ANNALS OF "DUNAREA DE JOS" UNIVERSITY OF GALATI
MATHEMATICS, PHYSICS, THEORETICAL MECHANICS
FASCICLE , YEAR X (XLI) 2018, No. 1

Article DOI: https://doi.org/10.35219/ann-ugal-math-phys-mec.2018.1.14

FIRST ORDER STATISTICS-BASED FEATURES SELECTION FOR CLUSTERING USING GAUSSIAN MIXTURE MODEL

Lucian Traian Dimitrievici[1,4], Simona Moldovanu[1,2,3], Luminița Moraru[1,*]

[1]*Faculty of Sciences and Environment, Modelling & Simulation Laboratory, "Dunarea de Jos "University of Galati, 47 Domneasca St., 800008, Romania*
[2]*Department of Computer Science and Engineering, Electrical and Electronics Engineering, Faculty of Control Systems, Computers, "Dunarea de Jos" University of Galati, Romania*
[3]*"Dumitru Motoc" High School, 15 Milcov St., 800509, Galati, Romania*
[4]*"Mihail Kogalniceanu" High School, 131B Brailei ST, 800379, Galati, Romania*

Abstract
In this study, a total of 60 brain DW-MRI images belonging to three patients showing health, multiple hemorrhagic areas in the left temporal lobe and ischemic stroke pathologies were analyzed with a Gaussian Mixture Model (GMM) for classification-based clustering. To optimize the clustering analysis and to investigate the performance of the classification, various first order statistical based features such as entropy, energy, kurtosis and skewness were used as distinguishable features. Also, the mixing proportion of the chosen components of GMM were investigated. Experiments are performed on DW-MRI images acquired with two magnetic field gradient values or b-value ($b = 500\ s/mm^2$ and $b = 1000\ s/mm^2$) and for non diffusion-weighted images ($b = 0\ s/mm^2$). The experimental results show that GMM classifier together with kurtosis and skewness first order statistics-based features better discriminate between studied classes and can be applied for components identification successfully.

Keywords: Gaussian Mixture Model GMM, DW-MRI, kurtosis, skewness

1. INTRODUCTION

The classification based on the maximum likelihood usually cover many clustering algorithms and, in this paper, we limit our approach to Gaussian distributions. Among they, mixture models allow for probabilistic modelling of data, being used in pattern recognition, machine learning, computer vision, and signal processing research areas [1-5]. If more information we gather about each dataset featured by m variables, the clustering method is expected to better perform.

The clustering-based methods used various models for clusters in order to optimize the correlation between the analyzed data and the model. Each cluster is characterized by a parametric distribution, in our case a Gaussian distribution is used, and the experimental data set is modelled by a mixture of these distributions. Bouveyron et al. [6] discussed on a proper Gaussian mixture family in order to simultaneously cluster the data and reduce their dimension. On the other hand, Law et al. [7] proposed a set of real-valued quantity for each variable in the mixture of the distribution with the final goal to evaluate the importance of the variable in the clustering process. The main advantage of model-based clustering consists of the possibility to evaluate the number of mixture components and to assess the role of each variable in a clustering process.

The individual distributions that are combined to form the mixture distribution are so called ‘mixture components’, and their associated probabilities (or weights) are known as ‘the mixture weights’ [8]. From the perspective of clustering optimization and classification accuracy, a better approach is to split the Gaussian mixture in their components [9].

2. EXPERIMENTAL METHODS

2.1 GAUSSIAN MIXTURE MODEL (GMM)

Gaussian mixture is a model-based approach in which the analyzed experimental classes are subjected to various distributions, such as the Gaussian distribution, whose parameters are optimized to match the experimental data. A Gaussian model assumes that the experimental data follows a Gaussian distribution (characterized by a mean and a standard deviation) and the use of a larger number of Gaussian distributions in a combination model allows the efficient characterization of complex databases. Let be a random variable X that supports the numerical values x and its probability is denoted P{}. The distribution function is defined as the probability that the realization of a particular situation of the random variable X is less than or equal to x [10]:

$$F_x(x) = P\{X \leq x\} \quad \text{and} \quad 0 \leq F_x(x) \leq 1$$

The probability density function is defined as follows:

$$f_X(x) = \frac{dF_X(x)}{dx}, f_X(x) \geq 0 \tag{1}$$

The probability can be seen as:

$$P\{X \leq x\} = F_x(x) = \int_{-x}^{x} f_X(t)dt$$

$$P\{x_1 < X \leq x_2\} = F_x(x) = \int_{x_1}^{x_2} f_X(t)dt \quad \text{or} \quad P\{X \approx x\} = f_x(x)dx = P\{x < X \leq x + dx\}$$

The probability density for Gauss distribution is as follows [11] :

$$N(X; \mu, \sigma^2): f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}}, \tag{2}$$

where μ represents the mean of values and σ is the standard deviation.

