Measurement 89 (2016) 78–86

Contents lists available at ScienceDirect

Measurement

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

A novel distributed variational approximation method for density estimation in sensor networks

Behrouz Safarinejadian *, Mahboobeh Estakhri Estahbanati

Control Engineering Department, Shiraz University of Technology, Modarres Blvd., P.O. Box: 71555-313, Shiraz, Iran

ARTICLE INFO

Article history:
Received 6 April 2015
Received in revised form 28 March 2016
Accepted 29 March 2016
Available online 4 April 2016

Keywords:
Sensor networks
Consensus filter
Density estimation
Mixture of Gaussians
Variational approximations

ABSTRACT

In this paper, a consensus filter based distributed variational Bayesian (CFBDVB) algorithm is developed for distributed density estimation. Sensor measurements are assumed to be statistically modeled by a finite mixture model for which the CFBDVB algorithm is used to estimate the parameters, including means, covariances and weights of components. This algorithm is based on three steps: (1) calculating local sufficient statistics at every node, (2) estimating a global sufficient statistics vector using a consensus filter, (3) updating parameters of the finite mixture model based on the global sufficient statistics vector. Scalability and robustness are two advantages of the proposed algorithm. Convergence of the CFBDVB algorithm is also proved using Robbins–Monro stochastic approximation method. Finally, to verify performance of CFBDVB algorithm, we perform several simulations of sensor networks. Simulation results are very promising.

1. Introduction

Advances in network technology, like peer-to-peer networks on the internet or sensor networks, have highlighted the need for efficient ways to deal with large amounts of data that are distributed over a set of nodes. Examples are financial data reported on the internet, weather data observed by a set of sensors, etc. In particular, in many data mining applications we are interested in learning a global model from such data, like a probability distribution or a clustering of the data, without transferring all the data to a central unit [1].

The usefulness of graphical models in distributed learning and data mining is evaluated in [2,3] using multivariate Gaussian probability density functions. A distributed EM (Expectation Maximization) algorithm is proposed in [4] for distributed density estimation. The observations are assumed to be statistically modeled by a Gaussian mixture model. Distributed density estimation is also considered in [5,6] in which a distributed incremental EM (DIEM) method is proposed for distributed networks with a ring topology. Due to the ring topology employed in these references, failing a node will cause a critical problem. Other fault tolerant variants of the distributed EM algorithm have also been proposed using other network topologies [7–11]. Singularity in the estimated parameters is an important problem of the EM algorithm. Singularity may happen particularly if the assumed model order is not proper. Variational Bayesian approximation has been used recently to learn finite mixture model parameters. The basis of the variational Bayesian approach is to optimize a lower bound on the likelihood function by an iterative method [12–20]. Distributed consensus algorithms have recently emerged as a class of low-complexity, iterative distributed algorithms where neighboring nodes communicate with each other to reach an agreement [21]. In particular, the average consensus algorithm computes the average of an initial set of state values in a scalable and fault tolerant manner [22].

Consensus was early studied in [23] and has received increased attention in different fields due to its wide range of applications such as load balancing in parallel computing [24], coordination of autonomous agents [25–28], distributed control [29,30,22] and data fusion [31–35].

In this paper, it is assumed that each node in the network senses an environment that can be described as a mixture of some elementary surroundings. The measurements are thus statistically modeled by a finite mixture model, where each component corresponds to one of the elementary conditions. The paper proposes a consensus filter based distributed variational Bayesian (CFBDVB) algorithm for estimating the mixture components, which are common to the environment and sensor network as a whole, as well as the mixing probabilities that may vary from node to node. The algorithm produces an estimate of the density of the sensor data

* Corresponding author. Tel.: +98 71 37354500 5x2574.
E-mail addresses: safarinejad@sutech.ac.ir (B. Safarinejadian), m.estakhri@sutech.ac.ir (M.E. Estahbanati).

http://dx.doi.org/10.1016/j.measurement.2016.03.074
0263-2241/

without requiring the data to be transmitted to and processed at a central location. Alternatively, the algorithm can be viewed as a distributed processing strategy for clustering the sensor data into components corresponding to predominant environmental features sensed by the network. The applicability of the CFBDVB algorithm is not limited to distributed sensor networks, but it can be considered as a distributed data mining algorithm that can be used in any distributed data warehouse.

