

基于数据挖掘的装备在役考核评估研究

钱昭勇, 曹裕华, 燕如意, 刘现明, 史增凯

引用本文:

钱昭勇, 曹裕华, 燕如意, 等. 基于数据挖掘的装备在役考核评估研究[J]. [兵器装备工程学报](#), 2020, 41(8): 158-162.

您可能感兴趣的其他文章

1. [基于灰色关联度的指控分系统在役考核指标体系构建](#)

引用本文: 张超, 彭文成, 闫耀东, 等. 基于灰色关联度的指控分系统在役考核指标体系构建[J]. [兵器装备工程学报](#), 2018, 39(6): 94-98.

2. [一种基于装备画像的武器装备数据化方法](#)

引用本文: 刘虎, 刘卫东, 杨萍. 一种基于装备画像的武器装备数据化方法[J]. [兵器装备工程学报](#), 2016, 37(3): 59-62,81.

3. [单兵光电侦察装备作战适用性试验评估方法](#)

引用本文: 杜锋, 刘帅, 史圣兵, 等. 单兵光电侦察装备作战适用性试验评估方法[J]. [兵器装备工程学报](#), 2016, 37(5): 36-40.

4. [基于特征量参数漂移的装备可靠性评估方法](#)

引用本文: 张兴有, 刘学君. 基于特征量参数漂移的装备可靠性评估方法[J]. [兵器装备工程学报](#), 2020, 41(8): 112-114+181.

5. [一种武器装备虚拟训练平台开发方法及其应用](#)

引用本文: 胡松伟, 雷波. 一种武器装备虚拟训练平台开发方法及其应用[J]. [兵器装备工程学报](#), 2015, 36(1): 35-37.

基于数据挖掘的装备在役考核评估研究

钱昭勇¹, 曹裕华², 燕如意³, 刘现明¹, 史增凯¹

(1. 航天工程大学, 北京 101416; 2. 国防大学联合勤务学院, 北京 100858,
3. 中国人民解放军93110部队, 北京 100036)

摘要:通过装备在役考核评估,可以充分挖掘装备实际应用中的问题短板,总结各种环境下人装结合的特点规律,为武器装备的改造升级、实战化训练的战法改进等提供辅助决策支持。在把握数据挖掘和在役考核的内涵基础上,针对装备寿命周期的数据特征,介绍了数据挖掘的预处理、挖掘实施和结果解释3个阶段在装备在役考核中的应用形式,全流程分析了基于数据挖掘技术的装备在役考核评估程序,详细给出了数据挖掘在在役考核中选择应用工具与算法的思路,然后在在役考核指标体系模式评估和挖掘知识表示两个方面,重点分析了在役考核中数据挖掘的应用效果,最后总结了数据挖掘在装备在役考核应用中的优势和未来发展趋势。

关键词:数据挖掘;装备;在役考核;算法;知识表示

本文引用格式:钱昭勇,曹裕华,燕如意,等.基于数据挖掘的装备在役考核评估研究[J].兵器装备工程学报,2020,41(08):158-162.

Citation format:QIAN Zhaoyong, CAO Yuhua, YAN Ruyi, et al. Research of Data Mining on Equipment Assessment in Service[J]. Journal of Ordnance Equipment Engineering, 2020, 41(08):158-162.

中图分类号:E92

文献标识码:A

文章编号:2096-2304(2020)08-0158-05

Research of Data Mining on Equipment Assessment in Service

QIAN Zhaoyong¹, CAO Yuhua², YAN Ruyi³, LIU Xianming¹, SHI Zengkai¹

(1. Space Engineering University, Beijing 101416, China;

