Non-negative matrix factorization

Non-negative matrix factorization (**NMF** or **NNMF**), also **non-negative matrix approximation**[1][2] is a group of algorithms in multivariate analysis and linear algebra where a matrix **V** is factorized into (usually) two matrices **W** and **H**, with the property that all three matrices have no negative elements. This non-negativity makes the resulting matrices easier to inspect.
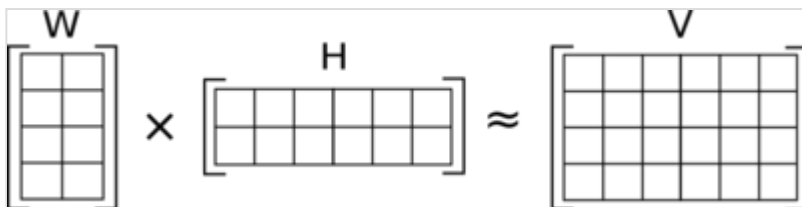
Illustration of approximate non-negative matrix factorization: the matrix **V** is represented by the two smaller matrices **W** and **H**, which, when multiplied, approximately reconstruct **V**.

Also, in applications such as processing of audio spectrograms or muscular activity, non-negativity is inherent to the data being considered. Since the problem is not exactly solvable in general, it is commonly approximated numerically.

NMF finds applications in such fields as astronomy,[3][4] computer vision, document clustering,[1] missing data imputation,[5] chemometrics, audio signal processing, recommender systems,[6][7] and bioinformatics.[8]

History

In chemometrics non-negative matrix factorization has a long history under the name "self modeling curve resolution".[9] In this framework the vectors in the right matrix are continuous curves rather than discrete vectors. Also early work on non-negative matrix factorizations was performed by a Finnish group of researchers in the 1990s under the name *positive matrix factorization*.[10][11][12] It became more widely known as *non-negative matrix factorization* after Lee and Seung investigated the properties of the algorithm and published some simple and useful algorithms for two types of factorizations.[13][14]

Background

Let matrix **V** be the product of the matrices **W** and **H**,

$$\mathbf{V} = \mathbf{WH}.$$

Matrix multiplication can be implemented as computing the column vectors of **V** as linear combinations of the column vectors in **W** using coefficients supplied by columns of **H**. That is, each column of **V** can be computed as follows:

$$\mathbf{v}_i = \mathbf{Wh}_i,$$

where $\mathbf{v}_i$ is the i-th column vector of the product matrix **V** and $\mathbf{h}_i$ is the i-th column vector of the matrix **H**.

When multiplying matrices, the dimensions of the factor matrices may be significantly lower than those of the product matrix and it is this property that forms the basis of NMF. NMF generates factors with significantly reduced dimensions compared to the original matrix. For example, if $\mathbf{V}$ is an $m \times n$ matrix, $\mathbf{W}$ is an $m \times p$ matrix, and $\mathbf{H}$ is a $p \times n$ matrix then p can be significantly less than both m and n.

Here is an example based on a text-mining application:

- Let the input matrix (the matrix to be factored) be $\mathbf{V}$ with 10000 rows and 500 columns where words are in rows and documents are in columns. That is, we have 500 documents indexed by 10000 words. It follows that a column vector $\mathbf{v}$ in $\mathbf{V}$ represents a document.
- Assume we ask the algorithm to find 10 features in order to generate a *features matrix* $\mathbf{W}$ with 10000 rows and 10 columns and a *coefficients matrix* $\mathbf{H}$ with 10 rows and 500 columns.
- The product of $\mathbf{W}$ and $\mathbf{H}$ is a matrix with 10000 rows and 500 columns, the same shape as the input matrix $\mathbf{V}$ and, if the factorization worked, it is a reasonable approximation to the input matrix $\mathbf{V}$.
- From the treatment of matrix multiplication above it follows that each column in the product matrix $\mathbf{WH}$ is a linear combination of the 10 column vectors in the features matrix $\mathbf{W}$ with coefficients supplied by the coefficients matrix $\mathbf{H}$.

This last point is the basis of NMF because we can consider each original document in our example as being built from a small set of hidden features. NMF generates these features.

It is useful to think of each feature (column vector) in the features matrix $\mathbf{W}$ as a document archetype comprising a set of words where each word's cell value defines the word's rank in the feature: The higher a word's cell value the higher the word's rank in the feature. A column in the coefficients matrix $\mathbf{H}$ represents an original document with a cell value defining the document's rank for a feature. We can now reconstruct a document (column vector) from our input matrix by a linear combination of our features (column vectors in $\mathbf{W}$) where each feature is weighted by the feature's cell value from the document's column in $\mathbf{H}$.

Clustering property

NMF has an inherent clustering property,[15] i.e., it automatically clusters the columns of input data $\mathbf{V} = (v_1, \ldots, v_n)$.

More specifically, the approximation of $\mathbf{V}$ by $\mathbf{V} \simeq \mathbf{WH}$ is achieved by finding W and H that minimize the error function (using the Frobenius norm)

$$\|V - WH\|_F, \text{ subject to } W \geq 0, H \geq 0.,$$

If we furthermore impose an orthogonality constraint on $\mathbf{H}$, i.e. $\mathbf{HH}^T = I$, then the above minimization is mathematically equivalent to the minimization of K-means clustering.[15]

Furthermore, the computed H gives the cluster membership, i.e., if $\mathbf{H}_{kj} > \mathbf{H}_{ij}$ for all $i \neq k$, this suggests that the input data v_j belongs to k-th cluster. The computed W gives the cluster centroids, i.e., the k-th column gives the cluster centroid of k-th cluster. This centroid's representation can be significantly enhanced by convex NMF.

When the orthogonality constraint $\mathbf{HH}^T = I$ is not explicitly imposed, the orthogonality holds to a large extent, and the clustering property holds too.

When the error function to be used is Kullback–Leibler divergence, NMF is identical to the probabilistic latent semantic analysis (PLSA), a popular document clustering method.[16]

Types

Approximate non-negative matrix factorization

Usually the number of columns of $\mathbf{W}$ and the number of rows of $\mathbf{H}$ in NMF are selected so the product $\mathbf{WH}$ will become an approximation to $\mathbf{V}$. The full decomposition of $\mathbf{V}$ then amounts to the two non-negative matrices $\mathbf{W}$ and $\mathbf{H}$ as well as a residual $\mathbf{U}$, such that: $\mathbf{V} = \mathbf{WH} + \mathbf{U}$. The elements of the residual matrix can either be negative or positive.

When $\mathbf{W}$ and $\mathbf{H}$ are smaller than $\mathbf{V}$ they become easier to store and manipulate. Another reason for factorizing $\mathbf{V}$ into smaller matrices $\mathbf{W}$ and $\mathbf{H}$, is that if one's goal is to approximately represent the elements of $\mathbf{V}$ by significantly less data, then one has to infer some latent structure in the data.

Convex non-negative matrix factorization

In standard NMF, matrix factor $\mathbf{W} \in \mathbf{R}_+^{m \times k}$, i.e., $\mathbf{W}$ can be anything in that space. Convex NMF[17] restricts the columns of $\mathbf{W}$ to convex combinations of the input data vectors $(v_1, \ldots, v_n)$. This greatly improves the quality of data representation of $\mathbf{W}$. Furthermore, the resulting matrix factor $\mathbf{H}$ becomes more sparse and orthogonal.

Nonnegative rank factorization

In case the nonnegative rank of $\mathbf{V}$ is equal to its actual rank, $\mathbf{V} = \mathbf{WH}$ is called a nonnegative rank factorization (NRF).[18][19][20] The problem of finding the NRF of $\mathbf{V}$, if it exists, is known to be NP-hard.[21]

Different cost functions and regularizations

There are different types of non-negative matrix factorizations. The different types arise from using different cost functions for measuring the divergence between $\mathbf{V}$ and $\mathbf{WH}$ and possibly by regularization of the $\mathbf{W}$ and/or $\mathbf{H}$ matrices.[1]

Two simple divergence functions studied by Lee and Seung are the squared error (or Frobenius norm) and an extension of the Kullback–Leibler divergence to positive matrices (the original Kullback–Leibler divergence is defined on probability distributions). Each divergence leads to a different NMF algorithm, usually minimizing the divergence using iterative update rules.

