Recent Advances in Statistical Mixture Models: Challenges and Applications

Sami Bourouis[a]
Department of Information Technology, College of Computers and Information Technology,

Keywords: Mixture Model, Finite And Infinite Models, Learning Techniques, Applications.

Abstract: This paper discusses current advances in mixture models, as well as modern approaches and tools that make use of mixture models. In particular, the contribution of mixture-based modeling in various area of researches is discussed. It exposes many challenging issues, especially the way of selecting the optimal model, estimating the parameters of each component, and so on. Some of newly emerging mixture model-based methods that can be applied successfully are also cited. Moreover, an overview of latest developments as well as open problems and potential research directions are discussed. This study aims to demonstrate that mixture models may be consistently proposed as a powerful tool for carrying out a variety of difficult real-life tasks. This survey can be the starting point for beginners as it allows them to better understand the current state of knowledge and assists them to develop and evaluate their own frameworks.

1 INTRODUCTION

Statistical machine learning (SML) has made great progress in recent years on supervised and unsupervised learning tasks including clustering, classification, and pattern identification of large multidimensional data. Data scientists may examine the connections and patterns across datasets with the use of statistical models, which are mathematical representations of observable data. It gives them a strong foundation on which to predict data for the near future. Additionally, analysts can approach data analysis systematically by applying statistical modeling to original data, which results in logical representations that make it easier to find links between variables and make predictions. By estimating the attributes of huge populations based on existing data, statistical models aid in understanding the characteristics of known data (Bouveyron and Girard, 2009; Fu et al., 2021). It is the main principle behind machine learning. For them to be effective, statistical models must have the capacity to handle data distribution appropriately. Pattern recognition, computer vision, and knowledge discovery are just a few of the research fields where SML has been used. Over the past few decades, there has been a lot of cutting-edge research on SML.

Mixture models (MM), one of the various unsupervised learning techniques already in use, are receiving more and more attention because they are effective at modeling heterogeneous data (Alroobaea et al., 2020). MM are widely used for modeling unknown distributions and also for unsupervised clustering tasks. In order to partition multimodal data and determine the membership of observations with ambiguous cluster labels, well-principled mixture models can effectively deployed to achieve this objective(Lai et al., 2018). Many distributions have been studied in the past to model multimodal data like Gaussian, Gamma, inverted Beta, Dirichlet, Liouville, von Mises, and many others. Nevertheless, some of model-based distributions such as Laplace or Gaussian entail making a strict hypothesis about the shape of components, which might result in poor performance. On the other hand, finding out the exact number of components might be difficult. When attempting to describe complex real-world problems, solving such problem can help and prevent issues with over- and under-fitting. As a consequence, more flexible mixtures have been developed to get around these restrictions and offering an accurate approximation to data that contains outliers. For example, some studies tried to develop infinite mixture models in order to tackle the limitations of finite instances. In reality, incorporating an infinite number of components may enhance the statistical model's performance. Moreover, various learning techniques (non-deterministic and deterministic inference methods) were imple-

[a] https://orcid.org/0000-0002-6638-7039

Bourouis, S.
Recent Advances in Statistical Mixture Models: Challenges and Applications.
DOI: 10.5220/0011660900003411
In *Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2023)*, pages 312-319
ISBN: 978-989-758-626-2; ISSN: 2184-4313
Copyright © 2023 by SCITEPRESS – Science and Technology Publications, Lda. Under CC license (CC BY-NC-ND 4.0)

mented and used to infer mixture model's parameters and so to make accurate prediction. The current article's objective is to give a short overview of the methodological advancements that support mixture model implementations. Many academics are interested in MM-based frameworks, and they have discovered a variety of fascinating applications for them. As a result, the literature on mixture models has greatly increased, and the bibliography included here can only give a limited amount of coverage. The structure of this article is designed to give a quick understanding of mixture models. In the next section, a taxonomy related to this area of research is provided. Then, Section 3 provides a summary of the principal methods now in use that employ mixture models. A concise and comprehensive explanation of main challenges is also provided, along with some suggestions for the future. Finally, this article is concluded.

2 TAXONOMY

The next paragraphs present a basic and understandable taxonomy related to mixture models.