2. Joint Service College, National Defence University, Beijing 100858, China;

3. The No. 93110th Troop of PLA, Beijing 100036, China)

Abstract: Through equipment assessment in service, the shortcomings in the practical applications of equipment can be discovered, and the characteristics of the combination of human and equipment in various environments can also be summarized. The transformation and upgrading of weapons and equipment, as well as the improvement of combat methods for combat training will be promoted to provide assistant decision-making. On the basis of expounding the connotation of data mining and equipment assessment in service, according to the data characteristics of equipment life cycle, the application forms of data mining preprocessing, mining implementation and result interpretation of the in-service assessment were introduced. The programs of equipment assessment in service based on data mining technology were analyzed step by step. The idea of selecting application tools and algorithms on the in-service assessment of data mining were expounded in detail. Then, from the evaluation of the in-service assessment index system model and the mining of knowledge representation, the application effect of data mining on in-service assessment was analyzed. Finally, advantages and future development trends of data mining used in the

收稿日期:2019-09-18;修回日期:2019-11-15

基金项目:军队一般课题(JY2019C213)

作者简介:钱昭勇(1987—),男,博士研究生,主要从事装备与装备技术研究。

application of in-service assessment were summarized.

Key words: data mining; equipment; assessment in service; algorithm; knowledge representation

当前实战化训练背景下,装备在役考核过程中产生和统计的数据量大面广、种类繁多、结构不一,对这些海量数据的深层次分析与处理需求日益凸显。然而,传统的数据分析处理方法对海量考核数据利用效率较低,很少揭示数据间的关联规则和相互关系,难以立足于历史和当前数据预测未来的发展趋势,很难对装备使用情况进行客观可信的评估。如何设计出科学、合理、好用的装备在役考核指标体系,以便从大量杂乱无序并受噪声干扰的各种装备数据中“智能地”和“自动地”挖掘出有价值的信息,是目前我军装备信息化和智能化发展面临的瓶颈之一,也是对数据挖掘应用于在役考核的技术倒逼和挑战。为此,本文开展了数据挖掘方法在装备在役考核领域的应用研究。从数据挖掘的全过程和装备在役考核评估的全流程出发,首先介绍了装备在役考核与数据挖掘的内涵;继而对比分析了挖掘工具和算法应用;最后立足于挖掘结果的解释,重点剖析了考核指标体系构建模型评估和知识表示等相关问题。

1 数据挖掘与装备在役考核的基本内涵

数据挖掘是指从大量的、不完全的、有噪声的、随机的数据中,获取隐藏在其中的、有价值的潜在知识的过程^[1]。数据挖掘可分为预处理、挖掘实施和结果解释3个阶段,如图1所示。

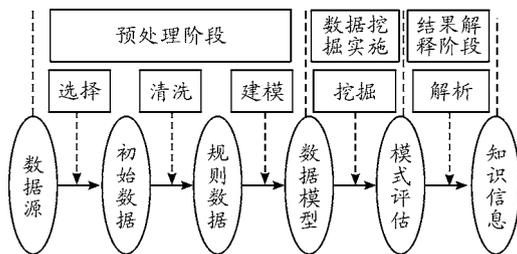


图1 数据挖掘过程框图

预处理阶段将高频数据从数据库中选取出来,进行转换后构建数据仓库,目的是为抽取足够且有效的数据;数据挖掘阶段通过选用有效的挖掘工具和算法,科学地运用挖掘模型对前期数据进行分析处理;结果解释阶段将依据模型的各项指标对比分析数据结果,并进行结果的可视化表示与知识的合理解释。

新装备在全寿命周期的试验可分为性能试验、作战试验和在役考核3类^[2]。性能试验是解决“能用”的问题,作战试验是解决“管用”的问题,在役考核是确保武器装备列装部队后“好用”。在役考核与美军后续的作战试验鉴定类似,主要立足部队实际条件,结合平时成建制、成体系的训练、演练,全时空、全领域、全要素地跟踪掌握部队装备使用、保障和维

修情况,既考核装备部署部队的适编性和服役期的经济性,又检测部分在性能试验和作战试验阶段难以考核的指标。装备在役考核的基本过程分为考核设计(编报在役考核大纲)、数据采集与存储、考核评估分析、问题反馈和处理等关键环节,如图2所示。

