Information Processing and Management 49 (2013) 123–137

Contents lists available at SciVerse ScienceDirect

Information Processing and Management

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

Deriving kernels from generalized Dirichlet mixture models and applications

Nizar Bouguila

Concordia Institute for Information Systems Engineering, Faculty of Engineering and Computer Science, Concordia University, Montreal, Qc, Canada H3G 2W1

A R T I C L E I N F O

Article history:
Received 22 January 2010
Received in revised form 6 February 2011
Accepted 25 June 2012
Available online 31 July 2012

Keywords:
Finite mixture
Generalized Dirichlet
Clustering
Agglomerative EM
SVM
Generative learning
Discriminative learning
Object detection
Image database

A B S T R A C T

In the last few years hybrid generative discriminative approaches have received increasing attention and their capabilities have been demonstrated by several applications in different domains. Hybrid approaches allow the incorporation of prior knowledge about the nature of the data to classify. Past work on hybrid approaches has focused on Gaussian data, however, and less attention has been given to other kinds of non-Gaussian data which appear in many applications. In this article we introduce a class of generative kernels based on finite mixture models for non-Gaussian data classification. This particular class is based on the generalized Dirichlet distribution which have been shown to be effective to model this kind of data. We demonstrate the efficacy of the proposed framework on two challenging applications namely object detection and content-based image classification via the integration of color and spatial information.

1. Introduction

Clustering techniques have found several applications in statistical data analysis, processing, and mining by offering the advantage that the categorization process is unsupervised (i.e a priori knowledge of categories is not needed) (Everitt, 2009; Helmbold, 1997). The main goal is to group together the data in such a way that data within the same cluster are more similar to each other than to those in other clusters. Clustering is a difficult task, since we do not usually know the number of clusters, which actually affects the quality of modeling, in the input data a priori. In fact, the selection of the number of clusters has to be performed directly from the data. Clustering is generally viewed as a density estimation problem (Fayyad, Reina, & Bradley, 1998; Jardine, 1971). There is currently a growing interest in the use of finite mixtures of distributions as generative models and as a principled formal approach to clustering (McLachlan & Peel, 2000). One of the most popular distributions preferred as mixture components is the Gaussian distribution which has been the subject of extensive theoretical and experimental studies and which has generally relied on the assumption of independence between the components in the feature space (i.e, the consideration of diagonal covariance matrices) when dealing with high-dimensional vectors. This assumption can severely compromise the modeling and classification accuracies in real life applications (Bouguila, Ziou, & Vaillancourt, 2004). Recent developments have shown that other choices may perform better especially when dealing with non-Gaussian data. Among the range of alternatives, the generalized Dirichlet distribution has been shown to be effective and flexible due to its favorable properties widely discussed in Bouguila and Ziou (2006); Bouguila, Ziou, and Hammoud (2009) which has justified its use in many applications (Bouguila & Ziou, 2004; Ziou, Hamri, & Boutemedjet, 2009). Given

E-mail address: bouguila@ciise.concordia.ca

0306-4573/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ipm.2012.06.002

a set of unlabeled data generated from a generalized Dirichlet mixture model, the main goals are to learn the parameters of each mixture component namely the parameters of each generalized Dirichlet and the importance of each component (i.e mixing parameter) (Bouguila & Ziou, 2006), and to select the optimal number of clusters which has been tackled in Bouguila and Ziou (2007). As a generative[1] approach which requires little manual supervision and labeling and having the ability to learn from data with missing information, finite mixture models have received a lot of attention. At the same time, discriminative approaches have the advantage to produce accurate classifiers by learning decision boundaries, but require large training sets. Indeed, it is well known that depending on the application and data (number of training samples) generative and discriminative approaches may outperform each other (Ng & Jordan, 2001; Raina, Shen, Ng, & McCallum, 2003). In recent years, Support Vector Machines (SVMs) have become the state-of-the-art classifier for supervised classification problems, and have demonstrated great successes in a broad range of applications such as document categorization, character recognition, and image classification (Schölkopf et al., 2001; Vapnik, 1998). For instance, the authors in Moghaddam and Yang (2000) have shown that SVM performs better than linear, quadratic, Fisher linear discriminant, and neighbor classifiers and it outperforms even human subjects for the specific problem of gender classification.