The factorization problem in the squared error version of NMF may be stated as: Given a matrix $\mathbf{V}$ find nonnegative matrices W and H that minimize the function

$$F(\mathbf{W}, \mathbf{H}) = \|\mathbf{V} - \mathbf{WH}\|_F^2$$

Another type of NMF for images is based on the total variation norm.[22]

When L1 regularization (akin to Lasso) is added to NMF with the mean squared error cost function, the resulting problem may be called **non-negative sparse coding** due to the similarity to the sparse coding problem,[23][24] although it may also still be referred to as NMF.[25]

Online NMF

Many standard NMF algorithms analyze all the data together; i.e., the whole matrix is available from the start. This may be unsatisfactory in applications where there are too many data to fit into memory or where the data are provided in streaming fashion. One such use is for collaborative filtering in recommendation systems, where there may be many users and many items to recommend, and it would be inefficient to recalculate everything when one user or one item is added to the system. The cost function for optimization in these cases may or may not be the same as for standard NMF, but the algorithms need to be rather different.[26][27]

Convolutional NMF

If the columns of $\mathbf{V}$ represent data sampled over spatial or temporal dimensions, e.g. time signals, images, or video, features that are equivariant w.r.t. shifts along these dimensions can be learned by Convolutional NMF. In this case, $\mathbf{W}$ is sparse with columns having local non-zero weight windows that are shared across shifts along the spatio-temporal dimensions of $\mathbf{V}$, representing convolution kernels. By spatio-temporal pooling of $\mathbf{H}$ and repeatedly using the resulting representation as input to convolutional NMF, deep feature hierarchies can be learned.[28]

Algorithms

There are several ways in which the $\mathbf{W}$ and $\mathbf{H}$ may be found: Lee and Seung's multiplicative update rule[14] has been a popular method due to the simplicity of implementation. This algorithm is:

initialize: $\mathbf{W}$ and $\mathbf{H}$ non negative.
Then update the values in $\mathbf{W}$ and $\mathbf{H}$ by computing the following, with n as an index of the iteration.

$$\mathbf{H}_{[i,j]}^{n+1} \leftarrow \mathbf{H}_{[i,j]}^{n} \frac{((\mathbf{W}^n)^T \mathbf{V})_{[i,j]}}{((\mathbf{W}^n)^T \mathbf{W}^n \mathbf{H}^n)_{[i,j]}}$$

and

$$\mathbf{W}_{[i,j]}^{n+1} \leftarrow \mathbf{W}_{[i,j]}^{n} \frac{(\mathbf{V}(\mathbf{H}^{n+1})^T)_{[i,j]}}{(\mathbf{W}^n \mathbf{H}^{n+1}(\mathbf{H}^{n+1})^T)_{[i,j]}}$$

Until $\mathbf{W}$ and $\mathbf{H}$ are stable.

Note that the updates are done on an element by element basis not matrix multiplication.

We note that the multiplicative factors for $\mathbf{W}$ and $\mathbf{H}$, i.e. the $\frac{\mathbf{W}^T \mathbf{V}}{\mathbf{W}^T \mathbf{W} \mathbf{H}}$ and $\frac{\mathbf{V} \mathbf{H}^T}{\mathbf{W} \mathbf{H} \mathbf{H}^T}$ terms, are matrices of ones when $\mathbf{V} = \mathbf{W} \mathbf{H}$.

More recently other algorithms have been developed. Some approaches are based on alternating non-negative least squares: in each step of such an algorithm, first $\mathbf{H}$ is fixed and $\mathbf{W}$ found by a non-negative least squares solver, then $\mathbf{W}$ is fixed and $\mathbf{H}$ is found analogously. The procedures used to solve for $\mathbf{W}$ and

H may be the same[29] or different, as some NMF variants regularize one of **W** and **H**.[23] Specific approaches include the projected gradient descent methods,[29][30] the active set method,[6][31] the optimal gradient method,[32] and the block principal pivoting method[33] among several others.[34]

Current algorithms are sub-optimal in that they only guarantee finding a local minimum, rather than a global minimum of the cost function. A provably optimal algorithm is unlikely in the near future as the problem has been shown to generalize the k-means clustering problem which is known to be NP-complete.[35] However, as in many other data mining applications, a local minimum may still prove to be useful.

In addition to the optimization step, initialization has a significant effect on NMF. The initial values chosen for **W** and **H** may affect not only the rate of convergence, but also the overall error at convergence. Some options for initialization include complete randomization, SVD, k-means clustering, and more advanced strategies based on these and other paradigms.[36]

Sequential NMF
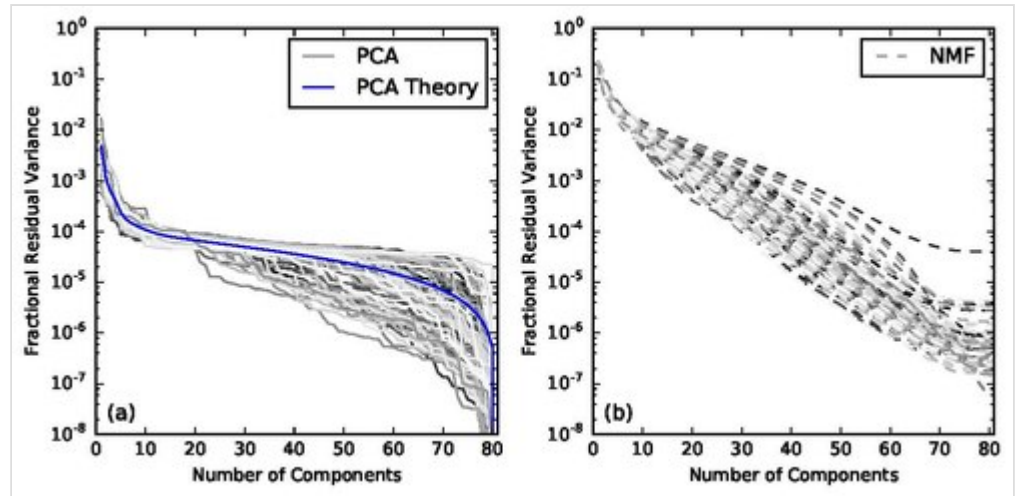
Fractional residual variance (FRV) plots for PCA and sequential NMF;[4] for PCA, the theoretical values are the contribution from the residual eigenvalues. In comparison, the FRV curves for PCA reaches a flat plateau where no signal are captured effectively; while the NMF FRV curves are declining continuously, indicating a better ability to capture signal. The FRV curves for NMF also converges to higher levels than PCA, indicating the less-overfitting property of NMF.

The sequential construction of NMF components (**W** and **H**) was firstly used to relate NMF with Principal Component Analysis (PCA) in astronomy.[37] The contribution from the PCA components are ranked by the magnitude of their corresponding eigenvalues; for NMF, its components can be ranked empirically when they are constructed one by one (sequentially), i.e., learn the $(n+1)$-th component with the first n components constructed.

The contribution of the sequential NMF components can be compared with the Karhunen–Loève theorem, an application of PCA, using the plot of eigenvalues. A typical choice of the number of components with PCA is based on the "elbow" point, then the existence of the flat plateau is indicating that PCA is not capturing the data efficiently, and at last there exists a sudden drop reflecting the capture of random noise

and falls into the regime of overfitting.[38][39] For sequential NMF, the plot of eigenvalues is approximated by the plot of the fractional residual variance curves, where the curves decreases continuously, and converge to a higher level than PCA,[4] which is the indication of less over-fitting of sequential NMF.

Exact NMF

Exact solutions for the variants of NMF can be expected (in polynomial time) when additional constraints hold for matrix $\mathbf{V}$. A polynomial time algorithm for solving nonnegative rank factorization if $\mathbf{V}$ contains a monomial sub matrix of rank equal to its rank was given by Campbell and Poole in 1981.[40] Kalofolias and Gallopoulos (2012)[41] solved the symmetric counterpart of this problem, where $\mathbf{V}$ is symmetric and contains a diagonal principal sub matrix of rank r. Their algorithm runs in $O(rm^2)$ time in the dense case. Arora, Ge, Halpern, Mimno, Moitra, Sontag, Wu, & Zhu (2013) give a polynomial time algorithm for exact NMF that works for the case where one of the factors W satisfies a separability condition.[42]

Relation to other techniques

In *Learning the parts of objects by non-negative matrix factorization* Lee and Seung[43] proposed NMF mainly for parts-based decomposition of images. It compares NMF to vector quantization and principal component analysis, and shows that although the three techniques may be written as factorizations, they implement different constraints and therefore produce different results.
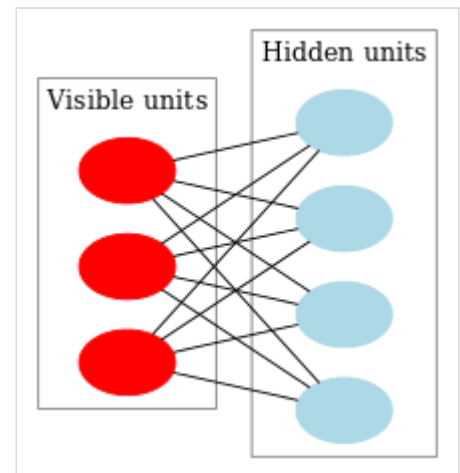
It was later shown that some types of NMF are an instance of a more general probabilistic model called "multinomial PCA".[44] When NMF is obtained by minimizing the Kullback–Leibler divergence, it is in fact equivalent to another instance of multinomial PCA, probabilistic latent semantic analysis,[45] trained by maximum likelihood estimation. That method is commonly used for analyzing and clustering textual data and is also related to the latent class model.

