Publishing House
AKAPIT

COMPUTER METHODS IN MATERIALS SCIENCE

Informatyka w Technologii Materiałów

Vol. 11, 2011, No. 1

TEMPERATURE PREDICTION IN ELECTRIC ARC FURNACE BY THE USE OF DECISION TREES

TADEUSZ WIECZOREK*, KRYSTIAN MĄCZKA, KRZYSZTOF GANSTY

Department of Management and Computer Science, Faculty of Material Engineering and Metallurgy, The Silesian University of Technology, Krasinskiego 8, 40-019 Katowice, Poland
**Corresponding author: tadeusz.wieczorek@polsl.pl*

Abstract

Decision trees are one of the computing intelligence methods which proved to be very reliable as far as solving complicated multidimensional problems is concerned. Therefore, these methods are often used for extracting rules and to predict variables, what makes them useful for production automation. In this paper authors discuss the possibility of the use of decision trees for electric arc steelmaking process. The main goal is to predict temperature in the electric arc furnace by the use of decision trees. Proper automatic temperature prediction may reduce the number of temperature measurements during the process and consequently, it may shorten the time of the process. Optimization of production processes leads to real benefits, which is, for example, lowering costs of production. Calculations were done by the use of six types of regression decisions trees available in Statistica Data Miner software. The algorithms were examined considering the minimum error rate of temperature prediction, but also less complicated tree structure. The structure of a decision tree is also important owing to computational complexity.

Key words: friction electric arc furnace, temperature prediction, decision trees

1. INTRODUCTION

The need to solve very complex practical issues which are hard to describe by the use of conventional mathematical methods causes that more advanced methods and calculating techniques, especially methods of computing intelligence (CI) have aroused researchers' interest. The use of such intelligent methods in industrial issues has become more popular due to their positive results. The progress in the field of computerization has caused a significant development of both theory and practical actions of CI methods. Moreover, these methods have grown in popularity due to their quick learning ability, their usefulness and higher accuracy than accuracy obtained by the conventional methods. Nowadays, artificial intelligence systems have become a well-developed discipline, which is applied in many branches of technology.

One of the methods of Computational Intelligence is so-called *decision trees*. It is hard to establish the origin of the conception of representing knowledge by the use of decision trees. However, this method is not new and moreover, it does not originate from researches over the artificial intelligence. Decision trees appeared in the professional literature in the context of social researches in 1963. They were used in statistical researches in 1984 and afterwards decision trees became a part of machine learning and gained popularity by dint of Quinlan's works e.g. (Quinlan, 1993). Undoubtedly, decision trees are at present the most popular and the most efficient method of Computational Intelligence and they are used for solving prediction problems, ex-

 ISSN 1641-8581

plaining data structure, extracting knowledge from data and also for explaining models obtained from the other methods (Wieczorek, 2008).

The aim of this paper is to analyze and compare the influence of decision trees parameters on accuracy of decision-making on the basis of industrial data. The investigation includes the analysis of possibilities of the use of the obtained model (decision tree) in real industrial process. Moreover, the aim of this work is to analyse the most popular decision trees algorithms (the analysis of parameters that influence on prediction accuracy), compare them and draw conclusions. Calculations were done in *Statistica Data Miner* software.

The paper is organized as follows: section 2 formulates the problem, section 3 introduces the main idea of decision trees algorithms and describes data used for computation. In section 4 we present the results and section 5 contains the final conclusions and recommendations.

