1 Paper Title

Learning the parts of objects by non-negative matrix factorization

2 Summary

This paper discusses the NMF algorithm and states its merits over methods like PCA (Principal Component Analysis) and VQ (Vector Quantization). The authors apply NMF to a database of facial images and compare results with the results from PCA and VQ. The authors also talk about other applications of NMF and give intuitive explanations of this widely-used algorithm.

3 Detailed Analysis

The NMF is used to learn a low-rank approximation of data with the constraint that the basis and coefficient matrices are all non-negative. Let V be the data matrix (nxm) and W be the basis matrix (nxr) and H be the weight matrix whose each column in called an encoding, which is coefficients for representing an image in the basis learned by W. The dimension r of the basis and coefficients is chosen such that $(n+m)r < nm$. The data matrix that the authors consider is one of images - containing n pixels for each image for m images. Both W and H are unknowns. The objective function discussed in this paper is:

$$F = \Sigma_{i=1}^{n} \Sigma_{\mu=1}^{m} [V_{i\mu} log(WH)_{i\mu} - (WH)_{i\mu}] \tag{1}$$

subject to the constraints that both W and H are non-negative. The update rules of W and H are given by:

$$W_{ia} \leftarrow W_{ia} \Sigma_{\mu} \frac{V_{i\mu}}{(WH)_{i\mu}} H_{a\mu} \tag{2}$$

$$W_{ia} \leftarrow \frac{W_{ja}}{\Sigma_j W_{ja}} \tag{3}$$

$$H_{a\mu} \leftarrow H_{a\mu} \Sigma_i \frac{W_{ia}}{(V)_{iu}} WH_{i\mu} \tag{4}$$

The above is the iterative algorithm for NMF which is an EM algorithm and is likely to converge on a local minima. It is to be noted that these update rules are such that the non-negativity of H and W is preserved as that is our primary goal when we do NMF.
Now let's see how the constraints in NMF are different from PCA and VQ. In PCA, the columns of the basis are orthogonal to each other, but the coefficients can be positive or negative, thereby adding or subtracting different basis vectors to get the image, which is not intuitive. One of the advantages of NMF in having non-negative coefficients is that the reconstruction in additive hence makes more sense. It is intuitive to combine different parts to construct let's say a face and this is how humans view an image as well. In VQ however, the H matrix is sparse having only one element equal to unity and the rest zero, which means each image is given by a basis.
The authors note that the basis matrix (W) and coefficient matrix (H) in case of NMF are both sparse. The basis images contain various versions of only a part of the image and hence is sparse whereas an image is a sparse combination of only a subset of such features, hence the H matrix is sparse. This sparse image encoding is in stark constrast to VQ and PCA that both contain whole face-like images (in case of face dataset that the authors consider in this paper).
The authors also generaize the above analysis to other problems like semantic analysis of documents. The V matrix is a bag of words of document vs frequency of words in this case. As NMF, PCA and VQ can be used for understanding hidden variables in the data (H), the hidden topics (feature set W) in each document can be known from applying this algorithm. In this case even, VQ is not the ideal choice because we get only one topic for each document while a document can be about multiple topics. PCA isn't intuitive because what does subtracting a topic mean from other topics to get a document. NMF is widely used for semantic analysis and clearly highlights which topics are being talked about in a particular document.
The demerits of NMF however, include data that lies on a non-linear manifold (like same images taken from different viewpoints). Also, NMF is not making any statistical assumptions and hence doesn't give more statistical information about the hidden variables.

4 Your critiques

4.1 Assumptions made

The authors introduce NMF without first introducing matrix approximation and that might confuse the uninitiated reader. Also, it is assumed that the readers are aware of PCA and VQ and there is a lot of emphasis on the comparison between these three models. The authors also introduce ICA, which is a more evolved technique and assume the reader's familiarity with all these techniques. However, these assumptions can be somewhat justified as this paper is published in ”Letters to Nature” which essentially publishes short reports on important findings and isn't much concerned with the completeness of the paper.

4.2 Pros and cons

The authors treat the topic very succintly and show the uses and impact of the NMF algorithm in a very practical way. However, they don't give a proper justification of the objective function for which the reader is referred to other papers. I feel some more explanation was needed for the actual algorithm that computes NMF and how it came about. The authors compare NMF to popular algorithms like PCA, ICA and VQ, thus helping the reader make more connections about where to use which algorithm which is absolutely fantastic for the reader who is aware of thise algorithms, but not so much for those who pick this paper up only to know about NMF. Such readers will be disappointed to find out that they need to understand three other algorithms in order to really enjoy this paper. The best part of this paper was to know how the same algorithm seamlessly translates from one area (images) to another (text) and is equally effective, which made me want to apply it in a real-world scenario.
As for NMF, the pros of this algorithm include giving compact and intuitive representations of real-world data. The representations are sparse which is a huge advantage as it makes many problems computationally tractable. But NMF cannot be blindly applied to any dataset and for more complex representations, can be combined with other algorithms for better working.

4.3 Possible future extensions

I would love to see NMF being applied in other fields like medical data [2], hyperspectral data (for feature extraction) and audio data. The optimization works great for smaller matrices, but for huge matrices, we need better optimization methods[3] to make this tractable. NMF can also be applied on clustering problems and that is an interesting application to talk about. Also, since the objective function converges at local minima sometimes, there is a lot of scope for better optimization algorithms to find global minima. Evolutionary algorithms[1] can be tried out to cover a larger space and find global minima.

5 Closing Remarks

5.1 What you have learned from reading this paper ?

1. I learned about a new algorithm and its real-world applications.
2. I also learned which algorithms are suitable in which cases, and the need to critically analyze and form connections with prior learnings when analyzing a new algorithm.

References

[1] Janecek, A., and Tan, Y. Iterative improvement of the multiplicative update nmf algorithm using nature-inspired optimization. In *2011 Seventh International Conference on Natural Computation* (2011), vol. 3, IEEE, pp. 1668–1672.

[2] Leuschner, J., Schmidt, M., Fernsel, P., Lachmund, D., Boskamp, T., and Maass, P. Supervised non-negative matrix factorization methods for maldi imaging applications. *Bioinformatics 35*, 11 (2019), 1940–1947.

[3] Lin, X., and Boutros, P. C. Optimization and expansion of non-negative matrix factorization. *BMC bioinformatics 21*, 1 (2020), 1–10.