Neurocomputing 134 (2014) 206–213

Contents lists available at ScienceDirect

Neurocomputing

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

Adaptive multiplicative updates for quadratic nonnegative matrix factorization

He Zhang*, Zhirong Yang, Erkki Oja

Department of Information and Computer Science, Aalto University, Espoo, Finland

ARTICLE INFO

Article history:
Received 8 March 2013
Received in revised form
7 May 2013
Accepted 25 June 2013
Available online 13 January 2014

Keywords:
Adaptive
Multiplicative updates
Quadratic
Nonnegative matrix factorization

ABSTRACT

In Nonnegative Matrix Factorization (NMF), a nonnegative matrix is approximated by a product of lower-rank factorizing matrices. Quadratic Nonnegative Matrix Factorization (QNMF) is a new class of NMF methods where some factorizing matrices occur twice in the approximation. QNMF finds its applications in graph partition, bi-clustering, graph matching, etc. However, the original QNMF algorithms employ constant multiplicative update rules and thus have mediocre convergence speed. Here we propose an adaptive multiplicative algorithm for QNMF which is not only theoretically convergent but also significantly faster than the original implementation. An adaptive exponent scheme has been adopted for our method instead of the old constant ones, which enables larger learning steps for improved efficiency. The proposed method is general and thus can be applied to QNMF with a variety of factorization forms and with the most commonly used approximation error measures. We have performed extensive experiments, where the results demonstrate that the new method is effective in various QNMF applications on both synthetic and real-world datasets.

1. Introduction

Nonnegative Matrix Factorization (NMF) has attracted a lot of research effort in recent years (e.g. [1–6]). NMF has a variety of applications e.g. in machine learning, signal processing, pattern recognition, data mining, and information retrieval (e.g. [7–11]). Given an input data matrix, NMF finds an approximation that is factorized into a product of lower-rank matrices, some of which are constrained to be nonnegative. The approximation error can be measured by a variety of divergences between the input and its approximation (e.g. [6,12–14]), and the factorization can take a number of different forms (e.g. [15–17]).

In most existing NMF methods, the approximation is linear with respect to each factorizing matrix, that is, these matrices appear only once in the approximation. However, such a linearity assumption does not hold in some important real-world problems. A typical example is graph matching, when it is presented as a matrix factorizing problem, as pointed out by Ding et al. [18]. If two graphs are represented by their adjacency matrices A and B, then they are isomorphic if and only if a permutation matrix P can be found such that $A-PBP^T=0$. Minimizing the norm or some other suitable error measure of the left-hand side with respect to P, with suitable constraints, reduces the problem to an NMF problem. Note that both adjacency matrices and permutation matrices are nonnegative, and the approximation is now quadratic in P.

Another example is clustering: if X is a matrix whose n columns need to be clustered into r clusters, then the classical K-means objective function can be written as [19] $\mathcal{J}_1=\mathrm{Tr}(X^TX)-\mathrm{Tr}(U^TX^TXU)$ where U is the $(n\times r)$ binary cluster indicator matrix. It was shown in [20] that minimizing $\mathcal{J}_2=\|X^T-WW^TX^T\|^2_{\text{Frobenius}}$ with respect to an orthogonal and nonnegative matrix W gives the same solution, except for the binary constraint. This is another NMF problem where the approximation is quadratic in W.

Methods for attacking this kind of problems are called Quadratic Nonnegative Matrix Factorization (QNMF). A systematic study on QNMF was given by Yang and Oja [21], where they presented a unified development method for multiplicative QNMF optimization algorithms.

The original QNMF multiplicative update rules have a fixed form, where an exponent in the multiplying factor in these rules remains the same in all iterations. Despite its simplicity, the constant exponent corresponds to overly conservative learning steps and thus often leads to mediocre convergence speed.

Here we propose new multiplicative algorithms for QNMF to overcome this drawback. We drop the restriction of constant exponent in multiplicative update rules, which relaxes the updates by using variable exponents in different iterations. This turns out to be an effective strategy for accelerating the optimization while

* Corresponding author.
E-mail addresses: he.zhang@aalto.fi (H. Zhang), zhirong.yang@aalto.fi (Z. Yang), erkki.oja@aalto.fi (E. Oja).

0925-2312/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.06.047

still maintaining the monotonical objective decrease. The acceleration for Projective NMF, a special case of QNMF, was presented in our preliminary work [22]. In this paper we demonstrate that the new method can bring improvement for many other QNMF optimizations. We generalize the adaptive multiplicative algorithm for a wide variety of QNMF problems. In addition to Projective Nonnegative Matrix Factorization, we also apply the strategy on two other special cases of QNMF, with corresponding application scenarios in bi-clustering and estimation of hidden Markov chains. Extensive empirical results on both synthetic and real-world data justify the efficiency advantage by using our method.

In the following, Section 2 recapitulates the essence of the QNMF objectives and their previous optimization methods. Section 3 presents the fast QNMF algorithm by using adaptive exponents. In Section 4, we provide empirical comparison between the new algorithm and the original implementation on three applications of QNMF. Section 5 concludes the paper.

2. Quadratic nonnegative matrix factorization

Nonnegative matrix factorization (NMF) finds an approximation $\widehat{X}$ to an input data matrix $X \in \mathbb{R}^{m\times n}$:

$$X \approx \widehat{X} = \prod_{q=1}^{Q} F^{(q)}. \tag{1}$$

and some of these matrices are constrained to be nonnegative. The dimensions of the factorizing matrices $F^{(1)},\ldots,F^{(Q)}$ are $m\times r_1, r_1 \times r_2,\ldots,r_{Q-1}\times n$, respectively. Usually $r_1,\ldots,r_{Q-1}$ are smaller than m or n.

