

大数据背景下的军事情报转型

黄傲林, 黎铁冰, 毛盾, 汪莉莉

(海军工程大学电子工程学院信息对抗系)

摘要: 大数据时代的到来, 给军事情报研究带来了机遇和挑战, 文章首先介绍了大数据的定义和特点, 然后从军事情报发展的角度分析了军事情报大数据产生的背景, 总结了在此背景下情报研究的发展趋势和相关技术问题, 最后从人才培养和大数据情报系统的构建等方面对我军军事情报的转型问题进行了思考。

关键词: 军事情报; 情报研究; 大数据; 情报系统

0. 引言

随着传感器、移动终端、云计算、物联网等技术的发展, 数据正以前所未有的速度在不断增长和积累。美国互联网数据中心指出, 目前世界上 90% 以上的数据是最近几年才产生的, 并且每年仍以 59% 的速率增长, 人类社会已经进入了大数据时代。

学术界、工业界、政府机构都已经开始密切关注大数据问题, 并对其产生了浓厚兴趣。2012 年 3 月, 奥巴马政府发布了“大数据研究与开发计划”^[1], 并宣布先期将投资超过 2 亿美元的资金, 用于研发大数据关键技术, 以推进从大量的、复杂的数据集合中获取知识和洞见的能力, 抢占数据资源开发利用的制高点。2012 年 5 月, 联合国“全球脉动”计划发布了《大数据开发: 机遇与挑战》报告^[2], 英国、德国、法国、日本、加拿大等发达国家积极响应。我国也于 2012 年 8 月由中科院启动了“面向感知中国的新一代信息技术研究”战略性先导科研专项, 任务之一即是研制用于大数据采集、存储、处理、分析和挖掘的未来数据系统, 并于同年 10 月成立了中国通信学会大数据专家委员会。种种迹象表明, 世界各国特别是发达国家都把大数据的发展摆到国家战略层面加以推动, 使大数据正在成为世界新的战略资源争夺的一个新焦点。

1. 大数据的基本概念与产生背景

1.1 大数据的基本概念

大数据缺少一个标准且普及性的定义, 至少不像 NIST 对云的定义那样, 能被人们广泛接受, 调研公司 IDC 的定义^[3]可能比较容易被人们所接受。它对大数据的定义是: 一种新一代的技术和架构, 具备高效率的捕捉、发现和分析能力, 能够经济地从类型繁杂、数量庞大的数据中挖掘出色价值, 即认为大数据应满足 4 个特点: 规模性 (volume)、多样性 (variety)、高速性 (Velocity) 和价值性 (Value), 如表 1 所示。

表 1 大数据的 4V 特性

特性	表现
规模性 Volume	非结构化数据的超大规模和增长, 占总数据量的 80~90%; 比结构化数据增长快 10 倍到 50 倍, 是传统数据仓库的 10 倍到 50 倍, 集中储存/集中计算已经无法处理如此巨大的数据量。
多样性 Variety	数据种类繁多, 在编码方式、数据格式、应用特征等多个方面存在差异性, 多信息源并发形成大量的异构数据; 数据无模式或者模式不明显以及不连贯的语法或句义
高速性 Velocity	要求能够对海量数据进行实时分析, 从大量非结构化、低相关的数据中发现有用信息
价值性 Value	大量的不相关信息, 不经过处理则价值较低, 需要对数据趋势与模式做预测性分析, 方法包括: 机器学习、人工智能等等

大数据的出现将对数据和信息的分析提升到了前所未有的高度, 使得传统的基于服务器进行数据挖掘的方式面临巨大挑战。

1.2 军事情报大数据产生的背景

军事情报的数据产生方式大致经历了 3 个阶段, 如图 1 所示。第一阶段为管理信息系统阶段, 该阶段

的主要特点是军事情报数据库的出现和大量使用, 情报数据往往伴随特定的军事活动(如: 侦察、监听、探测等)而产生, 此阶段的数据产生方式是被动的。第二阶段为用网络共享阶段, 该阶段的出现主要是因为互联网、移动通信等技术在军事领域被广泛应用, 情报侦察部门可以随时随地接入网络并将获取的情报通过网络进行共享, 此阶段情报数据的产生方式是主动的。第三阶段为动态感知阶段, 该阶段的主要特点是随着传感器、云计算等技术的发展, 感知式系统被大量运用, 这些设备会源源不断地产生情报数据, 而且大多为非结构化的视频、音频等信息, 情报分析人员需要借助相关工具从海量数据中快速发掘出有用信息。

