Signal Processing ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Non-informative hierarchical Bayesian inference for non-negative matrix factorization

Qingquan Sun [a,*], Fei Hu [b], Yeqing Wu [b], Jiang Lu [b], Xinlin Huang [c]

[a] School of Computer Science and Engineering, California State University San Bernardino, San Bernardino, CA 92407, USA
[b] Department of Electrical and Computer Engineering, The University of Alabama, Tuscaloosa, AL 35487, USA
[c] Department of Information and Communication Engineering, Tongji University, Shanghai 201804, China

ARTICLE INFO

Article history:
Received 30 April 2014
Received in revised form
29 July 2014
Accepted 2 September 2014

Keywords:
Non-negative Matrix Factorization (NMF)
Bayesian inference
Model determination
Ground-truth bases

ABSTRACT

Non-negative matrix factorization (NMF) is an intuitive, non-negative, and interpretable approximation method. Canonical NMF approach could derive some basic components to represent original data, while probabilistic NMF approaches try to introduce some reasonable constraints to optimize the canonical NMF model. However, both of them cannot handle ground-truth bases discovering and model order determination problems. In general, the model order of basis matrix needs to be pre-defined. The model order determines the capability and accuracy of data structure discovering. However, how to accurately infer the model order of basis matrix has not been well investigated. In this paper, we propose a method called non-informative hierarchical Bayesian non-negative matrix factorization (NHBNMF) to automatically determine the model order and discover the data structure. They are achieved through hierarchical Bayesian inference model, maximum a posteriori (MAP) criterion, and non-informative parameters. In NHBNMF method, we first introduce a structure with two-level parameters to enable the entire model to approach the distributions of ground-truth bases. Then we use non-informative parameter scheme to eliminate the hyper-parameter to enable automatic searching. Finally, the model order and ground-truth bases are discovered by using MAP criterion and L_2-norm selection. The experiments are conducted based on both synthetic and real-world datasets to show the effectiveness of our algorithm. The results demonstrate that our algorithm can accurately estimate the model order and discover the ground-truth bases. Even for the complicated FERET facial dataset, our algorithm still obtained interpretable bases and achieved satisfactory accuracy of the model order estimation.

1. Introduction

Non-negative matrix factorization (NMF) has become a popular technique since it was proposed by Lee and Seung [1] in 1999. NMF has demonstrated its power and capabilities in many research fields such as image and video processing [2–4], audio and acoustic signal processing [5–7], text and semantic analysis [8–10]. NMF is widely applied due to its non-negative, interpretable, and part-based representative properties. As we know, there is no negative values in the physical world. Compared with principal components analysis (PCA) [11] and independent components analysis (ICA) [12], NMF adds the non-negative constraint to all the elements. This is the most impressive feature of NMF to fit the physical world. In NMF, given a non-negative dataset X, we intend to find two non-negative factor matrices $W \in R^{M \times K}$ and $H \in R^{K \times N}$, which are named base matrix and feature matrix. In addition,

* Corresponding author.
E-mail addresses: quanqian12345@gmail.com (Q. Sun), fei@eng.ua.edu (F. Hu), ywu40@crimson.ua.edu (Y. Wu), jlu5@crimson.ua.edu (J. Lu), xlhuang@tongji.edu.cn (X. Huang).

http://dx.doi.org/10.1016/j.sigpro.2014.09.004
0165-1684/

W and H satisfy

$$X \approx WH \quad \text{s.t.} W \geq 0, H \geq 0 \tag{1}$$

K is an important parameter here, and its value is the model order. Additionally, K usually satisfies the inequality $K \leq MN/(M+N)$.

During the past several years, many variants of NMF algorithm have been proposed to improve its performance. Most of the variants can be classified into two categories. One is sparseness-oriented category, the other is manifold-oriented category. The sparseness-oriented algorithms aim to enhance the sparseness of basis by introducing certain constraints. Sparseness is consistent to the nature of NMF algorithm, which is part-based representation. Sparseness in NMF algorithm is different from that in sparse linear regression. In sparse linear regression, the sparseness only acts on H, while the direction W is fixed. Whereas, in NMF algorithms, sparseness refers to the total number of coefficients required to encode the data. The typical algorithms belonging to such category are sparse NMF algorithm proposed in [13–15] and localized NMF proposed in [16]. In comparison, manifold-oriented variants aim to find the low-dimension manifold of original data set. Such kind of algorithms often apply graph embedding approach to preserve the geometry information of original data into the surrogate low-dimension manifold. One typical algorithm is called non-negative graph embedding [17].

