Neurocomputing 150 (2015) 155–162

Contents lists available at ScienceDirect

Neurocomputing

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

An information geometric framework for the optimization on a discrete probability spaces: Application to human trajectory classification ☆

Jacinto C. Nascimento [a,*], Miguel Barão [b], Jorge S. Marques [a], João M. Lemos [c]

[a] ISR, Instituto Superior Técnico, Universidade de Lisboa, Portugal
[b] INESC-ID and Universidade de Évora, Portugal
[c] INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, Portugal

ARTICLE INFO

Article history:
Received 17 January 2014
Received in revised form
4 July 2014
Accepted 12 August 2014
Available online 8 October 2014

Keywords:
Fisher information metric
HMM
Human activity recognition
Natural gradient
Riemannian manifold
Surveillance

ABSTRACT

This paper presents an iterative algorithm using a information geometric framework to perform the optimization on a discrete probability spaces. In the proposed methodology, the probabilities are considered as points in a statistical manifold. This differs greatly regarding the traditional approaches in which the probabilities lie on a simplex mesh constraint. We present an application for estimating the switching probabilities in a space-variant HMM to perform human activity recognition from trajectories; a core contribution in this paper. More specifically, the HMM is equipped with a space-variant vector fields that are not constant but depending on the objects's localization. To achieve this, we apply the iterative optimization of switching probabilities based on the natural gradient vector, with respect to the Fisher information metric. Experiments on synthetic and realworld problems, focused on human activity recognition in long-range surveillance settings show that the proposed methodology compares favorably with the state-of-the-art.

1. Introduction

This paper presents a natural gradient method applied to the optimization on discrete (finite) probability spaces. Traditionally, this problem is tackled in a simplex probability constraint where standard gradient based methods are used. In this paper, we show that performing the optimization in a Riemannian space equipped with the Fisher metric provides several advantages over the standard methods. One of the advantages is that the Fisher metric is able to smoothly modify the gradient direction, so that it flows within the feasible region, i.e. the parameter space that satisfies all probability constraints. If some constraints become active, then the method behaves as the gradient projection method. Also, the Fisher metric exhibits a fast convergence since it behaves asymptotically as a Newton method. From the above, it will be shown that a formally correct interpretation of a natural gradient as the steepest-descent method is verified. With this approach the computational requirements are minimal: only marginally larger than the standard gradient; constant in time and space; and using rudimentary operations, i.e. additions and multiplications.

The novel contribution proposed in this paper is the application of the above framework in a new context: classification of human activities in far-field surveillance settings using the trajectories performed by pedestrians. Indeed, in the so-called far field scenarios, people are far from the camera, making it impossible to obtain detailed shape information and the system has to extract trajectories, or a rough shape description e.g., a bounding box or a coarse silhouette [7]. Models of typical trajectories may be estimated from training sets and then used to classify observed trajectories. This is one traditional problem that arises in outdoor surveillance systems and it will a focus of this paper. To model human trajectories in video sequences, we use a generative model of non-parametric vector fields proposed in [2]. The framework in [2] models the trajectories using a small set of vector or motion fields, estimated from observed trajectories. An advantage of using this model resides in its flexibility of modeling pedestrian's trajectories. More specifically, the trajectory is split into a sequence of segments, each of which generated by one vector field. Switching between models can occur at any point in the image but with

☆This work was supported by FCT Project PEst-OE/EEI/LA0009/2013, by projects "ARGUS"-PTDC/EEA-CRO/098550/2008, "PROBCONTROL"- PTDC/EEA-CRO/115038/2009 and by INESC-ID funding through PEst-OE/EEI/LA0021/2013.

* Corresponding author.

http://dx.doi.org/10.1016/j.neucom.2014.08.074
0925-2312/

probability that may depend on the spatial location. This provides a flexible tool to represent a wide variety of motion patterns. We present an expectation-maximization (EM) algorithm to learn the proposed model from sets of observed trajectories. The difference regarding the work in [2] is that here, the switching probabilities are estimated using the natural gradient instead of a projection simplex approach. These two methods, for computing the gradient, will be compared and the effectiveness of using the natural gradient will be illustrated.

2. Related work in human activity

Recognizing human activities in a quite diverse range of contexts and scenarios remains an up-to-date topic in image processing and computer vision communities. The goal is usually to interpret or classify human activities using tracked features. Related research in human activity analysis can be used in a wide variety of fields, such as intelligent environments [13], human machine interaction [14], surveillance [15,16], human computer interaction and sports analysis [17,18], to quote a few. Most of the work in this area falls into one of the two different settings, depending on which, the camera is close or far away to the person. In short range (SR) settings, the camera is close to the observed people, thus detailed information of human gestures, pose, gait can be computed. In long range or far field (FF) settings, the camera covers a wide area, thus no longer able to acquire a detailed type of information. Although, a large fraction of the related work on human activity recognition has been devoted to SR setup, we concentrate more to describe related work proposed in FF scenarios, that is the focus of the application presented herein, i.e. people are far from the camera and the trajectories are used as the information to perform classification/recognition.