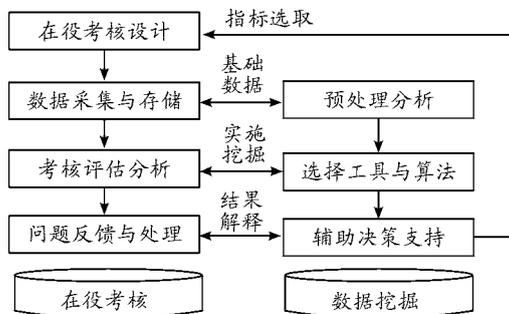


图2 数据挖掘介入装备在役考核过程框图

由图2可知,数据挖掘技术应用于装备在役考核过程中,将对考核流程里的每个关键节点发挥重要作用。依据装备鉴定定型试验总案,通过使用数据挖掘辅助设计合理的考核指标体系,对考核采集的基础数据进行预处理后,选择恰当的数据挖掘工具和算法进行考核评估分析,最后通过对挖掘模型评估和知识表示,总结规律,反馈问题,起到辅助决策的作用。

2 在役考核中数据挖掘的预处理

装备在役考核的数据挖掘预处理,相对于对传统统计数据预先处理,既重视数据的质,也注重数据的量。其本质就是通过一系列的方法和手段,确保装备在役考核数据的质和量,为实施数据挖掘奠定基础。

实施数据挖掘的首要前提是数据的量达到一定的程度,并且数据的维度需要尽可能完整全面。毕竟数据挖掘是使用局部样本来代替和反映整体,挖掘结果不可避免会出现局部样本无法完全正确反映整体而出现偏差。事实上,在役考核数据采集是一个长久持续的活动,绝大多数都不是短期的、一次性的,而是从装备列装服役到退役报废整个过程,涉及部队作战指挥、装备管理、维修保障和使用操作等所有与装备相关的要素,这就决定了考核指标采集数据的特点有:分布广泛、数据繁多、动态变化、类型多样等^[3]。为确保数据的量可以覆盖考核装备的全维度特性,可以把装备在役考核数据按照应用需求的方式总体分为连续型和离散型两类。连续型装备数据以时间为轴,主要统计考核装备从列装定型到退役报废所有相关性能参数变化情况,形成时间序列,进而分析装备的性能变化趋势,以及预测装备的剩余寿命等。离散型装备在役考核数据,更多立足于平时的演训,主要将

考核装备置于各种不同的应用场景,结合复杂环境的作用,丰富装备的应用模式,增强数据的层次范围,从而增加装备的评估维度,不仅仅统计演训中按照各种既定预案展开使用装备的情况,而是可以通过场景(实际或仿真)的加入使得每一次演训中装备的应用数据都能得到极大的扩展,从而真正有针对性地实现装备考核大数据的创造。

在役考核中装备数据的质量如何,是考核评估是否科学合理的关键,拥有较好质量的数据将较大提高挖掘的效率,取得更好的考核效果。为解决考核装备数据时间差异性、平台多样性、数据不稳定性等问题,消除噪声,纠正数据不一致性,不仅需要到装备在役考核数据严格执行 ETL (Extract-Transform-Load) 过程,结合考核目的和算法对数据削减和集成也是非常重要的预处理分析环节^[4]。数据削减是在不影响挖掘效果前提下,通过数据压缩、聚合、维数降减、数据块削减等方式,缩小数据规模。例如,卫星系统在役考核数据包括型号、批次、出厂日期、服役日期、组网数量、编制信息、功能性能、维修情况等具体指标信息,而在考核其作战效能的时候,传统的人机结合指标明显不符合也不适用于考核意图,为此与人机结合相关的指标属性应该舍弃,不予考虑。数据集成是指将来自多个不同数据源、格式性质各异的数据在逻辑上或物理上合并到一起构成一个完整的数据集,常使用联邦式、中间件和数据仓库等集成架构一个数据集成平台,该平台具有全面、统一、经济等特点,能更有效地进行数据挖掘操作和更快地做出数据驱动型业务决策。

