Optimizing Extreme Learning Machine Algorithm using Particle Swarm Optimization to Estimate Iron Ore Grade

Mahdi Fathi, Andisheh Alimoradi\(^*\), and Hamidreza Hemati Ahooi

Department of Mining Engineering, Faculty of Engineering, Imam Khomeini International University, Qazvin, Iran

1. Introduction

The geologists and mine exploration engineers always have a challenge in estimating the ore grade, especially in the metallic deposits. The reason is perhaps the dependence of the exploration and extraction projects on the exact amount of ore grade in order to model the reserve amount and design the mining activities. Often due to the surface and underground faults and complex geological structures, the simulation of deposits is associated with certain problems. It should be noted that simulation is one of the most important and difficult processes in geoscience [1]. In general, it can be found that the traditional or geometric methods suffer from over-simplifying, unlike the geostatistical methods that are complicated in the evaluation process. Selecting one of these methods is usually a choice between the speed and accuracy or between the low cost and attention to the deposit details [2].

The artificial intelligence-based methods can overcome this problem using more traditional methods with a more realistic strategy since these methods are capable of understanding the hidden relationships between the different input and output variables in complex and non-linear situations [2, 3]. As one of the widely used sub-branches of artificial intelligence and one of the general branches of artificial neural networks, machine learning has been organized and developed in order to understand the patterns in the surface and deep explorations [4].
Many researchers have employed different artificial neural networks for grade estimation in the recent years. For example, a Radial Basis Function (RBF) network has been used successfully for grade estimation in an iron deposit in England, and the results obtained have been compared with the geostatistical models [4]. In another case, a Four-Level Perceptron network (4L-MLP) has been used on the modified magnetic data in order to estimate the iron grade [6]. Badel has compared one of the newer kriging methods called the median indicator kriging with the artificial neural networks for grade estimation in an iron ore deposit [7]. The grade estimation results of a Choqart iron ore deposit obtained from a back-propagation neural network have been compared with the results of a Support Vector Machine (SVM) [8]. In another research work, the artificial neural networks and geo-statistics have been integrated using ANNMG in order to optimize the mineral reserve evaluation in the SW Sierra Leone [9]. Nezamolhosseini has applied a multi-layer perceptron (MLP) neural network in order to estimate the storage of the Chaghart mine using the exploratory boreholes data [10]. In the last article, the comparison between the Local Linear Radial Basis Function using Skewed Gaussian activation (LLRBF-SG) and older neural networks such as Differential Evolution (DE), Cuckoo Search (CS), Covariance Matrix Adaptation Evolution (CMAE), Artificial Bee Colony (ABC), Improved Artificial Bee Colony (IABC) were used in order to estimate the phosphate grade in Bafgh's Esfondi [11]. All of these research works have illustrated that artificial neural networks can be used as a reliable approach to obtain the most accurate grade estimations.
The feed-forward neural networks have a superior performance in approximating the complex non-linear relationships and providing the models that cannot be easily predicted by the classical parametric techniques. The extreme learning machine is a feed-forward neural network that has hidden neurons in one or more layers for classification, regression analysis, derivatives, Laplace transform comparison, and training of characteristics [11].

Machine learning is a science that makes a computer information plan for a particular subject to identify. The aim of machine learning in the artificial neural network is to find the effective efficiency in the information model by increasing the data in the network process. The range of research works that is used in machine learning in earth sciences is very wide and complex. Theoretically, the researchers are determined to find or create the new network learning methods that can improve the quality of learning for a research work, and on the other hand, some researchers have tried to apply the deep learning methods to new issues and identify their strengths and weaknesses. However, this spectrum is continuous, and most of the research works on both methods [12].
The extreme learning machine, as the main modeling tool in this work, is easy to use with the capability to reduce the training errors and make a generalized result. This method runs extremely fast, and all the characteristics differentiate it from the other popular SLFN learning algorithms [14].