2.1 Statistical Modelling

Making accurate decisions is now probabilistically possible using statistical modeling. The goal is to build a model that might logically explain the data. Mixture models (MM) are well-founded probabilistic models with the benefit of using several distributions to characterize their component elements. Mixture models offer an easy-to-use yet formal statistical framework for especially grouping and classification by using well-known probability distributions (such as Gaussian, Poisson, Gamma, and binomial). Therefore, we can evaluate the likelihood of belonging to a cluster and draw conclusions about the sub-populations, unlike conventional clustering algorithms. MM can successfully express multidimensional distributions and heterogeneous data in a finite (or infinite) number of classes, which makes them useful for modeling visual features (McLachlan and Peel, 2004). The core of mixture modeling is selecting the appropriate probability density functions (PDFs) of each component in the mixture of distribution. An example of clustering task that divides the input data set into three groups is shown in Fig.1. It is noted that a mixture model learns from the input data during the learning phase and then this model is evaluated using different set of data (testing phase). The obtained model will be applicable for further concerns including classification, grouping, and prediction especially if it can produce results with high precision. Several scientific researchers are interested in developing efficient unsupervised statistical learning approaches to address various data mining and ML problems.

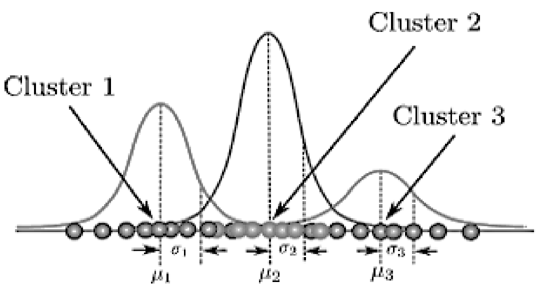

Figure 1: Example of density modelled by three Gaussian probability functions for data clustering (or classification).

2.2 Finite Mixture Models

Mixture models are intended to combine two or more distributions to produce a distribution with a more flexible shape than a single distribution. In a mixture model, observations are produced through a combination of many unique models, which is an example of a hidden model. A convex combination of two or more finite density functions is known as a finite mixture model (FMM). FMM is used to estimate the likelihood of being a member of each cluster, to calculate the parameters of each component, to group data into different classes, and to make inference. They offer a powerful framework for identifying latent patterns and studying ambiguous data. A general formulation of FMM is given as follow. Let y the observed dataset and that each instance is taken from one of K components. In a statistical setting, a FMM might be expressed as follow:

$$p(y|\Theta) = \sum_{k=1}^{K} \pi_k p(y|\theta_k) \tag{1}$$

where Θ is the full parameters of the mixture model; $p(y|\theta_k)$ is a probability density function (pdf) of the cluster k, and π_k represent the relative mixture weights (different proportions).

2.3 Infinite Mixture Models

Although finite MM are useful, determining the appropriate number of components for a given dataset and then estimating the parameters of that mixture are the two most difficult issues that must occasionally be solved (Brooks, 2001). In many circumstances, we claim that we have an infinite mixture if the number of components is equal to (or more than) the number of observations. In such cases, it is better to let the model adjust its complexity to the volume of data to prevent both underfitting and overfitting. Infinite models are a kind of non-parametric models. One of the most attractive methods for converting the finite mixture model into its infinite counterpart is to use Dirichlet process mixture model (Fan and Bouguila, 2020; Bourouis and Bouguila, 2021). When new data is received, infinite models may either add new groups or eliminate some of the current ones. By taking into account $K \to \infty$, we may overcome the problem of calculating K. As a result, the infinite mixture can be expressed as

$$p(y|\Theta) = \sum_{k=1}^{\infty} \pi_k p(y|\theta_k) \tag{2}$$

2.4 Model Learning

The development of parameter learning techniques has historically attracted a lot of attention. Estimating parameters of the mixture from data may be done in a number of ways, and this issue is still under investigation. There are several approaches known as estimators that are used for model's parameters estimation, including the least square method (LSM), maximum likelihood estimation (MLE), and so on. Mainly, there exist two kind of approaches: frequentist (known also as deterministic) and Bayesian. These two approaches may be distinguished from many others that have been derived out. For example, the maximum likelihood (ML) estimator is the foundation of the frequentist approach. One of the key purposes of machine learning is to make predictions using the parameters learnt from the training dataset. Depending on the kind of predictions and/or our past experience (prior knowledge), either a frequentist or a Bayesian technique could be adopted to accomplish the goal.

