Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 72 (2018) 1063–1068

www.elsevier.com/locate/procedia

51st CIRP Conference on Manufacturing Systems

Energy Consumption Modelling Using Deep Learning Technique — A Case Study of EAF

Chong Chen[a]*, Ying Liu[a], Maneesh Kumar[b] and Jian Qin[a]

[a]*Mechanical and Manufacturing Engineering, Cardiff University, Cardiff, CF24 3AA, UK*
[b]*Cardiff Business School, Cardiff University, Cardiff, CF10 3EU, UK*

* Corresponding author. Tel: +44(0) 7762930105; *E-mail address:* ChenC54@cardiff.ac.uk

Abstract

Energy consumption is a global issue which government is taking measures to reduce. Steel plant can have a better energy management once its energy consumption can be modelled and predicted. The purpose of this study is to establish an energy value prediction model for electric arc furnace (EAF) through a data-driven approach using a large amount of real-world data collected from the melt shop in an established steel plant. The data pre-processing and feature selection are carried out. Several data mining algorithms are used separately to build the prediction model. The result shows the predicting performance of the deep learning model is better than the conventional machine learning models, e.g., linear regression, support vector machine and decision tree.

Peer-review under responsibility of the scientific committee of the 51st CIRP Conference on Manufacturing Systems.

Keywords: Energy modelling; Intelligent manufacturing; Deep learning; Data mining

1. Introduction

Steel, as one of the world's important recycled materials, is widely used as engineering material [1]. Nowadays, steel is used in many aspects, such as machines, building structures and vehicles. The steel industry is one of the major consumers of energy and material in the form of electricity, oxygen, carbon, raw materials and various other earth metals [2]. Producing high-quality steel with lower electricity consumption is the aim of every steel company [3]. Furthermore, energy use is also restricted by national policy on issues such as carbon footprint and greenhouse gas emission. As a result, reducing energy consumption is one of the most critical challenges in the steel industry [4]. Therefore, if a steel plant can predict its energy usage given a specific formula of ingredients in advance, a proper combination of raw materials can be prepared in order to reduce energy spending without sacrificing steel quality.

The prediction is carried out by the analysis of the historical data. Traditionally, analysts generally use statistical techniques to deal with data in small quantities. Recently, data mining has offered advanced techniques, such as machine learning, to obtain the hidden patterns in a large dataset [5]. As a special type of machine learning technique, deep learning has become increasingly popular in recent years [6]. When the dataset becomes larger, the performance of deep learning becomes better due to its powerful capability to learn hidden patterns [6].

The data, in this case, are collected from the melt shop in an established steel plant in South Wales, UK. The plant buys more than 20 different grades of steel, which can be categorised into nine major types. The different steel batches have different masses and proportions of scraps. Hence, the energy consumption of each batch varies. Although the prediction system in the steel plant can provide reliable monthly energy consumption, it is difficult to predict energy consumption for each batch. However, it is possible to achieve this target with data-mining techniques, which are capable of carrying out big-data analysis [7].

This paper aims to establish an analytical model that focuses on the energy consumption modelling and prediction of an EAF using different types of scraps. In particular, deep learning is called upon jointly with other conventional machine learning techniques to reveal their technical performance.

2212-8271
Peer-review under responsibility of the scientific committee of the 51st CIRP Conference on Manufacturing Systems.
10.1016/j.procir.2018.03.095

2. Literature Review

2.1. Energy Modelling with EAF

Optimising energy efficiency and steel product quality is a big concern in the steel industry. One of the optimising methods is to consider the energy consumption of an EAF and its output product in order to standardise the scraps' blend [8]. Instead of using the standard method, making the process dynamic was another improvement of the optimisation since the variables used in modelling are not always constant [9].

A mathematical optimisation framework integrated with a mechanistic model was proposed to optimise the balance of the mass of scraps and energy consumption. The factors of chemical changes, ingredients and energy flow were fully considered in this case [10].

The process of an alternating current EAF was modelled based on the first law of thermodynamics. In this case, by using a developed computer program, the variation of energy consumption of an EAF was studied. Several parameters that have a strong relation to energy consumption were discussed. The performance of the model is evaluated by comparing the calculated value and the value reported in the literature [11].

