Neurocomputing 138 (2014) 238–247

Contents lists available at ScienceDirect

Neurocomputing

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

Recurrent neural network for approximate nonnegative matrix factorization

Giovanni Costantini [a], Renzo Perfetti [b,*], Massimiliano Todisco [a]

[a] Department of Electronic Engineering, University of Rome 'Tor Vergata', Italy
[b] Department of Electronic and Information Engineering, University of Perugia, Italy

ARTICLE INFO

Article history:
Received 11 July 2013
Received in revised form
30 October 2013
Accepted 2 February 2014
Communicated by L.C. Jain
Available online 17 February 2014

Keywords:
Recurrent neural networks
Lagrangian networks
Nonnegative matrix factorization
Features extraction
Clustering

ABSTRACT

A recurrent neural network solving the approximate nonnegative matrix factorization (NMF) problem is presented in this paper. The proposed network is based on the Lagrangian approach, and exploits a partial dual method in order to limit the number of dual variables. Sparsity constraints on basis or activation matrices are included by adding a weighted sum of constraint functions to the least squares reconstruction error. However, the corresponding Lagrange multipliers are computed by the network dynamics itself, avoiding empirical tuning or a validation process. It is proved that local solutions of the NMF optimization problem correspond to as many stable steady-state points of the network dynamics. The validity of the proposed approach is verified through several simulation examples concerning both synthetic and real-world datasets for feature extraction and clustering applications.

1. Introduction

The idea of using analogue circuits to solve mathematical programming problems can be traced back to the works of Pyne [1] and Dennis [2]. A canonical nonlinear programming circuit was proposed by Chua and Lin [3], later extended by Wilson [4]. Kennedy and Chua [5] recast the canonical circuit in a neural network framework and proved the stability. All the networks in [3–5] are based on the penalty function method, which gives exact solutions only if the penalty parameter tends to infinity, a condition impossible to meet in practice. To avoid the penalty functions, Zhang and Constantinides [6] proposed a Lagrangian approach to solve quadratic programming (QP) problems with equality constraints. The method can be extended to problems including both equality and inequality constraints converting inequalities into equalities by introducing slack variables. In addition, bound constraints on the variables, often arising in practical problems, can be treated in the same way at the expense of a huge number of variables. In the last decades several Lagrange neural networks have been proposed to solve specific optimization problems, handling both equality and inequality constraints as well as bounds on the variables [7–23].

Among the optimization problems of main interest in the context of machine learning and data, analysis there is nonnegative matrix factorization (NMF) [24]. The problem consists in finding reduced rank nonnegative factors to approximate a given nonnegative data matrix. This factorization can be interpreted as a representation of data using nonnegative basis vectors and nonnegative activation vectors. Like PCA, it can be used to accomplish the goal of reducing the number of variables required for data representation, with the additional constraint of non-negativity to enforce an additive, not subtractive, combination of parts. The idea of NMF can be traced back to Paatero and Tapper [25]. However, they were the seminal papers of Lee and Seung [26,27] which attracted the interest of many researchers. Applications of NMF have been proposed in diverse fields, e.g. text mining [28], document clustering [29,30], image reconstruction [31], human action recognition [32], discovering muscle synergies [33], EEG classification [34] and music transcription [35,36]. The relation between NMF and some clustering techniques has been proven [37,38], and several extensions and variants have been proposed in the literature [39–42].

Different algorithms can be used to solve the NMF problem. In particular, the most known are the multiplicative rules [26,27,42], and projected alternating least squares (ALS) algorithms [39]. With respect to other dimensionality reduction methods, probably the most intriguing feature of NMF is the capacity of finding the underlying parts-based structure of complex data. However, there is no explicit guarantee in the method to support this property,

* Corresponding author. Tel.: +39 0755853631.
E-mail addresses: costanti@uniroma2.it (G. Costantini), renzo.perfetti@unipg.it (R. Perfetti), massimiliano.todisco@uniroma2.it (M. Todisco).

http://dx.doi.org/10.1016/j.neucom.2014.02.007
0925-2312

which can be enforced introducing sparseness constraints as proposed by Hoyer [43] and Pascual-Montano et al. [44]. Due to the nonnegativity constraints, sparsity is strictly related to orthogonality among the basis vectors. Vice versa, imposing sparsity on the activation vectors, we can enforce an holistic representation of the data.

In the present paper we propose a neural network solver for the approximate NMF problem. It is a Lagrange programming neural network, using a projection operator to implement the nonnegativity constraints. A similar network has been proposed by the authors to solve convex optimization problems [14,17]. In this paper it is shown how this approach can properly work in a non-convex problem as the approximate NMF.

The rest of this paper is organized as follows. In Section 2, the NMF optimization problem is formulated. In Section 3, the proposed neural network is introduced and illustrated. Section 4 we investigate the network's dynamic behaviour. Section 5 presents the simulation results. Finally, some comments conclude the paper.

2. NMF optimization problem

Let $\Re_+$ denote the set of nonnegative real numbers. Given a nonnegative matrix $\mathbf{V}\in\Re_+^{m\times n}$ and an integer $p<\min(m,n)$, the NMF problem consists in computing a reduced rank approximation of $\mathbf{V}$ given by the product $\mathbf{WH}$ of nonnegative matrices $\mathbf{W}\in\Re_+^{m\times p}$ and $\mathbf{H}\in\Re_+^{p\times n}$. This problem can be formulated as the minimization of the objective function $J(\mathbf{W},\mathbf{H})=||\mathbf{WH}-\mathbf{V}||^2$ with non-negativity constraints on $\mathbf{W}$ and $\mathbf{H}$. The NMF optimization problem is not convex, so it admits multiple local minima and the solution found by iterative algorithms depends on initialization. Moreover the problem is characterized by an intrinsic invariance, since the product $\mathbf{WH}$ is unchanged by replacing matrices $\mathbf{W}$ and $\mathbf{H}$ by the nonnegative matrices $\mathbf{WD}$ and $\mathbf{D}^{-1}\mathbf{H}$, where $\mathbf{D}$ is any invertible nonnegative matrix; this implies the non-existence of isolated local minima of the objective function.

