NMF in Screening Some Spirometric Data, an Insight into 12-Dimensional Data Space
Anna Bartkowiak, Jerzy Liebhart

To cite this version:
Anna Bartkowiak, Jerzy Liebhart. NMF in Screening Some Spirometric Data, an Insight into 12-Dimensional Data Space. 16th IFIP International Conference on Computer Information Systems and Industrial Management (CISIM), Jun 2017, Bialystok, Poland. pp.155-166, 10.1007/978-3-319-59105-6_14. hal-01656197

HAL Id: hal-01656197
https://hal.inria.fr/hal-01656197
Submitted on 5 Dec 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
NMF in screening some spirometric data, an insight into 12-dimensional data space

Anna M. Bartkowiak\textsuperscript{1} and Jerzy Liebhart\textsuperscript{2}

\textsuperscript{1} Wroclaw University, Inst. of Computer Science, 50-383 Wroclaw PL, aba@cs.uni.edu.pl
\textsuperscript{2} Wroclaw Medical University, Department and Clinic of Internal Diseases and Allergology, Wroclaw PL

Abstract. We present the usage of the Non-negative Matrix Factorization (NMF), an unsupervised machine learning method, which learns normal and abnormal state of patient’s ventilatory systems. This is done using samples of patients having defects of obturative and restrictive kind and a control group. We show that the NMF method can identify patients being in the normal state and screen them off from the remaining patients; however the kind of the ventilatory disorder for the remaining patients is not recognized. This is confronted with clustering provided by the k-means method and visualization of the 12-dimensional data using heatmaps and Kohonen’s self-organizing maps. The data set can be reconstructed with a 0.9746 accuracy (fraction of explained variance) from 6 base vectors provided by the NMF and using appropriate encoders provided also by the NMF; while 3 factors yield an 0.8573 fraction of explained variance.

Keywords: healthy state, abnormal state, non-negative matrix factorization (NMF), heatmap, self-organizing map (SOM), inner factors, reconstruction of data

1 Introduction

Non-linear Matrix Factorization launched in \cite{9} is a method for approximating a given real data matrix by derived factor matrices of lower rank. It may by aligned with two others methods serving for this purpose for a centenary of years: Principal Components (PCA) and Singular Value Decomposition (SVD) \cite{6} based on spectral decomposition of a (preprocessed) data matrix. There are essential differences in the criteria of optimality of these methods, which lead to different factorization of the given data matrix. Generally speaking, NMF works on the assumption that both the analyzed data matrix and the derived factor matrices should be non-negative, that is to mean, their elements are allowed to show only non-negative values. It appears that this leads to better interpretability of the derived factors. The NMF method gains more and more popularity, especially that feasible algorithms working under the constraints of non-negativity
of all the derived elements have been elaborated (see, e.g., [4]). There are also algorithms for big data and data streams [14].

All the three mentioned data analysis methods (PCA, SVD and NMF) are frequently viewed also as dimensionality reducing methods. A satisfactory (‘good’) reduction to dimensionality equal 2 or 3 makes, that the data may be visualized in a 2D or 3D space, and usually we have some additional information on the fidelity of this visualization. By looking at the 2D or 3D representation, we may get insight into the multivariate data space. Such a visualization may also be the basis for finding some structure in the data, e.g. some clusters. This is difficult to achieve when working with PCA or SVD. However NMF, thanks to the non-negative constraints, may provide directly some information on the group membership of the data vectors contained in the analyzed data set. In the following we will consider data vectors that denote some subjects (e.g. patients), each of them characterized by a number of attributes, called in statistical data analysis variables. The clustering information is not always meaningful, and for some data sets we may get the clustering of the subjects, and for other not [15, 1]. Moreover, sometimes we may get even a bi-clustering, that is a simultaneous clustering both of subjects and of variables [10]. Of course, the NMF – as a priori un-supervised machine learning method – gets only the data for analysis, however it gets no a priori information on the putative groups contained in the analyzed data.

