Accepted Manuscript

Updating/downdating the NonNegative Matrix Factorization

P. San Juan, A.M. Vidal, V.M. García-Mollá

PII: S0377-0427(16)30590-8
DOI: http://dx.doi.org/10.1016/j.cam.2016.11.048
Reference: CAM 10921

To appear in: *Journal of Computational and Applied Mathematics*

Received date: 20 June 2016
Revised date: 28 October 2016

Please cite this article as: P.S. Juan, A.M. Vidal, V.M. García-Mollá, Updating/downdating the NonNegative Matrix Factorization, *Journal of Computational and Applied Mathematics* (2016), http://dx.doi.org/10.1016/j.cam.2016.11.048

Updating/Downdating the NonNegative Matrix Factorization

P. San Juan[a,*], A.M. Vidal[a], V.M. García-Mollá[a]

[a] *Department of Information Systems and Computing, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, SPAIN.*

Abstract

The Non-Negative Matrix Factorization (NNMF) is a recent numerical tool that, given a non-negative data matrix, tries to obtain its factorization as the approximate product of two nonnegative matrices. Nowadays, this factorization is being used in many science fields; in some of these fields, real-time computation of the NNMF is required. In some scenarios, all data is not initially available and when new data (as new rows or columns) becomes available the NNMF must be recomputed. Recomputing the whole factorization every time is very costly and not suitable for real time applications. In this paper we propose several algorithms to update the NNMF factorization taking advantage of the previously computed factorizations, with similar error and lower computational cost.

Keywords: NNMF, updating, downdating

1. Introduction

The Non-Negative Matrix Factorization (NNMF) is a very popular tool in fields such as document clustering, data mining, machine learning, image analysis, audio source separation or bioinformatics [1, 2, 3, 4, 5]. The goal of the NNMF of a nonnegative data matrix $A \in \mathbb{R}^{m\times n}, (a_{i,j} \geq 0\ \forall i,j)$ is to obtain two nonnegative matrices $W \in \mathbb{R}^{m\times k}$ and $H \in \mathbb{R}^{k\times n}$ with $k \leq \min(m,n)$, such that $A \approx WH$. The problem can be described as the minimization of the Frobenius norm based target function: $\|WH - A\|_F$ subject to $W, H \geq 0$.

In certain science fields, the NNMF is computed minimizing other target functions, based on alpha-divergence, beta-divergence, Kullback-Liebler divergence, etc. [6, 7, 8, 9, 10]. However, in this paper we will use as target function the Frobenius norm.

*Corresponding author.
Email addresses: p.sanjuan@upv.es (P. San Juan), avidal@dsic.upv.es (A.M. Vidal), vmgarcia@dsic.upv.es (V.M. García-Mollá)