Measurement 119 (2018) 1–10

Contents lists available at ScienceDirect

Measurement

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

Hierarchical diagnostics of analog systems based on the ambiguity groups detection

Piotr Bilski

Institute of Radioelectronics and Multimedia Technology, Warsaw University of Technology, ul. Nowowiejska 15/19, 00-665 Warsaw, Poland

ARTICLE INFO

Keywords:
Diagnostics of analog systems
Artificial intelligence
Fault detection

ABSTRACT

The paper presents the hierarchical approach to detect and identify faults in the analog system using combined Artificial Intelligence (AI) methods. The automated diagnostic system has two levels of fault identification, based on the unsupervised and supervised learning. The former is used in the initial stage to separate easily identifiable states of the analyzed system from the difficult ones. The latter are identified with the more sophisticated classifier. Because the difficulty of the fault identification is related with the existence of Ambiguity Groups, the Unsupervised Learning scheme is employed to detect them and decompose training data set into subsets, on which two stages of classifiers are trained. The first set (considered "simple") is processed by the simpler machine learning algorithm. The second set is used to train the more complex classifier (operating in the uncertainty conditions). The proposed scheme is generic, therefore various algorithms can be implemented. In the presented case, the Self Organizing Map (SOM) is used in the first stage, while Random Forest (RF) – in the second one. To verify the approach, the 3rd order Bessel highpass filter was analyzed. The architecture was confronted against the traditional approach (where the standalone classifiers are employed). Results confirm usefulness of the proposed solution, regarding the higher classification accuracy and smaller computational effort than its alternatives.

1. Introduction

Modern diagnostics of analog systems relies on data-driven methods, aimed at maximizing quality of knowledge extracted from available sets during the machine learning process. The Simulation Before Test (SBT) approach is typically used during the analysis of devices such as motors, actuators or circuits. This methodology exploits the extensive System Under Test (SUT) model simulations performed in the off-line mode, where the fault classification or parameter identification module is trained to gain knowledge about the SUT behavior for various faults. This knowledge is further used to detect, identify and locate faults in the actual SUT, based on the information extracted from accessible or partially accessible nodes [1]. The diagnostic operation may be performed on-line (when the SUT is working) or off-line (after disengaging it from the operating environment – so excitation signals can be applied to inputs of the SUT). The ability to determine work regime of the analyzed object based on the observable signals is related to its testability [2]. SUT responses bear the information about values of particular parameters. These symptoms are extracted from measured signals. Two aspects influence testability and must be studied for each SUT separately: informativeness of available symptoms and the set of accessible nodes. The former depends on the domain of analysis (time, frequency or mixed) and descriptors calculated from signals (such as dynamic range, spectral components, and wavelet coefficients). The optimal set of nodes contains their minimal number ensuring the maximum diagnostic accuracy.

Solving both problems are crucial for implementing data-driven approaches, as they influence the process of creating the training, testing and validating data sets, further used by machine learning algorithms. Quality of data determines further accuracy, therefore it must be measured during simulations to evaluate the difficulty of the classification or regression task for the specific SUT. Ambiguity Groups (AG) are well established measure of determining difficulty of the diagnostic procedure or complexity of the system. Their detection allows for identifying fault states similar to each other and therefore difficult to distinguish. Before implementation of the particular AI-based method to extract knowledge about relations between the observed symptoms and actual values of SUT parameters, AG detection should be performed.

Computational intelligence is often used to monitor the SUT because of its high accuracy (evaluated as the number of errors made by the fault classifier operating on the testing data set) or autonomous work regime. Among numerous methods available, the most important are the ones supplemented with machine learning algorithms (unlike Fuzzy

E-mail address: pbilski@ire.pw.edu.pl.

https://doi.org/10.1016/j.measurement.2018.01.029
Received 8 November 2017; Received in revised form 28 December 2017; Accepted 18 January 2018

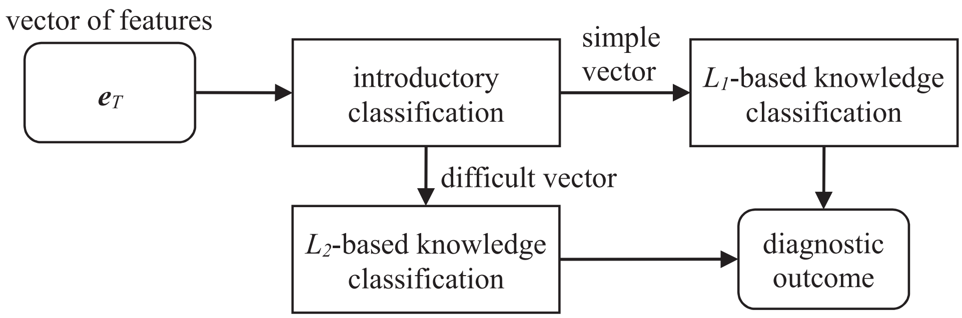

Fig. 1. Architecture of the hierarchical AI-based diagnostic scheme.

