2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Controlling Orthogonality Constraints for better NMF Clustering

REDKO Ievgen and BENNANI Younès

***Abstract*— In this paper we study a variation of a Non-negative Matrix Factorization (NMF) called the Orthogonal NMF(ONMF). This special type of NMF was proposed in order to increase the quality of clustering results of standard NMF by imposing orthogonality on clustering indicator matrix and/or the matrix of basis vectors. We develop an extension of ONMF which we call Weighted ONMF and propose a novel approach for imposing orthogonality on the matrix of basis vectors obtained via NMF using Gram-Schmidt process.**

I. Introduction

Clustering is a well-known machine learning technique used for unsupervised classification of patterns (observations, data items, or feature vectors) into groups of similar objects. The groups given by a clustering algorithm are called "clusters", each cluster consists of objects that are similar between themselves but different from objects in other clusters. There are three main types of machine learning algorithms:

- supervised learning (when data is labeled in both training and test sets)
- semi-supervised learning (data is labeled only in small training test)
- unsupervised learning (no labeled data available)

Clustering is usually associated with unsupervised learning. Unsupervised learning itself is extremely important setting of machine learning algorithms as it occurs in numerous real-world applications. Main reasons that show why unsupervised learning can prove beneficial are:

- labeling a set of objects manually can be hard or even impossible on large amounts of data
- it can be used to classify a huge amount of unlabeled data to further label it manually
- it can be used to find a set of variables that can be useful for further categorization

There exists numerous unsupervised learning methods that were applied in many contexts and by researchers in many disciplines. Typical applications of clustering are: statistics [1], pattern recognition [2], image segmentation and computer vision [3], multivariate statistical estimation [4]. Clustering is also widely used for data compression in image processing, which is also known as vector quantization [5]. Most general survey about clustering methods can be found in [6].

There exist lots of well-known clustering algorithms, notably: k-means, mixture models, hierarchical clustering, non-negative matrix factorization etc. Among all the methods used for clustering we will discuss the one called Non negative matrix factorization (NMF).

NMF is a group of algorithms in machine learning where a data matrix is factorized into (usually) two matrices with the property that all three matrices have no negative elements. This non-negativity makes the resulting matrices easier to interpret. We consider one of the matrices as a matrix containing the prototypes of a data set and the other one as a data partition matrix. Since this optimization problem is not convex in general, it is commonly approximated numerically.

A. Background

DIFFERENT kinds of constraints can be imposed on cluster's properties in order to achieve better clustering results. One of the most common constraints that is used for clustering is orthogonality of subspaces of clusters. Indeed, imposing orthogonality on the subspaces of clusters means that we try to find two clusters that are very different. In our case, orthogonality constraints imposed on matrices obtained with NMF is considered to be useful as it results in unique factorization and has a good clustering interpretation. The main goal of this work is to explore whether hard orthogonality constraints in NMF are really beneficial or there is some level of orthogonality that leads to a better clustering result.

B. Related works

The idea of Uni- and Bi-Orthogonal NMF was first described in [7] where it was introduced as a special form of a standard NMF [8] which increases the quality of clustering and provides an unique non negative matrix factorization (which is rare for this type of matrix factorizations). In [7], authors proposed a novel approach for solving this kind of optimization problems and showed that their update rules have a non-increasing property even though there was no robust proof of convergence. In [9], authors imposed orthogonality on matrices of Tri-NMF by adding supplementary terms directly into the cost-function instead of solving it as a constrained optimization problem (that is the case for [7]). Their approach has a robust convergence proof and it is mainly inspired by [10] but with its further generalization for matrices that have auxiliary constraints with mutually dependency between columns and/or rows.

C. Our contributions

In our work we can highlight two main contributions:

- We studied the effectiveness of orthogonality constraints in Uni-Orthogonal NMF
- We proposed a novel approach called Gram-Schmidt Orthogonal NMF and a modification of classical Uni- Orthogonal NMF called Weighted Uni-Orthogonal NMF

REDKO Ievgen and BENNANI Younès are with Laboratoire d'Informatique de Paris-Nord, CNRS (UMR 7030), Université Paris 13, Sorbonne Paris Cité, F-93430, Villetaneuse, France(email: {ievgen.redko, younes.bennani}@lipn.univ-paris13.fr).

