On the Application of Artificial Intelligence Techniques to the Quality Improvement of Industrial Processes

Pavlos Georgilakis[1] and Nikos Hatziargyriou[2]

[1] Schneider Electric AE, Elvim Plant, P.O. Box 59, 32011, Inofyta, Viotia, Greece
pavlos_georgilakis@mail.schneider.fr

[2] National Technical University of Athens, 9 Heroon Politexneiou str., 15780, Athens, Greece
nh@mail.ntua.gr

Abstract. In this paper, the combined use of decision trees and artificial neural networks is examined in the area of quality improvement of industrial processes. The main goal is to achieve a better understanding of different settings of process parameters and to be able to predict more accurately the effect of different parameters on the final product quality. This paper also presents results from the application of the combined decision tree - neural network method to the transformer manufacturing industry. In the environment considered, quality improvement is achieved by increasing the classification success rate of transformer iron losses. The results from the application of the proposed method on a transformer industry demonstrate the feasibility and practicality of this approach for the quality improvement of industrial processes.

1 Introduction

In this paper, the combined use of decision trees and artificial neural networks is examined in the area of quality improvement of industrial processes. The main goal is to achieve a better understanding of different settings of process parameters and to be able to predict more accurately the effect of different parameters on the final process (or product) quality.

A hybrid Decision Tree – Neural Network classifier is proposed in this paper. This approach combines the attractive features of two artificial intelligence techniques, namely the transparency and model interpretability of Decision Trees (DTs) and the information accuracy of multi layer perceptrons (MLPs). In the proposed method, first, DTs identify the critical parameters affecting the quality of the industrial process (or product) and express in a clear hierarchical fashion their influence on process (or product) quality. Second, the obtained trees are reformulated in terms of an equivalent four-layer feed-forward neural network (NN) [1] to which they provide structural information, i.e., number of neurons and topology.

This paper also presents results from the application of the combined decision tree - neural network method (hybrid method) to the transformer manufacturing industry. In

I.P. Vlahavas and C.D. Spyropoulos (Eds.): SETN 2002, LNAI 2308, pp. 473–484, 2002.

the environment considered, quality improvement is achieved by increasing the classification success rate (CSR) of transformer iron losses.

This paper is organized as follows: a short description of DTs and MLPs is given in Sections 2 and 3, respectively. In Section 4, the proposed hybrid DT - NN classifier is presented. In Section 5, this methodology is applied to transformer quality improvement. Conclusions are finally presented in Section 6.

2 Overview of DT Methodology

The Decision Tree methodology [2] is a non-parametric learning technique able to produce classifiers about a given problem in order to reduce information for new, unobserved cases. The DT is a tree structured upside down, built on the basis of a Learning Set (LS). The LS comprises a number of preclassified states defined by a list of candidate attributes. The construction of a DT starts at the root node with the whole LS of preclassified measurement sets (MS). These MS are analyzed in order to select the test T which splits them "optimally" into a number of most "purified" subsets. The test T is defined as:

$$T : A_i \leq t \ , \tag{1}$$

where A_i is the value of attribute i of a particular MS, and t is the optimal threshold value. Selection of the optimal test is based on maximizing the additional information gained through that test. A measure of the information provided by a test of form (1) is based on the entropy of the examined subset and is obtained from the normalized correlation measure between the test and the goal partition in the subset of the LS, as defined in [2]. The α-risk of the hypothesis test determines the amount of evidence required at each node in order to split it. The confidence level is defined as 1-α.

In order to detect if a node is terminal, i.e. "sufficiently" class pure, the classification entropy of the node with a minimum preset value H_{min} is compared. If it is lower than H_{min}, then the node is sufficiently class-pure and it is not further split. Such nodes are labeled LEAVES. Otherwise, a suitable test is sought to divide the node, by applying the optimal splitting rule. In the case that no test can be found with a statistically significant information gain, the node is declared a DEADEND and it is not split.

