Engineering Applications of Artificial Intelligence 43 (2015) 1–14

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Expectation propagation learning of a Dirichlet process mixture of Beta-Liouville distributions for proportional data clustering

Wentao Fan [a], Nizar Bouguila [b,*]

[a] Department of Computer Science and Technology, Huaqiao University, Xiamen, China
[b] The Concordia Institute for Information Systems Engineering (CIISE), Concordia University, Montreal, QC, Canada H3G 1T7

ARTICLE INFO

Article history:
Received 9 January 2014
Received in revised form
24 March 2015
Accepted 27 March 2015
Available online 18 April 2015

Keywords:
Unsupervised learning
Expectation propagation
Dirichlet process
Beta-Liouville distribution
Facial expression
Action recognition

ABSTRACT

We propose a nonparametric Bayesian model for the clustering of proportional data. Our model is based on an infinite mixture of Beta-Liouville distributions and allows a compact description of complex data. The choice of the Beta-Liouville as the basis of our model is justified by the fact that it has been shown to be a good alternative to the Dirichlet and generalized Dirichlet distributions for the statistical representation of proportional data. Using this infinite mixture, we show how a careful modeling can achieve good results by allowing the elicitation of prior belief about the parameters and the number of clusters through suitable learning. Indeed, we develop an efficient learning algorithm, based on expectation propagation, to estimate the parameters of our infinite Beta-Liouville mixture model. The feasibility and effectiveness of the proposed method are demonstrated by two challenging applications namely action and facial expression recognition.

1. Introduction

Statistical models are becoming increasingly important because of their role in providing a concise picture of the data, by taking uncertainty into account (Lewis and Catlett, 1994; Frey et al., 1995; Rosset and Segal, 2002; Keysers et al., 2004), and then in the development of useful algorithms for pattern recognition, computer vision, and image processing (Yildizer et al., 2012; Liao et al., 2013). Finite mixtures have been widely used in the past for statistical modeling and exploring data structure (Patrick, 1968; McLachlan and Peel, 2000; Nock and Nielsen, 2006). Images and videos modeling and clustering is a prime example of the role mixtures play. In practice, however, mixture-based modeling rely generally on simplistic assumptions that may compromise modeling and generalization capabilities. Examples of these assumptions include supposing that the number of clusters is known in advance, which implies that we have to rely on the practitioner ability to determine the optimal complexity, or using a multivariate normal distribution for modeling which disregard the nature of the data.

Infinite mixtures have been proposed to overcome the deficiencies related to finite mixtures and have been shown to be effective tools in data analysis, modeling and clustering (Lau and Green, 2007). Traditionally, there has been interest in infinite mixture models from a wide variety of disciplines including machine learning, data mining, pattern recognition, and computer vision. The prevalent assumption when using infinite mixture models has been to consider that the component densities are Gaussians. Unfortunately, the Gaussian assumption may not be met in practice and is often violated, producing poor modeling results. This is especially true in the case of proportional data (e.g. normalized histograms) which are largely present and naturally generated in several domains. Examples include the representation of textual (or visual) documents using histograms containing the normalized frequencies of words (or visual) words in a given dictionary (Bouguila, 2012a). The goal of this paper is to examine another alternative based on Beta-Liouville distribution to model suitably proportional data. Indeed, few applications of the Beta-Liouville mixture have appeared recently, and much of the potential of this model has not been realized yet (Bouguila, 2011, 2012a,b). In Bouguila (2012a), finite Beta-Liouville mixture models are applied on scene modeling and classification, and automatic image orientation detection. In Bouguila (2012b), infinite Beta-Liouville mixture models have been proposed and been successfully applied on text classification and texture discrimination.

Infinite mixture-based modeling belongs to the group of nonparametric Bayesian approaches which have been widely adopted recently (Hirano, 2002; Chib and Hamilton, 2002; Li et al., 2007; Ray and Mallick, 2006; Bouguila, 2012b). A challenging problem in this context is the development of efficient learning approaches.

