

文章编号: 2095-2163(2023)07-0187-05

中图分类号: TP391

文献标志码: A

# 联通公有云平台下军校多源数据融合清洗技术

李弘扬, 史令飞, 张利龙

(海军工程大学, 武汉 430033)

**摘要:** 通过对军队院校多源管理数据融合清洗, 提高院校多源管理数据的动态调度能力, 提出基于联通公有云平台调度的军队院校多源数据融合清洗技术。构建以科研管理数据、教学管理数据、学生画像数据以及师资分析数据等为一体的大数据挖掘模型, 采用基于物联网模型的数字孪生技术建立智慧校园动态管理数据的可视化调度模型, 采用冗余数据滤波技术实现对多源管理数据的滤波处理和强化跟踪数据, 根据多传感信息感知和模糊 C 均值聚类处理, 实现对公有云平台下军队院校多源数据融合清洗。构建以多模态院校管理资源数据为核心的军队院校数字化信息处理平台, 测试结果表明, 采用该方法进行军队院校多源管理数据融合清洗提高了数据的动态存储和调度能力, 数据调度过程的抗干扰性较强, 数据采集和检索的查准率较高。

**关键词:** 公有云平台; 军队院校; 多源数据; 融合清洗; 模糊 C 均值聚类; 物联网

## Multi-source data fusion and cleaning technology in military academies under public cloud platform

LI Hongyang, SHI Lingfei, ZHANG Lilong

(Naval University of Engineering, Wuhan 430033, China)

**[Abstract]** Based on the fusion cleaning of multi-source management data in military academies, the dynamic scheduling ability of multi-source management data in military academies is improved, and the fusion cleaning technology of multi-source data in military academies based on Unicom public cloud platform scheduling is proposed. A big data mining model integrating scientific research management data, teaching management data, student portrait data and teacher analysis data is constructed. A visual scheduling model of dynamic management data of smart campus is established by using digital twinning technology based on the Internet of Things model. Redundant data filtering technology is used to filter and strengthen tracking data of multi-source management data. According to multi-sensor information perception and fuzzy C-means clustering processing, multi-source data fusion and cleaning of military academies under the public cloud platform are realized. A digital information processing platform of military academies is constructed with multi-modal college management resource data as the core. The test results show that this method for multi-source management data fusion cleaning of military academies improves the dynamic storage and scheduling ability of data, has strong anti-interference in data scheduling process, and has high precision in data collection and retrieval.

**[Key words]** public cloud platform; military academies; multisource data; fusion cleaning; fuzzy C-means clustering; Internet of Things

## 0 引言

随着 5G 和物联网技术的发展, 采用数字化云平台体系构造的方法进行军队院校的数字化管理, 构建智慧化的校园管理体系, 通过对智慧教学支持平台构建, 结合大数据分析技术, 实现对军队院校管理过程中的数字化信息调度。军队院校管理的运维

数据主要有科研管理数据、教学管理数据、学生画像数据以及师资分析数据等多源数据, 这些多源数据的种类分布主要分为教学管理类、学生管理类和科研管理类以及资产管理类等<sup>[1]</sup>。采用大数据的云融合技术, 结合对海量数据的采集、清洗和调度处理, 提高军队院校的数字化信息管理水平。在进行军队院校的多源管理数据处理中, 处理数据存在冗

**作者简介:** 李弘扬(1989-), 男, 硕士, 讲师, 主要研究方向: 数字化教学管理与大数据分析; 史令飞(1992-), 男, 硕士, 讲师, 主要研究方向: 信息处理、大数据分析技术; 张利龙(1983-), 男, 本科, 工程师, 主要研究方向: 数据清洗、数据挖掘。