Most metaheuristics are characterized as follow:

- They are used as the strategies to guide the search process.
- They explore the search space to find the near-optimal solutions.
- They are simple local to complex learning procedures.
- They are approximate and non-deterministic estimators.
- They are not problem-specific [15].

Particle swarm optimization (PSO) is an iterative metaheuristic optimization method that can solve the problems that can be answered in a multi-dimensional space. The particles are assigned to an initial velocity per unit of time, and also the maximum and minimum communication channels between these particles are considered. Then these particles are motion-based in the general response space, and the evaluations based on a specified competency criterion are calculated for all particles with a constant speed after each repeating interval. Over time, some particles are accelerated towards other particles that are more attestable to the competency criterion and are placed in the equal communication group, and the search environment is in the best response [16].
2. Choghart Iron Deposit

Located 12 km NE of the city of Bafgh in Iran, the Choghart mine is one of the biggest iron mines in Iran. The main orebody at Choghart is in the form of a roughly vertical, discordant, pipe-shaped body plunging 73° NNW, and has been explored to a depth of 600 m. The orebody is roughly vertical, pipe-shaped, and asymmetric. Different types of volcanic (intrusive and extrusive alkali rhyolites) and metamorphous rocks occur in the vicinity of the deposit. Synite, pyroxenite, gabbro, granite, and alkali rhyolites are the major components of the volcanic rocks in the Choghart deposit.

Figure 5. PSO after N-iteration [15].

Figure 6. Satellite photo illustrating location of the drilling points.
Figure 7. 85 boreholes in the north Choghart anomaly. The blue ones are the validation boreholes that are not used in the learning model.

Figure 8. A) Central Iran’s position towards Zagros and Alborz, B) Landscaping map of central Iran blocks. The Choghart mine location is illustrated by a yellow star [17].
3. Methodology

In this section, the background surveys and observations about the borehole datasets are presented first. Then the hybrid neural network and how it works in order to estimate the ore grade is introduced.

3.1. Data preparation

After combining the assay, collar, and survey data from the raw data, the results obtained were composited into the 3-m samples. This was conducted in order to equalize the length of the samples, and also to minimize the distances of the drilling data. The total data number was 9067.

For two reasons the samples were composite in 3 m: The first reason was that most of the samples were approximately 3 m long, and therefore, the samples were less changed. The second reason was that the grade was also estimated for 1- and 2-m composites but the 3-m results were better than those for the 2-m ones.

3.2. Data statistical studies

The statistical studies on the composite data showed that there were many changes in the dataset both in depth and in different speculations. The mean value of this data was 9.15, the variance was 138.253, and the standard deviation (SD) was 11.75. Such variance and standard deviation indicate a high data distribution compared to the average. In the probability statistics, the symmetry indicates the asymmetry of the probability distribution. If the data is symmetric compared to the average, the density will be zero, where our density is positive and has a small amount of about 1. In the probability statistics, stretching describes the level of highness and flatness in a possible distribution. Table 1 illustrates the statistical parameters of the data.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Grade value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>9.14184</td>
</tr>
<tr>
<td>Median</td>
<td>1.47400</td>
</tr>
<tr>
<td>Mode</td>
<td>0.000</td>
</tr>
<tr>
<td>Std. deviation</td>
<td>11.758083</td>
</tr>
<tr>
<td>Variance</td>
<td>138.253</td>
</tr>
<tr>
<td>Skewness</td>
<td>1.327</td>
</tr>
<tr>
<td>Std. error of skewness</td>
<td>0.026</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>1.568</td>
</tr>
<tr>
<td>Std. error of kurtosis</td>
<td>0.051</td>
</tr>
<tr>
<td>Range</td>
<td>69.033</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.000</td>
</tr>
<tr>
<td>Maximum</td>
<td>69.033</td>
</tr>
</tbody>
</table>

![Figure 9. Iron ore grade histogram.](image)
3.3. Data normalization

Since the network can find the hidden patterns in a large dataset, the input and output parameters of this composit data should be placed in an equal size. For this purpose, the composite dataset should be placed in the normalized domain [0, 1]; in order to place the numbers estimated by the network in the position of analogy with the real outputs, the following fixed formula should be used:

\[ X_{\text{new}} = (X_{\text{old}} - X_{\text{mean}}) \cdot (1/X_{\text{std}}) \]

where \( X_{\text{mean}} \) is the average of each column of data, \( X_{\text{std}} \) is the standard deviation of each column of data, and \( X_{\text{old}} \) and \( X_{\text{new}} \) are the normal and converted values of each data, respectively.