2.4.1 Frequentist Learning

Deterministic approaches assume that observed data is drawn from a given distribution (Tissera et al., 2022). This distribution is referred to as the likelihood, or $P(Data \mid \theta)$ where the objective is to estimate θ (i.e. the model's parameters), which is assumed to be constant number, that might maximize the likelihood (MLE). In statistical modelling, when the model depends on unobserved latent variables, some fundamental algorithms (such as the famous iterative EM algorithm) could be applied to found the maximum likelihood or the maximum a posteriori. For this case, the procedure of maximizing the likelihood is formalized as the following optimization problem:

$$\hat{\Theta}_{ML} = \arg\max_{\Theta}\{\log p(\mathcal{Y}|\Theta)\} \tag{3}$$

Analytically, this equation cannot be solved. The expectation maximization (EM) algorithm or other related methods may be used to generate the MLE estimates of the mixture parameters. Indeed, the EM method generates a series of estimates $\{\Theta^t, t = 0,1,2\dots\}$ by alternately utilizing two stages until a convergence: Expectation and Maximization steps.

2.4.2 Bayesian Learning

Under the Bayesian method (Bourouis et al., 2021a), the parameter of the mixture denoted by Θ is viewed as a random variable with a specific probability distribution (the prior). The latter serves to represent our belief prior before seeing the data. For this case, the Bayes theorem is used to update the prior distribution based on the likelihood function. The information in the prior distribution as well as the data is summarized in a subsequent distribution known as the posterior distribution which is expressed as:

$$p(\Theta|\mathcal{Y}) \propto p(\mathcal{Y}|\Theta)p(\Theta) \tag{4}$$

The Markov chain Monte Carlo simulation technique (MCMC), the Gibbs sampler, and Laplace's method are some effective Bayesian approximation approaches that have been used to various machine learning applications (Husmeier, 2000).

2.4.3 Variational Learning

It should be emphasized that Bayesian techniques need a significant computing investment, particularly when working with massive data sets (Tan and Nott, 2014). For example, MCMCs are widely used to sample from distributions, however this method is occasionally computationally expensive. As a result, variational Bayes inference has been explored to solve these problems. It has actually been used as a more effective alternative than MCMC. The fundamental idea is to estimate the model posterior distribution by minimizing the Kullback-Leibler (KL) divergence between the true posterior and an approximation distribution.

2.4.4 Expectation Propagation

Expectation propagation (EP) learning (Minka, 2001) may be viewed as a recursive approximation method that minimizes a Kullback-Leibler (KL) divergence between an approximation and the exact posterior model. EP is a deterministic approach to Bayesian inference that generates the optimal posterior distribution through an iterative refining process. It is based on the so-called Assumed Density Filtering (ADF) (Minka, 2001). The EP inference differs from the ADF in that it does not rely on the input data's order, and it might be improved by utilizing more than one data point. Furthermore, the higher computing performance of EP over Gibbs sampling and Markov Chain Monte Carlo (MCMC) is one of its key advantages.

2.4.5 Batch/Online Learning Algorithms

Online algorithms enable the sequential processing of data instances, which is critical for real-time applications (Fujimaki et al., 2011). Online learning is more attractive than batch one and this for many applications especially when dealing with huge datasets. With mixture models, it is possible to save time and ensure performance since we need to update the model's parameters progressively. The settings must be adjusted appropriately without compromising flexibility and efficiency.

2.5 Mixture Model Selection

In order to find the best fit for modeling data, several model selection-based techniques have been proposed. Automated selection of the components number that best describes the observations has been the subject of several studies and investigations. Finding the optimal number of components to explain a given set of data is one of mixture models' most difficult tasks. This procedure is called model selection. Many successful information criteria have been considered in order to address this challenging issue. Among the well-known criteria, we may find the Akaike's information criterion (AIC) (Akaike, 1974) and Bayes information criterion (BIC) (Schwarz, 1978). It should be noted that these criteria are based on penalizing a mixture's log likelihood function. The Kullback-Leibler divergence between the probability density function and the mixture model is what the AIC seeks to reduce. By minimizing the influence of the prior, the BIC, on the other side, roughly approximates the marginal likelihood of the mixture. We refer to the literature for other additional information criteria such as the minimum description length (MDL) and the minimum message length (MML) (Azam and Bouguila, 2022). Indeed, several publications have suggested simultaneously estimating the parameters of the mixture and selecting the best optimal model using for example MML or MDL criteria.

