Mathematics and Statistics 9(3): 285-289, 2021
DOI: 10.13189/ms.2021.090309

http://www.hrpub.org

Reducing Approximation Error with Rapid Convergence Rate for Non-Negative Matrix Factorization (NMF)

Jayanta Biswas[1,*], **Pritam Kayal**[2], **Debabrata Samanta**[1]

[1]Department of Computer Science, Christ University, Bangalore, India

[2]Research Assistant, Don Bosco School, Park Circus, Kolkata, India

Received February 20, 2021; Revised April 5, 2021; Accepted April 25, 2021

Cite This Paper in the following Citation Styles

***(a)**: [1] Jayanta Biswas, Pritam Kayal, Debabrata Samanta, "Reducing Approximation Error with Rapid Convergence Rate for Non-Negative Matrix Factorization (NMF)," Mathematics and Statistics, Vol.9, No.3, pp. 285-289, 2021. DOI: 10.13189/ms.2021.090309.*

***(b)**: Jayanta Biswas, Pritam Kayal, Debabrata Samanta, (2021). Reducing Approximation Error with Rapid Convergence Rate for Non-Negative Matrix Factorization (NMF). Mathematics and Statistics, 9(3), 285-289. DOI: 10.13189/ms.2021.090309.*

Abstract Non-Negative Matrix Factorization (NMF) is utilized in many important applications. This paper presents development of an efficient low rank approximate NMF algorithm for feature extraction related to text mining and spectral data analysis. NMF can be used for clustering. NMF factorizes a positive matrix A to two positive matrices W and H matrices where $A = WH$. The proposal uses k-means clustering algorithm to determine the centroid of each cluster and assigns the centroid coordinates of each cluster as one column for W matrix. The initial choice of W matrix is positive. The H matrix is determined with gradient descent algorithm based on thin QR optimization. The performance comparison of the proposed NMF algorithm is illustrated with results. The accurate choice of initial positive W matrix reduces approximation error and the use of thin QR algorithm in combination with gradient descent approach provides rapid convergence rate for NMF. The proposed algorithm is implemented with the randomly generated matrix in MATLAB environment. The number of significant singular values of the generated matrix is selected as the number of clusters. The error and convergence rate comparison of the proposed algorithm with the current algorithms are demonstrated in this research. The accurate measurement of execution time for individual program is not possible in MATLAB. The average time execution over 200 iterations is therefore calculated with an increasing iteration count of the proposed algorithm and the comparative results are presented.

Keywords Non-Negative Matrix Factorization (NMF); Singular Value Decomposition (SVD), Clustering, Document Retrieval

1 Introduction

Clustering with heterogeneous data set [1] while retaining the original scale is a very important aspect of analyzing heterogeneous data set. The analysis of NMF is described in this context in the survey paper [2] and in the book chapter reported in [3]. NMF factorises positive matrix $A \equiv WH$ where W and H are also positive matrices. NMF forces W and H to be positive and hence allows only additive combinations. NMF basis images are localised features of the original image [4]. The basis images for vector quantization (VQ) [5] and principal component analysis (PCA) [6] are distorted representation of the complete image. A non-negative factorization can be used for clustering: the data vector a_j is assigned to cluster i, if h_{ij} is the largest element in column j of H. Each column of A represents a point in m dimension space. Where $A \in R^{m \times n}, W \in R^{m \times k}, H \in R^{k \times n}$. The constraint is given by,

$$\min_{W \geq 0, H \geq 0} ||A - WH||_F \tag{1}$$

which is equivalent to

$$\min_{h_j \geq 0} ||a_j - W^{(k)} h_j||_2 \tag{2}$$

where j = 1, 2....... n

The authors in [4] uses Singular Value Decomposition (SVD) to detect the most significant input basis vector for which all elements are positive. However, approximate methods are used to suppress the negative values to generate the remaining basis vectors. Negative values are replaced with zeros and SVD is used repetitively to generate all positive basis vectors which is written as W. Each basis vector represents a column in W. The initial choice of W determined by SVD algorithm is not very close to the actual solution as the negative values are replaced with zeroes while computing the initial basis vectors. Multiplicative algorithm is used in conjunction with gradient descent approach to determine H. Alternating least square in [7], and multiplicative algorithm in [4] are the best reported solutions for the problem.

