Pattern Recognition Letters 29 (2008) 1901–1905

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Robust fuzzy clustering using mixtures of Student's-*t* distributions

Sotirios Chatzis *, Theodora Varvarigou

Electrical and Computer Engineering, National Technical University of Athens, 9, Iroon Polytechneiou Street, 15773 Zografos, Athens, Greece

ARTICLE INFO

Article history:
Received 2 April 2007
Received in revised form 2 December 2007
Available online 26 June 2008

Communicated by S.K. Pal

PACS:
5.001
5.002

Keywords:
Fuzzy clustering
Fuzzy *c*-means
Finite mixture models
Student's-*t* distributions

ABSTRACT

In this paper, we propose a robust fuzzy clustering algorithm, based on a fuzzy treatment of finite mixtures of multivariate Student's-*t* distributions, using the fuzzy *c*-means (FCM) algorithm. As we experimentally demonstrate, the proposed algorithm, by incorporating the assumptions about the probabilistic nature of the clusters being dirived into the fuzzy clustering procedure, allows for the exploitation of the hard tails of the multivariate Student's-*t* distribution, to obtain a robust to outliers fuzzy clustering algorithm, offering increased clustering performance comparing to existing FCM-based algorithms. Our experimental results prove that the proposed fuzzy treatment of finite mixtures of Student's-*t* distributions is more effective comparing to their statistical treatments using EM-type algorithms, while imposing comparable computational loads.

1. Introduction

Fuzzy *c*-means (FCM) clustering has been successfully employed in a wide variety of fields and has demonstrated a high degree of adaptation to different data sets (Pedrycz, 2002; Kaufman and Rousseeuw, 1990; Dumitrescu et al., 2000). The FCM algorithm is an extension of the *k*-means algorithm (the hard *c*-means algorithm) and was first introduced in (Dunn, 1974) and generalized in (Bezdek, 1981), where the notion of the degree of fuzziness, taking values not less than one, is introduced. Fuzzy *c*-means (FCM) and its derivatives have been shown to be closely related to Gaussian mixture models (GMMs) in the algorithmic framework (Gath and Geva, 1989; Miyamoto and Mukaidono, 1997; Hathaway, 1986). Finite Gaussian mixture models (GMMs) are widely used in statistical pattern recognition and classification applications (McLachlan and Peel, 2000); they provide an appealing alternative to nonparametric density estimators for the approximation of unknown distributions, including distributions with multiple modes (Parzen, 1962). In (Hathaway, 1986), the expectation-maximization (EM) algorithm for GMMs is interpreted as a penalized version of the hard *c*-means clustering algorithm. In (Gath and Geva, 1989), an exponential distance for the FCM algorithm is defined to obtain a fuzzy clustering-based, FCM-type alternative to the EM algorithm for GMMs (Gath–Geva algorithm).

Despite the popularity of GMM-based or GMM-related parametric models, they have the major limitation of being extremely sensitive to outliers. Providing robustness to outlying data is crucial in many practical applications, where outliers comprise a significant proportion of the observable data, since they might affect severely the estimation of the model parameters as well as the model complexity, requiring additional components to capture the tails of the distribution (Kosinski, 1999). To mitigate this problem, in (Dave and Krishnapuram, 1997), the Noise Clustering (NC) method is proposed, which comprises the introduction of an extra cluster to represent the outliers. Even though the NC and related methods have become popular tools to provide protection to outliers in the context of fuzzy clustering techniques (Miyamoto and Alanzado, 2002; Honda and Ichihashi, 2004; Tran and Wagner, 1999), their heuristic nature remains a significant drawback.

Mixtures of Student's-*t* distributions (SMMs) have been proposed recently as an alternative to GMMs, providing the effective, non-heuristic means to mitigate the outlier vulnerability issues of GMMs (Shoham, 2002; Peel and McLachlan, 2000). The Student's-*t* distribution is a bell-shaped distribution with longer tails and one more parameter comparing to the normal distribution (the so-called degrees of freedom) and it tends to a normal distribution for big values of its degrees of freedom. This way, SMMs, exploiting the hard tails of the Student's-*t* distribution, provide a robust alternative to GMMs, allowing for the downweighting of the training data outliers, by means of a model-inherent, non-heuristic, sound statistical methodology.