A mathematical model of a direct current EAF was developed for predicting the heat transferred from the arc to the EAF bath and the shear stress on the bath surface. The model can also be used for determining the relevant parameters such as the current and electrode dimension [12].

Statistical models have been used in the past to study the effects of process parameters on electricity consumption in EAF. Statistical multivariate data analysis techniques such as principle component analysis (PCA), principle component regression (PCR), and partial least squares (PLS) have been used for EAF energy modelling [13]. However, those statistical models are not capable of precisely predicting the EAF energy consumption due to the increasing number of data, dimensions and impurities.

2.2. Machine Learning in Steel Industry

Nowadays, almost one-third of the steel in the world is produced by EAF. The EAF steel making process is influenced by a lot of factors, such as scrap price, energy consumption and electrodes. Traditionally, people strongly depend on empirical knowledge to make decisions [14]. When a problem is influenced by numerous factors, it is hard for engineers to figure out all the relationships between the factors. However, data mining, as it can extract the knowledge from the collected data, is becoming more and more popular in the industrial field [7]. Hence, it may offer a better solution.

Several linear models were developed to find the balanced points in electrical and chemical energy. Those models enable EAFs to achieve the lowest carbon dioxide emission and higher efficiency. This method does not find out the relationship between the input of an EAF and the consumption of natural gas, while it only indicates that the electrical energy efficiency can be improved by off-gas monitoring and extraction control [15]. Linear regression is the only technique used in this case; other machine learning methods may offer better performance.

In the past, EAF temperature control was mainly based on mathematical methods and finite element analysis. However, those methods cannot meet the expected success in EAF control and reduce energy consumption. A fuzzy neural network model was used to realise the online estimation of the EAF tap temperature. In this model, the neural network was used as a classifier and a fuzzy inference function was used to obtain predicted tap temperature [14]. The model in this case does not show excellent performance in the entire dataset. When the predicted value of the temperature at a high level (over 1640° C), the accuracy declines dramatically.

The random forest tool was used to build a grey box model to predict and control the temperature of molten steel in a tundish of a continuous caster. The model was used to determine the molten steel temperature in the Ruhrstahl-Heraeus degassing process. In this paper, the machine learning model shows better performance than the statistical model [16]. The statistical method is useful when the dataset is not large. When the size of the dataset is large enough, the machine learning method is definitely more advantageous.

A semantic model was formulated to optimise the global steel production workflow in order to decrease the inflexibility of cooperation and data exchange between different components in the steel industry. The seamless data integration is performed in the semantic model to achieve better product quality, cost efficiency, production flexibility and industrial performance. The model was proposed in the theoretical aspect and no has not realised in the real world [17]. It indicates the direction of the steel industry towards intelligent manufacturing.

A genetic programming method for predicting natural gas consumption was carried out in a steel plant. The program trees, which are set by the input variables, are randomly generated. Then, the crossover points and the genes of the parents are randomly selected. The ideal model, which has minimum deviation, can be obtained after the repeated test of the crossover process [18]. The statistical model was used as the only contrasting group which cannot show the advantages of the generic model in the machine learning field.

The steel plant, in this case, uses a Microsoft Excel-based software call solver. The process manager inserts the values of specific parameters and the solver, which calculate the best combination of scraps, energy and yield. Other parameters, such as residual elements are predicted by using a statistical method called multiple regression. The statistical methods can deal with the clean theoretical data, whereas the real-world industry data always contains impurities. Moreover, when the dataset becomes larger, the statistical method tends to be low in efficiency. Deep learning, as a popular data mining tool, is widely used in many fields such as image classification and speech recognition, which show deep learning techniques are advantageous in dealing with big data [6]. The collected data have 40 attributes with 10,990 instances. As the dataset used in this case is large and multidimensional, deep learning can be a useful tool for building the energy consumption model.

3. Methodology

In order to tackle the problem, the following methodology is adopted. The first step is to understand the steel making process. After scraps are collected, they will be added to the EAF with other additives. The next stage is ingredient melting and the impurities will be cleaned. Then, the injecting carbon will adjust the carbon content. Finally, with the continuous casting and hot rolling process, the final steel product will be obtained. The issue here is how to predict the daily consumption of energy by the EAF in this case.