2. FORMULATION OF THE PROBLEM

The electric arc steelmaking process is usually carried out in three stages: the melt process (EAF), the ladle furnace process (LHF or VAD) and the casting process (the second and the third stage of the process will not be considered in the paper). Scrap is melted in electric arc furnaces which serve only as melting units because liquefying must be done in the shortest time possible. Scrap, alloys and fluxes are charged into the furnace with baskets or by means of charging barrows. After charging, the furnace is closed to begin the meltdown stage. The liquefying of the charged materials is performed by electric energy but this stage of the process might be supported by the use of wall and door burners in order to deliver energy to areas which cannot be reached by the electric arc. This stage of the process is repeated by discharging additional baskets into the furnace and melting them until enough liquid steel is available. After meltdown, a sample is taken from the furnace and the temperature is measured. The steel sample is analyzed in the laboratory and according to the results of the analysis, some fluxes and alloys are added into the furnace or into the tapping stream during the tapping process. Afterwards the liquid steel is heated up to the required tapping temperature and additional metallurgical tasks, such as decarburization by oxygen blowing or other metallurgical operations (which depend on the steel grade to be produced) are performed. Before tapping steel to the ladle, deslagging of the bath through the opened furnace door is done. After deslagging, the liquid steel is tapped into the steel ladle and the EAF process starts again.

Improving modeling of EAF's parameters, e.g. final temperature of steel, requires a fundamental understanding of the physics of the process, i.e. heat transfer, mass transfer, fluid flow and the electromagnetic phenomena. To achieve this goal, mathematical models based on differential equations have become popular. Previous works relating to this topic discuss only simple models considering basic reactions and phenomena where plane geometry of furnace tank was assumed, what was less complicated (Billings et al., 1979). In 1981, it was Szekely who performed numerical simulations using the Turbulent Navier-Stokes equations, energy conservation equation and Maxwell equations. Calculations were done for both the arc and the bath where a parabolic current density distribution was assumed to simplify the magnetic problem (Alexis et al., 2000). In 1985 Szekely et al. used the magnetic diffusion equation to predict heat transfer and fluid flow in arc welding. In 1992 Choo et al. solved Laplace's equation for the electric potential to determine boundary conditions for a model of the arc. In 1996 Larsen and Bakken used the magnetic transport equation to predict the current and magnetic field in an AC arc. Later works were focused on attempts to describe the whole electric arc steelmaking process and to determine the most important quantities such as composition of gas emissions and composition of charge (Hayman et al., 1997; Cameron et al., 1998). Optimization of the furnace work was also discussed in these works. Denys et al. define conditions of chemical balance of the furnace for particular furnace working conditions (Denys et al., 1997). The basic model of electric arc steelmaking process was constructed by Matson and Ramirez and consists of a model of scrap melting process, chemical balance equations of the process and material and energy balance (Matson et al., 1999). The constructed model was optimized by the use of iterative dynamic programming. Taking into consideration the fact that for the effective temperature control, precise dynamic models are needed, there are many publications related to the modelling of a dynamics of the process or related to electrical equivalent circuit e.g. (Boulet et al., 2003). In 2000 (extended in 2005) Alexis made a 3-dimensional simulation model coupling heating and induction stirring in a ladle arc furnace (Alexis et al., 2000).

All authors, basing on differential and integral equations, tried to use different approaches to predict fluid flow, temperature and heat transfer and electromagnetic field distribution in the area of arc and bath. The methods usually do not include some phenomena occurring (e.g. induced electric field) and assume some simplifications (e.g. current distribution within the arc), what usually leads to decreasing of calculation precision. Classical modeling based on differential or integral equations, equilibrium chemistry and material and energy balances does not lead to the required results. Precise analytical modeling of three, asymmetrical arcs with process-dependent, nonlinear characteristics (such as highly variable reactances) is not possible because a number of nonlinear influences and temporary changes of the process must be taken into account. Therefore, artificial intelligence methods, including algorithms of neural networks, fuzzy sets and decision trees are of considerable interest.

3. ALGORITHMS OF DECISION TREES

Nowadays decision trees have become an essential machine learning method because of their great effectiveness and simple programming implementation. The idea of decision trees can be explained as a set of rules describing relations between features (attributes) and predefined or discovered classes. A number of different methods of constructing decision trees has already been published. Some of them are: Classification and Regression Trees (CART) (Breiman et al., 1984), ID3 and C4.5 (Quinlan, 1993), Separability of Split Value (SSV) Trees, Fast Algorithm for Classification Trees (FACT), Statistical Tree (QUEST), Cal5 (Jankowski & Grąbczewski, 2006). Apart from the conventional rules, there are also systems which are used for building alternative types of rules, for example fuzzy tree algorithm (Ichihashi, 1996). The methods are based on different ideas and use different model structures. Some of them use dichotomic splits (CART, SSV), the others allow the use of more complex branching. Some algorithms assume a particular data distribution and use parametric statistical tests (FACT, QUEST, Cal5), the others make no such assumptions.