NMF as a probabilistic graphical model: visible units ($\mathbf{V}$) are connected to hidden units ($\mathbf{H}$) through weights $\mathbf{W}$, so that $\mathbf{V}$ is generated from a probability distribution with mean $\sum_a W_{ia} h_a$

[13]:5

NMF with the least-squares objective is equivalent to a relaxed form of K-means clustering: the matrix factor $\mathbf{W}$ contains cluster centroids and $\mathbf{H}$ contains cluster membership indicators.[15][46] This provides a theoretical foundation for using NMF for data clustering. However, k-means does not enforce non-negativity on its centroids, so the closest analogy is in fact with "semi-NMF".[17]

NMF can be seen as a two-layer directed graphical model with one layer of observed random variables and one layer of hidden random variables.[47]

NMF extends beyond matrices to tensors of arbitrary order.[48][49][50] This extension may be viewed as a non-negative counterpart to, e.g., the PARAFAC model.

Other extensions of NMF include joint factorization of several data matrices and tensors where some factors are shared. Such models are useful for sensor fusion and relational learning.[51]

NMF is an instance of nonnegative quadratic programming, just like the support vector machine (SVM). However, SVM and NMF are related at a more intimate level than that of NQP, which allows direct application of the solution algorithms developed for either of the two methods to problems in both domains.[52]

Uniqueness

The factorization is not unique: A matrix and its inverse can be used to transform the two factorization matrices by, e.g.,[53]

$$\mathbf{WH} = \mathbf{WBB^{-1}H}$$

If the two new matrices $\tilde{\mathbf{W}} = \mathbf{WB}$ and $\tilde{\mathbf{H}} = \mathbf{B^{-1}H}$ are non-negative they form another parametrization of the factorization.

The non-negativity of $\tilde{\mathbf{W}}$ and $\tilde{\mathbf{H}}$ applies at least if $\mathbf{B}$ is a non-negative monomial matrix. In this simple case it will just correspond to a scaling and a permutation.

More control over the non-uniqueness of NMF is obtained with sparsity constraints.[54]

Applications

Astronomy

In astronomy, NMF is a promising method for dimension reduction in the sense that astrophysical signals are non-negative. NMF has been applied to the spectroscopic observations[55][3] and the direct imaging observations[4] as a method to study the common properties of astronomical objects and post-process the astronomical observations. The advances in the spectroscopic observations by Blanton & Roweis (2007)[3] takes into account of the uncertainties of astronomical observations, which is later improved by Zhu (2016)[37] where missing data are also considered and parallel computing is enabled. Their method is then adopted by Ren et al. (2018)[4] to the direct imaging field as one of the methods of detecting exoplanets, especially for the direct imaging of circumstellar disks.

Ren et al. (2018)[4] are able to prove the stability of NMF components when they are constructed sequentially (i.e., one by one), which enables the linearity of the NMF modeling process; the linearity property is used to separate the stellar light and the light scattered from the exoplanets and circumstellar disks.

In direct imaging, to reveal the faint exoplanets and circumstellar disks from bright the surrounding stellar lights, which has a typical contrast from 10^5 to 10^{10}, various statistical methods have been adopted,[56][57][38] however the light from the exoplanets or circumstellar disks are usually over-fitted, where forward modeling have to be adopted to recover the true flux.[58][39] Forward modeling is currently optimized for point sources,[39] however not for extended sources, especially for irregularly shaped structures such as circumstellar disks. In this situation, NMF has been an excellent method, being less over-

fitting in the sense of the non-negativity and sparsity of the NMF modeling coefficients, therefore forward modeling can be performed with a few scaling factors,[4] rather than a computationally intensive data re-reduction on generated models.

Data imputation

To impute missing data in statistics, NMF can take missing data while minimizing its cost function, rather than treating these missing data as zeros.[5] This makes it a mathematically proven method for data imputation in statistics.[5] By first proving that the missing data are ignored in the cost function, then proving that the impact from missing data can be as small as a second order effect, Ren et al. (2020)[5] studied and applied such an approach for the field of astronomy. Their work focuses on two-dimensional matrices, specifically, it includes mathematical derivation, simulated data imputation, and application to on-sky data.

The data imputation procedure with NMF can be composed of two steps. First, when the NMF components are known, Ren et al. (2020) proved that impact from missing data during data imputation ("target modeling" in their study) is a second order effect. Second, when the NMF components are unknown, the authors proved that the impact from missing data during component construction is a first-to-second order effect.

Depending on the way that the NMF components are obtained, the former step above can be either independent or dependent from the latter. In addition, the imputation quality can be increased when the more NMF components are used, see Figure 4 of Ren et al. (2020) for their illustration.[5]

Text mining

NMF can be used for text mining applications. In this process, a *document-term* matrix is constructed with the weights of various terms (typically weighted word frequency information) from a set of documents. This matrix is factored into a *term-feature* and a *feature-document* matrix. The features are derived from the contents of the documents, and the feature-document matrix describes data clusters of related documents.

One specific application used hierarchical NMF on a small subset of scientific abstracts from PubMed.[59] Another research group clustered parts of the Enron email dataset[60] with 65,033 messages and 91,133 terms into 50 clusters.[61] NMF has also been applied to citations data, with one example clustering English Wikipedia articles and scientific journals based on the outbound scientific citations in English Wikipedia.[62]

Arora, Ge, Halpern, Mimno, Moitra, Sontag, Wu, & Zhu (2013) have given polynomial-time algorithms to learn topic models using NMF. The algorithm assumes that the topic matrix satisfies a separability condition that is often found to hold in these settings.[42]

Hassani, Iranmanesh and Mansouri (2019) proposed a feature agglomeration method for term-document matrices which operates using NMF. The algorithm reduces the term-document matrix into a smaller matrix more suitable for text clustering.[63]

Spectral data analysis

NMF is also used to analyze spectral data; one such use is in the classification of space objects and debris.[64]

Scalable Internet distance prediction

NMF is applied in scalable Internet distance (round-trip time) prediction. For a network with N hosts, with the help of NMF, the distances of all the N^2 end-to-end links can be predicted after conducting only $O(N)$ measurements. This kind of method was firstly introduced in Internet Distance Estimation Service (IDES).[65] Afterwards, as a fully decentralized approach, Phoenix network coordinate system[66] is proposed. It achieves better overall prediction accuracy by introducing the concept of weight.

Non-stationary speech denoising

Speech denoising has been a long lasting problem in audio signal processing. There are many algorithms for denoising if the noise is stationary. For example, the Wiener filter is suitable for additive Gaussian noise. However, if the noise is non-stationary, the classical denoising algorithms usually have poor performance because the statistical information of the non-stationary noise is difficult to estimate. Schmidt et al.[67] use NMF to do speech denoising under non-stationary noise, which is completely different from classical statistical approaches. The key idea is that clean speech signal can be sparsely represented by a speech dictionary, but non-stationary noise cannot. Similarly, non-stationary noise can also be sparsely represented by a noise dictionary, but speech cannot.

The algorithm for NMF denoising goes as follows. Two dictionaries, one for speech and one for noise, need to be trained offline. Once a noisy speech is given, we first calculate the magnitude of the Short-Time-Fourier-Transform. Second, separate it into two parts via NMF, one can be sparsely represented by the speech dictionary, and the other part can be sparsely represented by the noise dictionary. Third, the part that is represented by the speech dictionary will be the estimated clean speech.

Population genetics

Sparse NMF is used in Population genetics for estimating individual admixture coefficients, detecting genetic clusters of individuals in a population sample or evaluating genetic admixture in sampled genomes. In human genetic clustering, NMF algorithms provide estimates similar to those of the computer program STRUCTURE, but the algorithms are more efficient computationally and allow analysis of large population genomic data sets.[68]

Bioinformatics

NMF has been successfully applied in bioinformatics for clustering gene expression and DNA methylation data and finding the genes most representative of the clusters.[24][69][70][71] In the analysis of cancer mutations it has been used to identify common patterns of mutations that occur in many cancers and that probably have distinct causes.[72] NMF techniques can identify sources of variation such as cell types, disease subtypes, population stratification, tissue composition, and tumor clonality.[73]

A particular variant of NMF, namely Non-Negative Matrix Tri-Factorization (NMTF),[74] has been use for drug repurposing tasks in order to predict novel protein targets and therapeutic indications for approved drugs[75] and to infer pair of synergic anticancer drugs.[76]

Nuclear imaging

NMF, also referred in this field as factor analysis, has been used since the 1980s[77] to analyze sequences of images in SPECT and PET dynamic medical imaging. Non-uniqueness of NMF was addressed using sparsity constraints.[78] [79] [80]

Current research

Current research (since 2010) in nonnegative matrix factorization includes, but is not limited to,

1. Algorithmic: searching for global minima of the factors and factor initialization.[81]
2. Scalability: how to factorize million-by-billion matrices, which are commonplace in Web-scale data mining, e.g., see Distributed Nonnegative Matrix Factorization (DNMF),[82] Scalable Nonnegative Matrix Factorization (ScalableNMF),[83] Distributed Stochastic Singular Value Decomposition.[84]
3. Online: how to update the factorization when new data comes in without recomputing from scratch, e.g., see online CNSC[85]
4. Collective (joint) factorization: factorizing multiple interrelated matrices for multiple-view learning, e.g. multi-view clustering, see CoNMF[86] and MultiNMF[87]
5. Cohen and Rothblum 1993 problem: whether a rational matrix always has an NMF of minimal inner dimension whose factors are also rational. Recently, this problem has been answered negatively.[88]

See also

- Multilinear algebra
- Multilinear subspace learning
- Tensor
- Tensor decomposition
- Tensor software

Sources and external links

Notes

1. Suvrit Sra; Inderjit S. Dhillon (2006). *Generalized Nonnegative Matrix Approximations with Bregman Divergences* (https://papers.nips.cc/paper/2757-generalized-nonnegative-matrix-approximations-with-bregman-divergences.pdf) (PDF). Advances in Neural Information Processing Systems. ISBN 978-0-262-23253-1. Wikidata Q77685465. {{cite book}}: |journal= ignored (help)

2. Tandon, Rashish; Sra, Suvrit (September 13, 2010). Sparse nonnegative matrix approximation: new formulations and algorithms (https://is.tuebingen.mpg.de/fileadmin/user_upload/files/publications/MPIK-TR-193_%5B0%5D.pdf) (PDF) (Report). Max Planck Institute for Biological Cybernetics. Technical Report No. 193.