Secondly, the history, availability and collection method of the data collected from the steel plant should be known and the reliability of the data needs to be identified as the dataset may include some incorrect data due to the improper operation or meter failure.

After the data has been thoroughly understood, data pre-processing is carried out in the Weka software [19]. The dataset is large and multidimensional. The number of variables is required to be lowered by using an appropriate feature selection technique to avoid overfitting during modelling. The extreme and missing values can be filled by using the average number rather than just deleting them, since the more data we use, the more accurate result will be.

The energy consumption prediction model will be built by using deep learning techniques and will be compared to the models built by other machine learning algorithms.

Finally, the evaluation process is carried out using a 5-fold cross validation for comparing several metrics, such as Model Correlation Coefficient and Mean Square Error, which are used to indicate the performances of various models.

4. Data Pre-processing

4.1. Data Collection

The steel plant uses software which name is system analysis and program development (SAP SE) to record the data and upload it to the cloud drive. In the EAF working stage, the situation and the data of EAF are monitored and uploaded by people manually. The data maintenance is carried out by another database which name is 'Level 2'. All the changes of the data in Level 2 will be updated in the SAP SE database. However, reversal is not available in this case. Therefore, the data in SAP SE database is more reliable. The first step is to reduce the attributes. All those variables can be classified into seven categories according to the attributes' properties. Table 1 shows the results of attribute classification.

Table 1. The classification of attributes.

Category	Attribute
Scraps	Clean bales 1, Clean bales 2, Merchant 1 & 2, Tin Can, Estructural, Fragmentized scrap, Steel turnings, Recovered Scrap, Total Scrap Mix
Additive	Main Oxygen, Secondary Oxygen, Natural Gas, Carbon injected, Lime, Dolomite
Index	Heat Number
Nominal	Steel Grade
Power	Average power LEVEL 2, PON time (min), TTT Level 2, Power factor, Apparent Power, Reactive Power, Current/kA, Foaming Index, Primary Volts (Onload), Varc
Temperature	T TAP (ºC), TLF entry (ºC), Ladle Energy Cons (KWh), PRECIPITATION (mm), Temperature (ºc)
Pressure	Average of cb1oxymainpressure, Average of cb2oxymainpressure, Average of cb3oxymainpressure
Output	EAF(MWh), Billet Tons

4.2. Feature Selection

The feature selection process is essential as the features selected will be used to build the predicting model. The target of this process is to increase the predicting accuracy, reduce the calculation loads and have a better understanding of the features of the problem domain [20].

The scraps and additives are the ingredients added to the EAF. The mass of the ingredient directly affects the energy consumption. EAF (Mwh) is the energy consumption so that it should be kept. Therefore, those features should be selected to build the prediction model.

The rest of the features, which are hard to determine their effects on the predicting outcome, will be run in the attribute selection function of Weka software. Weka is a powerful data mining software, which contains many functions such as data pre-processing, classifying, clustering and attribute selections. The feature selection function available in the dataset is CfsSubEval. There are two strategies which can be used in this function, which are BestFirst and GreedyStepwise. The outcome of these two procedures is the same in this case. The result shows that the PON time(min) and TTT level 2 have relevance to the EAF in different degrees. Those attributes will be added to the input features of the model. Therefore, the EAF model with features can be settled down. 16 attributes will be selected as input with EAF (Mwh) acting as the only output. Fig.1 shows the EAF model with features.

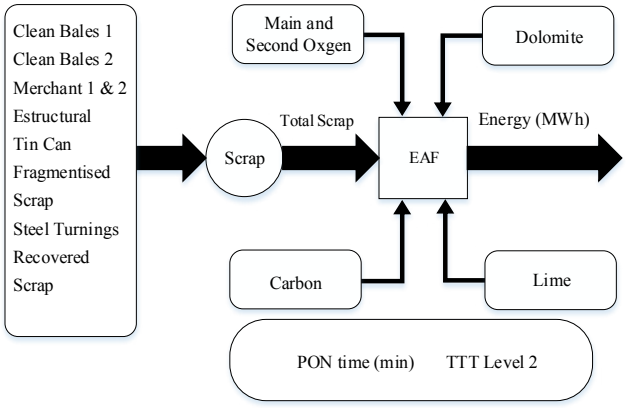

Fig. 1. EAF model with features

4.3. Data Quality and Filtering

Firstly, the data statistics of most are continuous except Dolomite. Nearly 80% of the values of Dolomite locate in the vicinity of 800kg while the rest are discrete from 0kg to 7000kg. In the steel making process, it is unreasonable to add 7000kg dolomite into EAF. The reason may be due to the out of order of instrument or meter. For these extreme data, they will be replaced by the mean value.