3 Multilayer Perceptrons

Multi layer perceptrons are feedforward neural networks consisting of one input layer, one or more hidden layers and one output layer (Fig. 5). Each layer is made out of neurons and each neuron is connected to the neurons in the adjacent layer with different weights.

The neurons in the input layer are passive; each simply broadcasts a single data value over weighted connections to the hidden neurons. The hidden and output neurons process their inputs in two steps. Each neuron multiplies every input by its weight, adds the product together with neuron's threshold value (bias) to a running total, and then passes the sum through a transfer (activation) function to produce its result. This transfer function is usually a steadily increasing S-shaped curve, commonly called a sigmoid function. The backpropagation (BP) algorithm [3] is the most frequently used training procedure for MLPs.

4 A Hybrid DT-NN Classifier

A MLP with two hidden layers used for classification performs the following functions. The first hidden layer is the partitioning layer that divides the entire feature space into several regions. The second hidden layer is the ANDing layer that performs ANDing of partitioned regions to yield convex decision regions for each class. The output layer is the ORing layer that combines the results of the previous layer to produce disjoint regions of arbitrary shape.

On the other hand, a binary DT induces a hierarchical partitioning over the decision space. Starting with the root node, each internal (test) node partitions its associated decision region into two half spaces. It is obvious that all the conditions along any particular path from the root to the terminal node of the DT must be satisfied in order to reach the particular terminal node. Thus, each path of a DT implements an AND operation on a set of half spaces. If two or more terminal nodes result in the same class, then the corresponding paths are in an OR relationship.

From the previous mentioned reasons, it is obvious that a DT and a four-layer perceptron are equivalent in terms of input-output mapping. In addition, a DT can be reformulated as a neural network by following the rules proposed in [1]. According to this technique the neural network, called entropy network (EN), has the following four-layer architecture.

- **a. The Input Layer (IL)** consists of one neuron per attribute selected and tested by the DT.
- **b. The Partitioning or Test Layer (TL)** consists of one neuron per DT test node.
- **c. The ANDing Layer (AL)** consists of one neuron per DT terminal node.
- **d. The ORing Lyer (OL)** consists of one neuron per DT class.

The connections between the neurons of the above four layers implement the hierarchy of the DT. In particular, each neuron of the TL is connected to the neuron of the IL corresponding to the tested attribute. In addition, each neuron of the AL is linked to the neurons of TL corresponding to the test nodes located on the path from the top node towards the terminal node. Finally, each neuron of the OL is connected to the neurons of AL corresponding to the DT terminal nodes. In comparison to the standard MLPs that are fully connected, the entropy network has fewer connections, or equivalently fewer number of parameters, reducing the time needed for training.

The entropy network can be used only for classification. However, some modifications to the structure of the EN are required in order to use it for prediction purposes [4]. In this case the OL layer would be replaced by a single output neuron, fully connected to all neurons of the AL and the resulted network should be trained again. This methodology is called hybrid DT - NN (HDTNN) approach. Since the entropy network is used only for classification, the comparison between the EN and the HDTNN can be done only on their classification performance. For that reason, after HDTNN convergence, the network is used to predict the test states and after that to classify them accordingly, providing the so-called hybrid DT - NN classifier (HDTNNC).

5 Industrial Applications

In this section, results from the application of DTs, ENs and the HDTNNC are used in order to improve transformer quality through better classification of both individual core and transformer specific iron losses.

In the specific industrial environment, accurate classification of iron losses is an important task, since the latter constitute one of the main parameters of transformer quality. In addition, accurate estimation of transformer iron losses protects the manufacturer of paying loss penalties. In order to avoid this risk, the transformer is designed at a lower magnetic induction, resulting in increase of the transformer cost, since more magnetic material is required. In case of wound core type transformers, classification of iron losses of individual cores is also desired. Satisfactory classification of iron losses however can be achieved only if various parameters, involved in the process, both qualitative and quantitative, are taken into consideration. Instead, in the current practice, only the loss curve is used, i.e. only the influence of the rated magnetic induction on iron losses, for each specific type of magnetic material, is considered. This is dictated by the fact that there is no analytical relationship expressing the effect of the other parameters on transformer iron losses.

