

# **Feature Selection from Iron Direct Reduction Data Based on Binary Differential Evolution Optimization**

Saleh SHAHBEIG <sup>1,\*</sup>, Khalil SADJAD <sup>2</sup>, Mohsen SADEGHI <sup>2</sup>

<sup>1</sup> Young Researchers and Elite Club, Najafabad Branch, Islamic Azad University, Najafabad, Iran

<sup>2</sup> Foolad Technic International Engineering Company (FIECo), Isfahan, Iran

\*Corresponding author: saleh\_shahbeig@yahoo.com

**Abstract**—Nowadays increasing growth in the production of the steel products makes automatic identification of effective parameters in determining the quality of the output product is very important. In this regard, level II automation plays an important role. In this study, a novel method has been proposed to identify effective parameters in determining the purity of sponge iron in the process of Iron Direct Reduction. In the proposed method, differential evolution (DE) optimization algorithm with the binary approach has been used in order to identify the subset of effective parameters with the lowest estimation error in determining the purity of sponge iron. The binary differential evolution algorithm is combined to the Least Squares- Support Vector Machine (LS-SVM) regression method to candidate a subset of the effective parameters. Implementation of the proposed algorithm on data obtained from the practical project (Bardsir steel complex) confirms the effectiveness of the proposed method so that by choosing the effective parameters, the ability to estimate the sponge iron purity with 98.8% accuracy (1.2% estimation error) has been attained.

**Keyword**—Differential evolution optimization algorithm, feature selection, binary approach, direct reduced iron, LS-SVM regression.

## **1. Introduction**

There are two conventional methods for the crude steel production in the industrial plants:

- Crude iron production in the Blast Furnaces (BF) and then steel production in the oxygen converters.
- Direct reduction of iron ore (pellet) and melting sponge iron and scrap iron in electric furnaces.

The crude steel production using the blast furnaces (indirect reduction) is an old method that has some significant disadvantages.

- The setting up costs in the specified capacity for the BF units is higher than the DR units.
- The BF method has more environmental restrictions than the DR method.
- Blast furnace at high capacities is economically feasible while the direct reduction method can be used in small units with good economic justification.
- Sponge iron is easy to transport while carrying agglomerates (blast furnace feed) are more difficult.

Considering implied shortcomings, spongy iron in the making iron industry is an acceptable alternative for the scrap iron. The spongy iron that is produced in the direct reduction process, in comparison with the scrap iron, has a proper karat. The Direct Reduced Iron (DRI) can be produced in the gas-based or coal-based technologies. In this regard, gas-based methods are preferred for the regions with the vast gas resources. One of the most conventional solutions is the midrex method. Since a large number of parameters effect on the sponge iron production process, so the automatic and optimal identification of subset of the parameters affecting the purity of sponge iron, is considered important in the steel industry. Therefore, identify the most effective subset of the available parameters is required in order to reduce the overall cost of the production and adjust the parameters quickly and automatically.

In this paper, a novel optimal feature selection method is proposed that the effective parameters can be elected based on binary DE optimization algorithm. The proposed methodology is applied to find the optimal subset of parameters (features) that is lead to the highest metallization degree. The selected parameters accurately estimate the system output with no needs to the excess variables.

The rest of the paper is organized as follows.

Section 2 explains the proposed algorithm and its preliminaries, briefly. Computerized simulations using the practical data are presented in section 3. Finally, our conclusions are presented in the section 4.

## 2. Proposed Algorithms

Block diagram of the proposed algorithm to select the optimal subset of parameters to determine the purity of sponge iron is shown in Fig. 1. If the initial population of the introduced optimization algorithm can be expressed in the form of a matrix, then each row of the matrix represents a member (answer). Each element of the selected member in the population consists of binary value "0" or "1", which represents the "existence" or "absence" of the parameter in the member. In this paper, by applying the proposed algorithm on the DRI data in several times all samples attend in the testing phase at least once. This means that random conditions of the plant will be reduced and the robustness of the algorithm will be improved.

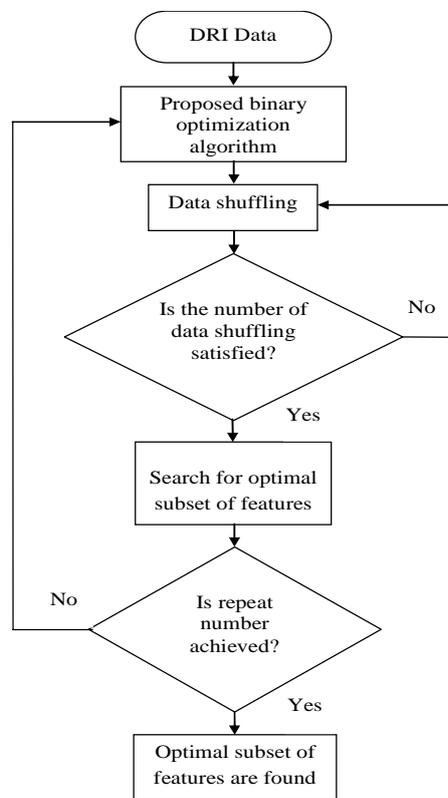


Fig. 1: Block diagram of the proposed algorithm.