Secondly, for the missing value, there are 3% of data in Second oxygen, Main oxygen, Natural gas and Carbon injected which are missing. It is interesting to note that the missing value is concentrated on the last 300 rows of data. Using the mean value to fill those missing values are available for the second oxygen since the variance of this data is not large. This method can also be applied to the missing value of other attributes since the number of missing value takes less than 0.05% of the dataset.

The third step is to randomise the dataset. The data was recorded continuous and there might be some hidden patterns which are caused by the sequence. However, those hidden patterns only show up locally, which cannot represent the characteristic of the whole dataset. Hence, those patterns should not be learnt by the model.

The last step is data normalisation. The deep learning model does not need to learn the patterns of the value, but the change rate of the value (i.e. the gradient of the attributes). Also, the units of the attributes are not the same. For the gas attributes, the unit is litre while the unit of scraps is ton. In the chemical process domain, it is necessary to standardise the units. However, if we shrink all the instances into the range from 0 to 1, the units are no longer a problem as the unit transfer is a linear process. Moreover, it can minimise the influence of the value to the model building process.

5. Model Building

5.1. The Development of Deep Learning Model

The neural network is a computing system which can process information by their dynamic state response to external inputs. It consists of numerous simple and highly interconnected processing elements (i.e. neuron) [21]. A neural network is organised in layers, which are made up of numerous nodes. Besides the input and output layers, the layers between them are called hidden layer. In deep learning field, there are several different types of the hidden layer which include the full connected layer, convolutional layer, pool layer, recurrent layer, etc. For example, the conventional layer is used to deal with the image data and the recurrent layer is used to deal with the data which is time domain. For the nodes, each of them uses an activation function to produce the output. The active functions include Sigmoid, ReLU, Tanh, Linear, etc. When different values are transmitted to the neural network, the input will be given a unique weight. Different values times their unique weight will be added together with a bias. The next step is that the sum will be imported into the activation function. The output will be transmitted to the next neuron [22].

Python is a software which can connect to a lot of the third-party software or programming languages and these are the reason why Python is widely used [23]. In this case, Keras, an advanced deep learning package of Python, will be used to develop the energy consumption prediction model [24].

The deep learning model we used, in this case, is deep Neural Network. For a neural network, the required elements are layers, neurons, activation function, loss function and optimiser. Those elements can be used directly in Keras and Keras Documentation has a detailed introduction on how to choose the proper elements. The data used for this paper is numerical and the general purpose is regression. Hence, the full connected layer– Dense layer in Keras can meet the requirement. ReLU is a prevailing activation function which produces 0 for the negative input and the same value as input for the positive input can be used as the activation function for the input and hidden layers [24]. Optimiser is used to solve the gradient descend problem. Nowadays, there are various optimisers can be used for deep learning such as Adadelta, RMSprop and Adam [25]. Adam has been believed to has excellent performance [25, 26]. Hence, Adam is selected as optimiser in model building. The loss function is used for measuring the compatibility between the actual value and the prediction value. The result which come from the function is called loss. For the regression model, the mean square error can be used to represent the loss. The computation expression is:

$$Loss = \frac{1}{n}(X_a - X_p)^2 \quad (1)$$

Where X_a is the actual value, X_p is the prediction value and n is the number of data.

For the parameters of the model parameters such as the number of hidden layers and node, learning rate and epsilon. Learning rate is a parameter to determine the weights in the direction of the gradient for a mini-batch, and epsilon is a fuzzy factor which is used to avoid the situation of the dividing zero mistakes. Most of them can be set according to Keras Documentation [24]. However, for the number of hidden layer and node, there is no recommendation. If the number of hidden layers and node is not sufficient, the capability of the neural network cannot meet the requirement. If the number of hidden layer and node are too much, the model building time may be too long. After the experiment, the number of hidden layers is set as 4 and the number of the node is set as 500. Fig. 2 shows the structure of deep learning model.