3. Blanton, Michael R.; Roweis, Sam (2007). "K-corrections and filter transformations in the ultraviolet, optical, and near infrared". *The Astronomical Journal*. **133** (2): 734–754. arXiv:astro-ph/0606170 (https://arxiv.org/abs/astro-ph/0606170). Bibcode:2007AJ....133..734B (https://ui.adsabs.harvard.edu/abs/2007AJ....133..734B). doi:10.1086/510127 (https://doi.org/10.1086%2F510127). S2CID 18561804 (https://api.semanticscholar.org/CorpusID:18561804).

4. Ren, Bin; Pueyo, Laurent; Zhu, Guangtun B.; Duchêne, Gaspard (2018). "Non-negative Matrix Factorization: Robust Extraction of Extended Structures" (https://doi.org/10.3847%2F1538-4357%2Faaa1f2). *The Astrophysical Journal*. **852** (2): 104. arXiv:1712.10317 (https://arxiv.org/abs/1712.10317). Bibcode:2018ApJ...852..104R (https://ui.adsabs.harvard.edu/abs/2018ApJ...852..104R). doi:10.3847/1538-4357/aaa1f2 (https://doi.org/10.3847%2F1538-4357%2Faaa1f2). S2CID 3966513 (https://api.semanticscholar.org/CorpusID:3966513).

5. Ren, Bin; Pueyo, Laurent; Chen, Christine; Choquet, Elodie; Debes, John H; Duechene, Gaspard; Menard, Francois; Perrin, Marshall D. (2020). "Using Data Imputation for Signal Separation in High Contrast Imaging" (https://doi.org/10.3847%2F1538-4357%2Fab7024). *The Astrophysical Journal*. **892** (2): 74. arXiv:2001.00563 (https://arxiv.org/abs/2001.00563). Bibcode:2020ApJ...892...74R (https://ui.adsabs.harvard.edu/abs/2020ApJ...892...74R). doi:10.3847/1538-4357/ab7024 (https://doi.org/10.3847%2F1538-4357%2Fab7024). S2CID 209531731 (https://api.semanticscholar.org/CorpusID:209531731).

6. Rainer Gemulla; Erik Nijkamp; Peter J. Haas; Yannis Sismanis (2011). *Large-scale matrix factorization with distributed stochastic gradient descent*. Proc. ACM SIGKDD Int'l Conf. on Knowledge discovery and data mining. pp. 69–77.

7. Yang Bao; et al. (2014). *TopicMF: Simultaneously Exploiting Ratings and Reviews for Recommendation* (http://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8273). AAAI.

8. Ben Murrell; et al. (2011). "Non-Negative Matrix Factorization for Learning Alignment-Specific Models of Protein Evolution" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245233). *PLOS ONE*. **6** (12): e28898. Bibcode:2011PLoSO...628898M (https://ui.adsabs.harvard.edu/abs/2011PLoSO...628898M). doi:10.1371/journal.pone.0028898 (https://doi.org/10.1371%2Fjournal.pone.0028898). PMC 3245233 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245233). PMID 22216138 (https://pubmed.ncbi.nlm.nih.gov/22216138).

9. William H. Lawton; Edward A. Sylvestre (1971). "Self modeling curve resolution". *Technometrics*. **13** (3): 617–633. doi:10.2307/1267173 (https://doi.org/10.2307%2F1267173). JSTOR 1267173 (https://www.jstor.org/stable/1267173).

10. Pentti Paatero; Unto Tapper; Pasi Aalto; Markku Kulmala (1991). "Matrix factorization methods for analysing diffusion battery data". *Journal of Aerosol Science*. **22**: S273–S276. doi:10.1016/S0021-8502(05)80089-8 (https://doi.org/10.1016%2FS0021-8502%2805%2980089-8). ISSN 0021-8502 (https://www.worldcat.org/issn/0021-8502). Wikidata Q58065673.

11. Pentti Paatero; Unto Tapper (June 1994). "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values" (https://onlinelibrary.wiley.com/doi/abs/10.1002/env.3170050203). *Environmetrics*. **5** (2): 111–126. doi:10.1002/ENV.3170050203 (https://doi.org/10.1002%2FENV.3170050203). ISSN 1180-4009 (https://www.worldcat.org/issn/1180-4009). Wikidata Q29308406.

12. Pia Anttila; Pentti Paatero; Unto Tapper; Olli Järvinen (1995). "Source identification of bulk wet deposition in Finland by positive matrix factorization". *Atmospheric Environment*. **29** (14): 1705–1718. Bibcode:1995AtmEn..29.1705A (https://ui.adsabs.harvard.edu/abs/1995AtmEn..29.1705A). doi:10.1016/1352-2310(94)00367-T (https://doi.org/10.1016%2F1352-2310%2894%2900367-T).

13. Daniel D. Lee & H. Sebastian Seung (1999). "Learning the parts of objects by non-negative matrix factorization". *Nature*. **401** (6755): 788–791. Bibcode:1999Natur.401..788L (https://ui.a dsabs.harvard.edu/abs/1999Natur.401..788L). doi:10.1038/44565 (https://doi.org/10.1038%2 F44565). PMID 10548103 (https://pubmed.ncbi.nlm.nih.gov/10548103). S2CID 4428232 (htt ps://api.semanticscholar.org/CorpusID:4428232).

14. Daniel D. Lee & H. Sebastian Seung (2001). *Algorithms for Non-negative Matrix Factorization* (http://papers.nips.cc/paper/1861-algorithms-for-non-negative-matrix-factorizati on.pdf) (PDF). Advances in Neural Information Processing Systems 13: Proceedings of the 2000 Conference. MIT Press. pp. 556–562.

15. C. Ding, X. He, H.D. Simon (2005). "On the Equivalence of Nonnegative Matrix Factorization and Spectral Clustering" (http://ranger.uta.edu/~chqding/papers/NMF-SDM2005.pdf). Proc. SIAM Int'l Conf. Data Mining, pp. 606-610. May 2005

16. Ding C, Li Y, Peng W (2008). "On the equivalence between non-negative matrix factorization and probabilistic latent semantic indexing" (https://web.archive.org/web/20160304070027/htt p://users.cis.fiu.edu/~taoli/pub/NMFpLSIequiv.pdf) (PDF). *Computational Statistics & Data Analysis*. **52** (8): 3913–3927. doi:10.1016/j.csda.2008.01.011 (https://doi.org/10.1016%2Fj.cs da.2008.01.011). Archived from the original (http://users.cis.fiu.edu/~taoli/pub/NMFpLSIequi v.pdf) (PDF) on 2016-03-04.

17. C Ding, T Li, MI Jordan, Convex and semi-nonnegative matrix factorizations, IEEE Transactions on Pattern Analysis and Machine Intelligence, 32, 45-55, 2010

18. Berman, A.; R.J. Plemmons (1974). "Inverses of nonnegative matrices". *Linear and Multilinear Algebra*. **2** (2): 161–172. doi:10.1080/03081087408817055 (https://doi.org/10.108 0%2F03081087408817055).

19. A. Berman; R.J. Plemmons (1994). *Nonnegative matrices in the Mathematical Sciences*. Philadelphia: SIAM.

20. Thomas, L.B. (1974). "Problem 73-14, Rank factorization of nonnegative matrices". *SIAM Rev*. **16** (3): 393–394. doi:10.1137/1016064 (https://doi.org/10.1137%2F1016064).

21. Vavasis, S.A. (2009). "On the complexity of nonnegative matrix factorization". *SIAM J. Optim*. **20** (3): 1364–1377. arXiv:0708.4149 (https://arxiv.org/abs/0708.4149). doi:10.1137/070709967 (https://doi.org/10.1137%2F070709967). S2CID 7150400 (https://ap i.semanticscholar.org/CorpusID:7150400).

22. Zhang, T.; Fang, B.; Liu, W.; Tang, Y. Y.; He, G.; Wen, J. (2008). "Total variation norm-based nonnegative matrix factorization for identifying discriminant representation of image patterns". *Neurocomputing*. **71** (10–12): 1824–1831. doi:10.1016/j.neucom.2008.01.022 (http s://doi.org/10.1016%2Fj.neucom.2008.01.022).