NMF is a cutting-edge feature extraction technique. NMF comes in handy when there are a lot of attributes and they're unclear or unpredictable. NMF may create meaningful patterns, subjects, or themes by combining attributes [8][9]. In text mining, NMF is often used. The same word appears in multiple places in a text document, each with a different meaning [10].

NMF decomposes multivariate data by producing a number of features that the user determines. The coefficients of these linear combinations are non-negative, and each function is a linear combination of the original attribute set [11] [12]. NMF decomposes a data matrix A into the product of two lower rank matrices W and H, yielding a result that is roughly equal to W times H. The initial values of W and H are modified by NMF using an iterative process until the product reaches A. The process ends [13][14] [15] when the approximation error converges or the required number of iterations is reached. An NMF model maps the original data into the new set of attributes (features) discovered by the model during model apply. initialize: W and H non negative.

Then update the values in W and H by computing the following, with n as an index of the iteration.

$$\mathbf{H}^{n+1}_{[i,j]} \leftarrow \mathbf{H}^{n}_{[i,j]} \frac{((\mathbf{W}^n)^T\mathbf{A})_{[i,j]}}{((\mathbf{W}^n)^T\mathbf{W}^n\mathbf{H}^n)_{[i,j]}} \tag{3}$$

and

$$\mathbf{W}^{n+1}_{[i,j]} \leftarrow \mathbf{W}^{n}_{[i,j]} \frac{(\mathbf{A}(\mathbf{H}^{n+1})^T)_{[i,j]}}{(\mathbf{W}^n\mathbf{H}^{n+1}(\mathbf{H}^{n+1})^T)_{[i,j]}} \tag{4}$$

Until W and H are stable.

NMF's numerical attributes are normalised by Automatic Data Preparation. When missing values occur in columns with simple data types (not nested), NMF considers them to be missing at random. Missing categorical values are replaced with the mode, and missing numerical values are replaced with the mean. When nested columns have missing values, NMF interprets them as sparse [16][17][18]. Sparse numerical data is replaced with zeros, and sparse categorical data is replaced with zero vectors.

NMF is a commonly used method for data dimensional reduction and feature extraction [19]. The key distinction between NMF and other factorization approaches, such as SVD, is that NMF allows only additive combinations of intrinsic parts,' i.e. hidden features. This is illustrated in, where NMF learns face parts and a face is naturally depicted as an additive linear combination of various parts. Negative combinations, on the other hand, are not as intuitive or natural as positive ones [20][21][22].

NMF is often used in bioinformatics to find 'metagenes' from expression profiles that are linked to biological pathways. NMF was used to derive trinucleotide mutational signatures from mutations present in cancer genomic sequences, and it was proposed that each cancer type's trinucleotide profile is a positive linear combination of these signatures[23][24].

For NMF decomposition, a variety of algorithms are available, including the multiplicative algorithms proposed in, gradient descent, and alternating non-negative least squares (ANLS) [25][26]. ANLS is gaining popularity because it guarantees a stationary point and is a faster non-negative least squares algorithm (NNLS). The resulting decomposed matrices have fewer entries than the original matrix as NMF is a dimension reduction process. This means that a decomposition does not include all of the entries in the original matrix and NMF should be able to accommodate missing entries in the target matrix [27][28] [29].

The authors in [4] have demonstrated that the performance of the final solution depends on the initial choice. The Multiplicative algorithm reported in [4] uses component wise division. If any element of (W^TWH) or (WHH^T) becomes zero, Multiplicative algorithm replaces zero by ϵ to overcome divide-by-zero problem where ϵ is a very small positive number. However, based on the value of the corresponding element in (W^TA) or (AH^T), the division by ϵ may generate high value which may suppress other values of W and H during normalization. This issue is addressed in this work by using thin QR decomposition [30][31] [32].

This paper aims to reduce the approximation error with rapid convergence rate of the NMF algorithm by addressing the shortcomings of the work reported in [4]. The structure of the paper is as follows. The details of the proposed algorithm are discussed in section 2. The results are presented in section 3 followed by discussion in Section 4. Section 5 concludes the paper.

2 Proposed NMF Algorithm

The number of significant clusters is denoted by k. The value of k is determined by SVD algorithm from the initial set of points or the input A matrix. The k clusters are determined and the centroid of the clusters are considered as one of the basis vectors based on the number of significant distinct singular values. Thus, k basis vectors are found to form W Matrix and W is positive. This is determined by k-means clustering algorithm.

