2012 International Conference on Image and Information Processing (ICIIP 2012)
IPCSIT vol. 46 (2012) © (2012) IACSIT Press, Singapore

Image Database Categorization based on a Novel Possibilistic Clustering and Feature Weighting Algorithm

Mohamed Maher Ben Ismail [+]

King Saud University

Abstract. We propose a novel image database categorization approach using robust unsupervised learning of finite generalized Dirichlet mixture models with feature weighting. The proposed algorithm, called Robust and Unsupervised Learning of Finite Generalized Dirichlet Mixture Models and Feature Weighting (RULe_GDM_FW), exploits a property of the Generalized Dirichlet distributions that transforms the data to make the features independents and follow Beta distributions. Then, it searches for the optimal relevance weight for each feature within each cluster. This property makes RULe_GDM_FW suitable for noisy and high-dimensional feature spaces. RULe_GDM_FW minimizes one objective function that combines learning two membership functions, the distribution parameters, and relevance weights for each feature within each distribution. These properties make RULe_GDM_FW suitable for noisy and high-dimensional feature spaces. RULe_GDM_FW is used to categorize a collection of color images. The performance of RULe_GDM_FW is illustrated and compared to similar algorithms.

Keywords: unsupervised learning, mixture models, feature weighting, possibilistic approach.

1. Introduction

Modern advances in technology and the accelerated development of the Internet have granted society the ability to capture, store, and view an everyday increasing number of digital images. Web communities are now a prevalent staple on the Internet and through sites such as Flickr [1], help demonstrate the scale of digital imagery available, and point toward the social and practical impact of viewing and interacting with these images. Navigation through these photo collections and finding photos of interest is naturally difficult due to their large sizes and to the computer's inability to capture the semantic meaning of images. This problem is known as the semantic gap [2]. To address this limitation, image database categorization based on the image content has become an active research topic [3]. One of the most used approaches is based on clustering techniques. Its goal is to categorize the image database into meaningful clusters based on their content. The resulting clusters are then used to index the image database and to reduce the search space during the retrieval process. They could also be used to help the user navigate through the database. Another application would involve using the clusters' representatives to create page zero in a Content Based Image Retrieval (CBIR) by example system. In this mode, instead of displaying a random sample of images, the system starts by displaying the representatives of the clustered database. Since these representatives provide a good summary of the database, the user would have a good idea about the content of the database before initiating the querying process.

Over the past few years, various clustering approaches have been applied to the problem of image database categorization and have proven to be effective [5]. The resulting clusters have been used to index the image database and reduce the search space during the retrieval process. They have also been used to help the user navigate through the database. However, despite recent progress, image database categorization remains a difficult research task. The problem is more acute when the high dimensional feature space,

[+] Corresponding author. Tel.: + 00966014695223.
E-mail address: mbenismail@ksu.edu.sa.

encoding the low level image features, is corrupted by noise. Moreover, most existing categorization algorithms assume that the data can be modelled by a mixture of Gaussian distributions [4]. However, this assumption rarely holds in a high-dimensional space and can affect the clustering performance. Generalized Dirichlet (GD) mixture has been adopted as a good alternative. However, noise points and outliers can drastically affect the estimate of such model parameters and, hence, the final clustering partition. Recently, we proposed a robust approach for GD mixture parameter estimation and data clustering [6] that uses possibilistic membership functions to reduce the effect of noise and outliers. Even though this approach has proved to be more robust and effective in image database categorization, it has two main limitations. First, when modeling an image collection that involves high dimensional data, not all of the extracted features are expected to be equally relevant to all clusters and some of them may be noisy or redundant. In fact, irrelevant features can bias the estimated parameters and subsequently, compromise the learned categories. To overcome these limitations, we propose an image categorization approach that relies on a Robust Unsupervised Learning of Finite Generalized Dirichlet Mixture Models and Feature Weighting (RULe_GDM_FW). This algorithm minimizes one objective function that combines learning of the membership functions, the distribution parameters, and feature relevance weights for each distribution.

