

# 军事科研范式转变与大数据技术支撑

张东江

(军事科学院 科研指导部, 北京 100091)

**摘 要:** 大数据是一类海量、高增长率和多样化的信息资产, 需要新处理模式对其进行处理才能提供更强的决策力、洞察力和流程优化能力。以数据采集、数据预处理、数据存储及管理、数据分析与挖掘、数据展示和应用等大数据相关技术为核心形成的数据挖掘模式, 可成为军事科研范式的第五种类型。借力大数据优化军事科研方式, 需要培育大数据思维和理念, 构建大数据环境下的信息资源保障体系, 构建科学、高效的大数据决策与应用平台, 增强大数据价值发现和数据洞察能力。

**关键词:** 大数据; 军事科研范式; 大数据技术服务

**中图分类号:** E917 **文献标志码:** A **文章编号:** 1672-8211(2015)03-0057-05

## 1 引言

“大数据”(Big Data)是继“云计算”之后 IT 行业最热的词汇之一。2012 年, 美国奥巴马政府发布了《大数据研究和发展倡议》, 宣布“将大力发展数字化接入、组织和挖掘的工具和技术”<sup>[1]</sup>。被 IT 行业誉为“大数据商业应用第一人”的舍恩伯格指出, “世界的本质是数据, 大数据将开启一次重大的时代转型。”这句话的意义同样适用于军事科研领域。在军事科研领域, 大数据应当助力实现从“拨云散雾查数据”到“腾云驾雾用数据”的转变, 从而带动军事科研范式的变革。

大数据可以而且应当与军事科研范式转变有必然联系。那么, 如何理解大数据时代? 如何借力大数据技术, 深入挖掘军事科研范式变革需求, 更好地为军事科研方式转变服务? 这需要我们深入思考, 认真研究, 提出应对之策。

## 2 世界大趋势: 大数据从理念到实践

大数据的概念最早可追溯到 20 世纪 80 年代。1980 年, 著名未来学家阿尔文·托夫勒在《第三次

浪潮》一书中曾做出预言: “如果说 IBM 的主机拉开了信息化革命的大幕, 那么大数据则是第三次浪潮的华彩乐章。”但是, 当时人们并没有给予托夫勒的大数据思想太多关注, 更没有多少人能“远眺”大数据在 30 年后会从理念变为现实。

### 2.1 大数据的内延与特征

大数据究竟是什么? 按照百度百科的定义, 大数据, 或称巨量资料, 指的是需要新处理模式才能发挥更强的决策力、洞察力和流程优化能力的数据集合。海量、高增长率和多样化的信息资产。“解构”是认知大数据内涵的最好方法。

#### 2.1.1 认知大数据的三个层面

第一是理论层面。理论是认知的必经途径, 也是被广泛认同和传播的基本方式。从理论层面认知大数据, 一是对大数据作整体描绘和定性分析; 二是探讨大数据的价值; 三是洞悉大数据发展趋势; 四是审视人与数据之间的互动关系。

第二是技术层面。技术是大数据存在与应用的基础, 离开技术大数据就毫无意义。从技术层面, 可以分别从云计算、分布式处理技术、存储技术和感知技术等方面来描述大数据的采集、处理、存储、形成结果等过程。

收稿日期: 2015-06-16

作者简介: 张东江(1958—), 男, 研究员, 博士, 博士生导师, 主要研究方向为军事管理与军事法制研究。

第三是实践层面。大数据的最终价值要通过实践来体现。所谓大数据实践,一般包括互联网的大数据、政府的大数据、企业的大数据和个人的大数据四个方面。

### 2.1.2 大数据的数据体量巨大

大数据有多大,至今没有权威说法。据美国政府估计,全球每年由各类设备产生的数据达到约1.2ZB(1ZB=12亿TB)。对信息化战争而言,大数据体量巨大的特征体现得尤为明显。例如,阿富汗战争期间,美军部署在太空、空中和地面的全方位情报侦察监视系统,24小时内所产生的数据就有53T。如此巨大的数据,常常会使指挥员和作战部队陷入数据的“海洋”中无所适从。例如,伊拉克战争爆发当日,美军驻卡塔尔和科威特前进指挥所由于无法处理各个相关机构提供的海量数据,不得不关闭设备,造成指挥所与部分突击方向一线部队通信联系中断。由此可见,大数据是远超出传统数据尺度的海量数据。这里,我们不需要给大数据之“大”定出一个具体“尺寸”,因为随着技术进步,这个“尺寸”还在不断增大。对于不同的领域,大数据“大”的定义也是不同的。