Recent studies have shown that a compromise is to take advantage of both approaches via hybrid generative/discriminative techniques which have several advantages by providing, for instance, lower test errors than either their purely generative or discriminative counterparts (Bouguila & Amayri, 2009; Raina et al., 2003). Our research focuses upon this same endeavor. Although some approaches have been proposed, most research have been made toward general approaches that do not take into account the nature of the data to classify by heavily relying on the Gaussian assumption. A good model should, however, reflect the nature of the data to analyze and be adapted as well as possible to the underlying structure of the data input space. This paper focuses instead on the development of approaches for non-Gaussian data classification which are present and appear naturally in several applications (Aitchison, 1986). In particular, we develop several SVM generative kernels from the generalized Dirichlet mixture which can be used for structured objects (i.e objects that can be represented as sets of vectors) such as images. These generated kernels make intelligent use of unlabeled data, which have been shown to have an important value for classification problems (Zhang & Oles, 2000), and have better generalization capabilities.

The outline of the paper is as follows. In Section 2, we briefly review the generalized Dirichlet mixture (GDM) model and then we propose a novel approach for GDM learning. Section 3 proposes several kernels generated from the GDM. In Section 4, we present experimental results on two challenging tasks: object detection and content-based image classification via the integration of color and spatial information. Finally, Section 5 presents our conclusions.

1.1. Relevant related works

In recent years, some researchers have voiced the concern that classic SVM kernels cannot be the best choice in all applications and that a better approach is to generate the kernels directly from the data (i.e. let the data speak for itself). These kernels are defined on probabilistic generative models learned from these data. Previous efforts include the Fisher kernel initially proposed in Jaakkola and Haussler (1998) and which main idea is to exploit the geometric structure on the statistical manifold by mapping a given individual sequence of vectors into a single feature vector, defined in the gradient log-likelihood space. The main motivation is that the gradient of the log-likelihood of the model captures the generative process of the data. In Jaakkola and Haussler (1998), it has been shown that the Fisher kernel performs asymptotically at least as good as the generative model from which is developed. The Fisher kernel has been used, for instance, in Wan and Renals (2005) where Gaussian mixture model-based Fisher kernels are developed for speaker identification and verification, in Sing and Beerenwinkel (2006) in computational biology and in Holub, Welling, and Perona (2008) for object recognition. A more general framework for defining kernel functions called "marginalized kernels" has been proposed in Tsuda, Kin, and Asai (2002) and applied to analyze biological sequences. An alternative to Fisher kernel called probability product kernel has been proposed in Jebara, Kondor, and Howard (2004) where two main special cases have been developed namely Bhattacharyya kernel and expected likelihood (or correlation) kernel. Several information divergence-based kernels have been proposed in Chan, Vasconcelos, and Moreno (2004), also. Examples include the Kullback–Leibler divergence kernel investigated in Moreno, Ho, and Vasconcelos (2003), Rényi and Jensen–Shannon Kernels. Many other kernels have been proposed also for other kinds of structured data such as strings, trees and graphs, but are beyond the scope of this paper. The reader is refereed to Gärtner (2003) for interesting and in depth discussions about kernels on structured data.

2. Fitting generalized Dirichlet mixture models

2.1. The generalized Dirichlet mixture model

A GDM model is defined as a convex combination of generalized Dirichlet distributions (GDDs). A GDD in a D-dimensional space, characterized by its $2D$ positive parameters $\alpha_1, \beta_1, \ldots, \alpha_D, \beta_D$, is defined as Connor and Mosimann (1969)

[1] Generative learning is also called informative learning (see, for instance, Rubinstein & Hastie, 1997).