* Corresponding author. Tel.: +30 2108823270; fax: +30 2107722569.
E-mail address: stchat@telecom.ntua.gr (S. Chatzis).

0167-8655/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2008.06.013

Motivated by the aforementioned observations, we introduce in this paper a fuzzy clustering scheme, that exploits the merits of mixtures of Student's-*t* distributions to offer high robustness against outliers in the context of fuzzy clustering techniques; we obtain the fuzzy mixture of Student's-*t* distributions model (FSMM) and we provide an efficient treatment of it using the FCM algorithm. We evaluate the efficacy of this novel fuzzy clustering algorithm using real and synthetic data, and we compare its classification performance with the performance of competing fuzzy and statistical clustering techniques.

The remainder of this paper is organized as follows: Section 2 begins with an overview of mixtures of Student's-*t* distributions; further the fuzzy mixture of Student's-*t* distributions model (FSMM) is introduced. In Section 3, we provide an effective treatment of the FSMM model using the FCM algorithm. In Section 4, the experimental evaluation of our model is conducted. The final section summarizes the results of this paper.

2. Model formulation

2.1. Student's-*t* mixture models

Let us consider a p-dimensional random sample $\mathbf{x}_1, \ldots, \mathbf{x}_n$. One way to broaden the normal distribution for potential outliers is to adopt the two-component Gaussian mixture density

$$(1-\epsilon)\mathcal{N}(\mathbf{x}_j; \boldsymbol{\mu}, \boldsymbol{\Sigma}) + \epsilon\mathcal{N}(\mathbf{x}_j; \boldsymbol{\mu}, k\boldsymbol{\Sigma}) \tag{1}$$

where, $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for a normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and k is large and ϵ is small, representing the small proportion of observations $\mathbf{x}_j$ that have a relatively large variance. The Gaussian scale mixture model (1) can be rewritten as

$$\int \mathcal{N}(\mathbf{x}_j; \boldsymbol{\mu}, \boldsymbol{\Sigma}/u)\mathrm{d}H(u) \tag{2}$$

where H is the probability distribution that places mass $1-\epsilon$ at the point $u = 1$ and mass ϵ at the point $u = \frac{1}{k}$. Let us replace H with the distribution of a χ^2 random variable on its degrees of freedom v; that is, we replace H with the random variable u distributed as

$$u \sim \mathcal{G}\left(\frac{v}{2}, \frac{v}{2}\right) \tag{3}$$

where $\mathcal{G}(\alpha, \beta)$ is the Gamma distribution, with probability density function (pdf) $\mathcal{G}(u; \alpha, \beta) = u^{\alpha-1}\frac{\beta^{\alpha}e^{-\beta u}}{\Gamma(a)}$. We have, hence, that,

$$p(\mathbf{x}_j; \boldsymbol{\mu}, \boldsymbol{\Sigma}, v) = \int \mathcal{N}(\mathbf{x}_j; \boldsymbol{\mu}, \boldsymbol{\Sigma}/u_j)\mathrm{d}\mathcal{G}\left(u_j; \frac{v}{2}, \frac{v}{2}\right) \tag{4}$$

where $p()$ is a generic notation for a probability function (pdf).

From (4), it follows that the observed data $\mathbf{x}_j (j = 1, \ldots, n)$ follows a multivariate Student's-*t* distribution, with mean vector $\boldsymbol{\mu}$, positive definite inner product matrix $\boldsymbol{\Sigma}$, and v degrees of freedom (Liu and Rubin, 1995), i.e.,

$$\mathbf{x}_j \sim t(\boldsymbol{\mu}, \boldsymbol{\Sigma}, v) \tag{5}$$

with pdf

$$t(\mathbf{x}_j; \boldsymbol{\mu}, \boldsymbol{\Sigma}, v) = \frac{\Gamma\left(\frac{v+p}{2}\right)|\boldsymbol{\Sigma}|^{-1/2}}{(\pi v)^{p/2}\Gamma(v/2)\{1+\delta(\mathbf{x}_j, \boldsymbol{\mu}; \boldsymbol{\Sigma})/v\}^{(v+p)/2}} \tag{6}$$

