

doi: 10.7690/bgzdh.2022.09.013

基于大数据分析技术的战场态势分析及预测

王秀娟¹, 曹 瑾², 王建强², 韩文华²

(1. 武警后勤学院基础部, 天津 300309; 2. 武警部队研究院建设发展研究所军事工作研究室, 北京 100020)

摘要: 针对信息化条件下联合作战产生的海量、多源、复杂的战场数据, 提出一种 Hadoop 分布式数据处理平台。收集海量数据进行战场态势 (battle field situation, BS) 要素分析, 用粒子群算法 (particle swarm optimization, PSO) 优化极限学习机 (extreme learning machines, ELM) 的方法对战场态势历史数据进行训练, 构建战场态势预测模型; 并采用 Matlab2018 对战场态势进行模拟仿真。仿真结果表明: Hadoop 处理海量战场数据效率更高, 可有效提高战场态势的预测精度, 为辅助指挥员快速掌握复杂战场态势提供新的方法和途径。

关键词: 联合作战; 战场态势; 粒子群; 极限学习机

中图分类号: TJ02 **文献标志码:** A

Battlefield Situation Analysis and Prediction Based on Big Data Analysis Technology

Wang Xiujuan¹, Cao Jin², Wang Jianqiang², Han Wenhua²

(1. Department of Basic, Logistics University of People's Armed Police Force, Tianjin 300309, China;

2. Military Work Laboratory, Construction and Development Research Institute,
Research Academy of PAP, Beijing 100020, China)

Abstract: Aiming at the massive, multi-source and complex battlefield data produced by joint operations under the condition of informationization, a Hadoop distributed data processing platform is proposed. Massive data are collected to analyze the elements of battle field situation (BS), and the particle swarm optimization (PSO) is used to optimize the extreme learning machines (ELM). The method of ELM is used to train the historical data of battlefield situation and construct the prediction model of battlefield situation, and Matlab 2018 is used to simulate the battlefield situation. The simulation results show that Hadoop is more efficient in processing massive battlefield data, and can effectively improve the prediction accuracy of battlefield situation, which provides a new method and way for assisting commanders to quickly grasp the complex battlefield situation.

Keywords: joint operation; battlefield situation; particle swarm optimization; extreme learning machine

0 引言

信息化条件下的联合作战是在陆、海、空、天、电、网和认知领域多维空间展开的诸军兵种联合实施的复杂体系作战, 战场态势信息时空范围广、种类多、信息量大, 作战节奏与进程加快, 战争迷雾已成为制约指挥决策的重要因素。面对复杂多变的战场环境, 随着信息技术的不断发展, 大量侦察探测和传感设备应用于战场, 战场态势数据以惊人的速度增长。对于这些海量、多源、复杂、异构且快速增长的战场数据, 如何加以采集、存储与分析利用, 结合大数据技术实现智能化战场态势认知, 辅助指挥员快速地掌握战场态势和战争趋势, 预测变化趋势, 成为战场信息化过程中亟待解决的问题。

战场态势要素分析及预测问题最早由美军提出, 国外在该领域的理论和系统的研究与开发取得

了较大成果, 而国内这方面的研究还处于起步阶段^[1-6]。文献[1]提出强制稀疏子编码神经网络的作战态势评估方法, 建立了自编码深度学习网络模型; 文献[2]提出了基于时空维度分析的战场态势预测方法; 文献[3]提出一种基于灰色神经网络的战场态势分析。目前, 针对联合作战条件下, 面对海量战场数据态势分析领域的理论研究还不系统, 多为综述性和构建框架成果, 缺乏相关理论模型; 因此, 笔者以联合作战条件下产生的海量、复杂、多源、异构的战场数据为背景, 对战场态势预测及分析进行了研究。

1 战场态势分析方案及设计流程

未来战争是体系与体系间对抗, 以战争复杂系统思想为指导, 面对复杂多变的战场状态, 系统地

收稿日期: 2022-05-09; 修回日期: 2022-06-23

基金项目: 武警后勤学院基础研究项目 (WHJ202101)