5.1 Wound Core Distribution Transformer

In order to construct a three-phase wound core distribution transformer, two small individual cores (width of core window equal to F1) and two large individual cores (width of core window equal to F2) should be assembled (Fig. 1). In general, the width F2 is twice F1.

The theoretical iron losses, say W1 (in Watt), of the small individual core are given by:

$$W1 = WPK_1 * CTW_1 \ , \tag{2}$$

where WPK_1 are the theoretical individual core specific iron losses at the rated magnetic induction (Fig. 2) and CTW_1 is the theoretical weight of the small core as defined in [5].

The theoretical iron losses, say W2 (in Watt), of the large individual core are:

$$W2 = WPK_1 * CTW_2 \ , \tag{3}$$

where CTW_2 is the theoretical weight of the large core.

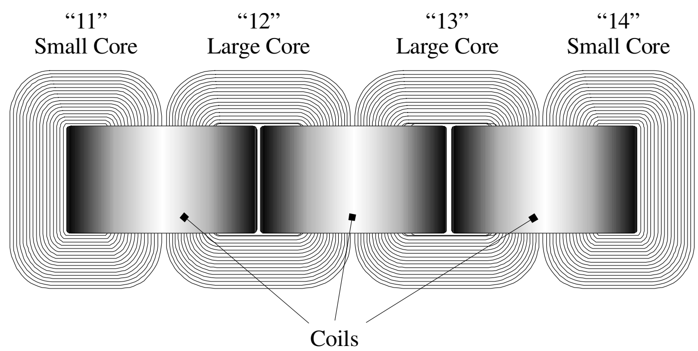

Fig. 1. Assembled active part of wound core distribution transformer.

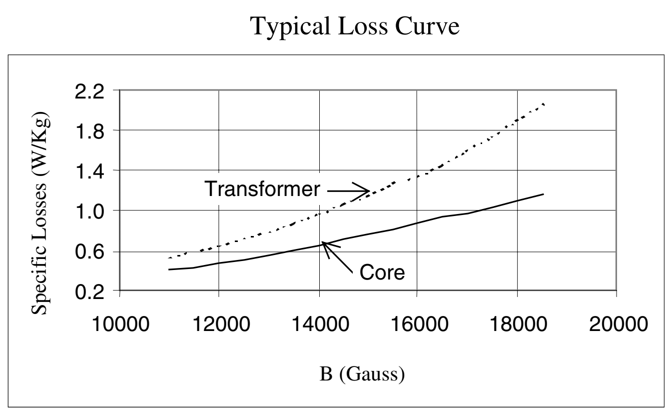

Fig. 2. Typical loss curve.

Consequently, the theoretical total iron losses, say $W1_{tot}$ (in Watt), of the four individual cores are:

$$W1_{tot} = 2 * (W1 + W2) \ . \tag{4}$$

The theoretical iron losses of the three-phase transformer, TFLosses, are:

$$TFLosses = WPK_3 * CTW \ , \tag{5}$$

where WPK_3 are the theoretical transformer specific iron losses at the rated magnetic induction, also obtained from Fig. 2 and CTW is the theoretical total weight of transformer.

5.2 Application on Individual Core

The objective is the improvement of the quality of individual cores. In particular, the impact of the annealing cycle, the divergence of the actual core weight from the theoretical value and the quality of core magnetic material are taken into consideration as input attributes (Table 1).

Table 1. Attributes for the classification of specific losses of individual cores.

