Information Sciences 257 (2014) 369–387

Contents lists available at SciVerse ScienceDirect

Information Sciences

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

Subtractive clustering for seeding non-negative matrix factorizations

Gabriella Casalino [a], Nicoletta Del Buono [b,*], Corrado Mencar [a]

[a] *Computer Science Department, Università degli Studi di Bari Aldo Moro, Via E. Orabona 4, I-70125 Bari, Italy*
[b] *Mathematics Department, Università degli Studi di Bari Aldo Moro, Via E. Orabona 4, I-70125 Bari, Italy*

ARTICLE INFO

Article history:
Available online 20 June 2013

Keywords:
Non-negative matrix factorization
Subtractive clustering
Initialization method

ABSTRACT

Non-negative matrix factorization is a multivariate analysis method which is proven to be useful in many areas such as bio-informatics, molecular pattern discovery, pattern recognition, document clustering and so on. It seeks a reduced representation of a multivariate data matrix into the product of basis and encoding matrices possessing only non-negative elements, in order to learn the so called part-based representations of data. All algorithms for computing non-negative matrix factorization are iterative, therefore particular emphasis must be placed on a proper initialization of NMF because of its local convergence. The problem of selecting appropriate starting matrices becomes more complex when data possess special meaning as in document clustering. In this paper, we propose the adoption of the subtractive clustering algorithm as a scheme to generate initial matrices for non-negative matrix factorization algorithms. Comparisons with other commonly adopted initializations of non-negative matrix factorization algorithms have been performed and the proposed scheme reveals to be a good trade-off between effectiveness and speed. Moreover, the effectiveness of the proposed initialization to suggest a number of basis for NMF, when data distances are estimated, is illustrated when NMF is used for solving clustering problems where the number of groups in which the data are grouped is not known a priori. The influence of a proper rank factor on the interpretability and the effectiveness of the results are also discussed.

1. Introduction

The modern era is characterized by a fast growing amount of data, which are often stored in huge non-negative matrices. Examples are documents in document collections, which can be stored as columns of the so called term-by-document matrix, whose elements count the number of times (possibly weighted) a corresponding term appears in a selected document. Similarly, digital images or photos can be represented by matrices whose elements correspond to the intensity and/or the color of each pixel. In recommender systems, the information for a purchase history of customers or ratings on a subset of items is stored as elements of non-negative sparse matrix. Further noteworthy examples are non-negative matrices storing data generated when decoding genes of living creatures or data acquired from the observations of space. Thus, it is natural to uncover the latent low-dimensional structure that is often inherent in these high-dimensional structured data. Moreover, due to the non-negativity character of these data, it should be useful to seek low-dimensional approximations that preserve this property—since non-negativity enhances meaningful interpretations of mined information [12].

[*] Corresponding author. Tel.: +39 0805442204.
E-mail addresses: gabriella.casalino@uniba.it (G. Casalino), delbuono@dm.uniba.it (N. Del Buono), mencar@di.uniba.it (C. Mencar).

0020-0255/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2013.05.038

Recently, non-negative matrix factorization (NMF) received an increasing attention from the data analysis community, due to its capabilities of obtaining a reduced representation of data only using non-negative restrictions [13,25,26]. These constraints led to a part-based representation, because they allow non-negative linear combination of a set of non-negative "bases" that may represent realistic "building blocks" for the original data. More formally, a NMF consists in approximating (in some sense) an initial set of non-negative data expressed by a $n \times m$ matrix X, whereas each entry X_{ij} represents in a broad sense the score obtained by the entity j on the variable i, with the product of two reduced rank (namely, the factor rank r) non-negative matrices W (the basis matrix), and H (the encoding matrix), so that $X \approx WH$.