作者简介: 王秀娟(1984—), 女, 山东人, 硕士, 从事数学建模研究。E-mail:306344931@qq.com。

研究联合作战条件下战场态势，为辅助指挥员准确认知和科学决策提供技术支撑。笔者基于 Hadoop 数据分析技术，构建如图 1 所示的分布式大数据分析建模系统框架。

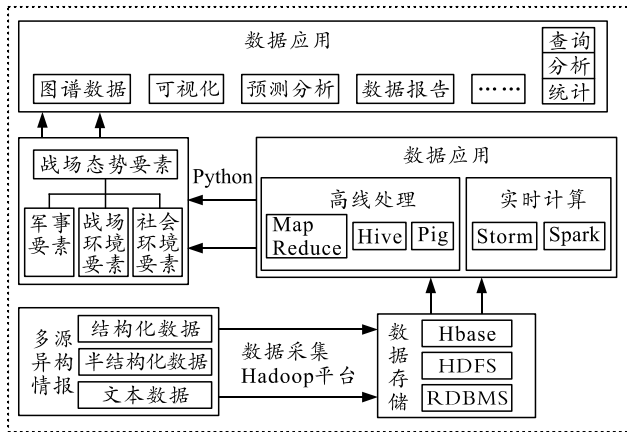


图 1 分布式大数据分析建模系统框架

对于结构化、半结构化和非结构化数据，通过建立 Hadoop 数据处理平台对战场态势要素特征进行抽取、清洗和转换，使用 HBase 和数据存储 (Hadoop distributed file system, HDFS) 存储非关系型数据，使用 RDBMS 存储关系型数据；使用 MapReduce、Pig 和 Hive 处理离线数据，再使用 Spark 和 Storm 高速处理实时数据。分析整理后的数据可以实现知识图谱、分析报告、可视化以及各种功能

性应用、专题应用和预测分析。

1.1 战场态势关键要素分析和提取

战场态势 (BS) 指战场中兵力分布及战场环境的当前状态和发展变化趋势。态势要素指构成战场态势的兵力、环境、事件和估计等诸类要素。战场数据复杂多样、异构多元，从海量战场数据中采集、提取关键要素是未来实现智能态势认知的关键。

战场态势要素也称为战场态势估计要素，通常战场态势主要由 3 类要素构成：

- 1) 军事要素：敌我双方的作战能力要素与动态目标要素。例如：空间战场中的固定设施 (雷达站、阵地、指挥中心等)。
- 2) 战场环境要素：社会/政治/经济环境，包括地理数据 (地形、对地位置、太阳风等) 及电磁气象环境等。
- 3) 社会环境要素：包括政治因素、经济状况、社会动态等。

随着信息技术不断发展，面对海量战场数据，如何快速准确提取战场态势关键要素 (如图 2 所示)，进而将信息优势转化为行动优势，已经成为智能辅助决策的关键。笔者提出一种基于 Hadoop 分布式数据处理平台，可根据战场态势关键要素进行分片并行处理，再结合 python 对数据进行统计分析。