In this study, the LS-SVM regression has been used to estimate the purity of the output production. Using the proposed binary differential evolution optimization algorithm in this study, the subset of effective parameters in determining the purity of sponge iron has been identified with high accuracy. The only limitation of the differential evolution optimization algorithm is the number of initial population that should be selected between 3 to 10 times the number of parameters (control variables). By applying the differential evolution optimization

algorithm on the DRI data, all the solutions have been saved in the repository matrix. In this case, selection of the optimum solution will be provided. In order to explain the steps of the proposed algorithm, each of the components are described separately as follows.

### **2.1. LS-SVM Regression**

Support vector machine is an intelligent machine that is learned based on the statistical fundamentals. The machine supervised learning procedure is implemented using input-output system data. Support vector machine uses supervised learning method [1-4]. At first the support vector machine was invented to classify data into two classes but gradually this tool was used for multi-class classification and regression. Support vector machine idea is firstly introduced to solve the linear classification problems but the kernel functions transfer the main problem solution space from a low finite- dimensional linear space to a higher order nonlinear space. This mapping technique, known as the kernel trick, modifies the hyper planes for the nonlinear problems. Namely, this methodology is extended for the nonlinear classification applications.

In the Support Vector Machine Regression (SVR) problem, a nonlinear function is estimated from the system input-output signals that keep the property of the maximum confidence margin. One of the most leading versions of SVR that has a significant capability in the modeling problems is the Least Squared- Support Vector Machine (LS-SVM) [5-7]. In this regard, the regression model is determined as follows.

$$g(a_i) = H^T f(a_i) + m \quad (1)$$

where  $i = 1, \dots, N, \{a_i, b_i\}$  determines the system input-output signals.  $f$  represents the nonlinear mapping function.  $m$  and  $H$  are the model parameters that are calculated from the minimization of the following function.

$$F = 0.5 \|H\|^2 + \xi F_1 \quad (2)$$

where  $\xi$  is the regularization parameter that establish a tradeoff between the output error minimization and the model complexity factors.  $F_1$  is considered for the experimental error. As a result, the optimization problem of the LS-SVM regression can be determined as follows.

$$\begin{cases} \min J[H, e_i] = 0.5\|H\|^2 + 0.5\xi P \\ b_i = H^T f(a_i) + m + e_i \end{cases}, i = 1, \dots, N \quad (3)$$

$$P = \sum_{i=1}^N e_i^2 \quad (4)$$

The Lagrange function is derived from Equation (3) and (4) as follows.

$$L(H, \lambda_i, m, e_i) = J + \sum_{i=1}^N \lambda_i [b_i - H^T f(a_i) - m - e_i] \quad (5)$$

where  $\lambda_i \geq 0$  and  $e_i$  are the Lagrange coefficients and the learning errors, respectively. Based on the Karush-Kuhn-Tucker (KKT) conditions, the first-order partial differentiate of the Lagrange function can be represented as follows.

$$\begin{cases} \frac{\partial L}{\partial H} = 0 \\ \frac{\partial L}{\partial \lambda_i} = 0 \\ \frac{\partial L}{\partial m} = 0 \\ \frac{\partial L}{\partial e_i} = 0 \end{cases} \quad (6)$$

which the above system of equations can be solved as:

$$H = \sum_{i=1}^N \lambda_i f(a_i), \sum_{i=1}^N \lambda_i = 0, \lambda = \xi e_i \quad (7)$$

$$b_i - H^T f(a_i) - m - e_i = 0 \quad (8)$$

and can be simplified into the following least squares problem.

$$\begin{bmatrix} 0 & 1 & \dots & 1 \\ 1 & k(a_1, a_1) + \frac{1}{r} & \dots & k(a_1, a_N) \\ \dots & \dots & \dots & \dots \\ 1 & k(a_N, a_1) & k(a_N, a_N) + \frac{1}{r} & \dots \end{bmatrix} \begin{bmatrix} m \\ \lambda_1 \\ \vdots \\ \lambda_N \end{bmatrix} = \begin{bmatrix} 0 \\ b_1 \\ \vdots \\ b_N \end{bmatrix} \quad (9)$$

Namely, the LS-SVM regression model is written as:

$$\tilde{b}(a) = \sum_{i=1}^N \lambda_i \varphi(a_i, a) + m \quad (10)$$

where  $\varphi(a_i, a) = f(a_i)^T f(a)$  is the kernel function and should be satisfied in the Mercer's Theorem. The most convenient kernel functions applied in the modeling problems are the linear functions, polynomial functions and Radial Basis Functions (RBF).

