Journal of Southeast University (English Edition) Vol. 29, No. 1, pp. 43 - 47 Mar. 2013 ISSN 1003—7985

Gaussian mixture model clustering with completed likelihood minimum message length criterion

Zeng Hong[1] Lu Wei[2] Song Aiguo[1]

([1] School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China)
([2] College of Engineering, Nanjing Agricultural University, Nanjing 210031, China)

Abstract: An improved Gaussian mixture model (GMM)-based clustering method is proposed for the difficult case where the true distribution of data is against the assumed GMM. First, an improved model selection criterion, the completed likelihood minimum message length criterion, is derived. It can measure both the goodness-of-fit of the candidate GMM to the data and the goodness-of-partition of the data. Secondly, by utilizing the proposed criterion as the clustering objective function, an improved expectation-maximization (EM) algorithm is developed, which can avoid poor local optimal solutions compared to the standard EM algorithm for estimating the model parameters. The experimental results demonstrate that the proposed method can rectify the over-fitting tendency of representative GMM-based clustering approaches and can robustly provide more accurate clustering results.

Key words: Gaussian mixture model; non-Gaussian distribution; model selection; expectation-maximization algorithm; completed likelihood minimum message length criterion

doi: 10.3969/j.issn.1003-7985.2013.01.009

The Gaussian mixture model (GMM) is commonly used as a basis for cluster analysis[1-2]. In general, the GMM-based clustering involves two problems. One is the estimation of parameters for the mixture models. The other is the model order selection for determining the number of components. The expectation-maximization (EM) algorithm is often used to estimate the parameters of the mixture model which fits the observed data. Popular model selection criteria in the literature include the Bayesian information criterion (BIC), Akaike's information criterion (AIC), the integrated likelihood criterion (ILC), etc.

However, most previous studies generally assume the Gaussian components for the observed data in the mixture model. If the true model is not in the family of the assumed ones, the BIC criterion tends to overestimate the correct model size regardless of the separation of the components. In the meantime, because the EM algorithm is a local method, it is prone to falling into poor local optima in such a case, leading to meaningless estimation. In order to approximate such a distribution more accurately, the feature weighted GMM, which explicitly takes the non-Gaussian distribution into account, is adopted in Refs. [3-8]. Nevertheless, the approaches in Refs. [3-8] assume that the data features are independent, which is often not the case for real applications. Based on the minimum message length (MML) criterion, Ref. [9] proposed an improved EM algorithm that can effectively avoid poor local optima. But we find that it still tends to select much more Gaussian components than necessary for fitting the data with uniform distribution, giving obscure evidence for the clustering structure of data.

We propose a novel method to address the model selection and parameter estimation problems in the GMM-based clustering method when the true data distribution is against the assumed one. In particular, we derive an improved model selection criterion for mixture models with an explicit objective of clustering. Furthermore, with the proposed criterion as the cost function, an improved EM algorithm is developed for estimating parameters. Ultimately, the proposed method is not only able to rectify the over-fitting tendency of some representative model selection criteria, but also able to avoid poor local optima of the EM algorithm.

1 Completed Likelihood of the Gaussian Mixture Model

Suppose that a D-dimensional sample follows a K-component mixture distribution, then the probability density function of $\boldsymbol{y}$ can be written as

$$p(\boldsymbol{y} \mid \boldsymbol{\Theta}) = \sum_{k=1}^{K} w_k p(\boldsymbol{y} \mid \boldsymbol{\theta}_k) \tag{1}$$

where w_k is the mixing probability for the k-th mixture

Received 2012-07-20.

Biography: Zeng Hong (1981—), male, doctor, lecturer, hzeng@seu.edu.cn.

Foundation items: The National Natural Science Foundation of China (No. 61105048, 60972165), the Doctoral Fund of Ministry of Education of China (No. 20110092120034), the Natural Science Foundation of Jiangsu Province (No. BK2010240), the Technology Foundation for Selected Overseas Chinese Scholar, Ministry of Human Resources and Social Security of China (No. 6722000008), and the Open Fund of Jiangsu Province Key Laboratory for Remote Measuring and Control (No. YCCK201005).