2. Possibilistic Clustering and Feature Weighting Based on Robust Modeling Of Finite Generalized Dirichlet Mixture

Let $\boldsymbol{Y} = (\boldsymbol{Y_1}, \boldsymbol{Y_2}, \dots \boldsymbol{Y_N})$ be a set of N points where $\mathbf{Y}_i \in R^d$. We assume that $\mathbf{Y}$ is generated by a mixture of M GD distributions, i.e.,

$$p(\mathbf{Y}|\Theta^*)=\sum_{j=1}^{M}\mathbf{p}_j p(\mathbf{Y}|\boldsymbol{\Theta}_j^*)=\sum_{j=1}^{M}\mathbf{p}_j\prod_{l=1}^{d}p_b(\mathbf{X}_{il}|\Theta_{jl}^*). \tag{1}$$

In (1), $\boldsymbol{\Theta}^*=(\boldsymbol{\Theta}_1^*,\boldsymbol{\Theta}_2^*,\dots,\boldsymbol{\Theta}_M^*,\mathbf{p}_1,\dots,\mathbf{p}_M)$, is the parameter vector of the j^{th} GD component and p_j are the mixing weights where $\sum_j p_j=1 \quad for\ j=1..M$

In the mixture-based clustering, each $\mathbf{Y}_i$ is assigned to each component, j, with a posterior probability $p(j|\mathbf{Y}_i)\propto\mathbf{p}_j p(\mathbf{Y}_i|\Theta_j^*)$. The GD distribution has a desirable property that allows the factorization of the posterior probability as $p(j|\mathbf{Y}_i)\propto\mathbf{p}_j\prod_{l=1}^{d}p_b(\mathbf{X}_{il}|\Theta_{jl}^*)$, $\mathbf{X}_{i1}=\mathbf{Y}_{i1}$, and $\mathbf{X}_{il}=\frac{\mathbf{Y}_{il}}{1-\sum_{k=1}^{l-1}\mathbf{Y}_{ik}}$ for $l>1$, and $p_b(\mathbf{X}_{il}|\Theta_{jl}^*)$ is a Beta distribution with $\Theta_{jl}^*=(\alpha_{jl}^*,\beta_{jl}^*), l=1,\dots,d.$ In other words, the clustering structure underlying $\mathbf{Y}$ is the same as that underlying $\mathbf{X}=(\mathbf{X}_1,\mathbf{X}_2,\dots,\mathbf{X}_N)$ governed by $p(\mathbf{X}_i|\Theta^*)=\sum_{j=1}^{M}\mathbf{p}_j\prod_{l=1}^{d}p_b(\mathbf{X}_{il}|\Theta_{jl}^*)$ with conditionally independent features $\mathbf{X}$. Thus, the problem of estimating the parameters of the Generalized Dirichlet mixture of $\mathbf{Y}$ is reduced to the estimation of the Beta mixture of **X.**

The objective function formulation proposed in our previous work [6] can be optimized to yield the parameters of the M distributions that best fit the data. However, in our approach we do not expect all d features to be relevant for all M components. Instead, we propose a modification to the objective function in [6] to learn the relevant features for each component. We consider the l^{th} feature as irrelevant to cluster j if its distribution is independent of the corresponding component, i.e., if it follows a common density denoted by $q(\mathbf{X}_l/\lambda_l)$.

Let $\varphi_j=(\varphi_{j1},\dots,\varphi_{jd})$ be a set of binary parameters, such that $\varphi_{jl}=1$ if feature l is relevant to cluster j and $\varphi_{jl}=0$ otherwise. Using an approach similar to the one in [9], we treat φ_{jl} as a missing variable and define the probability that the l^{th} feature is relevant to cluster j as the feature saliency $\varrho_{jl}=P(\varphi_{jl}=1)$. Thus, the likelihood function in [6] becomes

$$p(\mathbf{X}_i|\Theta^*)=\sum_{j=1}^{M}\mathbf{p}_j\prod_{l=1}^{d}(\varrho_{jl}p_b(\mathbf{X}_{il}|\Theta_{jl}^*)+(1-\varrho_{jl})q(\mathbf{X}_{il}|\lambda_l)). \tag{2}$$

where $\Theta^{*}=\{\{\mathbf{p}_j\},\{\Theta_{jl}\},\{\lambda_l\},\{\varrho_{jl}\}\}$ includes all model parameters. In our approach, irrelevant features are approximated by one distribution, q, that is common to all clusters. In particular, we consider the distribution of an irrelevant feature as a Beta distribution that is independent of the clusters.