Symbol	Description
ATTR1	Annealing final temperature
ATTR2	Temperature rising time
ATTR3	Furnace opening temperature
ATTR4	Duration of constant temperature
ATTR5	Position of core in the furnace
ATTR6	Protective atmosphere
ATTR7	Actual over theoretical core weight ratio
ATTR8	Specific losses of core magnetic material

768 samples were collected for the creation of the learning and test sets. The 3/4 (576) of them were used as learning set and the rest (192) as test set (TS).

The criterion considered for the classification of specific iron losses of individual core as non-acceptable (NA) is the actual specific iron losses being greater than 15% of the theoretical specific iron losses. Otherwise, individual core is acceptable (A).

In Fig. 3 a characteristic DT is illustrated, developed with the 8-attribute list and 0.999 confidence level. The notation used for the DT nodes is explained in Fig. 4.

The Acceptability Index of a node is defined as the ratio of the acceptable MS in the subset E_n of node n to the total number of MS in E_n. If the Acceptability Index of a terminal node is greater than 0.5, then the MS "falling" to this node are characterized as acceptable, otherwise as non-acceptable.

It should be noticed that the DT consists of 3 test and 4 terminal nodes, and has automatically selected only 3 attributes among the 8 candidate ones. These attributes in decreasing order of significance are ATTR8, ATTR2 and ATTR7. The DTs' classification success rate is 94%.

Based on the Decision Tree of Fig. 3 and the methodology described in section 4 the EN of Fig. 5 can be derived. The EN is composed of 3 input, 3 test, 4 ANDing and 2 ORing (output) neurons. The output discrete information is a two-class classification, i.e., acceptable (A) and non-acceptable (NA) transformers with respect to the DT acceptability criterion considered. The correspondence used between the DT nodes and EN neurons is described in Table 2.

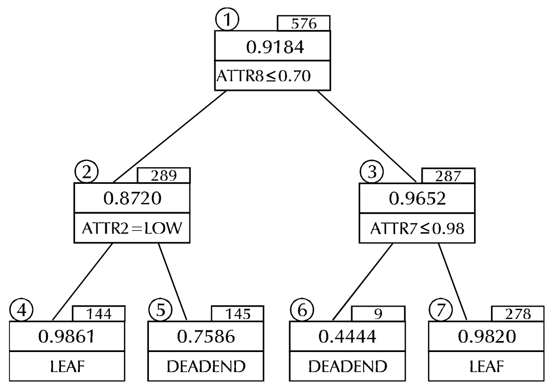

Fig. 3. DT developed using the 8-attribute set.

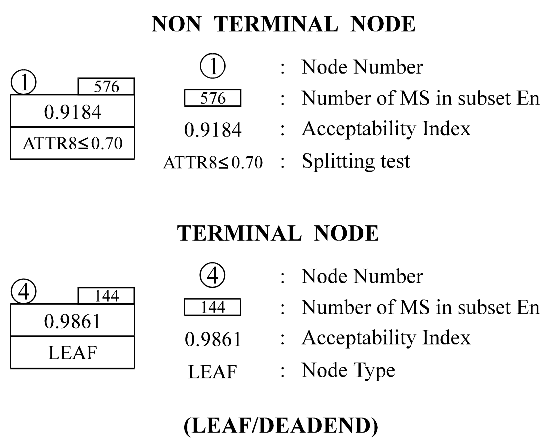

Fig. 4. Notation of the DTs' nodes.

If a very high value of the sigmoidal slope of the transfer function is selected, e.g. λ=20, the EN replicates as closely as possible the discrete classification of the DT. The EN is trained using the NNET package [6]. At convergence the CSR on the TS is 94.2%. Using smaller values for λ, it is possible to improve the CSR. For example, for λ=10, and after further adaptation of the weights the CSR of the EN is increased from 94.2% to 94.6%.

Furthermore, the output layer of the EN is replaced by a single neuron representing transformer specific iron losses and the HDTNN approach, described in section 4, is applied using a value of λ=0.5. After training with NNET and convergence, the NN is used to predict the transformer specific iron losses of the TS and classify them accordingly to the criterion used for DT building. The HDTNNC significantly improves the CSR to 95.7%. This important result is obtained due to the enhancement of the EN information.