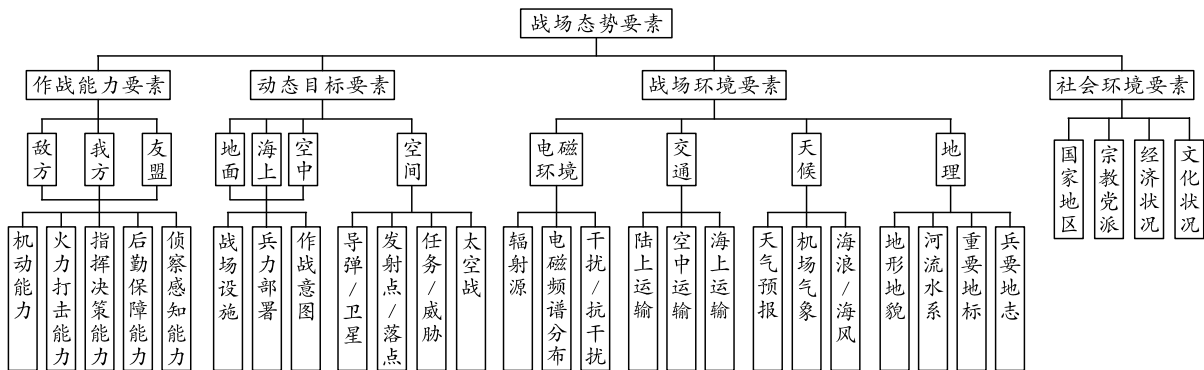


图 2 战场态势关键要素结构

1.2 Hadoop 分布式数据处理平台

如图 3 所示，Hadoop 是一个能够处理大数据集的软件程序框架。该平台实施案例主要用来处理来自多库多表的海量结构化数据。

具体流程如下：

- 1) 数据存储 (HDFS)。

Hadoop 特有的用来分布式存储海量数据文件系统。将大数据文件分割为小的数据文件 block，存储在集群中各个节点的硬盘中，每个 block 有 3 个副本数，由 Block 统一管理。

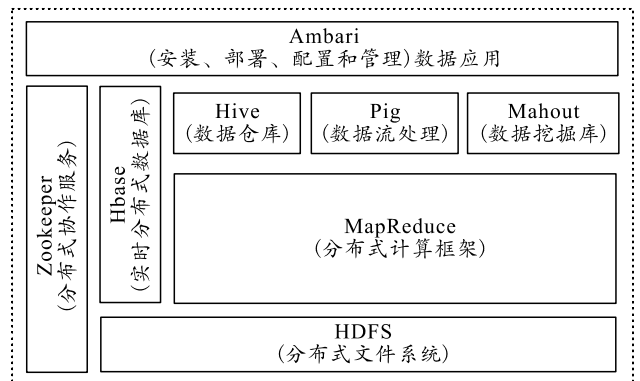


图 3 Hadoop 的生态系统

- 2) 资源管理(yet another resource negotiator, YARN)。
- 一个资源管理和任务调度的分布式框架。
- 3) 数据分析框架(MapReduce)。

图 4 用来分析海量数据框架。Map 阶段：将大数据文件分为很多小的数据文件，并行处理输入数据，对每一部分数据进行分析处理。Reduce 阶段：对 Map 结果进行汇总，合并 Map 处理数据的结果。

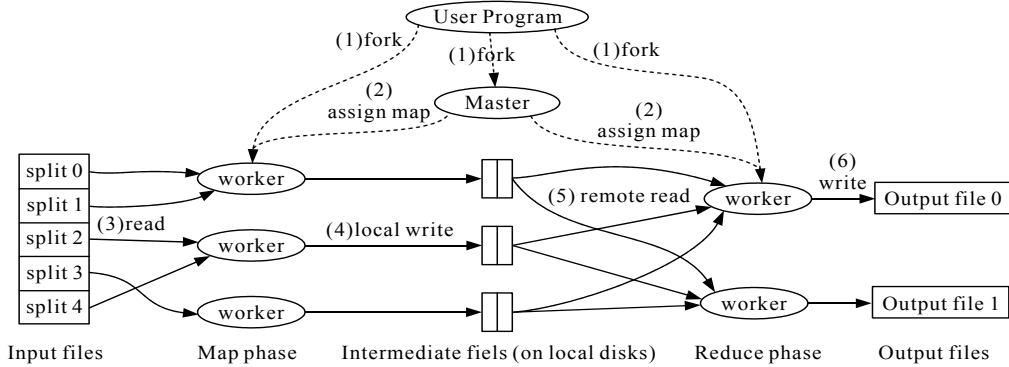


图 4 Map Reduce 模型计算

1.3 战场态势预测思路

基于大数据分析技术的战场安全态势预测思路：

- 1) 利用 Hadoop 收集战场态势历史数据，并对数据进行分片处理，数据字段信息主要有时间、敌我双方的兵力部署、作战能力、战场环境和社会环境等，使用逗号分割隔；
- 2) 数据分析：将要分析的数据存储到 HDFS 文件系统中，使用 Python 编码，通过数据挖掘算法(K-means、SVM、K-最近邻、分类和回归树等)进行数据整理和统计分析；
- 3) 采用 Matlab 软件，选择 PSO-ELM 算法实现对战场态势预测模型。