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

The rest of this paper is organized as follows: in section 2 we will briefly introduce basic notations of standard, Semi, Uni-Orthogonal and Bi-Orthogonal non-negative matrix factorizations, in section 3 we are introducing Weighted Orthogonal NMF and Gram-Schmidt NMF and derive both multiplicative and additive update rules for them. We will summarize the results in section 4. Finally, we will point out some ideas about the future extension of our method in section 5.

II. Preliminary knowledge

A. Standard and Semi-NMF

A standard NMF seeks the following decomposition:

$$X \simeq FG^T, X \in \mathbb{R}^{n\times m}, F \in \mathbb{R}^{n\times k}, G \in \mathbb{R}^{m\times k}$$
$$X, F, G \geq 0.$$

where

- X is an input data matrix
- columns of F can be considered as basis vectors
- columns of G are considered as cluster assignments for each data object
- k is the desired number of clusters

When the data matrix is unconstrained (i.e., it may have mixed signs), Semi-NMF is a factorization in which we restrict G to be non-negative while placing no restriction on the signs of F.

B. Uni- and Bi-Orthogonal NMF

The Bi-Orthogonal NMF (BONMF) seeks the following decomposition:

$$X \simeq FSG^T,$$
$$X \in \mathbb{R}^{n\times m}, F \in \mathbb{R}^{n\times k}, S \in \mathbb{R}^{k\times l}, G \in \mathbb{R}^{m\times l},$$
$$F^TF = I, G^TG = I, X, F, S, G \geq 0.$$

The multiplicative update rules for matrices F, G and S have the following form:

$$F = F \circledast \frac{XGS^T}{FF^TXGS^T}$$
$$S = G \circledast \frac{F^TXG}{F^TFSG^TG}$$
$$G = G \circledast \frac{X^TFS}{GG^TX^TFS}$$

The Uni-Orthogonal NMF (UONMF) imposes orthogonality constraint on either columns of F or rows of G. It is clear that this variation is just a special case of BONMF with $S = I$.

The authors of Orthogonal NMF mentioned that the full orthogonality of matrices F and G cannot be achieved using their algorithm because it uses an approximate solution for non diagonal elements of the Lagrange multipliers matrix. So, their solution of this optimization problem does not result in a set of fully orthonormalized vectors.

C. Gram-Schmidt process

There exists three well known approaches that can be used for orthonormalizing a set of vectors: Householder reflections [11], Givens rotations[12] and Gram-Schmidt process. In our work we will use the Gram-Schmidt process [13]. The Gram - Schmidt process takes a finite, linearly independent set $S = \{v_1, ..., v_k\}$ for $k \leq n$ and generates an orthogonal set $S' = \{u_1, ..., u_k\}$ that spans the same k-dimensional subspace of $\mathbb{R}^n$ as S.

The projection operator is defined by the following expression

$$proj_u(v) = \frac{\langle u, v\rangle}{\langle u, u\rangle}u$$

where $\langle u, v\rangle$ denotes the inner product of the vectors u and v. This operator projects the vector v orthogonally onto the line spanned by vector u.

The Gram-Schmidt process works as follows:

$$u_1 = v_1$$
$$u_2 = v_2 - proj_{u_1}(v_2)$$
$$u_3 = v_3 - proj_{u_1}(v_3) - proj_{u_2}(v_3)$$
$$.....$$
$$u_i = v_i - \sum_{j=1}^{k-1} proj_{u_j}(v_i)$$

The sequence $u_1, ..., u_k$ is the required system of orthogonal vectors, and the normalized vectors $e_1, ..., e_k$ form an orthonormal set. The calculation of the sequence $u_1, ..., u_k$ is known as Gram-Schmidt orthogonalization, while the calculation of the sequence $e_1, ..., e_k$ is known as Gram-Schmidt orthonormalization as the vectors are normalized ($e_i = \frac{u_i}{\|u_i\|}, \forall\, i = 1..k$)

III. Proposed approach

We would like to test an another way of imposing orthogonality on the basis vectors found by Standard NMF that is to perform the Gram-Schmidt process to obtain a set of vectors which is close to an orthonormal basis of the initial space.