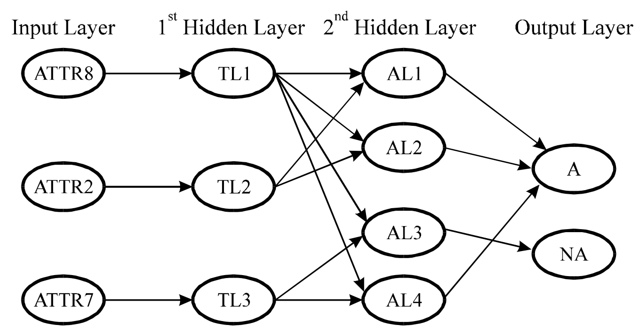

Fig. 5. Entropy network for the DT of Fig. 3.

Table 2. Correspondence between DT nodes (Fig. 3) and EN neurons (Fig. 5).

DT	1	2	3	4	5	6	7
EN	TL1	TL2	TL3	AL1	AL2	AL3	AL4

5.3 Application on Transformer

The objective is the improvement of the quality of transformer, by accurately classifying specific iron losses. However, in this case different attributes have been selected, since at the transformer level the specific iron losses of individual cores are taken for granted and geometrical characteristics are of primary importance. The attributes considered are shown in Table 3.

Table 3. Attributes for the classification of transformer specific iron losses.

Symbol	Attribute Name
ATTR1	Ratio of actual over theoretical total iron losses of the four individual cores
ATTR2	Ratio of actual over theoretical total weight of the four individual cores
ATTR3	Magnetic material average specific losses of the four individual cores (Watt/Kg at 15000 Gauss)
ATTR4	Rated magnetic induction
ATTR5	Thickness of core leg
ATTR6	Width of core leg
ATTR7	Height of core window
ATTR8	Width of core window
ATTR9	Transformer volts per turn

For the creation of the LS and TS 2595 actual industrial measurements were used, 2/3 (1730) of the MS were used as the LS, and the rest as the test set.

The criterion considered for the classification of specific iron losses of transformer as non-acceptable (NA) is the actual specific iron losses being greater than 9% of the theoretical specific iron losses. Otherwise, transformer is acceptable (A).

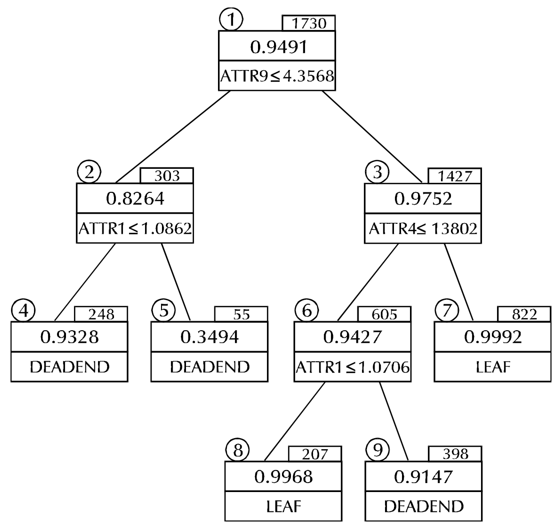

Fig. 6. DT developed using the 9-attribute set.

In Fig. 6 a characteristic DT is illustrated, developed with the 9-attribute list and 0.999 confidence level.

The DT of Fig. 6 consists of 4 test and 5 terminal nodes, and has automatically selected only 3 attributes among the 9 candidate ones. These attributes in decreasing order of significance are ATTR9, ATTR1 and ATTR4. The DTs' classification success rate on the TS is 96%.

Based on the Decision Tree of Fig. 6 the EN of Fig. 7 can be derived. The EN is composed of 3 input, 4 test, 5 ANDing and 2 ORing (output) neurons. The correspondence used between the DT nodes and EN neurons is described in Table 4.

