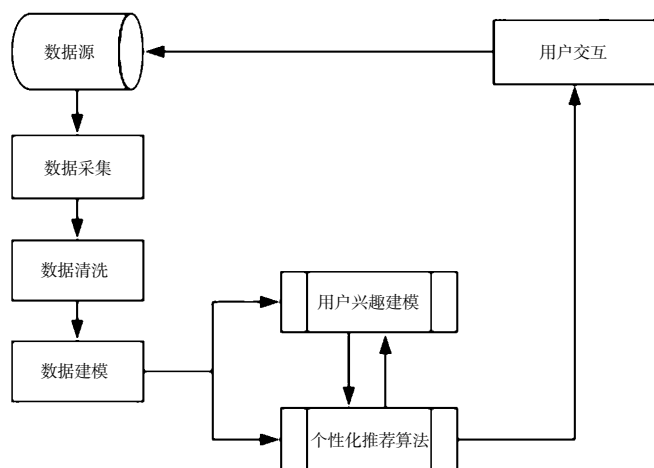


图1 个性化推荐系统的运行流程

(1) 数据源。不同企业的数据源不尽相同,如果企业拥有满足实际需求的各项数据,则无须向外界寻求数据源支持。但实际情况是,大多数企业掌握的数据十分有限,它们需要结合自身实际需求,通过各类数据库寻找所需的数据类型。数据库中存储着丰富的数据,既有结构化的交易数据、业务数据,又有传感器生成的半结构化数据等<sup>[5]</sup>。同时,数

据源也有很多,既可以是用户的特征属性,又可以是用户的操作行为,但不论是哪种数据源都应当与数据库的数据格式相符。在确定数据源之后,接着就要开展数据采集工作<sup>[6]</sup>。

(2) 数据采集。根据不同的分类标准可将数据分为多种类型,且各种数据所能实现的价值也不尽相同,由此要求企业在开展数据采集前,必须做好数据调查工作。在此基础上,企业在采集数据的过程中,应明确自身需求的数据种类、数据规模、数据价格等信息,并与数据库开展数据端口的规范对接,以实现对接需求数据的有效接收。

(3) 数据清洗。在数据采集结束后,企业应对数据进行清洗处理,以确保数据质量过关,找出数据存在的问题,避免出现数据错误、缺失等情况。在此过程中,可采用数据压缩、统一规格、修正逻辑等清洗方式。在数据清洗满意后,即可进入接下来的数据建模步骤<sup>[7]</sup>。

(4) 数据建模。系统设计人员在此环节应构建数据存放模型,依据相应规则或实际需求对相关数据开展优化整合,理清数据间的关系,为后续加工利用创造便利。在数据建模结束后,即可在所构建模型的基础上开展进一步的用户兴趣建模及制定个性化推荐算法。其中,用户兴趣建模是个性化推荐系统的核心环节,通过构建模型能够展现用户在不同阶段的兴趣,为接下来制定个性化推荐算法提供依据。设计人员基于用户兴趣模型可进一步优化个性化推荐算法,以确保可推测出用户的潜在偏好。

(5) 用户交互。在数据建模结束后,即可开展个性化推荐。个性化推荐系统可通过用户兴趣模型、个性化推荐算法开展相应的计算,获取用户可能感兴趣的信息内容,并进行推荐<sup>[8]</sup>。然后,用户的选择及兴趣偏好可再次被作为数据传输到数据库中,成为未来可能被利用的新型数据源。