The Gaussian Mixture Model (GMM) are based on the following assumptions:

- for k classes or data sources, each source generates data with an average value μ_i (Gaussian distribution);
- for a source i, characterized by a probability p_i, the data generated has the probability distribution ~N(μi, σ^2). For a set of characteristics λ_i, the Gaussian mixture is the weighted sum of the M components of the mixture:

$$p\left(\overrightarrow{x_t}/\lambda_j = \sum_{i=1}^{M} g_i N(\overrightarrow{x_t}; \overrightarrow{\mu_i}; \sum i)\right). \tag{3}$$

where g_i represents the weight of the model mixtures. The imposed condition is $\sum_{i=1}^{M} g_i = 1$, and $N(\overrightarrow{x_t}; \overrightarrow{\mu_i}; \sum i)$ represents the specific weights for each set of individual characteristics or component (characterized by mean and mean square deviation).

In the present study, the components means the main brain tissues: white matter, gray matter and cerebrospinal fluid.

2.2. CLUSTERING ALGORITHM

To determine optimally which component is best suited to the distribution of input data in the feature space (within the model - GMM), different classifiers are available [12]. Clustering means the method of segmenting a set of data into multiple clusters based on a certain similarity [13]. Clustering techniques are unsupervised algorithms in which the test data is composed of a set of input vectors that initially do not match any set of output vectors. Thus, the algorithm receives only unlabeled data, and its task is to find an adequate representation of the data distribution, to be grouped into relevant clusters. Clustering techniques group the input data according to distance separating representative points.

The classical clustering problem was applied to our data sets (60 images of three subjects and for gradients $b = 0\, s/mm^2, b = 500\ s/mm^2$ and $b = 1000\, s/mm^2$). It follows the steps:

(1) A set of data $X = \{x_1, x_2,..., x_n\}$ is given, where x_i is a non-labeled entity that is assigned a row-vector V_i of size m (with m characteristics). It is required to find the most natural partitions of X: $\{C_1, C_2, \dots, C_k\}$, with $k \leq n$, so that entities belonging to the same class are as close as possible, and those belonging to different classes or as far as possible [14].

(2) Gaussian Mixture Model uses the k-means method for clustering and optimize this process by using the Expectation Maximization EM algorithm. The EM optimization algorithm is a general method for estimating the maximum probability of distribution parameters in a given set when the data is incomplete or has missing values. There are two main applications of the EM algorithm. The first occurs when the data really has missing values due to problems or limitations of the observation process. The second occurs when the optimization of the probability function is analytically impossible to be solve, but when the probability of the function can be simplified, assuming the existence of additional but missing (or hidden) values. The algorithm works by following steps:

- First step:
    - Intialize k clusters: $C_1, \dots, C_k$
      $(\mu_j, \sum_j)$ and $P(C_j)$ for each cluster j.

- Iteration step:
    - Expectation Step:

$$\boldsymbol{p(x_i/C_j) = \frac{p(x_i/C_j).p(C_j)}{p(x_i)} = \frac{p(x_i/C_j).p(C_j)}{\sum_j p(x_i/C_j).p(C_j)}}$$

- Maximization Step:

$$\boldsymbol{\mu_j = \frac{\sum_i p(C_j/x_i).x_i}{\sum_i p(C_j/x_i).} \quad \sum j = \frac{\sum_i p(C_j/x_i).(x_i-\mu_j).(x_i-\mu_j)^T}{\sum_i p(C_j/x_i).} \quad p(C_j) = \frac{\sum_i p(C_j/x_i).}{N}}$$

The K-means clustering method classifies a set of input data into a preset number of K clusters. The basic idea is to define K weight centers, called centroids, one for each cluster. Then each element in the input data set is assigned to the nearest centroid. The first stage of the grouping ends when there are no items left unassigned. At this point, it is necessary to calculate the new centroid for the clusters determined in the previous step. The process continues until the positions of the new centroids no longer change significantly [15].