If a value of λ=20 is selected, and training the EN again, the CSR on the TS amounts to 96.3%. Using a value of λ=10, and after further adaptation of the weights the CSR of the EN is increased from 96.3% to 96.7%.

Furthermore, the output layer of the EN is replaced by a single neuron representing transformer specific iron losses and the HDTNN approach is applied using a value of λ=0.5. After training with NNET and convergence, the NN is used to predict the transformer specific iron losses of the TS and classify them accordingly to the criterion used for DT building. The HDTNNC significantly improves the CSR to 97.8%.

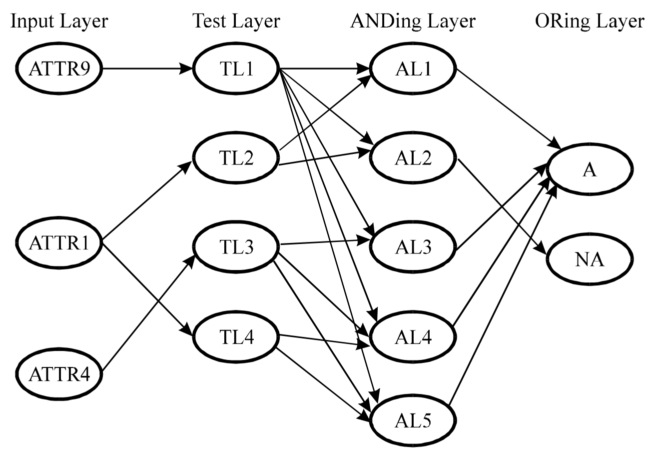

Fig. 7. Entropy network for the DT of Fig. 6.

Table 4. Correspondence between DT nodes (Fig. 6) and EN neurons (Fig. 7).

DT	1	2	3	4	5	6	7	8	9
EN	TL1	TL2	TL3	AL1	AL2	TL4	AL3	AL4	AL5

Table 5. Comparing methods for classification of transformer iron losses.

Method	Structure	*CSR* (%)
DT	-	96.0%
EN	6-6-7-2	96.7%
HDTNNC	6-6-7-1	97.8%
MLP (9 attributes)	9-5-2	98.6%
MLP (3 DT attributes)	3-7-2	96.8%

Moreover, two fully connected MLPs were constructed for the same classification problem. The first MLP comprises 9 input neurons corresponding to the candidate attributes of Table 3, while the second comprises the 3 attributes selected by the decision tree of Fig. 6. Both MLPs have only one single hidden layer and two output neurons corresponding to the acceptable and non-acceptable transformers. The first MLP comprises 5 hidden neurons, i.e. a 9-5-2 structure, and presents a classification success rate of 98.6%. The second MLP has a 3-7-2 structure and a 96.8% CSR.

Table 5 summarizes the results of classification of transformer iron losses. The EN, which is derived by translating the decision tree structure and the second MLP with the 3 attributes identified by the tree, provide very similar classification results. The HDTNNC is more accurate than the EN and the second MLP. The first MLP with the 9 attributes provides the best classification results. Concerning training computational performance, decision trees are by far the fastest method, while among the different neural network approaches the slowest method corresponds to the fully connected MLP.

5.4 Discussion of Results

The 3 attributes appearing in the node splitting tests of the DT of Fig. 3 in decreasing order of significance are ATTR8, ATTR2 and ATTR7. Parameter ATTR8 reflects the quality of the material, as it is equal to the specific losses (Watt/Kg at 15000 Gauss) of core magnetic material. Parameter ATTR2 represents the temperature rising time of the annealing cycle, while parameter ATTR7 expresses the actual over theoretical core weight ratio. The selection of these attributes is reasonable and expected, since they are all related to the quality of individual core. It is notable that the only variable, relevant to the annealing cycle that appears in the node splitting tests of the DT is ATTR2. This is due to the fact that ATTR2, ATTR4 and also the duration of the slow and fast cooling stages are strongly correlated, since the duration of the annealing cycle is considered to be constant. On the other hand ATTR5, which declares the position of core in the furnace, is not important.