3 MIXTURE MODELS AND APPLICATIONS: OVERVIEW AND DISCUSSION

Due to the widespread adoption of new technologies, which has led to millions of people producing enormous volumes of heterogeneous data through smart devices, there seems to be significant opportunity for expanding knowledge across a variety of scientific disciplines. The technological revolution has made it possible to quickly analyze and extract knowledge from massive datasets, which is particularly beneficial for a wide range of sectors. Nevertheless exploring the content of these sizeable multimedia databases is a crucial challenge that may be dealt with by statistical tools. Using statistical methods like mixture models (MM), problems like data clustering, object segmentation, image denoising, pattern recognition, and many more might be successfully handled. For various applications of data analysis, a fundamental mixture model-based architecture requires a number of processing stages as shown in Fig.2. It has recently been shown that MM can provide prospective capacity for addressing difficult machine learning problems. In particular, MM can effectively address the issue of imbalanced samples and insufficient training data (McLachlan and Peel, 2004). Typically, the first step in employing mixture models is to extract effective visual features (descriptors) from the input dataset. Different techniques, such as SIFT (Scale Invariant Feature transform) (Lowe, 2004), could be used to extract such characteristics. An additional step is to implement a visual vocabulary by quantifying the attributes into visual words using for example a bag-of-words (BOW) model and a simple clustering algorithm like K-means (Csurka et al., 2004). Furthermore, a probabilistic Latent Semantic Analysis (pLSA) may be utilized to reduce the dimensionality and to produce a d-dimensional proportional vector (Hofmann, 2001) (that takes any value from 0 and 1). The developed statistical model is then applied as a classifier to assign each input to the category with the highest posterior probability in accordance with Bayes' decision rule. Typically, datasets are randomly divided into two parts; the first is utilized for training and for creating the visual vocabulary, while the second is used for testing and assessment.

3.1 Visual Features Extraction

One of the critical step in the field of machine learning is feature extraction. In particular, the extraction of important and relevant visual characteristics is of great importance in the computational analysis

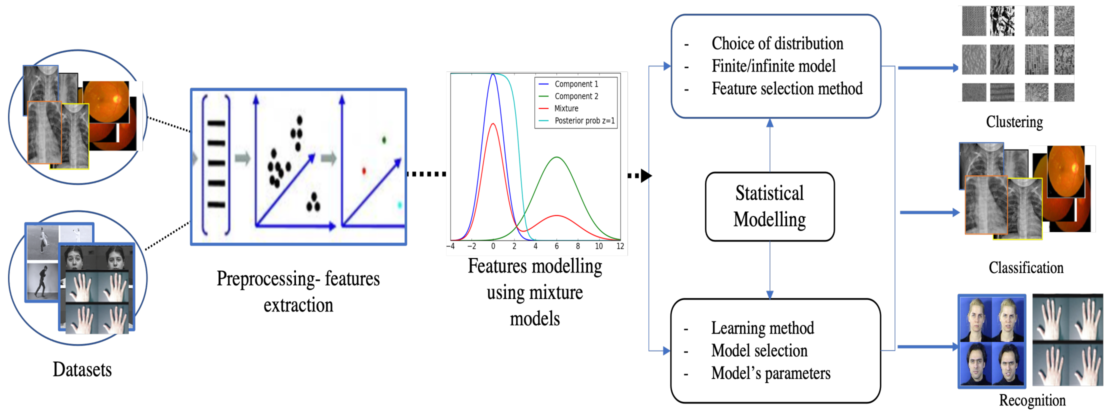

Figure 2: A basic mixture model-based framework for different data analysis applications.

of multimedia data. The first stage for any image processing framework is to convert the multimodal data (image, text, and video) into a collection of characteristics. In order to accomplish this, some existing techniques extract the most representative attributes, from the source data (data set), that aid in differentiating between input data. The task of feature extraction has mostly been addressed in the context of image analysis applications. Feature extraction may be thought of as dimensionality reduction since only less information is used than in the initial input data. In order to extract spatial characteristics from images and videos, one must take into consideration spatial-pixel and spatio-temporal information. Various techniques, including statistical ones, local or global methods have been published in the literature for feature extraction.They might be based on the gray level value's first order or higher order statistics. Local extraction techniques (eg. SIFT, SURF, and LBP) do not necessarily need background detection or subtraction since it is less prone to noise and partial occlusion.