A. Gram-Schmidt Orthogonal NMF (GS-ONMF)

First step consist in calculating a set of basis vectors A which arises from the Standard NMF:

$$X \simeq FG^T, X \in \mathbb{R}^{n\times m}, F \in \mathbb{R}^{n\times k}, G \in \mathbb{R}^{m\times k}$$
$$X, F, G \geq 0.$$

Performing the Gram-Schmidt process does not result in a set of positive vectors due to the properties of the inner product. So, we will need to use a Semi-NMF with its matrix A fixed to the orthonormal basis.

On the other hand, using fully orthonormal set of basis vectors for Semi-NMF gives very poor results because it

attempts to discover only one dominating class ignoring all the others. So, we need to find some appropriate level of orthogonality between vectors of matrix A.

To do that, we multiply the projection operator by a constant $\alpha_{GS} \in [0; 1]$:

$$proj_u^*(v) = \alpha_{GS} \frac{\langle u, v \rangle}{\langle u, u \rangle} u$$

so that the final set of vectors will be changing from the initial set of vectors ($\alpha_{GS} = 0$) till fully orthogonized vectors ($\alpha_{GS} = 1$). We denote by $F_{\alpha_{GS}}$ - a result of a Gram-Schmidt process obtained for some arbitrary α_{GS}.

We will change α_{GS} - value and use the matrix $F_{\alpha_{GS}}$ as a fixed matrix of basis vectors in Semi-NMF:

$$X \simeq F_{\alpha_{GS}} G^T, X \in \mathbb{R}^{n \times m}, F_{\alpha_{GS}} \in \mathbb{R}^{n \times k}, G \in \mathbb{R}^{m \times k}$$
$$G \geq 0.$$

All along this way the orthogonality measured by the following expression

$$Orthogonality = \| F_{\alpha_{GS}}^T * F_{\alpha_{GS}} \|$$

increases (approaching 1).

B. Weighted Uni-Orthogonal NMF

Inspired by the previous idea, we would like to introduce the Weighted Uni-Orthogonal NMF. As in UONMF the cost function rests the same but with some slight modification of the orthogonality condition. So, we try to solve the following optimization problem:

$$X \simeq FG^T,$$
$$X \in \mathbb{R}^{n \times m}, F \in \mathbb{R}^{n \times k}, G \in \mathbb{R}^{m \times k},$$
$$F^T F = I, X, F, G \geq 0.$$

By doing this, we try to find some appropriate angle between basis vectors while increasing orthogonality between them so that we can obtain some improvement in clustering result. Our main idea is to show that fully orthogonal constraints for NMF work worse than some kind of soft orthogonality assured by some parameter α .

To solve this optimization problem we rewrite it in the following form:

$$\min J = \|X - FG^T\|_F^2 + \|F^T F - \alpha I\|_F^2$$

Using the gradient descent approach and switching alternatively between three sets of parameters, we obtain the following update rules:

$$F = F \circledast \frac{XG^T + \alpha F}{FGG^T + FF^T F}$$
$$G = G \circledast \frac{F^T X}{F^T FG}$$

These update rules were derived by calculating partial differentials of J and using the following scheme:

$$X = X \circledast \frac{\left[\frac{\partial J}{\partial X}\right]_-}{\left[\frac{\partial J}{\partial X}\right]_+}.$$

where X represents all the variables involved in the cost function, $\left[\frac{\partial J}{\partial X}\right]_+$ stands for positive part of gradient and $\left[\frac{\partial J}{\partial X}\right]_-$ for negative.

Expression for α that minimizes the objective function is given by the following expression:

$$\alpha = \frac{tr(F^T F)}{k}$$

Using the gradients calculated before we can also determine now the additive update rules.

They are of the form:

$$F = F + \eta_F (FGG^T + FF^T F - XG^T - \alpha F)$$
$$G = F + \eta_G (F^T FG - F^T X)$$

where

$$\eta_F = \frac{F}{FGG^T + FF^T F}$$
$$\eta_G = \frac{G}{F^T FG}$$

As it was shown in [9] both multiplicative and additive update rules assure convergence of the algorithm. It is also worth mentioning that our update rules differ from the update rules presented before.