2.3 FIRST-ORDER TEXTURAL FEATURES

The texture features of an image are determined by how the gray levels are distributed across pixels in the region. It quantifies the properties of a certain region by exploiting the spatial relationships among pixels and by underlying the distribution of gray levels of a given image.

Suppose that the image is a function f(x, y), x = 0,1, ..., N-1 and y = 0,1, ..., M-1. The total number of intensity levels in the image is denoted by G, i = 0,1, ..., G-1. The histogram of the intensity level is a function that shows (for each level of intensity) the number of pixels from the entire image that have the same intensity:

$$h(i) = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \sigma(f(x,y), i) \tag{4}$$

where σ is the Kronecker delta function

$$\sigma(j,i) = \begin{cases} 1, & j = i \\ 0, & j \neq i \end{cases} \tag{5}$$

The probability density of the intensity levels is obtained by dividing the values h(i) by the total number of pixels from the image.

$$p(i) = \frac{h(i)}{NM}, i = 0,1, \dots, G-1 \tag{6}$$

The shape of the histogram gives many information about the nature of the image. For example, a narrow distributed histogram indicates an image with low contrast. A bimodal histogram often suggests that the image contained an object with a certain intensity range on a background of different intensity. The following first-order statistics allow the description of the various statistical features of the images [16]:

Mean
$$\mu = \sum_{i=0}^{G-1} ip(i) \tag{7}$$

Variance
$$\sigma^2 = \sum_{i=0}^{G-1} (i-\mu)^2 p(i) \tag{8}$$

Skewness
$$\mu_3 = \sigma^{-3} \sum_{i=0}^{G-1} (i-\mu)^3 p(i) \tag{9}$$

Kurtosis
$$\mu_4 = \sigma^{-4} \sum_{i=0}^{G-1} (i-\mu)^4 p(i) - 3 \tag{10}$$

Energy
$$E = \sum_{i=0}^{G-1} [p(i)]^2 \tag{11}$$

Entropy
$$H = -\sum_{i=0}^{G-1} p(i) log2^{[p(i)]} \tag{12}$$

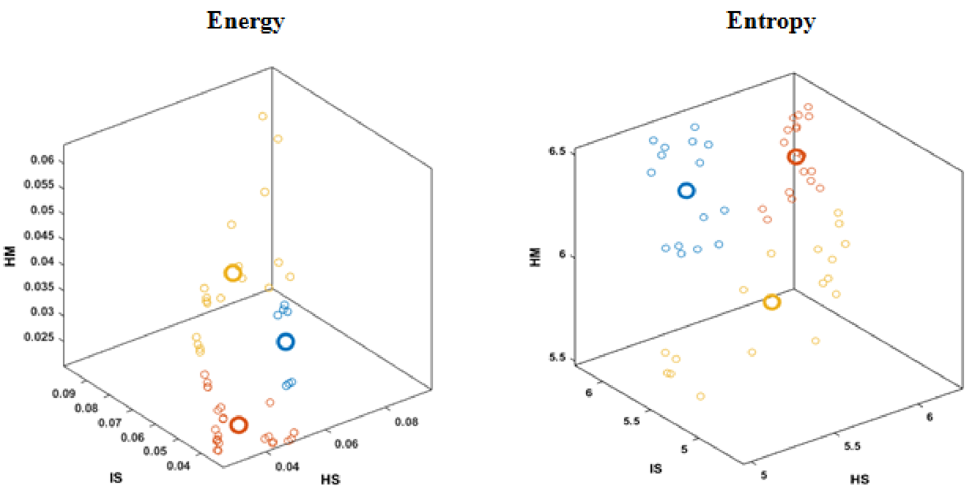
2.4. DATA ACQUISITION

The images were acquired by using a 1.5T Philips MRI scanner (Philips Medical Systems, Best, the Netherlands) with a six-channel sensitivity encoding (SENSE) for faster scanning (FS =1.5). The scanning parameters were: 3D gradient echo, echo time (TE) (from 83 to 110 ms), and repetition time (TR) (from 6500 ms to 7800 ms), bandwidth = 1070 Hz/pixel, flip angles = 2- and 6-, voxel resolution = 2.5 - 3.0 mm, slice thickness = 4 mm, and 128 slices. DW-MR images were obtained from three adult patients: one healthy, one with multiple hemorrhagic areas in the left temporal lobe and one suffering from stroke. DW-MRI images were acquired using two diffusion gradient values ($b = 500\ s/mm^2$ ***and*** $b = 1000\ s/mm^2$). Images were also acquired without using diffusion gradient ($b = 0\ s/mm^2$) but with identical imaging parameters.