Logic [3], which requires external algorithm for that). This way knowledge is inherently extracted and can be used in diagnostics. The main differences between various approaches lie in the form of stored knowledge (for Artificial Neural Networks – ANN it takes the form of the matrix of real numbers, while rule-based approaches rely on rules).

In the traditional approach to the AI-based automated diagnostics, supervised learning is used to train classifier or regression machines. In such a case, the relation between symptoms and fault states presented in the data set is used to guide the training phase of the algorithm. Methods often used in diagnostics include rule-based systems (such as decision trees), statistical methods (including Naïve Bayes Classifier) or Artificial Neural Networks (ANN), such as Multi-Layer Perceptron [4] or Radial Basis Function network [5]. These approaches are easily implemented, they are fast and have acceptable accuracy for many SUTs of simpler structure.

The presence of AG in the set deteriorates the quality of knowledge extracted from it. Therefore they must be first detected to isolate SUT states, which are difficult to distinguish. For this purpose, the Unsupervised Learning (UL) determines similarities in the set based only on the values of symptoms, disregarding the information about SUT states. After finding such “difficult” examples, more sophisticated approaches, working in the uncertainty conditions, are implemented. These include Random Forest (RF) [6], Fuzzy Neural Networks [7] or Support Vector Machines (SVM) [8]. Such methods are considered more useful during the analysis of difficult data, minimizing diagnostic errors. If the data set is considered difficult such a sophisticated approach should be used.

The paper presents the novel hierarchical diagnostic scheme based on both unsupervised and supervised learning. The former is used to find relations in the learning data set and find AG, which are the basis to create “simple” and “difficult” subsets, further exploited by the selected fault classifiers for the training. The diagnostic process is two-staged. In the first one, the vector of symptoms extracted from SUT responses is categorized as simple or difficult to process. Based on this information, the selected classifier is used to make a decision about the SUT state. To confirm usefulness of the approach, the highpass filter was analyzed. Parametric faults were considered here (i.e. related with gradual deviations of parameters from their nominal values, while the topology of the system remains intact).

The aims of the paper are as follows:

- Introduce the novel diagnostic approach, decomposing learning data sets into groups of examples easily distinguishable and difficult for the classifier, individually adjusted for the specific subset of faults.
- Compare effectiveness of the proposed approach against the traditional diagnostic methods, where the single classifier is trained on the whole available data set.
- Verify accuracy of the specific classifiers and UL method for the defined task.

The paper structure is as follows. In Section 2 the diagnostic architecture is presented. Section 3 presents works related to the presented solution. In Section 4 the data set used for training and testing the scheme is described. Section 5 presents unsupervised and supervised learning classifiers implemented in the research. In Section 6 the analyzed SUT (the analog highpass filter) is introduced. Section 7 contains experimental results, comparing effectiveness of various classifier configurations, also with the traditional diagnostic methods. Section 8 concludes the paper with the summary about the effectiveness of the approach.

2. Hierarchical AI-based diagnostics scheme

In the data-driven approaches, the crucial part of the system is knowledge extracted from the learning data set *L* during the classifier construction. The content of the set implies accuracy of the diagnostic module. If all faults occurring in the SUT are considered, the diagnostic module should be able to generalize, i.e. correctly detect problems related to particular deviations of parameters, which were not presented during the training. Verification of the classifier accuracy is performed using the testing set *T*, of the same form as *L*, but containing different fault cases. Traditionally, this pair of sets is used to optimize the classifier, trained on *L* and tested on *T*. The latter has the same purpose in the presented work, while the former is decomposed during the training, as is presented in the following subsections.

2.1. Diagnostic architecture

The proposed automated diagnostic scheme using AI methods (originally introduced in [9]) is presented in Fig. 1. It is assumed that intelligent classifiers were already trained on the available sets and the tested SUT's state is represented by the features vector $\boldsymbol{e}_T$, containing symptoms extracted from the measured signals. First, the introductory classification is performed. Its task is to make a binary decision, i.e. determine, which of the two available classifiers should be used to process the vector. As the result, the selected classifier is executed and its outcome returned as the diagnostic decision. Each classifier is trained on the separate data set, created after decomposing the original set *L* into subsets L_1 and L_2, containing simpler and more difficult examples for processing, respectively. Each example in *L* is the vector of symptoms supplemented by the fault code, describing the SUT's state. If the introductory classification module evaluates the vector as the simple one, the basic classifier, trained on L_1 is used. It is fast, its knowledge relatively compact and memory efficient. Examples of such methods include the Decision Tree (DT) [10] or Naïve Bayes Classifier [11]. The more sophisticated module is able to consider fault states difficult to diagnose. They are affected, for instance, by the limited sensitivity of the SUT to changes of parameters' values, or existence of AG. The module must also be able to work in the uncertainty conditions (such as the presence of noise). Such methods include SVM, RF or Fuzzy Logic. They can be used as the auxiliary classifier, ensuring the higher accuracy in identification of hardly distinguishable SUT states.

The decomposition of the data set *L* into subsets for training the simple and sophisticated classifier is performed based on the UL (processing only symptoms from *L* and disregarding the corresponding fault codes). It is used to detect AG [12] among examples, identified thanks