Kalman filtering in the problem of noise reduction in the absorption spectra of exhaled air

Kistenev[a] Yu.V., Shapovalov[a] A.V., Vrazhnov[b] D. A., Nikolaev[b] V.V. [1]

[a]Tomsk State University, Russian Federation, 634050, Tomsk, Lenina av., 36;
[b]Tomsklabs PTE LTD, 634055, Tomsk, Akademicheskiy av., 8/8;

Abstract

We examined possibilities of the Kalman filter for reducing the noise effects in the analysis of absorption spectra of gas samples, in particular, for samples of the exhaled air.

It has been shown that when comparing groups of patients with broncho-pulmonary diseases on the basis of the absorption spectra analysis of exhaled air samples the data preprocessing with the Kalman filtering can improve the classification sensitivity using a support vector kernel with mpl.

This work was partially supposed by the Federal Target Program for Research and Development, Contract No. 14.578.21.0082 (unique identifier of applied scientific research and experimental development RFMEFI57814X0082).

Kalman filter, exhaled air, noise reduction, support vector machine

Introduction

Non-invasive diagnostic methods of disease based on the analysis of exhaled air are now actively developing [1, 2], and as a patient feature one can use the absorption spectrum of the gas sample [3].

The process of measuring the absorption spectra of the gas samples is associated with exposure to the photoacoustic detector random factors. When using sensitive microphones and low-noise preamplifiers, the external acoustic noise has the crucial meaning [4].

Quite often a small change in the initial conditions can lead to drastic changes both in the intensity of the acting noise and its characteristics. Typically, each component of the test signal may have a complex structure, which is determined by many factors [5,6]. In general, the signal analysis is complicated in the presence of random noise with a priori unknown statistical characteristics.

Thus, a variety of the processing of initial conditions and waveforms describing the components of the analyzed process, determines variety of signal analysis methods. Decrease in the contribution of the noise component in the signal processing results is achieved by filtering algorithms and smoothing primary experimental signal.

In this task it is interesting to do an analysis of preprocessing the experimental spectra to see its effect on various patient's group classification accuracy based on the spectral analysis of exhaled air by an example of the laser

[1] *E-mail:* yuk@iao.ru, shpv@phys.tsu.ru, denis.vrazhnov@gmail.com, vik-nikol@bk.ru

22nd International Symposium Atmospheric and Ocean Optics: Atmospheric Physics, edited by Gennadii G. Matvienko, Oleg A. Romanovskii, Proc. of SPIE Vol. 10035, 100350A ·

· doi: 10.1117/12.2249139

photoacoustic infrared gas analyzer. The paper presents comparison results of the median filter, Kalman filter, and Savitzky-Golay filter.

The analyzed spectrum is a bar chart which describes the dependence of the absorption coefficient on the wavelength. Note that the sample limitation in presence of the noise component is the actual specificity of the experimental health data. The solution for this problem is either in the use of special statistical methods to obtain estimates for small samples (Bootstrap [12]) or in the noise suppression using analytic filters to identify not static, but deterministic regularities.

The carrying out of comparative analysis of the various signal filtering algorithms is also hampered because the classic criterion of "signal / noise" ratio assumes the presence of a reference signal, that is impossible in the framework of the conducted experiments. It has been proposed to assess the quality of the filtering algorithms according to their impact on the classification results.

In the absorption spectra classification problem the classifiers with the teacher are commonly used, e.g., (SIMCA [7], SVM [8], Random Forest [7] and others). However, such classifiers are susceptible to noise and the quality of separation is directly related to the data quality.

In studies on the subject [8,9] some success has been achieved by combining the principal components method to reduce the dimension of the test data and the SVM classifier.

In this paper we used the approach described above to assess the quality of filtration using the Kalman filter, the median filter and the Savitzky-Golay filter.

