Research on blast furnace stability evaluation model based on neural networks using dynamic sampling approach

Guolei Jiang, Linying Jiang * and Cuicui Shang

Software College, Northeastern University, Shenyang, Liaoning, 110004, China
*Corresponding author’s e-mail: jiangly@swc.neu.edu.cn

Abstract. The blast furnace iron-making is a complicated physical and chemical reaction and heat exchange process in a closed shaft furnace. The blast furnace’s operator must judge the condition of the furnace in time and accurately, and then take the adjustment measures based on the condition to ensure that the blast furnace is running smoothly. Not handling the fluctuation of the furnace condition in time, it may cause the failure of the furnace, which leads to the decrease of the blast furnace’s output, the increase of fuel consumption, and a serious effect on the production and operation of the enterprises. In order to solve the above problem, the study implement several methods to try to judge the stability of the blast furnace in the iron-making process. Results shows that the Neural Networks with dynamic sampling method, which has a great superiority among the class imbalance problems, has the highest anomaly detection rate compared to the other methods.

1. Introduction
Nowadays, the Blast Furnace Iron-making (BFI) is the most popular iron-making technology, and the mostly used method to judge the stability is manual detection. It is fairly important to adjust the furnace’s conditions when it is about to go abnormal, which may otherwise greatly reduce the yield and quality of the pig iron produced by BFI. Besides, if the operator could not judge the furnace’s conditions in time, quite a lot of raw materials would be wasted, given that the actual furnace conditions were abnormal.

For a long time, there have been problems existing in the BFI which have already resulted in terrible damages to the producing and operation of iron-making companies. These problems comprise experience over data, lack of uniform indicators of furnace conditions, fuzzy judgement, overdependence on personal experience, etc[1].

Based on these situations, it seems quite important to come up with a more accurate new method to detect the iron-making anomalies other than just by manual detection. By establishing a blast furnace stability evaluation model with a high anomaly detection rate, the company can save quite a lot economic cost due to the reduction of material waste and the improvement of yield rate.

2. Methods
In this section, we will introduce the feature extraction method for the original data, the main classification method for the processed data, as well as the evaluation metrics used in this paper.
2.1. Feature extraction algorithm - Relief

After the preprocessing of the original datasets, there are 461 features in total, but not all of them have a positive effect on the classification results. Therefore, a filter feature selection method called Relief (Relevant Features)[2] is chosen in this paper to extract features of the BFI datasets.

The Relief method designs a “correlation” to measure the importance of features. The correlation is a vector of which each component corresponds to an initial feature. The importance of a feature subset is determined by the sum of each feature’s corresponding correlation components in the subset.

Since the Relief has been proved to be very effective in many aspects, this paper will not discuss more details about it as you can find them in [3]. Relief is especially designed for binary classification problems. The expansion version, Relief-F, can solve multiclass problems[4].

During the experiments, we will do some tests to see if the Relief algorithm can actually improve the classification results of the classifier, and if it does, we will try to choose a proper number of features to be extracted to increase both the performance and efficiency of the classifiers.

2.2. Dynamic sampling method for MLPs

Multilayer perceptrons (MLPs) have been proved to possess great abilities to learn complex classification boundary and they can be directly used in binary classification problems. Therefore, MLP is adopted as our basic model and a data selection method called dynamic sampling (DyS) is developed for MLP to deal with class imbalance issues.

The basic idea of DyS is to select examples for training dynamically in the training process. We do not pre-delete any example to prevent information loss; however, we dynamically select examples for training to avoid redundant information and to make the best use of the training data.

The general flow of the DyS-MLPs algorithm can be seen in figure 1.

![Figure 1. Flowchart of the Dynamic sampling method for MLPs.](image)

Unlike the pre-sampling methods, DyS integrates the sampling and training processes, which can solve some drawbacks of a pretreatment procedure. More details of the Dynamic sampling method can be found in [5].