**通讯作者:** 史令飞 Email: 1198301462@qq.com

**收稿日期:** 2023-01-08

余,需要采用数据清洗技术提高对数据分析的针对性,因此,研究军校多源数据融合清洗技术,对提高数据调度和共享能力方面具有重要意义<sup>[2]</sup>。

对数据清洗是建立在对数据的融合滤波和抗干扰设计基础上,构建滤波检测模型,通过动态参数融合和线性化调度,结合数据行为特征参数分析和提取,实现对信息化管理数据的清洗处理,传统方法中,对信息化管理数据的清洗技术主要有基于 protocol buffers 调度的数据缓存清洗技术、基于 PSO 粒子群调度的数据聚类清洗技术以及基于网格调度的数据清洗技术等<sup>[3]</sup>,通过对待清洗数据的非线性时间序列分析和包络幅值检测,通过线性预测方法实现对数据的动态清洗和优化调度,但传统方法对复杂平台体系下的数据清洗聚类性好。针对上述问题,本文提出基于联通公有云平台调度的军队院校多源数据融合清洗技术。首先构建军队院校的信息化管理数据的大数据挖掘模型,然后采用冗余数据滤波技术实现对多源管理数据的滤波处理和数据强化跟踪,结合优化的数据聚类算法,实现数据融合清洗,最后进行仿真测试分析,展示了本文方法在提高军队院校多源管理数据动态清洗融合能力方面的优越性。

## 1 平台总体构架

为了实现联通公有云平台下军校多源数据融合清洗,构建基于 CiteSpace V 的军队院校多源数据融合调度模型,构建军队院校多源数据动态融合清洗数据库,结合 AD 信息采样进行军队院校多源数据采样,在联通 CMP 多云管理平台体系下,通过联通机房控制,结合动态融合的分布式云数据清洗技术<sup>[4]</sup>,采用交叉编译和 PCI 总线协议进行军队院校多源数据融合清洗的平台体系构建,将军队院校的信息化管理系统分为管理资源数据库模块、网络模块、人机交互模块和接口控制模块等,通过对管理员的权限配置和动态参数分析,结合管理用户分配,建立底层数据库,信息管理平台的底层结构如图 1 所示。

根据图 1 的底层结构模型,采用嵌入式的 BS 总线调度下进行军队院校多源数据动态融合清洗的数据库体系结构设计,对军队院校多源数据融合清洗平台设计分为数据采集输入单元、数据库模型调度单元、数据集成加工单元、数据三维可视化处理单元以及人机交互输出单元等<sup>[5]</sup>,得到联通公有云平台下军校多源数据融合调度模型的实现结构如图 2 所示。

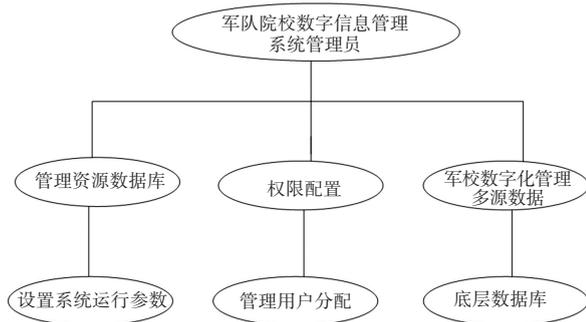


Fig. 1 Platform bottom structure model

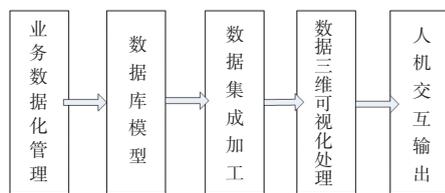


图 2 平台实现结构流程

Fig. 2 Platform implementation structure flow

通过 ADO.NET 组件库建立军队院校多源数据融合清洗平台的 3D 可视化处理模型,采用 ModelBuilder 3D 中间件技术进行军队院校多源数据动态融合过程中的私有云平台调度,在上述总体结构设计的基础上,进行数据清洗算法设计<sup>[6]</sup>。