Kalman filter

The Kalman filter has been considered to reduce the noise effects on the classification of spectra of the exhaled air samples. Also, as a comparison, the data of the median filter and the Savitzky-Golay filter are given. The Kalman filter also allows to correct the data based on evaluation of the system dynamic state. The filtered value is dependent linearly on all previous data, so this filter is called linear. A peculiarity of the Kalman filter is a possibility to establish any mathematical model of the object dynamics. In the ideal case, the model can describe various changes of the object that will reduce or eliminate measurement errors [10]. The system model is assumed to describe a current state at time k based on the state at time k-1.

The rule of changing the state can be expressed as the following:

$$X_k = AX_{k-1} + Bu_k + N(0, R),$$

where X_k is a current state, A is a temporal transformation matrix of states, B is a matrix describing a force action u_k.

This section presents the results of filtering the absorption spectra of the exhaled air, as well as the classification results on the base of the approach that combines principal component analysis and SVM algorithm before and after filtration.

In this paper we have used the experimental data obtained with the gas analyzer “LaserBreese” (produced by “Special Technologies, LTD., Novosibirsk, Russia). As the experimental data, we used the absorption spectra of the exhaled air samples (EAS) in range of 2.5-3.5 microns from 11 patients with chronic obstructive pulmonary disease (COPD) and from 10 healthy volunteers (total 77).

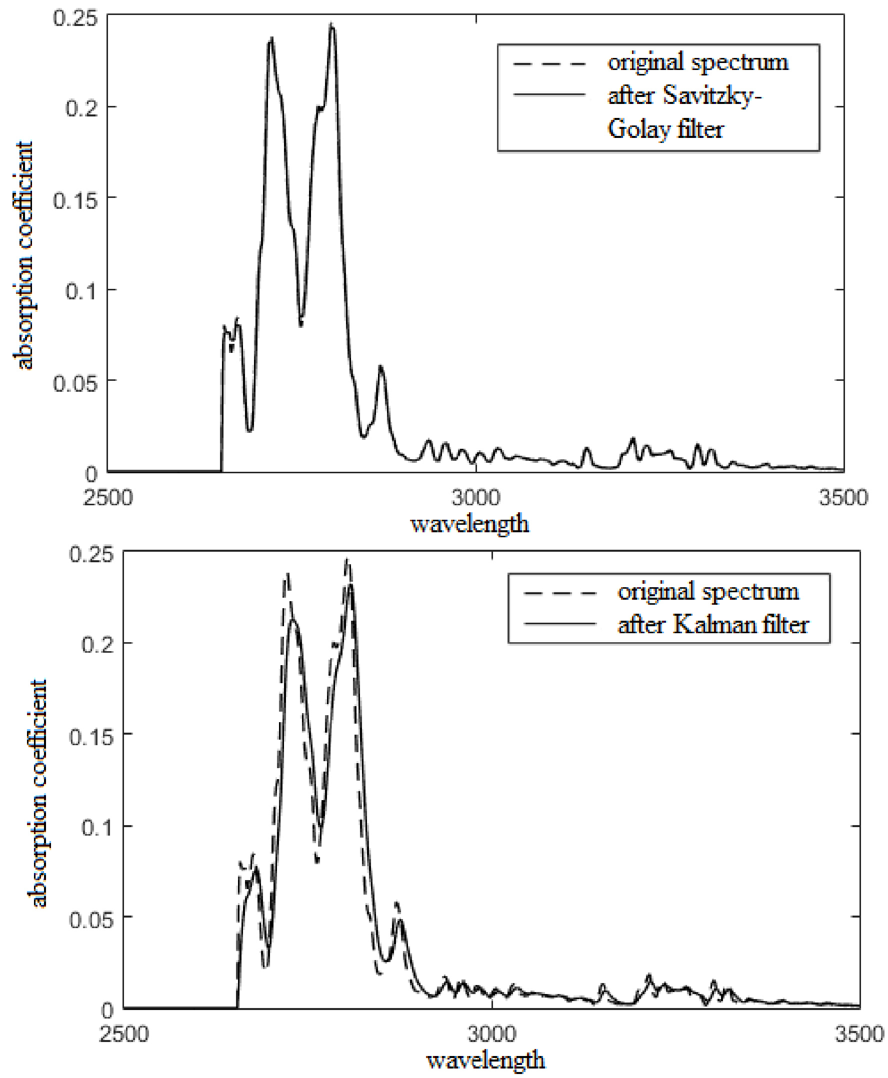

Figure 1. Example of the sample data before and after filtering with the Savitzky-Golay and Kalman filter.