The marix H is determined by thin QR decomposition after the matrix W is determined. $W = QR$, where Q is an orthogonal matrix, R is an upper triangular matrix and the marix H is expressed by (5) based on $WH = A$, $QRH = A$.

$$H = R^{-1}Q^TA \tag{5}$$

The marix H is determined but some of the elements are negative. The negative elements are replaced with zeros and the matrix is termed as H_1. Then applying thin QR decomposition on H_1 as below.

$$H_1 = QR \tag{6}$$

$$W_1 H_1 = A \tag{7}$$

$$W_1 QR = A \tag{8}$$

$$W_1 = AR^{-1}Q^T \tag{9}$$

At this point W_1 is normalised and the value of the error constraint given by equation (1) is computed. The process is repeated till the error is within the acceptable limit. The flowchart of the proposed algorithm is presented in Figure 1.

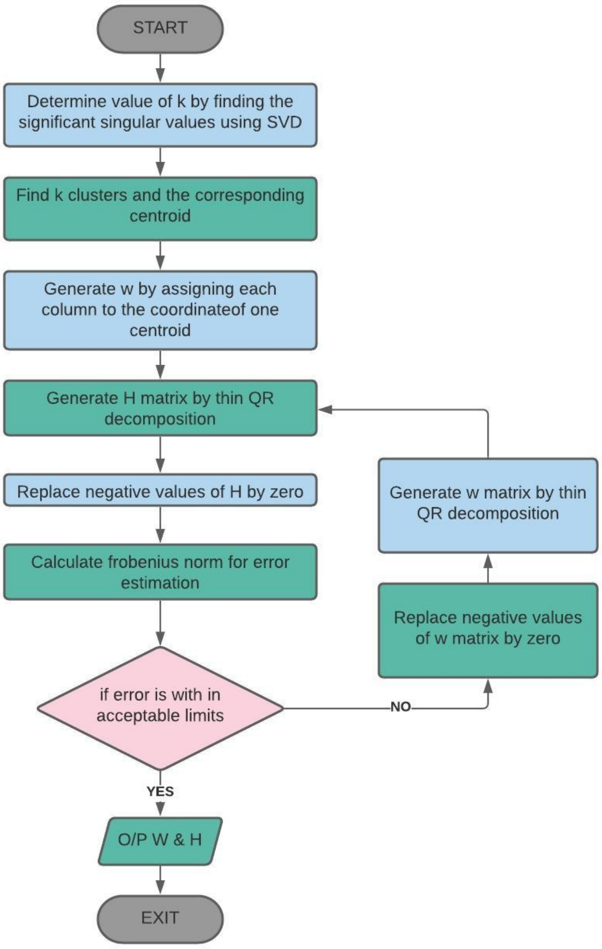
START
Determine value of k by finding the significant singular values using SVD
Find k clusters and the corresponding centroid
Generate w by assigning each column to the coordinateof one centroid
Generate H matrix by thin QR decomposition
Replace negative values of H by zero
Calculate frobenius norm for error estimation
if error is with in acceptable limits
NO
Replace negative values of w matrix by zero
Generate w matrix by thin QR decomposition
YES
O/P W & H
EXIT

Figure 1. Flow Chart of the Proposed Algorithm

3 Results

The proposed algorithm is implemented in MATLAB/Simulink environment for randomly generated matrix. The number of significant singular values of the generated matrix is chosen as the number of clusters. The comparison of error and the convergence rate of the proposed algorithm with the existing algorithms is depicted in Figure 2.

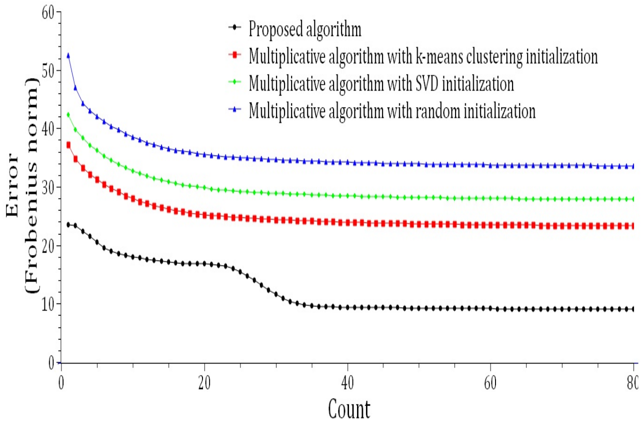

Figure 2. Approximation Error (Frobenius Norm) versus Iteration Number

It is observed that Multiplicative algorithm presented in [4] with different initial choice of W computed by SVD outperforms random initialization of W with respect to approximation error and convergence rate. This is because the chosen W matrix with SVD initialization is more closer to the actual in comparison with the randomly chosen W matrix. Finally, it is observed that the proposed algorithm outperforms all other variations of NMF reported in the literature with respect to relative approximation error and convergence rate.