By integrating the feature selection model in (2) into the objective function in [6], we minimize the following objective function

$$J = -\sum_{j=1}^{M}\sum_{i=1}^{N}(\mathbf{u}_{ji}^{m}log(\mathbf{p}_j)+\mathbf{u}_{ji}^{m}\sum_{l=1}^{d}\log\left[\varrho_{jl}p_b(\mathbf{X}_{il}|\Theta_{jl}^{*})+(1-\varrho_{jl})q(\mathbf{X}_{il}|\boldsymbol{\lambda}_l)\right])+\sum_{j=1}^{M}\eta_j\sum_{i=1}^{N}(1-\mathbf{u}_{ji})^{m}, \quad (3)$$

subject to the membership constraint. Setting the gradient of J with respect to $\mathbf{u}_{ji}$ to zero yields the following necessary condition to update the possibilistic membership:

$$\mathbf{u}_{ji}=\left[1-\left(\frac{\log\left[\mathbf{p}_j\prod_{l=1}^{d}(\varrho_{jl}p_b(\mathbf{X}_{il}|\Theta_{jl}^{*})+(1-\varrho_{jl})q(\mathbf{X}_{il}|\boldsymbol{\lambda}_l))\right]}{\eta_j}\right)^{\frac{1}{m-1}}\right]^{-1}. \quad (4)$$

Setting $\frac{\partial J}{\partial\varrho_{jl}}$ to zero, and assuming that ϱ_{jl} does not change significantly from iteration (t)to iteration (t+1)we obtain the following update equation for ϱ_{jl}:

$$\varrho_{jl}^{(t+1)}=\sum_{i=1}^{N}\mathbf{u}_{ji}^{m}\frac{p_b(\mathbf{X}_{il}|\Theta_{jl}^{*})-q(\mathbf{X}_{il}|\boldsymbol{\lambda}_l)}{p_b(\mathbf{X}_{il}|\Theta_{jl}^{*})-q(\mathbf{X}_{il}|\boldsymbol{\lambda}_l)+\frac{q(\mathbf{X}_{il}|\boldsymbol{\lambda}_l)}{\varrho_{jl}^{(t)}}}. \quad (5)$$

It can also be shown that minimizing J with respect to the GD mixture weights yields

$$\mathbf{p}_j=\frac{\sum_{i=1}^{N}\mathbf{u}_{ji}^{m}}{\sum_{j=1}^{M}\sum_{i=1}^{N}\mathbf{u}_{ji}^{m}}. \quad (6)$$

The presence of Gamma functions in the Beta distribution prevents obtaining a closed-form solution for Θ_{jl} that minimizes J. Thus, to minimize $J$with to Θ and **λ**, we use the gradient descent method estimate Θ and **λ** iteratively using

$$\Theta_{jl}^{(t+1)}=\Theta_{jl}^{(t)}-\xi_1\frac{\partial J}{\partial\Theta_{jl}}, \qquad \lambda_l^{(t+1)}=\lambda_l^{(t)}-\xi_2\frac{\partial J}{\partial\lambda_l} \quad (7)$$

The resulting RULe_GDM_FW algorithm is summarized below.

Algorithm 1RULe_GDM_FW Algorithm

Begin
Fix m∈]1,∞);
Let M be the number of clusters.
Initialize **U** ,Θ, λ, ϱ, and η,
Repeat
Compute log $\left[p_b(\mathbf{X}_{il}|\Theta_{jl})\right]$
Update Θ and λ for few iterations using (7);
Update the partition matrix **U** using (4);
Update the mixture weights **p** using (6);
Until (U stabilize)
End

3. Experimental results

To illustrate the ability of RULe_GDM_FW to model high dimensional data and cluster real data sets, we use it to categorize an image database. We use a subset of 3000 color images from the COREL image collection. This subset includes 30 categories with 100 images in each one. Each image in the collection is characterized by five generic MPEG-7 descriptors [13]. These features are the Color Structure Descriptor

(CSD) (32 dim), the Scalable Color Descriptor (SCD) (32 dim), the RGB Color Histogram (32 dim), the Wavelet Texture Descriptor (WTD) (20 dim), and the Edge Histogram Descriptor (EHD) (5 dim).