where, $\delta(\mathbf{x}_j, \boldsymbol{\mu}; \boldsymbol{\Sigma})$ is the squared Mahalanobis distance between $\mathbf{x}_j, \boldsymbol{\mu}$ with covariance matrix $\boldsymbol{\Sigma}$

$$\delta(\mathbf{x}_j, \boldsymbol{\mu}; \boldsymbol{\Sigma}) = (\mathbf{x}_j - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}_j - \boldsymbol{\mu}) \tag{7}$$

and $\Gamma(s)$ is the Gamma function, $\Gamma(s) = \int_0^{\infty} \mathrm{e}^{-t}z^{s-1}\mathrm{d}z$. From (4), it also follows that (Liu and Rubin, 1995)

$$\mathbf{x}_j|u_j \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}/u_j) \tag{8}$$

where the scalar u_j is distributed as

$$u_j \sim \mathcal{G}\left(\frac{v}{2}, \frac{v}{2}\right) \tag{9}$$

Then, the pdf of a c-component mixture of Student's-*t* distributions, with weights $w_1, \ldots, w_c$, is given by (Peel and McLachlan, 2000)

$$p(\mathbf{x}_j; \boldsymbol{\Theta}) = \sum_{i=1}^{c} w_i p(\mathbf{x}_j; \boldsymbol{\Theta}_i) \tag{10}$$

where

$$p(\mathbf{x}_j; \boldsymbol{\Theta}_i) = t(\mathbf{x}_j; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, v_i) \tag{11}$$

and the parameter vector $\boldsymbol{\Theta}_i$ consists of the elements of the $\boldsymbol{\mu}_i$, and $\boldsymbol{\Sigma}_i$, along with the degrees of freedom v_i, of the ith component distribution, and $\boldsymbol{\Theta} = \{\boldsymbol{\Theta}_i, w_i\}_{i=1}^{c}$.

2.2. Robust fuzzy clustering: the FSMM model

Each one of the component densities consisting a finite mixture of multivariate Student's-*t* distributions model, of the form (10), can be viewed as representing a cluster in the space of observable data. Concerning the a posteriori probability of the observation $\mathbf{x}_j$ deriving from the ith component distribution of the mixture model (10), $p(i|\mathbf{x}_j)$, it holds

$$0 \leqslant p(i|\mathbf{x}_j) \leqslant 1, \quad \sum_{i=1}^{c} p(i|\mathbf{x}_j) = 1 \tag{12}$$

Following Ruspini (1969), Eq. (12) implies that these clusters can be also considered as fuzzy sets; hence, the assumed mixture model can be regarded as defining a fuzzy c-partition R of the space of observations

$$R = \{r_{ij}\} \tag{13}$$

where $r_{ij}(i = 1, \ldots, c,\ j = 1, \ldots, n)$ represents the degree of the observation $\mathbf{x}_j$, belonging to the cluster represented by the ith component distribution. The function r_{ij} is called the fuzzy membership function and has the following properties

$$0 \leqslant r_{ij} \leqslant 1, \quad \sum_{i=1}^{c} r_{ij} = 1, \quad 0 < \sum_{j=1}^{n} r_{ij} < n \tag{14}$$

Under these considerations, the FSMM model is formulated. In the following section, we provide a (fuzzy) treatment of the FSMM model using the FCM algorithm.

3. FSMM parameters estimation: an FCM-based algorithm

The standard objective function minimized by an FCM-based fuzzy clustering algorithm (Gustafson and Kessel, 1979) is of the form

$$J_{\phi} \triangleq \sum_{i=1}^{c}\sum_{j=1}^{n} r_{ij}^{\phi} d_{ij} \tag{15}$$

where, d_{ij} is the dissimilarity between the jth data point and the ith cluster prototype, and $\phi \geqslant 1$ is a weighting exponent on each fuzzy membership function, r_{ij}, and is called the degree of fuzziness of the fuzzy clustering algorithm. Therefore, to conduct the FCM treatment of the FSMM model, we need first to define a suitable dissimilarity function, d_{ij}, taking into account the assumed probabilistic nature and the properties of the clusters (mixture components) being derived. Following Gath and Geva (1989), a suitable dissimilarity function, meeting these requirements, is

$$d_{ij} \triangleq -\log[w_i p(\mathbf{x}_j; \boldsymbol{\Theta}_i)] \tag{16}$$