Neurocomputing 184 (2016) 221–231

Contents lists available at ScienceDirect

Neurocomputing

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

Collective motion pattern inference via Locally Consistent Latent Dirichlet Allocation

Jialing Zou [a], Qixiang Ye [a], Yanting Cui [a], Fang Wan [a], Kun Fu [b], Jianbin Jiao [a,*]

[a] University of Chinese Academy of Sciences, Beijing 101408, China
[b] Institute of Electronics, Chinese Academy of Sciences, China

ARTICLE INFO

Article history:
Received 20 January 2015
Received in revised form
1 August 2015
Accepted 4 August 2015
Communicated by Chennai Guest Editor
Available online 1 February 2016

Keywords:
Collective motion
Local motion pattern
Topic model
Trajectory clustering

ABSTRACT

Extracting motion descriptors in crowd videos is highly challenging due to scene clutter and serious occlusions. In this paper, Locally Consistent Latent Dirichlet Allocation (LC-LDA) model is proposed to learn collective motion patterns using tracklets and bag-of-words as low level features. The LC-LDA model implements a graph Laplacian operator to impose neighboring constraints to tracklets on a local manifold, which enforces the spatial–temporal coherence of tracklets in a high dimensional bag-of-word feature space. With initialization of clustering on a manifold, LC-LDA model improves the unsupervised inference capability and compactness of learned collective motion patterns. Experimental results on three public datasets indicate that LC-LDA based motion patterns can improve the trajectory clustering performance effectively.

1. Introduction

Extracting motion patterns or object feature descriptors in computer vision is of great value to activity recognition [1,2], event detection [3,4], face detection [5,6], image classification [7], texture segmentation [8] and scene understanding [9–13]. In crowded video scenes, however, collective motion pattern or feature descriptor extraction remains a challenging task due to frequent occlusions among objects. With existing tracking algorithms [2,14–17], short trajectory segments (tracklets) are usually obtained, from which it is difficult to extract meaningful motion patterns by performing simple trajectory classification [18] or clustering [19]. Zhu et al. [20] propose tracking crowded groups in surveillance videos based on the theory of coherent motions [21] in crowded scenes. They obtain crowded motion patches through a KLT keypoint tracker and then integrate and update the patches into groups. The information in the same level of patches and among hierarchical levels of patches is used to regularize the tracking results.

To "connect" tracklets into meaningful motion patterns, unsupervised inference methods have been widely explored. In [9], a hierarchical unsupervised framework with k-means and GMMs is proposed to learn motion patterns from dense optical flow features. In [11], a Tensor Voting algorithm is employed to calculate the local geometry structure of tracklets. In the local geometry structure space, Bayesian inference is later used to perform motion pattern inference. In [22], Chen et al. apply a deep model, Replicated Softmax-based model, to cluster tracklets in crowded scenes. In [4], Wang et al. propose using a mixture of Latent Dirichlet Allocation (LDA) models to learn collective motion patterns. Motion patterns are reflected by a hierarchical Dirichlet processes, where collective motions are modeled as distributions over tracklet features, and interactions are modeled as distributions over collective motion patterns. The approach mentioned above explore the advantages of statistical inference; however, few of them consider the spatial–temporal coherence among tracklets. In Fig. 1, the red path denotes collective motion patterns. Although it can be seen the red tracklets and the blue tracklets in two clusters, respectively, are in the same spatial–temporal space, a single tracklet (the blue one in Fig. 1) may denote the wrong motion pattern.

