

Classification of non-normative errors in measuring instruments based on data mining

Milov Anton Vladimirovich

Dept. of Information Control Systems
Reshetnev Siberian State University of Science and
Technology
Krasnoyarsk, Russia
antnraven@ieee.org

Tynchenko Vadim Sergeevich

(a) Dept. of Information Control Systems
Reshetnev Siberian State University of Science and
Technology,
(b) Dept. of Production Machinery and Equipment for
Petroleum and Natural Gas Engineering
Siberian Federal University
Krasnoyarsk, Russia
vadimond@mail.ru

Kukartsev Vladislav Viktorovich

(a) Dept. of Informatics
Siberian Federal University,
(b) Dept. of Information Economic Systems
Reshetnev Siberian State University of Science and
Technology
Krasnoyarsk, Russia

Tynchenko Valeriya Valerievna

(a) Dept. of Informatics
Siberian Federal University,
(b) Dept. of Computer science and engineering
Reshetnev Siberian State University of Science and
Technology
Krasnoyarsk, Russia

Antamoshkin Oleslav Alexandrovich

(a) Dept. of Information Technologies
in Creative and Cultural industries
Siberian Federal University,
(b) Dept. of Information Economic Systems
Reshetnev Siberian State University of Science and Technology
Krasnoyarsk, Russia

Abstract— The article deals with the problem of classification of non-normative errors in measuring instruments. The use of such high-tech methods for creating integral joints, like induction soldering, electron beam welding, diffusion welding, is complicated by the presence of errors in measuring instruments associated with high temperatures, the use of contactless temperature sensors, different thickness of the connected parts, different emissivity coefficients and human factor. Errors that arise when managing the process of creating permanent joints have a significant negative impact on the quality of products in various areas of engineering. In order to adequately control the technological process, it is necessary to identify and classify all types of errors in measuring instruments related to both the features of the technical means of measurement and the human factor. To solve the problem of classifying non-normative errors in measuring instruments, this article proposes the use of intelligent methods. The most suitable, effective and powerful means of solving the problem of non-normative errors classification is the use of artificial neural networks, which make it possible to develop the most effective control actions to compensate the arising non-normative errors.

Keywords— *induction soldering; waveguide path; measurment error classification; flux; neural networks; joints*

I. INTRODUCTION

At present, in the field of aerospace engineering, the most common ways to create permanent joints are: induction soldering, electron beam welding, and also diffusion welding. Figure 1 shows the graph of the technological process of induction soldering.

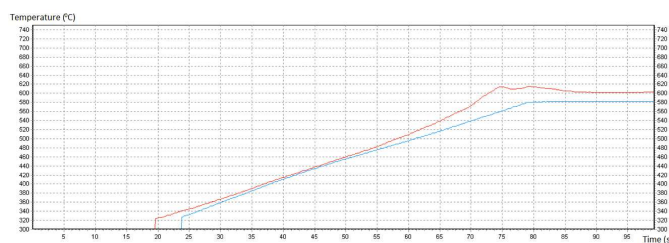


Fig. 1. Graph of temperatures during induction soldering: red graph - readings from pyrometric sensors, blue graph - readings from thermocouple.

The reported study was funded by the President of the Russian Federation grant for state support of young Russian scientists No MK-6356-2018.8

As can be seen from the graph, the pyrometric sensors give a more significant error related to the flux meltdown than the thermocouple. However, it is not possible to use a thermocouple to measure the temperature during induction soldering, since it is necessary to preserve the integrity of the parts to be soldered.

The use of such high-tech methods for creating integral connections is often complicated by the presence of various types of errors in measuring instruments, which include [1, 2]:

1. Errors caused by technical features of measuring instruments.
2. Errors associated with the adjustment of process parameters.
3. Errors associated with the calibration of measuring equipment.
4. Errors associated with the features of the technological process, such as: different thickness of the parts to be joined; different emissivity coefficients of the parts materials being joined; melting of flux, solder, etc.
5. Errors associated with the human factor.

In order to choose an adequate method for compensating the arising non-normative errors in measuring instruments, it is necessary to carry out a classification of the error data. The mathematical formulation of the non-normative errors classifying problem looks as follows.

