Neural Networks 71 (2015) 88–104

Contents lists available at ScienceDirect

Neural Networks

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

Incremental multi-class semi-supervised clustering regularized by Kalman filtering

Siamak Mehrkanoon *, Oscar Mauricio Agudelo, Johan A.K. Suykens

KU Leuven, ESAT-STADIUS, Kasteelpark Arenberg 10, B-3001 Leuven (Heverlee), Belgium

ARTICLE INFO

Article history:
Received 25 September 2014
Received in revised form 9 July 2015
Accepted 2 August 2015
Available online 14 August 2015

Keywords:
Incremental semi-supervised clustering
Non-stationary data
Video segmentation
Low embedding dimension
Kernel spectral clustering
Kalman filtering

ABSTRACT

This paper introduces an on-line semi-supervised learning algorithm formulated as a regularized kernel spectral clustering (KSC) approach. We consider the case where new data arrive sequentially but only a small fraction of it is labeled. The available labeled data act as prototypes and help to improve the performance of the algorithm to estimate the labels of the unlabeled data points. We adopt a recently proposed multi-class semi-supervised KSC based algorithm (MSS-KSC) and make it applicable for on-line data clustering. Given a few user-labeled data points the initial model is learned and then the class membership of the remaining data points in the current and subsequent time instants are estimated and propagated in an on-line fashion. The update of the memberships is carried out mainly using the out-of-sample extension property of the model. Initially the algorithm is tested on computer-generated data sets, then we show that video segmentation can be cast as a semi-supervised learning problem. Furthermore we show how the tracking capabilities of the Kalman filter can be used to provide the labels of objects in motion and thus regularizing the solution obtained by the MSS-KSC algorithm. In the experiments, we demonstrate the performance of the proposed method on synthetic data sets and real-life videos where the clusters evolve in a smooth fashion over time.

1. Introduction

In many real-life applications, ranging from data mining to machine perception, obtaining the labels of input data is often cumbersome and expensive. Therefore in many cases one encounters a large amount of unlabeled data while the labeled data are rare. Semi-supervised learning (SSL) is a framework in machine learning that aims at learning from both labeled and unlabeled data points (Zhu, 2006). SSL algorithms received a lot of attention in the last years due to rapidly increasing amounts of unlabeled data. Several semi-supervised algorithms have been proposed in the literature (Belkin, Niyogi, & Sindhwani, 2006; Chang, Pao, & Lee, 2012; He, 2004; Mehrkanoon & Suykens, 2012; Wang, Chen, & Zhou, 2012; Xiang, Nie, & Zhang, 2010; Yang et al., 2012). However, most of the SSL algorithms, operate in batch mode, hence requiring a large amount of computation time and memory to handle data streams like the ones found in real-life applications such as voice and face recognition, community detection of evolving networks and object tracking in computer vision. Therefore designing SSL algorithms that can operate in an on-line fashion is necessary for dealing with such data streams.

In the context of on-line clustering, due to the complex underlying dynamics and non-stationary behavior of real-life data, attempts have been made to design adaptive clustering algorithms. For instance, evolutionary spectral clustering based algorithms (Chakrabarti, Kumar, & Tomkins, 2006; Chi, Song, Zhou, Hino, & Tseng, 2007; Ning, Xu, Chi, Gong, & Huang, 2010), incremental K-means (Chakraborty & Nagwani, 2011), self-organizing time map (Sarlin, 2013) and incremental kernel spectral clustering (Langone, Agudelo, De Moor, & Suykens, 2014). However, in all above-mentioned algorithms the side-information (labels) is not incorporated and therefore they might underperform in certain situations.

A semi-supervised incremental clustering algorithm that can exploit the user constraints on data streams is proposed in Halkidi, Spiliopoulou, and Pavlou (2012). The user's prior information are presented to the algorithm in the form of must-link and cannot-link constraints. The authors in Kamiya, Ishii, Furao, and Hasegawa (2007) introduced an on-line semi-supervised algorithm based on a self-organizing incremental neural network.

