Information Fusion 20 (2014) 136–145

Contents lists available at ScienceDirect

Information Fusion

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

Composite distance based approach to von Mises mixture reduction

Mario Bukal *, Ivan Marković, Ivan Petrović

University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb, Croatia

A R T I C L E I N F O

Article history:
Received 23 July 2013
Received in revised form 6 December 2013
Accepted 6 January 2014
Available online 4 February 2014

Keywords:
Von Mises mixture
Mixture component number reduction
Rényi α-divergence
Composite distance measure
People trajectory shape analysis

A B S T R A C T

This paper presents a systematic approach for component number reduction in mixtures of exponential families, putting a special emphasis on the von Mises mixtures. We propose to formulate the problem as an optimization problem utilizing a new class of computationally tractable composite distance measures as cost functions, namely the composite Rényi α-divergences, which include the composite Kullback–Leibler distance as a special case. Furthermore, we prove that the composite divergence bounds from above the corresponding intractable Rényi α-divergence between a pair of mixtures. As a solution to the optimization problem we synthesize that two existing suboptimal solution strategies, the generalized *k*-means and a pairwise merging approach, are actually minimization methods for the composite distance measures. Moreover, in the present paper the existing joining algorithm is also extended for comparison purposes. The algorithms are implemented and their reduction results are compared and discussed on two examples of von Mises mixtures: a synthetic mixture and a real-world mixture used in people trajectory shape analysis.

1. Introduction

Many statistical and engineering problems [1–3] require modeling of complex multi-modal data, wherein mixture distributions became an inevitable tool. In this paper we draw attention to finite mixtures of a specific distribution on the unit circle, the von Mises distribution. Starting from 1918 and the seminal work of von Mises [4], where he investigated hypothesis on integrality of atomic weights of chemical elements, the proposed parametric density plays a pertinent role in directional statistics with wide range of applications in physics, biology, image analysis, neural science and medicine – confer monograms [5–7] and references therein.

Estimation of complex data by mixture distributions may lead to models with large or, in applications like target tracking, ever increasing number of components. In lack of efficient reduction procedures, such models become computationally intractable and lose their feasibility. Therefore, component number reduction in mixture models is an essential tool in many domains like image and multimedia indexing [8,9], speech segmentation [10], and it is an indispensable part of any tracking system with mixtures of Gaussian [11–13] or von Mises distributions [3]. The subject matter is particularly relevant to the information fusion domain since it relates to the following challenging problems in multisensor data fusion [14]: data dimensionality, processing framework, and data association. These problems are related to component reduction by the fact that measurement data as quantity of interest can be preprocessed (compressed) prior to communicating it to other nodes (in a decentralized framework) or the fusion center, thus effectively saving on the communication bandwidth and power required for transmitting data. For example, consider the problem of people trajectory analysis with von Mises mixtures [2] in a distributed sensor networks where the mixtures might need to be communicated between the sensor nodes. Motivated by [2,3], in this paper we study methods and respective algorithms for component number reduction in mixtures of von Mises distributions, but due to the general exposition of the subject in the framework of exponential family mixtures, the methods and findings easily extend to other examples like mixtures of Gaussian distributions, and von Mises-Fisher distributions [5].

Existing literature on mixture reduction schemes is mostly related to Gaussian mixture models. A reduction scheme for Gaussian mixtures in the context of Bayesian tracking systems in a cluttered environment, which successively merges the closest pair of components and henceforth referred to as the *joining algorithm*, was proposed in [11]. The main drawback of the scheme is its local character, which gives no information about the global deviation of the reduced mixture from the original one. In [15] the mixture reduction was formulated as an optimization problem for the integral square difference cost function. A better suited distance measure between probability distributions is the Kullback–Leibler (KL)

* Corresponding author. Tel.: +385 16129577.
E-mail addresses: mario.bukal@fer.hr (M. Bukal), ivan.markovic@fer.hr (I. Marković), ivan.petrovic@fer.hr (I. Petrović).

http://dx.doi.org/10.1016/j.inffus.2014.01.003
1566-2535/

distance [16], but it lacks a closed form formula between mixtures, what makes it computationally inconvenient. Several concepts have been employed to circumvent this problem. A new distance measure between mixture distributions, based on the KL distance, which can be expressed analytically was derived in [17], and utilized to solve the mixture reduction problem. In [12] an upper bound for the KL distance was obtained and used as dissimilarity measure in a successive pairwise reduction of Gaussian mixtures – henceforth we refer to it as the *pairwise merging algorithm*. Unlike the joining algorithm, this procedure gives a control of the global deviation of the reduced mixture from the original one. Introducing the notion of Bregman information, the authors in [18] generalized the previously developed Gaussian mixture reduction concepts to arbitrary exponential family mixtures. Further development of these techniques for exponential family mixtures can be found in [19–24]. Finally, we mention the variational Bayesian approach [25,26] as well as [27] as alternative concepts of mixture reduction developed for Gaussian mixtures.