Let X be the set of a description of the creating integral connections technological process in the form of a time series and tuning parameters of pyrometric temperature sensors, and Y is a finite set of class labels.

There is an unknown target dependence - mapping:

$$Y^* = X \rightarrow Y, \quad (1)$$

whose values are known only at the objects of the final training sample:

$$X^M = \{ (X_1, Y_1), \dots (X_M, Y_M) \}. \quad (2)$$

It is required to construct a mapping algorithm capable of classifying an arbitrary object from X .

II. MATERIALS AND METHODS

The task of classifying non-normative errors in measuring equipment refers to that section of machine learning, such as training with a teacher [3]. Machine learning is an extensive subsection of artificial intelligence that studies methods of constructing algorithms capable of learning [4, 5]. There are two types of training:

1. Learning from precedents, based on the identification of general patterns of private empirical data.

2. Deductive learning, which involves the formalization of experts' knowledge and their transfer to information systems as a knowledge base.

The task of classifying non-normative errors in measuring instruments is related to problems with learning by use of precedents. The task of classification is a task with training with the teacher.

Training with the teacher is one of the sections of machine learning devoted to solving the following problem. There are many objects (situations) and many possible answers (responses, reactions). There is some relationship between responses and objects, but it is not known. Only a finite set of precedents is known - the pairs "object, answer", called training sample. On the basis of these data, it is required to restore the dependence, that is, to construct an algorithm capable of giving an accurate enough answer for any object. To measure the accuracy of answers, the quality functional is introduced in a certain way. [6]

According to the type of input data, the problem of classifying the non-normative errors of measuring instruments refers to tasks whose input is provided with a time series representing a sequence of measurements of the process parameters in time. In general, each measurement is an indicative description of the technological process.

The main methods for solving the classification problem are:

1. The method of k-nearest neighbor.
2. Artificial neural networks.
3. Decision trees.

A. k-nearest neighbor algorithm

The k-nearest neighbor method is the simplest metric classifier, based on an estimation of the objects similarity degree. The object to be classified belongs to the class to which the training sample objects closest to it belong. The method of k-nearest neighbor is designed to increase the reliability of classification. The object belongs to the class, to which most of its neighbors belong, i.e. k objects of the training sample are nearest to it(1) (1)

In binary classification problems, the number of neighbors is taken odd to avoid ambiguity situations when the same number of neighbors belong to different classes. [7, 8]

An example of the k-nearest neighbor algorithm is shown in Figure 2.

In the figure: x_1 and x_2 are input data for classification, yellow indicates the values belonging to the first class, and purple is for the second class. An example which must be classified is marked with a star. As can be seen from the figure, with the parameter $k = 3$, the example should be assigned to the second class, and in the case of $k = 5$, the example refers to the first class.

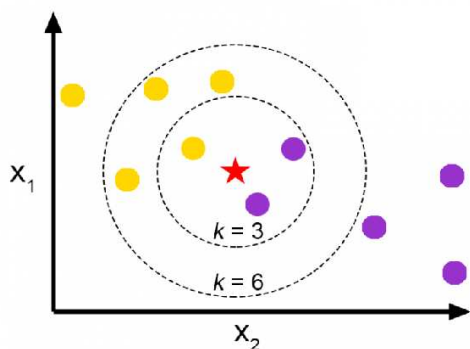


Fig. 2. Example of the k-nearest neighbor algorithm.

On the basis of the experimentally obtained data of the technological process of induction soldering, the classifier model based on the k-nearest neighbor method with $k = 5$ was trained, the conjugacy table of this model is presented in Table 1.

TABLE I. CONJUGACY TABLE OF THE MODEL BASED ON THE K-NEAREST NEIGHBOR METHOD

Predicted value	Actual Value				
	Non-normative errors exist	Non-normative errors do not exist	Normative errors exist	Normative errors do not exist	Total
Non-normative errors exist	69	2	5	6	82
Non-normative errors do not exist	2	78	6	4	90
Normative errors exist	16	0	96	8	120
Normative errors do not exist	5	0	38	127	170
Total	92	80	145	145	462

It can be seen from the conjugacy table that the efficiency of the k-nearest neighbor method for solving the assigned classification problem is 80.08%.