Most decision tree algorithms split the nodes depending on the values of a single feature (attribute). Selecting the most proper for the split feature is often closely bound up with the selection of the splitting points (CART, C4.5, SSV). An alternative strategy is applied in FACT, where the split feature is defined as the one that maximizes the value of F statistic (Fisher-Snedecor statistics used in ANOVA method) (Jankowski & Grąbczewski, 2006). In C4.5 algorithm the information entropy approach is used. Information (entropy) of a set of learning examples E is:

$$I(E) = -\sum_{i=1}^{C} \frac{|E_i|}{|E|} \cdot \log_2 \left(\frac{|E_i|}{|E|} \right) \tag{1}$$

where: $|E_i|$ – number of examples which describe i-th class (one of the existing C classes), $|E|$ – number of examples in set E

Expected information value after splitting the set E into subsets $E^{(m)}$, $m = 1,...,|v_f|$, for which feature "f" has got value v_m, is defined as (Quinlan, 1993):

$$I(E,f) = \sum_{m=1,K,|V_a|,E^{(m)}\neq 0} \frac{|E^{(m)}|}{|E|} \cdot I(E^{(m)}) \tag{2}$$

where: $|E^{(m)}|$ – number of examples after splitting the set E for value „m" of feature „f",

The most common split selection criterion is called purity gain or impurity reduction (3). For a split node and for a chosen feature „f", it is defined as:

$$\Delta I(E,f) = I(E) - I(E,f) \tag{3}$$

where I is a node impurity measure.

The impurity measure used in CART algorithm, called Gini index, is different from (1):

$$I_G(E) = 1 - \sum_{i=1}^{C} \left(\frac{|E_i|}{|E|} \right)^2 \tag{4}$$

Maximal fitting of a decision tree to the training data leads to overfitting of the model. The tree becomes large and complicated and some of its leaves just describe a few data vectors (examples). To simplify a tree and make it more general, stopping criteria, crossvalidation and pruning methods can be used. Better generalization is obtained when at first, overfitted tree is obtained and then pruned (Jankowski & Grąbczewski, 2006). However, the pruning criteria are not obvious. Combining cross-validation with good results of pruning can be obtained. Minimizing crossvalidation error leads to stopping tree's growth (Wieczorek, 2008).

The CHAID algorithm is the most popular modification of well-known statistics, the AID method (Automatic Interaction Detection), introduced in 1980. This method allows splitting the set of examples into exhaustive and disjoint subsets. However, splitting does not have to be binary (The CHAID trees can have any number of branches). Index Gini (4) is used as a criterion. Regression CHAID trees are analogical to CART trees.

– Exhaustive CHAID for Regression,
– General CHAID for Regression,
– Boosted Regresion Trees,
– Advanced CHAID Regression.