IV. Experimental results

We used 30-fold cross-validation and evaluated the results using two different measures: entropy and purity [14]. These are the standard measures of clustering quality in supervised setting. Entropy measures how the various semantic classes are distributed within each cluster. Given a particular cluster S_r of size n_r , the entropy of this cluster is defined to be:

$$E(S_r) = -\frac{1}{q} \sum_{i=1}^{q} \frac{n_r^i}{n_r} log \frac{n_r^i}{n_r}$$

where q is the number of classes in the data set, and n_r^i is the number of elements of the ith class that were assigned to the rth cluster. The entropy of the entire clustering is then the sum of the individual cluster entropies weighted according to the cluster size:

$$entropy = \sum_{r=1}^{k} \frac{n_r}{n} E(S_r)$$

The clustering solution is perfect if clusters only contain words from one single class; in that case the entropy of the clustering solution is zero. Smaller entropy values indicate better clustering solutions.

Using the same mathematical definitions, the purity of a cluster is defined as:

$$Pu(S_r) = \frac{1}{n_r} \max_i n_r^i$$

The purity gives the fraction of the overall cluster size that the largest class of elements assigned to that cluster represents. The purity of the clustering solution is then again the weighted sum of the individual cluster purities:

$$purity = \sum_{r=1}^{k} \frac{n_r}{n} Pu(S_r)$$

Larger purity values indicate better clustering solutions.

We will also study how orthogonality constraints influence the sparsity of F. We use the sparseness function [15] given by the following expression:

$$sparseness(x) = \frac{\sqrt{n} - \frac{\|x\|_1}{\|x\|_2}}{\sqrt{n} - 1}$$

where n is a size of vector x. For a given prototype matrix, we calculate the average sparseness of all the columns.

In Table 1, 2 and 3 we can see the results of the experimental tests of our approach for data sets from UCI repository.

The following data sets were chosen:

- Iris dataset: this is perhaps the best known database to be found in the pattern recognition literature. Fisher's paper is a classic in the field and is referenced frequently to this day. The dataset contains 3 classes of 50 instances each described by 4 features (sepal length, sepal width, petal length, petal width), where each class refers to a type of iris plant (Setosa, Versicolour, Virginica). One class is linearly separable from the other 2; the latter are not linearly separable from each other [16].
- Wine dataset: these data are the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. This dataset consists of 178 instances described by 13 features and divided in three classes [17].
- Glass dataset: the glass identification database is composed of 214 examples with 9 variables, the data is divided in seven classes [18].
- Yeast dataset: this database contains information about a set of Yeast cells and consists of 1484 observations divided into 10 classes and described by 10 features. The task is to determine the localization site of each cell [19].
- Hepatitis dataset: the clynical observations of people that were suffering from hepatitis is composed of 155 examples with 19 variables, the data is divided in 2 classes [20].
- Ecoli dataset: the prokaryotic gram-negative bacterium Escherichie Col (E.coli) 336 patterns data are classified to eight classes and described by 8 features. Classes of this data set are drastically imbalanced [19].
- Australian Credit Approval dataset: this database concerns credit card applications. There are 6 numerical and 8 categorical attributes. All data represents two classes of applications: accepted and declined [21].
- Breast Cancer Wisconsin: this breast cancer databases was obtained from the University of Wisconsin Hospitals with 10 attributes divided into two classes [22].
- Heart disease: these data are the results of 303 observation of heart diseases. This database is composed of 13 fetures and is divided into 5 classes [23].

TABLE I

NMF, ONMF(F), W-ONMF(F) AND GS-NMF(F) PURITY VALUES ON VARIOUS DATA SETS

Data set	NMF	ONMF(F)	GS-NMF(F)	W-ONMF
Iris(3)	0.7876	0.6567	0.7944	**0.8453**
Ecoli(8)	0.6831	0.6255	0.7043	**0.7559**
Wine(3)	0.6333	0.4391	0.6348	**0.6463**
Glass(7)	0.5847	0.5760	**0.7055**	0.6732
Australian(2)	0.6784	0.6783	0.6784	**0.6792**
Hepatitis(2)	0.7935	0.7935	**0.8013**	0.7935
Breast Wisconsin(2)	0.7694	0.7670	0.8215	**0.8406**
Heart(5)	0.5413	0.5414	**0.5501**	0.5425
Yeast(10)	0.4378	0.3610	0.4679	**0.4860**