Our proposed CFBDVB algorithm calculates the local sufficient statistics firstly. The global sufficient statistics vector is estimated afterwards. Finally, the estimated global sufficient statistics are used to update the mixture model parameters. The estimation of global sufficient statistics is achieved by using an average consensus filter. The consensus filter can diffuse the local sufficient statistics over the entire network through communication with the neighboring nodes [21]. Therefore, the global sufficient statistics vector is estimated using the local information of each node and also neighbors' information. Each node, then updates the parameters in the same way as in the standard variational Bayesian algorithm. Scalability and robustness are two advantages of the proposed method. Because the consensus filter only requires local communication, the proposed CFBDVB algorithm is scalable. Furthermore, since no critical problem occurs in the case of any node failure, this algorithm is also scalable. Convergence properties of the CFBDVB algorithm are also studied here and it is shown that the CFBDVB converges to a maximum negative free energy point.

The rest of the paper is organized as follows. Section 2 discusses the variational Bayesian approach for finite mixture probabilistic models. Section 3 develops a consensus filter based distributed variational Bayesian algorithm to compute the maximum likelihood estimate. The proof of the stochastic approximation is presented in Section 4. Section 5 provides the results of simulations. Concluding remarks are given in Section 6.

2. Variational Bayesian algorithm for Gaussian mixtures

The variational Bayesian (VB) algorithm is an iterative method based mainly on the maximum likelihood principle. The idea behind the VB algorithm is as follows [38,39]. Assume observation $\mathcal{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N)$ drawn from a population with density function $f(\boldsymbol{y}; \boldsymbol{\psi})$. $\boldsymbol{\psi}$ is a vector of unknown parameters. In this paper, we assume that f can be represented by a Gaussian mixture model in which the components of the mixture are multivariate normal distributions $\mathcal{N}(\boldsymbol{y}; \boldsymbol{\mu}, \mathbf{T})$ where $\boldsymbol{y}$ is a multidimensional variable, and $\boldsymbol{\mu}$ and $\mathbf{T}$ are the mean and inverse covariance parameters, respectively. Therefore, assume that the distribution of the measurements is represented by

$$f(\boldsymbol{y}; \pi, \boldsymbol{\varphi}) = \sum_{j=1}^{J} \pi_j \mathcal{N}(\boldsymbol{y}; \boldsymbol{\mu}_j, \boldsymbol{T}_j) \tag{1}$$

where $J \in \mathbb{Z}_{\geqslant 1}$ is the number of mixture components, $j \in \mathbb{Z}_{\geqslant 1}$ represents the component index, $\pi = \{\pi_j\}_{j=1}^{J}$ are the *mixture probabilities* and $\boldsymbol{\phi}_j = \{\boldsymbol{\mu}_j, \Sigma_j\}$ is the set of parameters defining the jth component. Define the set of parameters $\boldsymbol{\phi} = \{\boldsymbol{\phi}_j\}_{j=1}^{J}$, $\boldsymbol{\psi} = \{\pi, \boldsymbol{\phi}\}$.

The observed data likelihood function is defined by

$$L(\boldsymbol{\psi}) = f(\mathcal{Y}|\boldsymbol{\psi}) = \prod_{i=1}^{N} \left[\sum_{j=1}^{J} \pi_j \mathcal{N}(\boldsymbol{y}_i; \boldsymbol{\mu}_j, \boldsymbol{T}_j) \right] \tag{2}$$

The VB algorithm is an iterative procedure to find the $\boldsymbol{\psi}$ that maximizes $f(\mathcal{Y}|\boldsymbol{\psi})$ by data augmentation.

Consider a set of missing variables $\mathcal{Z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N)$ corresponding to $\mathcal{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N)$. Each $\boldsymbol{z}_i = [z_i^1, \ldots, z_i^J]$ is a binary vector indicating by which component the data $\boldsymbol{y}_i$ is produced. We would say $\boldsymbol{y}_i$ is produced by the jth component of the mixture if for all $r \neq j$, $z_i^r = 0$ and $z_i^j = 1$.