Figure 1 shows examples of the absorption spectra of EAS before and after filtering using the Savitzky-Golay filter with the following parameters: the filter order k = 2, the window size f = 5, and the Kalman filter with the following parameters: A = 1, u = 0, R = 0.05.

The spectrum is seen to be changed significantly after filtering. For the Kalman filter due to its specificity the change occurs with a "delay", however, for the given filter the main peaks remain. The filtering was applied to the original spectrum. After that the principal components were found, which were then used in the step of classification.

The Kalman filter work in the space of principal components is represented in Figure 2 as an example. Figure 2 displays COPD patients' projection onto the plane of 4 and 8 principal components. It can be seen that the scatter of points after filtering becomes smaller.

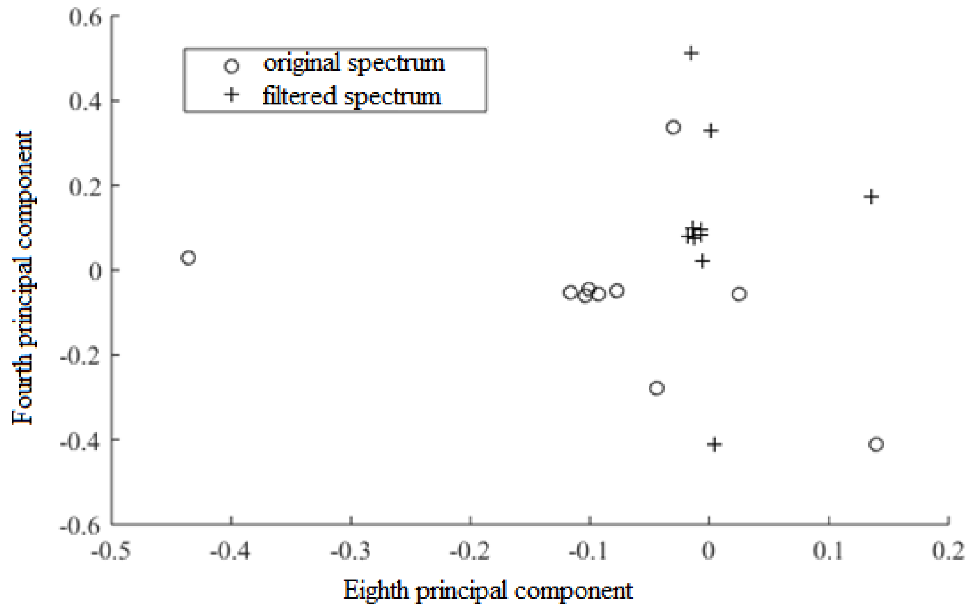

Figure 2. The projections of the objects onto the plane with 4 and 8 main components for the input and filtered spectra.

Figure 2 displayed COPD patients' projection onto the plane with 4 and 8 principal components. It can be seen that the scatter of points after filtering becomes smaller.

As a classifier, we used a support vector machine (SVM) with mpl kernel [8, 11], which applied only to the principal components of a test spectrum.

Because the SVM method includes the step of training on the classified data, 40% of the initial data were used for training and 60% for testing.

Table 1 shows the results of the data classification before and after preprocessing. Here, the sensitivity of the classifier is the ability to select the patients with COPD, and the specificity is the ability to select healthy volunteers.

It can be seen that the sensitivity of the classifier after filtering increases significantly indicating the improvement of the quality of the data separation. Note that the median filter increases both sensitivity and specificity, it is best suited for filtering the analyzed experimental data. The Savitzky-Golay filter does not improve the data separability. The classification accuracy does not become higher for different values of parameters (one of the sets of parameters is shown in the table).