TABLE II

NMF, ONMF(F), W-ONMF(F) AND GS-NMF(F) CLUSTER ENTROPY VALUES ON VARIOUS DATA SETS

Data set	NMF	ONMF(F))	GS-NMF(F)	W-ONMF
Iris(3)	0.4179	0.6050	0.5186	**0.2024**
Ecoli(8)	0.3880	0.4653	0.1948	**0.1847**
Wine(3)	0.7300	0.9424	0.3358	**0.1746**
Glass(7)	0.5166	0.5862	**0.1148**	0.1954
Australian(2)	0.9035	0.9049	**0.0555**	0.0557
Hepatitis(2)	0.6433	0.7145	**0.0191**	0.7935
Breast Wisconsin(2)	0.7302	0.7247	**0.0865**	0.0914
Heart(5)	0.7760	0.7889	0.0845	**0.0247**
Yeast(10)	0.6110	0.6755	0.3216	**0.1240**

TABLE III

NMF, ONMF(F), W-ONMF(F) AND GS-NMF(F) PROTOTYPES SPARSENESS VALUES ON VARIOUS DATA SETS

Data set	NMF	ONMF(F))	GS-NMF(F)	W-ONMF
Iris(3)	0.3702	**0.7694**	0.4222	0.5051
Ecoli(8)	0.4472	**0.8854**	0.4550	0.4759
Wine(3)	0.7300	**0.9424**	0.3358	0.1746
Glass(7)	0.8206	**0.9563**	0.8281	0.8203
Australian(2)	**0.9953**	0.9758	0.9540	0.9653
Hepatitis(2)	0.2655	**0.5009**	0.2659	0.2705
Breast Wisconsin(2)	0.2185	**0.3605**	0.2362	0.1600
Heart(5)	0.7286	**0.8852**	0.7580	0.7582
Yeast(10)	0.4989	**0.9277**	0.5212	0.4876

It is easy to see that two proposed approaches outperform NMF and ONMF for all data sets in terms of entropy and purity. At the same time, the values of sparseness obtained for each method are rather surprising. ONMF gives the most sparse representations of objects but fails to give the best result in terms of quality. It means that high sparseness of prototypes is not always beneficial for clustering using NMF. This result shows that forcing the prototype-matrix to be very sparse changes the initial nature of data and as a result spoils clustering quality. Nevertheless, very sparse prototype matrices can be very useful in such tasks as blind source separation or part-based representations of objects. The difference in results can be also explained by the fact that in [7] and [9] the approaches used to solve the optimization problem were different. In Figures 1-9 we show that orthogonality of F approaches 1 while we change α - multiplier of a projection operator and reaches it for almost every data set. The blue bars indicate on what level of orthogonality the best clustering results were achieved. It is worth to be mentioned that for data sets with a quite high number of classes the process of orthonormalization sometimes results in a set which is only approximately orthonormalized (as for ecoli and yeast). Nevertheless, it does not affect the results because we are still able to increase the orthogonality monotonically for some sub interval.