The main goal of this paper is to explore the clustering ability of NMF when applied to a real medical data set concerned with diagnosing normal and abnormal state of a sample of patients, part of them suffering from pulmonary diseases, and the other part being in normal state. The data set is not big; it contains 77 patients, each characterized by 12 variables. We will show in detail that NMF is able to screen off the normal state patients, and even find some outliers in the data. We will confirm our results by using alternate visualization methods dealing with multivariate data. This statement ends the first Section of the paper.

In next Section, 2, we will show in more detail the pulmonary data used in our elaboration. They will be visualized by a heatmap. Then, Section 3 shows the details of the NMF model and its work when reducing the dimensions of the data to rank 2 and rank 3. The indication of the clusters will be discussed. Section 4 shows analogous results obtained by the k-means method, when splitting directly the data into $k = 2$ and $k = 3$ clusters. Section 5 shows an alternate multivariate visualization of the of the data using Kohonen’s self-organizing maps. Also the placement of the true group structure obtained from an independent medical diagnosis is depicted. Finally, Section 6 contains a global summary and final conclusions.

2 The analyzed data set

The data were gathered in the Department and Clinic of Internal Medicine and Allergology, Wroclaw Medical University. We consider a multivariate data
sample containing patients being either in the normal or the abnormal state of their ventilatory system. The abnormal state was diagnosed as obturation or restriction in ventilatory disorders. In the elaborated sample counting \( n = 77 \) patients we have:

- Group 1, \( n_1 = 28 \) patients with obturative disorders,
- Group 2, \( n_2 = 21 \) patients with restrictive disorders,
- Group 3, \( n_3 = 28 \) patients with normal functioning, serving as control.

Each patient has its record (data vector) composed of 15 values: Patient’s ID no., residual value (RV), total lung capacity (TLC), weight, height and 10 various spirometric variables derived from the patient’s spirogram. The data were gathered with the aim to predict the RV and TLC variables (they need measurements using a different device) just from the spirometric variables considered as predictors [11, 3]. In the following we will use from these records only 12 variables defined in Table 1.

### Table 1. Labels and definitions of the variables used in our analysis

| X1: Age [years], | X2: Height [cm], | X3: FVC, Forced Vital Capacity [cm³], | X4: FVC% evaluated as \( \frac{FVC}{FVC_{predicted}} \times 100 \) [%], | X5: \( FEV_1 \), Forced Expiratory Volume in one second [cm³], | X6: Ratio \( FEV_1/FVC \times 100 \) [%], | X7: \( FEF_{0.2-1.2} \), Forced Expiratory Flow at the level of 0.2 – 1.2 dm³ of FVC [dm³/min], | X8: \( MMFR \), Maximal Mid-expiratory Flow Rate [dm³/min], | X9: \( MMFT \), Maximal Mid-expiratory Flow Time [sec], | X10: Ratio \( MMFR/MMFT \) [dm³/min²], | X11: Ratio \( \frac{FEF_{0.2-1.2}/FVC}{1000 \times X7/X3} \) [min⁻¹], | X12: Ratio \( FVC/MMFT \) [cm³/sec]. |

The twelve variables displayed in Table 1 will be hereafter called 'spirometric variables'. The first two of them are weight and height, the remaining ten – as obtained from a spirogram – are sensu stricto spirometric variables.

The recorded data are memorized as the \( n \times m \) matrix \( X \) (with \( n = 77 \) and \( m = 12 \)) containing in its rows the patients and in its columns the values of the variables characterizing the patients. The rows of \( X \) contain firstly the \( n_1 = 28 \) patients of group 1 (obturation), next the \( n_2 = 21 \) patients from group 2 (restriction), and finally the \( n_3 = 28 \) patients from group 3 (normal, control).