3.4. Extreme learning machine (ELM) algorithm

As a single hidden layer feedforward neural network, ELMs were extended in order to the generalize SLFNs. There is no need in tuning the hidden layer in ELMs. The output function is as follow:

\[ f_i(x) = \sum_{i=1}^{L} \beta_i h_i(x) = h(x) \beta \]

where \( \beta = [\beta_1, ..., \beta_L]^T \) is the output weights in the hidden layer of \( L \) nodes, and \( h(x) = [h_1(x), \ldots, h_L(x)]^T \) is the output vector of the hidden layer; \( h(x) \) transfers the data from the input space to the hidden-layer feature space. And therefore, \( h(x) \) is a feature mapping. The decision function of ELM for the classification applications should be:

\[ f_i(x) = \text{sign}(h(x) \beta) \]

Opposite to the traditional learning techniques, ELM makes both the training error and the norm of output weights as small as possible. According to the Bartlett’s theory, the smaller the norms of the weights are, the better generalization performance the networks tend to have.

Minimize: \( ||H\beta - T||^2 \) and \( ||\beta|| \)

Here, \( H \) is the hidden-layer output matrix.

\[ H = \begin{bmatrix} h_1(x_1) \\ \vdots \\ h_L(x_N) \end{bmatrix} = \begin{bmatrix} h_1(x_1) & \ldots & h_L(x_1) \\ \vdots & \ddots & \vdots \\ h_1(x_N) & \ldots & h_L(x_N) \end{bmatrix} \]

According to the ELM’s capability in the target function approximation, the classification problem for the proposed constrained-optimization-based ELM with a single-output node can be formulated as follows:

Minimize \( L_{\text{Primal-ELM}} = \frac{1}{2} ||\beta||^2 + C \frac{1}{2} \sum_{i=1}^{N} \xi_i^2 \)

subject to: \( h(x_i)\beta = t_i - \xi_i \ i = 1, \ldots, N \)

Based on the Karush–Kuhn–Tucker (KKT) theorem, the following optimization problem should be solved in order to train the ELM algorithm:

Minimize \( L_{\text{Dual-ELM}} = \frac{1}{2} ||\beta||^2 + C \frac{1}{2} \sum_{i=1}^{N} \xi_i^2 - \sum_{i=1}^{N} \alpha_i (h(x_i)\beta - t_i + \xi_i) \)

In order to apply ELM as a multiclass algorithm, there should be multi-output nodes instead of a single-output node. For the original class label of \( p \), the expected output vector of the \( m \) output nodes is

\[ t_i = \begin{bmatrix} 0, \ldots, 1, \ldots, 0 \end{bmatrix}^T \]

Therefore, the classification problem for ELM with multi-output nodes can be formulated as follows:

Minimize \( L_{\text{Primal-ELM}} = \frac{1}{2} ||\beta||^2 + C \frac{1}{2} \sum_{i=1}^{N} \xi_i^2 \)

subject to: \( h(x_i)\beta = t_i^T - \xi_i^T \ i = 1, \ldots, N \)

According to the KKT theorem, the following dual optimization problem should be solved in order to train ELM:

Minimize \( L_{\text{Dual-ELM}} = \frac{1}{2} ||\beta||^2 + C \frac{1}{2} \sum_{i=1}^{N} \xi_i^2 - \sum_{i=1}^{N} \sum_{j=1}^{m} \alpha_{i,j} \left( h(x_i)\beta_j - t_{i,j} + \xi_{i,j} \right) \)

Different from the Support Vector Machine (SVM), the hidden-layer output vector is known in ELM. All the non-linear piecewise continuous functions such as sigmoid, hard-limit, Gaussian, and multi-quadric functions can be used as the hidden-node output functions.