* Corresponding author. Tel.: +1 5148482424; fax: +1 5148483171.
E-mail addresses: fwt@hqu.edu.cn (W. Fan), nizar.bouguila@concordia.ca (N. Bouguila).

http://dx.doi.org/10.1016/j.engappai.2015.03.016
0952-1976/

Many works have shown that principled Bayesian approaches to learning lead generally to improvement in modeling (Andrieu et al., 2001; Utsugi and Kumagai, 2001). For a broader perspective, we refer the interested reader to Marin and Robert (2007). Thus, Bayesian inference has been widely used to learn infinite mixture models and has been shown to be effective in many data modeling problems. However, a practical disadvantage of pure Bayesian approaches is that they require multiple integrations over variables, which are usually analytically intractable. Generally, Monte Carlo Markov Chain (MCMC) techniques have been adopted in this case, yet they unfortunately involve huge computational costs (Marin and Robert, 2007; Bouguila, 2012b). Several deterministic approximation techniques have been then proposed to overcome this problem (see, for instance, Ghahramani and Beal, 2000; Xing et al., 2003). Among these techniques, expectation propagation (Minka, 2001) has been shown to provide good generalization capabilities as well as computational tractability in several real life applications.

Expectation propagation is a deterministic approximation scheme based on the minimization of a Kullback–Leibler (KL) divergence between the true model's posterior and an approximation (Minka, 2001; Minka and Lafferty, 2002). It is an extension to assumed-density filtering (ADF) (Maybeck, 1982) which is a one pass, sequential approximation method. In contrast to the ADF, the order of the input data points is not important in the expectation propagation inference and its inference accuracy is improved by reusing the data points many times. In Stern et al. (2009), a probabilistic model based on expectation propagation learning method was proposed for generating personalized recommendations of items to users of a web service. Thus, we shall adopt expectation propagation in this paper within a learning framework that we will develop to estimate the parameters of infinite Beta-Liouville mixture models. The proposed approach manages to overcome the problem of model complexity selection which has been the topic of extensive research in the past (Hammond et al., 1993; Cherkassky et al., 1999) and which is, in our case, automatically determined in the learning process. In order to show that the proposed approach suits the needs of real-life problems very well, we validate it using two challenging applications namely action and facial expression recognition.

The paper proceeds as follows. In Section 2, we present the infinite Beta-Liouville mixture model. In Section 3, the expectation propagation learning framework is developed. The experimental results are presented, analyzed, and discussed in Section 4. Finally, in Section 5, we give the conclusion.

2. Infinite Beta-Liouville mixture model

Finite mixture models are often adopted as effective tools to capture the multimodality of the data and to reason under uncertainty. However, one major concern regarding clustering in general, and finite mixture modeling in particular, is the selection of the optimal number of mixture components (Yang et al., 2011; Bouguila, 2012a). This obstacle can be removed by assuming that there is an infinite number of components through a Bayesian nonparametric framework known as the Dirichlet process mixture model. In this section, we shall first briefly review the finite Beta-Liouville mixture model. Then, we extend it to the infinite case by using the Dirichlet process mixture framework with a stick-breaking representation.

2.1. Finite Beta-Liouville mixture model

Assume that a D-dimensional random vector $\vec{X} = (X_1, \ldots, X_D)$ follows a Liouville distribution of the second kind with positive parameters $(\alpha_1, \ldots, \alpha_D)$ and density generator $g(\cdot)$. Then, the probability density function of $\vec{X}$ is defined by Bouguila (2012a)

$$p(\vec{X} \mid \alpha_1, \ldots, \alpha_D) = g(u) \prod_{l=1}^{D} \frac{X_l^{\alpha_l - 1}}{\Gamma(\alpha_l)} \tag{1}$$

where $u = \sum_{l=1}^{D} X_l < 1$ and $X_l > 0$, $l = 1, \ldots, D$. The mean and covariance of the Liouville distribution are given by