In most conventional NMF approaches, the factorizing matrices $F^{(q)}$ in Eq. (1) are all different, and thus the approximation $\widehat{X}$ as a function of them is *linear*. However, there are useful cases where some matrices appear more than once in the approximation. In this paper we consider the case that some of them may occur twice, or formally, $F^{(s)} = F^{(t)^T}$ for a number of non-overlapping pairs $\{s,t\}$ and $1 \le s < t \le Q$. We call such a problem and its solution *Quadratic Nonnegative Matrix Factorization* (QNMF) because $\widehat{X}$ as a function is quadratic to each twice appearing factorizing matrix.[1]

When there is only one doubly occurring matrix W in the QNMF objective, the general approximating factorization form is given by [21]

$$\widehat{X} = AWBW^TC, \tag{2}$$

where the products of the other, linearly appearing factorizing matrices are merged into single symbols. QNMF focuses on the optimization over W, while learning the matrices that occur only once can be solved by using the conventional NMF methods of alternative optimization over each matrix separately [11].

As stated by the authors in [21], The factorization form unifies many previously suggested QNMF objectives. For example, it becomes the *Projective Nonnegative Matrix Factorization* (PNMF) when $A=B=I$ and $C=X$ [17,20,23–25]. If X is a square matrix and $A=C=I$, the factorization can be used in two major scenarios if the learned W is highly orthogonal: (1) when B is much smaller than X, the three-factor approximation corresponds to a blockwise representation of X [26,27]. If B is diagonal, then the representation becomes diagonal blockwise, or a partition. In the extreme case $B=I$, the factorization reduces to the *Symmetric Nonnegative Matrix Factorization* (SNMF) $\widehat{X} = WW^T$ [16]. (2) When X and B are of the same size, the learned W with the constraint $W^TW=I$ approximates a permutation matrix and thus QNMF can be used for learning order of relational data, for example, graph matching [18]. Alternatively, under the constraint that W has column-wise unitary sums, the solution of such a QNMF problem provides parameter estimation of hidden Markov chains (see Section 4.3).

The factorization form in Eq. (2) can be recursively applied to the cases where there are more than one factorizing matrices appearing quadratically in the approximation. For example, the case $A=C^T=U$ yields $\widehat{X}=UWBW^TU^T$, and $A=B=I$, $C=XUU^T$ yields $\widehat{X}=WW^TXUU^T$. An application of the latter example is shown in Section 4.2, where the solution of such a QNMF problem can be used to group the rows and columns of X simultaneously. This is particularly useful for the biclustering or coclustering problem. These factorizing forms can be further generalized to any number of factorizing matrices. In such cases we employ alternative optimization over each doubly occurring matrix.

It is important to notice that quadratic NMF problems are not special cases of linear NMF [21]. In linear NMF, the factorizing matrices are different variables and the approximation error can alternatively be minimized over one of them while keeping the others constant. In contrast, the optimization of QNMF is harder because matrices in two places vary simultaneously, which leads to higher-order objectives. For example, given the squared Frobenius norm (Euclidean distance) as approximation error measure, the objective of linear NMF $\|X-WH\|_F^2$ is quadratic with respect to W and H, whereas the PNMF objective $\|X-WWX\|_F^2$ is quartic with respect to W. Minimizing such a fourth-order objective with the nonnegativity constraint is considerably more challenging than minimizing a quadratic function.

3. Adaptive QNMF

The difference between the input matrix X and its approximation $\widehat{X}$ can be measured by a variety of divergences. Yang and Oja [21] presented a general method for developing optimization algorithms with multiplicative updates for QNMF based on α-divergence, β-divergence, γ-divergence or Rényi divergence. These families include the most popular used NMF objectives, for example, the squared Euclidean distance ($\beta=1$), Hellinger distance ($\alpha=0.5$), χ^2-divergence ($\alpha=2$), I-divergence ($\alpha\to1$ or $\beta\to0$), dual I-divergence ($\alpha\to0$), Itakura-Saito divergence ($\beta\to-1$) and Kullback-Leibler divergence ($\gamma\to0$ or $r\to1$). Multiplicative update rules can be developed for more QNMF objectives, for example the additive hybrids of the above divergences, as well as many other unnamed Csiszár divergences and Bregman divergences.

In general, the multiplicative update rules take the following form:

$$W_{ik}^{\text{new}} = W_{ik}\left[\frac{(A^TQC^TWB^T+CQ^TAWB)_{ik}}{(A^TPC^TWB^T+CP^TAWB)_{ik}}\cdot\theta\right]^{\eta}, \tag{3}$$

where P, Q, θ, and η are specified in Table 1. For example, the rule for QNMF $X\approx WBW^T$ based on the squared Euclidean distance ($\beta\to1$) reads

$$W_{ik}^{\text{new}} = W_{ik}\left[\frac{(XWB^T+X^TWB)_{ik}}{(WBW^TWB^T+WB^TW^TWB)_{ik}}\right]^{1/4}. \tag{4}$$

Such multiplicative learning rules have the essential advantage that matrix W always stays nonnegative. Note especially the role of the exponent η in the algorithms. The value given in the table guarantees that the objective function will be non-increasing in the

[1] Though equality without matrix transpose, namely $F^{(s)}=F^{(t)}$, is also possible, to our knowledge there are no corresponding real-world applications.