3.2 Overview

A valuable tool for examining diverse types of data in many computer science fields is the combination of statistical approaches with machine learning. A foundation of machine learning is statistics. Without it, it is difficult to fully comprehend and apply machine learning. The basic goal of statistical machine learning is the design and optimization of probabilistic models for data processing, analysis, and prediction. Statistical machine learning advancements have a big influence on lots of different sectors including artificial intelligence, signal/image processing, information management, as well as fundamental sciences. It is important to offer new sophisticated statistical machine learning (SML) approaches based on a family of flexible distributions in order to solve significant issues with standard machine learning algorithms and data modeling. SML has made great progress in recent years on both supervised and unsupervised learning tasks including clustering, classification, pattern recognition, and many other data analysis-based application.

3.2.1 Data Classification with MM

Multimodal data classification, which includes text, images, and videos, is one of the most important tasks for many computer vision applications. It involves properly allocating objects to one of a number of specified classes. When using mixture models, it is reasonable to assume that the observed data are drawn from several distributions, and the classification problem is then seen as an estimation of the parameters of these distributions. In recent years, mixture models have been used to develop efficient computer-aided systems that successfully classify input scans and/or video sequences. For instance, they have been used to classify retinal images and detect diabetic retinopathy (Bourouis et al., 2019) and lung disease in CXR images (Alharithi et al., 2021). They have been also exploited to classify biomedical data (Bourouis et al., 2021a; Bourouis et al., 2021c). It is notable that a number of models were constructed for various purposes, including texture categorization (Norah Saleh Alghamdi, 2022).

3.2.2 Data Clustering with MM

Data clustering is a fundamental and quite well unsupervised learning technique with use in data mining and information retrieval, among other areas (McLachlan and Peel, 2004). Often, a clustering technique is used to discover hidden patterns in data. By fitting a variety of probability distributions to the underlying data and then continually modifying their parameters until they best match it, mixture models (MM) have the potential to locate clusters. When we simultaneously have low inter-cluster and significant intra-class similarities, an MM-based clustering method will produce high quality clusters. Numerous research on clustering high-dimensional data have been done and published in the literature (Melnykov and Wang, 2023; Tissera et al., 2022; Jiao et al., 2022; Hu et al., 2019; Lai et al., 2018).

3.2.3 Pattern Recognition with MM

The usage of mixture models in the field of unsupervised pattern recognition has become common. In fact, a pattern is viewed as an entity, represented by a feature vector, and may then be described using a combination of distributions. To deal with the problem of recognizing complex patterns, such as human activities, hand gestures, face expressions, and so forth, several unsupervised-based mixture models have been established (Yang et al., 2013; Najar et al., 2020; Bouguila, 2011; Bourouis et al., 2021b; Alharithi et al., 2021). In (Yang et al., 2013), a method for simulating articulated human movements, gestures, and facial expressions is suggested. Additionally, an expectation propagation inference approach based on inverted Beta-Liouville mixture models was suggested to handle diverse pattern recognition applications (Bourouis and Bouguila, 2022).

3.2.4 Data Segmentation with MM

Image segmentation is the process of partitioning an image's pixels into a number of homogeneous subgroups (segments). Image segmentation presents a variety of challenges, including choosing the right number of segments (Su et al., 2022). Smooth object segmentation may be challenging, particularly if the image contains noise, a complex foreground, poor contrast, and irregular intensity. The main factors for grouping pixels that we seek for in image segmentation are proximity, color, and textures, which are frequently present in pixels. Different mixture model-based approaches were implemented to tackle the issue of object segmentation, and some of them have shown to produce better outcomes (Cheng et al., 2022; Channoufi et al., 2018; Allili et al., 2008).