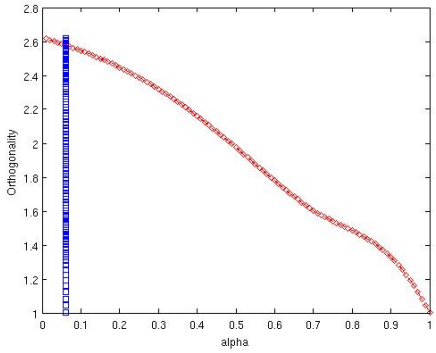

Fig. 1. Iris data set

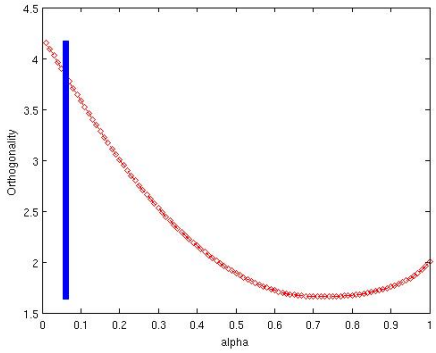

Fig. 2. Ecoli data set

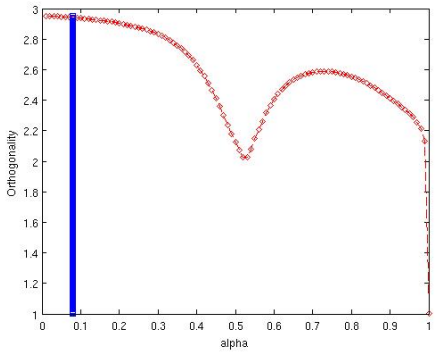

Fig. 3. Wine data set

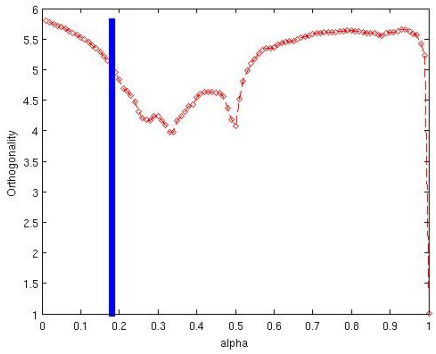

Fig. 4. Glass data set

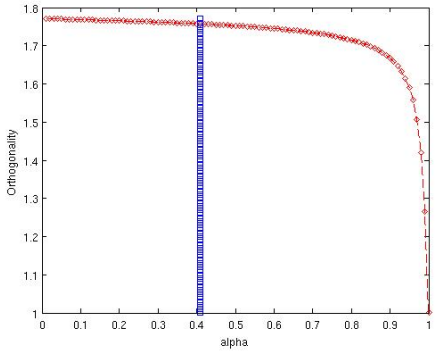

Fig. 5. Australian Credit data set

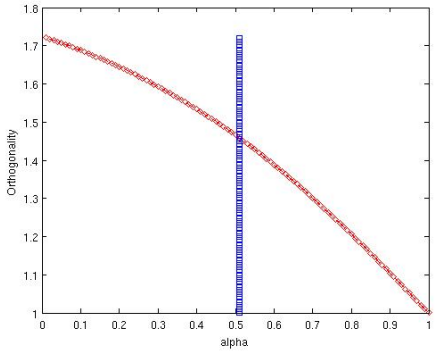

Fig. 6. Hepatitis data set

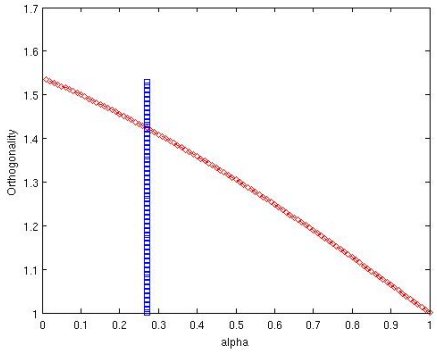

Fig. 7. Breast Cancer Wisconsin data set

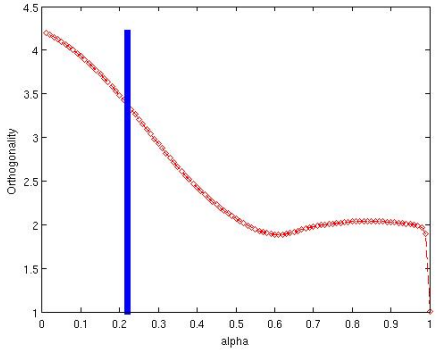

Fig. 8. Heart data set

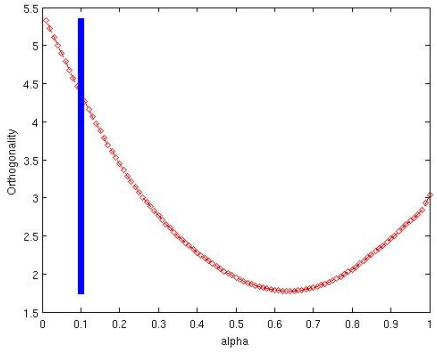

Fig. 9. Yeast data set

These figures clearly show that in almost all cases the best solution was found when vectors were not fully orthonormal.

V. Conclusions

In this paper we explore the orthogonality constraints of Uni-Orthogonal NMF. We use the Gram-Schmidt process to orthonormalize basis vectors obtained via Standard NMF to further use them as the basis vectors in Semi-NMF. We introduce also the Weighted Orthogonal NMF where the orthogonality of basis vectors depends on the parameter. Considering the results obtained during our experiments, we can assume that imposing hard orthogonality constraints is not as effective as searching an appropriate level of orthogonality between vectors of the prototype matrix.