Here we adopt the recently proposed multi-class semi-supervised kernel spectral clustering (MSS-KSC) algorithm

* Corresponding author. Tel.: +32 16328653; fax: +32 16321970.
E-mail addresses: siamak.mehrkanoon@esat.kuleuven.be, mehrkanoon2011@gmail.com (S. Mehrkanoon), mauricio.agudelo@esat.kuleuven.be (O.M. Agudelo), johan.suykens@esat.kuleuven.be (J.A.K. Suykens).

http://dx.doi.org/10.1016/j.neunet.2015.08.001
0893-6080/

(Mehrkanoon, Alzate, Mall, Langone, & Suykens, 2015) and make it applicable for an on-line data clustering/classification. In MSS-KSC the core model is kernel spectral clustering (KSC) algorithm introduced in Alzate and Suykens (2010). MSS-KSC is a regularized version of KSC which aims at incorporating the information of the labeled data points in the learning process. It has a systematic model selection criterion and the out-of-sample extension property. Moreover, as it has been shown in Mehrkanoon and Suykens (2014), it can scale to large data.

In contrast to the methods described in Adankon, Cheriet, and Biem (2009), Belkin et al. (2006), Chang et al. (2012), Xiang et al. (2010), Yang et al. (2012), in the MSS-KSC approach a purely unsupervised algorithm acts as a core model and the available side information is incorporated via a regularization term. In addition, the method can be applied for both on-line semi-supervised classification and clustering and uses a low-dimensional embedding. In the MSS-KSC approach, one needs to solve a linear system of equations to obtain the model parameters. Therefore with n number of training points, the algorithm has $\mathcal{O}(n^3)$ training complexity with naive implementations. The MSS-KSC model can be trained on a subset of the data (training data points) and then applied to the rest of the data in a learning framework. Thanks to the previously learned model, the out-of-sample extension property of the MSS-KSC model allows the prediction of the membership of a new point. However, in order to cope with non-stationary data-stream one also needs to continuously adjust the initial MSS-KSC model.

To this end, this paper introduces the Incremental MSS-KSC (I-MSS-KSC) algorithm which takes advantage of the available side-information to continuously adapt the initial MSS-KSC model and learn the underlying complex dynamics of the data-stream. The proposed method can be applied in several application domains including video segmentation, complex networks and medical imaging. In particular, in this paper we focus on video segmentation.

There have been some reports in the literature on formulating the object tracking task as a binary classification problem. For instance in Teichman and Thrun (2012) a tracking-based semi-supervised learning algorithm is developed for the classification of objects that have been segmented. The authors in Badrinarayanan, Budvytis, and Cipolla (2013) introduced a tree structured graphical model for video segmentation.

Due to the increasing demands in robotic applications, Kalman filtering has received significant attention. In particular Kalman filter has been applied in wide applications areas such as robot localization, navigation, object tracking and motion control (see Chen, 2012 and references therein). The authors in Suliman, Cruceru, and Moldoveanu (2010) use the Kalman filter for monitoring a contact in a video surveillance sequence. In Zhong and Sclaroff (2003), a Kalman filter based algorithm is presented to segment the foreground objects in video sequences given non-stationary textured background. An adaptive Kalman filter algorithm has been used for video moving object tracking in Weng, Kuo, and Tu (2006).

In case of the video segmentation, we show how Kalman filter can be integrated into the I-MSS-KSC algorithm as a regularizer by providing an estimation of the labels throughout the whole video sequences. This paper is organized as follows. In Section 2, the kernel spectral clustering (KSC) algorithm is briefly reviewed. In Section 3, an overview of the multi-class semi-supervised clustering (MSS-KSC) algorithm is given. The incremental multi-class semi-supervised clustering regularized by Kalman filtering approach is described in Section 4. In Section 5, experimental results are given in order to confirm the validity and applicability of the proposed method. The experimental findings and the demonstrative videos are provided in the supplementary material (see Appendix A) of this paper.[1]