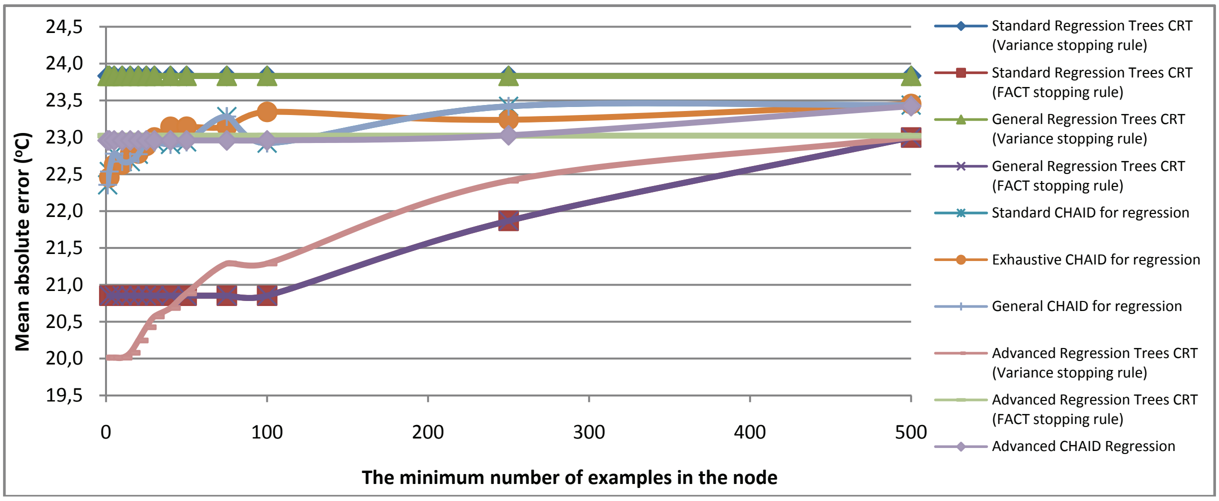

Fig. 1. *Mean absolute error (^{o}C) of predicting the temperature depending on minimum number of examples in a node.*

The boosted trees algorithm developed from the use of statistical method, called boosting. The main aim of this method is to create a sequence of simple trees, from which every next tree is used to predict remainders obtained from the previous ones. This method creates binary trees. It can be proved that this procedure, called *additive weighted extension*, leads to a appropriate adjustment of predicted values to the observed ones, even if the nature of relation between predictors and dependent variable is very complicated (for example non-linear one).

The data used for the calculations presented in this paper comes from the real industrial process. The aim of the calculations was to predict the final temperature of the EAF process basing on the following features:

– weight of charge of particular baskets,
– electric energy used while melting particular baskets,
– electric energy used during reheating period,
– energy used during the whole process (including gas burners),
– the amount of carbon inserted to the furnace during the process,
– temperature measured on the furnace shell and in some chosen areas of the furnace tank.

Owing to the fact that temperature prediction is a regression problem, the analysis was carried out by the use of the following algorithms of regression trees available in Statistica Data Miner software:

– Standard Regression Trees (CART),
– Standard Regression Trees CHAID,

4. CALCULATION RESULTS

2888 results of industrial attempts were used in the researches. All calculations were done by the use of ten-fold crossvalidation. Minimal variation was chosen as a criterion for stopping the pruning. In case of Standard Regression Trees (CART), General Regression Trees (GCART) and Advanced Regression Trees additionally stopping of FACT type was used because of specificity of these algorithms. The obtained error of predictions of particular types of trees presents figure 1. Algorithm prediction error is calculated according to this formula:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|T_i^{(m)} - T_i^{(p)}\right| \qquad (5)$$

where: MEA – mean absolute error (^{0}C), N – number of learning examples, $T_i^{(m)}$, $T_i^{(p)}$ – measured and predicted temperature

The following chart (figure 2) was compiled to present the influence of tree pruning criterion, (amongst two possible ones to chose in Statistica Data Miner software: stopping the pruning on the basis of a variation and of the FACT type) comparing types of stopping for tree algorithms which enable the choice of stopping criterion.

In case of CART regression trees, there is a possibility to decide on a number of nodes in a tree by determining their maximum quantity. Figure 3 presents mean absolute prediction error depending on the minimum number of nodes.