## 2 多源数据挖掘和特征分析

### 2.1 军队校园多源分布大数据挖掘模型

构建以科研管理数据、教学管理数据、学生画像数据以及师资分析数据等为一体的大数据挖掘模型,根据对数据采集的活跃状态节点分布式部署,结合传感信息融合跟踪识别<sup>[6]</sup>,建立军队校园多源分布大数据分布的样本集,得到  $K$  个近邻样本值为:

$$P_{ij} = \sum_{d_i \in kNN} Sim(x, d_i) y(d_i, C_j) \quad (1)$$

其中,  $x$  表示多源数据调度节点汇聚链路增益;  $d_i$  表示科研管理数据的动态调度特征分布集;  $C_j$  表示联合自相关分配参数;  $Sim(x, d_i)$  为相似度;  $y(d_i, C_j)$  为校园全域感知的动态信息特征分配参数。基于可视化的数据特征重构技术,建立军队院校数字化信息管理的云平台处理模型,得到数字孪生的动态稀疏特征分配模型表示为:

$$d(t) = a(t)c(t) = \sum_{n=0}^{\infty} d_n g_c(t - nT_c) \quad (2)$$

其中,

$$d_n = \begin{cases} +1 & a_n = c_n \\ -1 & a_n \neq c_n \end{cases} (n-1)T_c \leq t \leq nT_c \quad (3)$$

采用线性化的均衡调度方案,实现军校多源数据融合过程中的边缘分布式检测,得到边缘特征分布检测模型为:

$$Z_n = \sum_{m=-\infty}^{\infty} |\operatorname{sgn}[x(m)] - \operatorname{sgn}[x(m-1)]| w(n-m) \quad (4)$$

其中,

$$\operatorname{sgn}[x] = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0 \end{cases} \quad (5)$$

$$w(n) = \begin{cases} \frac{1}{2N} & 0 \leq n \leq N-1 \\ 0 & \text{其他} \end{cases} \quad (6)$$

通过稀疏性特征重组,结合线性均衡调度和分块特征匹配,实现对军队校园多源分布大数据挖掘,根据数据挖掘结果进行样本特征点匹配<sup>[7]</sup>。

## 2.2 数据异构特征分析

在构建了大数据挖掘模型的基础上,结合对分配数据的联合关联相似度分析,采用模糊自相关特征匹配技术,建立军队校园多源管理数据时间序列的检测统计量,得到检测统计特征分配模型参数表示为:

$$x_n = a_0 + \sum_{i=1}^{M_{AR}} a_i x_{n-i} + \sum_{j=0}^{M_{MA}} b_j \eta_{n-j} \quad (7)$$

其中,  $a_0$  为数据融合调度过程中的特征匹配系数;  $x_{n-i}$  为任意某时刻点  $p$  处的语义相似度分配的条件概率密度;  $b_j$  为每个节点采集的军校信息化管理的匹配数据。采用回归分析方法,建立数据异构特征匹配模型,得到特征匹配函数为:

$$\max_{x_{a,b,d,p}} \sum_{a \in A} \sum_{b \in B} \sum_{d \in D} \sum_{p \in P} x_{a,b,d,p} V_p \quad (8)$$

$$\text{s.t.} \quad \sum_{a \in A} \sum_{d \in D} \sum_{p \in P} x_{a,b,d,p} R_p^{bw} \leq K_b^{bw}(S), b \in B \quad (9)$$

其中,  $x_{a,b,d,p}$  为第  $m$  类匹配数据的边缘联合特征分配特征向量;  $V_p$  为整合移动平均自回归参数;  $R_p^{bw}$  为军队院校多源管理数据融合的联合加权;  $K_b^{bw}(S)$  为大数据模糊隶属度;  $B$  为学习行为的多维特征分布集。根据上述分析,采用基于物联网模型的数字孪生技术建立智慧校园动态管理数据的可视化调度模型,采用冗余数据滤波技术实现对多源管理数据的滤波处理和强化跟踪数据。