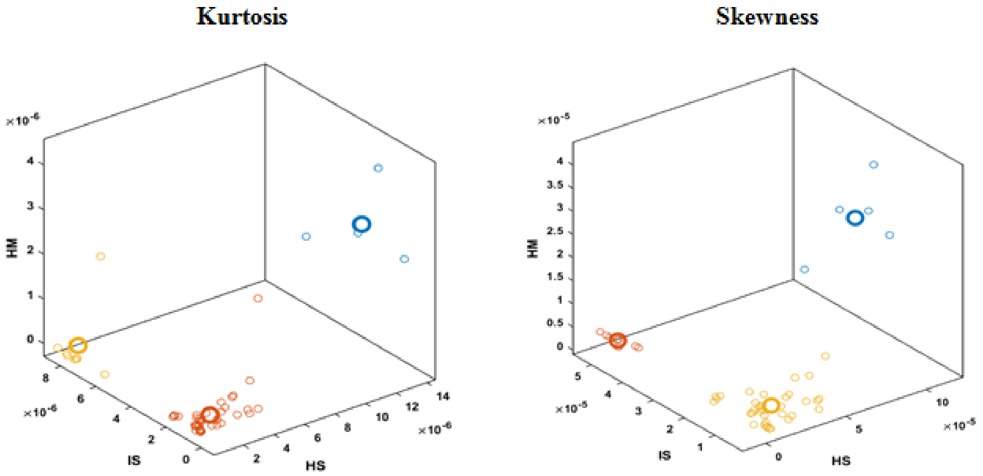
Subsequent data processing was performed using the Matlab programming language software (The MathWorks, Inc., Natick, MA, USA) based on the methods described previously in [17].

3. RESULTS AND DISCUSSION

The clustering results for the chosen features, i.e. energy, entropy, skewness and kurtosis are represented for 20 of the MRI images from the healthy subject HS, 20 from the multiple hemorrhagic areas in the left temporal lobe HM and 20 from IS ischemic stroke IS (Fig. 1).



Energy clustering for 20 images of the 3 subjects **Entropy clustering for 20 images of the 3 subjects**



Kurtosis clustering for 20 images of the 3 subjects **Skewness clustering for 20 images of the 3 subjects**

Figure 1: The clusters found by GMM method and 3D correlation for HS, HM and IS subjects

The correlation between the corresponding variables/analyzed classes is strong if the data is very close packed so that, there is a clear linear trend in any direction in the three-dimensional space of the 3D scatter plot. According to data in figure 1, the energy and entropy features meet this requirement; energy better than entropy. Important dissimilarities between analyzed classes were determined for kurtosis and skewness features. An important finding is that, based on the EM algorithm, the features like skewness and kurtosis show the best clustering results, for all three subjects. The data were well classified into components and a few data is “misclassified” being somewhere in between the clusters.

The mixing proportion of each component and for each features (energy, entropy, kurtosis and skewness) for the analyzed subjects (i.e. healthy subject, multiple hemorrhagic area in the left temporal lobe and subject with ischemic stroke) are displayed in Table 1. The best results were obtained for kurtosis feature and the third component.

Table 1. The mixing proportion of each component for energy, entropy, kurtosis and skewness. The GMM model uses three components, namely white matter, gray mater and CSF

	Energy	Entropy	Kurtosis	Skewness
Component 1:	0.239	0.341	0.100	0.184
Component 2:	0.299	0.256	0.160	0.390
Component 3:	0.260	0.402	0.741	0.424

The mixing proportion of the components indicate the density of the input data into the GMM. The energy feature approach shows a balanced distribution between the components while kurtosis feature encapsulates more density for component 3. The energy feature does not strongly associated with either component. On the contrary, the kurtosis feature strongly associated with one component. These results can be interpreted as a posterior probability of belonging to the three components and they allow us to *a priori* chose and *a posterior* validate the number of components. In the present case, we *a priori* chose three components as the brain is characterize by three main tissues, but the results of clustering analysis indicated a good separation between components for kurtosis and skewness. This is the *a posterior* validation.