23. Hoyer, Patrik O. (2002). *Non-negative sparse coding*. Proc. IEEE Workshop on Neural Networks for Signal Processing. arXiv:cs/0202009 (https://arxiv.org/abs/cs/0202009).

24. Leo Taslaman & Björn Nilsson (2012). "A framework for regularized non-negative matrix factorization, with application to the analysis of gene expression data" (https://www.ncbi.nlm. nih.gov/pmc/articles/PMC3487913). *PLOS One*. **7** (11): e46331. Bibcode:2012PLoSO...746331T (https://ui.adsabs.harvard.edu/abs/2012PLoSO...746331T). doi:10.1371/journal.pone.0046331 (https://doi.org/10.1371%2Fjournal.pone.0046331). PMC 3487913 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3487913). PMID 23133590 (h ttps://pubmed.ncbi.nlm.nih.gov/23133590).

25. Hsieh, C. J.; Dhillon, I. S. (2011). *Fast coordinate descent methods with variable selection for non-negative matrix factorization* (http://www.cs.utexas.edu/~cjhsieh/nmf_kdd11.pdf) (PDF). Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '11. p. 1064. doi:10.1145/2020408.2020577 (https://doi.org/10.114 5%2F2020408.2020577). ISBN 9781450308137.

26. Fung, Yik-Hing; Li, Chun-Hung; Cheung, William K. (2 November 2007). *Online Discussion Participation Prediction Using Non-negative Matrix Factorization* (http://dl.acm.org/citation.cfm?id=1339264.1339709). Wi-Iatw '07. IEEE Computer Society. pp. 284–287. ISBN 9780769530284 – via dl.acm.org.

27. Naiyang Guan; Dacheng Tao; Zhigang Luo & Bo Yuan (July 2012). "Online Nonnegative Matrix Factorization With Robust Stochastic Approximation". *IEEE Transactions on Neural Networks and Learning Systems*. **23** (7): 1087–1099. doi:10.1109/TNNLS.2012.2197827 (https://doi.org/10.1109%2FTNNLS.2012.2197827). PMID 24807135 (https://pubmed.ncbi.nlm.nih.gov/24807135). S2CID 8755408 (https://api.semanticscholar.org/CorpusID:8755408).

28. Behnke, S. (2003). "Discovering hierarchical speech features using convolutional non-negative matrix factorization" (https://ieeexplore.ieee.org/document/1224004). *Proceedings of the International Joint Conference on Neural Networks, 2003*. Vol. 4. Portland, Oregon USA: IEEE. pp. 2758–2763. doi:10.1109/IJCNN.2003.1224004 (https://doi.org/10.1109%2FIJCNN.2003.1224004). ISBN 978-0-7803-7898-8. S2CID 3109867 (https://api.semanticscholar.org/CorpusID:3109867).

29. Lin, Chih-Jen (2007). "Projected Gradient Methods for Nonnegative Matrix Factorization" (http://www.csie.ntu.edu.tw/~cjlin/papers/pgradnmf.pdf) (PDF). *Neural Computation*. **19** (10): 2756–2779. CiteSeerX 10.1.1.308.9135 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.308.9135). doi:10.1162/neco.2007.19.10.2756 (https://doi.org/10.1162%2Fneco.2007.19.10.2756). PMID 17716011 (https://pubmed.ncbi.nlm.nih.gov/17716011). S2CID 2295736 (https://api.semanticscholar.org/CorpusID:2295736).

30. Lin, Chih-Jen (2007). "On the Convergence of Multiplicative Update Algorithms for Nonnegative Matrix Factorization". *IEEE Transactions on Neural Networks*. **18** (6): 1589–1596. CiteSeerX 10.1.1.407.318 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.407.318). doi:10.1109/TNN.2007.895831 (https://doi.org/10.1109%2FTNN.2007.895831). S2CID 2183630 (https://api.semanticscholar.org/CorpusID:2183630).

31. Hyunsoo Kim & Haesun Park (2008). "Nonnegative Matrix Factorization Based on Alternating Nonnegativity Constrained Least Squares and Active Set Method" (http://www.cc.gatech.edu/~hpark/papers/simax-nmf.pdf) (PDF). *SIAM Journal on Matrix Analysis and Applications*. **30** (2): 713–730. CiteSeerX 10.1.1.70.3485 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.70.3485). doi:10.1137/07069239x (https://doi.org/10.1137%2F07069239x).

32. Naiyang Guan; Dacheng Tao; Zhigang Luo; Bo Yuan (June 2012). "NeNMF: An Optimal Gradient Method for Nonnegative Matrix Factorization". *IEEE Transactions on Signal Processing*. **60** (6): 2882–2898. Bibcode:2012ITSP...60.2882G (https://ui.adsabs.harvard.edu/abs/2012ITSP...60.2882G). doi:10.1109/TSP.2012.2190406 (https://doi.org/10.1109%2FTSP.2012.2190406). S2CID 8143231 (https://api.semanticscholar.org/CorpusID:8143231).

33. Jingu Kim & Haesun Park (2011). "Fast Nonnegative Matrix Factorization: An Active-set-like Method and Comparisons". *SIAM Journal on Scientific Computing*. **58** (6): 3261–3281. Bibcode:2011SJSC...33.3261K (https://ui.adsabs.harvard.edu/abs/2011SJSC...33.3261K). CiteSeerX 10.1.1.419.798 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.419.798). doi:10.1137/110821172 (https://doi.org/10.1137%2F110821172).

34. Jingu Kim; Yunlong He & Haesun Park (2013). "Algorithms for nonnegative matrix and tensor factorizations: A unified view based on block coordinate descent framework" (https://smallk.github.io/papers/nmf_review_jgo.pdf) (PDF). *Journal of Global Optimization*. **33** (2): 285–319. doi:10.1007/s10898-013-0035-4 (https://doi.org/10.1007%2Fs10898-013-0035-4). S2CID 11197117 (https://api.semanticscholar.org/CorpusID:11197117).

35. Ding, C.; He, X. & Simon, H.D. (2005). "On the equivalence of nonnegative matrix factorization and spectral clustering". *Proc. SIAM Data Mining Conf*. Vol. 4. pp. 606–610. doi:10.1137/1.9781611972757.70 (https://doi.org/10.1137%2F1.9781611972757.70). ISBN 978-0-89871-593-4.

36. Hafshejani, Sajad Fathi; Moaberfard, Zahra (November 2022). "Initialization for Nonnegative Matrix Factorization: a Comprehensive Review" (http://arxiv.org/abs/2109.03874). *International Journal of Data Science and Analytics*. **16** (1): 119–134. arXiv:2109.03874 (https://arxiv.org/abs/2109.03874). doi:10.1007/s41060-022-00370-9 (https://doi.org/10.1007%2Fs41060-022-00370-9). ISSN 2364-415X (https://www.worldcat.org/issn/2364-415X).

37. Zhu, Guangtun B. (2016-12-19). "Nonnegative Matrix Factorization (NMF) with Heteroscedastic Uncertainties and Missing data". arXiv:1612.06037 (https://arxiv.org/abs/1612.06037) [astro-ph.IM (https://arxiv.org/archive/astro-ph.IM)].

38. Soummer, Rémi; Pueyo, Laurent; Larkin, James (2012). "Detection and Characterization of Exoplanets and Disks Using Projections on Karhunen-Loève Eigenimages". *The Astrophysical Journal Letters*. **755** (2): L28. arXiv:1207.4197 (https://arxiv.org/abs/1207.4197). Bibcode:2012ApJ...755L..28S (https://ui.adsabs.harvard.edu/abs/2012ApJ...755L..28S). doi:10.1088/2041-8205/755/2/L28 (https://doi.org/10.1088%2F2041-8205%2F755%2F2%2FL28). S2CID 51088743 (https://api.semanticscholar.org/CorpusID:51088743).

39. Pueyo, Laurent (2016). "Detection and Characterization of Exoplanets using Projections on Karhunen Loeve Eigenimages: Forward Modeling" (https://doi.org/10.3847%2F0004-637X%2F824%2F2%2F117). *The Astrophysical Journal*. **824** (2): 117. arXiv:1604.06097 (https://arxiv.org/abs/1604.06097). Bibcode:2016ApJ...824..117P (https://ui.adsabs.harvard.edu/abs/2016ApJ...824..117P). doi:10.3847/0004-637X/824/2/117 (https://doi.org/10.3847%2F0004-637X%2F824%2F2%2F117). S2CID 118349503 (https://api.semanticscholar.org/CorpusID:118349503).

40. Campbell, S.L.; G.D. Poole (1981). "Computing nonnegative rank factorizations" (https://doi.org/10.1016%2F0024-3795%2881%2990272-x). *Linear Algebra Appl*. **35**: 175–182. doi:10.1016/0024-3795(81)90272-x (https://doi.org/10.1016%2F0024-3795%2881%2990272-x).