1.4 ELM 算法

极限学习机(ELM)是一种针对单隐含层前馈神经网络的新算法，是在 Moore-Penrose 矩阵理论基础上所提出的快速学习算法，相对于传统的前馈神经网络训练速度慢、容易陷入局部极小值点等缺点，ELM 算法可以随机产生输入层和隐含层的权重和阈值，只需设置隐含层神经元的个数，便可获得唯一最优解。在保证学习精度的前提下，比传统的学习算法速度更快、泛化性能更好。ELM 结构图如图 5 所示。

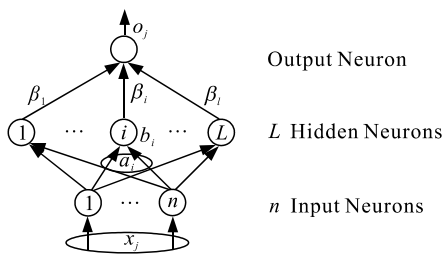


图 5 ELM 结构

假设有 N 个任意的样本 (x_i, o_i) ，其中： $x_i=[x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n$ ； $o_i=[o_{i1}, o_{i2}, \dots, o_{in}]^T \in R^n$ 。

假设一个含有 L 个隐含层的单隐层神经网络可表示为：

$$\sum_{i=1}^L \beta_i g(\omega_i \cdot x_j + b_i) = o_j, j=1, 2, \dots, N. \quad (1)$$

式中： $g(x)$ 为隐含层的激活函数； $\omega_i=[\omega_{i1}, \omega_{i2}, \dots, \omega_{in}]^T \in R^n$ 为输入层第 i 个神经元到隐含层的权重； $\beta_i=[\beta_{i1}, \beta_{i2}, \dots, \beta_{in}]^T \in R^n$ 为隐含层中第 i 个神经元到输出层的输出权重； b_i 为第 i 个隐层单元的偏置值。

上述 N 个方程可简化为：

$$H\beta=T. \quad (2)$$

式中： H 为隐层节的输出矩阵； β 为输出权重； T 为期望输出。其具体形式如下：

$$H(\omega_1, \dots, \omega_L, b_1, b_2, \dots, b_L, x_1, x_2, \dots, x_L) = \begin{bmatrix} g(\omega_1 \cdot x_1 + b_1) & \dots & g(\omega_L \cdot x_L + b_L) \\ \vdots & & \vdots \\ g(\omega_1 \cdot x_N + b_1) & \dots & g(\omega_L \cdot x_N + b_L) \end{bmatrix}_{N \times L};$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m}; \quad T = \begin{bmatrix} T_1^T \\ \vdots \\ T_N^T \end{bmatrix}_{N \times m}.$$

为了能够训练单隐层神经网络，希望得到 $\hat{\omega}_i, \hat{b}_i$ 和 $\hat{\beta}_i$ ，使得

$$\|H(\hat{\omega}_i, \hat{b}_i)\hat{\beta}_i - T\| = \min_{\omega, b, \beta} \|H(\omega_i, b_i)\beta_i - T\|. \quad (3)$$

其中， $i=1, 2, \dots, L$ ，等价于最小化损失函数：

$$E = \sum_{j=1}^N (\sum_{i=1}^L \beta_i g(\omega_i \cdot x_j + b_i) - o_j)^2. \quad (4)$$

