STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **13**, April 2025, pp 1404–1412.
Published online in International Academic Press (www.IAPress.org)

Density-Adaptive Clustering of Multivariate Angular Data Using Dirichlet Process Mixture Models with Circular Normal Distribution for Artificial Intelligence Applications

Said Benlakhdar [1,2,*], Saralees Nadarajah [3] Mohammed Rziza [2], Rachid Oulad Haj Thami [4]

[1] *SmartiLab Laboratory, Moroccan School of Engineering Sciences (EMSI), Rabat 10000, Morocco*
[2] *LRIT URAC 29, Faculty of Sciences, Mohammed V University, Rabat, Morocco*
[3] *Department of Mathematics, University of Manchester, Manchester M13 9PL, UK*
[4] *RIITM, ENSIAS, Mohammed V University, Rabat, Morocco*

Abstract Data clustering is an essential technique for organizing unsupervised data, extracting subjects automatically, and swiftly retrieving or filtering information. In this study, we approach the task of clustering multivariate angular distributions using nonparametric Bayesian mixture models featuring von Mises distributions. Our approach operates within a nonparametric Bayesian framework, specifically leveraging the Dirichlet process. Unlike finite mixture models, our approach assumes an infinite number of clusters initially, inferring the optimal number automatically from the data. Morever, our paper introduces a unified approach, leveraging Ward's algorithm, Dirichlet process, and von Mises Mixture distributions (DPM-MvM), to effectively capture both the structure and variability inherent in the data. We've developed a variational inference algorithm for DPM-MvM enabling automatic determination of the number of clusters. Our experimental results showcase the efficiency and accuracy of our method for analyzing multivariate angular data with state of the art approaches.

Keywords Dirichlet Process, Mixture distributions, Non parametric Bayesian model, Clustering, Ward's algorithm, Modeling Multivariate Angular Data

DOI: 10.19139/soic-2310-5070-2146

1. Introduction

The Bayesian methodology has gained considerable traction in recent years due to its widespread application in finite mixture models, which are utilized across various fields [1, 2]. These mixture models have proven invaluable in a plethora of tasks including image organization, color segmentation, restoration, texture processing, anomaly detection, sentiment analysis, and recommendation systems [3, 4, 5, 6]. Probabilistic methods play a crucial role in deciphering the underlying patterns within such data [7], with the Bayesian approach being particularly noteworthy for its ability to estimate model uncertainty, encompassing both model fit uncertainty and parameter estimation uncertainty [8]. By merging the Bayesian approach with finite mixture models, potent probabilistic modeling tools are forged for both univariate and multivariate data [9], which have found widespread application in modeling diverse practical scenarios where data stems from multiple mixed populations.
Despite the extensive theoretical exploration of the Dirichlet distribution, its practical applications, especially in parameter estimation, have received limited attention [10]. Current studies tend to focus on individual distributions or are restricted to the two-parameter Beta distribution, possibly due to the distribution's unfamiliarity among many researchers [11]. Thus, this study presents an algorithm aimed at estimating the parameters of a Dirichlet

*Correspondence to: Said Benlakhdar (Email: said-benlakhdar@um5s.net.ma).Laboratory of Research in Information Technology and Telecommunication (LRIT), Rabat IT Center, Faculty of Sciences, Mohammed V University in Rabat, Rabat B.P. 1014 RP, Morocco

ISSN 2310-5070 (online) ISSN 2311-004X (print)

mixture, with the goal of bridging the gap between theoretical underpinnings and practical implementation [12]. The proposed algorithm integrates a penalty term into the objective function to ascertain the optimal number of components necessary for accurate data modeling.
In this investigation, we center our attention on the Dirichlet distribution as a superior option for data modeling [13]. Serving as the multivariate extension of the Beta distribution, the Dirichlet distribution offers exceptional flexibility and ease of use [14]. Notably, it permits multiple symmetric and asymmetric modes, thereby accommodating data exhibiting right, left, or symmetric skewness. Furthermore, to capture the underlying structure and variability in multivariate angular data, we employ Mixture von Mises distributions. Estimating a hierarchical von Mises model without specifying the number of clusters poses a challenge [15, 16, 17]. To tackle this, we employ an exploratory approach utilizing Ward's algorithm to group observations into homogeneous clusters based on similarities. Subsequently, upon obtaining the clusters, we utilize the Dirichlet process to estimate cluster proportions [18]. A Mixture von Mises distributions Model is then employed to estimate cluster means and dispersions. This approach offers flexibility, robustness, and computational feasibility in capturing the complex structure and variability of the data.
The subsequent sections of this paper are organized as follows: Section 2 delves into the Dirichlet distribution, while Section 3 presents the Dirichlet Process Mixture by von Mises Mixture Distributions Method. Section 4 is dedicated to describing the estimation algorithm for fitting the model, and Section 5 presents and discusses the experimental results. Finally, conclusions are drawn in Section 6.