### 2.1.3 大数据的形态多样

大数据的“大”也体现在处理的数据类型多样化,远远超出传统数据格式和分析工具能处理的范畴。随着互联网、物联网、多媒体等技术的快速发展和普及,各种非结构化数据每年都以60%的速度增长。有专家预计,非结构化数据将占数据总量的80%以上。

### 2.1.4 大数据产生和处理的速度极快

大数据处理模式遵循“1秒定律”,可从各种类型的数据中快速获得高价值的信息。这一点与传统的数据挖掘有本质不同。大数据的数据流往往为高速实时数据流,而且需要快速、持续的实时处理,才能在第一时间抓住有价值信息。

## 2.2 大数据关键技术

大数据相关技术,就是从各种类型的数据中快速获得有价值信息的技术。大数据关键技术一般包括:数据采集、数据预处理、海量数据存储及管理、数据分析与挖掘、数据展示和应用(数据检索、数据可视化、数据应用、数据安全)等。

### 2.2.1 数据采集技术

数据采集技术一般分为两个层次:①数据智能感知层,主要包括数据传感体系、网络通信体系、传感适配体系、智能识别体系及软硬件资源接入系统,实现对结构化、半结构化、非结构化的海量数据的智能化识别、定位、跟踪、接入、传输、信号转换、监控、初步处理和管理等。②基础支撑层,主要是提供数据服务平台所需的虚拟服务器,结构化、半结构化及非结构化数据的数据库及物联网络资源等基础支撑环境。

### 2.2.2 数据预处理技术

这类技术主要用于对已接收数据进行辨析、抽取、清洗等操作。①抽取,因获取的数据可能具有多种结构和类型,数据抽取过程可将这些复杂的数据转化为单一的或者便于处理的结构,以达到快速分析处理的目的。②清洗,数据并不全是有价值的,有些数据并不是我们所关心的内容,还有些数据则是完全错误的干扰项,因此要对数据通过过滤去噪从而提取出有效数据。

### 2.2.3 大数据存储与管理技术

这类技术主要用于把采集到的数据存储起来,建立相应的数据库,并进行管理和调用,主要解决数据的可存储、可表示、可处理、可靠性及有效传输等几个关键问题。

### 2.2.4 数据分析挖掘技术

这类技术主要是从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中,提取隐含在其中人们事先不知道、但又是潜在有用的信息和知识。

大数据挖掘技术的开发,将在五个方面有所突破:①可视化分析。数据图像化可以让数据自己说话,让用户直接感受到结果。②数据挖掘算法。图像化是将机器语言翻译给人看,而数据挖掘就是机器的母语。分割、集群、孤立点分析等多种算法可以精炼数据,挖掘价值。③预测性分析。预测性分析可以让分析师根据图像化分析和数据挖掘的结果做出一些前瞻性判断。④语义引擎。语义引擎需要设计到具有足够人工智能以足以从数据中主动地提取信息。语言处理技术包括机器翻译、情感分析、舆情分析、智能输入、问答系统等。⑤数据质量和数据管理,通过标准化流程和机器对数据进行处理可以确保获得一个预设质量的分析结果。