In this way, the perception of the whole, being it an image or a document in a collection, becomes a combination of its parts represented by basis vectors. In contrast to other traditional dimensionality reduction algorithms, such as Principal Component Analysis (PCA) [22], Independent Component Analysis (ICA) [20] or Singular Value Decomposition (SVD) [17], NMF decomposition implies that the original data are reconstructed using only additive and no subtractive combinations of the basic elements. Therefore, the NMF basis is able to extract localized features that correspond with intuitive notions of parts of the original data. This correspondence is, in fact, a kind of knowledge about the underlying problem, which can be used not only to guide learning procedures, but also to provide interpretability to users. For instance, in the standard vector space model, when the data matrix X represents a term-by-document matrix, the column vectors of the basis matrix W identify a set of words denoting a particular concept or topic and the elements of H represent the weights used to linearly combine the basis vectors in order to approximate each column of X. Hence, each document is viewed as combination of basis vectors and it can be categorized as belonging to one or more topics. In this way, non-negative factors of NMF can be directly applied to perform clustering that identifies semantic features in a document collection and groups the documents into clusters on the basis of shared semantic features. Clustering capabilities of NMF have been highlighted in several works [25,29,31,41]; moreover, some theoretical results showing the relationship between NMF and the classical k-means clustering have been proved in [16,27].

All algorithms for computing NMF are iterative and require initialization of the basis and the encoding matrices [4,38,39]. Therefore, the efficiency of many NMF algorithms is affected by the selection of the starting matrices: poor initialization often results in slow convergence or lower error reduction. Furthermore, the problem of selecting an appropriate initialization becomes more complicated when additional structures or constraints are imposed on the factorized matrices, or when the data possess special meaning. Different initialization mechanisms have been proposed in literature: some of them lead to rapid error reduction and faster convergence of the adopted NMF algorithm, others lead to a good overall error accuracy at convergence [5,6]. However, there does not exist a definitive suggestion about the best initialization strategy to be adopted for different NMF algorithms [14]. The choice of the proper factor rank r is still an open issue, as in the case of k-means algorithm, which is strictly related to a NMF as highlighted in [16].

This paper contains three main parts. The first part addresses the problem of NMF initialization schemes. In this way, in Section 3, we provide a kind of review where an original and purposely conceived categorization of a number of initialization schemes is proposed. This review is useful for highlighting the main advantages and disadvantages in the use of each particular initialization method.

The second part of the paper is devoted to the proposal of an additional method for generating the initial matrices $W^{(0)}$ and $H^{(0)}$ for any NMF algorithm, based on the subtractive clustering scheme [7]. Our proposal differs from the reviewed schemes in the sense that it enables the suggestion of a rank factor, whenever a distance statistics distance between initial data is provided. Such a feature appears to be particularly relevant, since it represents a tentative of solving a crucial point in the context of NMF, that is the choice of the most appropriate rank factor r. Furthermore, this peculiarity would be very beneficial to the common user who has no justification for the a priori selection of a particular rank r of a low rank approximation when dealing with data in a more general unsupervised learning scenario.

The last part of the paper describes a experimental session performed on text-by-document data chosen among the datasets most frequently employed in literature. The experiments have been conducted by adopting different NMF algorithms and different initialization methods. This led to a large number of experiments which proved to be useful both to assess the capability of the proposed approach (when compared with initial algorithms belonging to the class of clustering based initialization) to improve the performance of some NMF algorithms (either for consideration of computational complexity and for suggesting the rank factor value) and to provide some insights on the performance of NMF algorithms, when initialized by different schemes.

2. Background of non-negative matrix factorization and algorithms

Non-negative matrix factorization aims to obtain a linear representation of multivariate data under non-negativity constraints. These constraints lead to a part-based representation because only additive, not subtractive, combinations of the original data are allowed [26].

From a mathematical point of view, given an initial set of data expressed by a $n \times m$ matrix X, NMF seeks a set of non-negative "bases" column vectors $W = [W_1, \ldots, W_r]$, that are non-negatively (conically) combined to approximate each column of the matrix $X = [X_1, \ldots, X_m]$. Particularly, NMF looks for non-negative W_k and H_{kj}, such that:

$$X_j \approx \sum_{k=1}^{r} W_k H_{kj}, \quad H_{kj} \geqslant 0, \quad j = 1, \ldots, m, \tag{1}$$