All elements of the data matrix \( X \) were strictly positive, however its columns were expressed in different scales. To annihilate the effects of largely differentiated scales of variables, the columns of the data matrix were standardized by
'range' to take values from the [0,1] interval. The data matrix with standardized elements will be referred to as $X_s$.

A holistic view of a not very big data matrix with non-negative elements may be obtained by displaying its values in the form of a heatmap. As such, the data matrix is considered as an image in the scale of an assumed color palette defined in computer graphics as colormap. Figure 1 shows such a heatmap of the transposed matrix $(Xs)^T$ of our data, when using the colormap $JET$ (the default in Matlab).

![Fig. 1. Heatmap for $(Xs)^T$, the transposed standardized data matrix](image)

The vertical constituents (columns) of the image are 77 columns numbered 1, 2, …, 77. They correspond to the subsequent patients constituting our data set. The no.s in the x-axis denote patients. Each column $j$ visualizes the 12-dimensional data vector recorded for the $j$-th patient. In particular, no.s 1-28 indicate patients from group 1 (obturation), no.s 29-49 from group 2 (restriction), and no.s 50-77 from group 3 (normal, control).

Looking at Fig. 1 one may state that the patient group 3 (no.s 50-77) is clearly different as the groups 1 and 2 with no.s 1–49. The conclusion is: the group 3 of patients differs much in its inner structure from the the remaining two groups which look rather similar. Therefore, the ‘normal’ group should be rather easily identifiable in the data, while for the disease groups (obturation and restriction) this is doubtful.

3 The NMF method

The NMF (non-negative matrix approximation) method was launched in [9] with the idea to find inner (hidden, unobservable) structure of the data which could be parameterized by lower rank factor matrices permitting to reconstruct the observed data matrix $V$. It was assumed that the approximation (alias reconstruction) will be of the form $V \approx WH$, with $W$ and $H$ denoting the inner structure
factors found using an appropriate goodness of fit criterion. The methodology proved to yield very interesting applications and was subsequently developed using other variants of approach (see, e.g., [4, 16, 15, 14, 17]) and exploiting its properties (see, e.g., [10, 1, 12, 2]). In the following we use matrix denotation $V, W, H$ introduced in [9].

Suppose, we have data observed for $n$ subjects characterized by $m$ variables denoting attributes of the subjects. If we are more interested in the subjects (in their relationship, clustering) then we should put them in the columns of the matrix $V$ to be dealt with by NMF. In such a case the model NMF is given as

$$V_{m \times n} \approx W_{m \times r} H_{r \times n},$$

(1)

where $r$ is usually much smaller as $\min(n, m)$ and is called rank of approximation.

The column vectors in $W$ are called base vectors, they play the role of a kind of prototypes (representatives) of the subjects included into the data. If the data set contains $r$ groups (clusters), then the derived base vectors are expected to be representatives of these groups [5].

The column vectors in $H = [h_1, \ldots, h_n]$ are called encoders or coefficient vectors, they play a crucial role in reconstructing the matrix $V$ from the base vectors $W = [w_1, \ldots, w_r]$.

We will carry out the NMF approximation of the observed (visible) data matrix $V_{12 \times 77}$ by rank $r = 2$ and $r = 3$ factors. Computations will be done using the Matlab function snmf available at http://mikkelschmidt.dk/code.html. The criterium of goodness of fit of the model is the explained variance defined as follows

$$\text{explained variance} = (sst - sse)/sst,$$

(2)

where $sst$ denotes the total sum of squares of deviations of elements $v_{ij}$ from their overall mean, and $sse$ is the sum of squared terms of errors $e_{ij} = v_{ij} - (WH)_{ij}$.