To assess the performance of RULe_GDM_FW, we assume that the ground truth is known and we compute the overall accuracy of the partition as the average of the clusters rates weighted by the clusters cardinality. In addition, we use the Jaccard coefficient and Folkes–Mallows index [8] to compare each generated partition to the ground truth partition.

We set the fuzzyfier$m$ to 1.1 and estimate the scale parameter η_j for each cluster j as suggested in [11]. The results were compared with those obtained using the basic FCM [10], and PCM [11] algorithms and the method proposed in [12]. We run each algorithm 30 times and compute its average classification accuracy and standard deviation. Since these algorithms require the specification of the number of clusters, first, we set the initial number of clusters $C$to 30 and measure the performance of the different algorithms as shown in Table 1. All methods achieved reasonable performance with RULe_GDM_FW outperforming the method in [12].

Table. 1: Comparison of the FCM [11], PCM [12], the method proposed in [14], and RULe_GDM_FW

on the COREL data

	PCM	FCM	Method in [12]	RULe_GDM_FW
Accuracy	41.71±.013	46.56±.01	50.25±.027	56.14±.01
Folkes-Mallows	28.7±.012	30.2±.01	42.5±.019	50.3±.007
Jaccardcoef	12.9±.012	9.1±.01	19.8±.02	23.7±.005

As it can be seen, RULe_GDM_FW and the method in [12] outperform the PCM and FCM algorithms with respect to all performance measures. This confirms what has been reported in the literature [7] that Generalized Dirichlet distributions are more suitable to model high dimensional data than Gaussian distributions. In fact, by analysing the content of the different clusters, we observed that the FCM splits many categories over several clusters. This is because these categories have large intra-cluster color variations and do not have spherical shapes in the high dimensional feature space.

To illustrate this advantage, in Fig. 1, we show three sample clusters obtained using RULe_GDM_FW and the corresponding relevance weights assigned to the 32 dimensions of the RGB color histogram features (The remaining features and their weights cannot be easily illustrated). In Fig. 1(a), we show 25 representative images from the "horse" cluster. For this cluster, RULe_GDM_FW assigned relatively higher relevance weights to the histogram features that correspond to the green, brown, and white bins (Fig. 1(b)). This matches the visual appearance of the images in this cluster where the green background that represents the grass, and the presence of brown and white objects, corresponding to horses, characterize the images within this cluster. Similarly, in Fig. 1(c), we display representative images from cluster #21 that corresponds to "ski scene". From Fig. 1(d), it can be seen that for this cluster, RULe_GDM_FW identified white, gray, and blue bins, as the most relevant features. Clearly, one set of feature weights is not sufficient to capture the variations among the clusters.

The second advantage of RULe_GDM_FW that helped it to outperform the method in [12], is its robustness against noise and outliers. The image collection includes few images that are not visually similar to the remaining images in their categories. RULe_GDM_FW was able to identify most of these images and assign low possibilistic membership to them in all clusters. In Fig.2, we display samples of these noisy images. For instance, the "butterfly" image (#4), with black background, is quite different than the majority of "butterfly" images because has a different background than the rest of the images in this category.