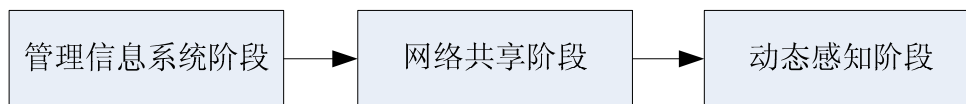


图1 军事情报数据产生的三个阶段

从军事情报数据产生的三个阶段可以看出军事情报大数据的产生是科技发展的产物, 是现代科技在军事情报学领域应用的结果。从技术角度来讲, 新型的硬件与数据中心、分布式计算、云计算、大容量数据存储与处理技术、社会化网络、移动终端设备、多样化的数据采集方式使海量数据的产生和记录成为可能。从情报学研究的对象来讲, 情报数据的来源日趋多元, 网络环境中敌人的日常行为很多都可以作为数据被记录, 人人都可能成为情报数据的提供方, 战场情报与非战争情报的界限正变得模糊^[4]。

2. 大数据分析方法

虽然大数据是一个重大问题, 但真正的问题是让大数据更有意义。传统的关系数据管理技术经过了将近 40 年的发展, 在扩展性方面遇到了巨大的障碍, 无法胜任大数据分析的任务; 而以 MapReduce 为代表的非关系数据管理和分析技术则以其良好的扩展性、容错性和大规模并行处理的优势, 在互联网信息搜索、数据分析等领域得到了广泛应用。

MapReduce^[5]是 2004 年由 Google 提出的面向大数据集处理的编程模型, 起初主要用作互联网数据的处理, 例如文档抓取、倒排索引的建立等。但由于其简单而强大的数据处理接口和对大规模并行执行、容错及负载均衡等实现细节的隐藏, 该技术一经推出便迅速在机器学习、数据挖掘、数据分析等领域得到广泛应用。

MapReduce 技术框架包含 3 个层面的内容: (1)分布式文件系统; (2)并行编程模型; (3) 并行执行引擎。分布式文件系统(Google file system)运行于大规模集群之上, 集群使用廉价的机器构建。MapReduce 将数据处理任务抽象为一系列的 Map(映射)-Reduce(化简)操作对, 如图 2 所示。Map 主要完成数据的过滤操作, Reduce 主要完成数据的聚集操作。输入输出数据均以 $\langle \text{key}, \text{value} \rangle$ 格式存储。首先将用户的原始数据源进行分块, 然后分别交给不同的 Map 任务区处理, Map 任务从输入中解析出 $\langle \text{key}, \text{value} \rangle$ 对集合, 然后对这些集合执行用户自定义的 Map 函数得到中间结果, 并将结果写入本地硬盘。Reduce 任务从硬盘上读取数据之后会根据 key 值进行排序, 将具有相同 key 值的组织在一起。最后用户自定义的 Reduce 函数作用于这些排好序的结果并输出最终结果。

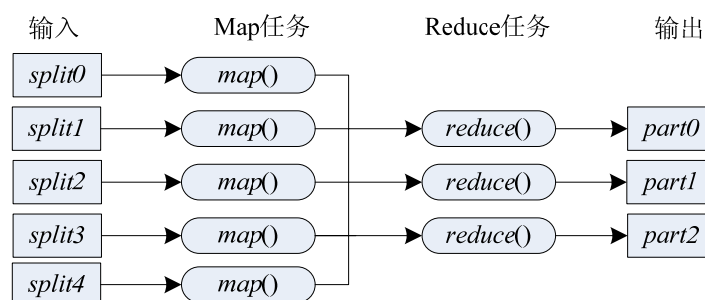


图2 MapReduce 并行计算过程

MapReduce 是面向由数千台中低端计算机组成的大规模机群而设计的，其扩展能力得益于其 shared-nothing 结构、各个节点间的松耦合性和较强的软件级容错能力：节点可以被任意地从机群中移除，而几乎不影响现有任务的执行。这是一种简洁的并行计算模型，它在系统层面解决了扩展性、容错性等问题，通过接受用户编写的 Map 函数和 Reduce 函数，自动地在可伸缩的大规模集群上并行执行，从而可以处理和分析大规模的数据。