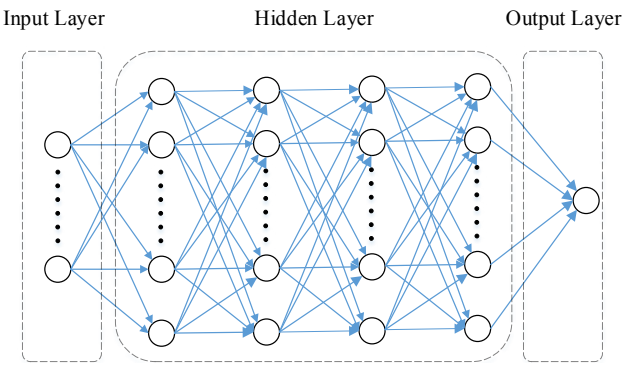

Fig. 2. The structure of deep learning model

5.2. The Development of Contrasting Machine Learning Model

In order to find out the performance of the deep learning model, the contrasting model is needed. In this case, three machine learning algorithms, which are linear regression (LR), support vector machine (SVM) and decision tree (DT) will be used as the contrasting model. Those models will use the same input and output features like the deep learning model. Python Scikit-learn package offers various machine learning functions. The first step is to set up the input of the models. Then different algorithms which are LR, SVM and DT will be called in Python. The Final step is to set relevant training and output parameters.

Linear regression is a method that uses data to estimate the linear function parameters. It is used for finding the linear relation between variable y and one or more variables x [27]. SVM can build the model by marking belonging to one or the other of two categories from the dataset, then decide the new instance into the previous setting categories. SVM can efficiently carry out the nonlinear classification and regression mission [28]. DT is decision support tool which can generate a tree-like model of decision and consequence [29]. All the three algorithms have their advantages. They are good at dealing with the linear, nonlinear and decisional problems respectively.

6. Evaluation

6.1. The Index and Method Used for Evaluation

For numeric classes, it is not possible to calculate the accuracy of the algorithm based on correctly or incorrectly classified instances as there are no defined classes. Hence the performance is calculated by other indicators. In this case Model Correlation Coefficient, Mean Absolute Error, Absolute Error distribution and Maximum Error will be used as the metrics to evaluate the performance of the prediction models.

For the Model Correlation Coefficient, the mathematical expression of the model correlation coefficient is:

$$\text{Model Correlation Coefficient} = \frac{S_{PA}}{\sqrt{S_P S_A}}, \tag{2}$$

where,

$$S_{PA} = \frac{\sum_i (p_i - \bar{p})(a_i - \bar{a})}{n-1}; \quad S_P = \frac{\sum_i (p_i - \bar{p})^2}{n-1};$$

$$S_A = \frac{\sum_i (a_i - \bar{a})^2}{n-1};$$

In the formula, p_i is the prediction value and $\bar{p}$ is the average of the prediction value. a_i is the actual value and the $\bar{a}$ is the average value. n is the number of the training data.

The Mean Absolute Error can be used to calculate the average of the absolute error for all the instances divided by the number of the data. The mathematic expression of Mean Absolute Error is:

$$\text{Mean Absolute Error} = \frac{|p_1 - a_1| + \ldots + |p_n - a_n|}{n} \tag{3}$$

For the Maximum error and the Absolute Error distribution, they can be obtained by simple statistical methods.

In order to get the comprehensive performance of the models, *K*-fold validation is used. The dataset will be separated into *K* parts and each time used *K*-1 parts as training set and the last part as the testing set. Repeatedly training the model for K times and changing the testing set every time will ensure the full use of the data. For operation convenience, the K will be set as 5. The model will be trained five times by using different testing sets, which takes 20% of the dataset.

6.2. The Comparison of Deep Learning Model and the Contrasting Machine Learning Model

Table 2 shows the comparison of Model Correlation Coefficient and Mean Absolute Error for deep learning and the contrasting models. For the result analysis of the deep learning model, the Model Correlation Coefficient is up to 0.854. The Mean Absolute Error of the deep learning model is 1.5Mwh, with the average value of the actual instance is 54.4Mwh. Fig. 3 shows the scatter diagram of actual and prediction instance of deep learning Model. It can be seen from the diagram that most of the points are situated in the diagonal line. The points in the diagonal line are the actual value equals to the predicting value. Only a few points are located far away from the diagonal line. Those errors may come from the impurities of the dataset.