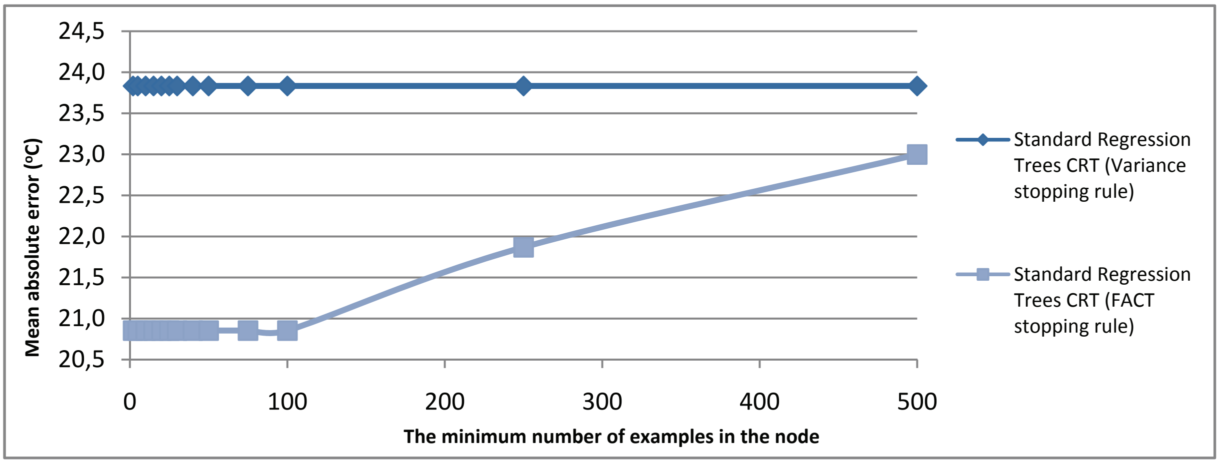

Fig. 2. *Mean absolute error (0C) of predicting the temperature depending on minimal number of examples in a node. Analysis made for CART algorithm with pruning based on variance and FACT method.*

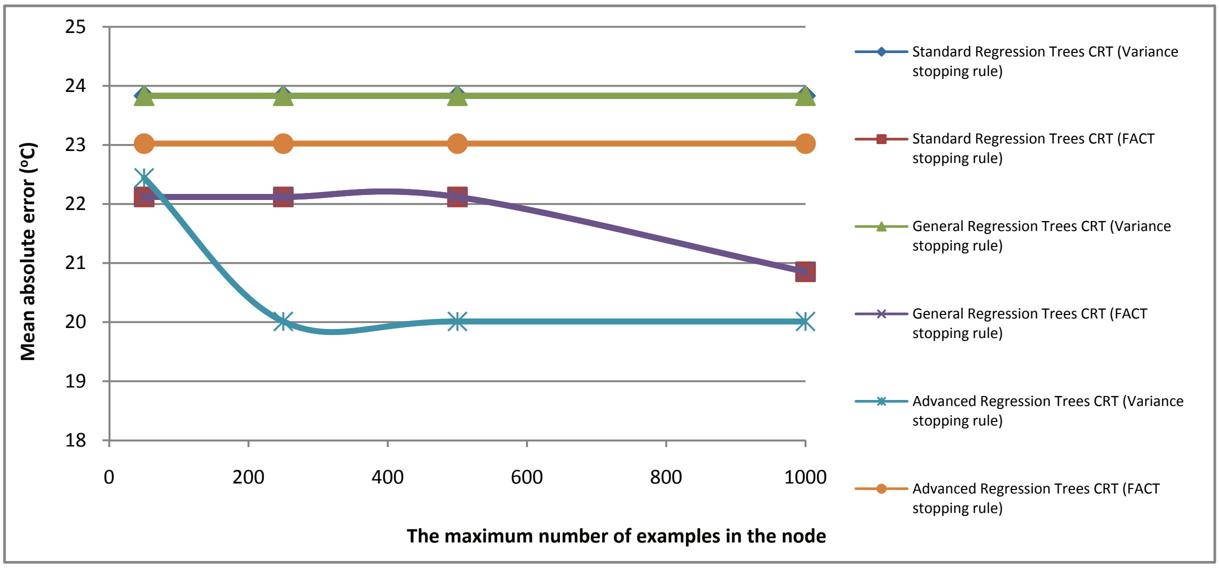

Fig. 3. *Mean absolute error (°C) of temperature prediction depending on the maximum number of nodes for CART algorithm with pruning based on FACT method.*