### 3.2 个性化推荐系统的实践应用

随着大数据技术的不断进步,个性化推荐系统也获得不断发展,并应用在电子商务、新闻行业、广告行业等领域,众多企业纷纷开展了相关的探索研究。

(1) 电商推荐。20世纪90年代末,随着网络信息技术的飞速进步,电子商务蓬勃发展,并逐渐转变成网络信息技术的重要应用领域。电子商务凭借其高效、便捷的优势,迅速吸引一大批用户。中国互联网络信息中心发布的第47次《中国互联网络发展状况统计报告》显示,截至2020年12月,我国的网民规模达9.89亿人,网络购物用户规模达7.82亿人。庞大的用户群体不仅促进了电子商务的长足发展,还在各大网络平台上留下了海量的信息数据。与此同时,广大用户总是会迷失在繁杂的信息数据中,不得不投入大量的时间寻找自身需求的信息内容。针对这一问题,诸如淘宝、拼多多、京东等电商平台纷纷建立了自身平台的推荐系统,由此让个性化推荐系统在电商领域得到广泛推广。除去传统以协同过滤算法及其改进算法为基础的个性化推荐系统之外,相关研究人员还提出了进一步提升电商推荐系统体验的一系列新型

想法<sup>[9]</sup>。比如,一些研究人员将马斯洛需求层次理论与电子商务推荐系统相结合,建立了以电商特征为基础的用户体验需求层次模型,并通过回归分析采集影响用户体验的需求因素,进一步提炼影响推荐系统体验好坏的一系列因素。

(2) 新闻推荐。近年来,个性化推荐在新闻推荐领域也呈现出良好的应用前景。有别于电商推荐,新闻推荐不仅要解决数据量庞大、结构复杂的问题,还会面临用户兴趣变化的问题。针对这一问题,相关研究人员从用户位置标签入手,提出了以位置为影响因素的新闻推荐算法,也就是最初通过聚类算法对用户位置标签数据进行聚类,接着借助基于LDA算法的协同过滤技术,为各聚类位置构建兴趣模型,进一步实现新闻推荐的目标。另外,新闻推荐极易受各种因素影响,进一步会导致用户浏览行为不易被预测的问题。为此,有研究人员提出基于用户历史浏览日志中留下的上下文相关特征,来构建 Logistic 回归模型,以预测用户可能感兴趣的新闻。实践表明,用户浏览行为与上下文信息相关联,并且在预测模型中引入上下文数据可提升新闻推荐效果<sup>[10]</sup>。

#### 4 结 语

在大数据技术快速发展的背景下,个性化推荐系统在开发应用方面获得长足突破。该系统不仅拥有诸多强大、通用的功能,还可以为用户提供良好的体验,所以一经应用推广便受到用户的青睐。随着科学技术的不断发展及业务需求的逐步增加,对个性化推荐系统提出了越来越严格的要求,为

此在未来研究探讨中还需要不断对其进行创新、完善。

#### 参考文献

- [1] 乔岚. 基于大数据技术的个性化推荐系统的设计与实现[J]. 信息与电脑, 2017(21):70-72.
- [2] 石岩松. 基于大数据的个性化推荐系统研究[J]. 通讯世界, 2019, 26(4):81-82.
- [3] 陶庆凤. 基于 Apriori 算法的电子商务个性化推荐系统研究[J]. 重庆科技学院学报(自然科学版), 2020, 22(6):62-64.
- [4] 龙虎, 彭志勇. 大数据驱动的个性化服务智能推荐系统研究[J]. 信息与电脑, 2019, 13(10):140-142.
- [5] 罗雨馨, 姜攀. 基于大数据技术的商品精准推送系统设计与实现[J]. 电脑编程技巧与维护, 2019(12):83-87.
- [6] 王腾宇. 大数据时代下电子商务个性化推荐系统的分析与设计[J]. 智库时代, 2020, 2(8):132-133.
- [7] 梁袁媛. 基于大数据分析的音乐个性化推荐系统应用研究[J]. 艺术品鉴, 2019(7):153-154.
- [8] 张军. 基于大数据技术的有线电视节目推荐系统[J]. 电脑知识与技术, 2019, 15(25):303-304.
- [9] 唐晓凤. 基于大数据的 3D 虚拟学习环境个性化推荐系统的研究[J]. 数字通信世界, 2020, 185(5):102.
- [10] 龙虎, 彭志勇. 大数据驱动的个性化服务智能推荐系统研究[J]. 信息与电脑, 2019(10):140-142.