$\mathcal{X} = \langle \mathcal{Y}, \mathcal{Z} \rangle$ is regarded as the complete data with density function $f_c(\mathcal{X}; \boldsymbol{\psi})$. The complete data log likelihood is $\log L_c(\boldsymbol{\psi}) = \log f_c(\mathcal{X}; \boldsymbol{\psi})$.

Variational methods involve the introduction of a distribution $g(\boldsymbol{\phi})$ which provides an approximation to the true posterior distribution. Using Jensen's inequality [38], we can write

$$\begin{aligned} \ln f(\mathcal{Y}|\pi) &= \ln \int f(\mathcal{Y}, \boldsymbol{\varphi}|\pi) d\boldsymbol{\varphi} = \ln \int g(\boldsymbol{\varphi}) \frac{f(\mathcal{Y}, \boldsymbol{\varphi}|\pi)}{g(\boldsymbol{\varphi})} d\boldsymbol{\varphi} \\ &\geqslant \int g(\boldsymbol{\varphi}) \ln \frac{f(\mathcal{Y}, \boldsymbol{\varphi}|\pi)}{g(\boldsymbol{\varphi})} d\boldsymbol{\varphi} = L(g) \end{aligned} \tag{3}$$

If the g distribution is chosen suitably in the above equation, then the quantity $L(g)$ will be tractable to compute, although the original log-likelihood function is not. A variational approach seeks to choose a suitable form for $g(\boldsymbol{\varphi})$ that not only is sufficiently simple but also is sufficiently flexible. Indeed, we choose some family of g distributions and then try to find the best approximation by maximizing $L(g)$. Therefore, the parameters π and $\boldsymbol{\phi}$ are assumed to have conjugate priori pdf's as

$$\begin{aligned} f_\pi(\pi) &= Dir(\pi; \alpha_1^0, \ldots, \alpha_J^0) \\ f_{\mu|T}(\mu|T) &= \prod_{j=1}^{J} \mathcal{N}\left(\mu_j; m_j^0, (\beta_j^0 T_j)^{-1}\right) \\ f_T(T) &= \prod_{j=1}^{J} \mathcal{W}(T_j; \upsilon_j^0, \Sigma_j^0) \end{aligned} \tag{4}$$

where *Dir* and *W* represent *Dirichlet* and *Wishart* pdf's, respectively, $\mu = (\mu_1, \ldots, \mu_J)$ and $T = (T_1, \ldots, T_J)$.

The quantities α_j^0, m_j^0, β_j^0, υ_j^0 and Σ_j^0 are called hyperparameters. The hyperparameters of the posterior distributions obtained using the variational approximation method are then

$$\begin{aligned} \hat{f}_\pi(\pi) &= Dir(\pi; \alpha_1, \ldots, \alpha_J) \\ \hat{f}_{\mu|T}(\mu|T) &= \prod_{j=1}^{J} \mathcal{N}\left(\mu_j; m_j, (\beta_j T_j)^{-1}\right) \\ \hat{f}_T(T) &= \prod_{j=1}^{J} \mathcal{W}(T_j; \upsilon_j, \Sigma_j) \end{aligned} \tag{5}$$

with hyperparameters given by

$$\begin{aligned} \alpha_j &= \alpha_j^{(0)} + \sum_{i=1}^{N} q_{i,j} \\ \beta_j &= \beta_j^{(0)} + \sum_{i=1}^{N} q_{i,j} \\ \boldsymbol{m}_j &= \frac{\beta_j^{(0)} \boldsymbol{m}_j^{(0)} + \sum_{i=1}^{N} q_{i,j} \boldsymbol{y}_i}{\beta_j} \\ \Sigma_j &= \Sigma_j^{(0)} + \sum_{i=1}^{N} q_{i,j} \boldsymbol{y}_i \boldsymbol{y}_i^T + \beta_j^{(0)} \boldsymbol{m}_j^{(0)} \boldsymbol{m}_j^{(0)^T} - \beta_j \boldsymbol{m}_j \boldsymbol{m}_j^T \\ \upsilon_j &= \upsilon_j^{(0)} + \sum_{i=1}^{N} q_{i,j} \end{aligned} \tag{6}$$

In the above

$$q_{i,j} = \frac{\varphi_{i,j}}{s_i} \tag{7}$$

with