Optimization of Aluminum Alloy Cleaning Agent Formulation by Combining LHD Algorithm and MLP Neural Network

Authors: Lin Ruixiang, He Peihao, Lin Ruixiang

Date: 2024-01-23T00:00:00+00:00

Abstract

To address the deficiencies of traditional formulation design methods—namely, being cumbersome and time-consuming, exhibiting high material consumption, and yielding poor optimization effectiveness—a novel method for aluminum alloy cleaning agent formulation optimization was established by integrating the Latin Hypercube Design (LHD) algorithm with Multi-Layer Perceptron (MLP) neural networks. By leveraging the LHD algorithm's capability for comprehensive parameter space exploration and the efficient modeling capacity of MLP neural networks, the aluminum alloy cleaning agent formulation was successfully optimized. The optimized formulation exhibited an increase in cleaning power from 87.9% to 98.24% and a reduction in corrosion amount from 4.2 mg to 0.3 mg. Linear regression fitting between model-predicted and experimental data yielded a coefficient greater than 0.98, thereby demonstrating consistency between the predicted and experimental results.

Full Text

Optimization of Aluminum Alloy Cleaner Formulations by Combining LHD Algorithm with MLP Neural Network

Peihao He[1], **Ruixiang Lin**[1] [1]Dagao Industrial Technology Research Institute (Guangzhou) Co., Ltd., Guangzhou 510700, China

Abstract

Conventional formulation design methods suffer from being cumbersome, time-consuming, material-intensive, and yielding suboptimal results. To address these limitations, we propose a novel optimization approach for aluminum alloy cleaner formulations that integrates Latin Hypercube Design (LHD) algorithm

with Multilayer Perceptron (MLP) neural network. By leveraging LHD's ability to comprehensively explore parameter spaces and MLP's efficient modeling capabilities, we successfully optimized an aluminum alloy cleaner formulation. The optimized formulation achieved a substantial performance improvement: cleaning efficiency increased from 87.9% to 98.24%, while corrosion amount decreased from 4.2 mg to 0.3 mg. Linear regression fitting between model predictions and experimental data yielded R^2 values exceeding 0.98, confirming strong consistency between predicted and measured results.

Keywords: aluminum alloy cleaner; Latin Hypercube Design (LHD) algorithm; Multilayer Perceptron (MLP) neural network
Classification: TP393

Introduction

Traditional cleaner formulation design relies heavily on extensive trial-and-error experimentation and empirical knowledge accumulation, making it both time-consuming and costly. In recent years, computer-aided methods and machine learning techniques have garnered significant attention for experimental design in complex parameter spaces [1]. Machine learning methods such as neural networks and genetic algorithms demonstrate powerful capabilities for efficient optimization within complex parameter spaces, substantially reducing the time and costs associated with conventional experimental approaches [1].

Traditional response surface methodology and orthogonal experiments are often constrained by their limited exploration of specific parameter ranges, failing to provide comprehensive coverage of complex parameter spaces [2], which may restrict the discovery of truly optimal formulations. This paper demonstrates that combining Latin Hypercube Design (LHD) algorithm [3] with neural network methods offers even more pronounced advantages [4]. LHD algorithm samples the entire parameter space more efficiently and uniformly, enabling more comprehensive exploration. Meanwhile, neural networks, as powerful machine learning tools, possess exceptional capabilities for learning and modeling complex data [3]. Compared to conventional methods, neural networks can more effectively process datasets and capture complex nonlinear relationships between parameters. This high accuracy enables more precise prediction of formulation performance during the optimization of aluminum alloy organic alkali cleaners, providing a more reliable basis for final optimization results [8]. This combination offers more comprehensive and accurate data exploration and performance prediction, facilitating breakthroughs in cleaner formulation design.

We selected the combination of LHD algorithm and Multilayer Perceptron (MLP) neural network for optimizing aluminum alloy organic alkali cleaner formulations to fully exploit LHD's strength in comprehensively exploring parameter spaces and MLP's efficiency in modeling and prediction. This approach accelerates the discovery of optimal formulations while achieving efficient and environmentally friendly cleaning performance. This integration

provides industry with more innovative and sustainable solutions, advancing clean production and sustainable manufacturing.

Our workflow follows the flowchart shown in [Figure 1: see original paper], comprising three core components: training data collection, MLP neural network model training, and LHD algorithm-based data generation to identify optimal cleaning scores, followed by reverse derivation of the corresponding input formulation. Although this workflow requires more experiments than traditional methods like orthogonal arrays, it covers more complex parameter spaces and can theoretically identify the best-performing formulation. Furthermore, the trained MLP neural network model can be used for process monitoring to track cleaning efficiency and corrosion amounts. All code in this study was written in Python, making it easily portable to other platforms or systems for training, thus facilitating broader application.

2.1 Training Set Preparation for MLP Neural Network

Training data were collected experimentally. The input data comprise five dimensions representing raw materials in the base cleaner formulation [6,7,11]: monoethanolamine, sodium potassium tartrate, deionized water, magnesium-aluminum corrosion inhibitor DX5819, and a self-prepared nonionic surfactant. The output data consist of two dimensions: substrate cleaning efficiency and corrosion amount.