B. Artificial neural networks

Artificial neural networks (ANN) are the mathematical models, as well as their software or hardware implementations, built on the principle of the organization and functioning of biological neural networks - nerve cell networks of a living organism. This concept arose when studying the processes occurring in the brain, and when trying to simulate these processes. The first such attempt was the neural networks of McCulloch and Pitts. Subsequently, after the development of learning algorithms, the resulting models began to be used for practical purposes: forecasting problems, pattern recognition, control tasks, etc.

ANN is a system of connected and interacting simple processors (artificial neurons). Such processors are usually

quite simple, especially in comparison with processors used in personal computers. Each processor of such a network only deals with the signals it periodically receives, and the signals it periodically sends to other processors. And yet, when they are connected to a sufficiently large network with controlled interaction, such locally simple processors together are able to perform rather complex tasks.

From the point of view of machine learning, a neural network is a special case of image recognition methods, discriminant analysis, clustering methods, etc. From a mathematical point of view, learning neural networks is a multiparameter problem of nonlinear optimization. From the point of view of cybernetics, the neural network is used in problems of adaptive control and as algorithm for robotics. From the point of view of the development of computer technology and programming, a neural network is a way of solving the problem of effective parallelism. And from the point of view of artificial intelligence, ANN is the basis of the philosophical trend of connectivism and the main direction in the structural approach to study the possibility of constructing or at least modelling natural intelligence using computer algorithms.

Neural networks are not programmed in the usual sense of the word, they are trained. The possibility of learning is one of the main advantages of neural networks over traditional algorithms. Technically, training is to find the coefficients of connections between neurons. In the process of learning, the neural network is able to identify complex dependencies between input and output data, and perform generalization. This means that, if successful, the network will be able to return the correct result based on data that was not available in the training sample, as well as incomplete and/or "noisy", partially distorted data. [9-11]

An example of a typical structure of an artificial neural network is shown in Figure 3.

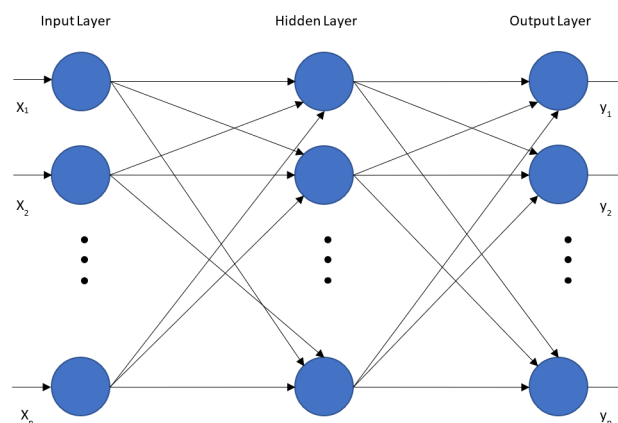


Fig. 3. Visualization of the typical structure of an artificial neural network.

In the figure: x_1, x_2, \dots, x_n are the inputs of the neural network, y_1, y_2, \dots, y_n are the outputs of the neural network.

On the basis of the experimentally obtained data of the technological process of induction soldering, the classifier model based on the artificial neural network was trained.

Table 2 presents the conjugacy table for the ANN with the structure:

1. At the input of the neural network, 62 variables are given:
 - a) Inputs 1 to 60 are the time series.
 - b) Inputs 61 and 62 are readings of the emission coefficient for the 1st and 2nd pyrometers.
2. The hidden layer contains 16 neurons.
3. At the output of the neural network there are 2 variables: the 1st class corresponds to the presence of an error; the second class specifies the error class.

TABLE II. CONJUGACY TABLE OF THE MODEL BASED ON ARTIFICIAL NEURAL NETWORKS

Predicted value	Actual Value				
	Non-normative errors exist	Non-normative errors don't exist	Normative errors exist	Normative errors don't exist	Total
Non-normative errors exist	88	1	1	2	92
Non-normative errors don't exist	1	75	1	3	80
Normative errors exist	2	3	141	5	151
Normative errors don't exist	1	1	2	135	139
Total	92	80	145	145	462

From the conjugacy table it can be seen that the effectiveness of ANN in solving the assigned classification problem is 95.02%.