In future, our work can be extended in the multiple directions. The main question that remains open is how to choose the level of orthogonality that assure the best clustering result. To answer this question further theoretical studies should be conducted in order to found an analytic expression for the optimal value of α parameter. Another way of solving this problem is to choose a suitable technique for " optimization. This can be done, for example, by constructing a generative model of ONMF and by imposing some arbitrary prior to control the orthogonality of the prototypes matrix.

References

[1] Arabie, P., Hubert, L.J., *An overview of combinatorial data analysis*, World Scientific Publishing Co., 1996.

[2] Duda, R. and Hart, P., *Pattern Classification and Scene Analysis*, 1973.

[3] Jain, A.K. and Flynn, P.J., "Image segmentation using clustering," *In Advances in Image Understanding: A Festschrift for Azriel Rosenfeld*, IEEE Press, 65-83, 1966.

[4] Scott, D.W., *Multivariate Density Estimation*, Wiley, New York, NY, 1992.

[5] Gersho, A. and Gray, "Vector Quantization and Signal Compression," *Communications and Information Theory*, Kluwer Academic Publishers, 1992.

[6] Han, J. and Kamber, M., *Data Mining*, Morgan Kaufmann Publishers, 2001.

[7] C. H. Ding, T. Li, and M. I. Jordan, "Convex and SemiNonnegative Matrix Factorizations," *IEEE Trans. on Pattern Analysis and Machine Intelligence.* vol. 99, no. 1, 2006.

[8] D. Lee and H. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, 401(6755), pp. 788-91, 1999.

[9] Andri Mirzal, "Converged Algorithms for Orthogonal Nonnegative Matrix Factorizations," *Computing Research Repository*, Vol. 1010. 2010

[10] C.J. Lin, "On the convergence of multiplicative update algorithms for nonnegative matrix factorization," *IEEE Transactions on Neural Networks*, Vol. 18(6), 2007.

[11] Anderson, E., "Discontinuous Plane Rotations and the Symmetric Eigenvalue Problem," *LAPACK Working Note 150*, December 4, 2000.

[12] Householder, A. S., "Unitary Triangularization of a Nonsymmetric Matrix," *Journal of the ACM*, 5 (4): 339342.

[13] Golub, G. H. and Van Loan, C. F., "Matrix Computations," *3rd ed. Baltimore, MD: Johns Hopkins*, 1996.

[14] Y. Zhao and G. Karypis, "Empirical and theoretical comparisons of selected criterion functions for document clustering," *E. Machine Learning*, 55(3):311331, 2004.

[15] Hoyer, P. O., "Non-negative Matrix Factorization with sparseness constraints," *Journal of Machine Learning Research*, vol.5, pp. 1457-1469, 2004.

[16] Gates, G.W., "The Reduced Nearest Neighbor Rule," *IEEE Transactions on Information Theory*, 431-433, May 1972.

[17] S. Aeberhard, D. Coomans and O. de Vel, "The classification performance of RDA," *Tech. Rep. no. 92-01*, Dept. of Computer Science and Dept. of Mathematics and Statistics, James Cook University of North Queensland, 1992.

[18] Ian W. Evett and Ernest J. Spiehler, "Rule Induction in Forensic Science," *Central Research Establishment*, Home Office Forensic Science Service. Aldermaston, Reading, Berkshire RG7 4PN.

[19] Paul Horton and Kenta Nakai, "A Probablistic Classification System for Predicting the Cellular Localization Sites of Proteins," *Intelligent Systems in Molecular Biology*, 109-115. St. Louis, USA 1996.

[20] Diaconis,P. and Efron,B., *Computer-Intensive Methods in Statistics*, Scientific American, Volume 248, 1983.

[21] Ross Quinlan, "Simplifying decision trees," *Int J Man-Machine Studies 27*, pp. 221-234, 1987.

[22] Wolberg, W.H., and Mangasarian, O.L., "Multisurface method of pattern separation for medical diagnosis applied to breast cytology," *In Proceedings of the National Academy of Sciences*, pp. 9193–9196, 1990.

[23] Gennari, J.H., Langley, P, and Fisher, D., "Models of incremental concept formation," *Artificial Intelligence*, vol. 40, pp. 11-61, 1989.