3 在役考核中数据挖掘的应用分析

3.1 挖掘工具选用分析

针对装备在役考核数据特征,选取具有代表性的数据挖掘工具,分别从数据存取、处理、分析建模、可视化展示、编程语言、是否开源等方面对数据挖掘软件进行比较分析和评估,有关情况如表 1 所示。

表 1 数据挖掘工具能力

挖掘工具	数据存取	数据处理	分析建模	可视化	编程语言	是否开源
SPSS	良	优	良	优	Java	否
SAS	良	优	优	良	C/C++ 等	否
ODM (Oracle Data Mining)	优	良	优	良	SQL	否
Rattle	良	良	良	优	R	是
Rapid Miner	优	良	良	优	Java	是
Orange	良	良	优	优	C++/Python	是
Weka	良	良	优	良	Java	是
Knime	优	优	良	良	Java	是
Apache Mahout	良	优	良	—	Java	是

通过表 1 比较分析,装备在役考核中,须对各种不同类型的考核数据分类采集,使用 ODM、Rapid Miner、Knime 等工具进行数据存取更加快捷高效,便于维护;在数据处理方面,SPSS、SAS、Knime、Apache Mahout 等工具在对数据排序、转换、检索、传送等方面功能强大,兼容性好,容错性高;在分析建模方面,SAS、ODM、Orange、Weka 等工具拥有完整的集成模型,且满足差异化建模诉求,有较好的可扩展性;在可视化展示方面,SPSS、Rattle、Rapid Miner、Orange 等工具常用表格或图形的方式反映考核数据的特征和模式,甚至还可以按照用户自定的维度探索数据背后的信息,更加直观,增强了人机交互和知识感知的能力^[5]。

数据挖掘工具较多,大多继承了人工智能和模式识别等许多新技术的最新发展应用成果,把让数据“说话”的思想和相关算法进行了科学的平台整合。装备在役考核中,选择更加恰当的工具,实现海量数据的科学分析与评估,将为在役考核提供高效的处理方法和可信的决策支持。

3.2 挖掘算法应用分析

在役考核中,数据挖掘的核心在于合理科学地运用各种算法,基于装备产生的海量数据,在考核评估装备适用性、适配性、稳定性、作战效能、经济性、体系贡献率等方面发挥明显优势。

数据挖掘的主要任务在于描述和预测。描述性任务依据数据基本特征,按照相似程度进行分组,进而归纳总结数据所蕴含的规则;预测性任务立足历史和当前数据,通过学习和训练,对未来发展情况作出预判和推理。描述性任务通常使用聚类分析方法(包括 K-means 算法、K-medoids 算法、FCM 算法、神经网络聚类算法等);预测性任务通常包括关联分析(Apriori 算法、FP-growth 算法)、分类(KNN 算法、Ada-boost 算法、Bayes 分类算法、决策树算法、SVM 算法)、回归分析(线性/非线性回归、Logistic 算法)、演变分析(时间序列分析算法)等^[6]。装备在役考核应用过程中通常采用多种数据挖掘算法,或采用有效集成的技术,结合若干方法的优点。例如,在装备可靠性的在役考核方面,目前涉及装备的各类应用信息库、数据库管理系统中存储的数据量急剧增加。具体包括历年的装备统计实力、装备财务实力、装备训练数据、装备维修数据等数字与非数字数据库。无论从时间还是空间因素考虑,传统的数据分析手段(方差均值、专家打分、层次分析等)只能获得这些数据的表层信息,很难对数据进行深层次的处理,造成大量数据资源的浪费。然而,通过时间序列算法,就能充分利用装备从列装部队服役以来的各种数据,获得数据属性之间的内在关联和隐含的规则与模式^[7]。以对卫星装备的在役考核为例,其中一个重要的途径就是通过对在轨卫星的工作状态及健康状况保持持续监控,根据历史每天积累下来的各种类型遥测参数数据,进行相关性检验,通过对卫星这些海量异构数据的变化规律进行关联分析,判断哪些参数类型的变化与对应卫星分系统真正故障紧密相关;通过时间序列算法预测未来相关参数的演变趋势,可以掌握部件或分系统随时间的衰减情况,判断在轨卫星以