C. Decision trees method

The decision trees reproduce the logical schemes allowing one to obtain the final decision on the classification of the object by means of answers to a hierarchically organized system of questions. And the question asked at the subsequent hierarchical level depends on the answer received at the previous level.

Such logical models have long been used in botany, zoology, mineralogy, medicine and other fields. Each of the vertices of a tree, with the exception of leaves, corresponds to a certain question, which implies several variants of answers corresponding to outgoing edges. Depending on the selected answer option, the next level is reached. The end vertices are assigned labels that indicate the relation of the recognized object to one of the classes.

A decision tree is called binary if each inner or root vertex is incident with only two outgoing edges. Binary trees are convenient to use in machine learning models. [12, 13]

Visualization of the classifier on the basis of decision trees is shown in Figure 4.

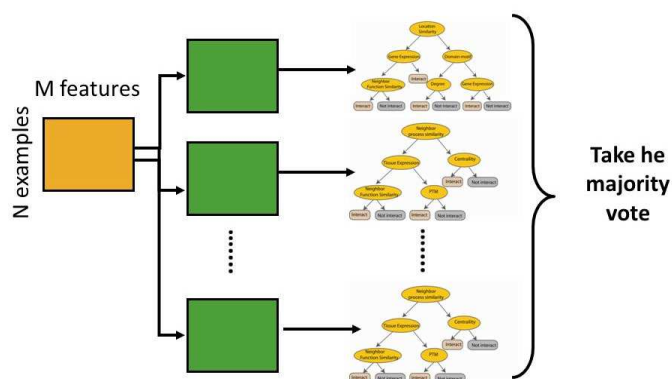


Fig. 4. Classifier based on decision trees.

On the basis of the experimentally obtained data of the technological process of induction soldering, the classifier model based on the decision trees was trained, the conjugacy table of this model is presented in Table 1.

TABLE III. CONJUGACY TABLE OF THE MODEL BASED ON DECISION TREES

Predicted value	Actual Value				
	Non-normative errors exist	Non-normative errors don't exist	Normative errors exist	Normative errors don't exist	Total
Non-normative errors exist	73	1	3	6	83
Non-normative errors don't exist	2	76	5	4	87
Normative errors exist	15	2	96	8	121
Normative errors don't exist	2	1	41	127	171
Total	92	80	145	145	462

It can be seen from the conjugacy table that the effectiveness of the decision trees method for solving the assigned classification problem is 80.52%.

In order to choose the most effective method for classifying non-normative errors in measuring instruments, it is necessary to conduct a comparative analysis of the selected methods. The results are presented in Table 4.

Based on the comparison results, it was concluded that it is expedient to use classification methods based on the artificial neural networks to solve the problem of classifying non-normative errors of measuring instruments.

Artificial neural networks are the most promising and powerful way to solve the classification problem. The classifier based on that algorithm can show the best results from all three classification methods presented.

TABLE IV. COMPARATIVE ANALYSIS OF THE CLASSIFICATION METHODS EFFECTIVENESS

Method of Classification	Probability of Correct Classification				
	Non-normative errors exist	Non-normative errors don't exist	Normative errors exist	Normative errors don't exist	Total
k-nearest neighbor	75%	97.5%	66.2%	87.6%	80.08%
Artificial neural network	95.6%	93.8%	97.2%	93.1%	95.02%
Decision trees	79.3%	95%	66.2%	87.6%	80.52%

III. EXPERIMENTAL STUDY

For the experiments, two artificial neural networks with different structures were trained. The experiments were carried out on the accumulated database of the real technological processes results.

The first series of experiments was carried out with the following artificial neural network structure:

1. At the input of the neural network are given 62 variables:

- Inputs 1 to 60 are the time series.
- Inputs 61 and 62 are readings of the emission coefficient for the 1st and 2nd pyrometers.

2. The hidden layer contains 16 neurons.

3. At the output of the neural network, there are 2 variables: the 1st class corresponds to the presence of an error; the second class specifies the error class.

As a learning method, the back-propagation algorithm was used. Figure 5 shows a graph of the error in the learning process, and Figure 6 shows the accuracy graph in the model learning process.