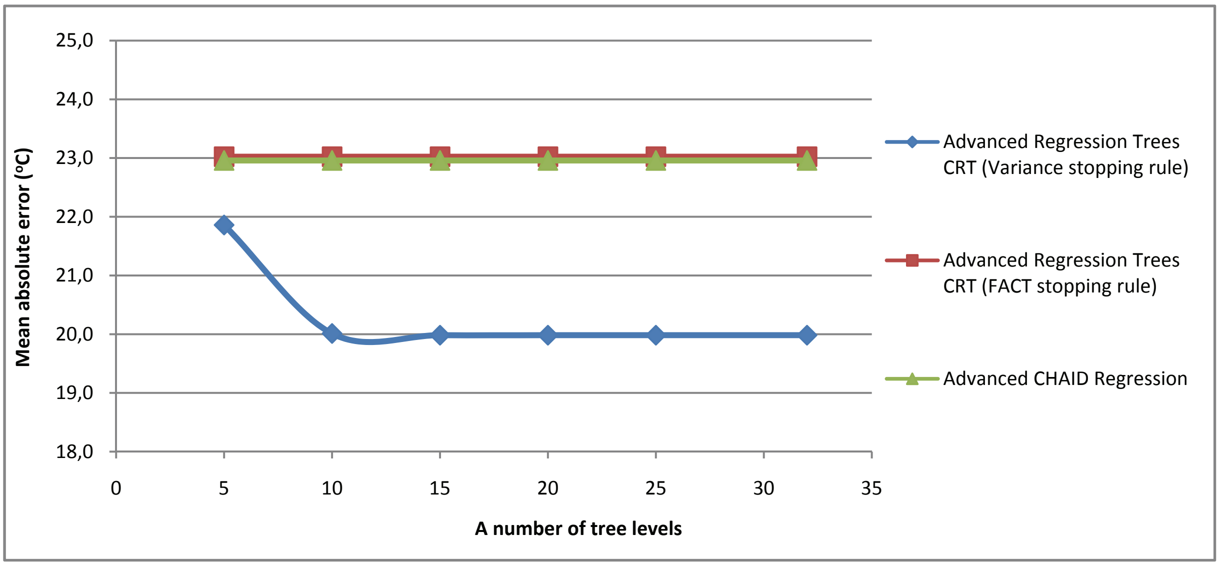

Fig. 4. *Mean absolute error (°C) of temperature prediction depending on the number of tree levels.*

Advanced CHAID for Regression and Advanced CART Regression Trees give a chance of a choice of maximum number of tree levels as one of criteria of stopping (figure 4).

In CART, GCART and Advanced Regression Trees algorithms, it is possible to carry out divisions by the use of stopping of FACT type, while the last nodes are pure. A parameter for this method is impurity of a node (figure 5).

CART where a tree with the pruning stopped on the basis of variation was constructed 2 times longer than with the pruning of FACT type. The fact that a tree was constructed longer did not affect on quality improvement of prediction because mean error in case of CART amounted to 23,8°C and it did not change whereas a tree with the pruning of FACT type gave an error of 20,9°C. In case of Advanced Regression Trees CART, the time of tree construction was twofold longer in comparison with the pruning of FACT type and it gave a smaller error rate of prediction (20°C).

