Pattern Recognition Letters 80 (2016) 172–178

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Toward semantic attributes in dictionary learning and non-negative matrix factorization☆

Mohammadreza Babaee*, Thomas Wolf, Gerhard Rigoll

Institute for Human-Machine Communication, Technische Universität München, Munich, Germany

A R T I C L E I N F O

Article history:
Available online 9 July 2016

Keywords:
Dictionary
Matrix factorization
Attributes
Learning

A B S T R A C T

Binary label information is widely used semantic information in discriminative dictionary learning and non-negative matrix factorization. A Discriminative Dictionary Learning (DDL) algorithm uses the label of some data samples to enhance the discriminative property of sparse signals. A discriminative Non-negative Matrix Factorization (NMF) utilizes label information in learning discriminative bases. All these technique are using binary label information as semantic information. In contrast to such binary attributes or labels, relative attributes contain richer semantic information where the data is annotated with the strength of the attributes. In this paper, we utilize the relative attributes of training data in non-negative matrix factorization and dictionary learning. Precisely, we learn rank functions (one for each predefined attribute) to rank the images based on predefined semantic attributes. The strength of each attribute in a data sample is used as semantic information. To assess the quality of the obtained signals, we apply k-means clustering and measure the performance for clustering. Experimental results conducted on three datasets, namely PubFig (16), OSR (24) and Shoes (15) confirm that the proposed approach outperforms the state-of-the-art discriminative algorithms.

1. Introduction

Image content representation plays a key role in computer vision and pattern recognition. The idea is to transform an image from its original representation into a new representation suitable for a desired task (e.g. classification). Modern techniques such as Bag-of-Words models of local features (e.g., SIFT [21], Weber [8], Gabor [19]) represent an image by a very high-dimensional feature vector. Although this representation leads to relatively high accuracy in visual recognition and search, it increases the computation time and, consequently, is improper for real-time applications. Therefore, developing new algorithms that generate a compact and informative representation of image content is highly needed. Perhaps, the most common way to tackle this problem is learning a subspace of the original feature space and using this representation for the recognition tasks.

For several years, the representation of images with visual attributes has been studied intensively by researchers in the fields of clustering, classification, object recognition, and face verification. Farhadi et al. [10] proposed a shift from naming images to describing images. Instead of a naming an animal a "dog" it can be described as a "spotty dog". This means a shift from traditional approaches, where each instance was labeled with one label, to a model with more semantic information. This information can be crucial to model and learn inter- and intraclass relations. Kumar et al. [17] have developed an attribute classifier which focuses on the similarity regions in an image, associates classed depending on them. In Silberer et al. [28] images were described by attributes of 8 different categories, such as *shape_size*, *color_patterns* and *structure*. Generally, these attributes are observable semantic cues, which can be learned from low-level features. For example, "smiling" and "dry" can be considered as attributes of a face or a scene, respectively. Recently, it has been proposed that *relative attributes* provide a richer source of semantic information in images [14,25] than binary attributes. They depict the strength of attributes in an image and can be predicted by prelearned rank functions. For each attribute, a single rank function, which is a rank-SVM, is learned from a set of training data [25]. In this work, we use predicted relative attributes, as discriminant constraints to guide a NMF to generate a new subspace of images. More precisely, the relative attributes are embedded in a regularizer coupled with the NMF objective function. We call our proposed method Attribute constrained NMF (ANMF).

Also we present a Dictionary Learning approach, utilizing relative attributes to find a discriminative sparse representation for

☆ This paper has been recommended for acceptance by G. Moser.

* Corresponding author.
E-mail address: reza.babaee@tum.de (M. Babaee).

http://dx.doi.org/10.1016/j.patrec.2016.06.020

images. In Dictionary Learning we consider a set of n input signals $\mathbf{Y} \in \mathbb{R}^{p\times n}$ and the goal is to find a dictionary $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_k] \in \mathbb{R}^{p\times k}$ and sparse representations $\mathbf{X} \in \mathbb{R}^{k\times n}$ such that $\mathbf{Y} \cong \mathbf{DX}$, where the term *over-complete* indicates $k > n$. Dictionaries can either be predefined as in the form of wavelets [23], or be learned from observations [1,7,32]. Additionally, many approaches have been developed to impose discriminative capabilities onto the dictionary learning process. Those methods often use binary label information to acquire discriminative behavior. In this work, we present an approach that utilizes relative attributes instead of binary labels to enhance the discriminative property of the dictionary. Just as previous discriminative dictionary learning approaches use binary label information to enhance their discriminative capabilities, we incorporate relative attributes into the dictionary learning process as semantic information.

The rest of the paper is organized as follows. In Section 2 related work in the field of dictionary learning and non-negative matrix factorization is presented. Section 3 gives an in-depth explanation of the problem solved by the dictionary learning approach. In Sections 4 and 5 the details for the dictionary learning and non-matrix factorization algorithms are given. Afterwards in Section 6 the concluded experiments are described together with the obtained results. The report concludes with a discussion and summary in Section 8.

2. Related work

The first approaches in the field of reconstructive dictionary learning are the K-SVD algorithm [1] and the Method of Optimal Direction (MOD) [9], where no semantic information is used in the learning process. An additional example for the usage of sparse representation is the Sparse Representation based Classification (SRC) [29] where the dictionary is built directly from the training data.

Another large field in dictionary learning is called Discriminative Dictionary Learning (DDL), where either the discriminative property of the signal reconstruction residual, or the discriminative property of the sparse representation itself is enhanced. Approaches with a focus on the reconstruction residual are the work of Ramirez et al. [26], which includes a structured incoherence term to find independent sub-directories for each class, and the work of Gao et al. [11], where sub-dictionaries for the different classes are learned as well as a shared dictionary over all classes.