1.5 PSO-ELM 算法步骤

粒子群算法 (PSO) 是一种模拟鸟群觅食行为的研究, 最早由 Eberhart 和 Kennedy 于 1995 年提出。每个鸟就是 PSO 中的粒子, 这些鸟在寻找食物的过程中, 通过不断改变位置与速度寻找食物。PSO 算法就从这种生物种群行为特性中得到启发并用于求解优化问题。PSO-ELM 算法是通过 PSO 优化 ELM 的权值和阈值, 其适应度函数^[7-11]用式(3)表示。PSO 优化 ELM 的战场态势预测算法流程图如图 6 所示。

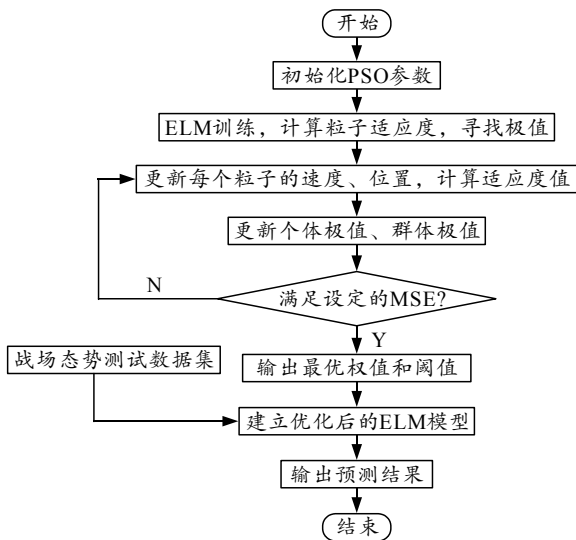


图 6 PSO-ELM 算法流程

- 1) 初始化 PSO 中的基本参数, 主要包括种群大小、学习因子、惯性权重的最大最小值和最大迭代次数;
- 2) 将经过筛选后的战场态势样本数据归一化, 建立训练样本和测试样本;
- 3) 将构建样本输入 ELM 进行训练学习, 计算出每个粒子的目标函数值, 并找出每个粒子的个体和全局最优粒子的位置和最值;
- 4) 更新粒子的速度、位置及惯性权重;
- 5) 计算适应度的同时更新速度和位置;
- 6) 检查 PSO 的终止条件, 如果满足最大迭代次数或者最优解时已经停止而不再变化, 则输出最优解位置, 否则, 返回第 4) 步;
- 7) 通过 PSO 优化 ELM 获取的最优参数 $\hat{\omega}_i, \hat{b}_i$ 和 $\hat{\beta}_i$, 实现战场态势预测。

2 战场态势预测及仿真

2.1 测试数据

为了增强仿真结果的说服力, 在当前缺少大规

模历史作战数据支撑条件下, 依靠深度学习、对抗网络和数据挖掘技术自主生成训练样本。笔者采用 5 个不同样本数量 (如表 1 所示) 的战场态势要素数据模拟战场环境, P_1 (机动能力)、 P_2 (火力打击能力)、 P_3 (天候气象)、 P_4 (战术水平)、 P_0 (作战效果) 的部分数据如表 2 所示。模拟仿真, 粒子群算法优化极限学习机参数设置如下: 最大迭代次数为 200, 种群规模为 40, 学习因子 $c_1=c_2=2$, 输入神经元个数为 6, 隐含层神经元个数为 30, 输出层神经元为 1, 训练次数为 20 000 次, 学习速率为 0.02, 训练误差为 0.000 01, PSO-ESM 的训练结果和预测对比如图 7、8 所示。

表 1 战场态势历史数据

编号	样本数量	编号	样本数量
1	500	4	3 000
2	1 000	5	5 000
3	1 500		

表 2 战场态势要素数据

序号	P_1	P_2	P_3	P_4	...	P_0
T_1	67	74	65	0.54	...	62.87
T_2	70	78	74	0.57	...	71.54
T_3	89	85	86	0.76	...	81.53
T_4	82	76	73	0.66	...	73.24
T_5	71	82	71	0.85	...	74.68

