DIMENSIONALITY REDUCTION IN COMPUTATIONALLY INTELIGENT PHOTOACOUSTIC MEASUREMENT DATA PROCESSING

Miroslava Jordović-Pavlović[1], Marica Popović[2], Slobodanka Galović[3], Katarina Đorđević[3], Mioljub Nešić[3], Dragana Markushev[2], Dragan Markushev[2]

[1] Western Serbia Academy of Applied Studies, Užice, Serbia
e-mail: miroslava.jordovic-pavlovic@vpts.edu.rs

[2] Institute of Physics, University of Belgrade, Belgrade - Zemun, Serbia
e -mail: dragan.markushev@ipb.ac.rs, marica.popovic@ipb.ac.rs, dragana.markushev@ipb.ac.rs

[3] Vinca Institute of Nuclear Sciences - National Institute of the Republic of Serbia, Serbia
e-mail: bobagal@vin.bg.ac.rs, mioljub.nesic@vin.bg.ac.rs, katarina.djordjevic@vin.bg.ac.rs

Abstract:

The research presented in this paper is part of an effort to improve the photoacoustics measurement signal autocorrection method, based on computational intelligence. Autocorrection here deals only with the elimination of distortion of the experimental signal due to the influence of the measuring system. The method implies two already developed models - regression and classification, which in cooperation recognize the transfer function of the microphone in the frequency domain, as an accurate representation of the microphone response. The obtained data are significant for the signal correction procedure, the result of which is an undistorted signal, used in the process of material characterization, for accurate, precise, and reliable determination of the parameters of the tested sample. Further research goes in the direction of dimensionality reduction for both regression and classification models without losing the quality of measurements. This paper presents the results of classification model improvement. Namely, testing the model under different conditions (theoretical or experimental signals, with and without noise, different types of microphones, different samples) we found that the accuracy of the model is high and that the processing speed of measured data does not change seriously by reducing the number of measurement points and thus reducing dimension of the models' input vector. Principal component analysis, discussion of feature correlations and expert knowledge were used to determine the number and measurement points frequency. It was proved that the procedure of measuring and classification of microphones can be performed simply and quickly by measuring at 1 defined point.

Keywords: dimensionality reduction, machine learning, neural networks, photoacoustics.

1. Introduction

In open-cell configuration photoacoustic spectroscopy, used in the inverse determination of sample properties, a microphone with associated electronics and a phase-frequency (lock-in) amplifier plays a major role. For the sake of simplicity, the microphone is considered as the main part of the measurement system with the greatest impact on distortions. A complete description of the microphone frequency response is therefore necessary for the signal correction procedure. It requires specific details that can be classified into several different classes and are related to the shape of the transfer function and different levels of signal amplification and attenuation. There are flat response microphone and shaped response microphone. A flat response curve is an almost straight line, indicating that the measuring chain is almost equally sensitive along the entire frequency range, so that the photoacoustic signal is transmitted without distortion. The shape response curve has a structure with peaks and valleys indicating that the microphone is more sensitive at certain frequency bands compared to others, so that the signal is transmitted with a certain degree of distortion.