We start computing the NMF factorization for rank $r = 2$, obtaining the approximation $V_{12 \times 77} \approx W_{12 \times 2} H_{2 \times 77}$. In such a case there are only two base vectors $w_1, w_2$. Simple linear algebra shows, that these two vectors, combined in various proportions (each proportions taken as one column vector from the set $H = [h_1, \ldots, h_n]$), permit to reconstruct in sequence the subjects vectors $v_1, \ldots, v_n$ contained in $V$. For example, the 1st and 2nd subjects from $V$, that is the subject vectors $v_1$ and $v_2$ are reconstructed as

$$v_1 = w_1 h_{11} + w_2 h_{21}, \quad v_2 = w_1 h_{12} + w_2 h_{22}, \ldots$$

(3)

and so on. The elements of the coefficient vector $h_j$ say how important are subsequent base vectors $w_1, \ldots, w_r$ in reproducing just the $j$-th subject vector. This may be seen globally when inspecting the row profiles of $H$ depicted in Fig. 2 (for $r = 2$) and in Fig. 3 (for $r = 2$).

We formulate the following Principle of group assignment in NMF:

Find for given $j$, ($j = 1, 2, \ldots, n$), of which the coefficients $h_{ij}$, ($i = 1, \ldots, r$) is the largest. Say this is the $i_{\text{max}}$ coefficient. Then assign the $j$-th data vector to the group no. $i_{\text{max}}$. 
Fig. 2. NMF for $r = 2$. Top: Profiles of the coefficient matrix $H_{2 \times 77}$ obtained when using the entire data matrix $V$. Notice the outstanding green pics for no.1 and no.32. Bottom: Profiles of the coefficient matrix $H_{2 \times 75}$ obtained when using the data matrix $V$ from which the outlying vectors no.1 and no.32 were removed. Notice in both exhibits the discrepancy of the green and blue curves for the last 28 indices in the x-axis indicating patients from the control group.

Figure 2, top exhibit, shows the row profiles plot of the coefficient matrix $H$ obtained for $r = 2$. We see the values $h_{1j}$ and $h_{2j}$, for the indices $j = 1, 2, \ldots, 77$, connected into two curves colored in blue and green respectively. The blue and green curves show - for the index $j$ - the prevalence of the subject coefficient $h_{ij}$ to the first or second basic vector. This prevalence means that the respective base vector has a larger contribution in reconstructing the object no.$j$. In such a way the subjects are assigned to one of the two groups represented by the two base vectors.

Looking at the top exhibit in Fig. 2 we see there a particular intertwining of the two curves. For $j = 1$ to 49 (patients with obturation or restriction) the two curves are intermingled with a slight prevalence of the blue curve. Moreover, for $j = 1$ and $j = 32$ there are two big eruptions of the green curve. For $j = 50$ to
77 (patients from the normal group) the green curve connected with the base vector \( w_2 \) is decidedly dominating.

The analysis was repeated with the data set reduced to 75 patients (patients no.1 and no.32 notified as outliers were removed). Results of the new analysis, without the two outliers, are shown in bottom exhibit of Fig. 2 and look more consistent. The essential pattern noticed in the top panel is present also here: the control group is decidedly connected with the base vector \( w_2 \); the obturation and restriction groups show the same intermingled pattern with a mild preference of the base vector \( w_1 \).

**Fig. 3.** NMF for rank \( r = 3 \). Top: Profiles of the coefficient matrix \( H_{3 \times 77} \) obtained when using the entire data matrix \( V \). Bottom: Profiles of the coefficient matrix \( H_{3 \times 75} \) obtained when using the data matrix \( V \) from which the subjects no.1 and no.32 were removed. Notice the big discrepancy of the red curve from both the green and the blue curve when observed for the last 28 indices in the x-axis – the red curve indicating the control group - is for no.s starting from no.49 decidedly dominating. Notice also the appearance of two new pics of the red curve for no.s covering the first group (obturation).
When looking at Fig. 2 one may come to the following conclusion: Similarly as in the heatmap, one sees here clearly a subdivision of the entire data into two groups: the ‘normal’ state group (no.s 50 through 77) and the abnormal group designated by no.s 1-49. The group assignment for the normal group is good: 26 out of 28 patients are correctly located. Thus the first kind error is small. However the second kind error is considerable: several patients from the disease group are assigned to the normal, i.e. healthy group. Removing some putative outliers reduces the erroneous assignments in the disease group.