## 2.2. Binary Differential Evolution Optimization Algorithm

Differential evolution algorithm is one of the most powerful optimization algorithms which is based on the non-human fertility. The member that is combined with the mother (current member) to produce the next generation is the mutated member that can be achieved by combining several members of the previous population [8-9]. In the differential evolution algorithm, the produced member remains only if that quality is better than mother (previous generation). Otherwise, the newly produced member gets destroyed. In the differential evolution algorithm, all members of the current population play the role of mother and father (mutated member) is formed by combination of the mothers. It should be noted that the mutated member (father) for every member of the population (for each mother) is different. In this paper, in order to produce the mutated member for each member of the current population ( $X_i^t$ ), the following equation is proposed.

$$X_{mut} = X_{best} + F_1(X_b - X_{k1}) + F_2(X_{k2} - X_{k3}) \quad (11)$$

Which  $F_1$  and  $F_2$  are the mutation factors and usually are in the range of  $0.1 \leq F_1, F_2 \leq 0.9$ .  $X_{k1}$ ,  $X_{k2}$  and  $X_{k3}$  are randomly selected from the current population which should satisfy the following equation  $X_{k1} \neq X_{k2} \neq X_{k3} \neq X_i^t$ .  $X_{best}$  and  $X_b$  are the best global and local solutions, respectively.

If  $X_i^t = (X_{i,1}^t, X_{i,2}^t, \dots, X_{i,n}^t)$  is the position of the i-th member in the iteration t with n variables and  $X_{mut} = (X_{mut,1}, X_{mut,2}, \dots, X_{mut,n})$  is the generated mutated member for the member  $X_i$ , then the generated member (born fetus) is produced according to the following equations.

$$X_i^{t+1} = (X_{i,1}^{t+1}, X_{i,2}^{t+1}, \dots, X_{i,n}^{t+1}) \quad (12)$$

$$X_{ij}^{t+1} = \begin{cases} X_{mut,i,j}^t & \text{if } Cr > rand() \text{ or } j == I \\ X_{ij}^t & \text{otherwise} \end{cases}, I \in [1, \dots, n] \quad (13)$$

$Cr$  is the combination constant, which is usually in the range  $0.2 \leq Cr \leq 0.9$ . In the above equations, it is clear that the generated member in the differential evolution algorithm should never be exactly the same as his mother.

By calculating the fitness function for the generated member and compare it with the fitness function related to the mother, the best member will remain and the worst member will be removed. This process is done for all members of the current population (first iteration). By examining the convergence condition, the above process is repeated until the optimal solution is obtained.

In order to use the binary DE algorithm in this paper, the following equations are introduced to determine the binary value of the elements of the generated  $X_{mut}$ .

$$Sig(X_{mut,i,j}) = \frac{1}{1 + e^{-X_{mut,i,j}}} \quad (14)$$

If the result of this relationship is more than a random value ( $rand$ ), the value of "1" is replaced with  $X_{mut,i,j}$  and otherwise the value of "0" is replaced.

### 3. Implementation Results

#### 3.1. Database

The suggested methodology has been implemented on the Bardsir steel complex database. This database is a set of 176 selected data sample that each one comprised of 40 controllable parameters. Each data sample is gathered from the data record of 23 temperature variables, 8 gas flow meters, 8 gas analyzers and one weight transmitter.

#### 3.2. Validation Parameters

In this paper, the proposed algorithm has been validated using the Inverse Mean Square Error (IMSE) criterion. IMSE criterion is defined as follows.

$$\frac{1}{m \text{ean}[(Y_{test} - Y_{real})^2]} \quad (15)$$

where  $Y_{real}$  and  $Y_{test}$  are the real system output and the estimated output, respectively. In other words, the maximum value of the criterion defined in (15) is sought. As a result, the proposed algorithm has been utilized to reach an effective subset of the parameters.