2. The Dirichlet distribution

Let Δ_n denote the $(n-1)$- dimensional probability simplex, representing the set of vectors in $\mathbb{R}^n$ with non-negative components that sum up to one

$$\Delta_n = \{q = (q_1, \dots, q_n) \in \mathbb{R}^n, \sum_{i=1}^{n} q_i = 1, q_i, i = 1, \dots, n\} \tag{1}$$

The family of Dirichlet distributions constitutes a collection of probability distributions on parametrized by n positive scalars $x_1, ..., x_n > 0$ (see Figure 1), which encompass the following probability density function with respect to the Lebesgue measure [19, 20]. The probability density of the Dirichlet distribution of order $k \geq 2$ and parameter α is expressed as follows

$$P(y \mid \alpha) = \frac{1}{\beta(\alpha)} \prod_{j=1}^{d} y_j^{\alpha_j - 1}, \left(\alpha = \frac{\prod_{j=1}^{d} \Gamma(\alpha_j)}{\Gamma(1^T \alpha)}\right), 1^T y = 1, y > 0 \tag{2}$$

$$P(y; \alpha) = \frac{\Gamma(\alpha_0)}{\prod_i^k \Gamma(\alpha_i)} \prod_i^k y^{\alpha_i - 1} \tag{3}$$

Where $\alpha > 0$ and the normalization factor $\beta(\alpha)$ corresponds to the beta function, formulated as a function of the gamma function. When setting $\theta = \alpha$, the parameters of the exponential family are delineated as follows: $T(y) = \log y$, $b(\theta) = \beta(\theta)$, $h(y, \theta) = \frac{1}{\prod_{j=1}^{d} y_j}$, $a(\Phi) = 1$. For the sake of enhanced generality and to underscore the specific properties of this distribution required for our findings, we will proceed by considering a broader Dirichlet mixture in the subsequent analysis. We will only impose the essential assumptions on this mixture to facilitate our study.

3. Dirichlet Process Mixture by von Mises Mixture Distributions (DPM-MvM)

The Dirichlet Process Mixture (DPM) is a non-parametric extension of mixture models where the number of mixture components is not predetermined but rather inferred from the data [21, 22]. The performance of the model

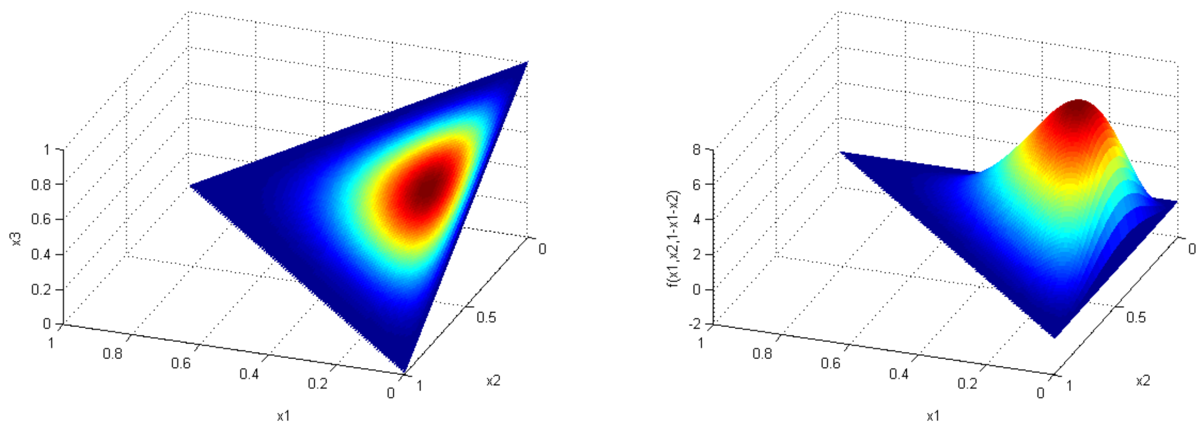


Figure 1. Visualization of a 3D Dirichlet Distribution: Mapping Dimensions onto Axes and Contours Depicting Density at (x, y, 1-x-y).