The analysis of the results obtained when assuming 3 groups of data (see Fig. 3) shows clearly that the assumption on 3 groups of data cannot be sustained.

Concerning the efficacy of the approximation measured by the fraction of explained variance (2): The models using \( r = 2 \) and \( r = 3 \) explain 79.18% and 87.43% of total variance. Model with \( r = 6 \) yields 97.46% of explained variance.

4 Clusters by classic k-means

For comparison with the above results by NMF, search for \( k = 2 \) and \( k = 3 \) clusters using the classical k-means algorithm was carried out. Below the group classification indices obtained from this method. The calculations were performed twice:

- For the full data matrix \( X_s \) size 77 × 12,
- For the reduced data matrix \( Z_s \) size 77 × 12 (without two outliers discovered by the NMF).

The results are shown in Table 2 (for \( k = 2 \) and Table 3 (for \( k = 3 \)) below. Notice that the k-means algorithm has established its own cluster denotations which is different from our group denotation linked with the diagnostic of obstructive ("O") or restrictive ("R") ventilatory disorders, or normal ("N") state of the patient.

| Table 2. Results of k-means for \( k = 2 \) |
|---------------------------------------------|
| group ↓ subdividing data \( X_s \) into \( k=2 \) subgroups | k-means indicators | frequencies |
| g1 "O" n=28 | 12222222222221212222222222222222 | 1(5), 2(23) |
| g2 "R" n=21 | 212122222222222222221 | 1(3), 2(18) |
| g3 "N" n=28 | 112211111111111111111211122 | 1(23), 2(5) |
| \( Z_s \), i.e. \( X_s \) with two outliers removed | | |
| g1 "O" n=27 | 222222222222221222222122222222 | 1(4), 2(23) |
| g2 "R" n=20 | 21222222222222222222221 | 1(2), 2(18) |
| g3 "N" n=28 | 1111111111111111111111111122 | 1(26), 2(2) |
Table 3. Results of k-means for $k = 3$

| group | subdividing data $X_s$ into $k=3$ subgroups | k-means indicators | frequencies |
|-------|-------------------------------------------|-------------------|-------------|
| g1 "O" n=28 | 12223223222312223331232333 | 1(3), 2(15) 3(10) |
| g2 "R" n=21 | 322132232332232323 | 1(1), 2(7) 3(13) |
| g3 "N" n=28 | 122211111111112112112 | 1(19), 2(9) 3(0) |

$Z_s$, i.e. $X_s$ with two outliers removed

| group | subdividing data $X_s$ into $k=3$ subgroups | k-means indicators | frequencies |
|-------|-------------------------------------------|-------------------|-------------|
| g1 "O" n=27 | 3332332233221322322322 | 1(2) 2(10) 3(15) |
| g2 "R" n=20 | 23322232222222222222 | 1(0) 2(13) 3(7) |
| g3 "N" n=28 | 133311311111111113113113131 | 1(19) 2(0) 3(9) |

Working with $k = 2$ clusters the k-means yielded reasonable results similar to those obtained by NMF: the control group was diagnosed with 5 and 2 false assignments. They appear in Table 2 under the heading 'frequencies'.

For $k = 3$ clusters the results from k-means were much worse as those by NMF: the control group was diagnosed with 9 false assignments both in $X_s$ and $Z_s$. The assignment to groups 2 and 3 was a mess of '2' and '3' assignments.

5 Visualization of the data using Kohonen’s SOM

Now we visualize the data matrix $X_s$ using Kohonen’s SOMs. The SOM (self-organizing map) is de facto a neural network which reflects specifically the positions of data points in the multi-dimensional data space with preserving the data neighborhood topology. The principles of the method and the tricky algorithm are described in [8]. Maps obtained for our data - obtained using the free Matlab SOM toolbox [13] - are shown in Fig. 4. One may notice there that the maps have rectangular shape and are composed from $M = 11 \times 4 = 44$ equal size hexagons.