$$E(X_l) = E(u) \frac{\alpha_l}{\sum_{l=1}^{D} \alpha_l} \tag{2}$$

$$Cov(X_a, X_b) = \frac{\alpha_a \alpha_b}{\sum_{a=1}^{l} \alpha_a} \left(\frac{E(u^2)}{\sum_{l=1}^{D} \alpha_l + 1} - \frac{E(u)^2}{\sum_{l=1}^{D} \alpha_l} \right) \tag{3}$$

where $E(u)$ and $E(u^2)$ are the first and second moments of a random variable u. The variable u follows a probability density function $f(\cdot)$ namely the generating density, and is related to the density generator $g(\cdot)$ in the form

$$g(u) = \frac{\Gamma(\sum_{l=1}^{D} \alpha_l)}{u^{\sum_{l=1}^{D} \alpha_l - 1}} f(u) \tag{4}$$

Therefore, we can rewrite the Liouville distribution of the second kind in Eq. (1) as

$$p(\vec{X} \mid \alpha_1, \ldots, \alpha_D) = \frac{\Gamma(\sum_{l=1}^{D} \alpha_l)}{u^{\sum_{l=1}^{D} \alpha_l - 1}} f(u) \prod_{l=1}^{D} \frac{X_l^{\alpha_l - 1}}{\Gamma(\alpha_l)} \tag{5}$$

Two important properties regarding the Liouville distribution are noticeable: first, it has a more general covariance structure than the Dirichlet distribution; second, similar to the Dirichlet, the Liouville distribution is conjugate to the multinomial (Bouguila, 2012a). Motivated by the fact that the Beta distribution has a flexible shape and can approximate nearly any arbitrary distribution, it is adopted as the generating density for u with positive parameters α and β:

$$f(u \mid \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} u^{\alpha - 1} (1 - u)^{\beta - 1} \tag{6}$$

We then obtain the so-called Beta-Liouville distribution by substituting Eq. (6) into Eq. (5):

$$\mathrm{BL}(\vec{X} \mid \vec{\theta}) = \frac{\Gamma(\sum_{l=1}^{D} \alpha_l)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \prod_{l=1}^{D} \frac{X_l^{\alpha_l - 1}}{\Gamma(\alpha_l)} \left(\sum_{l=1}^{D} X_l \right)^{\alpha - \sum_{l=1}^{D} \alpha_l} \times \left(1 - \sum_{l=1}^{D} X_l \right)^{\beta - 1} \tag{7}$$

where $\vec{\theta} = (\alpha_1, \ldots, \alpha_D, \alpha, \beta)$ are the parameters of the Beta-Liouville distribution. It is noteworthy that when the density generator has a Beta distribution with parameters $\sum_{l=1}^{D} \alpha_l$ and α_{D+1}, Eq. (1) is reduced to the Dirichlet distribution with parameters $\alpha_1, \ldots, \alpha_{D+1}$. Therefore, the Beta-Liouville distribution includes the Dirichlet as a special case. Given a set of N vectors $\mathcal{X} = \{\vec{X}_1, \ldots, \vec{X}_N\}$, where each D-dimensional vector $\vec{X}_i = (X_{i1}, \ldots, X_{iD})$ follows a finite Beta-Liouville mixture model with M components, then

$$p(\vec{X}_i \mid \vec{\pi}, \vec{\theta}) = \sum_{j=1}^{M} \pi_j \mathrm{BL}(\vec{X}_i \mid \theta_j) \tag{8}$$

where $\theta_j = (\alpha_{j1}, \ldots, \alpha_{jD}, \alpha_j, \beta_j)$ are the parameters of the Beta-Liouville distribution corresponding to component j. Moreover, $\vec{\pi} = (\pi_1, \ldots, \pi_M)$ represents the vector of mixing probabilities which are positive and sum to one.

2.2. Infinite Beta-Liouville mixture model via Dirichlet process

A conventional finite mixture model can be extended to have an infinite number of mixture components using a Dirichlet process with a stick-breaking representation (Sethuraman, 1994; Ishwaran and