In FF scenarios, it is usually impossible to obtain detailed descriptions of the observed persons, thus most methods rely only on the use of trajectories, taking for instance the center of the bounding box extracted by some region detection algorithm. Several trajectory analysis problems such as the one addressed herein, i.e. classification, have been addressed using pairwise similarity or dissimilarity measures between trajectories; these include Euclidean [4] and Hausdorff distances [5]. Because trajectories may have different lengths, techniques to face trajectories alignment have also been proposed. For instance, the use of dynamic time warping [6] or longest common subsequence [19] have been suggested to perform such comparisons. The class of approaches adopted in this paper models the trajectories as being produced by a probabilistic generative mechanism, usually an HMM or one of its variants [8–12]. These approaches have the key advantage of not requiring trajectory alignment or registration; moreover, they allow building a well grounded probabilistic inference formulation, based on which model parameters may be obtained from observed data. In that same class of approaches, in [20] the authors proposed a set of behavioral maps based on Markovian trajectory models, however, their application context is orthogonal to ours, since their goal is to improve tracking results by reconstructing full trajectories from fragments thereof.

All these contributions have been reported in [1], but in this paper, we provide a more comprehensive literature review, explanations and experimental results.

The paper is organized as follows. Section 3 describes the natural gradient proposed herein. Section 4 presents the generative model from which trajectory classification is performed. Section 5 describes how the generative model is learned with the EM algorithm using the natural gradient. Section 6 presents simulation results highlighting the superiority of the natural gradient and provides results using the proposed framework for human activity classification. Section 7 concludes the paper.

3. Discrete probability distributions in Riemannian space

This section provides detailed description of the proposed natural gradient. An usual premise is to assume that the probabilities lie on a simplex probability mesh. Contrasting with the above approach, we bring a new methodology, based on the information geometric framework [3,23,24], in which the probabilities are considered as points in a statistical manifold. We describe next how to parameterize the switching probabilities in the presented context.

One way to parameterize the probability mass function (p.m.f) of $p(x)$ defined over the set of p.m.f., $\mathcal{P}$, is to use the probabilities $\theta^k \overset{\text{def}}{=} Pr\{X=k\} = p(k)$, for $k=0,\dots,K$. Since the probabilities satisfy the *partition of unity*, the probability θ^0 is defined as $\theta^0 \overset{\text{def}}{=} 1-\sum_{k=1}^{K}\theta^k$. This parameterization defined above provides a global coordinate system of $\mathcal{P}$, where θ^k are the coordinates.

3.1. The Fisher metric

The set $\mathcal{P}$ can be seen as a manifold, where each member $p \in \mathcal{P}$ is a p.m.f. and as an associated tangent space $T_p(\mathcal{P})$. One suited metric than can be introduced on the tangent space [3,23,24] uses the Fisher information matrix for defining the inner product on the manifold. The Fisher information matrix $\mathbf{G}_\theta$ has its entries defined as follows:

$$g_{ij}(\theta) \overset{\text{def}}{=} \mathbb{E}\left[\frac{\partial \log\ p(x)}{\partial \theta^i}\frac{\partial \log\ p(x)}{\partial \theta^j}\right] \tag{1}$$

It can be straightforwardly seen [3] that the components of the Fisher information matrix are given by

$$g_{ij}(\theta) = \frac{1}{1-\sum_{k=1}^{K}\theta^k} + \frac{\delta_{ij}}{\theta^i} \tag{2}$$

where δ_{ij} is the Kronecker delta function ($\delta_{ij}=1$ if i=j, $\delta_{ij}=0$ otherwise) and the corresponding Fisher information matrix $\mathbf{G}_\theta$ is given by

$$\mathbf{G}_\theta = \mathbf{I}_\theta + \left(1-\sum_{k=1}^{K}\theta^k\right)^{-1} \mathbf{1}\times\mathbf{1}^\top \tag{3}$$

where $\mathbf{I}_\theta$ is the $K\times K$ matrix having a diagonal structure with the i-th diagonal entry equal to $(\theta^i)^{-1}$ and $\mathbf{1}$ is a $K\times 1$ unit vector.

3.2. The natural gradient and its relation with Euclidean spaces

When performing optimization on $\mathcal{P}$, it is necessary to define a cost function F defined over $p \in \mathcal{P}$, i.e. $F(p)$. Since we are given the parameterization introduced in Section 3.1, it is possible to define a new function in these parameters, that is F_θ. Then, we can iterate over these parameters, until the convergence is reached.

Here, the parameter iteration that gives the gradient of F_θ in a arbitrary v-direction is the vector ∇F_θ such that the following equality holds:

$$\langle \nabla F_\theta, v\rangle = dF_\theta(v), \quad \forall v \neq 0. \tag{4}$$

Comparing with the standard Euclidean space, the inner product $\langle\cdot,\cdot\rangle$ (i.e. the gradient with respect to the Euclidean metric) is simply the product, and (4) becomes

$$\nabla F_\theta = \left[\frac{\partial F_\theta}{\partial \theta^1},\dots,\frac{\partial F_\theta}{\partial \theta^K}\right]^\top \tag{5}$$