Although sparseness constraint and manifold learning can improve the performance of NMF algorithm, the determination of model order is even more important to improve NMF's performance. Unfortunately, this issue has not received sufficient attention and investigation.

From machine learning and data mining perspective, we always attempt to extract the hidden structure of data. More accurate hidden structure extraction can achieve better representation and recognition. On one hand, the hidden structure indicates the real composition of data; on the other hand, it enables the factorization to be interpretable. For instance, suppose a human face can be represented only by four basic components: eyebrows, eyes, nose and mouse, namely, the four basic components are the ground-truth bases to represent a face. So if we can determine that the model order is 4 and can find the true bases, then we can accurately represent the face; On the contrary, if we determine the model order as other numbers rather than 4, then we have to use other parts to represent the face. Obviously, other parts are not the intrinsic features of a face, it is not practical to use them to accurately represent the face. The model order of factorized basis is the most important parameter to evaluate the accuracy of structure extraction. Furthermore, the accurate structure could help us get better understanding and analysis of data, thus improving the performance in applications.

The main challenge of model order determination problem is little prior knowledge available, thus it is hard to approach the real distributions of bases. Consequently, the real model order cannot be discovered. Usually, the model order and cost function need to be pre-defined. There are no more prior knowledge introduced to the algorithm in previous methods. That is why the canonical NMF method and traditional Bayesian method (ML, MAP) cannot handle the model order determination problem. Although fully Bayesian method is a choice to achieve model order determination, its computation cost is too high. Moreover, the accuracy of this approach is also dependent on the hyper-parameter's distribution. If the choice of hyper-parameter's distribution can not indicates the real condition, we can not obtain the expected results.

In order to overcome the dilemma of discovering model order and high-computation cost, motivated by the model order selection method used in Bayesian PCA [18], we propose a hierarchical Bayesian inference method (in which we introduce two level parameters into the inference model) to seek the correct model order of factorized basis. Furthermore, we utilize non-informative prior as the parameter of the hyper-parameter (second level parameter) to enable our model to approach the real distributions of basis automatically. Then we use L_2-norm as the selection function to obtain the value of model order. Experimental results on three datasets demonstrate the efficiency of our algorithm.

The rest of this paper is organized as follows. Section 2 provides a brief review of related works on model order determination in NMF. In Section 3, we describe our non-informative hierarchical Bayesian inference algorithm in details. The analysis and evaluation of experimental results are provided in Section 4. Section 5 concludes the paper.

2. Related works

Although sparseness optimization and manifold learning are different techniques, they are consistent to the part-based representation principle of NMF. To some extents, sparseness optimization, manifold learning and model order determination are identical, that is, to use a subset of localized features or structures to represent original data. As in localized non-negative matrix factorization (LNMF) [16,19–21], some local features are learned to represent data. While in projective non-negative matrix factorization (PNMF) [22], the locality is further enhanced. The bases become smaller and sharper. Such kind of optimization can improve the performance because during the sparseness optimization procedure, some less-important features are abandoned and only the distinct features are preserved.

Instead of focusing on the locality of data, manifold learning technique concerns the structure and relationship among data points. Since for some recognition applications such as face recognition, the distinct local structure is more powerful than the local features [23]. The local structure here involves some geometry and distance information. Usually graphs are constructed to embed geometry and distance information into manifold. For NMF algorithm, an approach called non-negative graph embedding (NGE) is proposed in [17], which extends the general graph embedding framework to matrix factorization problem. But this method involves a high computation cost. Wang et al. [24] improved this work through multiplicative updating rule and proposed multiplicative non-negative graph embedding (MNGE) algorithm. However, both of the above techniques are unable to find the ground-truth basis. It means that they are unable to interpret the real composition of data or provide accurate representations.

Actually, there are few literatures discussing ground-truth basis discovering and model order selection issues. Sparse

Please cite this article as: Q. Sun, et al., Non-informative hierarchical Bayesian inference for non-negative matrix factorization, Signal Processing (2014), http://dx.doi.org/10.1016/j.sigpro.2014.09.004