### 3 大数据:应当为改变军事科研方式提供技术支持

所谓范式(paradigm)的概念,是美国著名科学哲学家托马斯·库恩提出并在《科学革命的结构》(The Structure of Scientific Revolutions)(1962)中系统阐述的。它指的是一个共同体成员所共享的信仰、价值、技术等集合,指常规科学所赖以运作的理论基础和实践规范,是从事某一科学的研究者群体所共同遵从的世界观和行为方式。所谓军事科研范式是指开展军事科学研究的基本方式。研究军事科研范式转变与大数据之间的关系,首先要了解现代军事科学的学科分类。

#### 3.1 现代军事科学的学科分类

现代军事科学的学科分类,是随着人们对战争和军事活动认识深化而不断发展的,同时还因各国军事科学研究传统、研究水平的不同而不同。就我国而言,在长期的革命战争和国防军队建设实践中,我们形成了具有中国特色的、以毛泽东军事思想为代表的现代军事科学体系,但对军事科学学科具体分法却没有统一定论。一般来说,有三种分法:一是军事科研的分法;二是军事教育的分法;三是军事百科编撰的分法。三种分法总体上一致,但在具体学科的分类上各有不同。本文主要针对军事科研的学科分类方法。

从近几年来我国军事学术界关于军事科学学科分类的研究情况来看,无论是主张哪一种分类法,在军事科学应包括军事思想、军事学术、军事技术等三个组成部分(三分法)这一点上是一致的。而持有四分法以上主张的学者,则认为武装力量建设理论、军事历史及军事地理等三类应作为军事科学体系中的独立学科门类,不应列入军事学术范畴。主要理由是,现行的三分法中的军事学术的内涵过于宽泛无边,而且随着军事科学分化与综合趋势的发展,新兴学科的不断涌现,军事学术的研究对象和内容如不科学界定,将会变得更加无所不包。还有的学者主张将军事社会学、军事管理学、军事心理学等学科也列入单独的军事学术门类。此外,还有的学者主张把研究战争史、军事史、军事思想史、军事学术史、军事技术史和军事文献编纂学等研究内容的军事历史,把研究国家、战区的兵

要地志、地理环境与气象条件等学科知识的军事地理,分别从军事学术学科门类分离出来,使其成为独立的学科门类。据此,军事科学理论体系应由军事思想、军事学术、武装力量建设理论、军事技术、军事历史、军事地理等六个学科组成。笔者认为,这种分法是比较完整、科学的,能够如实地反映我国现代军事科学发展的实际水平。

#### 3.2 军事科研范式面临转变

一般来说,军事科研范式有四种:第一种就是我们通常所使用的辩证思维的研究方式。毛泽东同志在军事名著《论持久战》中所运用的就是这种方式。第二种是实验科学范式,在战争开始之前通过反复的实兵对抗演习来论证和改进作战方案。第三种是理论科学范式,采用数学公式描述交战的过程,如经典的兰彻斯特方程。第四种是计算科学范式,基于计算机开发出模拟系统来模拟不同作战单元之间的交战场景。上述四种研究范式有两个共同的不足:一是它们都只能使军事科研人员感知交战的过程和结果,并未有效提高对海量数据的管理、存储和分析能力;二是四种范式之间相互借鉴和交融不够。

笔者认为,以大数据相关技术为核心的数据挖掘模式应被称为第五种范式。大数据研究范式利用软件处理各种传感器或模拟实验产生的大量数据,利用包括量子计算机在内的各种高性能计算机对海量信息进行挖掘,寻找隐藏在数据中的关联,从而发现未知规律,捕获有价值的情报信息,促使军事情报侦察和决策能力产生质的飞跃,更符合“未来战争不是大吃小,而是快吃慢”的制胜规律。可以预见,大数据必将推动战争形态的演变,基于大数据的实时、无人化作战,将彻底改变人类几千年来以歼灭有生力量为主要目的战争形态。军事科研从来都是为战争和军队建设服务的。大数据对军事科研范式转变的影响将是双重的:它既是新一轮军事革命的战略制高点,又对军事科研范式转变具有重要推动作用。