is influenced by the choice of parameters for the Dirichlet Process. The parameter α is a vector that controls the concentration of clusters [23]. In other words, it governs the model's propensity to form either a large number of small clusters or a small number of large clusters. A higher value of α promotes the creation of more clusters by increasing the probability of introducing new clusters. Conversely, a lower value of α encourages the concentration of data into fewer clusters. To determine the optimal α, we employed an empirical approach based on cross-validation by testing various α values. This approach involves partitioning the data into subsets, training the model on certain subsets, and validating it on the remaining subsets to ascertain the best value of α [24].
We employ a probabilistic model to depict collections of multivariate N samples, denoted as θ_i, drawn from a mixture of von Mises distributions. This mixture encompasses an unknown number, M, of independent components, each characterized by parametersk_k and μ_k. Here, $\pi_m = P(c_i = k)$ signifies the probability of sample I originating from a particular cluster, while indicates the assignment of sample I to one of the M clusters.

$$P(\theta_i; \mu_1, \ldots, \mu_M, k_1, \ldots, k_M, \pi_1, \ldots, \pi_M) = \sum_{m=1}^{M} \pi_m \prod_{i=1}^{N} f(\theta_i; \mu_m, k_m) \tag{4}$$

Estimating a von Mises Mixture model through the Dirichlet distribution in a supervised manner, without a predetermined number of clusters, presents a challenge due to the often ambiguous parameterization of cluster count. However, there are strategies to tackle this complexity using exploratory methods. In our approach, we employ an unsupervised statistical modeling technique, utilizing Ward's algorithm to explore potential group structures within the data. This method facilitates the grouping of observations based on their similarities, obviating the need to predefine the number of groups. Once these exploratory groups are identified, the Dirichlet process is leveraged to estimate group proportions. Additionally, a von Mises Mixture distribution model is applied to estimate the parameters of each group, offering a comprehensive insight into the underlying structure of the data. In this section, we present a novel approach, combined by mixture model utilizing the Dirichlet Process Mixture (DPM). This model is tailored to autonomously ascertain the number of topics, enabling the precise revelation of latent topics within each label. Our starting point is a dataset $X = x_1, x_2, \ldots, x_N$, comprising d-dimensional vectors, each representing individual data instances.

Let denote the set of labels $V = v_1, v_2, \ldots, v_N$, where each $v_i$corresponds to the label of x_i. The label v_n is generated using a multinomial distribution: $v_n Mult(\delta)$, where $\delta = (\delta_1, \delta_2, \ldots, \delta_{v_n})$ represents each von Mises distributions proportion. By utilizing separate DPMs, our model can autonomously determine the optimal number of topics for each class. As a result, our model acquires the unique topics essential for distinguishing between classes and interpreting labels. Let z_n denote an assignment variable representing the mixture topics linked to a data instance x_n. When the label z_n is known, z_n is generated via a multinomial distribution: $Z_n \sim Mult(\pi(\mu_{u_{v_n}})$.

The conditional distribution of x_n given v_n, $\mu_1, \mu_2, \ldots, \mu_n$ and k is $P(x_n \mid v_n, k, \mu_1, \mu_2, \ldots) = MvM(x_n \mid \mu_{v_n,z_n}, k_{v_n,z})$. Here, $\mu_{v,t}$ (where $t = 1, 2, \ldots, \infty$) represents the mean direction vectors, which are the specific topics associated with label (see Figure2) .[26] Mardia and El–Atoum (1976) recognized the von Mises (vM) distribution as the conjugate prior for the mean direction.

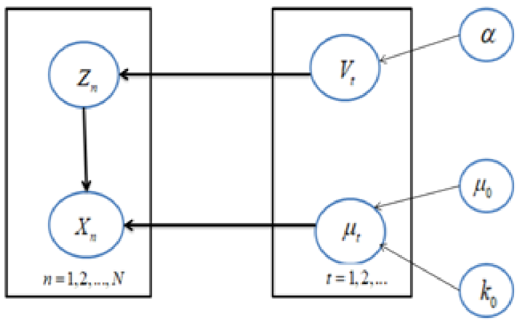


Figure 2. Graphical Model of DPM by MvM

4. Variational Inference for DPM of MvM distributions

During the training process, each instance x_n is associated with a labelv_n . Let X_v and Z_v denote the instances and their assignment variables, respectively, in class . The posterior distribution can then be represented as

$$P(Z, U, \mu \mid X, V, \alpha_v, k_v) = \prod_{v=1}^{V} P(Z^v, U_v, \mu_v \mid X^v, \alpha_v) \tag{5}$$