The problem formulation is often extended to include auxiliary constraints on $\mathbf{W}$ and/or $\mathbf{H}$, in order to avoid the invariance problem, limit the number of local minima and enforce some desired characteristics of the solution. Sparsity of $\mathbf{W}$ is sometimes required to enforce a parts-based decomposition [24,39,43,44]; sparsity of $\mathbf{H}$ is required to improve the performance in clustering applications. It has been shown that imposing L_1 normalization on rows or columns is a straightforward way to enforce sparsity; L_1 normalization of nonnegative vectors simply requires a constraint on the sum of elements. In this paper we take into account NMF with the following additional constraints: L_1 normalization of columns of $\mathbf{W}$; L_1 normalization of rows of $\mathbf{H}$.

The NMF optimization problem, with L_1 normalization of $\mathbf{W}$ columns, can be stated as follows:

minimize

$$J(\mathbf{W},\mathbf{H})=||\mathbf{WH}-\mathbf{V}||^2 \tag{1a}$$

such that

$$\mathbf{W}\geq 0 \tag{1b}$$

$$\mathbf{H}\geq 0 \tag{1c}$$

$$||\mathbf{w}_j||_1=\sum_{i=1}^{m} w_{ij}=1,\quad j=1,\ldots,p \tag{1d}$$

where $\mathbf{w}_j\in\Re_+^m$ denotes the jth column of $\mathbf{W}$.

The Lagrangian function corresponding to problem (1) is [45]:

$$\mathcal{L}=J(\mathbf{W},\mathbf{H})+\sum_{j=1}^{p}\alpha_j\left(\sum_{i=1}^{m}w_{ij}-1\right)-\sum_{i=1}^{m}\sum_{j=1}^{p}\lambda_{ij}w_{ij}-\sum_{j=1}^{p}\sum_{k=1}^{n}\mu_{jk}h_{jk} \tag{2}$$

where λ_{ij} and μ_{jk} are the Lagrange multipliers corresponding to inequality constraints (1b) and (1c), respectively; α_j is the Lagrange multiplier of the jth equality constraint (1d).

The Karush–Khun–Tucker (KKT) first order conditions for the existence of a local minimizer of problem (1) are the following [45]:

$$\frac{\partial\mathcal{L}}{\partial w_{ij}}=\frac{\partial J}{\partial w_{ij}}+\alpha_j-\lambda_{ij}=0 \tag{3a}$$

$$\frac{\partial\mathcal{L}}{\partial h_{jk}}=\frac{\partial J}{\partial h_{jk}}-\mu_{jk}=0 \tag{3b}$$

$$\lambda_{ij}\geq 0 \tag{3c}$$

$$\mu_{jk}\geq 0 \tag{3d}$$

$$\lambda_{ij}w_{ij}=0 \tag{3e}$$

$$\mu_{jk}h_{jk}=0 \tag{3f}$$

$$w_{ij}\geq 0 \tag{3g}$$

$$h_{jk}\geq 0 \tag{3h}$$

$$\sum_{i=1}^{m}w_{ij}-1=0 \tag{3i}$$

In relations (3) we assume $i=1,\ldots,m$, $j=1,\ldots,p$, and $k=1,\ldots,n$.

Since the objective function (1a) is non-convex, KKT conditions (3) are only necessary [45].

3. Neural network model

For a convex constrained optimization problem, Lagrangian duality can be used to obtain the global solution [6,45]. The basic idea is to find the saddle point of the Lagrangian function, which is maximized with respect to the Lagrange multipliers (dual variables) and minimized with respect to the primal variables. Here, we propose the same strategy to find a (local) solution of non-convex problem (1). To limit the number of variables, we adopt a *partial dual* approach introducing the following *reduced* Lagrangian function:

$$L(\mathbf{W},\mathbf{H},\boldsymbol{\alpha})=J(\mathbf{W},\mathbf{H})+\sum_{j=1}^{p}\alpha_j\left(\sum_{i=1}^{m}w_{ij}-1\right) \tag{4}$$

where $\boldsymbol{\alpha}=[\alpha_1\ \ldots\ \alpha_p]^{\mathrm{T}}$ is the vector of Lagrange multipliers (dual variables) corresponding to the equality constraints (1d). Constraints (1b) and (1c) are not included in (4), avoiding $p(m+n)$ additional dual variables. To fulfill constraints (1b) and (1c), avoiding the drawbacks of the penalty function approach, we introduce the auxiliary variables $\omega_{ij},\eta_{ij}\in\Re$, being $w_{ij}=P(\omega_{ij})$, $h_{jk}=P(\eta_{jk})$ and $P(.)$ is the piecewise linear function defined as follows (Fig. 1):

$$P(x)=\begin{cases}0 & \text{if}\quad x<0\\ x & \text{if}\quad x\geq 0\end{cases} \tag{5}$$

Function (5) is a projection operator: the auxiliary variables can vary in $\Re$ according to the gradient of the Lagrangian function (4) while the true variables w_{ij}, h_{jk} are confined in $\Re_+$.

To find a saddle point of the Lagrangian function (4) a dynamical system can be used such that, along a trajectory, function L is