#### 3.3 大数据技术能够为军事科研提供的技术服务

大数据能够为军事科研转变提供至少以下两类支持和服务。

##### 3.3.1 按需提供的个性化服务

依据军事科研人员个性化服务需求进行服务

内容的定制、服务模式选择和服务实时推送,可以提升大数据阅读服务内容的个性化定制比例、增强读者知识获取效率和价值量的重要途径。首先,图书资料保障可多源采集结构化、半结构化和非结构化的军事科研人员大数据资源,并对数据进行科学的处理、转化和整合。同时,可对不同知识层次、不同性别、不同年龄、不同爱好的军事科研人员群体进行 360 度用户视图的读者特征分析,实现服务内容的个性化定制。其次,大数据时代的军事科研人员个性化服务,是由通信运营商、互联网信息服务商、大数据技术提供商和图书馆等众多 IT 运营与服务商共同提供的、基于大数据支持的个性化阅读增值服务。各运营商可通过大数据资源的共享实现服务的优势互补,为军事科研人员提供诸如高速移动阅读服务、地理位置查询、读者群体的片区化管理、信息实时查询链接等增值服务。

### 3.3.2 海量数据分析

19 世纪初叶,军事战略家克劳塞维茨以人的认知局限为由,提出了“战争迷雾”问题。未来依托大数据分析处理技术建构模型,通过数据挖掘模式,可以从海量数据中挖掘出有价值的信息,为其他四种军事科研范式提供支撑,从而拨开“战争迷雾”,达成运筹于帷幄之中,决胜于千里之外的作战目的。

## 4 借力大数据:优化军事科研方式提高军事科研效率

优化军事科研方式提高军事科研效率,是运用大数据转变科研范式的根本目的。而要达到这一目的,需要做好以下工作:

### 4.1 培育大数据思维和理念

大数据不仅是一种技术,也是一种价值观、方法论,它带来了一场思维和理念的大变革。大数据时代强调分析与某事物相关的全体数据,而不是少量的数据样本,接受事物的混杂性;不再探求难以捉摸的因果关系,转而关注事物的相关关系。以前,因为数据不足,图书资料保障很难为军事科研人员提供个性化的服务。而大数据带来了近似“上帝般”的全知,而且上帝身边还多了一个精于计算的会计,这势必减少图书资料为军事科研人员服务

时的不确定性。大数据构建了图书馆服务科研的一切要素,也必将培育我们用全新的数据化思维创新军事科研服务。这就要求图书馆提供军事科研服务时必须注重“用数据说话”,强调基于数据的精确分析和恰到好处的服务。这样的思维方式,将推动我们这个习惯于靠“差不多”运行的社会发生大变革。要充分意识到数据是科研的宝贵资源,高度重视对信息的收集、存储和处理,为开展军事科研奠定扎实的数据基础。

### 4.2 构建大数据环境下的信息资源保障体系

信息资源保障体系是指一个国家或地区通过信息资源整体化建设,建立能在一定范围内有效地满足社会在科学、文化、经济等方面信息需求的信息资源和服务系统。目前我国已经开始了全国性信息资源保障体系建设的探索,如中国高等教育文献保障系统(CALIS)的成功建设,形成了全国中心、地区中心、省中心、成员馆和数字图书馆基地的服务体系。对于军事科研信息保障而言,构建基于专题和产业的信息资源保障体系具有重要的意义,尤其是在大数据环境下,将多源数据充分地进行集成,将各种类型的信息资源进行有效整合,可形成信息资源建设与军事科研服务的良性互动。

### 4.3 构建科学、高效的大数据决策与应用平台

在军事科研资源的采集、存储、处理、应用和决策中,应摆脱传统 IT 环境下以数据小样本采样、苛求计算结果精确度、探究事物本质关系的传统做法。在保证所采集军事科研资源精确度满足大数据决策需求的前提下,实现对所有军事科研资源的挖掘、整合和价值提取。我们仅需要将大数据内部蕴涵的价值应用到军事科研服务与决策中,而不必过分强调军事科研资源之间的因果关系。因此,大数据决策与应用平台的构建,应在坚持安全、高效、智能和经济原则的前提下,重点加强大数据平台的网络传输带宽、存储设备容量与性能、平台管理与决策智能性和应用决策经济性等方面的建设,保证大数据平台决策所依据的信息具备较高的价值密度、可用性、信息对称性和可控性。此外,还应加强图书资料的大数据管理、大数据挖掘和深度分析专业队伍的建设,提高大数据平台在整合图书资料的多源异构以及跨域关联海量数据的能力,实现人才、大数据分析平台和用户服务系统的最优化结