Citation: Zeng Hong, Lu Wei, Song Aiguo. Gaussian mixture model clustering with completed likelihood minimum message length criterion [J]. Journal of Southeast University (English Edition), 2013, 29(1): 43-47. [doi: 10.3969/j.issn.1003-7985.2013.01.009]

component with $0 \leqslant w_k \leqslant 1$ and $\sum_{k=1}^{K} w_k = 1$; $\boldsymbol{\theta}_k$ is the internal parameters describing the k-th mixture component. $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K; w_1, \ldots, w_K\}$ denotes the D_K dimensional vector describing the complete set of parameters for the mixture model. $p(\cdot \mid \boldsymbol{\theta}_k)$ defines the k-th Gaussian density. The GMM is typically an incomplete data structure model. N independent and identically distributed samples of the incomplete data Y are denoted as $Y = \{\boldsymbol{y}_1, \cdots \boldsymbol{y}_N\}$, and the complete data are $\bar{Y} = \{Y, Z\} = \{(\boldsymbol{y}_1, \boldsymbol{z}_1), \cdots (\boldsymbol{y}_N, \boldsymbol{z}_N)\}$, where the missing data are $Z = \{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N\}$, with $\boldsymbol{z}_n = \{z_{n1}, \cdots, z_{nK}\}$ being the binary label vector such that $z_{nk} = 1$ if and only if $\boldsymbol{y}_n$ belongs to the k-th mixture component and $z_{nk} = 0$ otherwise. Z is normally unknown, and it must be inferred from Y. The observed log-likelihood of $\boldsymbol{\Theta}$ for the incomplete data Y is

$$\log p(Y \mid \boldsymbol{\Theta}) = \sum_{n=1}^{N} \log \sum_{k=1}^{K} w_k p(\boldsymbol{y}_n \mid \boldsymbol{\theta}_k) \tag{2}$$

The completed log-likelihood of $\bar{Y}$ is

$$\begin{aligned} \log p(\bar{Y} \mid \boldsymbol{\Theta}) &= \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} \log(w_k p(\boldsymbol{y}_n \mid \boldsymbol{\theta}_k)) = \\ &\log p(Y \mid \boldsymbol{\Theta}) + \log p(Z \mid Y, \boldsymbol{\Theta}) = \\ &\sum_{n=1}^{N} \log \sum_{k=1}^{K} w_k p(\boldsymbol{y}_n \mid \boldsymbol{\theta}_k) + \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} \log p_{nk} \end{aligned} \tag{3}$$

where p_{nk} is the conditional probability of $\boldsymbol{y}_n$ belonging to the k-th component and can be computed as

$$p_{nk} = \frac{w_k p(\boldsymbol{y}_n \mid \boldsymbol{\theta}_k)}{\sum_{j=1}^{K} w_j p(\boldsymbol{y}_n \mid \boldsymbol{\theta}_j)} \tag{4}$$

In practice, the true parameter $\boldsymbol{\Theta}$ in Eqs. (2) and (3) is replaced using the maximum likelihood (ML) estimate $\hat{\boldsymbol{\Theta}}$, and then the completed log-likelihood is rewritten as

$$\log p(\bar{Y} \mid \hat{\boldsymbol{\Theta}}) = \sum_{n=1}^{N} \log \sum_{k=1}^{K} \hat{w}_k p(\boldsymbol{y}_n \mid \hat{\boldsymbol{\theta}}_k) + \sum_{n=1}^{N} \sum_{k=1}^{K} \hat{z}_{nk} \log p_{nk} \tag{5}$$

where

$$\hat{z}_{nk} = \begin{cases} 1 & \text{if } \arg\max_j p_{nj} = k \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

2 Clustering with Completed Likelihood Minimum Message Length (CL-MML) Criterion

2.1 Completed likelihood minimum message length criterion

The MML criterion defines a goodness measure for a model with an inherent bias towards simple models[10]. Based on the formulation of the MML criterion for a general density model in Ref. [9], the MML criterion for the GMM of the complete data $\bar{Y}$ can be written as follows:

$$\begin{aligned} \mathrm{MML}(K) = &-\log p(\hat{\boldsymbol{\Theta}}) - \log(\bar{Y} \mid \hat{\boldsymbol{\Theta}}) + \\ &\frac{1}{2} \log \mid \boldsymbol{I}_c(\hat{\boldsymbol{\Theta}}) \mid + \frac{D_k}{2}\left(1 + \log \frac{1}{12}\right) \end{aligned} \tag{7}$$

