

doi: 10.7690/bgzdh.2020.05.004

陆军演训数据分析系统建设

杜 伟, 张成举

(陆军指挥学院演训中心, 南京 210045)

摘要: 为解决当前陆军演训数据在积累、整合和运用等方面存在的问题, 对陆军演训数据分析系统建设进行研究。在分析演训数据建设现状的基础上, 针对演训数据来源、应用定位与流程等方面需求, 构建陆军演训数据分析系统, 对系统的功能模块进行设计, 并对系统研制中关键技术进行分析, 实现演习知识的不断积累。分析结果表明: 该研究能够为后续系统的设计开发提供参考依据, 对提高陆军军事训练信息化水平具有现实指导意义。

关键词: 演训; 数据分析; 系统建设; 功能需求

中图分类号: TP274 **文献标志码:** A

Army Exercise Training Data Analysis System Construction

Du Wei, Zhang Chengju

(Exercise Training Center, Army Command College, Nanjing 210045, China)

Abstract: In order to solve a series of current problems, such as army exercise training data accumulating, conformity and operation, research on army exercise-training data analysis system construction. Based on analyzing current situation of exercise-training data construction, aiming at requirements of source, application orientation and process in exercise-training, establish army exercise-training data analysis system, design system function module, analyze key technology in system researching, and realize exercise-training knowledge accumulating. Analysis results show that the research can provide the foundation for subsequent system design, and has practical guiding significance for improving the informationization level of army military training.

Keywords: exercise-training; data analysis; system construction; functional requirement

0 引言

当前信息条件下, 数据仓库、联机分析处理(online analytical processing, OLAP)、数据挖掘、知识图谱等数据管理与分析技术在商业领域、政务领域、机械制造等领域中已经被广泛应用^[1-3]。陆军演训数据是指陆军部队产生于或服务于各类演习训练以及训练考核活动, 能够依托人工或者自动化设备进行采集、传输、处理的各种数据。在陆军演训活动中, 通过对反映训练诸要素的特征、行为和结果等信息进行加工处理, 形成可度量的数据和数值, 构造陆军部队演训数据体系, 对于揭示陆军作战规律、指导训练实践具有重要的作用。

陆军演训数据分析系统的设计, 主要为解决当前建设陆军部队演训数据体系过程中信息化、智能化手段不足的问题, 在大量演训数据积累的基础上, 对数据进行科学的处理分析, 挖掘训练数据深层次信息和潜在价值, 为改进陆军演训质量、引领训练和战法创新发展提供支撑。

1 陆军演训数据建设现状分析

从美军对于训练数据运用的能力、机制以及方式来看, 我军在军事训练数据资源建设方面, 还存在数据积累难、整理难、运用难的现实问题。这些问题不仅造成了大量训练资源浪费, 而且严重制约了军事训练乃至战斗力建设向更高层次发展。

1) 在数据积累上, 由于缺乏顶层设计, 在不同类型的演训活动中, 采集数据种类、采集的权限划分不够明确, 有很多训练数据产生却难以正常采集并积累。不同系统之间, 由于数据标准和协议的差异, 难以实现自动化的采集, 人工采集效率和质量非常低。共享机制不够健全, 跨单位、跨部门进行数据的采集困难重重, 造成演训数据的共享和集成非常困难。

2) 在数据整合上, 当前进行的各类数据资源建设项目中, 普遍存在着重采集、轻处理的问题。大量的演训数据耗费了大量的人力、物力, 在管理过程中缺乏数据清洗, 难以进行有效的分析和运用。

收稿日期: 2020-01-09; 修回日期: 2020-02-23

基金项目: 国家社会科学基金(16GJ003-055)

作者简介: 杜 伟(1986—), 男, 山东人, 硕士, 助理研究员, 从事作战体系、军事训练研究。E-mail: duweithinker@163.com。

此外，受存储和计算能力等因素的限制，传统数据分析方法不能满足海量数据以及非格式化数据的分析要求^[4-5]，收集的各类文档、视频、音频数据基本处于原始状态。

3) 在数据运用上，由于存在数据积累难等客观问题，各类数据分析技术在军事领域的应用较少，主要集中于资源的管理和情报数据的整编融合^[6-9]方面。如何根据演训数据的特点构建相应的数据模型，对大量乃至海量的训练数据进行关联规则挖掘^[9]，分析各作战要素、训练要素之间的关联关系，提取有价值的作战和训练信息，是当前演训数据在分析和运用上的一个难点问题。