The substrate selected was 2A12 aluminum alloy, commonly used in aerospace and automotive manufacturing. Cleaning efficiency and corrosion amount were measured and calculated according to procedures specified in sections 5.7 and 5.8 of GB/T35759-2017. The experimental training set comprised 150 data points, which generally satisfies the minimum data requirement for neural network training.

2.2 Training Set Preprocessing

Due to the relatively small training set, we applied Z-Score normalization to clean the data and remove outliers before training. The Z-Score method determines a data point's relative position in the dataset by expressing its distance from the mean in standard deviations. The Z-Score is calculated as:

$$Z = \frac{X - \mu}{\sigma} \tag{1.1}$$

where X is the data point value, μ is the dataset mean, and σ is the dataset standard deviation. Larger absolute Z-Score values indicate greater deviation from the mean. Common practice defines data points with Z-Scores exceeding a certain threshold as outliers. We set this threshold at 2, removing all values more than 2 standard deviations from the mean. This reduced the training set from 150 to 143 data points.

Since formulation coefficients sum to 1, no additional normalization was required. However, due to the limited data volume, we applied data augmentation before MLP training. This involved generating noise from a normal distribution and adding it to each sample to create new augmented samples, thereby reducing overfitting risk.

MLP Neural Network Model Construction

To further enhance model generalization, we constructed the MLP neural network using the Sequential model in TensorFlow 2.15 [9], with architecture shown in [Figure 2: see original paper]. By configuring fully connected layers with multiple hidden and output layers, the model can more effectively learn and represent features. Specifically, our input layer comprises five features representing the five raw materials in the cleaner formulation, while the output layer contains two nodes for predicting cleaning efficiency and corrosion amount.

The fully connected layer design leverages connections between neurons across layers, enabling the model to better understand and capture data features and patterns. This structure allows the neural network to more accurately predict required performance metrics during formulation optimization, providing a reliable foundation for formulation design and performance prediction.

The MLP neural network model contains five features in the input layer, three hidden layers, and two nodes in the output layer [Figure 2: see original paper]. We employed a neural network architecture with three hidden layers containing 8, 5, and 3 neurons, respectively. Following each hidden layer, we applied ReLU (Rectified Linear Unit) activation, added Dropout layers, and introduced L2 regularization (weight decay) to limit model complexity and reduce overfitting.

Given the characteristics of cleaning efficiency and corrosion data, we applied Sigmoid activation in the output layer to ensure values remained between 0 and 1, keeping outputs within a reasonable range appropriate for the problem. For model optimization, we used the Adam optimizer with learning rate decay to better balance convergence speed and optimization effectiveness [10]. During training, we implemented early stopping to reduce overfitting risk and employed K-fold cross-validation for multiple training and evaluation cycles. We monitored model performance using loss function, accuracy, and F1 score to achieve good generalization [5].

MLP Neural Network Model Evaluation

We designated 10 data points as a test set to further evaluate model generalization by comparing predicted and experimental values for cleaning efficiency and corrosion through regression fitting curves.

LHD Algorithm for Data Generation and MLP Input

The LHD algorithm uniformly distributes sample points in multidimensional space, ensuring adequate coverage across all dimensions. This avoids the potential unevenness of random sampling, provides more comprehensive datasets, and requires no initial input values. In practice, since formulation degrees of freedom equal 4, we used LHD algorithm to randomly sample four raw materials (excluding deionized water). Thanks to LHD's efficiency, we could easily generate over 100,000 formulations as input data. Feeding these into the MLP neural network rapidly produced predictions for cleaning efficiency and corrosion amount.

Since higher cleaning efficiency is desirable while lower corrosion is preferred, we defined a dimensionless cleaning score as:

$$\text{Cleaning Score} = \frac{w_1 \cdot \text{Cleaning Efficiency} + w_2 \cdot (1 - \text{Corrosion Amount})}{w_1 + w_2} \tag{1.2}$$

where w_1 and w_2 are weight coefficients. In this study, to balance magnitudes and give equal importance to cleaning efficiency and corrosion amount, we set $w_1 = w_2 = 1$. These weights can be adjusted to modify the relative importance of each factor. The cleaning score can also be adapted for different research objectives, such as incorporating cost factors.

As shown in equation (1.2), the formulation with the minimum cleaning score represents the optimal solution under the current weight coefficients.

3 Results and Discussion

MLP Neural Network Model Training and Evaluation During MLP neural network training, we observed that as training epochs increased, the loss function consistently decreased on both training and validation sets, while accuracy increased continuously until reaching 1, as shown in [Figure 3: see original paper]. This trend indicates successful model training. records the results of 5-fold cross-validation, including loss function, accuracy, and F1 score for each fold. The minimal performance variation across different validation sets demonstrates good model generalization. These results show steady performance improvement on both training and validation sets with increasing training iterations, while cross-validation results confirm model robustness.