3. 大数据促使军事情报转型

3.1 大数据时代的美军军事情报战略

2012 年 3 月 12 日，美国 Forbes 杂志网站发表一篇文章，题目是“Military Intelligence Redefined: Big Data in the Battlefield（重新定义军事情报：战场上的大数据）^[6]”，其中提到自从 911 事件以来美军通过无人机和其他信息采集工具收集到的数据增长了 16 倍，目前美军拥有的计算机已经达到 700 万台，而预期到 2020 年还将增加 1 倍，大数据将彻底改变美军以往的情报处理方式。

美军的情报转型是在美国军事转型全面展开的大背景下进行的，同时又是美国军事转型的重要内容和主要支柱。美军认为，未来战争是利用计算机信息网络，将分布在广阔区域内的各种探测、传感装置、指挥中心和各种武器系统高效集成，实现战场态势的信息共享和武器的高效使用。美国的情报体系非常庞大，拥有 16 个机构、20 万名员工。即使如此庞大的体系，也越来越难以及时处理无人机、侦查机器人等各种平台传送回来的海量信息。据悉美军的战斗机、轰炸机、坦克和运输机的数量在过去 10 年减少了 11%，但情报、监视和侦察（即 ISR）平台（主要是无人机）的数量却增加了近 3 倍，以阿富汗战场为例，美军的 ISR 采集系统每天产生的数据超过 53 TB，整合情报网络体系、提高情报信息处理能力的需求日益迫切。因此在美国政府 2012 年 3 月 29 日发布的 Big Data Across the Federal Government^[1]报告中，提到了大量与国防部相关的项目，下表是报告中提到的部分项目及其介绍。

项目名称	用途
Anomaly Detection at Multiple Scales (ADAMS)	从常规监测的海量数据中识别出个体的异常行为
Cyber-Insider Threat (CINDER)	开发创新的方法，以监测军事计算机网络中的网络间谍活动，并提高监测的精确度和速度
Insight	通过分析图像或非图像信息，来识别威胁网络和非常规军事行动威胁的资源管理系统，以弥补当前情报、监视与侦察系统存在的不足
Machine Reading	开发能够处理自然语言的学习系统，并将由此产生的语义表征插入到知识库中，实现人工智能应用以替代目前主要依靠专家和领域工程师的昂贵处理途径
Mind's Eye	为无人系统研发“可视智能”的能力，通过对系统采集到的流媒体数据进行分析，发现并及时报告重要的情报信息
Mission-oriented Resilient Clouds	构造一种抗攻击的云，用以检测、诊断和应对攻击，即使个别计算机宿主失效也无碍系统的运行
XDATA	利用开源技术与架构开发计算技术和软件工具，用于分析大量数据，包括半结构化数据和非结构化数据，建立有效的人机交互工具，为各种任务快速提供定制的视觉推理

3.2 大数据对我军军事情报建设的启示

大数据时代的到来迫使我们必须对传统的军事情报收集和处理方式进行变革，需要集成人工智能、数学算法、自然语言理解、信息技术等多个跨学科领域的技术成果，技术难度大，建设周期长。通过分析美军的军事情报战略，可以获得以下启示。

(1) 树立大数据的情报意识

尽管大数据的迅速增长及相关技术的出现正在给各个行业带来全新的发展机遇与广阔应用前景，但目前人们对大数据及其潜在价值的认识显然还不够。有相当一部分人不明白什么是大数据，还有一种认识认

为大数据只是一些诸如社交网络、电子商务平台等大型的互联网公司需要关注的问题，更有人认为大数据只是上述商家用来吸引眼球的噱头。

在军事领域，对大数据认识的不足，无疑会使军事情报工作落后于时代的发展。正如美国国家科学基金会发布的报告^[7]所说，一个国家在科学和工程领域的领先地位将越来越取决于利用数字化科学数据以及借助复杂的数据挖掘集成分析与可视化工具将其转换为信息和知识的能力。军事情报的大数据战略需要解决三个关键问题：一是情报、监视、侦察的大数据处理；二是从大数据中高效提炼出决策和执行人员所需的知识；三是强化基于大数据的信息实时融合。

(2) 加强具有大数据分析能力的情报人才培养

大数据能够转化为可操作情报的前提条件是具有大数据分析能力的人才，麦肯锡公司预测，到 2019 年，全球将缺少高达 19 万可处理大数据的科学家。目前我军在军事情报的大数据研究领域几乎空白，为了适应未来战争的需要，加快军事情报转型，必须培养一批高素质的军事情报大数据分析人才。