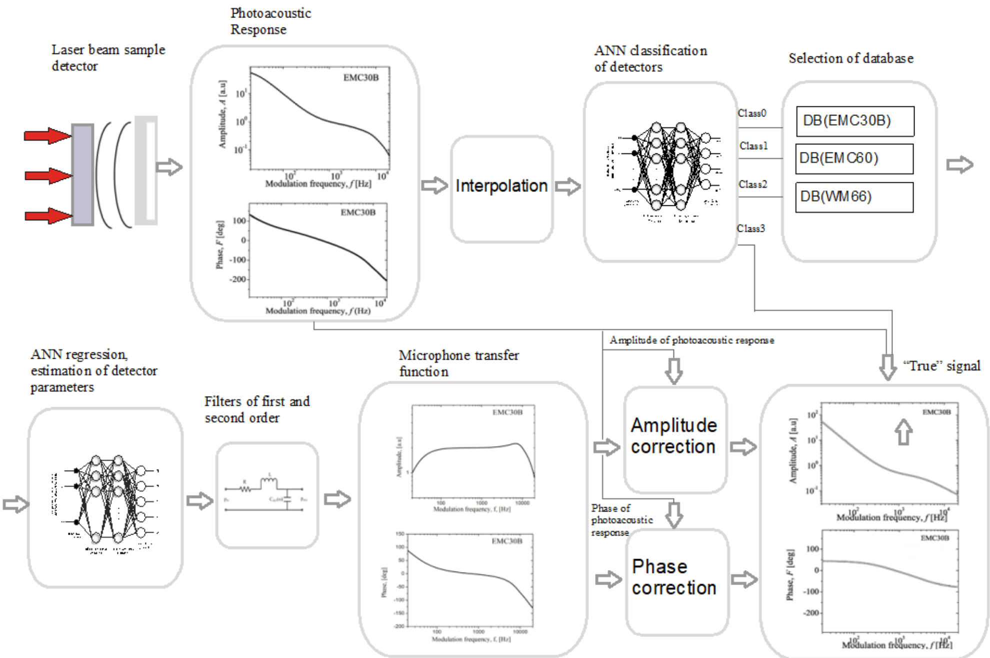

Fig. 1. Photoacoustics measurement signal autocorrection method [1]

Photoacoustics measurement signal autocorrection method (Figure 1) recognises the transfer function of the microphone in the frequency domain from the experimental signal [1]. Here the regression and classification model are coupled in such a way that the classification model first recognizes the form of transfer function corresponding to a specific microphone type, and then the regression model predicts numerical values of parameters attributed to the shape of the transfer function [1]. Signal correction procedure uses derived data. The output of the method is undistorted signal. Obtained signal is used in the process of material characterisation, for accurate, precise and reliable determination of the parameters of the tested sample.

In the previously mentioned method, the classification model (Figure 2) accelerates the processing photoacoustics measurement data by recognising the type of frequency response curve: if a flat curve is detected, the photoacoustic signal is not distorted and is automatically forwarded to the output. If the signal is distorted, the classification model has the task to recognise which type of microphone has such a response. A regression model (Figure 2) [1] is trained on a selected database, recognised by the classification. Database for each observed type of microphone is obtained by a known theoretical model [2].

The input vector of classification model has 400 elements, whereby 200 of them are the samples of amplitude and the remaining 200 are the samples of phase characteristics of the analysed photoacoustics response in a definite number of points on the frequency axis, which means that 200 measurements must be performed. It is a well-known fact that multilayer neural networks (regression model is a three-layer network), and especially deep learning, are not sensitive to the number of input parameters and do not require prior data engineering [3]. Therefore, it can be concluded that dimensionality reduction is not important for optimisation of regression and classification model but it is important for solving the problem of a large number of measurements.

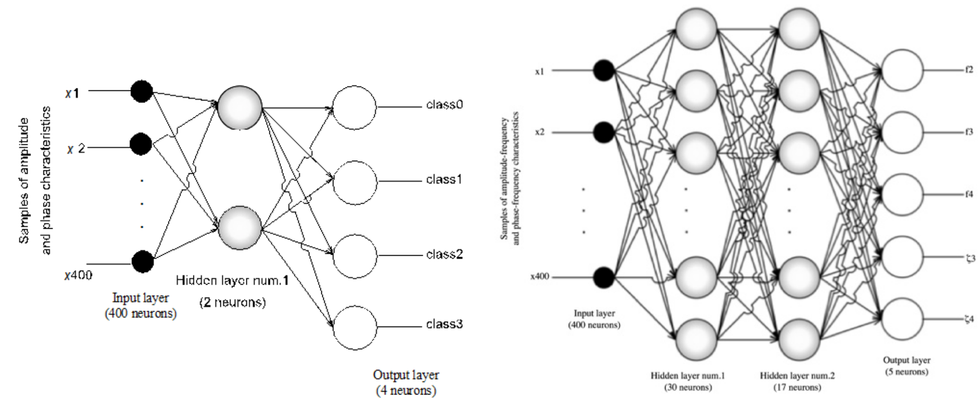

Fig. 2. Classification model [4] and regression model [1]

In this research, the PCA algorithm and data set filtering combined with expert knowledge were used to provide an answer - what is the number of points and what are the frequencies of the chosen measurement points of the photoacoustic experiment.

2. Dimensionality reduction of the input vector of classification model

The PCA algorithm applied to the database used to train the classification model (270 000 records) [2] gives the distribution of the remaining variance in relation to the number of features as shown in Figure 3. The remaining variance for 5 measuring points or 10 features is 99.978%, for 4 features or 2 measuring points it is 99,772%, while for 2 features or 1 measuring point it is 98.849%. The high values of the remaining variance indicate the possibility of a significant simplification of the measurement procedure in order to classify the detector.