CONCLUSIONS

This study analyzed the usefulness of the first order features of clustering algorithms, such as k-means and EM, respectively. The standard was the healthy subject (HS), compared to a subject with multiple hemorrhagic areas in the left temporal lobe (HM) and another subject who underwent ischemic stroke (IS). The correlation analysis for analyzed HS, HM and IS classes allows to conclude that kurtosis and skewness features are the “well-clustered” data. Moreover, the GMM model gives the possibility to decide *a posteriori* the best way for data clustering.

References

1. Dempster A. P., Laird N. M., Rubin D. B., Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society (B)*, 39:1–38, 1977.
2. Duda R. O., Hart P. E., Stork D. G., *Pattern Classification*, Wiley-Interscience Publication, 2000.
3. Fukunaga K., *Statistical Pattern Recognition*, Academic Press, San Diego, 2nd edition, 1990.
4. Hastie T., Tibshirani R., Discriminant analysis by gaussian mixtures, *Journal of the Royal Statistical Society (B)*, 58:155–176, 1996.
5. Maugis C., Celeux G., Martin-Magniette M.L., Variable selection for clustering with gaussian mixture models. *Biometrics* 65(3):701–709, 2009.
6. Bouveyron C., Brunet-Saumard C., Model-based clustering of high-dimensional data: A review, https://doi.org/10.1016/j.csda.2012.12.008, 2014.

7. Law M.H., Jain A.K., Figueiredo M.A.T., Simultaneous Feature Selection and Clustering Using Mixture Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence - PAMI*, 26, 1154-1166, 2004.
8. S. Wilfried, Mixture models, *International Encyclopedia of Statistical Science*, Heidelberg: Springer, 2010.
9. Ge Z., Song Z., Maximum-likelihood mixture factor analysis model and its application for process monitoring, *Chemometrics and Intelligent Laboratory Systems,* 102(1):53-61, 2010.
10. Namita A., Agrawal R. K., First and Second Order Statistics Features for Classification of Magnetic Resonance Brain Images, *Journal of Signal and Information Processing*, 146-153 http://dx.doi.org/10.4236/jsip.2012.32019, 2012.
11. Moldovanu S., Moraru L., Mass Detection an Classification in Breast Ultrasound Image using K-means Clustering Algorithm, *Proceedings of the 3rd International Symposium on Electrical and Electronics Enginering (ISEEE)*, Galati University, INSPEC Accession Number: 11660606, DOI: 10.1109/ISEEE.2010.5628516, 2010.
12. Dahshan E.-S. A., Hosny T., Salem A. B. M., A Hybrid Technique for Automatic MRI Brain Images Classification, *Digital Signal Processing*, Vol. 20, No. 2, pp. 433-441. doi:10.1016/j.dsp.2009.07.002, 2010.
13. Materka A., Strzelecki M., *Texture analysis methods - A review., COST B11 Technical Report*, Lodz-Brussels, Technical University of Lodz. Available at http://www.eletel. p. lodz.pl/cost/publications. html, 2003.
14. Lerski R. A., Straughan K., Schad L. R., Boyce D., Bluml S., Zuna I., MR Image Texture Analysis—An approach to Tissue Characterization, *Magnetic Resonance Imagin*g, Vol. 11, No. 6, pp. 873-887. doi:10.1016/0730-725X(93)90205-R, 1993.
15. Gan G., Ma C., Wu J., *Data Clustering: Theory, Algorithms and Applications,* Philadelphia : SIAM, 2007.
16. Moldovanu S., Moraru L., Bibicu D., Characterization of myocardium muscle biostructure using first order features, *Digest Journal of Nanomaterials and Biostructures*, 6(3): 1357-1365, 2011.
17. Gonzalez R. C., Woods R. E., Wavelet and Multiresolution Processing, in: *Digital Image Processing,* 2nd Edition, Pearson Education, Upper Saddle River, pp. 349-408, 2004