Thomas Fingar^[8]指出军事情报界需要综合利用人际情报、信号情报、图像情报和外部情报进行全资源分析，即利用多种不同的信息资源来评估、揭示、解释事物的发展，发现新知识或解决政策难题，从这段话中我们可以看出大数据下的军事情报人才培养应突出两点：

一是拓宽情报人员情报知识的广度。在大数据背景下情报研究对象既包含传统的战场雷达、通信、图像等信号，也可能包含微博等社会化媒体信息，大量非结构或半结构化数据的涌入使得情报研究问题更为综合，涉及要素更为多元，需要情报人员除了具备专业技能外还应具备心理学、社会学等知识。

二是加强情报人员数据分析的能力。传统的统计分析已不能满足现代军事情报活动决策的需求，依靠人力从海量信数据中发掘有用信息工程浩大且时效性差。因此必须借助于数据挖掘、机器学习等智能化技术自动进行高级、复杂的信息处理分析工作，把情报研究人员从繁琐的体力劳动中解放出来。

(3) 构建基于云计算的战场情报系统

美军一直将提高信息融合处理与分析能力，实现对各类海量信息数据的有效融合作为 ISR 领域建设的重点。相比之下我军的战场情报系统的发展还很滞后，主要体现在以下三个方面：一是情报来源渠道较少，和美军相比，我军的战场情报主要来自观通部门，战术无人机、无线传感器网络等由于起步较晚，目前大多停留在试验摸索阶段，不是情报的主要来源；二是情报信息总量有限，还达不到大数据的量级；三是缺乏海量信息存储、分析的平台。信息系统的信息处理速度、目标态势获取时间、决策周期以及快速响应时间决定着军事行动的成败。在“大数据”背景下，加快战场信息流转，缩短从“传感器到射手”的周期，实现“发现即摧毁”的作战目标，是信息系统建设必须解决的首要问题。可以想见，在不久的将来，随着我军 ISR 领域建设的快速进步，现有的情报中心无论在数据存储还是在信息融合、分析处理上都难以满足需求，因此迫切需要构建基于云计算的军事情报分析处理系统。

4. 总结与展望

未来影响、决定军事行动的最大核心在数据，数据的积累量、数据分析和处理能力、数据主导决策将是获得战场优势的关键。在军事情报领域，不断增长的数据搜集能力和有限的数据处理能力之间的鸿沟正在逐步扩大，面对大数据时代的来临，我们必须认清特点，把握走向，积极应对，高度重视其潜在战略价值，时刻关注其前沿技术，加快推进其实际应用，不断提高从数据到决策的能力，实现由数据优势向决策优势的转化，确保在新一轮信息化浪潮中赢得主动，占得先机。

参考文献：

- [1] Big Data Across the Federal Government [EB/OL]. [2012-10-02]. http://www.whitehouse.gov/sites/default/files/microsites/stp/big_data_fact_sheet_final_1.pdf
- [2] UN Global Pulse. Big Data for Development: Challenges & Opportunities [R/OL]. [2012-10-02]. <http://www.unglobalpulse.org/projects/BigDataforDevelopment>

- [3] Barwick H. The “four Vs” of Big Data. Implementing Information Infrastructure Symposium [EB/OL]. [2012-10-02]. http://www.coumputerworld.Com.au/article/396198/iiis_four_vs_big_data/
- [4] 孟小峰, 慈祥. 大数据管理: 概念、技术与挑战, 计算机研究与发展[J]. 50(1): 146-169, 2013
- [5] Dean J, Ghem aw at S . MapReduce: Simplified data processing on large clusters// Proceedings of the 6th Symposium on Operating System Design and Implementation (OSDI' 04) . SanFrancisco, California, USA, 2004: 137 – 150
- [6] <http://www.forbes.com/sites/teconomy/2012/03/12/military-intelligence-redefined-big-data-in-the-battlefield/>
- [7] NSF's Cyberinfrastructure Vision for 21st Century Discovery [EB/OL]. 2012-09-16. [http// www.nsf.gov/od/oci/ci_v5.pdf](http://www.nsf.gov/od/oci/ci_v5.pdf).
- [8] Thomas Fingar. A Guide to All-Source Analysis EB/OL . 2012-09-07. http://www.afio.com/publications/Fingar_All_Source_Analysis_in_AFIO_INTEL_WinterSprg2012.pdf.