## 3 数据融合清洗算法实现

### 3.1 数据冗余滤波和强化跟踪学习

采用模板匹配技术,对军队院校多源数据进行动态增益控制,构建双向链路管控模型,表示为  $(k,$

$i)$ , 提取军队院校多源数据融合相关性特征量,采用联通公有云平台调度方案,结合 3DStudio MAX 软件进行军队院校多源数据融合清洗的过程分配,得到清洗过程的过程分配特征子集:

$$\eta_{comm} = \frac{k_1 \cdot l}{E_{comm}} \cdot (1 - p_{drop}) \quad (10)$$

其中,  $p_{drop}$  为军队校园多源分布大数据清洗的冗余特征分量;  $k_1$  为自相关特征匹配系数;  $l$  为数据分布的长度;  $E_{comm}$  是动态约束参数。在重连通图中,通过数据融合聚类处理,结合聚类中心的强化跟踪学习,实现对数据冗余滤波<sup>[8]</sup>。

### 3.2 数据模糊 C 均值聚类及清洗

采用无线传感器网络的共享调度的方法,建立军队校园多源管理数据匹配模型,通过模糊 C 均值聚类进行节点参数动态调节<sup>[9]</sup>,得到模糊 C 均值聚类中心分别为  $p$  和  $q$ ,链路  $l$  的联合特征匹配模型参数通过频率间隔指数  $T_{L \times 1}(l < L)$  调节,数据清洗的动态调节模型表示如下:

$$T_{l1} = \sqrt{F_{p1}^2 + F_{q1}^2} \quad (11)$$

其中,  $F_{p1}$  和  $F_{q1}$  为军队院校多源数据分布节点的能量匹配系数。在  $N$  个簇首节点中,通过大数据融合、预测和自适应处理,得到冗余滤波系数为  $p \in [0, p_1, p_2, \dots, p_{max}]$ 。根据计算数据的状态差异度函数,得到匹配周期矩阵  $\mathbf{P}_{N \times 1}$ ,数据清洗输出为:

$$\Gamma_i = G \frac{\lambda_i}{(n-i)\lambda_i + \sigma^2} \quad i = 1, 2, \dots, n \quad (12)$$

其中,  $i$  为数据清洗的冗余参数配置节点;  $G$  为连通图;  $\lambda_i$  为冗余数据的分布网格单元;  $\sigma$  为协方差。基于冲突调节机制,采用模糊 C 均值聚类,得到数据特征参数提取和信息融合输出为:

$$p_i = \frac{\sigma^2}{1 - \sum_{j=k+1}^n \frac{\gamma_j}{\gamma_j + G}} \frac{\gamma_i}{h_i(\gamma_i + G)} \quad (13)$$

其中,  $\sigma$  为 sink 节点周围的统计分析数据;  $G$  为模糊 C 均值聚类簇;  $h_i$  为最大跳数;  $\gamma_i = \gamma_{ih}$  为阈值。根据上述算法设计,实现对数据模糊 C 均值聚类及清洗,提高数据的调度和检索能力<sup>[10]</sup>。

## 4 仿真实验测试

为了验证本文方法在实现军队院校多源管理数据动态融合清洗的性能,数据采样的动态分布覆盖范围为  $240 \times 120$  个网格节点,数据采样的频谱分布特征量为  $P_e = 0.12$ ;种子节点集分布概率  $P_d = 0.09$ ,云平

台虚拟机初始能量为 90 KJ,吞吐量为 23.4 B/S/HZ,多源数据管理平台任务分配参数分布见表 1。

表 1 多源数据管理平台任务分配参数

Tab. 1 Task allocation parameters of multi-source data management platform

平台	内存缓冲能力/MB	检索任务队列长度/Kbps
平台 1	408.667	784.207
平台 2	851.481	462.762
平台 3	545.359	452.850
平台 4	485.959	272.411
平台 5	703.434	166.493
平台 6	97.342	63.900
平台 7	788.709	85.422
平台 8	940.875	787.406
平台 9	965.489	880.421
平台 10	786.069	148.859