Fig. 2: Sample of images detected as noise points by RULe_GDM_FW

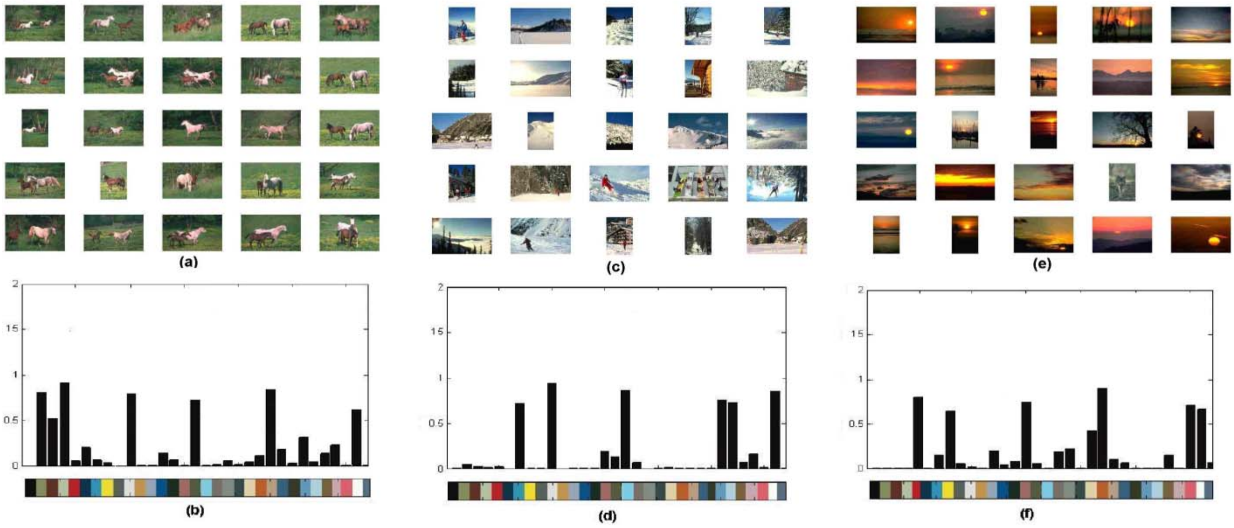

Fig. 1: Sample clusters and Relevance weights assigned to the 32-bins color histogram features by RULe_GDM_FW. (a) 25 samples from the "horse" cluster (category #4). (b) Features weights assigned to RGB histogram features for the "horse" cluster. (c) 25 samples from the"ski scene" cluster (category #21), (d) Features weights assigned to RGB histogram features for the "ski scene" cluster (category #21). (e) 25 samples from the "sunset" cluster (category #22). (f) Features weights assigned to RGB histogram features for the "sunset" cluster.

4. Conclusions

We proposed an image database categorization approach using clustering and feature weighting algorithm based on robust modeling of the Generalized Dirichlet (GD) finite mixture. Robustness to noisy and irrelevant features is achieved by two main features. First, transforming the data to make the features independent and follow Beta distribution, and learning optimal relevance weight for each feature within each cluster. Second, by learning two types of membership degrees. The first membership is a posterior probability that indicates the degree to which each sample fits the estimated distribution. The second membership represents the degree of "typicality" and is used to indentify and discard noise points. We have shown that RULe_GDM_FW outperforms similar existing algorithms and provides a richer description of the image collection.

5. References

[1] Ludicorp, Flickr, http://www.Flickr.com.

[2] A. W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, Content-based image retrieval at the end of the early years, in IEEE Trans. PAMI, 2000.

[3] G. Sheikholeslami, W.Chang, and Aidong Zhang, Sem- query: Semantic clustering and querying on

[4] G. McLachlan and D. Peel, Finite Mixture Models. New York: John Wiley & Sons, 2000.

[5] H. Frigui, N. Boujemaa, and S. A. Lim, "Unsupervised clustering and feature discrimination with application to image database categorization", in Proceedings of NAFIPS, 2001.

[6] M.M.Ben Ismail, H.Frigui: "Possibilistic Clustering based on Robust Modeling of Finite Generalized Dirichlet Mixture", ICPR 2010.

[7] N. Bouguila and D. Ziou, "A Hybrid SEM Algorithm for High- Dimensional Unsupervised Learning Using a

Finite Generalized Dirichlet Mixture", IEEE Trans. Image Processing, 2006.

[8] H.Frigui, C.Hwang, and F.C. Rhee, "Clustering and aggregation of relational data with applications to image database categorization", Pattern Recognition, 2007.

[9] M.H.C. Law, M.A.T. Figueiredo, and A.K. Jain, "Simultaneous Feature Selection and Clustering Using Mixture Models," IEEE Trans. Pattern Analysis and Machine Intelligence, Sept. 2004.

[10] J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum Press, New York, 1981.

[11] R. Krishnapuram and J. Keller, "A possibilistic approach to clustering," IEEE TFS, vol. 1, pp. 98-110, May 1993.

[12] S.Boutemedjet, N.Bouguila, D.Ziou: "A Hybrid Feature Extraction Selection Approach for High-Dimensional Non-Gaussian Data Clustering". IEEE Trans. PAMI. 2009.

[13] P. Salembier, B.S. Manjunath, T. Sikora, "Introduction to MPEG 7: Multimedia Content Description Language", Wiley, New York, 2002.