Where $U_v = \{\mu_{v,1}, \mu_{v,2}, \ldots, \mu_{v,t}, \ldots\}, \mu_v = \{\mu_{v,1}, \ldots, \mu_{v,t}, \ldots\}$. We utilize hidden variables, denoted as $U_v = \{\mu_{v,1}, \mu_{v,2}, \ldots, \mu_{v,t}, \ldots\}$ and $\mu_{v,} = \{\mu_{v,1}, \ldots, \mu_{v,t}, \ldots\}$, in order to delineate specific topics associated with label v, we independently learn the Dirichlet Process Mixture (DPM) for each label v , separate from other classes. To facilitate the learning process of the DPM for class$v$, we introduce a mean-field variational method. Let N_v denote the number of instances labeled as v. However, in DPM models, the value of z_n can be unbounded. Therefore, the variational distribution requires truncation. We set the truncation level **T** to a fixed value and set $q(\mu_{v,T} = 1) = 1,$, indicating that the mixture proportions $\pi_{v,t}(u) = 0$ for $t > 1$. It is crucial to highlight that the model retains its full Dirichlet process nature and is not truncated; solely the variational distribution undergoes truncation. For inferring latent variables and estimating proportions, we utilize the mean-field variational inference method. This approach endeavors to pinpoint a distribution within a simple family that closely mimics the true posterior. We approximate the fully factorized family of distributions over the hidden variables:

$$q(U_v, \mu_v, Z^v \mid \gamma_v, \tilde{\mu}, \tilde{k}, \Phi) = \prod_{t=1}^{T-1} q(\mu_{v,t} \mid \gamma_{v,t} \prod_{t=1}^{T} q(\mu_{v,t} \mid \tilde{\mu}_{v,t}, \tilde{k}_{v,t}) \prod_{n=1}^{N_v} q(z_n \mid \Phi_n) \tag{6}$$

With, $q(\mu_{v,t} \mid \gamma_{v,t}) = \beta(\mu_{v,t}|\gamma_{v,t_1}, \gamma_{v,t_2}), q(\mu_{v,t}|\tilde{\mu}_{v,t}, \tilde{k}_{v,t}) = MvM(\mu_{v,t}|\tilde{\mu}_{v,t}, \tilde{k}_{v,t}),$
$q(z_n|\Phi_n) = Mult(z_n|\Phi_n)(q(z_n = t|\Phi_n) = \Phi_{n,t}).$
Here,$\gamma_{v,t_1}, \gamma_{v,t_2}, \tilde{\mu}_{v,t}, \tilde{k}_{v,t}, \Phi_n$. are the free variational parameters. Utilizing this factorization, we obtain a lower bound $L(\gamma_v, \tilde{\mu}_v, \tilde{k}_v, \Phi)$ for the log likelihood.

$$\begin{aligned} L(\gamma_v, \tilde{\mu}_v, \tilde{k}_v, \Phi) = E_q[\log P(X^v \mid Z^v, U_v, \mu_v)] + E_q[\log P(Z^v \mid U^v)] + E_q[\log P(U_v \mid \alpha_v)] - \\ E_q[\log q(U_v \mid \gamma_v)] - E_q[\log q(Z^v \mid \Phi)] - E_q[\log q(\mu_v \mid \tilde{\mu}_v, \tilde{k}_v)] \end{aligned} \tag{7}$$

To optimize the lower bound of the log-likelihood, we utilize an EM algorithm to iteratively train the model. This entails iteratively executing two steps, namely the E-step and the M-step, until convergence is attained. In the E-step, the lower bound is optimized with respect to each of the free parameters $\gamma_v, \tilde{\mu}_v, \tilde{k}_v, \Phi$ as

$$\gamma_{v,t_1} = 1 + \Sigma_{i=1}^{N_v} \Phi_{i,t} \tag{8}$$

$$\gamma_{v,t_2} = \alpha_v + \Sigma_{i=1}^{N_v} \Sigma_{j=t+1}^{T} \Phi_{i,j} \tag{9}$$

$$\Phi_{n,t} \propto \exp(S_{n,t}) \tag{10}$$

$$\tilde{k}_{v,t} = \sum_{n=1}^{N} k_{v,t} \Phi_{n,t} \tilde{\mu}_{v,t}^T x_n \tag{11}$$