The accurate measurement of execution time for individual program is not possible in MATLAB. Hence, the average execution of time over 200 iteration with increasing iteration count of the proposed algorithm is measured. A comparison of average execution time for the proposed algorithm is illustrated in Figure 3. It is observed that the proposed algorithm consumes 10% more execution time. All algorithms are of the same order in terms of execution time.

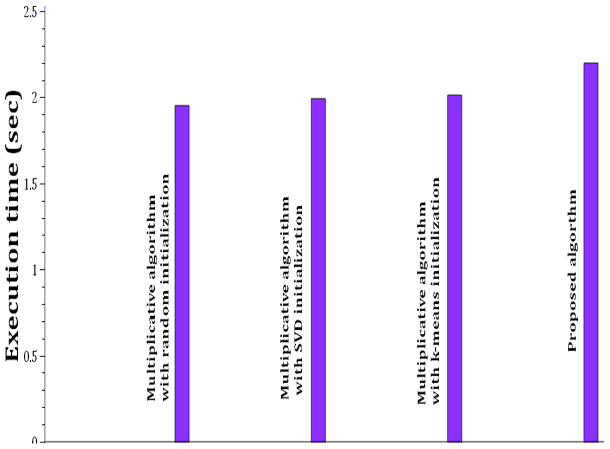

Figure 3. Comparison of Average Execution Time over multiple Iteration

4 Discussion

In the proposed algorithm, W is better approximated with k-means clustering as no approximation has taken place to suppress negative numbers. The value of k is chosen by SVD algorithm in the proposal which is one of the major contribution in reducing the error and increasing the convergence rate.

5 Conclusion

In this work an efficient NMF algorithm is proposed to reduce approximation error and achieve fast convergence rate. The major contribution is the selection of number of clusters by using SVD algorithm which determines the number of column in W matrix. The accurate initial choice of W matrix corresponding to significant singular values provides less error with rapid convergence rate in associate with gradient decent thin QR optimization. Simulation results are illustrated to validate the performance of the proposed NMF algorithm and compared with other NMF variants available in the literature.

REFERENCES

[1] N. R. Shamsuddin and N. I. Mahat, "Investigation on the clusterability of heterogeneous dataset by retaining the scale of variables," Mathematics and Statistics, vol. 7, no. 4A, pp. 49–57, 2019. DOI: 10.13189/ms.2019.070707

[2] M. Berry and et.al, "Algorithms and applications for approximate nonnegative matrix factorization," Technical report, Department of Computer Science, University of Tennessee, 2006.

[3] L. Elden, "Matrix methods in data mining and pattern recognition," SIAM Philadelphia, 2007.

[4] D. Lee and H. Seung, "Learning the parts of objects by nonnegative matrix factorization," Nature, vol. 401, pp. 788–791, 1999.

[5] R. Gray, "Vector quantization," IEEE ASAP Magazine, vol. 1, no. 2, pp. 4–29, 1984.

[6] I. T. Jolliffe, "Principal component analysis," Springer Series in Statistics, New York: Springer-Verlag, 2002.

[7] P. Paatero and U. Tapper, "Positive matrix factorization: A nonnegative factor model with optimal utilization of error estimates of data values," Environmetrics, vol. 5, pp. 111–126,1994.

[8] A. C. Fialkowski,. SimMultiCorrData: Simulation of correlated data with multiple variable types, Comprehensive R Archive Network (CRAN), 2017.

[9] C. Carmona, L. Nieto-Barajas, A. Canale. Model-based approach for household clustering with mixed scale variables, Advances in Data Analysis and Classification, Vo.13, No.2,559, 2019.

[10] E. C. de Assis,. R. M. C. R. de Souza. A k-medoids clustering algorithm for mixed feature-type symbolic data, EEE International Conference on Systems, Man and Cybernetics, 527-531, 2011.