后发生故障的时间预测区间,为检验考核系统或部件质量稳定性提供重要参考依据;通过搭建使用分类算法为主的多层感知器模型,综合考虑卫星分系统或元部件的健康状态,为下一步卫星系统维护和系统级容错控制策略提供决策支持。不仅如此,卫星的各种组网运行控制方式,其中的过程整体行为远比构成这个整体的部分复杂,从某种角度也可以视为一种复杂的自适应系统,通过恰当运用神经网络、模糊C均值聚类、集成学习等各种挖掘算法,甚至可以评估和探索卫星体系的涌现现象,进而对卫星体系的整体考核评估提供理论参考。

根据反映影响装备战斗力发挥的关键问题或困扰装备建设的实际问题需求分类,在役考核的内容主要包括适用性、适编性、适配性、稳定性、作战效能、经济性、体系贡献率等方面。按照这些内容指标的分类,分析匹配更加合适的挖掘算法,构建起数据挖掘在武器装备在役考核中的应用,有关框图如图3所示。

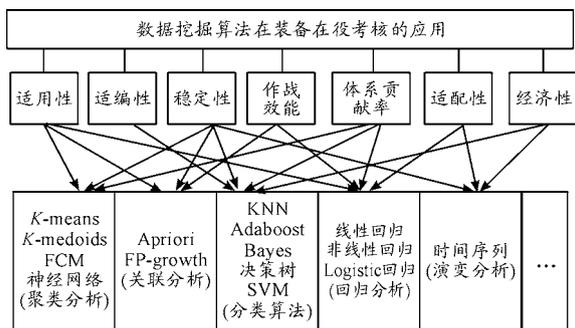


图3 数据挖掘算法在装备在役考核中的应用框图

4 在役考核中数据挖掘的结果解释

在役考核中数据挖掘的结果解释包括两个方面:模式评估和知识表示。模式评估是针对装备在役考核的目的(某种兴趣度度量),识别表示知识的真正有趣的模式;知识表示是使用可视化等知识表示技术,有效展示问题求解的结果。

4.1 装备在役考核指标体系设计模式评估

设计良好的在役考核指标体系是后续数据采集和分析处理的基本依据,也是数据分析挖掘的基础需求,直接决定在役考核的结果。

在役考核评估的本质是对装备全寿命周期产生的海量信息数据进行科学分析处理。采用传统的考核指标体系设计方法(大多仅仅考虑考核各阶段任务分工、完成的主要功能等)和简单的统计分析手段,只能获得数据表层信息,很难进行深层次的处理,而且不能获得数据属性之间的内在关联和隐含规则,导致“数据爆炸但知识贫乏”现象^[8]。将数据挖掘应用于在役考核评估指标设计,就是发挥数据挖掘分类、聚类、关联分析、异常侦测等算法功能,在基于统计学、系统工程、应用数学、图形图像等理论上,选择适合的工具

进行分析,找出能够更好反映考核整体效果的指标,如表2所示。

表2 在役考核指标体系设计模式的考核指标

考核指标	主要传统模式	数据挖掘模式
选择牵引	需求驱动	数据驱动
数据源处理形式	一般统计分析	海量数据处理
设计常用方法	专家打分、目标管理、平衡计分、关键指标	关联规则、分类、聚类分析
权重分配	专家打分、AHP、德尔菲法、灰色关联	神经网络、分层聚类、集成聚类
冗余度	较高	较低
预期效果	满足特定任务	在满足特定任务基础上挖掘探索内在规律与联系