Methods aiming at finding discriminative coding vectors learn the dictionary and a classifier simultaneously. In the work of Zhang et al. [32], the K-SVD algorithm is extended by a linear classifier. Jiang et al. [13] included an additional discriminative regularizer to come up with the so called Label Consistent K-SVD (LC-KSVD) algorithm. Both of these algorithms show good results for classification and face recognition tasks. The approach of Yang et al. [31] combines the two types of DDL by taking the discriminative capabilities of the reconstruction residual and the sparse representation into account. Therefore, class specific sub-dictionaries are learned while maintaining discriminative coding vectors by applying the Fisher discrimination criterion. In the recent work of Cai et al. [7] a method called Support Vector Guided Dictionary Learning (SVGDL) is presented, where the discrimination term consists of a weighted summation over squared distances between the pairs of coding vectors. The algorithm automatically assigns non-zero weights to critical vector pairs (the support vectors) leading to a generalized good performance in pattern recognition tasks.

Inspired by the part-based perception behavior of the human brain (i.e., combining the perceptions of an object to perceive it as a whole) non-negative Matrix Factorization (NMF) is a widely used matrix factorization method [6,18,20]. A parts-based representation can be achieved by applying a non-negativity constraint to the matrix factors, only allowing additive combinations of original data.

In order to find a robust data representation, further methods are needed besides the non-negativity constraint [20]. One approach is to focus on preserving the intrinsic geometry of the data space by defining new objective functions. Cai [6] proposed the GNMF, constructing a nearest neighbor graph and encoding the geometrical information of the data space, and therefore considering the local invariance. Another approach is the CNMF, introduced by Liu and Wu [30] who constrain the NMF to only use the prior annotation of the data, enforcing similar encoding for points from the same class. In the work of Gu and Zhou [12] local linear embedding assumptions are used to propose the so called NPNMF. A new constraint was presented allowing each data point to be presented by its neighbors.

3. Background

For the general problem formulation we assume $\mathbf{Y} = [\mathbf{y_1}, \mathbf{y_2}, \ldots \mathbf{y_n}]$ to be the set of p-dimensional input signals, each belonging to one of C (hidden) classes, $\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}]$ to be their corresponding k-dimensional sparse representation and $\mathbf{D} \in \mathbb{R}^{n\times k}$ to be the dictionary. As a consequence, the standard dictionary learning method is defined by

$$\langle \mathbf{D}, \mathbf{X} \rangle = \arg\min_{\mathbf{D},\mathbf{X}} \|\mathbf{Y} - \mathbf{DX}\|_2^2 + \lambda_1 \|\mathbf{X}\|_1, \tag{1}$$

with the regularization parameter λ_1. In order to take the relative attributes into account the objective function has to be extended with an additional term $\mathcal{L}(\mathbf{X})$.

$$\langle \mathbf{D}, \mathbf{X} \rangle = \arg\min_{\mathbf{D},\mathbf{X}} \|\mathbf{Y} - \mathbf{DX}\|_2^2 + \lambda_1 \|\mathbf{X}\|_1 + \lambda_2 \mathcal{L}(\mathbf{X}) \tag{2}$$

As additional information, the strength of M predefined attributes, the so called relative attributes [25], for the input signals are available.

3.1. Relative attributes

The idea in learning relative attributes, assuming there are M attributes $\mathcal{A} = \{a_m\}$, is to learn M ranking functions $\mathbf{w_m}$ for $m = 1..$M. Therefore, the predicted relative attributes are computed by

$$r_m(\mathbf{x_i}) = \mathbf{w_m}^\intercal \mathbf{x_i}, \tag{3}$$

such that the maximum number of the following constraints is satisfied:

$$\forall (i, j) \in \mathcal{O}_m : \mathbf{w_m}^\intercal \mathbf{x_i} > \mathbf{w_m}^\intercal \mathbf{x_j}, \tag{4}$$

$$\forall (i, j) \in \mathcal{S}_m : \mathbf{w_m}^\intercal \mathbf{x_i} \approx \mathbf{w_m}^\intercal \mathbf{x_j} \tag{5}$$

whereby $\mathcal{O}_m = \{(i, j)\}$ is a set of ordered signal pairs with signal i having a stronger presence of attribute a_m than signal j and $\mathcal{S}_m = \{(i, j)\}$ being a set of un-ordered pairs where signal i and j have about the same presence of attribute a_m. It is possible to approximate this objective with the introduction of non-negative slack variables, similar to an SVM classifier:

$$\min\left(\frac{1}{2}\|\mathbf{w_m}^\intercal\| + c\left(\sum \xi_{ij} + \sum \gamma_{ij}\right)\right) \tag{6}$$

$$\text{s.t.}\mathbf{w_m}^\intercal(\mathbf{x_i} - \mathbf{x_j}) \geq 1 - \xi_{ij}; \quad \forall (i, j) \in \mathcal{O}_m \tag{7}$$

$$|\mathbf{w_m}^\intercal(\mathbf{x_i} - \mathbf{x_j})| \leq \gamma_{ij}; \quad \forall (i, j) \in \mathcal{S}_m \tag{8}$$

The work of Parikh et al. [25] provides us with a convenient *RankSVM* function that returns the ranking vector $\mathbf{w_m}$ for a set of input images and their relative ordering. This information can further be used in the objective function in Eq. (2).