2 应用需求分析

通过构建陆军演训数据分析系统，提供高效的数据采集和分析手段，为不同用户提供科学的数据分析问题的解决方案。应用需求分析主要分析陆军演训数据的来源与格式，明确系统的用户、应用定位和应用流程。

2.1 演训数据的来源

1) 请领与下发的基础数据。主要包括人员、武器装备性能、地理信息等基础类数据。这些数据一般是从相应的数据库中导出，具有清晰而固定的格式，便于系统进行数据的描述、获取。

2) 人工制定的文档数据。各类训练和考核活动中，由组织人员制定的演练方案、导调文书、军事想定，演练中涉及各类人员编制和武器装备的参数数据，受训人员在演练中制定的作战构想、战斗决心、作战方案、作战计划，导调评估人员在演练中人工记录的各类文档以及最终的总结、讲评文档。

3) 信息系统运行与输出的数据。包括：受训者在运用指挥信息系统进行作战指挥过程中下达的指令、接收的情报；模拟训练系统(器材)用于模拟训练过程中的基础数据、交互数据、运行数据、结果数据、配置数据；评估分析系统中的指标数据，以及由人工录入、自动采集等方式形成的评估分析数据；考核系统中的基础数据和配置数据，考核过程中的运行数据以及成绩数据^[10]。

4) 演训活动中采集的图片、视频、音频数据。主要有各类训练活动中的组织人员、保障人员以及演习活动中导调评估人员运用摄像机、照相机、录音笔等人工采集的各类图片、视频和音频，布置在各类训练场所的监控设备自动采集的各类图片、视

频和音频。

2.2 系统应用定位

陆军演训数据分析系统，通常部署在陆军各训练基地以及院校的演训中心。系统主要用户为演训活动的组训人员，也可是军事理论研究人员和指挥员使用，分析对象是大量演训活动积累形成的数据。在系统的具体应用中，针对陆军部队在各类演训活动中产生的数据，以及组织活动所需的各类型基础数据进行集成、处理、分析，支持用于历史演训数据的查询，辅助进行演训活动的组织与实施。着重分析各类数据之间的关联分析，发现作战要素之间的影响关系并进行形式化的描述，在可能的条件下构建量化的影响关系模型，为提高组织质量和效益提供支撑。