where $\log p(\bar{Y} \mid \hat{\boldsymbol{\Theta}})$ is given in Eq. (5); $\boldsymbol{I}_c(\hat{\boldsymbol{\Theta}}) = -E[\partial^2 \log p(\bar{Y} \mid \hat{\boldsymbol{\Theta}}) / \partial \hat{\boldsymbol{\Theta}} \partial \hat{\boldsymbol{\Theta}}^{\mathrm{T}}]$ is the expected Fisher information matrix associated with the complete data $\bar{Y}$, and $\mid \boldsymbol{I}_c(\hat{\boldsymbol{\Theta}}) \mid$ denotes its determinant. By differentiating $\log p(\bar{Y} \mid \hat{\boldsymbol{\Theta}})$ in Eq. (5), $\boldsymbol{I}_c(\hat{\boldsymbol{\Theta}})$ has a block-diagonal structure $\boldsymbol{I}_c(\hat{\boldsymbol{\Theta}}) = N$ block-diag $\{\hat{w}_1 \boldsymbol{I}^{(1)}(\hat{\boldsymbol{\theta}}_1), \cdots, \hat{w}_K \boldsymbol{I}^{(1)}(\hat{\boldsymbol{\theta}}_K), \boldsymbol{A}\}$ where $\boldsymbol{I}^{(1)}(\hat{\boldsymbol{\theta}}_k)$ is the Fisher matrix for a single observation produced by the k-th component, and $\boldsymbol{A}$ is the Fisher matrix of a multinomial distribution with $\mid \boldsymbol{A} \mid = (\hat{w}_1 \hat{w}_2 \cdots \hat{w}_K)^{-1}$. Since we have no knowledge about the parameters, we adopt the non-informative Jeffrey's priors as in Ref. [9], i.e.,

$$p(\hat{\boldsymbol{\Theta}}) = p(\hat{w}_1, \cdots, \hat{w}_K) \prod_{k=1}^{K} p(\hat{\boldsymbol{\theta}}_k) \tag{8}$$

where $p(\hat{\boldsymbol{\theta}}_k) \propto \sqrt{\mid \boldsymbol{I}^{(1)}(\hat{\boldsymbol{\theta}}_k) \mid}$, $p(\hat{w}_1, \cdots, \hat{w}_K) \propto \sqrt{\mid \boldsymbol{A} \mid}$. After substituting $p(\hat{\boldsymbol{\Theta}})$ and $\mid \boldsymbol{I}_c(\hat{\boldsymbol{\Theta}}) \mid$ into Eq. (7) and dropping the constant items, we obtain the explicit form of CL-MML for the GMM of the complete data as follows:

$$\begin{aligned} \text{CL-MML}(K) = &-\log(\bar{Y} \mid \hat{\boldsymbol{\Theta}}) + \left(-\sum_{n=1}^{N} \sum_{k=1}^{K} \hat{z}_{nk} \log p_{nk}\right) + \\ &\frac{M}{2} \sum_{k=1}^{K} \log \hat{w}_k + \frac{D_k}{2}(1 + \log N) \end{aligned} \tag{9}$$

where M is the number of parameters in each component. The first item on the right hand side of Eq. (9) emphasizes the goodness-of-fit of the candidate GMM. The third and the fourth items control the complexity of the GMM. Compared to the standard MML for the GMM of incomplete data in Ref. [9], CL-MML has an extra non-negative penalty item, i.e., the second item on the right side of Eq. (9). This item is essentially a measure of the K-component GMM to provide a relevant partition of the data Y. If the mixture components are well separated (i.e., p_{nk^*} is close to 1 with $\hat{z}_{nk^*} = 1$), such an item will be close to 0. But if the mixture components are poorly separated, such an item will have a large value, implying that such an unreasonable partition cannot discover the clustering structure of data. By minimizing this item, CL-MML prefers smaller K compared to the MML on the same data set. In other words, CL-MML is expected to be able to rectify the over-fitting tendency of the MML, favoring mixtures which lead to a clustering result of the data with the greatest evidence.

2.2 Estimation of GMM parameters

For the GMM, each component follows the Gaussian

distribution, i.e., $p(\boldsymbol{y} \mid \boldsymbol{\theta}_k) = G(\boldsymbol{y} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, where $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are the mean and the covariance matrix of the k-th Gaussian components. For a fixed model order K, we estimate the GMM parameters $\boldsymbol{\Theta}$ by an improved EM algorithm, with CL-MML in Eq. (9) as the cost function. The proposed EM algorithm alternatively applies the following two steps in the t-th iteration until convergence:

E-step: Compute the conditional expectation:

$$p_{nk}^{(t)} = \frac{\hat{w}_k^{(t)} p(\boldsymbol{y}_n \mid \hat{\boldsymbol{\theta}}_k^{(t)})}{\sum_{j=1}^{K} \hat{w}_j^{(t)} p(\boldsymbol{y}_n \mid \hat{\boldsymbol{\theta}}_j^{(t)})} \tag{10}$$