3.3 Discussion

Standard distributions, such as the Gaussian, have been employed for many years as the primary distributions to address data analysis concerns. These probabilities are unfortunately not the most accurate approximation when dealing with non-Gaussian data. Effectively modeling data vectors requires both choosing the best approximation for the treated dataset and the most effective inference technique for learning mixture models. Recent studies have shown that mixtures based on the Dirichlet, inverted Dirichlet (ID), and generalized Dirichlet distributions outperform the conventional Gaussian for a range of data analysis tasks. Meanwhile, certain distributions continue to have a number of shortcomings (such as the restrictive covariance matrix structure) that restrict their application in a number of different real-world scenarios. Furthermore, other distributions fail to find the right number of components to adequately characterize the input vectors without over-fitting and/or under-fitting. To overcome these issues, even when doing so requires expensive computing resources, it is strongly advised to employ other methods and incorporate additional criteria, such as the Akaike information criterion (AIC), MML, and MDL. Future studies should go deeply into the fundamental issues with MM-based models. First, determining the optimal number of components is a difficult problem that needs more investigation. Then, when dealing with model parameter estimation using various methodologies, such as the EM algorithm, initial values selection and convergence issues are frequently encountered. On the other side, to reproduce the optimal log-likelihood, non-normal distributed mixtures require more random starting values than do normal distributed mixtures. Additionally, although they are quite time-consuming, Bayesian approaches based on the MCMC methodology have gained popularity and may be able to assist prevent issues with EM algorithms. Each component must be specified before the entire mixture model can be constructed. As a result, if the model is not properly defined, inconsistent estimations for the set of parameters may happen. Furthermore, when several statistics must be calculated simultaneously, choosing the best model is a desired but challenging problem. It is also necessary to make more adjustments in order to compare potential models and evaluate the fit quality. On the other hand, it should be emphasized that while many probabilistic models may successfully categorize comparable data, they occasionally fail when the data is severely influenced by noise and outliers. In light of these shortcomings, it is theoretically possible that discriminative classifiers, in particular the support vector machine (SVM), might be used. Therefore, it is advisable to investigate hybrid methods that take into consideration both the benefits of probabilistic and discriminative models in order to gain superior performance. For instance, designing robust mixture-based probabilistic SVM kernels can help with this. Last but not least, it is crucial to emphasize that the output of various statistical MM models frequently depends on the dataset's sample size.

4 CONCLUSION

Mixture models (MM), an emerging statistical method for modeling complex multimodal data, is discussed in this paper. The current study presents a recent brief review of the advances in MM models. Although there hasn't been much research on MM-based methods, and only a few publications have time-varying indicators, we are optimistic that more significant and insightful results will soon be made available to the public.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Alharithi, F. S., Almulihi, A. H., Bourouis, S., Alroobaea, R., and Bouguila, N. (2021). Discriminative learning approach based on flexible mixture model for medical data categorization and recognition. *Sensors*, 21(7):2450.

Allili, M. S., Bouguila, N., and Ziou, D. (2008). Finite general gaussian mixture modeling and application to image and video foreground segmentation. *Journal of Electronic Imaging*, 17(1):013005–013005.

Alroobaea, R., Rubaiee, S., Bourouis, S., Bouguila, N., and Alsufyani, A. (2020). Bayesian inference framework for bounded generalized gaussian-based mixture model and its application to biomedical images classification. *Int. J. Imaging Systems and Technology*, 30(1):18–30.

Azam, M. and Bouguila, N. (2022). Multivariate bounded support asymmetric generalized gaussian mixture model with model selection using minimum message length. *Expert Systems with Applications*, 204:117516.

Bouguila, N. (2011). Bayesian hybrid generative discriminative learning based on finite liouville mixture models. *Pattern Recognit.*, 44(6):1183–1200.

Bourouis, S., Alroobaea, R., Rubaiee, S., Andejany, M., Almansour, F. M., and Bouguila, N. (2021a). Markov chain monte carlo-based bayesian inference for learning finite and infinite inverted beta-liouville mixture models. *IEEE Access*, 9:71170–71183.

Bourouis, S., Alroobaea, R., Rubaiee, S., Andejany, M., and Bouguila, N. (2021b). Nonparametric bayesian learning of infinite multivariate generalized normal mixture models and its applications. *Applied Sciences*, 11(13).

Bourouis, S. and Bouguila, N. (2021). Nonparametric learning approach based on infinite flexible mixture model and its application to medical data analysis. *Int. J. Imaging Syst. Technol.*, 31(4):1989–2002.

Bourouis, S. and Bouguila, N. (2022). Unsupervised learning using expectation propagation inference of inverted beta-liouville mixture models for pattern recognition applications. *Cybernetics and Systems*, pages 1–25.

Bourouis, S., Sallay, H., and Bouguila, N. (2021c). A competitive generalized gamma mixture model for medical image diagnosis. *IEEE Access*, 9:13727–13736.

Bourouis, S., Zaguia, A., Bouguila, N., and Alroobaea, R. (2019). Deriving probabilistic SVM kernels from flexible statistical mixture models and its application to retinal images classification. *IEEE Access*, 7:1107–1117.