Table 2. The comparison of Model Correlation Coefficient and Mean Absolute Error for deep learning and the contrasting models.

	DL	LR	SVM	DT
Model Correlation Coefficient	0.854	0.785	0.762	0.775
Mean Absolute Error	1.5	2.1	1.7	1.9

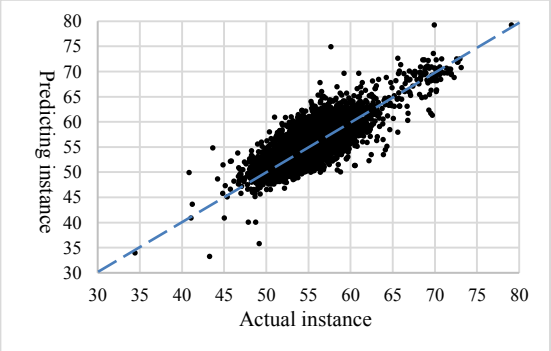

Fig. 3. The diagram of actual and prediction instance of deep learning model

For the contrasting models, the Model Correlation Coefficient of the contrasting models are 0.785 for LR model, 0.762 for SVM model and 0.775 for DT model. The model correlation coefficient of LR, SVM and DT is evidently lower than the coefficient of the deep learning model.

The Mean Absolute Error of LR, SVM and DT model are 2.1Mwh, 1.7Mwh and 1.9Mwh respectively. None of the errors of the contrasting models is smaller than the error of deep learning model which is 1.5Mwh.

Table 3 shows the comparison of the deep learning and contrasting models' Absolute Error distribution. In the error range of 0.5 to 1 and under 0.5, the deep learning model does not show advantage. The instances in deep learning model which Absolute Error higher than 3Mwh are 10.6%, which is dramatically smaller than the ratio of LR, SVM and DT is

22.3%, 14.9% and 19.9% respectively. It indicates that only a small part of the predicting values in the deep learning model have a large error which is over 3Mwh. The maximum Absolute Error of deep learning model is 17.36Mwh, which is evidently smaller than 28.23Mwh of LR model, 26.22Mwh of SVM model and 29.49Mwh for DT model. It is evident that most of the Absolute Errors of deep learning model are situated in the low range.

Table 3. The comparison of Absolute Error distribution for the deep learning (DL) and contrasting models

Error(Mwh)	DL	LR	SVM	DT
0—0.5	2407	2257	2393	2499
0.5—1	2232	1993	2171	2239
1—3	5182	4289	4792	4060
>3	1169	2451	1634	2192
Maximum Error	17.36	28.23	26.22	29.49

In this case, the Model Correlation Coefficient and Mean Absolute Error of deep learning model is evidently better than that of the conventional machine learning model. The Absolute Error distribution indicates that deep learning model produces the higher percentage of instances which Mean Absolute errors are situated in the low range. Therefore, deep learning model shows advantages in different metrics.

7. Conclusion

In nowadays steel plant, the existing statistical approach to solve the data problem had its limitations. This paper suggests a data mining-based approach to solve the energy prediction problem at an established steel plant. The data collected from the steel plant is well pre-processed with the feature selection being carried on. Several machine learning algorithms are developed to build the prediction model. The performance evaluation of the models is based on 5-fold cross validation and carried out by comparing several metrics. Among those models, the deep learning model presents excellent performance. In the future, the model developed in this paper can serve as an effective tool to predict the daily energy consumption of the EAF. It is evident from the experiment that the data mining techniques have the powerful ability in tackling the problem with regards to big data.