M-step: Update the parameters of the GMM by

$$\hat{w}_k^{(t+1)} = \frac{\max\left\{0, \left(\sum_{n=1}^{N} p_{nk}^{(t)}\right) - \frac{M}{2}\right\}}{\sum_{j=1}^{K} \max\left\{0, \left(\sum_{n=1}^{N} p_{nj}^{(t)}\right) - \frac{M}{2}\right\}} \tag{11}$$

$$\hat{\boldsymbol{\mu}}_k^{(t+1)} = \frac{\sum_{n=1}^{N} p_{nk}^{(t)} \boldsymbol{y}_n}{\sum_{n=1}^{N} p_{nk}^{(t)}} \tag{12}$$

$$\hat{\boldsymbol{\Sigma}}_k^{(t+1)} = \frac{\sum_{n=1}^{N} p_{nk}^{(t)} (\boldsymbol{y}_n - \hat{\boldsymbol{\mu}}_k^{(t)})(\boldsymbol{y}_n - \hat{\boldsymbol{\mu}}_k^{(t)})^{\mathrm{T}}}{\sum_{n=1}^{N} p_{nk}^{(t)}} \tag{13}$$

In Eq. (11), $\sum_{n=1}^{N} p_{nk}^{(t)}$ can be viewed as the evidence for the k-th component from the data points. Then according to Eq. (11), when one of the components becomes too weak, namely it is not supported by the data, it will then be driven into extinction. Such a modification to the standard EM can be expected to suppress spurious solutions.

3 Experiments

We present experimental results to illustrate the effectiveness of CL-MML for GMM-based clustering (denoted as GMM + CL-MML), compared to that of BIC (denoted as GMM + BIC), MML (denoted as GMM + MML), as well as the method utilizing the feature-weighted GMM and the integrated likelihood criterion (FWGMM + ILC) for clustering[3].

3.1 Synthetic data

We consider a synthetic 2D data set where data from each cluster follow the uniform random distribution:

$$u_{\mathrm{r}}(y_1, y_2) = \begin{cases} \dfrac{1}{(r_2 - r_1)(r_4 - r_3)} & r_1 \leqslant y_1 \leqslant r_2; r_3 \leqslant y_2 \leqslant r_4 \\ 0 & \text{otherwise} \end{cases}$$

where $\boldsymbol{r} = \{r_1, r_2, r_3, r_4\}$ are the parameters of the distribution. 1 000 data points are generated using a 5-component uniform mixture model. Its parameters are as follows:

$$w_1 = 0.1, w_2 = w_4 = w_5 = 0.2, w_3 = 0.3$$
$$\boldsymbol{r}_1 = \{-1.89, 4.07, 4.89, 7.94\}$$
$$\boldsymbol{r}_2 = \{1.11, 5.11, 2.47, 3.53\}$$
$$\boldsymbol{r}_3 = \{5.17, 6.53, 2.77, 5.77\}$$
$$\boldsymbol{r}_4 = \{4.31, 6.49, 6.29, 6.71\}$$
$$\boldsymbol{r}_5 = \{5.58, 8.42, -0.77, 2.23\}$$

The Gaussian components are adopted to fit such a uniform mixture data set, for which the true distribution models are very different from the assumed ones. The models with the number of components K varying from 1 to $K_{\max}$, a number that is considered to be safely larger than the true number (i.e., 5), are evaluated. $K_{\max}$ is set to be 30 in this case. We evaluate these methods by the accuracy in estimating the model order and structure. Tab. 1 illustrates the number of times that each order is selected over 50 trials. Fig. 1 shows typical clustering results by these four methods.

It can be observed that for such a data set, the GMM + BIC approach not only fails to yield a good estimation of model order (see Tab. 1), but also leads to a meaningless mixture model by the standard EM (see Fig. 1(a)). Although the MML criterion generates a GMM which fits the data well, it suffers from severe over-fitting as shown in Fig. 1(b) and Tab. 1. Since the features are assumed to be independent in FWGMM, it also tends to select more components in order to approximate the distribution of data accurately (see Fig. 1(c) and Tab. 1).