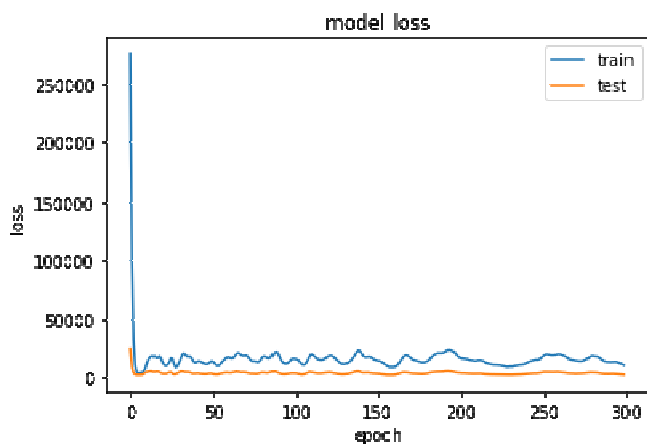


Fig. 5. Graph of the error in the learning process of an artificial neural network with a structure 62x16x2.

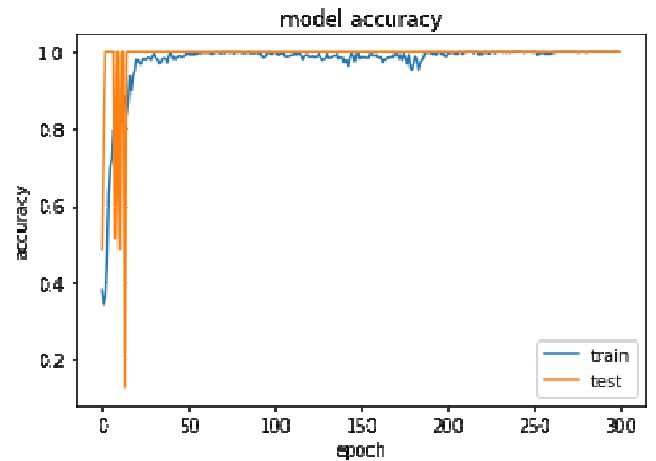


Fig. 6. Graph of accuracy in the learning process of an artificial neural network with a structure 62x16x2.

The second series of experiments was carried out with the following artificial neural network structure:

1. At the input of the neural network are given 62 variables:

- Inputs 1 to 60 are the time series.
- Inputs 61 and 62 are readings of the emission coefficient for the 1st and 2nd pyrometers.

2. Two hidden layers of 8 neurons each.

3. At the output of the neural network, there are 2 variables: the 1st class corresponds to the presence of an error; the second class specifies the error class.

Figure 7 shows a graph of the error in the learning process, and Figure 8 shows the accuracy graph in the model learning process.

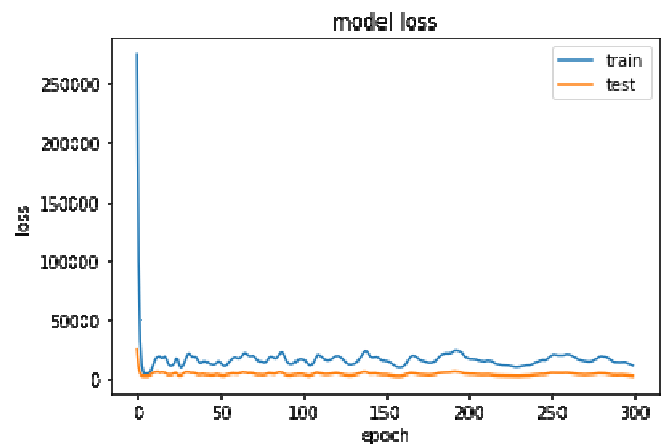


Fig. 7. Graph of the error in the learning process of an artificial neural network with a structure 62x8x8x2.