41. Kalofolias, V.; Gallopoulos, E. (2012). "Computing symmetric nonnegative rank factorizations" (https://infoscience.epfl.ch/record/198764/files/main.pdf) (PDF). *Linear Algebra Appl*. **436** (2): 421–435. doi:10.1016/j.laa.2011.03.016 (https://doi.org/10.1016%2Fj.laa.2011.03.016).

42. Arora, Sanjeev; Ge, Rong; Halpern, Yoni; Mimno, David; Moitra, Ankur; Sontag, David; Wu, Yichen; Zhu, Michael (2013). *A practical algorithm for topic modeling with provable guarantees* (http://jmlr.csail.mit.edu/proceedings/papers/v28/arora13.html). Proceedings of the 30th International Conference on Machine Learning. arXiv:1212.4777 (https://arxiv.org/abs/1212.4777). Bibcode:2012arXiv1212.4777A (https://ui.adsabs.harvard.edu/abs/2012arXiv1212.4777A).

43. Lee, Daniel D.; Sebastian, Seung, H. (1999). "Learning the parts of objects by non-negative matrix factorization" (http://www.columbia.edu/~jwp2128/Teaching/E4903/papers/nmf_nature.pdf) (PDF). *Nature*. **401** (6755): 788–791. Bibcode:1999Natur.401..788L (https://ui.adsabs.harvard.edu/abs/1999Natur.401..788L). doi:10.1038/44565 (https://doi.org/10.1038%2F44565). PMID 10548103 (https://pubmed.ncbi.nlm.nih.gov/10548103). S2CID 4428232 (https://api.semanticscholar.org/CorpusID:4428232).

44. Wray Buntine (2002). *Variational Extensions to EM and Multinomial PCA* (http://cosco.hiit.fi/Articles/ecml02.pdf) (PDF). Proc. European Conference on Machine Learning (ECML-02). LNAI. Vol. 2430. pp. 23–34.

45. Eric Gaussier & Cyril Goutte (2005). *Relation between PLSA and NMF and Implications* (https://web.archive.org/web/20070928032454/http://eprints.pascal-network.org/archive/00000971/01/39-gaussier.pdf) (PDF). Proc. 28th international ACM SIGIR conference on Research and development in information retrieval (SIGIR-05). pp. 601–602. Archived from the original (http://eprints.pascal-network.org/archive/00000971/01/39-gaussier.pdf) (PDF) on 2007-09-28. Retrieved 2007-01-29.

46. Ron Zass and Amnon Shashua (2005). "A Unifying Approach to Hard and Probabilistic Clustering (http://www.cs.huji.ac.il/~zass/papers/cp-iccv05.pdf)". International Conference on Computer Vision (ICCV) Beijing, China, Oct., 2005.

47. Max Welling; et al. (2004). *Exponential Family Harmoniums with an Application to Information Retrieval* (http://papers.nips.cc/paper/2672-exponential-family-harmoniums-with-an-application-to-information-retrieval). NIPS.

48. Pentti Paatero (1999). "The Multilinear Engine: A Table-Driven, Least Squares Program for Solving Multilinear Problems, including the n-Way Parallel Factor Analysis Model". *Journal of Computational and Graphical Statistics*. **8** (4): 854–888. doi:10.2307/1390831 (https://doi.org/10.2307%2F1390831). JSTOR 1390831 (https://www.jstor.org/stable/1390831).

49. Max Welling & Markus Weber (2001). "Positive Tensor Factorization". *Pattern Recognition Letters*. **22** (12): 1255–1261. Bibcode:2001PaReL..22.1255W (https://ui.adsabs.harvard.edu/abs/2001PaReL..22.1255W). CiteSeerX 10.1.1.21.24 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.21.24). doi:10.1016/S0167-8655(01)00070-8 (https://doi.org/10.1016%2FS0167-8655%2801%2900070-8).

50. Jingu Kim & Haesun Park (2012). *Fast Nonnegative Tensor Factorization with an Active-set-like Method* (http://www.cc.gatech.edu/~hpark/papers/2011_paper_hpscbook_ntf.pdf) (PDF). High-Performance Scientific Computing: Algorithms and Applications. Springer. pp. 311–326.

51. Kenan Yilmaz; A. Taylan Cemgil & Umut Simsekli (2011). *Generalized Coupled Tensor Factorization* (http://books.nips.cc/papers/files/nips24/NIPS2011_1189.pdf) (PDF). NIPS.

52. Vamsi K. Potluru; Sergey M. Plis; Morten Morup; Vince D. Calhoun & Terran Lane (2009). *Efficient Multiplicative updates for Support Vector Machines*. Proceedings of the 2009 SIAM Conference on Data Mining (SDM). pp. 1218–1229.

53. Wei Xu; Xin Liu & Yihong Gong (2003). *Document clustering based on non-negative matrix factorization* (http://portal.acm.org/citation.cfm?id=860485). Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval. New York: Association for Computing Machinery. pp. 267–273.

54. Eggert, J.; Korner, E. (2004). "Sparse coding and NMF". *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541)*. Vol. 4. pp. 2529–2533. doi:10.1109/IJCNN.2004.1381036 (https://doi.org/10.1109%2FIJCNN.2004.1381036). ISBN 978-0-7803-8359-3. S2CID 17923083 (https://api.semanticscholar.org/CorpusID:17923083).

55. Berné, O.; Joblin, C.; Deville, Y.; Smith, J. D.; Rapacioli, M.; Bernard, J. P.; Thomas, J.; Reach, W.; Abergel, A. (2007-07-01). "Analysis of the emission of very small dust particles from Spitzer spectro-imagery data using blind signal separation methods" (https://www.aanda.org/articles/aa/abs/2007/26/aa6282-06/aa6282-06.html). *Astronomy & Astrophysics*. **469** (2): 575–586. arXiv:astro-ph/0703072 (https://arxiv.org/abs/astro-ph/0703072). Bibcode:2007A&A...469..575B (https://ui.adsabs.harvard.edu/abs/2007A&A...469..575B). doi:10.1051/0004-6361:20066282 (https://doi.org/10.1051%2F0004-6361%3A20066282). ISSN 0004-6361 (https://www.worldcat.org/issn/0004-6361).

56. Lafrenière, David; Maroid, Christian; Doyon, René; Barman, Travis (2009). "HST/NICMOS Detection of HR 8799 b in 1998". *The Astrophysical Journal Letters*. **694** (2): L148. arXiv:0902.3247 (https://arxiv.org/abs/0902.3247). Bibcode:2009ApJ...694L.148L (https://ui.adsabs.harvard.edu/abs/2009ApJ...694L.148L). doi:10.1088/0004-637X/694/2/L148 (https://doi.org/10.1088%2F0004-637X%2F694%2F2%2FL148). S2CID 7332750 (https://api.semanticscholar.org/CorpusID:7332750).

57. Amara, Adam; Quanz, Sascha P. (2012). "PYNPOINT: an image processing package for finding exoplanets". *Monthly Notices of the Royal Astronomical Society*. **427** (2): 948. arXiv:1207.6637 (https://arxiv.org/abs/1207.6637). Bibcode:2012MNRAS.427..948A (https://ui.adsabs.harvard.edu/abs/2012MNRAS.427..948A). doi:10.1111/j.1365-2966.2012.21918.x (https://doi.org/10.1111%2Fj.1365-2966.2012.21918.x). S2CID 119200505 (https://api.semanticscholar.org/CorpusID:119200505).

58. Wahhaj, Zahed; Cieza, Lucas A.; Mawet, Dimitri; Yang, Bin; Canovas, Hector; de Boer, Jozua; Casassus, Simon; Ménard, François; Schreiber, Matthias R.; Liu, Michael C.; Biller, Beth A.; Nielsen, Eric L.; Hayward, Thomas L. (2015). "Improving signal-to-noise in the direct imaging of exoplanets and circumstellar disks with MLOCI". *Astronomy & Astrophysics*. **581** (24): A24. arXiv:1502.03092 (https://arxiv.org/abs/1502.03092). Bibcode:2015A&A...581A..24W (https://ui.adsabs.harvard.edu/abs/2015A&A...581A..24W). doi:10.1051/0004-6361/201525837 (https://doi.org/10.1051%2F0004-6361%2F201525837). S2CID 20174209 (https://api.semanticscholar.org/CorpusID:20174209).

59. Nielsen, Finn Årup; Balslev, Daniela; Hansen, Lars Kai (2005). "Mining the posterior cingulate: segregation between memory and pain components" (http://orbit.dtu.dk/ws/files/3936747/imm3661.pdf) (PDF). *NeuroImage*. **27** (3): 520–522. doi:10.1016/j.neuroimage.2005.04.034 (https://doi.org/10.1016%2Fj.neuroimage.2005.04.034). PMID 15946864 (https://pubmed.ncbi.nlm.nih.gov/15946864). S2CID 18509039 (https://api.semanticscholar.org/CorpusID:18509039).

60. Cohen, William (2005-04-04). "Enron Email Dataset" (https://www.cs.cmu.edu/~enron/). Retrieved 2008-08-26.

61. Berry, Michael W.; Browne, Murray (2005). "Email Surveillance Using Non-negative Matrix Factorization". *Computational and Mathematical Organization Theory*. **11** (3): 249–264. doi:10.1007/s10588-005-5380-5 (https://doi.org/10.1007%2Fs10588-005-5380-5). S2CID 16249147 (https://api.semanticscholar.org/CorpusID:16249147).