Tab. 1 Number of times for selected model orders over 50 trials on synthetic data

Model order	GMM + BIC	GMM + MML	FWGMM + ILC	GMM + CL-MML
5 (true)	0	0	0	31
6	22	0	0	13
7	13	2	0	6
8	15	3	0	0
9	0	3	0	0
10	0	4	0	0
11	0	5	2	0
12	0	5	6	0
13	0	7	12	0
14	0	9	15	0
15	0	12	11	0
16	0	0	4	0

In contrast, due to the introduction of an extra penalty to the MML criterion, the proposed CL-MML criterion-based GMM clustering favors much fewer but more "powerful" components which successfully detect the clusters. The clustering result in a typical successful trial of CL-MML is shown in Fig. 1(d).

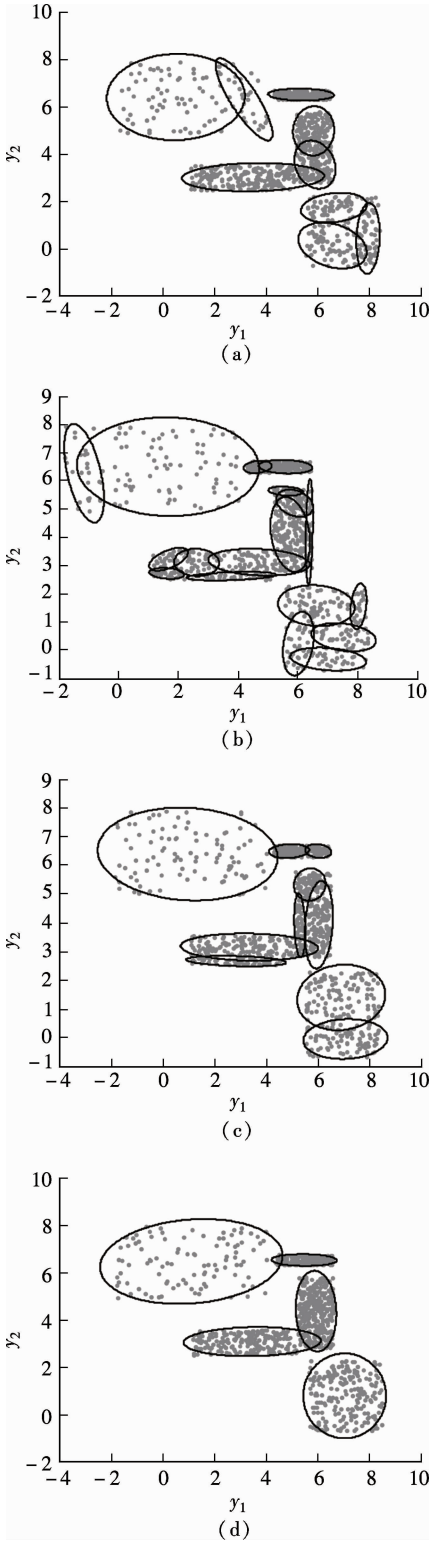

Fig. 1 Typical clustering results of different methods on the synthetic data. (a) GMM + BIC; (b) GMM + MML; (c) FWGMM + ILC; (d) GMM + CL-MML

3.2 Real data

We also measure performance on four real-world data sets from the UCI repository. The number of classes, the number of samples and the dimensionality of each data set are summarized in Tab. 2. For each data set, we randomly split the data 50 times into training and test sets. Training sets are created from 50% of the overall data points. We do not use any label in the training stage. $K_{\max}$ is still set to be 30. After model learning, we label each component by majority vote using the class labels provided for the test data, and we measure the test set classification accuracy as the matching degree between such obtained labels and the original true labels. The means and the standard deviations of the classification accuracy, as well as the number of components for each data set, over 50 trials are summarized in Tab. 2. The best results are marked in bold.

Tab. 2 Comparison of different clustering approaches on real data sets

Data set	Method	Accuracy	K
Heart	GMM + BIC	0.645 9 ± 0.063 5	2.4 ± 0.89
	GMM + MML	0.672 5 ± 0.058 4	2.9 ± 0.89
	FSGMM + ILC	0.674 0 ± 0.096 5	4.2 ± 2.28
	GMM + CL-MML	**0.701 4 ± 0.042 9**	**2.4 ± 0.87**
Zernike	GMM + BIC	0.482 4 ± 0.038 6	21.6 ± 2.46
	GMM + MML	0.671 4 ± 0.034 1	13.0 ± 1.15
	FSGMM + ILC	0.680 5 ± 0.002 5	14.0 ± 0.00
	GMM + CL-MML	**0.701 6 ± 0.057 0**	**11.2 ± 1.33**
Landsat	GMM + BIC	0.664 0 ± 0.064 0	15.0 ± 1.78
	GMM + MML	0.790 0 ± 0.012 8	11.0 ± 1.0
	FSGMM + ILC	0.636 8 ± 0.057 4	13.8 ± 4.76
	GMM + CL-MML	**0.813 4 ± 0.010 5**	**9.1 ± 1.79**
Image	GMM + BIC	0.618 3 ± 0.035 0	19.9 ± 3.78
	GMM + MML	0.826 4 ± 0.012 5	19.2 ± 0.83
	FSGMM + ILC	0.854 0 ± 0.023 8	23.1 ± 1.74
	GMM + CL-MML	**0.876 1 ± 0.015 8**	**17.7 ± 1.41**

