Knowledge-Based Systems 101 (2016) 48–59

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Molten steel temperature prediction model based on bootstrap Feature Subsets Ensemble Regression Trees

Xiaojun Wang[a,*], Ping Yuan[b], Zhizhong Mao[b], Mingshuang You[b]

[a] Key Laboratory of Advanced Design and Intelligent Computing, (Dalian University) Ministry of Education, Dalian, 116622, China
[b] Institute of Automatization, Northeastern University, Shenyang 110004, China

ARTICLE INFO

Article history:
Received 28 April 2015
Revised 24 February 2016
Accepted 25 February 2016
Available online 19 March 2016

Keywords:
Ladle furnace
Molten steel temperature prediction
Large-scale data and noise data
Ensemble method

ABSTRACT

Molten steel temperature prediction is important in Ladle Furnace (LF). Most of the existing temperature models have been built on small-scale data. The accuracy and the generalization of these models cannot satisfy industrial production. Now, the large-scale data with more useful information are accumulated from the production process. However, the data are with noise. Large-scale and noise data impose strong restrictions on building a temperature model. To solve these two issues, the Bootstrap Feature Subsets Ensemble Regression Trees (BFSE-RTs) method is proposed in this paper. Firstly, low-dimensional feature subsets are constructed based on the multivariate fuzzy Taylor theorem, which saves more memory space in computers and indicates “smaller-scale” data sets are used. Secondly, to eliminate the noise, the bootstrap sampling approach of the independent identically distributed data is applied to the feature subsets. Bootstrap replications consist of smaller-scale and lower-dimensional samples. Thirdly, considering its simplicity, a Regression Tree (RT) is built on each bootstrap replication. Lastly, the BFSE-RTs method is used to establish a temperature model by analyzing the metallurgic process of LF. Experiments demonstrate that the BFSE-RTs outperforms other estimators, improves the accuracy and the generalization, and meets the requirements of the RMSE and the maximum error on the temperature prediction.

1. Introduction

Ladle furnace (LF) steel refining technology plays a substantial role in secondary metallurgic process. When LF is taken over at downstream secondary metallurgy units or at a continuous caster, the main purpose of LF refining processing is to get the qualified molten steel temperature and composition [1–5]. In practice, the molten steel temperature cannot be measured continuously, making it difficult to realize the precise controlling. Therefore, it is important to build a precise molten steel temperature prediction model.

Several molten steel temperature models of LF based on thermodynamics and conservation of energy are developed in earlier works. [6] “However, these models cannot be used efficiently for online accurate prediction because the parameters are hard to obtain. It is attributed to the harsh operating environment of ladle metallurgy, especially the high temperatures and corrosive slag associated with the process. Practically, some of the parameters are estimated according to experience. Consequently, it is hard to ensure the accuracy of mechanism models” [1].

To overcome the limitations of mechanism models, various data-driven modeling methods have been applied to establish the temperature prediction models. For example, Sun et al. [7] build a temperature model based on the Neural Networks (NN). Later, a PLS-SVM based temperature model is proposed by Wang [8], where input variables are firstly dealt with Partial Least Squares (PLS) algorithm to get rid of the linear dependency and then Support Vector Machine (SVM) is utilized to establish the prediction model. However, these temperature models only learn from a single model, and their performance are hard to improve when the object is complicated or the samples are with high noise [9].

Compared with a single model, an ensemble model can further improve the accuracy and the generalization [10,11]. In the past decade, ensemble methods have been applied to establish temperature models. Tian & Mao [1] propose to establish an ensemble temperature prediction model based on a modified AdaBoost.RT algorithm. Lv et al. [9] propose a hybrid model in which the pruned Bagging model based on the negative correlation learning is used to predict the unknown parameters and the undefined function of the thermal model simultaneously.

Today, the accuracy and the generalization of most existing molten steel temperature models cannot satisfy industrial production. They have been estimated on small-scale data. With the

* Corresponding author. Tel.: +15040150903.
E-mail address: wxjjessicaxj0903@126.com (X. Wang).