由表2可知,在考核指标的数据源处理模式上,相对于一般统计分析,利用数据挖掘技术处理海量数据更加高效准确。在考核指标选择模式上,传统方法主要以当前的需求驱动,往往难以更新,而数据挖掘技术的核心在于数据驱动、建立模型、提取动态规则模式,可以随时更新“升级”指标库,随着数据仓库的不断更新,数据挖掘方法就可以及时发现数据库所最新隐含的规则、知识、模式,这也可以防止装备在役考核评估指标脱离实际。在考核指标设计方面,开发并使用关联规则、分类、聚类分析等算法优势,与传统的专家打分等方法相比,指标的冗余度较低,更具有客观性和可操作性。在考核指标权重分配上,通过神经网络、分层聚类、集成聚类赋权聚类的自优化功能,真正用数据“说话”,大大减少了人为的主观判断影响,避免了采用专家打分、AHP、德尔菲法等确定权重时主观性噪音、过拟合等缺点,确保了指标体系系统性和全局一致性^[9]。在预期效果上,传统的在役考核指标体系旨在满足特定的考核任务需求(例如,检验评估装备在役使用的适用性、可维修性、作战效能、经济性等),而通过数据挖掘方法还期待能探索发现装备生命周期里内在规律与联系,找到隐含的知识、规律和模式(例如,无人机集群或卫星组网里的涌现等),这一点也是数据挖掘技术在装备在役考核指标体系设计时特别突出的应用优势。

4.2 装备在役考核中数据挖掘知识表示

数据挖掘的结果往往是模式和规则,而新颖有效的、易于理解的、潜在有用的模式或规则就是知识。知识的表示方法决定了知识应用的形式,将直接影响决策者利用知识的程度和范围。为此,采取适当的知识表示方法,不仅可以提高装备在役考核决策分析的效率,而且可以更好地确保考核结果的准确性与完备性。

知识表示是知识的符号化和形式化的过程,表示方法有很多,主要包括语义网络、框架、谓词逻辑、产生式、本体等。

语义网络使用带属性的实体节点之间构成有向图网络

来传递和表达知识,可以清晰直观地表达事物间的关联关系。例如,在对卫星的质量稳定性考核评估过程中,使用关联规则挖掘的结果知识,通过语义网络解析表示为接收的异常数据集与卫星各分系统故障状态之间的关联关系,便于准确定位故障设备并加以及时处置。

框架表示是基于槽和侧面等形式来展现层次结构明确的知识,具有一定的推理与匹配功能。例如,卫星故障模式进行分类时,往往是根据已知的故障信息,通过与故障知识库中预先存储的框架进行匹配,以决定接受或放弃预选框架,便于形式化描述故障模式。

谓词逻辑将对象、特征、状态和关系通过逻辑公式进行表示,采用符号对确定事实或规则进行描述和推理。例如,在考核评估装备的适配性时,利用决策树算法得出的结果知识,可以通过一阶谓词逻辑的方式进行表示,使推理过程更加严密规范,实现模型精确表达。

产生式以“条件-结论”的方式揭示因果关系或启发性知识,不仅能表示确定性知识,而且能表示不确定性知识。例如,常使用四元组(对象,属性,值,可信度)表示不确定性知识,“卫星姿态控制分系统很可能存在2个故障”可以表示为(姿态控制,故障,2,0.8)。产生式知识表示结构相对固定,形式简单,已是人工智能中应用最多的一种知识表示方法^[10]。