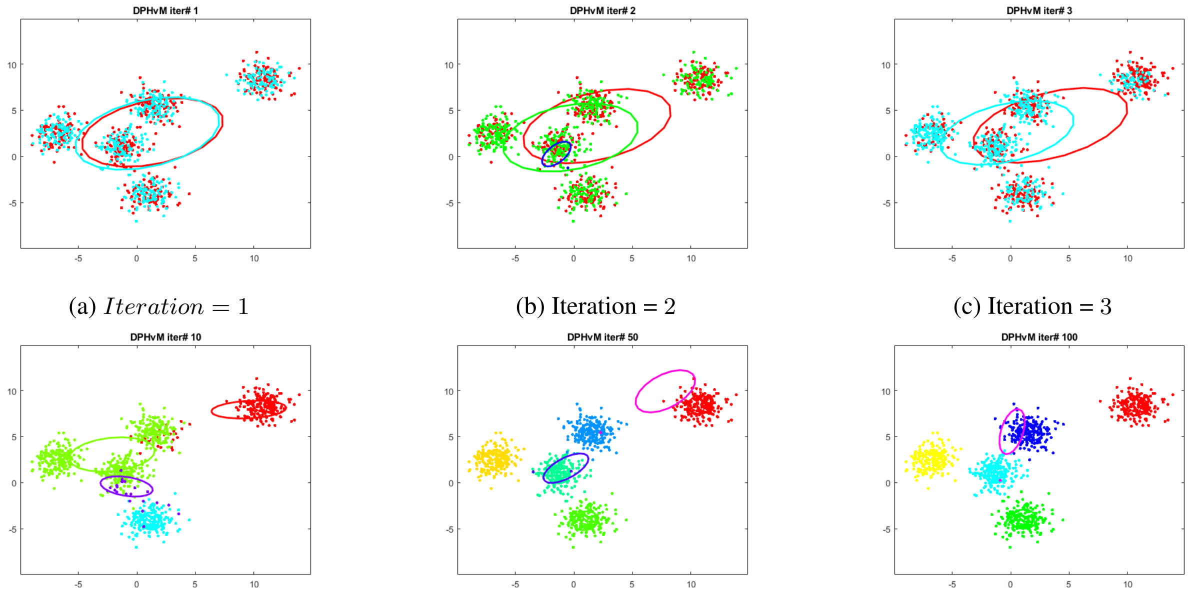
The EM procedure involves iteratively alternating between E and M steps until a suitable convergence criterion is met. Post-training, our model uncovers inherent topics represented by $\tilde{\mu}_{v,t}$ and $\tilde{k}_{v,t}$ within each label v. Furthermore, the concentration parameter $\tilde{k}_{v,t}$ signifies the density of instances around $\tilde{\mu}_{v,t}$. Notably, in the Dirichlet Process Mixture (DPM), a majority of instances tend to cluster around a limited number of topics, as indicated by T_v, representing the number of topics for each label. Consequently,$T = \sum_{v=1}^{V} T_v$ topics are identified from the training set (see algorithm 1).

Algorithm 1 : DPM-MvM

Require: Set χ of data points
Ensure: Inilialize randomly $\mu_0, k_0, \gamma_{t1}, \gamma_{t2}, \mu_t, k_t, \Phi(t = 1 \ldots T; n = 1 \ldots N)$

repeat
The E Step (Expectation)
  for $t = 1$ to T do **do**
    $\gamma_{t1} = 1 + \sum_{i=1}^{N} \Phi_{i,t}$
    $\gamma_{t2} = \alpha + \sum_{i=2}^{N} \sum_{j=t+1}^{T} \Phi_{i,j}$
    $\mu_t = \frac{\sum_{n=1}^{N} k\Phi_{n,t} x_n + k_0\mu_0}{\| \sum_{n=1}^{N} k\Phi_{n,t} x_n + k_0\mu_0 \|}$
    $k_t = \sum_{n=1}^{N} k\Phi_{n,t} x_n + k_0 \mu_t^T x_n + k_0 \mu_t^T \mu_0$
    for $n = 1$ to T do **do**
      Compute $\Phi_{n,t}$ in (10)
    end for
  end for
  repeat
Step M (Maximization)
    Compute μ_0,k_0 in (11)
    Until Convergence

During the testing phase, each unlabeled instance is generated from a mixture of T von Mises distributions learned during training. Probabilities of x_m pertaining to topics $\theta_m = \{\theta_1, ..., \theta_T\}$ are established based on prior studies. For classification, the label of x_m is deduced from the cumulative probabilities assigned to the topics of each label. Additionally, our approach introduces a novel representation θ for instances within the topical space, offering potential applications like dimension reduction and data visualization. Each label is distinguished by its specific topics, thus highlighting the discriminative nature within the topical space through our methodology.



(a) $Iteration = 1$ (b) Iteration = 2 (c) Iteration = 3

(d) Iteration = 10 (e) Iteration = 50 (f) Iteration = 100

Figure 3. Graphical Representation of Data Using the DPM-MvM Approach Across Six Iterations

5. Experimental Design

In this section, we present a series of experiments aimed at assessing the merits of our approach from various perspectives. We aim to furnish concrete evidence of its capability to discern patterns in the thematic space and to classify with efficiency. Furthermore, we seek to highlight its proficiency in outlier detection during the learning process. Additionally, we will conduct a comparative analysis of our approach against some of the most advanced supervised methods. Our primary objective with the simulated data is to validate the accuracy of our implementations by generating a dataset that adheres to specified von Mises (vM) distributions. We introduce a synthetic dataset comprising 500 points in a two-dimensional space, divided into a total of 5 components, each containing 100 points. The mean directions of the components are defined as follows: $\mu_1 = (1, 0)$,$\mu_2 = (0, \sqrt{-1})$,$\mu_3 = (0, \frac{-1}{\sqrt{2}})$,$\mu_4 = (\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}})$,$\mu_5 = (1, 1)$, with a concentration parameter $k = 5$ for each component. We utilize this dataset to demonstrate the effectiveness of our algorithm in identifying cluster structures (see Figure 3).

