Neurocomputing 124 (2014) 22–32

Contents lists available at ScienceDirect

Neurocomputing

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

Sparse coding of human motion trajectories with non-negative matrix factorization

Christian Vollmer [a,b,*], Sven Hellbach [a,b], Julian Eggert [b], Horst-Michael Gross [a]

[a] Ilmenau University of Technology, Neuroinformatics and Cognitive Robotics Lab, Helmholtz-Platz 5, 98693 Ilmenau, Germany
[b] Honda Research Institute Europe GmbH Carl-Legien-Str. 30, 63073 Offenbach/Main, Germany

ARTICLE INFO

Article history:
Received 16 September 2011
Received in revised form
10 October 2012
Accepted 3 December 2012
Available online 4 April 2013

Keywords:
Non-negative matrix factorization
Time series
Human motion
Sparse coding
Prediction

ABSTRACT

We use shift-invariant Non-negative Matrix Factorization (NMF) for decomposing continuous-valued time series into a number of characteristic primitives, i.e. the basis vectors, and their activations, which results in a model-independent and fully data driven parts-based representation. We interpret the basis vectors as short parts of motion that are shared between all trajectories in the data set, and the activations as onset times of those parts. The extension of the shift-invariant NMF by a new competition term between adjacent activations allows to gain temporally isolated activation events, which further supports this interpretation. We show that the resulting sparse and compact representation can be used for the prediction of motion trajectories, and that it can be beneficial for classification, because it allows the application of simple standard classification models with few parameters. In this paper we show that basis vectors can be extracted, which can be interpreted as short motion segments. We present results on trajectory prediction, and show that the sparse representation can be used for classification of trajectories of a single joint, like the one of a hand, obtained by motion capturing.

1. Introduction

The understanding and interpretation of movement trajectories is a crucial step for the analysis of dynamic visual scenes with moving items. For example, consider the simple task for a robot of grasping an object which is handed over by the human interaction partner. Most approaches for describing motion patterns, like [1], rely on a kinematic model for the observed human motion. Using a particular model makes it difficult to adapt the approach to new tasks including other types of motion. Here, we aim at a generic, model-independent framework for decomposition, classification, and prediction. In this paper, we present an approach which finds prototypical segments of point trajectories that, in sequential combination, model the entire trajectory. We demonstrate our approach using one-dimensional components of trajectories of a single joints of a human, like the hand or foot, as shown in Fig. 1.

Non-negative Matrix Factorization (NMF) [2] is a blind source separation approach in concept similar to PCA and ICA that provides an efficient parts-based factorization for decomposing data under non-negativity constraints. We use an extension of NMF with additional sparsity constraints as presented in [3] for decomposing a time series. The time series is decomposed into basis vectors and activations where the basis vectors can be interpreted as parts of motions that are shared amongst all trajectories in the data set, and thus, constitute a set of primitive motions. This yields a kind of representation, which is sometimes called *piano model* [4], where the sequence of activations can be understood in analogy to the keys of a piano, pressed by a piano player over time. Each pressed piano key then triggers a characteristic sound wave, the analog to our primitive, and the superposition of the primitives generates the melody, i.e. the analog to the time series.

A different interpretation, often used in the neural sparse coding literature, is to view the activations as spike-trains that mark the response of some kind of receptor. Both interpretations emphasize the fact that given the basis functions, we can transform the data into a more compact representation, where only the amplitudes and locations of the activations, or spikes, have to be stored to characterize a trajectory. Further, this representation is shift-invariant in the sense that the signal is characterized by the relative positions of the events. This property has been studied in early work on sparse coding of temporal signals. E.g. in [5], the authors aim at computing a sparse representation of natural audio signals in form of spike trains, where the spikes mark activations of a fixed and hand-crafted over-complete set of basis functions. Given this set of basis functions, the amplitude and timing of the activations of those basis primitives are learned. The authors argue

* Corresponding author at: Ilmenau University of Technology, Neuroinformatics and Cognitive Robotics Lab, Helmholtz-Platz 5, 98693 Ilmenau, Germany. Tel.: +49 3677 694178.
E-mail addresses: christian.vollmer@tu-ilmenau.de (C. Vollmer), hellbach@htw-dresden.de (S. Hellbach), julian.eggert@honda-ri.de (J. Eggert), horst-michael.gross@tu-ilmenau.de (H.-M. Gross).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.12.054