2.3. Evaluation metrics

In class imbalance problems, the overall classification accuracy does not work as a good metric for measuring the performance of the classifier. This metric overlooks the low recognition rate of the minority class in the violation of the iron-making company’s purpose—pursuing as high an anomaly detection rate as possible. The anomaly detection rate is known as the Sensitivity (also called the true...
positive rate or Recall), which corresponds with the Specificity (also called the true negative rate). These two metrics can be calculated by (1) and (2).

\[
\text{Sensitivity} / TPR = \frac{TP}{TP + FN} \quad (1)
\]

\[
\text{Specificity} / TNR = \frac{TN}{TN + FP} \quad (2)
\]

Furthermore, to evaluate the binary classification performance in detail, the geometric means (G-mean) of the classification accuracy of the Sensitivity and Specificity will be employed as the main metric in our experimental study. G-mean is defined as (3)[6].

\[
G\text{-mean} = \left( \frac{TP}{TP + FN} \times \frac{TN}{TN + FP} \right)^{1/2} \quad (3)
\]

3. Model design
After preprocessing the origin data and before further experiments, the whole datasets are separated into training, validation, and test sets in the ratio of 7:1:2. In this section, all parameters of the model are designed or optimized by the results on the validation set.

3.1. Data labeling
The data provided by the iron-making company do not have direct labels, but supervised learning algorithms such as the Neural Networks need labels to train models. The useful data we have are the overall times of material hanging, slipping, and collapsing in the furnace through each day’s iron-making process.

Formula (4) shows the computational formula of the antegrade index inspired by [7] where ‘a’ is the count of material hanging, ‘b’ is the count of material slipping, and ‘c’ is the count of material collapsing in each piece of data.

\[
f = \frac{1}{1 + 0.95 \times a + 0.02 \times b + 0.03 \times c} \quad (4)
\]

After the calculation, this formula will get a numerical value between 0 and 1, defined as the antegrade index, which can measure the conditions during iron-making. Furthermore, this paper will label the value into a negative example (denoted by 0) if the value is greater than 0.89 or a positive anomaly example (denoted by 1). At the end of the data labeling, there are totally 2205 examples, including 634 positive examples and 1571 negative examples.

The coefficient combination (0.95, 0.02, 0.03) in Formula (4) and the threshold value 0.89 are determined based on the clustering result of an unsupervised K-Means model. Figure 2 shows four different coefficient combinations and their best thresholds as well as the labeling accuracy on the training set. It turns out that the coefficient combination (0.95, 0.02, 0.03) and the threshold value 0.89 result in a best accuracy of 99.55% on training set, 99.61% on validation set, and 99.22% on test set.
3.2. Feature extraction
As mentioned before, there are totally 461 different features in the preprocessed iron-making datasets and it is necessary to choose a proper number of features for further training because too many features will slow down the learning speed of algorithm, not to mention an MLP model which has many layers or hidden units. In addition, some trivial features may deteriorate the final classification performance of the classifiers, so we had better remove these features.

From figure 3 we can see that the Sensitivity and G-mean tend to become stable after 150 features, approximately, and start to decline after about 220 features selected based on the Relief algorithm. The Specificity stays around 92% after 100 features are selected. Therefore, we will choose the first 200 features of the whole datasets with higher weights, and this number of features has been proved to improve both the performance and speed of other classifiers as well on the validation set.

3.3. Optimization of the MLP model
As is known to us all, neuron networks have many hyper parameters which can affect their learning speed and final classification performance. We can determine some parameters by experience, but the best way is to optimize them by specific validation results. Figure 3 and figure 4 show the classification results of some different parameters combinations on the validation set.

From figure 4, we can easily see that a two-hidden-layers MLP model always outperforms a one-hidden-layer MLP model with the same number of neurons in the first hidden layer. Further experiments on a three-hidden-layers MLP model show that it cannot improve the G-mean value a lot and will slow down the training, so we will use a two-hidden-layers MLP model as our final choice.
As for the number of the neurons in each hidden layer, since too few neurons cause a poor classification performance while too many neurons cause overfitting, it seems that 150 units in the first and 70 units in the second hidden layer is quite a reasonable choice from the validation results shown in figure 4 and figure 5.