2.3 系统应用流程

陆军演训数据分析系统应用流程如图 1 所示，包括以下步骤：

1) 对来自不同数据源、不同介质和格式的演训数据进行采集、抽取、转换、清洗、装载到数据仓库数据库，实现演训数据的集成。

2) 对各类集成数据建立索引，实施聚合、汇总等操作，按照不同的组织主题，将数据分布到不同的数据集市。

3) 分别运用不同的数据分析工具和数据挖掘工具，对数据仓库和数据集市中的数据进行分析和挖掘，生成包含各类信息的演训数据形式。

4) 运用展示工具对新生成的信息数据进行可视化展示，呈现在用户面前，便于用户理解和运用。

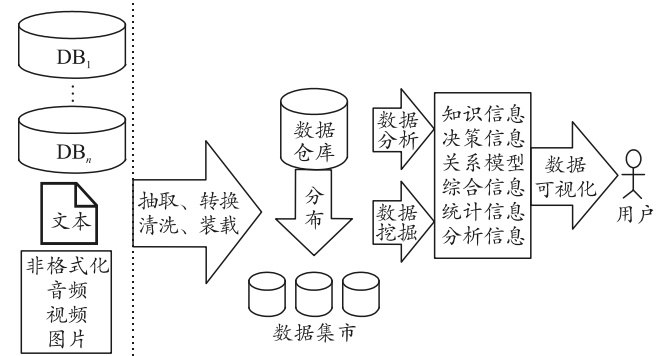


图 1 系统应用流程

3 系统架构与功能分析

系统体系架构如图 2 所示，陆军演训数据分析系统由数据层、支撑层和功能层以及相关的底层设施构成。

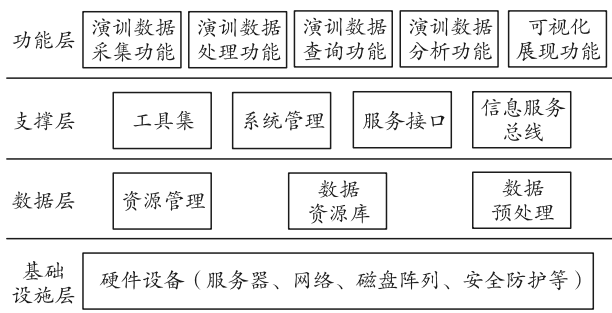


图2 系统体系架构

3.1 基础设施层

基础设施层包括服务器、网络硬件、磁盘阵列、硬件设备、计算设施等配套设施，为系统运行提供计算、传输、存储等底层的基础物理条件支撑。

3.2 数据层

数据层汇聚不同来源和格式的演训数据，对演训数据进行加工与处理构建演训数据资源，为演训数据的查询、分析与应用提供数据支撑和保障。数据预处理是依照格式规范，针对某些演训数据集，通过预先指定格式的方法进行预处理。数据资源库对系统涉及的采集数据、业务处理数据、经过分析得到知识数据以及支撑功能的各种模型和方法进行统一存储。数据资源管理是对构成数据资源库的各类数据进行审计、评估和数据质量分析，提高演训数据的及时性、准确性和完整性。

3.3 支撑层

支撑层各类模块主要为功能层提供服务总线、公共组件及工具、服务及接口等公共支撑服务，为演训数据的查询、关联分析、统计分析、数据挖掘等技术提供应用支撑。该层集成了各类型数据工具集，如 OLAP 工具、ETL (extract transform and load) 工具^[11]和大数据分析工具等。提供与其他应用系统、数据库管理系统的接口，为用户提供了系统配置管理、权限管理等服务。

3.4 功能层

功能层以数据层数据资源和基础实施层为基础，以支撑层为支点，为陆军演训数据分析系统的各级用户提供不同类型的功能：

1) 演训数据采集功能。

系统能够按照不同来源和格式的数据，包括规范化的文档数据、异构的数据库数据、非格式化的图片和音视频数据、系统生成的格式化数据等，采用固定模版、语义分析、自动转化、物理整合等机

制和方法，实现对各类型有效演训数据的抽取和归档。

2) 演训数据处理功能。

系统能运用集成的 ETL 工具，对采集的演训数据进行抽取、转换、清洗等操作，形成质量演训数据，并将其装载到目标数据库中。数据装载完成后，有些数据并非是军事研究人员所需要的数据。该模块能够对质量演训数据进行概念化、定量化的加工整理，得到符合军事概念的量化值。

3) 演训数据查询功能。

系统能够为用户提供高效、快速、准确的数据查询功能，通过对历年、历次演训活动数据以及相关基础数据的采集和处理，形成高质量的训练数据集，能够满足用户对于不同场景、不同演训内容、不同组训任务条件下对于演训数据的查询需求。在更高层次上，系统通过对用户基础和行为数据的分析，能够在一定程度上实现个性化、智能化的训练数据推荐。

4) 演训数据分析功能。

系统能够运用集成的数据分析、数据挖掘、大数据分析等工具，对演训数据进行统计描述、统计推断、相关分析、回归分析、维度分析、轨迹分析、趋势分析、关联分析、演化推理，并能够从大量的演训数据中发现、提取或“挖掘”出新的知识，包括关于作战、训练方面未知、潜在和有趣的概念、规则、模式、规律、约束等。

5) 可视化展示功能。

系统能够通过集成各类展示工具，运用静态与动态相结合的数据展现方式，将不同类型的信息和数据展现给用户，便于用户快速、准确理解数据中包含的信息和意义。静态可视化以多种图表形式展示，动态可视化支持用户对演训数据的动态属性进行链接、融合和显示，并可数据产品加载到二维态势图中，重构场景动态。

4 关键技术分析

陆军演训数据分析系统主要用于演习训练数据的分析、挖掘与评估，需要综合运用成熟的数据分析与处理技术，实现演训数据产品的不断丰富，最终实现演习知识的不断积累。在系统研制过程中，主要解决以下关键技术。

4.1 基于数据仓库的数据集成

演训数据采集管理过程中，数据来源复杂多样，

数据采集机制不同,实际使用过程中存在数据误差。基于演训数据的特点以及系统的应用定位,在系统建设过程中,采用数据仓库方法进行演训数据的集成。通过数据仓库对不同来源、类别、时间的演训数据进行有效集成,并按照人员、装备、演训活动等主题进行重组,实现较高的查询处理性能,便于用户进行分析和决策支持。通过构建元数据及应用数据归一模型,实现指定统一量纲、输入格式、必填项、数据保存验证等,约束数据输入,制定数据字典,使各类数据严格按照格式内容进行规范交互。