References

[1] Singh RK, Murty HR., Gupta SK, Dikshit AK. Development of compositesustainability performance index for steel industry. Ecological Indicators.2007; 7(3): 565-588.
[2] Fornai B, Vannucci M, Colla V, Amato A, Porzio FG, Bjork, A., and Westling K. A software tool for comparing the economic value of different scrap mixes in steel production. ISMS. 2013:315–320.
[3] Delgado C, Ferreira M, and Castelo Branco M.The implementation of lean Six Sigma in financial services organizations. Journal of Manufacturing Technology Management.2010; 21(4) :512–523.
[4] Pardo N., Moya, J. A., and Vatopoulos, K. , Prospective scenarios on energy efficiency and CO2 emissions in the EU iron & steel industry.1st ed. 2012.
[5] Ian H. W，Eibe F，Mark A. H，Christopher J. P. Data Mining: Practical Machine Learning Tools and Techniques.1st ed. Cambride. 2017.
[6] Yann L, Yoshua B and Geoffrey H. Deep Learning. Nature. 2015; 521:436-438.
[7] Richard D. V, Roger W. H & Ronald D. S. Big data and the missing links. Statistical Analysis and Data Mining. 2016; 9(6): 411-416.
[8] Swartz. E and MacRosty M. Dynamic optimization ofelectric arc furnaceoperation. American Chemical Society.2007; 53(3): 640–653.
[9] Snell J.J. Improved modeling and optimal control of an electric arc furnace. Dissertations & Theses – Gradworks.2010. 1-88.
[10] MacRosty M .R and Swartz E.C. Dynamic optimization of electric arc furnace operation. Process Systems Engineering. 2007;53 (3) : 640-653.
[11] Camdalı U and Tunc M. Modelling of electric energy consumption in the AC electric arc furnace. International Journal of Energy Resarch. 2002; 26(10): 935–947.
[12] Aleis J, Ramirez M, Trapaga G and Jonsson P. Modeling of a DC Electric Arc Funace—Heat Transfer from the Arc. ISIJ International.2000; 40(11):1089-1097.
[13] Köhle S. Recent improvements in modelling energy consumption of electric ac furnaces. 7th European Electric Steelmaking Conference. 2002; 1305-1314.
[14] Fena´ndez J, Cabal V, Montequin V and Balsera J. Online estimation of electric arc furnace tap temperature by using fuzzy neural networks. Engineering Applications of Artificial Intelligence.2008;21(7):1001–1012.
[15] Kirshen M, Risonarta V and Pfeifer H. Energy efficiency and the influence of gas burners to the energy related carbon dioxide emissions of electric arc furnaces in steel industry. Energy. 2009; 34(9) : 1065–1072.
[16] Ahmad I, Kano M, Hasebe S, Kitada H and Murata N. Gray-Box Modeling for Prediction and Control of Molten Steel Temperature in Tundish. Journal of Process Control.2014; 24 (4) :375-382.
[17] Zillner S, Ebel A, Schneider M. Towards intelligent manufacturing, semantic modelling for the steel industry. IFAC-PapersOnLine. 2016 Jan 1;49(20):220-225.
[18] Kovačič M, Šarler B, Genetic programming prediction of the natural gas consumption in a steel plant. Energy.2014; 66:273-284.
[19] Weka 3: Data Mining Software in Java. [online] Available at: https://www.cs.waikato.ac.nz/ml/weka/ [accessed 20 Dec.2017].
[20] Sabzevari H, Soleymani M and Noorbakhsh M. A comparison between statistical and Data Mining methods for credit scoring in case of limited available data. Proceedings of the 3rd CRC Credit Scoring Conference; 2007. 1-5.
[21] Caudill M. Neural networks primer. part I, AI Expert. 1987;3(6): 53-59.
[22] Woodford C. How neural networks work - A simple introduction. [online] Available at: http://www.explainthatstuff.com/introduction-to-neural-networks.html.2017 [Accessed 8 Oct. 2017].
[23] Dubois FP. Python: Batteries Included. Computing in Science & Engineering .IEEE/AIP. 2007. 5-27.
[24] Keras Documentation. [online] Available at: https://keras.io/ [Accessed 20 Dec. 2017].
[25] Ruder S. An overview of gradient descent optimization algorithms.2016. 1-14.
[26] Kingma PD and Ba J. Adam: a Method for Stochastic Optimization. International Conference on Learning Representations.2015; 1-13.
[27] Seal LH. The historical development of the Gauss linear model. Biometrika.1967; 54 (1/2): 1–24.
[28] Nello C and John S. An Introduction to Support Vector Machines and other kernel-based learning methods. Cambridge University Press.2000; 1-173.
[29] Kamiński, B, Jakubczyk and Szufel P M. A framework for sensitivity analysis of decision trees. Central European Journal of Operations Research. 2017; 1 : 1-25.