Table 1: Sensitivity and specificity for the SVM-classifier before and after filtering with the Kalman filter.

	Data without filtering	The data after the processing with the Kalman filter	The data after the processing with the Savitzky-Golay filter
Sensitivity	0.8462	0.9231	0.8462
Specificity	0.5000	0.5000	0.5000

Summary

Due to the fact that the classification result depends on the data, and the number of the analyzed data is small, these results make it possible to estimate quality of the filtering algorithm on a small sample.

The results obtained show the increase of sensitivity of the SVM classification when applying the Kalman filter for patients with COPD and healthy volunteers on the basis of the analysis of the absorption spectra of the EAS.

Acknowledgements

This work was partially supposed by the Federal Target Program for Research and Development, Contract No. 14.578.21.0082 (unique identifier of applied scientific research and experimental development RFMEFI57814X0082).

References

[1] Bajtarevic A, Ager C, Pienz M, Klieber M, Schwarz K, Ligor M, Ligor T, Filipiak W, Denz H, Fiegl M, Hilbe W, Weiss W, Lukas P, Jamnig H, Hackl M, Haidenberger A, Buszewski B, Miekisch W, Schubert J, Amann A. Noninvasive detection of lung cancer by analysis of exhaled breath. BMC Cancer 2009;9:348.

[2] Cao W. Breath Analysis: Potential for Clinical Diagnosis and Exposure Assessment, Clinical Chemistry. — №52., 5. — 2006. — P. 800-811.

[3] Bukreeva E.B., Bulanov A.A., Tuzikov S.A., Yumov E.L., Kistenev Y., Kuzmin D.A. Nikiforov O.Y. Features of the absorption spectra of exhaled air of patients with contagious chronic diseases: chronic obstructive pulmonary disease and lung cancer. Biotechnosphere. -2012. - №3-4 (22-23). - P.39-48.

[4] Ponomarev YN, Ageev BG, M. Siegrist, Captains V.A., Courtois D., Nikiforova O.Yu. Laser photoacoustic spectroscopy of intermolecular interactions in gases. -Tomsk: IHP "Rasco", 2000. -200 p.

[5] Sergienko A.B. Digital signal processing, - SPb .: Publisher: BHV-Petersburg, 2011. - 759 p.

[6] Ifeachor E.C., Barrie W. Digital Signal Processing: A Practical Approach (Electronic Systems Engineering) Jervis ebook ISBN: 020154413X, 9780201544138 Publisher: Addison-Wesley Page: 779 2002, 2nd Edition.

[7] Kowalski B. Chemometrics: Theory and Application, American chemical society 1977.

[8] Cortes C., Vapnik V. Support-vector networks, Machine Learning 1995 v. 20, issue 3, pp 273.

[9] Breiman L. Random Forests, Machine Learning 2001 v. 45, issue 1, pp 5-32.

[10] Kistenev Yu.V., Shapovalov A.V., Borisov A.V., Vrazhnov D.A., Nikolaev V.V., Nikiforova O.Y. Applications of principal component analysis to breath air absorption spectra profiles classification, Proceedings of SPIE - The International Society for Optical Engineering. 2015. Vol. 9810. P. 98101Y

[11] Kistenev Yu.V., Borisov A.V., Shapovalov A.V., Kuzmin Dm.A., Vrazhnov D.A., Nikolaev V.V. Applications of laser spectroscopy and data mining methods for complex gas mixture analysis, The 12th International Conference "Atomic and Molecular Pulsed Lasers": Abstracts. Tomsk: Publishing House of IAO SB RAS, 2015. p. 130.

[12] Kalman R.E. A new approach to linear filtering and prediction problems, Journal of Basic Engineering 1960 v. 82, issue 1, pp. 35—45.

[13] Aizerman M. A. Braverman Emmanuel M. Rozoner L. I. Theoretical foundations of the potential function method in pattern recognition learning, Automation and Remote Control 1964 v. 25, pp 821–837.

[14] Efron B., Tibshirani R. An Introduction to the Bootstrap Boca Raton, FL: Chapman & Hall/CRC 1993