合,从而能为军事科研提供更高标准的服务。

#### 4.4 增强大数据价值发现和数据洞察能力

首先,军事科研人员群体具有专业类型复杂、知识层次多样、年龄与性别差异大、阅读终端多样化和阅读周期不规律的特点,因此对军事科研人员群体进行大数据信息采集时,在加强对具有相同特征用户群的公共特征大数据信息采集的同时,还要兼顾具有相同特征用户群中不同个体的差异性,保证所采集的数据具有多样性和大价值属性。其次,对大数据资源使用前管理的科学性和效率,是关系数据可用和可控的关键。按照数据类型划分,大数据资源可分为服务系统配置与管理数据、用户行为数据、社会关系数据、用户位置和阅读模式数据、用户阅读评论与反馈数据等。通常情况下,这些数据以割裂、离散和孤立的状态呈现。应消除大数据库内部和不同服务系统之间的“数据孤岛”现象,实现大数据的自由流动和关联分析,努力挖掘其中蕴含的潜在数据价值,从而使军事科研人员能够了解数据之间的关联。再次,大数据的应用过程应坚持“信息—知识—智慧”的步骤原则,从大数据资源中发现数据规律和获取知识,提升大数据知识获取、预测、判断和决策能力,最终通过智慧决策和自动化管理过程实现图书资料保障的“智慧”服务,使得军事科研人员在最短时间内找到需要的数据。最后,大数据阅读活动是一个多模式和随机的过程,应细化所采集军事科研人员实时阅读行为数据的类型、次数、频率、时间、阅读习惯和评估等级,通过服务推荐系统、电子邮件、短信等方式,实现基于大数据的读者个性化精准服务。

## 5 结束语

一位西方军事家曾经说过,“胜利向那些能预见战争特性变化的人微笑,而不是向那些等待变化发生后才去适应的人微笑。”军事科研范式的变革与创新既有自发的动力,也有来自外界的推力。军事科学成果推动了大数据时代的到来,而大数据又为军事科研范式变革提供了新的手段与方法。目前,西方军事强国均把大数据视为潜在的巨大战斗力,都想抢占这一“数据红利”,从而推动大数据相关军事科研活动展开。我军应该积极顺应时代要求,把握历史机遇,借力大数据优化军事科研方式、提高军事科研效率,加快推进国防和军队信息化建设。

#### 参考文献

- [1] K TOM. Big date is a big deal[EB/OL]. [2014-03-21] <http://www.whitehouse.gov/blog/2012/03/29big-date-big-deal>.
- [2] 迈尔·舍恩伯格,库胡耶.大数据[M].盛杨燕,周涛,译.浙江:浙江人民出版社,2013.
- [3] 伊恩·艾瑞斯.大数据思维与决策[M].北京:人民邮电出版社,2014.
- [4] 埃里克·西格尔.大数据预测[M].周昕,译.北京:中信出版社,2014.
- [5] 马晓亭.基于可信大数据的图书馆个性化服务平台构建[J].图书馆理论与实践,2015(1):85-87.
- [6] 王晓燕,吴应清.数据挖掘技术分析及其应用评价[J].办公自动化,2007(6):52-54.
- [7] 高世光.基于模糊聚类的数据挖掘技术研究[D].长沙:国防科学技术大学,2002.
- [8] 陈源.数据挖掘在高校档案管理中的应用研究[J].办公室业务,2013(22):144-145.
- [9] 胡晓峰.大数据时代对建模仿真的挑战与思考[J].军事运筹与系统工程,2013,27(4):5-12.