The CNAE-9 [27] dataset is extracted from a text mining problem. The dataset contains 1080 free text business descriptions of Brazilian companies. The goal is to classify these descriptions in 9 categories. The features are 856 word frequency records. IoT customers are typically enterprises. Therefore, its description is quite useful in order to classify target customers. The CNAE-9 dataset also had the instances ordered. Thus the samples were shuffled randomly.

Table 1. Detection performance by different approaches for the CNAE-9 dataset and our simulated dataset for the DPL-MvM approach

Methods	F	Accuracy (%)
Deep Mixtures of Dirichlet-Multinomials (k2 = 5) [28]	92.50	0.975
VMM-RJMCMC [29]	74.19	76.99

Methods	F	Accuracy (%)
FMcDBMM [30]	85.01	90.94
VMM-EM [31]	73.90	75.18
InVMM-MCMC [32]	81.81	83.91
GDMM-En [33]	76.87	78,93
InVMM [34]	79.34	81.23
DPM-MvM	95.95	98.82

Our DPM-MvM approach has clearly outperformed all other evaluated methods in terms of performance, as illustrated in the table below. With an exceptional categorization accuracy of 98.82% and a highly precise estimation of the number of clusters, our method has demonstrated remarkable results (refer to Table 1). The DPM-MvM method achieves an accuracy of 95.95%, surpassing all other evaluated methods. Competing methods, such as FMcDBMM (85.01%) and Mixtures of Dirichlet-Multinomials (88.39%), exhibit lower performance, thereby affirming the robustness of our approach. Regarding the estimation of the number of clusters, DPM-MvM attains an impressive accuracy of 98.82%. This result is significantly higher than those of other methods, including Mixtures of Dirichlet-Multinomials, which, despite its strong performance (97.14%), does not match the efficacy of our approach. These results underscore the significant advantages of employing a combined Dirichlet Process Model (DPM) and a mixture of von Mises distributions (MvM). The DPM provides flexibility by allowing a non-parametric estimation of the number of clusters, while von Mises distributions are particularly well-suited for modeling angular or circular data, thereby enhancing overall precision.

Our DPM-MvM approach demonstrated superior performance compared to all other evaluated methods, achieving the highest categorization accuracy of 98.82% and the most precise estimation of the number of clusters. These findings underscore the significant benefits of employing both the Dirichlet process model and a mixture of von Mises distributions.

Analysis of Computational Complexity

The initialization of the model involves assigning each data point to an initial cluster. This step typically has a linear complexity with respect to the number of data points, $O(N)$, where N is the number of data points. This initialization is crucial for commencing the iterative process of parameter estimation. Subsequently, the parameters are estimated using the Expectation-Maximization (EM) Algorithm: The EM algorithm, or its variants (such as Variational EM), are often used to estimate parameters in Dirichlet Process Mixture Models (DPMMs). For each iteration:

• **E-Step:** Calculation of responsibilities (posterior probabilities) for each data point with respect to each cluster. This step requires O(KN) operations, where K is the number of clusters and N is the number of data points.

• **M-Step:** Updating the parameters of each von Mises distribution based on the current responsibilities. This step also involves a complexity of O(KN).

As the Dirichlet Process evolves, the number of mixture components K can dynamically increase. This increase in K results in higher computational costs for likelihood calculations and parameter updates. When K becomes large, the complexity of updating cluster assignments and computing likelihoods can become $O(K^2N)$ due to the increased number of components to consider.

Challenges for Large-Scale Datasets

For large-scale datasets, several complexity issues arise:

• As the dataset size N and the number of components K increase, computational and memory requirements grow significantly. The O(KN) complexity per iteration can become prohibitive, especially if K increases with the dataset size. Additionally, storing parameters and intermediate results for a large number of clusters and data points can lead to substantial memory consumption. This requirement can be a limiting factor when managing large datasets. Furthermore, the time required for the algorithm to converge may increase with the dataset size and the number of components. Longer training times may be necessary to ensure that the algorithm has sufficiently converged.