4.2 基于知识图谱的自然语言处理

对于想定文书、导调文书、配置方案以及受训者产生的作战方案、计划等,均以自然语言形成的文档形式存在。系统主要采用构建知识图谱进行自然语言的处理。演训领域的知识图谱构建主要包括数据采集、知识抽取、知识链接和融合、知识应用等过程。在知识数据采集阶段,对维基百科等来源提供的通用知识,OpenKG.CN 等网站提供的自然语言处理公开数据库,演训专家提供的问题领域数据进行收集、存储和查询。在知识抽取阶段,通过语言分词、词性标注、命名实体识别、句法分析等步骤实现实体-关系三元组的构建。在知识链接和融合阶段,根据三元组进行抽象层面的融合生成实体,学习推理,对关系进行处理并检测关系冲突。最后按照知识图谱网络对自然语言进行处理,抽取有效的演训数据。

4.3 面向演训数据生产的 ETL

在数据集成阶段,将演训数据通过 ETL 技术转化为高质量数据,是演训数据分析的重要基础。ETL 贯穿于数据生产流程,包括数据清洗、整合、转换、加载等过程。系统构建过程中,在运用 ETL 技术时,针对演训数据的特点,制定 ETL 转换策略,包括数据清洗策略、数据整合策略和数据转换策略等。构建灵活数据转换方法,保证 ETL 对所抽取的数据进行灵活的计算、合并、拆分等转换操作。通过分析数据产品的实际数据值特点,如数据类型值的精度、数据量,重点是数据采集的频率,具体到年月日的哪个粒度为分析基准,建立统一 ETL 数据标准。

4.4 面向演训数据分析的 OLAP 和数据挖掘

在系统构建中,数据分析关键技术主要是 OLAP 技术和数据挖掘技术。OLAP 技术主要通过多维的方式来对演训数据进行分析、查询以及生成

报表,通过对当前和历史的演训数据进行多维的切面和切块、钻取和旋转等操作,实现对于演训数据的时间序列分析、复杂查询、细化分析、随机查询等功能,辅助用户进行决策。数据挖掘技术主要针对演训数据的特征,综合运用统计学、聚类分析、模式识别、决策树分类技术、人工神经网络、遗传算法、贝叶斯网络等手段,按照理解业务、理解数据、准备数据、建模、模型评估、模型部署的步骤,进行演训数据的关联规则挖掘、聚类分析、分类预测、偏差分析等,挖掘出演训数据中隐含的高阶信息和知识。

5 结束语

笔者通过系统建设需求分析,指导系统开发人员理解和分析问题,对构建陆军演训数据资源体系、推动陆军军事训练的信息化水平具有一定的现实意义。由于数据资源建设是一个长期积累的过程,陆军演训数据分析系统的建设需求分析也是一个不断发展和完善的过程,在后续研究中,需要针对演训背景下的应用进行更深入的分析研究,尤其是对于海量演训数据的分析方法和技术手段的研究。

参考文献:

- [1] 施国良,刘芸,郑德俊.企业竞争对手产品分析的数据仓库模型研究[J].情报杂志,2012,31(5):7-11.
- [2] 袁丽娜.警务综合信息系统数据仓库的建设与实践[J].计算机科学,2012,39(6A):291-292.
- [3] 刘娇,李杨,段宏,等.知识图谱构建技术综述[J].计算机研究与发展,2016,53(3):582-600.
- [4] 曹冠平,王跃利,张立韬.关联规则挖掘在作战实验数据分析中的作用[J].指挥控制与仿真,2019,41(2):70-74.
- [5] 刘丙杰,杨继锋,冀海燕.基于大数据分析的潜射武器作战运用[J].兵工自动化,2019,38(3):1-5.
- [6] 尹承督,贾红丽,李卫东.大数据在部队信息资源管理中的应用[J].兵器装备工程学报,2016(8):118-121.
- [7] 张强,张宏军.基于作战数据分析的协同计划拟定及调整[J].火力与指挥控制,2014,39(5):6-9.
- [8] 唐志军,方晨,汪跃,等.基于数据仓库的情报数据分析系统设计[J].现代电子工程,2015,124(3):10-15.
- [9] 顾苗.关联规则挖掘算法及应用研究[D].南京:南京邮电大学,2015.
- [10] 赵倩,董冬梅,姜桂河.作战仿真数据的量化与分析[M].北京:国防工业出版社,2015:61-62.
- [11] 张宏军,郝文宁,陈刚,等.作战仿真数据工程[M].北京:国防工业出版社,2014:125-127.