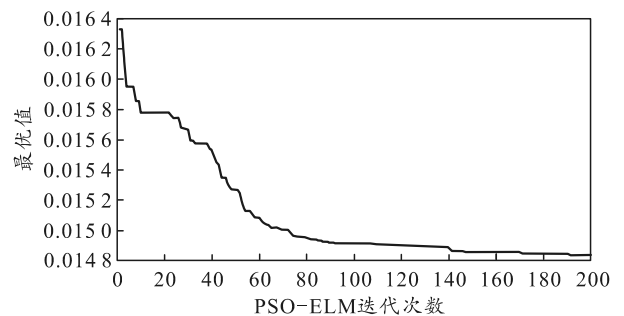


图 7 PSO-ELM 适应度曲线

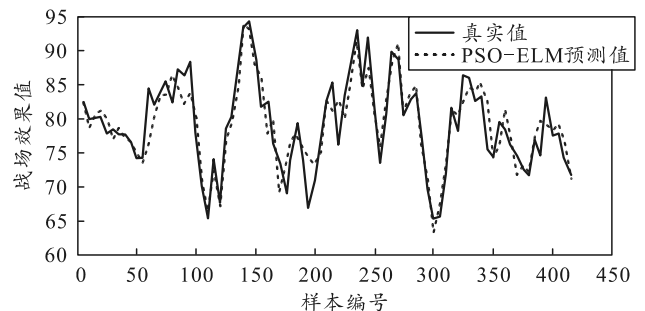


图 8 PSO-ELM 预测值和真实值对比

2.2 评价指标

为了对战场态势预测效果进行科学的评价 (结果如表 3 所示), 选择均方根误差 (root mean square error, RMSE)、绝对百分比误差 (mean absolute percentage error, MAPE)、绝对误差 (mean absolute

error, MAE) 作为预测效果的评价指标, 评价公式为:

$$MAPE = \frac{1}{N} \sum_{i=1}^N |(y_i - \hat{y}_i) / y_i| \times 100\%; \quad (5)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N ((y_i - \hat{y}_i) / y_i)^2}; \quad (6)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \times 100\%。 \quad (7)$$

式中: y_i 为某一时刻战场态势作战效果实际值; \hat{y}_i 为某一时刻战场态势作战效果预测值; N 为预测时间长度。

表 3 不同算法的对比结果

方法	RMSE	MAE/%	MAPE/%
PSO-ELM	1.527 8	0.745 8	0.642 5
ELM	2.425 1	1.954 2	2.124 6
BP 神经网络	3.542 3	2.346 7	4.232 1

2.3 结果及分析

2.3.1 战场态势的预测精度对比

采用 BP 神经网络、ELM、粒子群优化 ELM 分别对战场态势数据进行建模, 预测精度如图 9 所示。

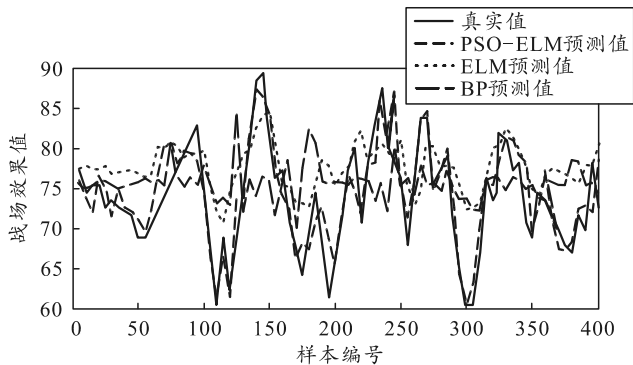


图 9 BP 神经网络、ELM、PSO-ELM 预测值对比

1) BP 神经网络、ELM 对战场态势数据进行建模, 预测结果偏差较大、精度低, BP 神经网络出现了过拟合现象。

2) PSO-ELM 对战场态势预测精度要高于 BP 神经网络、ELM, 这是因为极限学习机具有训练参数少、学习速度快、泛化能力强的优点。如图 10 所示, 给出了隐含层神经元个数对 PSO-ELM 性能的影响, 可以看到当隐含层神经元个数为 50 左右时, PSO-ELM 训练效果最佳。