Implementation of DPM-MVM in R

In this study, we employed the R programming language along with its associated libraries to implement the Dirichlet Process Mixture Model with von Mises distributions (DPM-MVM). The implementation utilizes several key R packages, each serving a crucial role:
DirichletReg: This package provides advanced tools for modeling mixtures based on Dirichlet processes. It offers essential functionalities for specifying and fitting Dirichlet Process Mixture Models (DPMMs), which are critical for handling the non-parametric nature of the DPM-MVM approach. MCMCpack: This library is used for Bayesian modeling and Monte Carlo simulations. It facilitates the estimation of posterior distributions through Markov Chain Monte Carlo (MCMC) methods, which are vital for approximating the parameters of the DPM-MVM model.
circular: This package is designed specifically for analyzing angular data and von Mises distributions. It includes a variety of functions for managing and interpreting circular data, which are indispensable for the accurate implementation of von Mises distributions within the DPM-MVM framework. By integrating these R packages, we effectively implemented the DPM-MVM algorithm, leveraging the specialized tools and functions each package provides. This comprehensive approach addresses multiple facets of the model, including Dirichlet process mixture modeling, Bayesian inference, and circular data analysis.

6. Conclusion

The novel DPM-MvM model offers a comprehensive representation for a vector of multivariate angular data, incorporating varying degrees of variation and capitalizing on dependencies among the random functions. It serves as a fundamental and valuable tool for handling multivariate angular data. Additionally, our approach introduces a promising and resilient framework for analyzing intricate circular datasets. By harnessing the adaptability of the Dirichlet Process Mixture, the model can automatically adjust to the underlying structure of the data, while the von Mises Mixture Distributions adeptly capture the circular dependence and variability within the multivariate angular observations. This innovative methodology not only tackles the challenges inherent in angular data but also presents a probabilistic framework facilitating uncertainty quantification, robustness, and enhanced model fitting. Consequently, this research makes a substantial contribution to the realms of circular statistics and Bayesian modeling.

Conflicts of interest

The author declares that there is no conflict of interest.

REFERENCES

1. W. Alhakami, A. ALharbi, S. Bourouis, R. Alroobaea, and N. Bouguila, *Network anomaly intrusion detection using a nonparametric Bayesian approach and feature selection*, IEEE Access, vol. 7, pp. 52181-52190, 2019.
2. R. van de Schoot, S. Depaoli, R. King, B. Kramer, K. Märtens, M. G. Tadesse, and C. Yau, *Bayesian statistics and modelling*, Nature Reviews Methods Primers, vol. 1, no. 1, p. 1, 2021.
3. Y. Chen, J. Zhang, and C. K. Yeo, *Privacy-preserving knowledge transfer for intrusion detection with federated deep autoencoding Gaussian mixture model*, Information Sciences, vol. 609, pp. 1204-1220, 2022.
4. K. Goel, N. Michael, and W. Tabib, *Probabilistic Point Cloud Modeling via Self-Organizing Gaussian Mixture Models*, IEEE Robotics and Automation Letters, vol. 8, no. 5, pp. 2526-2533, 2023.
5. O. J. Fisher, A. Rady, A. A. El-Banna, N. J. Watson, and H. H. Emaish, *An image processing and machine learning solution to automate Egyptian cotton lint grading*, Textile Research Journal, vol. 93, no. 11-12, pp. 2558-2575, 2023.
6. T. Zhou, K. Law, and D. Creighton, *A weakly-supervised graph-based joint sentiment topic model for multi-topic sentiment analysis*, Information Sciences, vol. 609, pp. 1030-1051, 2022.
7. S. Zhang, L. Yao, A. Sun, and Y. Tay, *Deep learning based recommender system: A survey and new perspectives*, ACM Computing Surveys (CSUR), vol. 52, no. 1, pp.
8. A. Garre, M. H. Zwietering, and H. M. den Besten, *Multilevel modelling as a tool to include variability and uncertainty in quantitative microbiology and risk assessment: Thermal inactivation of Listeria monocytogenes as proof of concept*, Food Research International, vol. 137, p. 109374, 2020.
9. M. A. T. Figueiredo and A. K. Jain, *Unsupervised learning of finite mixture models*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 3, pp. 381-396, 2002.
10. F. Culasso, B. Gavurova, E. Crocco, and E. Giacosa, *Empirical identification of the chief digital officer role: A latent Dirichlet allocation approach*, Journal of Business Research, vol. 154, p. 113301, 2023.
11. P. Darling (Ed.), *SME mining engineering handbook*, vol. 1, SME, 2011.
12. D. M. Blei and M. I. Jordan, *Variational inference for Dirichlet process mixtures*, 2006.