根据上述仿真条件设定,进行军队院校多源管理数据采样,得到原始采集的数据如图 3 所示。

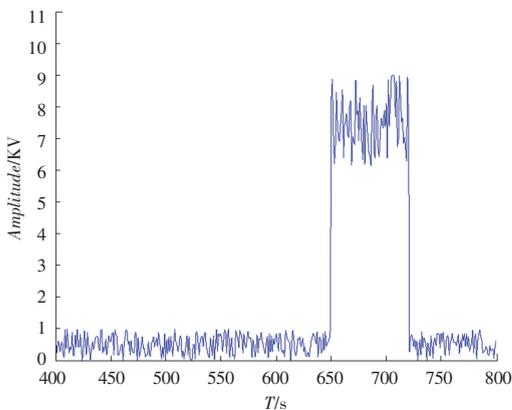


图 3 原始采集的数据

Fig. 3 Original collected data

以图 3 的军队院校多源管理数据为测试对象,采用本文方法进行多源数据融合清洗调度冗余数据并封装后返回到 docker 中,得到输出的数据如图 4 所示。

分析图 4 得知,本文方法能有效实现对军队院校多源管理数据冗余数据,提高了数据的动态提取能力。测试数据清洗过程收敛性,得到测试结果如图 5 所示。

图 5 采用多元回归联合自相关分析方法进行数据清洗性能的收敛性测试,分析得知,随着嵌入维的增大,清洗过程逐步收敛,测试数据的查准性,对比结果见表 2。分析得知,本文方法对数据的查准率比传统方法提升 12.3%,

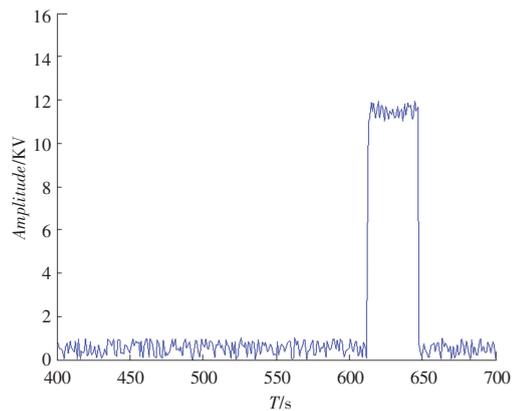


图 4 冗余清洗输出

Fig. 4 Redundant cleaning output

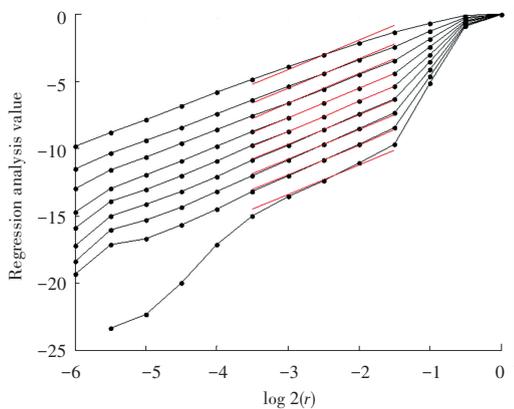


图 5 收敛性测试

Fig. 5 Convergence test

表 2 查准性测试

Tab. 2 Comparison of precision test

迭代次数	本文方法	文献[3]	文献[5]
10	0.976	0.778	0.676
20	0.940	0.795	0.544
30	0.968	0.709	0.572
40	0.957	0.768	0.603
50	0.977	0.694	0.648
60	0.948	0.678	0.674
70	0.970	0.780	0.556
80	0.916	0.768	0.524
90	0.954	0.604	0.589
100	0.962	0.735	0.508

## 5 结束语

采用大数据的云融合技术,结合对海量数据的采集、清洗和调度处理,提高军队院校的数字化信息管理水平。本文提出基于联通公有云平台调度的军队院校多源数据融合清洗技术。构建军队院校的信  
(下转封三)