Bouveyron, C. and Girard, S. (2009). Robust supervised classification with mixture models: Learning from data with uncertain labels. *Pattern Recognition*, 42(11):2649–2658.

Brooks, S. P. (2001). On bayesian analyses and finite mixtures for proportions. *Stat. Comput.*, 11(2):179–190.

Channoufi, I., Bourouis, S., Bouguila, N., and Hamrouni, K. (2018). Spatially constrained mixture model with feature selection for image and video segmentation. In *Image and Signal Processing - 8th International Conference, ICISP France*, pages 36–44.

Cheng, N., Cao, C., Yang, J., Zhang, Z., and Chen, Y. (2022). A spatially constrained skew student's-t mixture model for brain mr image segmentation and bias field correction. *Pattern Recognition*, 128:108658.

Csurka, G., Dance, C., Fan, L., Willamowski, J., and Bray, C. (2004). Visual categorization with bags of keypoints. In *Workshop on statistical learning in computer vision, ECCV*, volume 1, pages 1–2. Prague.

Fan, W. and Bouguila, N. (2020). Modeling and clustering positive vectors via nonparametric mixture models of liouville distributions. *IEEE Trans. Neural Networks Learn. Syst.*, 31(9):3193–3203.

Fu, Y., Liu, X., Sarkar, S., and Wu, T. (2021). Gaussian mixture model with feature selection: An embedded approach. *Computers and Industrial Engineering*, 152:107000.

Fujimaki, R., Sogawa, Y., and Morinaga, S. (2011). Online heterogeneous mixture modeling with marginal and copula selection. In *International Conference on Knowledge Discovery and Data Mining, SIGKDD, USA*, pages 645–653.

Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. *Mach. Learn.*, 42(1/2):177–196.

Hu, C., Fan, W., Du, J., and Bouguila, N. (2019). A novel statistical approach for clustering positive data based on finite inverted beta-liouville mixture models. *Neurocomputing*, 333:110–123.

Husmeier, D. (2000). The bayesian evidence scheme for regularizing probability-density estimating neural networks. *Neural Computation*, 12(11):2685–2717.

Jiao, L., Denœux, T., ga Liu, Z., and Pan, Q. (2022). Egmm: An evidential version of the gaussian mixture model for clustering. *Applied Soft Computing*, 129:109619.

Lai, Y., Ping, Y., Xiao, K., Hao, B., and Zhang, X. (2018). Variational bayesian inference for a dirichlet process mixture of beta distributions and application. *Neurocomputing*, 278:23–33.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.*, 60(2):91–110.

McLachlan, G. J. and Peel, D. (2004). *Finite mixture models*. John Wiley & Sons.

Melnykov, V. and Wang, Y. (2023). Conditional mixture modeling and model-based clustering. *Pattern Recognition*, 133:108994.

Minka, T. P. (2001). Expectation propagation for approximate bayesian inference. In Breese, J. S. and Koller, D., editors, *Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence, Washington, USA*, pages 362–369. Morgan Kaufmann.

Najar, F., Bourouis, S., Bouguila, N., and Belghith, S. (2020). A new hybrid discriminative/generative model using the full-covariance multivariate generalized gaussian mixture models. *Soft Comput.*, 24(14):10611–10628.

Norah Saleh Alghamdi, Sami Bourouis, N. B. (2022). Effective frameworks based on infinite mixture model for real-world applications. *Computers, Materials & Continua*, 72(1):1139–1156.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, pages 461–464.

Su, H., Zhao, D., Elmannai, H., Heidari, A. A., Bourouis, S., Wu, Z., Cai, Z., Gui, W., and Chen, M. (2022). Multilevel threshold image segmentation for COVID-19 chest radiography: A framework using horizontal and vertical multiverse optimization. *Comput. Biol. Medicine*, 146:105618.

Tan, S. L. and Nott, D. J. (2014). Variational approximation for mixtures of linear mixed models. *Journal of Computational and Graphical Statistics*, 23(2):564–585.

Tissera, D., Vithanage, K., Wijesinghe, R., Xavier, A., Jayasena, S., Fernando, S., and Rodrigo, R. (2022). Neural mixture models with expectation-maximization for end-to-end deep clustering. *Neurocomputing*, 505:249–262.

Yang, Y., Saleemi, I., and Shah, M. (2013). Discovering motion primitives for unsupervised grouping and one-shot learning of human actions, gestures, and expressions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(7):1635–1648.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS