Sparse Symmetric Nonnegative Matrix Factorization Applied to Face Recognition

Hennadii Dobrovolskyi, Nataliya Keberle, Yehor Ternovyy
Zaporizhzhya National University

The 9 th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, 21-23 September, 2017, Bucharest, Romania

Motivation

Symmetric nonnegative matrix factorisation (SNMF) is used successfully in different fields and was proved to provide the fine results.

Advantages:

- utilization of the pairwise similarity instead of metrics, i.e. feature space can be not metric one,
- any shape of cluster in the feature space.

Drawbacks:

- It is not unique
- The dimensionality of the multiplies is unknown

To the best of our knowledge there is no published attempts to achieve sparsity in SNMF providing more understandable and compact results.

Objectives are

1. Implementation of the Sparse SNMF (SSNMF)
2. Study of the SSNMF
    a. application to face recognition task;
    b. comparison to known clustering approaches such as k-means and spectral clustering.

Problem Statement

SSNMF optimization problem

$$||A - HH^T||_F^2 + \lambda \sum_{i,j} |H_{ij}| \to min,$$

$$H_{ij} \geq 0, A_{ij} = A_{ji}$$

where $||Z||_F^2$ denotes squared Frobenius norm of a matrix *Z*,

the parameter *λ* affects both sparsity level and factorization accuracy.

The matrix *A* is assumed to be symmetric similarity matrix.

Optimization is achieved with appropriate choice of the factor H_{ij} .

Projected Gradient Descent

Update rule

$$H_{ab}^{(n+1)} = max\left(0, H_{ab}^{(n)} + \delta\nabla_{ab}\right)$$

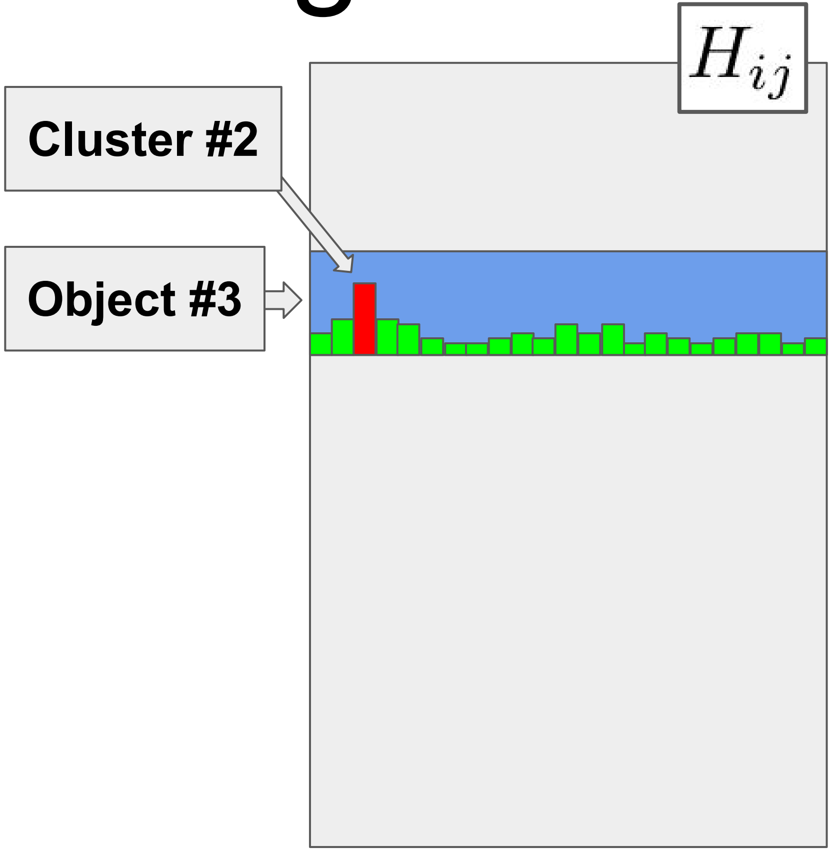
where

$$\nabla_{ab} = 4\left[\sum_j \left(\sum_p H_{ap}H_{jp} - A_{aj}\right) H_{jb}\right] + 1$$

δ is a variable step size

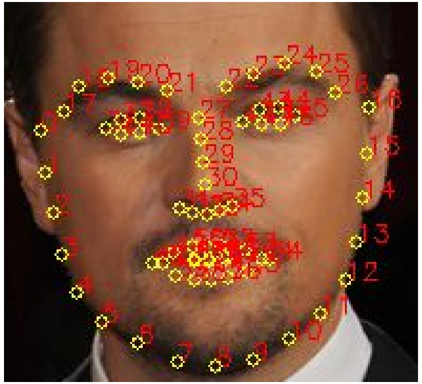
SSNMF. Clustering

The cluster number
for object *j*
is the position of
maximal element in
j-th row of the matrix H_{ij}



Application to face recognition

Face localization (histogram of oriented gradients, HOGs)



Retrieving a vector of face features from image (pre-trained neural facial landmark detector inside Dlib library)

Sparse
Symmetric
Nonnegative
Martix
Factorization

classifier training and authentication of a facial image

Dataset

The Yale Face Database

Contains 165 grayscale images in GIF format of 15 individuals. There are 11 images per subject, one per different facial expression or configuration: center-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised, and wink.

Clustering quality

- **Adjusted Rand Index (ARI)**
  is a measure of agreement between two partitions: one given by the clustering process and the other defined by external criteria.

- **Homogeneity score**
  shows the chance that each cluster contains only members of a single class.

- **Completeness score**
  is related to the probability of the fact that the members of a given class are assigned to the same cluster.

Clustering quality

Method	Scores		
	ARI	Homogeneity	Completeness
SSNMF (λ = 0.1)	**0.746**	**0.987**	**0.830**
K-means	0.729	0.979	0.822
Spectral clustering	0.656	0.920	0.774

Sparsity parameter λ	**Scores**		
	ARI	**Homogeneity**	**Completeness**
0	**0.753**	0.983	**0.833**
0.01	0.723	**0.987**	**0.837**
0.02	**0.758**	0.978	**0.842**
0.03	0.721	**0.987**	0.819
0.04	0.726	0.981	0.820
0.10	**0.746**	**0.987**	0.830

SSNMF-based PCA

Base of the principal component analysis (PCA) is a factorization

$$\boldsymbol{A} = \boldsymbol{W}^{T} \boldsymbol{P} \boldsymbol{W}$$

In a common PCA, $\boldsymbol{W}$ is the set of eigenvectors, $\boldsymbol{P}$ is diagonal matrix of eigenvalues.

In a SSNMF-based PCA, we set

$$\boldsymbol{H} = \sqrt{\boldsymbol{P}}\, \boldsymbol{W}$$

Then $\boldsymbol{W}$ is a factor, the sequence of a positive vectors having unit length,
$\boldsymbol{P}$ is diagonal matrix of the weights that are squares of the vector norms.

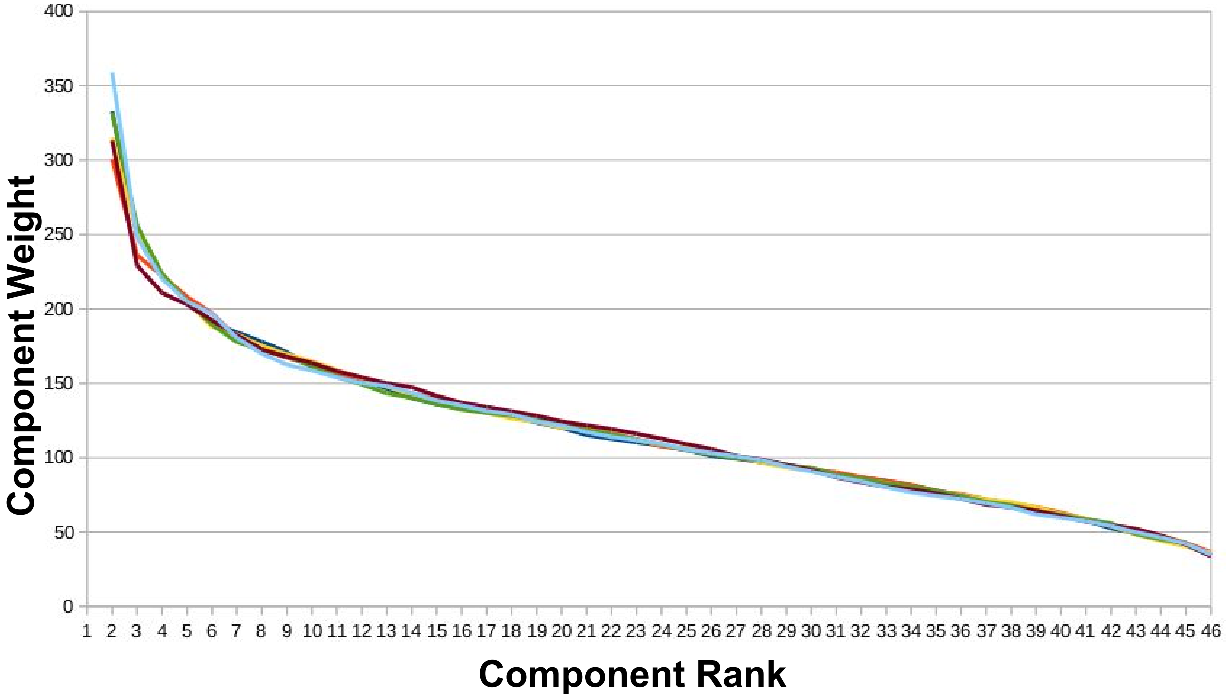
In both cases, we set $\boldsymbol{P}_{ii}\mathbf{=0}$ if $\boldsymbol{P}_{ii} < \boldsymbol{\varepsilon}$ for some small $\boldsymbol{\varepsilon}$.

SSNMF+PCA. Face clustering

Principal component weights P_{ii} ordered by weight for different values sparsity parameter λ.

Relative value of sparsity term is 0.46 . Natural threshold does not exist.

Suggestion: the feature selection is optimal and we cannot get more sparse representation.



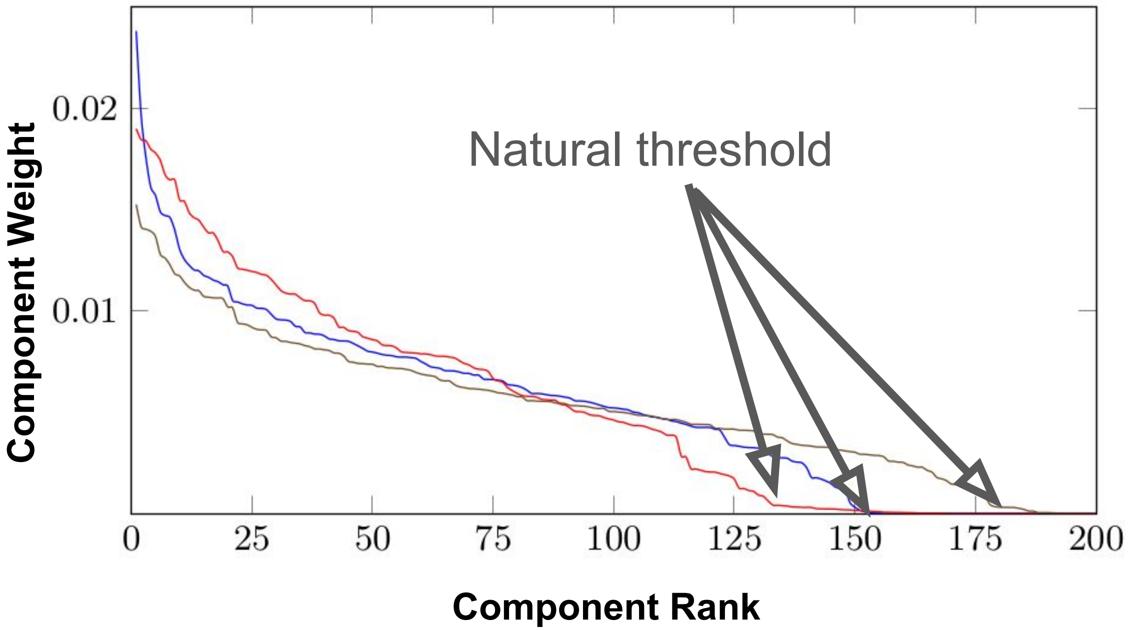
SSNMF+PCA. Topic model

Principal component weights P_{ii} ordered by weight value.

Relative value of sparsity is 0.87

Natural threshold does exist.

Suggestion: the feature selection is not optimal and we can increase the sparsity.



Conclusions

SSNMF

- **can compete** in clustering quality with other well-known algorithms.
- can detect clusters of any form;
- has the wide range of applications **utilizing any type of object similarity**;
- In the current **implementation**, it has the **low scalability** because of matrix operations;
- allows **adjusting the level of sparsity** on account of factorization accuracy leading to easier understanding and robustness;
- If the feature set is far from optimal it allows application of Principal Component Analysis to combine the features.