62. Nielsen, Finn Årup (2008). *Clustering of scientific citations in Wikipedia* (http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=5666). Wikimania. arXiv:0805.1154 (https://arxiv.org/abs/0805.1154).

63. Hassani, Ali; Iranmanesh, Amir; Mansouri, Najme (2019-11-12). "Text Mining using Nonnegative Matrix Factorization and Latent Semantic Analysis". arXiv:1911.04705 (https://arxiv.org/abs/1911.04705) [cs.LG (https://arxiv.org/archive/cs.LG)].

64. Berry, Michael W.; Browne, Murray; Langville, Amy N.; Paucac, V. Paul; Plemmonsc, Robert J. (15 September 2007). "Algorithms and Applications for Approximate Nonnegative Matrix Factorization". *Computational Statistics & Data Analysis*. **52** (1): 155–173. doi:10.1016/j.csda.2006.11.006 (https://doi.org/10.1016%2Fj.csda.2006.11.006).

65. Yun Mao; Lawrence Saul & Jonathan M. Smith (2006). "IDES: An Internet Distance Estimation Service for Large Networks". *IEEE Journal on Selected Areas in Communications*. **24** (12): 2273–2284. CiteSeerX 10.1.1.136.3837 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.136.3837). doi:10.1109/JSAC.2006.884026 (https://doi.org/10.1109%2FJSAC.2006.884026). S2CID 12931155 (https://api.semanticscholar.org/CorpusID:12931155).

66. Yang Chen; Xiao Wang; Cong Shi; et al. (2011). "Phoenix: A Weight-based Network Coordinate System Using Matrix Factorization" (https://web.archive.org/web/20111114191220/http://www.cs.duke.edu/~ychen/Phoenix_TNSM.pdf) (PDF). *IEEE Transactions on Network and Service Management*. **8** (4): 334–347. CiteSeerX 10.1.1.300.2851 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.300.2851). doi:10.1109/tnsm.2011.110911.100079 (https://doi.org/10.1109%2Ftnsm.2011.110911.100079). S2CID 8079061 (https://api.semanticscholar.org/CorpusID:8079061). Archived from the original (http://www.cs.duke.edu/~ychen/Phoenix_TNSM.pdf) (PDF) on 2011-11-14.

67. Schmidt, M.N., J. Larsen, and F.T. Hsiao. (2007). "Wind noise reduction using non-negative sparse coding (http://orbit.dtu.dk/files/3848474/Schmidt.pdf)", *Machine Learning for Signal Processing, IEEE Workshop on*, 431–436

68. Frichot E, Mathieu F, Trouillon T, Bouchard G, Francois O (2014). "Fast and efficient estimation of individual ancestry coefficients" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3982712). *Genetics*. **196** (4): 973–983. doi:10.1534/genetics.113.160572 (https://doi.org/10.1534%2Fgenetics.113.160572). PMC 3982712 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3982712). PMID 24496008 (https://pubmed.ncbi.nlm.nih.gov/24496008).

69. Devarajan, K. (2008). "Nonnegative Matrix Factorization: An Analytical and Interpretive Tool in Computational Biology" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2447881). *PLOS Computational Biology*. **4** (7): e1000029. Bibcode:2008PLSCB...4E0029D (https://ui.adsabs.harvard.edu/abs/2008PLSCB...4E0029D). doi:10.1371/journal.pcbi.1000029 (https://doi.org/10.1371%2Fjournal.pcbi.1000029). PMC 2447881 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2447881). PMID 18654623 (https://pubmed.ncbi.nlm.nih.gov/18654623).

70. Hyunsoo Kim & Haesun Park (2007). "Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis" (https://doi.org/10.1093%2Fbioinformatics%2Fbtm134). *Bioinformatics*. **23** (12): 1495–1502. doi:10.1093/bioinformatics/btm134 (https://doi.org/10.1093%2Fbioinformatics%2Fbtm134). PMID 17483501 (https://pubmed.ncbi.nlm.nih.gov/17483501).

71. Schwalbe, E. (2013). "DNA methylation profiling of medulloblastoma allows robust sub-classification and improved outcome prediction using formalin-fixed biopsies" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4313078). *Acta Neuropathologica*. **125** (3): 359–371. doi:10.1007/s00401-012-1077-2 (https://doi.org/10.1007%2Fs00401-012-1077-2). PMC 4313078 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4313078). PMID 23291781 (https://pubmed.ncbi.nlm.nih.gov/23291781).

72. Alexandrov, Ludmil B.; Nik-Zainal, Serena; Wedge, David C.; Campbell, Peter J.; Stratton, Michael R. (2013-01-31). "Deciphering signatures of mutational processes operative in human cancer" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3588146). *Cell Reports*. **3** (1): 246–259. doi:10.1016/j.celrep.2012.12.008 (https://doi.org/10.1016%2Fj.celrep.2012.12.008). ISSN 2211-1247 (https://www.worldcat.org/issn/2211-1247). PMC 3588146 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3588146). PMID 23318258 (https://pubmed.ncbi.nlm.nih.gov/23318258).

73. Stein-O'Brien, Genevieve L.; Arora, Raman; Culhane, Aedin C.; Favorov, Alexander V.; Garmire, Lana X.; Greene, Casey S.; Goff, Loyal A.; Li, Yifeng; Ngom, Aloune; Ochs, Michael F.; Xu, Yanxun (2018-10-01). "Enter the Matrix: Factorization Uncovers Knowledge from Omics" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6309559). *Trends in Genetics*. **34** (10): 790–805. doi:10.1016/j.tig.2018.07.003 (https://doi.org/10.1016%2Fj.tig.2018.07.003). ISSN 0168-9525 (https://www.worldcat.org/issn/0168-9525). PMC 6309559 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6309559). PMID 30143323 (https://pubmed.ncbi.nlm.nih.gov/30143323).

74. Ding; Li; Peng; Park (2006). "Orthogonal nonnegative matrix t-factorizations for clustering". *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. pp. 126–135. doi:10.1145/1150402.1150420 (https://doi.org/10.1145%2F1150402.1150420). ISBN 1595933395. S2CID 165018 (https://api.semanticscholar.org/CorpusID:165018).

75. Ceddia; Pinoli; Ceri; Masseroli (2020). "Matrix factorization-based technique for drug repurposing predictions". *IEEE Journal of Biomedical and Health Informatics*. **24** (11): 3162–3172. doi:10.1109/JBHI.2020.2991763 (https://doi.org/10.1109%2FJBHI.2020.2991763). PMID 32365039 (https://pubmed.ncbi.nlm.nih.gov/32365039). S2CID 218504587 (https://api.semanticscholar.org/CorpusID:218504587).

76. Pinoli; Ceddia; Ceri; Masseroli (2021). "Predicting drug synergism by means of non-negative matrix tri-factorization". *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. **PP** (4): 1956–1967. doi:10.1109/TCBB.2021.3091814 (https://doi.org/10.1109%2FTCBB.2021.3091814). PMID 34166199 (https://pubmed.ncbi.nlm.nih.gov/34166199). S2CID 235634059 (https://api.semanticscholar.org/CorpusID:235634059).

77. DiPaola; Bazin; Aubry; Aurengo; Cavailloles; Herry; Kahn (1982). "Handling of dynamic sequences in nuclear medicine". *IEEE Trans Nucl Sci*. **29** (4): 1310–21. Bibcode:1982ITNS...29.1310D (https://ui.adsabs.harvard.edu/abs/1982ITNS...29.1310D). doi:10.1109/tns.1982.4332188 (https://doi.org/10.1109%2Ftns.1982.4332188). S2CID 37186516 (https://api.semanticscholar.org/CorpusID:37186516).

78. Sitek; Gullberg; Huesman (2002). "Correction for ambiguous solutions in factor analysis using a penalized least squares objective". *IEEE Trans Med Imaging*. **21** (3): 216–25. doi:10.1109/42.996340 (https://doi.org/10.1109%2F42.996340). PMID 11989846 (https://pubmed.ncbi.nlm.nih.gov/11989846). S2CID 6553527 (https://api.semanticscholar.org/CorpusID:6553527).

79. Boutchko; Mitra; Baker; Jagust; Gullberg (2015). "Clustering Initiated Factor Analysis (CIFA) Application for Tissue Classification in Dynamic Brain PET" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4640278). *Journal of Cerebral Blood Flow and Metabolism*. **35** (7): 1104–11. doi:10.1038/jcbfm.2015.69 (https://doi.org/10.1038%2Fjcbfm.2015.69). PMC 4640278 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4640278). PMID 25899294 (https://pubmed.ncbi.nlm.nih.gov/25899294).

80. Abdalah; Boutchko; Mitra; Gullberg (2015). "Reconstruction of 4-D Dynamic SPECT Images From Inconsistent Projections Using a Spline Initialized FADS Algorithm (SIFADS)" (https://escholarship.org/uc/item/0b95c190). *IEEE Trans Med Imaging*. **34** (1): 216–18. doi:10.1109/TMI.2014.2352033 (https://doi.org/10.1109%2FTMI.2014.2352033). PMID 25167546 (https://pubmed.ncbi.nlm.nih.gov/25167546). S2CID 11060831 (https://api.semanticscholar.org/CorpusID:11060831).

