

# 大数据背景下媒体数据挖掘的应用

刘越,刘宁

(河北传媒学院,河北 石家庄 050000)

**摘要:**随着科技的进步,媒体产业也在发生着巨大的变化。在大数据作背景下,曾经主流的传统媒体随机受到新媒体的冲击,面临着技术、发展空间被挤压等诸多挑战。伴随信息技术改革应运而生的新媒体也在经历着技术多元化带来的洗礼。本文来分析不断更新中的媒体在大数据背景下如何运用好海量信息进行数据分析从中提取出有价值的信息。

**关键词:**大数据;媒体;数据挖掘;应用

自从人类诞生以来,信息技术经历了五次革命。随着互联网的高速发展,暨前面语言,文字,造纸术和印刷术,电话、电视、广播电报,计算机技术与现代通信技术的普及应用五次革命之后,我们迎来了多媒体信息和网络化高速发展的大数据时代。在如今的大数据背景下,媒体领域的更新速度之快的背后,隐含着的即是大量的信息。

## 1 大数据与媒体领域的更新及联系

无论是媒体还是大数据,在其发展过程中都各自受到相关性的影响,而两者之间的契合也是基于互相的容纳和运用。数据通常指信息的载体,把信息进行转化以便于保存和处理。信息技术的更新换代,带动了媒体的发展,信息量倍速增加的时候,形成了大数据模式。大数据是指无法在可承受的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。在维克托·迈尔·舍恩伯格编写的《大数据时代》中指出大数据的5V特点:Volume、Velocity、Variety、Value、Veracity,阐明其类型繁多,结构复杂,数量庞大的特性<sup>[1]</sup>。

时代的发展进步,某种程度上体现在电子科技产品的更新换代中,像第四次信息技术革命一样,同时硬件的发展带来推动社会进步的信息。新媒体就在这样的迭代中应运而生。相对于传统媒体而言,新媒体主要体现在三方面:时间上的新、技术上的新和观念上的新。在环境及技术互相推动进步的过程中,面对大数据这样的特性,信息产生的数量越来越多,速度也越来越快。在海量的数据中提取能用于媒体的一些知识规则,这本是一件门槛极高的事,在这之前的BI是办不到的,媒体及各大企业目前急需将自己保存的数据加以规范化的处理和提取。而现在统计学的发展和数据挖掘算法的兴起和快速更新让一项核心技术数据挖掘具备了客观条件和技術上的条件。

## 2 数据挖掘的理解

(1) 数据挖掘的定义。数据挖掘又称数据库中的知识发现(Knowledge Discover in Database, KDD),是目前人工智能和数据库领域研究的热点问题,指从相应数据库的海量数据中将潜在的未知的可以为人理解的并且较大可利用性的有价值的信息揭示并进行高等处理的流程,主要依靠计算机进行检索收集处理数据,进而发现新发展方向及新模式的非平凡过程。

(2) 数据挖掘技术的流程。数据挖掘的流程可根据不同项目的特点分为很多阶段,但是整体的步骤遵循数据挖掘标准流程中的商业理解、数据理解、数据准备、建立模型、模型评估、发布模型6个阶段,按顺序执行。首先是要进行了解用户需求的工作,挖掘师要了解项目的商业特点及用户需求;将商业信息进行拟定之后,对于数据有了明确的定义及功能性分类。进入到数据准备工作后确定项目进行中的有效数据,对于其他连带的部分如本次挖掘工作覆盖不到的数据要进行清理工作,形成在挖掘过程中很重要的一个数据即数据源。随即到下一个很关键的环节也是主要步骤,集合了智能化分析技术及提取技术,根据业务的特点选定对应有效的分析方法进行数据模型的建立。模型建立后,像很多程序在正式工作前的测试一样,

挖掘工作中的模型评估是检测可用性很重要的步骤,此时模型中真正涉及的数据及内容都是行业领域中用到的,由此对挖掘技术提炼的数据进行相应的评价。评估通过即可进行模型的发布。

## 3 数据挖掘在媒体行业的应用

媒体领域中,如今随着传统媒体和新媒体的融合进程加快,信息传递的边界正在消融,每个人都成了新的信息发声口。因此,及时获取用户关注信息对于媒体宣传来说显得尤为重要。当前,大数据挖掘已经成为媒体应对行业竞争,打破发展瓶颈的新工具。下面看数据挖掘在媒体领域几个场景中的应用:

(1) 关注数据运营。通过对观众观看时间、观看内容、停留时长等个人喜好类数据进行抓取,对于源数据进行挖掘分析,得出可行性策略能够帮助电视台或网络媒体等媒体平台更加了解用户群体,及对应关注的内容,然后进行精准的活动广告运营。

(2) 节目内容定位。通过对大数据的收集及分析可以从不同分类标准体现出的不同需求的角度,调整节目编排,提高收视率,以至于增加流量。数据处理的再生性过程,即可帮助平台根据当前社会、媒体等大环境的特点,开发出新类型的节目。

(3) 信息收集。大数据挖掘技术通过不同的算法不同的功能,可以实现从海量的信息中过滤提取出有价值的内容,进而根据受众的喜好来满足受众对信息的需求。其次在内容生产中利用大数据进行信息汇聚,从而使内容生产具有多样性。

## 4 结束语

大数据时代已经来临,它将在众多领域掀起变革的巨浪。我们要冷静的看到,大数据背景下,信息整合的核心在于为客户挖掘数据中蕴藏的价值,而不是软硬件的堆砌,随着移动互联网技术的普及和深入,互相推动的媒体领域也会迅速发展,媒介形态会不断推陈出新向更加精细化发展,数据挖掘技术会表现出得天独厚的优势及潜力,在媒体行业衡量成果的过程中的也会产生深远的影响。

## 参考文献

- [1] 英[维克托·迈尔·舍恩伯格·大数据时代[M].浙江人民出版社,2012.
- [2] 金京,周志飞.大数据挖掘中的自媒体应用[J].电子技术与软件工程,2017(21).
- [3] 华艳.基于新媒体信息的数据挖掘研究[J].电脑编程技巧与维护,2019(12).
- [4] 马泽鑫.数据挖掘在网络安全未知威胁检测中的应用价值[J].网络安全技术与应用,2020(3).

**作者简介:**刘越(1991,08-),女,汉族,籍贯:河北保定,单位:河北传媒学院,研究方向:新媒体数据分析,本科,助教;刘宁(1982,06-),男,汉族,籍贯:天津,单位:河北传媒学院,研究方向:新媒体与影视融合研究,研究生,副教授。