Besides, in our MLP model, the SGD optimizer may sometimes run into a local optimization solution, so we take the ADAM optimizer instead. Also, using the sigmoid activation function will slow down the training a lot, so we use the rectification (ReLU)[8] non-linearity in all layers (except the final soft-max layer).

4. Experiments and analysis

For a binary classification problem, there are many mature models nowadays which can be considered such as the logistic regression model (LR) or SVM. MLP (without DyS) will also be adopted to see the effectiveness of the dynamic sampling approach. In our experiments, we will compare these models with the DyS-MLP model, and just to mention that, all hyper parameters of these models have been optimized based on the validation results just like the optimization of the MLP model in Section 3.3.

| Model  | Sen  | Spe  | G-Mean |
|--------|------|------|--------|
| LR     | 0.5714 | 0.9110 | 0.7215 |
| SVM    | 0.6000 | 0.8813 | 0.7272 |
| MLP    | 0.7048 | 0.8510 | 0.7745 |
| DyS-MLP | 0.7467 | 0.8301 | 0.7876 |

Table 1 shows values of classification metrics of these classifiers over the testing set after training. To be mentioned, the metric values of the MLP and DyS-MLP are mean values of five separate results in the aim of error reduction.

It’s quite clear from Table 1 that the DyS-MLP model has the best performance on Sensitivity while a slightly worse performance on Specificity. In other words, it improves the anomaly detection rate at the cost of a relatively bad normal detection rate. However, the DyS-MLP model still has the highest G-mean value and it improves the Sensitivity quite a lot compared to the other models.
5. Conclusion
In this paper, by combining the MLP model with the dynamic sampling method, the anomaly detection rate of the blast furnace iron-making can reach nearly 75%, which is a great improvement compared to other classifiers. However, due to the lack of iron-making data we have collected, the neural network has not reached its best performance.

Also, to be mentioned, the DyS was proposed to solve multiclass imbalance problems at first and was implemented with an oversampling process at the beginning of every epoch to avoid the class bias. Nevertheless, in this case, the imbalance ratio was not so big that we did not implement the oversampling process. In the future study, we will try to collect more data and apply more details to try to improve the classification rate of the DyS-MLP model.

References
[1] Chen, W. (2016) Application and Practice of Antegrade Physical Examination System in Masteel Blast Furnace. Journal of Anhui Vocational College of Metallurgical Science and Technology, 27(2): 89-93
[2] Zhou, Z. (2016) Machine Learning. Tsinghua University Press, Beijing.
[3] Kira, K and Rendell, L.A. (1992) The feature selection problem: Traditional methods and a new algorithm. In: Proceedings of the 10th National Conference on Artificial Intelligence (AAAI). San Jose. pp.129-134.
[4] Kononenko, I. (1994) Estimating attributes: Analysis and extensions of RELIEF. In: Proceedings of the 7th European Conference on Machine Learning (ECML). Catania. pp.171-182.
[5] Lin, M., Tang, K. and Yao X. (2013) Dynamic sampling approach to training neural networks for multiclass imbalance classification. IEEE Transactions on Neural Networks and Learning System, 24(4): 647-660.
[6] Sun,Y., Kamel, M.S. and Wang, Y. (2006) Boosting for learning multiple classes with imbalanced class distribution. In: Conference on Data Mining. Las Vegas. pp.592-602.
[7] Liu, Y. (1998) Recommended to use the anterograde index to analyze the forward status of the blast furnace. Ironmaking, 17(6): 50-51.
[8] Krizhevsky, A., Sutskever, I. and Hinton G. (2012) ImageNet Classification with Deep Convolutional Neural Networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe. pp.1097-1105.