2.3.2 战场态势预测效率对比

分别计算 BP 神经网络、ELM、粒子群优化 ELM 训练时间和测试时间, 如图 11、12 所示。从图中可以发现, BP 神经网络的测试时间相差不大, 而 ELM 和粒子群优化 ELM 的建模训练时间差异很大。其

中: BP 神经网络的训练时间最长; ELM 训练时间次之; PSO-ELM 训练时间最短, 而且大大提升了战场态势预测建模时间和预测效率。

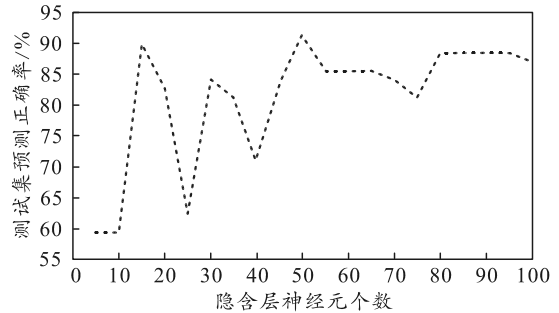


图 10 隐含层神经元个数对 PSO-ELM 性能影响

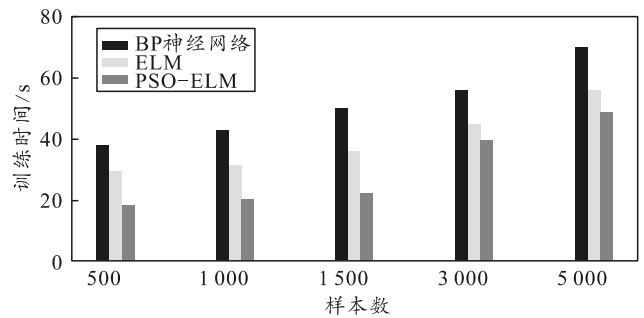


图 11 战场态势训练时间

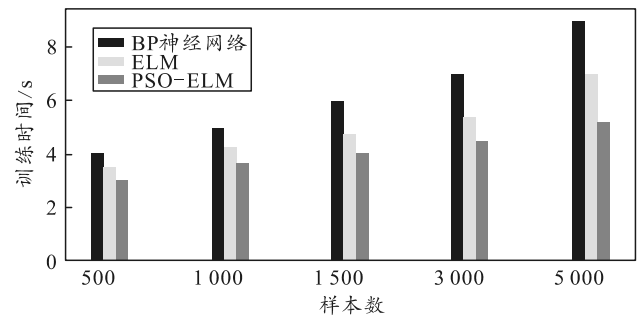


图 12 战场态势测试时间

2.3.3 Hadoop 分布式数据处理平台的优越性

为了比较 Hadoop 分布式数据处理平台和传统单节点数据处理平台在处理大数据方面的优越性, 建模方法采用 PSO-ELM, 得到不同数据集下战场态势训练和预测时间, 如图 13 所示。

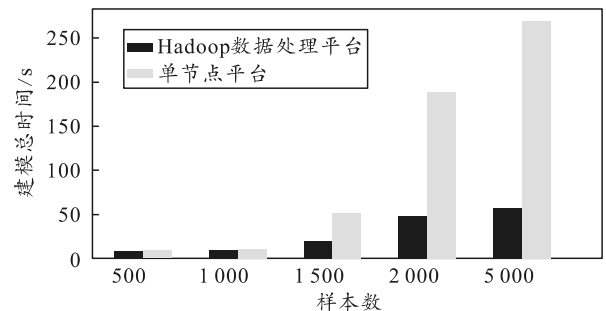


图 13 Hadoop 平台和单节点平台的建模时间