[11] H. Ralambondrainy. A conceptual version of the K-means algorithm, Pattern Recognition Letters, Vol.16, No.11, 1147-1157, 1995.

[12] A. S. Shirkhorshidi, S. Aghabozorgi, T. Y. Wah. A comparison study on similarity and dissimilarity measures in clustering continuous data, PLoS ONE, Vol.10, No.12, e0144059, 2015.

[13] J. Ji, T. Bai, C. Zhou, C. Ma, Z. Wang. An improved k-prototypes clustering algorithm for mixed numeric and categorical data, Neurocomputing, Vol.120, 590-596, 2013.

[14] J. C. Bezdek, N. R. Pal. Some new indexes of cluster validity, Part B (Cybernetics) IEEE Transactions on Systems, Man, and Cybernetics, Vol.28, No.3, 301-315, 1998.

[15] S. Saitta, B. Raphael, I. F.C. Smith. A bounded index for cluster validity, Machine Learning and Data Mining in Pattern Recognition, Vol.30, 174-187, 2007.

[16] C. Hennig, T. F. Liao. How to find an appropriate clustering for mixed-type variables with application to socio-economic stratification, Journal of the Royal Statistical Society: Series C (Applied Statistics), Vol.62, No.3, 309-369, 2013.

[17] Karvanen, J., Cichocki, A., 2003. Measuring sparseness of noisy signals. In: Proceedings of the Fourth International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), Nara, Japan.

[18] Keila, P., Ski005. Detecting unusual and deceptive communication in email. Technical Report, School of Computing, Queen's University, Kingston, Ont., Canada.

[19] 2003. A survey of spectral unmixing algorithms. Lincoln Laboratory J. 14 (1), 55–77.

[20] Langville, A., Meyer, C., Albright, R., Cox, J., Duling, D., 2006. Algorithms, initializations, and convergence for the nonnegative matrix factorization, preprint. International Conference on Data Mining, April 22–24. SIAM, Lake Buena Vista, FL.

[21] Pauca, P., Piper, J., Plemmons, R., 2006a. Nonnegative matrix factorization for spectral data analysis. Linear Algebra Appl. 416 (1), 29–47.

[22] Pauca, V., Plemmons, R., Abercromby, K., 2006b. Nonnegative matrix factorization methods with physical constraints for spectral unmixing, in preparation.

[23] Piper, J., Pauca, V., Plemmons, R., Giffin, M., 2004. Object characterization from spectral data using nonnegative factorization and information theory. In: Proceedings of the 2004 AMOS Technical Conference, Maui, HI, September.

[24] Plaza, A., Martinez, P., Perez, R., Plaza, J., 2004. A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data. IEEE Trans. on Geoscience and Remote Sensing 42 (3), 650–663.

[25] Polak, E., 1971. Computational Methods in Optimization: A Unified Approach. Academic Press, New York.

[26] Powell, M., 1964. An efficient method for finding the minimum of a function of several variables without calculating derivatives. Comput. J. 7, 155–162.

[27] Cichocki, A., Zdunek, R., 2006. NMFLAB for Signal Processing, available at http://www.bsp.brain.riken.jp/ICALAB/nmflab.html .

[28] Cichocki, A., Zdunek, R., Amari, S., 2006. Csiszar's divergences for non-negative matrix factorization: family of new algorithms. In: Proceedings of the Sixth International Conference on Independent Component Analysis and Blind Signal Separation, Charleston, SC, March 5–8.

[29] de Leeuw, J., Young, F., Takane, Y., 1976. Additive structure in qualitative data: an alternating least squares method with optimal scaling features. Psychometrika 41, 471–503.

[30] Dhillon, I., Sra, S., 2005. Generalized nonnegative matrix approximations with bregman divergences. In: Proceeding of the Neural Information Processing Systems (NIPS) Conference, Vancouver, BC.

[31] Ding, C., He, X., Simon, H., 2005. On the equivalence of nonnegative matrix factorization and spectral clustering. In: Proceedings of the Fifth SIAM International Conference on Data Mining, Newport Beach, CA.

[32] Donoho, D., Stodden, V., 2003. When does non-negative matrix factorization give a correct decomposition into parts? In: Seventeenth Annual Conference on Neural Information Processing Systems.