To comprehensively evaluate generalization performance, we prepared 10 datasets not used during training as a test set. As observed in [Figure 4: see original paper] and [Figure 5: see original paper], although some differences exist between MLP predictions and experimental data, linear regression fitting of predicted versus experimental results yielded R^2 values exceeding 0.98 for both cleaning efficiency and corrosion amount. This indicates high consistency between predictions and measurements with acceptable deviation. Considering

these results, we conclude that the MLP neural network model demonstrates good accuracy and generalization capability based on our collected training set.

3.1 Optimization Results Comparison After validating the MLP neural network's generalization performance, we implemented the Latin Hypercube Design (LHD) algorithm with appropriately defined data ranges to generate datasets of varying scales: 10,000, 100,000, and 1 million groups as inputs. Feeding these datasets into the MLP neural network rapidly produced predicted cleaning efficiency and corrosion values, enabling quick identification of minimum cleaning scores and their corresponding input formulations.

Results from the 1-million-data input set showed the cleaning score decreasing from 61.08 in the experimentally best formulation to 51.20. Cleaning efficiency improved from 87.9% to a maximum of 98.24%, while corrosion amount decreased from 4.2 mg to a minimum of 0.31 mg, representing significant performance improvements.

Comparing optimization results across 10,000, 100,000, and 1 million datasets revealed a trend of gradual improvement with increasing dataset size. However, performance gains were modest when comparing 100,000 to 1 million datasets, suggesting diminishing marginal returns beyond 100,000 data points. Given the efficiency of LHD algorithm and MLP neural network, we recommend utilizing these methods to maximize dataset size for optimal results.

presents the experimentally best formulation alongside predicted optimal formulations from 10,000, 100,000, and 1 million datasets, with corresponding cleaning efficiency, corrosion amount, and cleaning scores.

To ensure reliability of optimization results, we conducted three replicate experiments on optimized formulations and averaged the results. Linear regression fitting of predicted versus experimental values yielded R^2 values close to 1, as shown in , confirming consistency between experimental and predicted results.

Conclusion

This study successfully achieved efficient optimization of aluminum alloy cleaner formulations by combining LHD algorithm's uniform sampling across parameter spaces with MLP neural network's efficient modeling capabilities. The optimized formulation demonstrated significant performance improvements over traditional experimental results, with cleaning efficiency increasing from 87.9% to 98.24% and corrosion amount decreasing from 4.2 mg to 0.3 mg. The optimization process was highly efficient, providing industry with more innovative and sustainable solutions. Moreover, the high prediction accuracy of the MLP neural network model enables convenient adjustment of cleaning score weights or addition of other concern factors to achieve desired optimal results. This method can play important roles not only in aluminum alloy cleaner formulation development but also in formulation development across other fields and performance monitoring during application.

References

[1] You Yang. Method and Application of Experimental Design in Complex Test Space[D]. National University of Defense Technology, 2022.
[2] Xiaobing Shang, Tao Chao, Ping Ma, et al. An efficient local search-based genetic algorithm for constructing optimal Latin hypercube design[J]. Engineering Optimization, 2020, 6(1).
[3] Joseph, V. R. & Hung, Y. Orthogonal-maximin Latin hypercube designs[J]. Statistica Sinica, 2008, 18: 171–186.
[4] LeCun, Y., Bengio, Y., Hinton, G. Deep learning[J]. Nature, 2015, 521: 436–444.
[5] Sengul Mert Y., Song Yao, Van Duin Adri C. T. INDEEDopt: a deep learning-based ReaxFF parameterization framework[J]. npj Computational Materials, 2021, 7(1).
[6] LIU N H, GUO Y H, LAI X Q, et al. Development of an environmentally friendly cleaning agent for 7-series aluminum alloys[J]. Electroplating & Finishing, 2023, 42(14): 37-41.
[7] DENG Y C, XU X Y, WANG F, et al. Corrosion behaviors of three metals used in heat exchangers after being treated with an alkaline cleaner[J]. Electroplating & Finishing, 2023, 42(12): 17-22.
[8] HAN W, WANG Y, SU X, et al. A multi-objective optimization problem research for amphibious operational mission of shipboard helicopters[J]. Chinese Journal of Aeronautics, 2023, 36(09): 256-279.
[9] Joseph, V. R., Dasgupta, T., Tuo, R., Wu, C. F. Sequential exploration of complex surfaces using minimum energy designs[J]. Technometrics, 2015, 57: 64–74.
[10] Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization[EB/OL]. https://arxiv.org/abs/1412.6980, 2015.
[11] JI W, MENG F D, PAN K Y, et al. Preparation and Corrosion Research of a Water-Based Cleaning Agent for Aluminum Alloy[J]. Plating and Finishing, 2021, 43(02): 38-43.

Corresponding Author: Ruixiang Lin, E-mail: 20162432007@m.scnu.edu.cn

Author Contributions:
Ruixiang Lin: Conceptualization, methodology design, code development, data processing and analysis, manuscript writing;
Peihao He: Methodology design, experimentation.

Note: Figure translations are in progress. See original paper for figures.