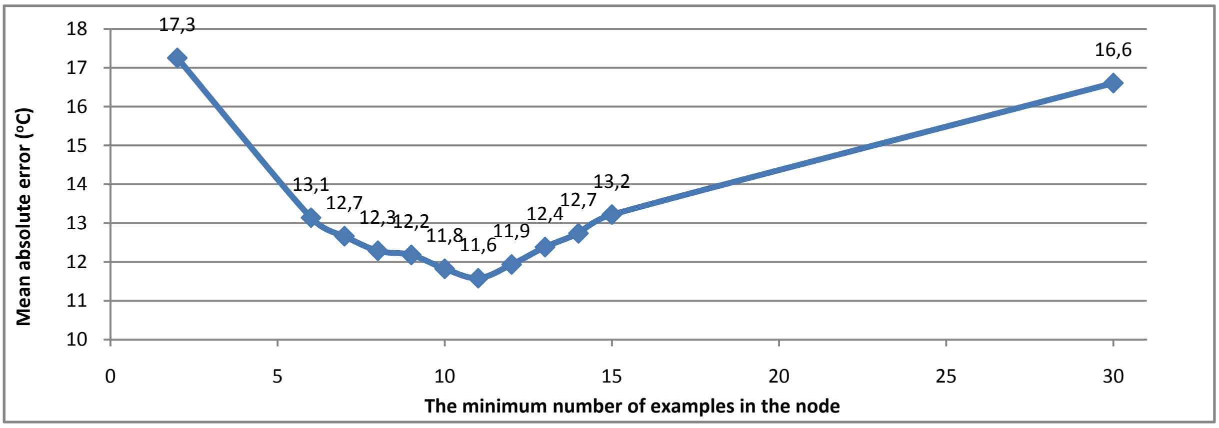

Fig. 5. *Mean absolute error (°C) of temperature prediction by the use of CART algorithm with FACT pruning and pure nodes depending on the minimum number of examples in the node.*

5. ANALYSIS OF THE RESULTS

Basing on the conducted researches, it can be assumed that the best algorithm in temperature prediction was CART. Mean absolute error of 11,6°C was obtained (by the use of ten-fold crossvalidation) where, for this algorithm, FACT stopping criterion was used, purity of nodes was assumed and the number of examples in the node was set to 11.

It can be stated that General GCART Regression Trees predicate with the same error rate as standard CART Regression Trees, what was confirmed during the researches. In case of General CHAID for Regression, we can choose between standard and exhaustive CHAID. The user may decide on which type of optimal division searching in each node will be used by the algorithm. A default value is set for standard CHAID. Therefore, the results obtained by the general CHAID are identical to the results obtained by standard CHAID.

Stopping the pruning criterion has the biggest influence on precision because in case of CART (and also GCART) a tree pruned on the basis of variation is constructed 15 times longer than a tree where FACT stopping is used. A similar relation can be observed in case of Advanced Regression Trees

6. CONCLUSIONS

The aim of this paper is to compare the influence of particular decision trees parameters on the quality of prediction and to find out if it is possible to use the obtained models in real industrial process. The studies consisted of analysis of the most popular algorithms of decision trees (the analysis of parameters that influence on the prediction quality) and their comparison. The investigation was carried out in Statistica Data Miner software by the use of three modules: general models of classification and regression trees, general CHAID models and interactive classification and regression trees. The data was gathered for analysis in the real industrial process. It follows from this paper that some of the analyzed algorithms are more successful in temperature prediction. For temperature prediction, Standard CART Regression Tree with stopping the pruning criterion of FACT type should be used and also purity of a node should be assumed and the number of examples in a node should be set to 11 because as it follows from the experiments, they predicted the temperature of liquid steel more accurately. Considering the obtained results (where mean error of prediction amounts to 11,6°C), this algorithm can be used for industrial purposes.

ACKNOWLEDGEMENTS

The project is partially sponsored by the grant No 4866/B/T02/2010/38 from the Polish Ministry of Education and Science (MNiSZW).

REFERENCES

Alexis, J., Ramirez, M., Trapaga, G., Jonsson, P., 2000, Modeling of a DC electric arc furnace – heat transfer from the arc, *ISIJ International,* 40 (11), 1089-1097.

Billings, S., Boland, F., Nicholson, H., 1979, Electric arc furnace modeling and control, *Automatica*, 15 (2), 137-148.

Blachnik, M., Mączka, K., Wieczorek, T., 2010, A model for temperature prediction of melted steel in the electric arc furnace (EAF), *Lecture Notes in Computer Science*, 4839, 371-378.

Boulet, B., Lalli, G., Ajersch, M., 2003, Modeling and control of an electric arc furnace, *Proc. American Control Conf.,* Denver, IEEE Press, 3060-3064.

Breiman, L., Friedman, J.H., Oslhen, R.A., Stone, C.J., 1984, *Classification and Regression Trees*, Wadsworth International Group, Belmont CA.