The SOM methodology performs a vector quantization and subdivides the data space into $M$ so called Voronoi regions. The regions contain all the data points and reflect their density in the data space; some regions may be empty. There is a strict correspondence between the hexagons in the map and the Voronoi regions in the multivariate data space. We can find which data points are very near each other, and which are distant. Each Voronoi region has a representative data point (called codebook vector), which is in one-to-one correspondence to the center of hexagon associated with the respective Voronoi Region. There are two fitness indices: the quantization error $vqe$ (means the average distance of data points in one Voronoi region to their representative codebook vector located in that region; for our data $vqe=0.304$) and the topological error $tpe$ (says, how many hexagons being neighboring hexagons in the map are not neighboring Voronoi regions in the data space; for our map we got $tpe = 0.0000$).
Fig. 4. Data groups assignment in Kohonen SOM. Letters O, R, N denote data points identified with patients belonging to the Obturation, Restriction or Normal group according the definition in Section 2. Colors red, green, blue are linked with groups "O", "R", "N" appropriately. Notice, that the "N"s appear in the southern part of the maps, while the "O"s and the "R"s reside in the northern part of the maps.

Now, where are our patient vectors from the categories "O", "R", and "N"? We found and identified all of them using the hit utility of the toolbox. The respective information (how many items from each group are represented by each hexagon) is inscribed as ordinary text onto the map located in the right exhibit of Fig. 4. E.g., one may notice in the displayed map, that the utmost south-west hexagon (coordinates 11,1) contains three "Ns" and one "R" item (this "R" item happens to be the outlier no. 32 found by NMF and visible in Fig. 2, top exhibit).
Having the allocation of each data vector, one may display them group-wise, assigning to them specific colors and drawing in each map hexagon a smaller colored one, with radius proportional to the square root frequency of items linked to the given hexagon. The result is shown in the map appearing in Fig. 4, left exhibit. Looking at that exhibit, it is interesting to find that items of the 3 groups of our data are spread out over the entire map. Blue color (denoting ”N” items) concentrates rather in the south of the map, however the region is not pure. In particular, there are two outliers at the south edge (no.s 1 and 32 notified previously and removed from part of our analyzes shown in Section 3). There is also one other ”O” and one other ”R” outlier located at the south-west edge. The ‘disease’ items (linked with red and green) appear intermixed in the upper part of the same map, however some of them invade the blue territory of the ”N” items. This observation is in agreement with the former comment on the results of NMF, that the results displayed in Fig. 2 and Fig. 3 are supporting the assumption on only \( r = 2 \) discernible clusters in the analyzed data.

6 General summary and concluding remarks

The aim of the investigation was to explore the role of the NMF (non-negative matrix factorization) method when applied to real medical data. We have considered for this purpose a medical data set obtained from spiographic investigation of 3 groups of patients diagnosed as having two kinds of ventilatory disorders (g1, obturation, \( n=28 \), and g2, restriction, \( n=21 \)) and additionally a control group (g3, normal, \( n=28 \)). Each patient was characterized by 12 variables, 10 of them being spirogram characteristics.

The NMF is a model based approach which can find in data some hidden factors (unobservable inner structure) permitting to reconstruct the entire data set with a hopefully small error. It was reported, that this inner structure might be connected with some groups into which the investigated data samples can be subdivided. In medical context this might mean subdivision into subgroups of various subtypes, e.g. various types of a disease. Our question was: Can the various types (groups of patients appearing in our data) be recognized – in a unsupervised way – by the NMF algorithm?

Our results. We have applied the NMF method seeking for \( k=2 \) and \( k=3 \) inner factors. The results from NMF were confronted with the classic k-means and two graphical methods: Kohonen SOM and heatmaps permitting to get a holistic vision of the 12-dimensional data vectors contained in the analyzed data set.