Sigmoid and Gaussian are two major hidden layer output functions for ELM, and hard-limit and multi-quadric are the most generalized ones.
3.5. Particle swarm optimization (PSO)

PSO is a collective, anarchic (in the original sense of the term), iterative method with the emphasis on the cooperation in its historical version; it is partially random and without selection. Eberhard and Kennedy have introduced PSO as an optimization algorithm that derives its inspiration from the social behavior of the birds and fishes. PSO can be used in order to solve a wide range of optimization problems, from the non-linear continuous functions to the most complex engineering problems.

In this work, the PSO algorithm was used, which could be described as follows. If the search space of optimization is considered as a D-dimensional space, then a D-dimensional vector (X_i) can be the representative of the ith particle of the swarm, X_i = (x_{i1}, x_{i2}, ..., x_{id})^T. Another D-dimensional vector can be considered as the velocity or position change vector for each particle, V_i = (v_{i1}, v_{i2}, ..., v_{id})^T. The parameter g is defined as the best particle of the swarm index, and obviously, the superscripts will show the iteration number. After each iteration, the position of every particle is updated based on the particle best exploration, best exploration among all the swarms, and also the previous velocity vector of the particle using the following two equations:

\[ v_{id}^{n+1} = v_{id}^n + c_1 r_1^n (p_{id}^n - x_{id}^n) + c_2 r_2^n (p_{gd}^n - x_{id}^n) \]
\[ x_{id}^{n+1} = x_{id}^d + v_{id}^{n+1} \]

where w is called the inertia weight, c_1 and c_2 are the constant values called the cognitive and social parameters, respectively, and \( \theta \) is a constriction factor. PSO is extremely computationally inexpensive in terms of both the memory requirements and speed. It is also flexible to integrate with other optimization and soft-computing techniques to form hybrid tools. In addition to these advantages, it was never coded as the training algorithm of a neural network to be used for the grade estimation before. Thus the performance of this optimizer in these kinds of operations can be tested for the first time in this research work [18].

4. Grade Estimation by PSO-ELM Algorithm

Applying the iterative metaheuristic algorithms to optimize the output of the single-layer extreme learning regression (ELM’s regression) is a work that affects the accuracy of the results and a better regression performance. For this purpose, the training and test processes are conducted once by the single-layer extreme learning machine, and then the whole process is conducted and the regression algorithm of the single-layer extreme learning machine and its inputs in PSO is used in order to identify the inputs and find the optimized values for them, and to familiarize them with the extreme learning machine’s regression.

The cost function in the PSO algorithm is the average absolute error (MAE) of the grade estimated extreme learning machine with a real grade. In this work, the conversion function of the single-layer extreme learning machine was the Elliot symmetric sigmoid transfer function based on the sigmoid relationships [19].

The results of the hybrid algorithm for the anomaly grade estimation are in the following chart:
Figure 10. The $R$, RMSE, and MAE changes in different iterations for the hybrid algorithm training sector.

![Graph showing the $R$, RMSE, and MAE changes in different iterations for the hybrid algorithm training sector.]

Figure 11. The $R$, RMSE, and MAE changes in different iterations for the hybrid algorithm testing sector.

![Graph showing the $R$, RMSE, and MAE changes in different iterations for the hybrid algorithm testing sector.]

The point-to-point validation method and the borehole-to-borehole validation method were used in order to investigate the ability of the hybrid algorithm in the grade estimation. In the point-to-point validation method, about 10% of the data was randomly set aside, and after learning the algorithm, they entered the network to be estimated by the algorithm and compared with the real values. The point-to-point validation result is shown in Figure 12. As shown, the optimized network was completely capable of simulating the trends and values of the grade change in a point-to-point validation case.
In the borehole-to-borehole validation method, the borehole numbers 1-4 were set aside, and after learning the algorithm, they entered the network to be estimated by the algorithm and compared with the real values of the ore grade. Figures 13-16 illustrate the simulation results for the boreholes 1-4 (2 boreholes in the central part of the grid and 2 others in the margins).
Figure 13. Comparison between the real and estimated grade values in a borehole-to-borehole validation case (borehole 2).