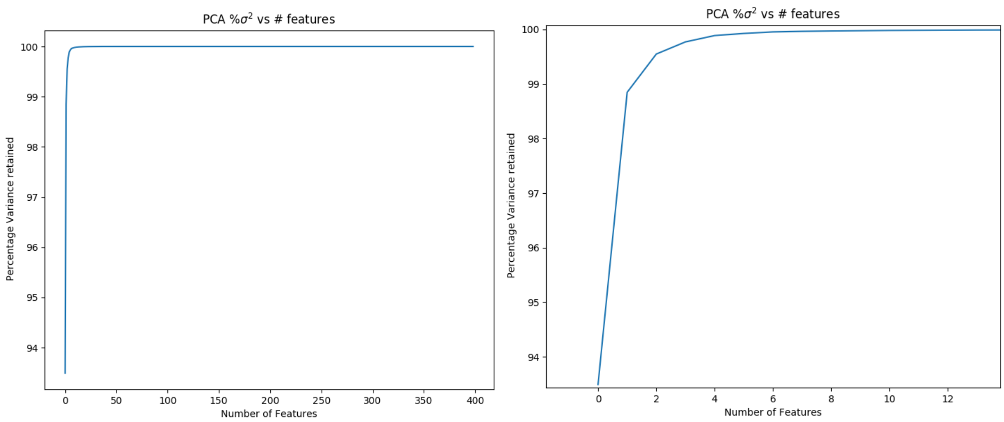

Fig. 3. Remaining variance in relation to the number of features, data set 400 x 270 000

Analysing the amplitude correlation diagram on the entire data set(Figure 4) we conclude that there are 5 areas with high correlation of the included features (light fields). The same situation is with the phase correlation diagram. This fact, as well as the PCA algorithm discussion, leads to the idea that if any measurement point from these 5 areas is selected, it will be 5 required measuring points. Expert knowledge about the microphone behaviour is consistent with the hypothesis. The results of the regression model dimensionality reduction confirmed the hypothesis [4]. The performance of the classification model was satisfying for 5, that is good for coupling with the regression model, and even for 1 measurement point, as shown in Tables 1 and 2.

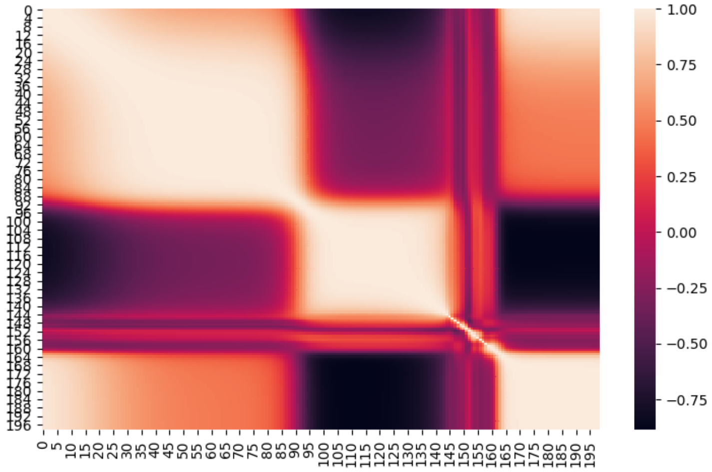

Fig. 4. Amplitude correlation diagrams on the entire data set

Input vector dimension	# layers	# neurons	Accuracy	Training time
10	1	2	99.99%	100 epochs

Table 1. Performance of the classification model with the reduced dimension of the input vector, 5 measurement points

Input vector dimension	# layers	# neurons	Accuracy	Training time
2	1	2	99.99%	100 epochs

Table 2. Performance of the classification model with the reduced dimension of the input vector, 1 measurement point, frequency 36.54 Hz

3. Conclusions

A neural network of simple structure trained for a short time (100 epochs) on a large data set with only two features, determined by statistical methods, accurately, reliably and in real-time recognises the type of microphone as a detector in the photoacoustics experiment. The classification is performed by simple measurement at 1 point in the low-frequency range, instead of the previous 200.

References

[1] M. I. Jordovic-Pavlovic et al., Computationally intelligent description of a photoacoustic detector. Opt. Quantum Electron., 52(5):1–14, 2020.

[2] M. Jordović-Pavlović, M., Kupusinac A., Galovic, S., Markushev, D., Nesic, M., Djordjevic, K., Popovic, Potential of Using Simulated Data in Processing Photoacoustic Measurement Data, in *Proceedings of 8th International Conference on Electrical, Electronic, and Computing Engineering (IcETRAN)*, 2021.

[3] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[4] M. Jordovic Pavlovic, Machine Learning-Based Software Framework for the Automation of Photoacoustic Measurement Data Processing, Univerisity of Novi Sad, 2020.