To enforce the spatial–temporal coherence of tracklets, scene or object motion priors are used. Zhou et al. [10], Alahi et al. [23] and Zou et al. [24,25] incorporate the scene priors into their frameworks, respectively. Their priors are scene sources and sinks, estimated from the starting and ending points of tracklets. In [26], a Mixture model of Dynamic pedestrian-Agents (MDA) is proposed to learn collective behavior patterns of pedestrians in crowded scenes. Each pedestrian in the crowd is driven by a dynamic pedestrian-agent with a motion prior, and the whole crowd is subsequently modeled as a mixture of dynamic pedestrian-agents. Once the model is learned, MDA can be used to simulate crowd motion patterns. Although the usage of scene or object motion

* Corresponding author.
E-mail address: jiaojb@ucas.ac.cn (J. Jiao).

http://dx.doi.org/10.1016/j.neucom.2015.08.108
0925-2312/

Fig. 1. Illustration of tracklets in same space sharing different motion patterns. (For interpretation of the references to color in this figure, the reader is referred to the web version of this paper.)

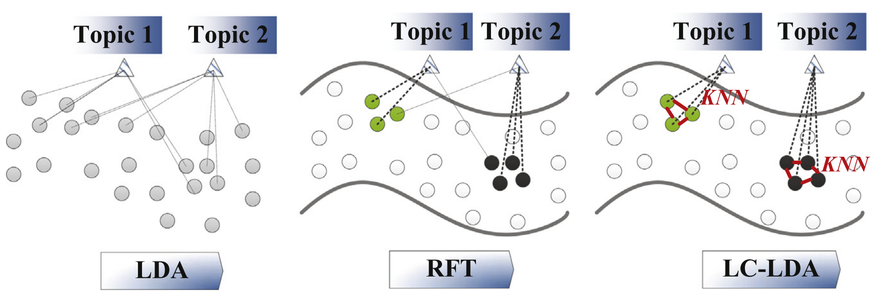

Fig. 2. Mapping among tracklets and topics of LDA, RFT, and LC-LDA. Points of different colors denote tracklets of different clusters. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

priors significantly improves the inference performance, it also reduces the robustness of the model when considering various scenes. In addition, when priors are introduced into the statistical inference framework, the discriminative power of the learned motion patterns may decrease [27], as the sample size increases.

In this paper, we propose a Locally Consistent LDA (LC-LDA) topic model for collective motion pattern inference using tracklets as low level features. The LC-LDA model carries out the cluster initialization on a manifold, updates correlation among topics, and coherence among tracklets with expectation-maximization iteration. We define a symmetrical KL divergence for the distance metric in a probability space. The metric is added to the objective function as a penalty term by mathematical approximation for the convergence of Variational Breaking Algorithm [28]. This term is used to impose neighbor constraints on tracklets and to enforce the spatial–temporal coherence of tracklets that belong to the same motion pattern. In the implementation of LC-LDA algorithm, the penalization term is approximated in the form of graph Laplacian, which is a discrete approximation of Laplace Beltrami operator [29].

Fig. 2 exhibits three types of LDA topic models for motion pattern inference. In the original LDA model, the input data has no class information, and the LDA model just considers the general likelihood. Therefore, topics (motion patterns) inferred by the LDA model are not discriminative. To obtain discriminative topics, a Random Field Topic (RFT) model [10] is used to introduce data class information by scene priors. However, class information cannot be transferred to topics in the model inference procedure. In [10], it has been proven that the usage of priors could be invalid to the improvement of the discriminative capability of learned topics. In the proposed LC-LDA, a Laplacian manifold embedding is used to capture data class information, which keeps the compactness of initial topics. In the model inference procedure, local consistence constraints are embedded to keep the discriminative capability of learned topics.

Based on LC-LDA topic model, we propose an unsupervised approach for learning collective motion patterns, as well as performing trajectory clustering, as shown in Fig. 3. Given a crowded video, tracklets are firstly computed. The tracklets are used to compute linking weights through a spanning tree algorithm [10] and are described with bag-of-word features. Low level motion features are fed to the LC-LDA model to infer motion patterns, on which we perform tracklet clustering using an entropy clustering algorithm [10].

The contributions of this paper are summarized as follows:

1. We propose a Locally Consistent LDA (LC-LDA) model to improve the inference performance of LDA by introducing locally consistent constraints.

2. We propose using a manifold embedding approach to initialize the LC-LDA model and using an iterative procedure to optimize it, which make it possible to learn collective motion patterns without any scene or object motion prior.