81. C. Boutsidis & E. Gallopoulos (2008). "SVD based initialization: A head start for nonnegative matrix factorization". *Pattern Recognition*. **41** (4): 1350–1362. Bibcode:2008PatRe..41.1350B (https://ui.adsabs.harvard.edu/abs/2008PatRe..41.1350B). CiteSeerX 10.1.1.137.8281 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.137.8281). doi:10.1016/j.patcog.2007.09.010 (https://doi.org/10.1016%2Fj.patcog.2007.09.010).

82. Chao Liu; Hung-chih Yang; Jinliang Fan; Li-Wei He & Yi-Min Wang (2010). "Distributed Nonnegative Matrix Factorization for Web-Scale Dyadic Data Analysis on MapReduce" (http://research.microsoft.com/pubs/119077/DNMF.pdf) (PDF). *Proceedings of the 19th International World Wide Web Conference*.

83. Jiangtao Yin; Lixin Gao & Zhongfei (Mark) Zhang (2014). "Scalable Nonnegative Matrix Factorization with Block-wise Updates" (http://rio.ecs.umass.edu/mnilpub/papers/ecmlpkdd2014-yin.pdf) (PDF). *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*.

84. "Apache Mahout" (https://mahout.apache.org/). *mahout.apache.org*. Retrieved 2019-12-14.

85. Dong Wang; Ravichander Vipperla; Nick Evans; Thomas Fang Zheng (2013). "Online Non-Negative Convolutive Pattern Learning for Speech Signals" (https://web.archive.org/web/20150419072552/http://cslt.riit.tsinghua.edu.cn:8081/homepages/wangd/public/pdf/cnsc-tsp.pdf) (PDF). *IEEE Transactions on Signal Processing*. **61** (1): 44–56. Bibcode:2013ITSP...61...44W (https://ui.adsabs.harvard.edu/abs/2013ITSP...61...44W). CiteSeerX 10.1.1.707.7348 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.707.7348). doi:10.1109/tsp.2012.2222381 (https://doi.org/10.1109%2Ftsp.2012.2222381). S2CID 12530378 (https://api.semanticscholar.org/CorpusID:12530378). Archived from the original (http://cslt.riit.tsinghua.edu.cn:8081/homepages/wangd/public/pdf/cnsc-tsp.pdf) (PDF) on 2015-04-19. Retrieved 2015-04-19.

86. Xiangnan He; Min-Yen Kan; Peichu Xie & Xiao Chen (2014). "Comment-based Multi-View Clustering of Web 2.0 Items" (https://web.archive.org/web/20150402103346/http://www.comp.nus.edu.sg/~xiangnan/files/www2014-he.pdf) (PDF). *Proceedings of the 23rd International World Wide Web Conference*. Archived from the original (http://www.comp.nus.edu.sg/~xiangnan/files/www2014-he.pdf) (PDF) on 2015-04-02. Retrieved 2015-03-22.

87. Jialu Liu; Chi Wang; Jing Gao & Jiawei Han (2013). "Multi-View Clustering via Joint Nonnegative Matrix Factorization". *Proceedings of the 2013 SIAM International Conference on Data Mining* (http://jialu.cs.illinois.edu/paper/sdm2013-liu.pdf) (PDF). pp. 252–260. CiteSeerX 10.1.1.301.1771 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.301.1771). doi:10.1137/1.9781611972832.28 (https://doi.org/10.1137%2F1.9781611972832.28). ISBN 978-1-61197-262-7. S2CID 4968 (https://api.semanticscholar.org/CorpusID:4968). `{{cite book}}`: |journal= ignored (help)

88. Chistikov, Dmitry; Kiefer, Stefan; Marušić, Ines; Shirmohammadi, Mahsa; Worrell, James (2016-05-22). "Nonnegative Matrix Factorization Requires Irrationality". arXiv:1605.06848 (https://arxiv.org/abs/1605.06848) [cs.CC (https://arxiv.org/archive/cs.CC)].

Others

- J. Shen; G. W. Israël (1989). "A receptor model using a specific non-negative transformation technique for ambient aerosol" (https://doi.org/10.1016%2F0004-6981%2889%2990190-X). *Atmospheric Environment*. **23** (10): 2289–2298. Bibcode:1989AtmEn..23.2289S (https://ui.adsabs.harvard.edu/abs/1989AtmEn..23.2289S). doi:10.1016/0004-6981(89)90190-X (https://doi.org/10.1016%2F0004-6981%2889%2990190-X).

- Pentti Paatero (1997). "Least squares formulation of robust non-negative factor analysis". *Chemometrics and Intelligent Laboratory Systems*. **37** (1): 23–35. doi:10.1016/S0169-7439(96)00044-5 (https://doi.org/10.1016%2FS0169-7439%2896%2900044-5).

- Raul Kompass (2007). "A Generalized Divergence Measure for Nonnegative Matrix Factorization". *Neural Computation*. **19** (3): 780–791. doi:10.1162/neco.2007.19.3.780 (https://doi.org/10.1162%2Fneco.2007.19.3.780). PMID 17298233 (https://pubmed.ncbi.nlm.nih.gov/17298233). S2CID 5337451 (https://api.semanticscholar.org/CorpusID:5337451).

- Liu, W.X.; Zheng, N.N. & You, Q.B. (2006). "Nonnegative Matrix Factorization and its applications in pattern recognition". *Chinese Science Bulletin*. **51** (17–18): 7–18. Bibcode:2006ChSBu..51....7L (https://ui.adsabs.harvard.edu/abs/2006ChSBu..51....7L). doi:10.1007/s11434-005-1109-6 (https://doi.org/10.1007%2Fs11434-005-1109-6). S2CID 15445516 (https://api.semanticscholar.org/CorpusID:15445516).

- Ngoc-Diep Ho; Paul Van Dooren & Vincent Blondel (2008). "Descent Methods for Nonnegative Matrix Factorization". arXiv:0801.3199 (https://arxiv.org/abs/0801.3199) [cs.NA (https://arxiv.org/archive/cs.NA)].

- Andrzej Cichocki; Rafal Zdunek & Shun-ichi Amari (2008). "Nonnegative Matrix and Tensor Factorization". *IEEE Signal Processing Magazine*. **25** (1): 142–145. Bibcode:2008ISPM...25R.142C (https://ui.adsabs.harvard.edu/abs/2008ISPM...25R.142C). doi:10.1109/MSP.2008.4408452 (https://doi.org/10.1109%2FMSP.2008.4408452). S2CID 9997603 (https://api.semanticscholar.org/CorpusID:9997603).

- Cédric Févotte; Nancy Bertin & Jean-Louis Durrieu (2009). "Nonnegative Matrix Factorization with the Itakura-Saito Divergence: With Application to Music Analysis". *Neural Computation*. **21** (3): 793–830. doi:10.1162/neco.2008.04-08-771 (https://doi.org/10.1162%2Fneco.2008.04-08-771). PMID 18785855 (https://pubmed.ncbi.nlm.nih.gov/18785855). S2CID 13208611 (https://api.semanticscholar.org/CorpusID:13208611).

- Ali Taylan Cemgil (2009). "Bayesian Inference for Nonnegative Matrix Factorisation Models" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2688815). *Computational Intelligence and*

Neuroscience. **2009** (2): 1–17. doi:10.1155/2009/785152 (https://doi.org/10.1155%2F2009%2F785152). PMC 2688815 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2688815). PMID 19536273 (https://pubmed.ncbi.nlm.nih.gov/19536273).

- Andrzej Cichocki, Morten Mrup, et al.: "Advances in Nonnegative Matrix and Tensor Factorization", Hindawi Publishing Corporation, ISBN 978-9774540455 (2008).
- Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan and Shun-ichi Amari: "Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation", Wiley, ISBN 978-0470746660 (2009).
- Andri Mirzal: "Nonnegative Matrix Factorizations for Clustering and LSI: Theory and Programming", LAP LAMBERT Academic Publishing, ISBN 978-3844324891 (2011).
- Yong Xiang: "Blind Source Separation: Dependent Component Analysis", Springer, ISBN 978-9812872265 (2014).
- Ganesh R. Naik(Ed.): "Non-negative Matrix Factorization Techniques: Advances in Theory and Applications", Springer, ISBN 978-3662517000 (2016).
- Julian Becker: "Nonnegative Matrix Factorization with Adaptive Elements for Monaural Audio Source Separation: 1 ", Shaker Verlag GmbH, Germany, ISBN 978-3844048148 (2016).
- Jen-Tzung Chien: "Source Separation and Machine Learning", Academic Press, ISBN 978-0128177969 (2018).
- Shoji Makino(Ed.): "Audio Source Separation", Springer, ISBN 978-3030103033 (2019).
- Nicolas Gillis: "Nonnegative Matrix Factorization", SIAM, ISBN 978-1-611976-40-3 (2020).