Conclusions from the investigation:
1. The NMF is able to recognize the normal group (92.86 perc accuracy), however it can not recognize properly the abnormal state groups.
2. The NMF has an ability to recognize outliers hidden in the data.
3. Comparing to clustering abilities of the classic k-means, the NMF is superior.
4. The overall inspection of the results leads to the conclusion that the NMF has some abilities to recognize the group structure and can be considered as a ‘weak
learner’, that - along with others weak learners - might be useful in ensemble learning.

References

1. Bartkowiak A.M.: Classic and convex non-negative matrix visualization in clustering two benchmark data. Przegląd Elektrotechniczny R93, 1, 53-59 (2017)
2. Bartkowiak A.M., Zimroz R.: NMF and PCA as applied to gearbox data, K. Jankowski, et al., (Eds): Intelligent Data Engineering and Automated Learning - IDEAL 2015 LNCS 9375, pp. 199-206 (2015)
3. Bartkowiak, A., Liebhart, E.: Estimation of the spirometric residual volume (RV) by a regression built from Gower distances. Biometrical Journal 37, 2, 131-149 (1995)
4. Cichocki A., Zdunek R., Phan A.H., Amari Sh.: Nonnegative matrix and tensor factorizations. Applications to exploratory multi-way data analysis and blind source separation. Wiley, Chichester U.K. (2009)
5. Ding C., Li T., Jordan M.: Convex and semi-nonnegative matrix factorizations. IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI), 32, 45-55 (2010)
6. Jolliffe I.T.: Principal Component Analysis, 2nd edn. Springer, New York (2002)
7. Kasim A. et al. (eds), Applied Biclustering Methods for Big and High-dimensional data using R. CRC Press, Taylor & Francis Group, A Chapman & Hall Book, Boca Raton FL (2017)
8. Kohonen T.: Self-organizing Maps, Springer, Heidelberg, Third Extended Edition in 501 pp. (2001)
9. Lee D.D., Seung H.S.: Learning the parts of objects by nonnegative matrix factorization. Nature, 401, 788–791 (1999)
10. Li Y., Ngom A.: The non-negative matrix factorization toolbox for biological data mining. BMC Source Code for Biology and Medicine, 8:10, 1–15 (2013)
11. Liebhart J., Bartkowiak A., Liebhart E.: The impact of outliers in in the regression estimating TLC from age and some spirometric observations. Modelling, Simulation & Control, C, AMSE Press, 15, 1–19 (1989)
12. Schmidt M.N., Larsen J., and Hsiao F-T.: Wind Noise Reduction using Nonnegative Sparse Coding, in Machine Learning for Signal Processing, IEEE International Workshop on (MLSP), 431-436, Aug. (2007).
13. Vesanto J. et al.: SOM Toolbox for Matlab 5, Som Toolbox Team, HUT, Finland. Libella Oy, Espoo, Version 0beta 2.0, 1–54 November (2001)
14. Zdunek R.: Extraction of nonnegative features from multidimensional nonstationary signals. Y. Tan and Y. Shi (Eds), DMBD 2016, LNCS 9714, 557–566 (2016).
15. Zdunek R.: Convex non-negative matrix factorization with Rank-1 update for clustering. Rutkowski L., et al. (Eds.): ICAISC 2015, LNAI 9120, 59-68, Springer-Verlag Berlin Heidelberg (2015)
16. Zdunek R.: Data clustering with semi-binary non-negative matrix factorization. Rutkowski L., et al. (Eds.): ICAISC 2008, LNAI 5097, 705-716, Springer-Verlag Berlin Heidelberg (2008).
17. Zurada J.M., Ensari T., Asi E.H., Chorowski J.: Nonnegative matrix factorization and its application to pattern recognition and text mining. Proc. of the 13th Federated Conference on Computer Science and Information Systems, Cracow, 11–16, (2013)