Contributions of the present paper are as follows. Firstly, we formulate the problem of component number reduction in exponential family mixtures as an optimization problem utilizing a new class of composite distance measures as cost functions. These distance measures are constructed employing Rényi α-divergences as ground distances, and it is shown that the composite distance bounds the corresponding Rényi α-divergence from above (see Lemma 1 below). This inequality is very important since it provides an information on the global deviation of the reduced mixture from the original one measured by the Rényi α-divergence. Secondly, we synthesize previously developed reduction techniques [12,18,24] in the sense that they can all be interpreted as suboptimal solution strategies to the proposed optimization problem. For the purpose of computational complexity and accuracy comparisons, the joining algorithm is extended using the scaled symmetrized KL distance as a dissimilarity measure between mixture components. Thirdly, special attention is given to von Mises mixtures for which we present analytical expressions for solving the component number reduction problem and analyze them on two examples: a synthetic 100-component mixture with several dominant modes and a real-world mixture stemming from the work on people trajectory analysis in video data [2].

Outline of the paper is as follows. The general framework of exponential family mixtures is introduced in Section 2 together with a brief survey on distance measures between probability distributions and definition of composite distance measures. Section 3 presents the component number reduction in exponential family mixtures as a constrained optimization problem. In Section 4 we discuss two suboptimal solution strategies and additionally consider the joining algorithm. Numerical experiments on two examples of circular data are performed and obtained results are discussed in Section 5. Finally, Section 6 concludes the paper by outlining main achievements and commenting on possible extensions.

2. General background

In this section we introduce exponential family distributions and the von Mises distribution as their subclass, we recall the notion of finite mixture distributions and discuss variety of distance measures between probability distributions emphasizing on *composite distance measures* between mixtures.

2.1. Exponential family distributions

A parametric set of probability distributions defined on a sample space $\mathcal{X}$ and parametrized by the natural parameter $\theta \in \Theta \subset \mathbb{R}^d$ is called *exponential family* if their probability densities admit the following canonical representation

$$p_F(x;\theta) = \exp(T(x)\cdot\theta - F(\theta) + C(x)), \quad x \in \mathcal{X}. \tag{1}$$

Map $T : \mathcal{X} \to \mathbb{R}^d$ is called the minimal sufficient statistics, and functions F and C denote the log-normalizer (or log-partition) and the carrier measure, respectively. It can be proved that $\Theta = \mathrm{Dom}(F)$ is a nonempty convex set, and F is convex and unique up to an additive constant [28]. Moreover, if the exponential family is *regular* (i.e. Θ is open), then F is strictly convex and differentiable on Θ [18]. In further, the exponential family accompanied with the convex function F will be denoted by $\mathcal{E}_F$.

Many well known parametric distributions, like Gaussian, Poisson, Gamma, and Dirichlet, are exponential families [6]. For the reader's convenience, recall the simplest example of the univariate Gaussian distribution

$$p(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-(x-\mu)^2/2\sigma^2\right),$$

with standard parameters (μ,σ^2), which is an exponential family with natural parameter $\theta = (\mu/\sigma^2, 1/2\sigma^2) \in \mathbb{R}^2$, sufficient statistics $T(x) = (x, -x^2)$, log-normalizer $F(\theta) = \theta_1^2/4\theta_2 + \log(\pi/\theta_2)/2$, and $C(x) = 0$. Canonical parametrizations (1) for other exponential families can be found in [29], and in the sequel we focus on our study example – the von Mises distribution.

2.1.1. Von Mises distribution

The von Mises distribution is a probability distribution defined on the unit circle, or equivalently on the interval $[0,2\pi)$, with density function given by

$$p(x;\mu,\kappa) = \frac{1}{2\pi I_0(\kappa)}\exp\{\kappa\cos(x-\mu)\}, \quad 0 \leqslant x < 2\pi, \tag{2}$$

where $\mu \in [0,2\pi)$ denotes the mean angle, $\kappa \geqslant 0$ is the concentration parameter, and I_0 is the modified Bessel function of the first kind and of order zero [5]. Recall, the modified Bessel function of the first kind and of order $n \in \mathbb{N}$ is defined by

$$I_n(\kappa) = \frac{1}{2\pi}\int_0^{2\pi}\exp(\kappa\cos\xi)\cos(n\xi)\,\mathrm{d}\xi. \tag{3}$$

In many ways von Mises distribution is considered as the circular analog of the univariate Gaussian distribution: it is unimodal, symmetric around the mean angle μ, and the concentration parameter κ is analogous to the inverse of the variance. Furthermore, it is characterized by the maximum entropy principle in the sense that it maximizes the Boltzmann–Shannon entropy under prescribed circular mean [5].

From (2) it can be readily derived that von Mises distribution with standard parameters (μ,κ) is an exponential family parametrized by the natural parameter $\theta = (\kappa\cos\mu, \kappa\sin\mu) \in \Theta = \mathbb{R}^2$. The minimal sufficient statistics is the standard parametrization of the unit circle $T(x) = (\cos x, \sin x)$, the log-normalizer is given by

$$F(\theta) = \log\left(2\pi I_0\left(\sqrt{\theta_1^2+\theta_2^2}\right)\right), \tag{4}$$

and the carrier measure is trivial, $C(x) = 0$.

2.2. Exponential family mixtures

A finite exponential family mixture distribution is a weighted normalized sum of distributions belonging to the same exponential family $\mathcal{E}_F$. Its density function is given by

$$p(x) = \sum_{i=1}^{K} w_i p_F(x;\theta_i), \quad x \in \mathcal{X}, \tag{5}$$