Figure 14. Comparison between the real and estimated grade values in a borehole-to-borehole validation case (borehole 3).
The results obtained indicate that the network could simulate the boreholes in the center of the grid (boreholes 2 and 3 according to Figure 7) better than the margin boreholes (boreholes 1 and 4). In the central boreholes, the network modeled the grade trend surprisingly but in the margin boreholes, we see some displacements in the grade values. It is due to the weak capability of the model in the extrapolation simulation.

5. Conclusions

Grade estimation is one of the most difficult steps in mine development and management. It is very hard and almost impossible to achieve a reliable grade estimation without proper methods and a valuable data; however, it gets even more complex when the lack of a proper data is the issue.

The estimations reported in this work were conducted with the data that had no geological and lithological information. However, the proposed method could be one of the effective methods for estimating the iron ore grade. Estimations with the clustered data of six clusters based on the borehole intervals proved that more accurate results could be obtained for each cluster regardless of all the structural complexities of the iron ore. It should be noted that the single-layer ELM network optimized by the PSO algorithm had a superior performance compared to the main network without PSO optimization. Therefore, the PSO algorithm is one of the most appropriate algorithms to optimize a neural network for the grade estimation using statistically low data. One of the most important understandings from this experiment is that regular drilling, even if carried out in a small number of boreholes and even when the study case is very random, can play a major role in the accuracy of the grade estimation because it gives the simulator the ability to make an accurate estimate in smaller parts, and then integrate the results in order to achieve a good and relatively reliable estimate for the whole region.

In the borehole-to-borehole modeling, there are some displacements in the grade estimation in margin boreholes due to the weak capability of the network to extrapolate the results. We recommend using other data such as the lithological data, which can make a logical relation between the grade data and the coordinate. This will probably lead us to make a better simulation of the margin boreholes.

Acknowledgments

The authors would like to express their gratitude to the Central Iran mines industries and the Kavoshgaran Consulting Engineers Company for their collaboration and support in permitting us to use the data that made this research work possible.

References


به هنگام سازی ماسیون‌های یادگیری حذف‌های به‌کم‌بینه ساز ازدحام ذرات به منظور تخمین عیار آهن

مهدی فتحی، اندیشه علی مراذی، حمیدرضا همتی آهوی

گروه مهندسی معدن، دانشگاه بین‌المللی امام خمینی، قزوین، ایران

ارسال: 23/03/2021

alimoradi@eng.ikiu.ac.ir

چکیده

عدم قطعیت‌های علمی، عملیات تخمین عیار در کاوش‌های فازی را همواره با مشکل مواجه می‌کند. این مقاله به معرفی یک روش تلفیقی با استفاده از دو تکنیک هوش مصنوعی در تعیین عیار آهن می‌پردازد. این روش تلفیقی بر اساس یک ماسیون‌های یادگیری حذف‌های به‌کم‌بینه ساز ازدحام ذرات است. این روش تلفیقی محسوب می‌شود که بر اساس موقعیت گمانه‌ای اکتشافی، محاسبه‌ها و اطلاعات موجود در آن می‌باشد. روش مورد بحث در این مطالعه، در تخمین عیار آهن در آنومالی شمالي چغاپ در ایران مصرف مورد استفاده قرار گرفته است. نتایج آزمون و آزمون‌های تکراری، نشان دهنده مشاهده‌های علمی به‌ینه شده در تخمین عیار است.

کلمات کلیدی: تخمین عیار، هوش مصنوعی، به‌ینه ساز ازدحام ذرات، ماسیون‌های یادگیری حذف‌های به‌ینه، اطلاعات گمانه اکتشافی.