13. N. Bouguila and D. Ziou, *A Dirichlet process mixture of generalized Dirichlet distributions for proportional data modeling*, IEEE Transactions on Neural Networks, vol. 21, no. 1, pp. 107-122, 2009.
14. S. Bourouis, M. Al Mashrgy, and N. Bouguila, *Bayesian learning of finite generalized inverted Dirichlet mixtures: Application to object classification and forgery detection*, Expert Systems with Applications, vol. 41, no. 5, pp. 2329-2336, 2014.
15. S. Benlakhdar, M. Rziza, and R. O. H. Thami, *In-depth Analysis of von Mises Distribution Models: Understanding Theory, Applications, and Future Directions*, Statistics, Optimization Information Computing, vol. 12, no. 4, pp. 1210-1230, 2024.
16. S. Benlakhdar, M. Rziza, and R. O. H. Thami, *Statistical modeling of directional data using a robust hierarchical von mises distribution model: perspectives for wind energy*, Computational Statistics, pp. 1-21, 2022.
17. M. Wang, Z. B. Abrams, S. M. Kornblau, and K. R. Coombes, *Thresher: determining the number of clusters while removing outliers*, BMC Bioinformatics, vol. 19, pp. 1-15, 2018.
18. P. K. Elango and K. Jayaraman, *Clustering images using the latent Dirichlet allocation model*, University of Wisconsin, pp. 1-18, 2005.
19. J. W. Miller and M. T. Harrison, *Mixture models with a prior on the number of components*, Journal of the American Statistical Association, vol. 113, no. 521, pp. 340-356, 2018.
20. J. A. Martín-Fernández, K. Hron, M. Templ, P. Filzmoser, and J. Palarea-Albaladejo, *Bayesian-multiplicative treatment of count zeros in compositional data sets*, Statistical Modelling, vol. 15, no. 2, pp. 134-158, 2015.
21. Y. Li, *Flexible Bayesian Nonparametric Modeling for Time-to-Event Data*, Doctoral dissertation, University of California, Santa Cruz, 2024.
22. H. Hihn and D. A. Braun, *Online continual learning through unsupervised mutual information maximization*, *Neurocomputing*, vol. 578, 127422, 2024.
23. S. Han, Y. Yan, J. Bao, and B. Huang, *A big data-driven predictive control approach for nonlinear processes using behaviour clusters*, *Journal of Process Control*, vol. 140, 103252, 2024.
24. J. Deppner and M. Cajias, *Accounting for spatial autocorrelation in algorithm-driven hedonic models: A spatial cross-validation approach*, *The Journal of Real Estate Finance and Economics*, vol. 68, no. 2, pp. 235–273, 2024.
25. D. Wang, *Statistical Models for Alternative Splicing with Applications to Heterogeneous Disease*, Doctoral dissertation, University of Pennsylvania, 2024.
26. K. V. Mardia and E. S. A. M. Atoum, *Bayesian inference for the von Mises-Fisher distribution*, Biometrika, vol. 63, pp. 203–206, 1976.
27. P. M. Ciarelli and E. Oliveira, *Agglomeration and elimination of terms for dimensionality reduction*, In 2009 Ninth International Conference on Intelligent Systems Design and Applications, pp. 547-552, IEEE, November 2009.
28. L. Anderlucci and C. Viroli, *Mixtures of Dirichlet-multinomial distributions for supervised and unsupervised classification of short text data*, Advances in Data Analysis and Classification, vol. 14, no. 4, pp. 759-770, 2020.
29. O. Amayri and N. Bouguila, *A Bayesian analysis of spherical pattern based on finite Langevin mixture*, Applied Soft Computing, vol. 38, pp. 373-383, 2016.
30. D. Forouzanfar, *Unsupervised Learning with Feature Selection Based on Multivariate McDonald's Beta Mixture Model for Medical Data Analysis*, Doctoral dissertation, Concordia University, 2023.
31. O. Amayri and N. Bouguila, *On online high-dimensional spherical data clustering and feature selection*, *Engineering Applications of Artificial Intelligence*, vol. 26, no. 4, pp. 1386–1398, 2013.
32. O. Amayri and N. Bouguila, *RJMCMC learning for clustering and feature selection of L2-normalized vectors*, In 2016 International Conference on Control, Decision and Information Technologies (CoDIT), pp. 269-274, IEEE, 2016.
33. W. Fan, F. R. Al-Osaimi, N. Bouguila, and J. Du, *Proportional data modeling via entropy-based variational bayes learning of mixture models*, *Applied Intelligence*, vol. 47, pp. 473–487, 2017.
34. W. Fan and N. Bouguila, *Spherical data clustering and feature selection through nonparametric Bayesian mixture models with von Mises distributions*, Engineering Applications of Artificial Intelligence, vol. 94, p. 103781, 2020.