Several trends are apparent. First, the numbers of components determined by the proposed method are generally less than those by the compared counterparts. This may be due to the reason that the distribution of a real data set often does not strictly follow the Gaussian mixture model, and most GMM-based clustering approaches tend to generate more components than necessary in order to better fit the data. However, it is found that the CL-MML can rectify the over-fitting tendency of the compared methods under such circumstances. This can be explained by the reason that it takes the separation among components into account. Secondly, the proposed method yields the most accurate results among all the approaches on these four data sets. This justifies that the proposed approach can estimate the GMM parameters more properly than the compared ones.

4 Conclusion

In this paper, by taking the capability of the candidate GMM to provide a relevant partition to the data into account, an improved GMM-based clustering approach is developed for the difficult scenario where the true distri-

bution of data is against the assumed GMM. The experimental results show that the proposed method is not only able to rectify the over-fitting tendency of the compared methods for performing the model selection, but also able to obtain higher clustering accuracy compared to the existing methods.

References

[1] Zeng H, Cheung Y M. Feature selection and kernel learning for local learning based clustering [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(8):1532 - 1547.

[2] Jain A K. Data clustering: 50 years beyond *K*-means [J]. *Pattern Recognition Letters*, 2010, **31**(8):651 - 666.

[3] Bouguila N, Almakadmeh K, Boutemedjet S. A finite mixture model for simultaneous high-dimensional clustering, localized feature selection and outlier rejection [J]. *Expert Systems with Applications*, 2012, **39**(7): 6641 - 6656.

[4] Law M H C, Figueiredo M A T, Jain A K. Simultaneous feature selection and clustering using mixture models [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(9):1154 - 1166.

[5] Markley S C, Miller D J. Joint parsimonious modeling and model order selection for multivariate Gaussian mixtures [J]. *IEEE Journal of Selected Topics in Signal Processing*, 2010, **4**(3):548 - 559.

[6] Li Y, Dong M, Hua J. Localized feature selection for clustering [J]. *Pattern Recognition Letters*, 2008, **29**(1):10 - 18.

[7] Allili M S, Ziou D, Bouguila N, et al. Image and video segmentation by combining unsupervised generalized Gaussian mixture modeling and feature selection [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2010, **20**(10):1373 - 1377.

[8] Fan W, Bouguila N, Ziou D, Unsupervised hybrid feature extraction selection for high-dimensional non-Gaussian data clustering with variational inference [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2012, in press.

[9] Figueiredo M A F, Jain A K. Unsupervised learning of finite mixture models [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(3): 381 - 396.

[10] Wallace C S, Dowe D L. MML clustering of multi-state, Poisson, von Mises circular and Gaussian distributions [J]. *Statistics and Computing*, 2000, **10**(1): 73 - 83.

基于完整似然最短信息长度准则的高斯混合模型聚类

曾 洪[1] 卢 伟[2] 宋爱国[1]

([1] 东南大学仪器科学与工程学院, 南京 210096)

([2] 南京农业大学工学院, 南京 210031)

摘要:针对数据真实的概率分布不符合事先假设的高斯混合模型的情形,提出了一种鲁棒的基于高斯混合模型的聚类方法. 首先,提出了一种新的模型选择准则,即完整似然最短信息长度准则. 该准则不仅能衡量模型对数据的拟合优度,还能度量该模型对数据分组的性能. 然后,将该准则作为聚类的代价函数,提出了一种新的期望最大化算法来估计模型参数. 与标准的期望最大化算法相比,新算法能较好地避免不理想的局部最优解. 实验结果表明:当数据概率分布模型不符合假设的高斯混合模型时,所提方法可克服现有的基于高斯混合模型聚类方法过拟合的缺点,鲁棒地得到准确的聚类结果.

关键词:高斯混合模型;非高斯分布;模型选择;期望最大化算法;完整似然最短信息长度准则

中图分类号:TP181