http://dx.doi.org/10.1016/j.knosys.2016.02.018
0950-7051/

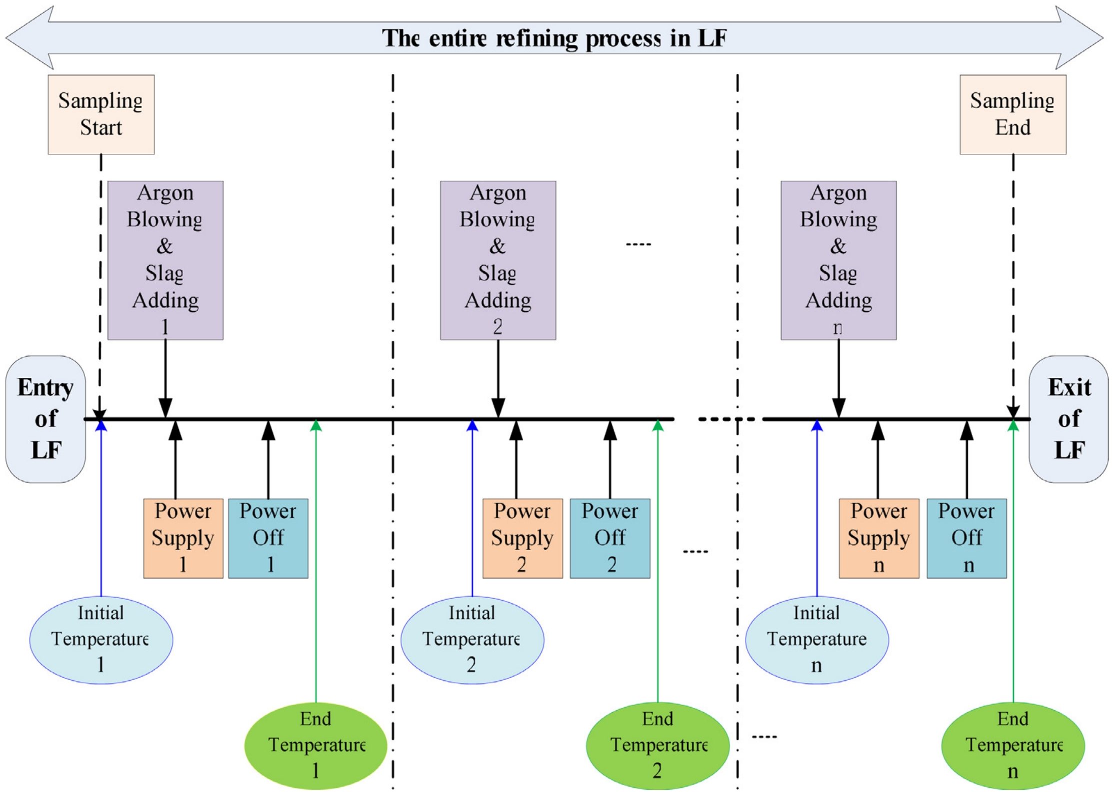

Fig. 1. The diagram of the entire refining process in LF.

development of the information and the computer techniques, large-scale data are accumulated from the production process in LF. They contain more useful information, and make it possible to improve the accuracy and the generalization of the temperature prediction.

However, large-scale data impose more restrictions on modeling and increase the costs of building models [12]. Furthermore, high complexity models increase the computational burden in the phase of actual use [13] and cannot be used efficiently for online accurate prediction. The existing ensemble temperature models are not suitable for large-scale data. In [1], the Extreme Learning Machine (ELM) [14] is selected as the base learner and the modified AdaBoost.RT is employed as the ensemble structure. Although the learning speed of the ELM is extremely fast, the AdaBoost.RT is a serial ensemble in which every new sub-model relies on previously built sub-models. In [9], the pruning process of sub-models is also one by one, and Bagging is changed into a serial ensemble. A serial ensemble is often more complex than a single model [15], especially on large-scale data. Additionally, data sampled from the production process are with noise, which reduces the accuracy and the generalization of the temperature prediction.

To deal with the large-scale and the noise issues, the Bootstrap Feature Subsets Ensemble Regression Trees (BFSE-RTs) method is proposed in this paper. Firstly, low-dimensional feature subsets are constructed based on the multivariate fuzzy Taylor theorem [16], which saves more memory space in computers and indicates "smaller-scale" data sets are used. Secondly, to eliminate the noise data, the bootstrap sampling approach [17] of independent identically distributed data is introduced into the feature subsets. Bootstrap replications consist of smaller-scale and lower-dimensional samples. Thirdly, considering its simplicity the Regression Tree (RT) [18] is employed as the base learner, and a RT is built on each bootstrap replication. The BFSE-RTs method is expected to successfully utilize the large-scale data accumulated from the production process in LF, to improve the accuracy and the generalization of the temperature prediction, and to meet the requirements that the Root Mean Square Errors (RMSE) of the temperature is less than 3 °C and the maximum error of it is less than 10 °C.

The remainder of this paper is organized as follows. Section 2 introduces the mechanism of the production process of LF. In Section 3, existing ensemble methods are reviewed. In Section 4, the BFSE-RTs method is proposed, and the differences from other ensembles are introduced. In Section 5, the experimental investigations are brought out, and the BFSE-RTs temperature model is compared with the FSE-RTs, the RF, the stacking trees, the modified AdaBoost.RT, and the pruned Bagging temperature models. In Section 6, the conclusion of this paper is summarized.

2. The mechanism of production process of LF

To establish the data-driven based model of the temperature in LF, the energy change during the refining process in Fig. 1 is considered [1,6,14].

The refining process starts from the entry of LF and ends at the exit of LF. The data are sampled in the entire refining process that is subdivided into many different sampling periods. In any sampling period, the same steps are executed, and they are the argon bowing and slag adding, the power supply, and the power off.

In any sampling period,

(i) The initial temperature is measured before the argon bowing and slag adding.
(ii) The power supply of the refining process in LF is from the electric arc.
(iii) The power off mainly contains three parts: ① The heat loss from the ladle refractory wall and the top surface. In this part, the heat loss is relatively stable, increases as the time goes on, and may be reflected by refining time. ② The heat