Cameron, A., Saxena, N., Broome, K., 1998, Optimizing EAF operations by dynamic process simulation. *Proc. of 56th Electric Furnace Conference*, New Orleans, 689-696.

Deneys, A., Peaslee, K., 1997, Post-combustion in the EAF – a steady-state model, *Proc. of 55th Electric Furnace Conference*, Chicago, 417-427.

Hayman, C., Agarwal, S., Heyd, L., Gordon, I., 1997, A dynamic model for the optimization of the electric steelmaking process, *Proc. of 55th Electric Furnace Conference*, Chicago, 409-415.

Ichihashi, H., Shirai, T., Nagasaka, K., Miyoshi, T., 1996, Neuro-fuzzy ID3: a method of inducing fuzzy decision trees with linear programming for maximizing entropy and an algebraic method for incremental learning, *Fuzzy Sets and Systems*, 81, 157-167.

Jankowski, N., Grąbczewski, K., 2006, Learning machines. In: Guyon, I., Gunn, S., Nikravesh, M., Zadeh, L., (Eds), *Feature Extraction, Foundations and Applications*, Physica Verlag, Springer, Berlin, Heidelberg, New York.

Matson, S., Ramirez, W.F., 1999, Optimal operation of an electric arc furnace. *Proc. of 57th Electric Furnace Conference*, USA, 719-730.

Quinlan, J.R., 1993, *C4.5 – programs for machine learning*, Morgan Kaufman, San Mateo.

Wieczorek, T., 2006, Intelligent control of the electric-arc steelmaking process using artificial neural networks, *Computer Methods in Material Science*, 6 (1), 9-14.

Wieczorek T., Blachnik M., Mączka K., 2008, Building model for time reduction of steel scrap meltdown in the electric arc furnace. Lecture Notes in Artificial Intelligence, vol. 5097, 1149-1159.

Wieczorek, T., Kordos, M., 2010**,** Neural Network-based Prediction of Additives in the Steel Refinement Process, *Computer Methods in Material Science*, 10, 9-14.

Wieczorek, T., Mączka, K., 2008, Modeling of the AC-EAF process using computational intelligence methods, *Electrotechnical Review*, 11, 184-188.

Wieczorek, T., Pilarczyk, M., 2008, Classification of steel scrap in the EAF process using image analysis methods, *Archives of Metallurgy and Materials*, 53 (2), 613-618.

Wieczorek, T., 2008, *Neuronowe modelowanie procesów technologicznych*, Politechnika Śląska, Gliwice (in Polish).

PRZEWIDYWANIE TEMPERATURY W ELEKTRYCZNYM PIECU ŁUKOWYM Z ZASTOSOWANIEM DRZEW DECYZJI

Streszczenie

Drzewa decyzji są jednymi z metod inteligencji obliczeniowej, które okazały się niezawodne przy rozwiązywaniu skomplikowanych, wielowymiarowych problemów obliczeniowych. Dlatego też, metody te są często stosowane do ekstrakcji reguł oraz do przewidywania wartości zmiennych, co czyni je szczególnie użytecznymi w problemach automatyzacji produkcji. W niniejszej pracy autorzy zaprezentują możliwość zastosowania drzew decyzji podczas procesu elektrołukowego. Głównym celem jest predykcja temperatury w elektrycznym piecu łukowym przy użyciu drzew decyzji. Poprawne i automatyczne przewidywanie temperatury może pozwolić na redukcję liczby wykonywanych pomiarów podczas procesu, a co za tym idzie, może pozwolić na skrócenie czasu całego procesu. Optymalizacja procesu daje wymierne korzyści, którymi mogą być na przykład niższe koszty produkcji. Obliczenia wykonane zostały przy użyciu sześciu typów drzew regresyjnych dostępnych w pakiecie Statistica Data Miner. Algorytmy były testowane pod względem osiągania jak najmniejszego błędu predykcji temperatury, ale także pod względem jak najmniej skomplikowanej struktury drzewa, która jest także ważnym elementem pod względem złożoności obliczeniowej.

Received: October 1, 2010
Received in a revised form: October 22, 2010
Accepted: October 22, 2010