Based on the DTs methodology, practical rules, useful for the industrial process, are derived.

In case of individual core, it is concluded from Fig. 3 that it is desirable to construct cores leading to nodes 4 and 7, if it is technically and economically feasible. These nodes have Acceptability Indices greater than 98%. The measurement sets following the rule ATTR8>0.7 and ATTR7≤0.98 are led to node 6, and characterised as non-acceptable. In order to avoid this, the Production Department must increase ATTR7. This is equivalent to increasing the real weight of core by adding more magnetic material, so that the actual over theoretical core weight ratio (ATTR7) is greater than 0.98.

In case of transformer, it is concluded from the DT of Fig. 6 that the measurement sets following the rule ATTR9≤4.3568 and ATTR1>1.0862 are led to node 5, and characterized as non-acceptable. In order to avoid this, ATTR1 should be reduced. The method is to reduce the actual total iron losses of individual cores, by removing from the transformer cores set one or more cores with high iron losses, and add cores with lower ones. The measurement sets following the rule ATTR9>4.3568 and ATTR4>13802 are led to node 7, and characterized as acceptable. This is equivalent to increasing the volts per turn (ATTR9), and also increasing the rated magnetic induction (ATTR4). Design engineers determine both these parameters. In fact, the rated magnetic induction offers enough flexibility, therefore it is desirable to design transformers leading to this node, if it is technically and economically feasible.

Regarding the iron loss classification problem, for the individual core as well as for the transformer, it is concluded that the EN provides very similar classification results with the DT. The HDTNNC is more accurate than the EN. The fully connected MLP provides the best classification results. Concerning training computational performance, decision trees are by far the fastest method, while among the different neural network approaches the slowest method corresponds to the fully connected MLP.

Using the hybrid method, instead of the standard MLP, the tedious task of network structure optimization is avoided. Moreover, training time is significantly reduced (more than a factor 10). Furthermore, the HDTNNC approach can be used to increase the classification success rate, as shown in subsections 5.2 and 5.3, resulting in transformer quality improvement.

6 Conclusions

In this paper, a hybrid decision tree - neural network classifier is proposed for the quality improvement of industrial processes. The method is applied for the increase of the classification success rate of the specific iron losses of both individual core and transformer. The basic steps in the application of the method, like the selection of candidate attributes, the generation of the learning and test sets and the derivation of the appropriate DTs and ENs are presented. Using the HDTNNC the CSR is increased from 94% to 95.7% for the individual core problem, and from 96% to 97.8% for the transformer problem. This significant result is obtained because the HDTNNC combines the advantages of DTs and MLPs while bypasses their weaknesses. Consequently, in the industrial environment considered, the HDTNNC is very suitable for classification of specific iron losses for both individual core and transformer.

References

1. Sethi, I.K.: Entropy Nets: From Decision Trees to Neural Networks. Proceedings of the IEEE, 78(10), October 1990, pp. 1605-1613
2. Wehenkel, L., Pavella, M.: Decision trees and transient stability of electric power systems. Automatica, 27 (1), 1991, 115-134
3. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning internal representation by error propagation. In Rumelhart, D.E., McClelland, J.L., (eds.): Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations, Cambridge, MA: M.I.T. Press, 1986
4. Wehenkel, L.: Automatic Learning Techniques in Power Systems. Kluwer Academic (1998)
5. Hatziargyriou, N.D., Georgilakis, P.S., Spiliopoulos, D.S., Bakopoulos, J.A.: Quality improvement of individual cores of distribution transformers using decision trees. Int. Journal of Engineering Intelligent Systems, 6 (3), September 1998, 29-34
6. AI Ware “NNET 210 User’s Manual”, AI Ware Incorporated, Cleveland, OH, 1989