2. Brief overview of KSC

The KSC method corresponds to a weighted kernel PCA formulation providing a natural extension to out-of-sample data i.e. the possibility to apply the trained clustering model to out-of-sample points. Given training data $\mathcal{D} = \{x_i\}_{i=1}^n$, $x_i \in \mathbb{R}^d$, the primal problem of kernel spectral clustering is formulated as follows (Alzate & Suykens, 2010):

$$\begin{aligned} \min_{w^{(\ell)},b^{(\ell)},e^{(\ell)}} \quad & \frac{1}{2}\sum_{\ell=1}^{N_c-1} w^{(\ell)^T} w^{(\ell)} - \frac{1}{2n}\sum_{\ell=1}^{N_c-1} \gamma_\ell e^{(\ell)^T} V e^{(\ell)} \\ \text{subject to} \quad & e^{(\ell)} = \Phi w^{(\ell)} + b^{(\ell)} 1_n,\ \ell = 1,\ldots,N_c-1 \end{aligned} \tag{1}$$

where N_c is the number of desired clusters, $e^{(\ell)} = [e_1^{(\ell)},\ldots,e_n^{(\ell)}]^T$ are the projected variables and $\ell = 1,\ldots,N_c-1$ indicates the number of score variables required to encode the N_c clusters. $\gamma_\ell \in \mathbb{R}^+$ are the regularization constants. Here

$$\Phi = [\varphi(x_1),\ldots,\varphi(x_n)]^T \in \mathbb{R}^{n\times h}$$

where $\varphi(\cdot) : \mathbb{R}^d \to \mathbb{R}^h$ is the feature map and h is the dimension of the feature space which can be infinite dimensional. A vector of all ones with size n is denoted by 1_n. $w^{(\ell)}$ is the model parameters vector in the primal. $V = \mathrm{diag}(v_1,\ldots,v_n)$ with $v_i \in \mathbb{R}^+$ is a user defined weighting matrix.

Applying the Karush–Kuhn–Tucker (KKT) optimality conditions one can show that the solution in the dual can be obtained by solving an eigenvalue problem of the following form:

$$V P_v \Omega \alpha^{(\ell)} = \lambda \alpha^{(\ell)}, \tag{2}$$

where $\lambda = n/\gamma_\ell$, $\alpha^{(\ell)}$ are the Lagrange multipliers and P_v is the weighted centering matrix:

$$P_v = I_n - \frac{1}{1_n^T V 1_n} 1_n 1_n^T V,$$

where I_n is the $n \times n$ identity matrix and Ω is the kernel matrix with ij-th entry $\Omega_{ij} = K(x_i,x_j) = \varphi(x_i)^T\varphi(x_j)$. In the ideal case of N_c well separated clusters, for a properly chosen kernel parameter, the matrix $VP_v\Omega$ has $N_c - 1$ piecewise constant eigenvectors with eigenvalue 1.

The eigenvalue problem (2) is related to spectral clustering with random walk Laplacian. In this case, the clustering problem can be interpreted as finding a partition of the graph in such a way that the random walker remains most of the time in the same cluster with few jumps to other clusters, minimizing the probability of transitions between clusters. It is shown that if

$$V = D^{-1} = \mathrm{diag}\left(\frac{1}{d_1},\ldots,\frac{1}{d_n}\right),$$

where $d_i = \sum_{j=1}^n K(x_i,x_j)$ is the degree of the ith data point, the dual problem is related to the random walk algorithm for spectral clustering.

From the KKT optimality conditions one can show that the score variables can be written as follows:

$$\begin{aligned} e^{(\ell)} &= \Phi w^{(\ell)} + b^{(\ell)} 1_n = \Phi\Phi^T\alpha^{(\ell)} + b^{(\ell)}1_n \\ &= \Omega\alpha^{(\ell)} + b^{(\ell)}1_n, \quad \ell = 1,\ldots,N_c-1. \end{aligned}$$

The out-of-sample extensions to test points $\{x_i\}_{i=1}^{n_{\mathrm{test}}}$ is done by an Error-Correcting Output Coding (ECOC) decoding scheme. First

[1] ftp://ftp.esat.kuleuven.be/stadius/siamak/155-spt.zip.