本体表示是指共享概念模型明确的形式化规范说明^[11]。通过显式地描述概念和行为,本体支持知识的重用和互操作,其核心在于知识共享。数据挖掘结果使用本体描述,就是用概念对知识进行形式化表示,不仅可反映知识之间内在关系,还可以为装备在役考核工作小组提供一个统一的规范模型,使来自不同岗位、完成不同任务分工的人员之间都可以在语义一致的基础上进行交流。

当然,以上知识表示方法都有各自的特点,可以针对数据挖掘的内容和目的进行选择,用于结果知识的合理解释。但它们都属于知识的替代表示方法,不能直观地、可交互地对知识进行理解探索和优化应用。为此,需要结合知识的可视化进行处理。知识可视化是指所有可以用来建构和传递复杂见解的图解手段,就是将数据挖掘结果中的模式、关系、规则、趋势等用最常见的各种图表进行展示,具有直观化、结构化、深度化和高效化等特点^[12]。知识可视化立足于人类强大的视觉处理能力,通过创建知识形象视图,融入人机交互技术,使数据挖掘结果生动直观地呈现出来,大大增强决策者对知识的感知和判断能力,有效促进决策者对知识的获取吸收与整合转化。

5 结论

装备在役考核蕴含大量的数据信息,这些数据复杂多

样、种类繁多,而数据挖掘技术在应对海量数据采集、数据存储、数据安全、数据分析与处理等多方面具有独特的优势。因此,基于数据挖掘的装备在役考核评估方法具有很强的应用潜力和重要的军事价值。本文重点介绍了数据挖掘的预处理、挖掘实施和结果解释3个阶段在装备在役考核中的应用形式,全流程分析了基于数据挖掘技术的装备在役考核评估程序。数据挖掘技术能够引入更有效的评估指标体系,有效避免了传统方法带来的资源浪费、人为主观因素影响等不足。后续仍需从数据挖掘的流程出发,更深入地研究武器装备在役考核过程中的数据特征,选择适当挖掘工具,充分应用各类算法模型解决具体问题,使在役考核更加科学、高效。

参考文献:

- [1] HAN J W, KAMBER M, PEI J. Data Mining: Concepts and Techniques [M]. 3rd Edition, Morgan Kaufmann, 2012: 5-8.
- [2] 刘党辉. 卫星在役考核相关问题研究[J]. 国防科技, 2017(6): 49-50.
- [3] 李亚楠, 田雪颖, 王志梅, 等. 基于大数据的航天装备试验鉴定数据管理及分析应用研究[J]. 航天工业管理, 2018(10): 35-36.
- [4] 吴照林, 张胜. 对大数据资源开发利用的几点建议[J]. 军事通信学术, 2013(1): 89-90.
- [5] 刘智慧, 张林泉. 大数据技术研究综述[J]. 浙江大学学报(工学版), 2014, 48(6): 957-972.
- [6] 毛国君. 数据挖掘原理与算法[M]. 北京: 清华大学出版社, 2007.
- [7] 董林. 时空关联规则挖掘研究[D]. 武汉: 武汉大学, 2014.
- [8] 孟庆均, 曹玉坤, 张宏江, 等. 装备在役考核的内涵与工作方法[J]. 装甲兵工程学院学报, 2017(5): 19-20.
- [9] 张超, 彭文成, 闫耀东, 等. 基于灰色关联度的指控分系统在役考核指标体系构建[J]. 兵器装备工程学报, 2018, 39(06): 94-98.
- [10] 马雯雯, 裴道方, 林丹萍. 产生式规则在岸桥电动机选型中的应用[J]. 辽宁工程技术大学学报(自然科学版), 2015(34): 880.
- [11] 张弛, 彭丹华, 黄柯棣. 武器装备运用知识表示方法及其仿真教学应用[J]. 系统仿真实报, 2015(04): 42-50.
- [12] 马昱欣. 结合可视化与数据挖掘的数据分析方法探究[D]. 杭州: 浙江大学, 2017.

科学编辑 周钢 博士(海军工程大学讲师)

责任编辑 杨梅梅