### 3.3. Case Study

Considering the limitations of the differential evolution algorithm, the initial population size is selected as ( $N = 3.n$ ) 120 members. The proposed algorithm is repeated 10 times, iteratively. All of the operators in equation (11) are considered as binary operators. It is implied that selected or not-selected parameters in the proposed binary optimization algorithm is represented as “1” or “0”. The proposed algorithm is implemented on the DRI database (Fig. 2). As it is shown, the proposed algorithm is able to achieve 98.8% estimation accuracy (1.2% estimation error).

The most important effective parameters in determining the purity of sponge iron can be mentioned as follows.

- The flow of natural gas parameter
- The flow of process gas parameter
- The output  $CO_2$  gas analyzer parameter

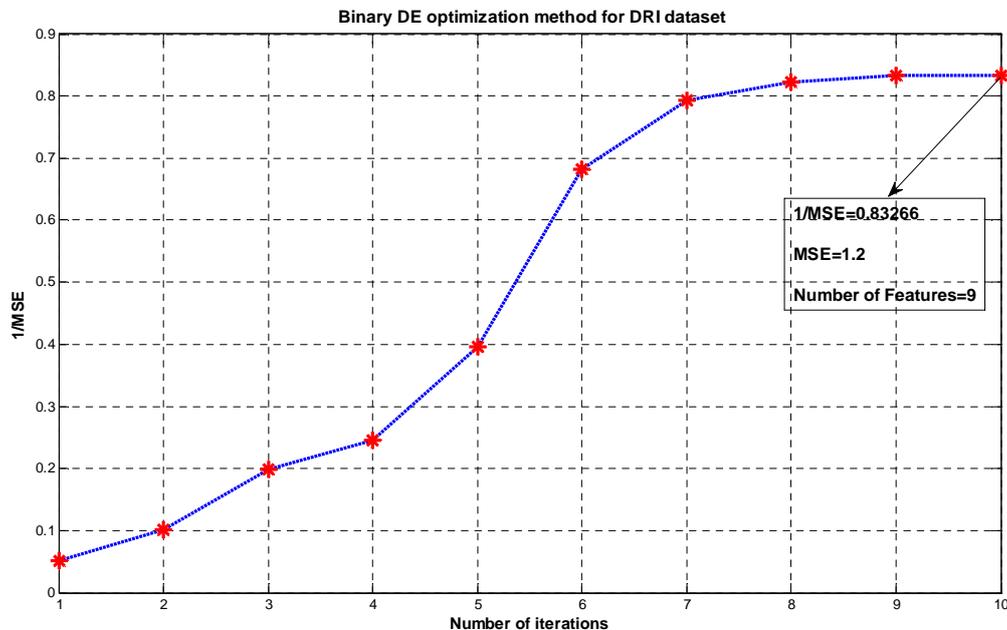


Fig. 2: DRI data analysis using the proposed algorithm.

## **4. Conclusion**

In this paper, an efficient algorithm based on binary optimization was proposed to select a subset of the parameters with the greatest influence in determining the purity of sponge iron. For this purpose, for the first time in this research by applying the differential evolution optimization algorithm in binary form, the effective subset of the parameters in determining the purity of sponge iron was determined. By applying the proposed algorithm on the DRI data in several times, random conditions of the plant was reduced and the robustness of the algorithm was improved.

## **5. References**

- [1] S. Amari and S. Wu, "Improving support vector machine classifiers by modifying kernel functions," *Neural Networks*, vol. 12, pp. 783-789, 1999.
- [2] G.M. Fung and O.L. Mangasarian, "Multicategory proximal support vector machine classifiers," *Machine learning*, vol. 59, pp. 77-97, 2005.
- [3] A. Widodo and B. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 21, pp. 2560-2574, 2007.
- [4] J. Wu and Z.H. Zhou, "Sequence-based prediction of microRNA-binding residues in proteins using cost-sensitive laplacian support vector machines," *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, vol. 10, pp. 752-759, 2013.
- [5] S. An, W. Liu, and S. Venkatesh, "Fast cross-validation algorithms for least squares support vector machine and kernel ridge regression," *Pattern Recognition*, vol. 40, pp. 2154-2162, 2007.
- [6] S.H. Hwang, D.H. Ham, and J.H. Kim, "Forecasting performance of LS-SVM for nonlinear hydrological time series," *KSCCE Journal of Civil Engineering*, vol. 16, pp. 870-882, 2012.
- [7] J.A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, pp. 293-300, 1999.
- [8] A.K. Qin, V.L. Huang, and P.N. Suganthan, "Differential evolution algorithm with strategy adaptation for global numerical optimization," *Evolutionary Computation, IEEE Transactions on*, vol. 13, pp. 398-417, 2009.
- [9] R. Storn and K. Price, "Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces," *Journal of global optimization*, vol. 11, pp. 341-359, 1997.