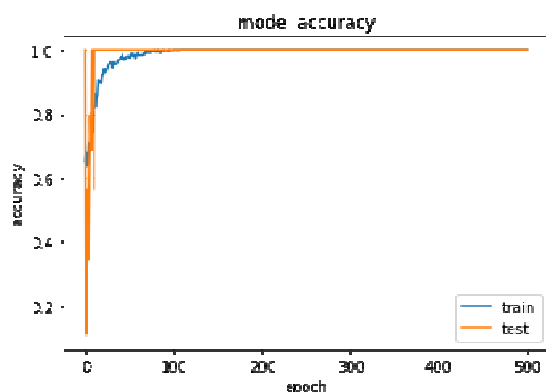


Fig. 8. Graph of accuracy in the learning process of an artificial neural network with a structure 62x8x8x2.

A comparative analysis of the effectiveness of ANN structures is presented in Table 5.

TABLE V. COMPARATIVE ANALYSIS OF THE ARTIFICIAL NEURAL NETWORKS EFFICIENCY FOR SOLVING THE PROBLEM OF NON-NORMATIVE ERRORS CLASSIFICATION

ANN Type	Probability of Correct Classification				
	Non-normative errors exist	Non-normative errors don't exist	Normative errors exist	Normative errors don't exist	Total
ANN with 1 hidden layer	95.6%	93.8%	97.2%	93.1%	95.02%
ANN with 2 hidden layer	91.3%	95%	89.7%	95.2%	92.64%

As can be seen from the graphs presented above, the method of non-normative errors classification based on the artificial neural networks shows a sufficiently high accuracy of recognizing the measuring instruments errors types. The artificial neural network with one hidden layer shows a recognition accuracy of 95.02%, and the artificial neural network with 2 hidden layers shows a recognition accuracy of 92.64%. Thus, it can be concluded that the artificial neural network for such classification problem shows good results.

IV. CONCLUSION

Within the framework of solving the non-normative measuring errors classification problem, three methods were

chosen: the k-nearest neighbor algorithm, the decision trees and artificial neural networks.

The effectiveness of the k-nearest neighbor method and decision trees was lower than that of artificial neural networks. Their effectiveness was 80.08% and 80.52%, respectively. The method of classification based on artificial neural networks showed better results. An artificial neural network with 1 hidden layer of 16 neurons showed a classification accuracy of 95.02%, and with two hidden layers of 8 neurons each - 92.64%.

Thus, artificial neural networks have proved to be an effective method of classifying non-normative errors in measuring technology. The next step is to develop methods for compensating the non-normative errors based on their classification.

References

- [1] A. V. Murygin, V. S. Tynchenko, V. D. Laptinok, O. A. Emilova, and Y. N. Seregin, "Modeling of thermal processes in waveguide tracts induction soldering," IOP Conference Series: Materials Science and Engineering, Vol. 173, No. 1, 2017, p. 012026.
- [2] A. V. Murygin, V. S. Tynchenko, V. D. Laptinok, O. A. Emilova, and A. N. Bocharov, "Complex of automated equipment and technologies for waveguides soldering using induction heating," IOP Conference Series: Materials Science and Engineering. Vol. 173. No. 1, 2017, p. 012023.
- [3] J. Friedman, T. Hastie, and R. Tibshirani, "The elements of statistical learning," Springer series in statistics, Vol. 1, 2001, pp. 337-387.
- [4] I.H. Witten, Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2016.
- [5] T. Mitchell, Machine Learning. Portland, OR : Science/ Engineering/ Math, 1997.
- [6] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, Vol. 86, No 11, 1998, pp. 2278-2323
- [7] D.T. Larose, and C.D. Larose, Discovering knowledge in data: an introduction to data mining. John Wiley & Sons, 2014.
- [8] T. Cover, and P. Hart, "Nearest neighbor pattern classification," IEEE Transactions on Information Theory, Vol. 13, No. 1, 1967, pp. 21-27.
- [9] M. Caudill, and C. Butler, "Understanding Neural Networks: Computer Explorations," Vols. 1 and 2, Cambridge, MA: the MIT Press, 1992.
- [10] C. Charalambous, "Conjugate gradient algorithm for efficient training of artificial neural networks," IEEE Proceedings, Vol. 139, No. 3, 1992, pp. 301-310.
- [11] D. D. Ripley, Pattern recognition and Neural Networks. Cambridge University Press, 1996.
- [12] L. Breiman Classification and regression trees. Routledge, 2017.
- [13] W.Y. Loh, "Classification and regression trees," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Vol. 1, No 1, 2014, pp. 14-23.