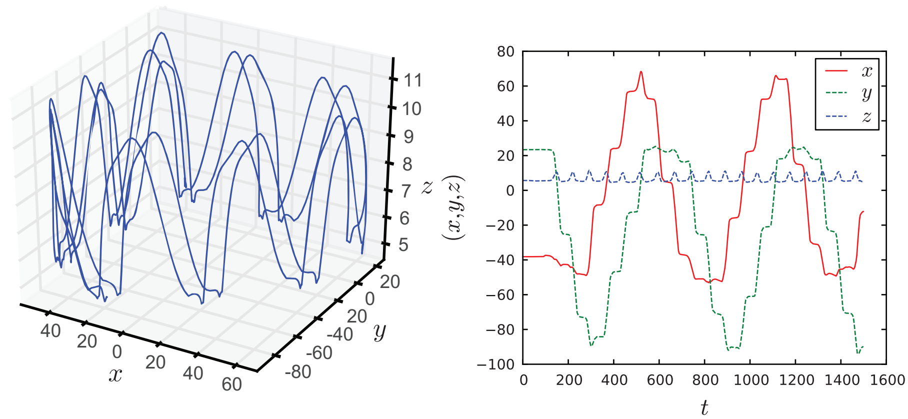

Fig. 1. Example of the kind of trajectories analyzed in this paper. Shown is the trajectory of the right foot of a human walking in a circle, recorded using a motion capturing device with a sampling rate of 60 Hz. Left: Trajectory in 3D space. Right: One-dimensional projections over time. For our decomposition, we approximate the problem by regarding all dimensions as independent. For the sake of clarity, in the following we concentrate on a single dimension (of the spatial dimensions x, y and z shown in the right graph) (the formal extension to multiple dimensions is straightforward and will be explored more in depth elsewhere).

that such a representation provides a very efficient encoding and uncovers the underlying event-like structure of the signal. This work has been extended (e.g. in [6]) to also learn the basis functions to find an optimal dictionary or code of the signal. The authors show that the emerging basis functions can be compared to auditory receptors in animals and thus are naturally interpretable. Such methods target to achieve local sparsity (the activation events should be sparse and occur isolated in time); for this purpose, the authors use sequential selection heuristics like Matching Pursuit, where the subset of the activations is selected one after another by correlation and thresholding. A similar kind of decomposition has also been done for music in [7] to find a shift-invariant and sparse representation and to uncover latent structure in the data. Here the authors also use a (slightly different) heuristic to select the coefficients with the largest magnitude gradient, and get isolated and well-localized activities. We show in this paper, however, that spatio-temporally isolated activities can also be achieved without selection heuristics, but instead by directly formulating a penalty for adjacent activities and including this penalty as an additional energy-term for the basis vector decomposition model. During optimization, the penalty term leads naturally to a competition between rivaling activities, eliminating spurious activity traces which are detrimental for further trajectory processing, as e.g. for classification. Further work that uses similar kinds of shift-invariant decomposition models include [8–10]. For recent publications, see [11,12].

A different formulation of such a model that, however, results in similar representations and properties but at higher computational costs has been proposed in [13]. In that work, the motion primitives are described by a strict left-to-right Hidden Markov Model (HMM). The probability of triggering a primitive at time t is described by an onset probability, very analog to the activation times in our approach (see Section 4.6). The primitives are then superimposed by means of a factorial Hidden Markov Model (fHMM), where each factor is a primitive HMM as described above. In [13] it was observed that the triggering probabilities are characteristic for a class in the data and provide a compact representation as a 'timing code'. Although very similar in concept, here we present a different method that allows the application of the efficient multiplicative learning and decomposition algorithms as gained from NMF. The resulting activations have very similar characteristics to the HMM approaches and can thus also be interpreted as a timing code of the trajectories.

The paper is organized as follows. In Section 2, we introduce the basics of the shift-invariant NMF-based decomposition, outline the necessary extensions of the NMF algorithm for the application on motion trajectories, and describe a penalty term that allows to improve the representational properties of the decomposition by getting decomposition results which have a smaller number of more isolated activity events than a straightforward NMF decomposition. In Section 3 we describe in more detail the preprocessing as well as the training and application phases for the decomposition of motion trajectories. We also explain how to use the algorithm continuously for online decomposition of a trajectory, and for trajectory prediction. In Section 4, we show the effects of the additional penalty term, typical results of the gained basis vectors, an evaluation of the predictive capabilities of our approach, and how classification can be done based on the sparse sets of activations.

2. Non-negative matrix factorization

Similar to other decomposition approaches like, e.g. PCA and ICA, Non-negative Matrix Factorization (NMF) [2] can be used to solve the source separation problem, where a set of training data $\mathbf{V}$ has to be decomposed into basis vectors $\mathbf{W}\geq 0$ and activations $\mathbf{H}\geq 0$

$$\mathbf{V}\approx\mathbf{W}\cdot\mathbf{H}$$

Each training data sample is represented as a column vector $\mathbf{V}_i$ of the matrix $\mathbf{V}$. Each column $\mathbf{W}_j$ of the matrix $\mathbf{W}$ is a basis vector. In the activation matrix $\mathbf{H}$ the element $H_i^j\geq 0$ determines to which extent each basis vector $\mathbf{W}_j$ is activated to reconstruct a training sample $\mathbf{V}_i$.

Unlike PCA or ICA, however, NMF aims at a decomposition which only consists of non-negative elements. Thus, superposition of basis vectors always means accumulation of positive parts. There exists no primitive which is able to erase a 'wrong' superposition of other primitives, and different primitives cannot cancel out each other. This leads to basis vectors with different properties: They are often spatially more localized (as e.g. for PCA) and it has been argued that this supports parts-based representation schemes and has advantages for certain applications, like reported e.g. for face recognition [14].

For calculating the decomposition, optimization-based methods are used. For this purpose, an energy function E is used, which targets a good reconstruction of the data. Within the energy function, additional desired constraints can be formulated as it is shown for an activation sparsity constraint in [15] and for transformation invariance of the basis vectors in [16]. This leads