

# Unsupervised and Semi-Supervised Learning

**Series Editor**

M. Emre Celebi, Computer Science Department, Conway, Arkansas, USA

Springer's Unsupervised and Semi-Supervised Learning book series covers the latest theoretical and practical developments in unsupervised and semi-supervised learning. Titles – including monographs, contributed works, professional books, and textbooks – tackle various issues surrounding the proliferation of massive amounts of unlabeled data in many application domains and how unsupervised learning algorithms can automatically discover interesting and useful patterns in such data. The books discuss how these algorithms have found numerous applications including pattern recognition, market basket analysis, web mining, social network analysis, information retrieval, recommender systems, market research, intrusion detection, and fraud detection. Books also discuss semi-supervised algorithms, which can make use of both labeled and unlabeled data and can be useful in application domains where unlabeled data is abundant, yet it is possible to obtain a small amount of labeled data.

Topics of interest include:

- Unsupervised/Semi-Supervised Discretization
- Unsupervised/Semi-Supervised Feature Extraction
- Unsupervised/Semi-Supervised Feature Selection
- Association Rule Learning
- Semi-Supervised Classification
- Semi-Supervised Regression
- Unsupervised/Semi-Supervised Clustering
- Unsupervised/Semi-Supervised Anomaly/Novelty/Outlier Detection
- Evaluation of Unsupervised/Semi-Supervised Learning Algorithms
- Applications of Unsupervised/Semi-Supervised Learning

While the series focuses on unsupervised and semi-supervised learning, outstanding contributions in the field of supervised learning will also be considered. The intended audience includes students, researchers, and practitioners.

More information about this series at <http://www.springer.com/series/15892>

Nizar Bouguila • Wentao Fan  
Editors

# Mixture Models and Applications

 Springer

*Editors*

Nizar Bouguila  
Concordia Institute for Information  
Systems Engineering  
Concordia University  
QC, Montreal, Canada

Wentao Fan  
Department of Computer Science  
and Technology  
Huaqiao University  
Xiamen, China

ISSN 2522-848X                      ISSN 2522-8498 (electronic)  
Unsupervised and Semi-Supervised Learning  
ISBN 978-3-030-23875-9              ISBN 978-3-030-23876-6 (eBook)  
<https://doi.org/10.1007/978-3-030-23876-6>

© Springer Nature Switzerland AG 2020

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG.  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface

Increasingly, business, government agencies, and scientists are confronted with large amounts of heterogenous data that are critical for the daily activities, but not well enough analyzed to get the valuable information and knowledge that they potentially hide. The availability of large data sets has changed the scientific approaches to data mining. This has given rise to the need to develop efficient data modeling tools. Among these approaches, mixture models have become a tool of choice in the last years in many scientific domains [1–3]. This is mainly due to their ability to offer a well-principled approach to clustering. New challenges (e.g., Big Data), new approaches (e.g., deep learning), and new technologies (e.g., cloud computing, Internet of Things, etc.) have added new problems when deploying mixture models in real-life scenarios. And several new frameworks based on mixture models have been proposed. The importance of mixture models as a powerful learning machine is evident by the great plethora of papers dedicated to this subject. Such models are finding applications in almost any area of human endeavor. This includes applications in engineering, science, medicine, and business, just to name a few. At the same time, however, there are a lot of challenges related to the development and application of mixture models. Indeed, very few books present a comprehensive discussion about the application of such models to many real-life domains. The present edited book shows clearly that mixture models may be applied successfully in a variety of applications if well deployed.

The book contains 14 chapters that are grouped into 5 parts, namely, Gaussian-based models (3 chapters), generalized Gaussian-based models (2 chapters), spherical and count data clustering (3 chapters), bounded and semi-bounded data clustering (3 chapters), and image modeling and segmentation (3 chapters). In the first chapter, Parsons presents a Gaussian mixture model approach to classify response types. The parameter estimates obtained from fitting the proposed Gaussian model are used in a naive Bayesian classifier to perform the classification task. In Chap. 2, Berio et al. use Gaussian mixtures for the interactive generation of calligraphic trajectories. The authors exploit the stochastic nature of the Gaussian mixture combined with an optimal control to generate paths with natural variation. The merits of the approach are tested by generating curves and traces that are

similar from a geometrical and dynamical point of views to the ones that can be observed in art forms such as calligraphy or graffiti. In Chap. 3, Calinon presents an interesting overview of techniques used for the analysis, edition, and synthesis of continuous time series, with emphasis on motion data. The author exploits the fact that mixture models allow the decomposition of time signals as a superposition of basis functions. Several applications with radial, Bernstein, and Fourier basis functions are presented in this chapter. A generalization to the Gaussian mixture called multivariate bounded asymmetric Gaussian mixture model is proposed by Azam et al. in Chap. 4. The proposed model is learned via expectation maximization and applied to several real-life applications such as spam filtering and texture image clustering. Another generalization is proposed in Chap. 5 by Najar et al. and applied for online recognition of human action and facial expression as well as pedestrian detection from infrared images. In Chap. 6, Fan et al. tackle the problem of spherical data clustering by developing an infinite mixture model of von Mises distributions. A localized feature selection approach is integrated within the developed model to detect relevant features. The resulting model is learned via variational inference and applied to two challenging applications, namely, topic novelty detection and image clustering. A hybrid generative discriminative framework, based on an exponential approximation to two distributions dedicated to count data modeling, namely, the multinomial Dirichlet and the multinomial generalized Dirichlet, is developed in Chap. 7 by Zamzami and Bouguila. Several SVM kernels are developed within this hybrid framework and applied to the problem of analyzing activities in surveillance scenes. A challenging problem when considering the multinomial Dirichlet and the multinomial generalized Dirichlet distribution in statistical frameworks is the computation of the log-likelihood function. This problem is tackled in Chap. 8 by Daghyani et al. by approximating this function using Bernoulli polynomials. The approach is validated via two clustering problems: natural scene clustering and facial expression recognition. A unified approach for the estimation and selection of finite bivariate and multivariate beta mixture models is developed in Chap. 9 by Manouchehri and Bouguila. The approach is based on minimum message length and deployed to several problems (e.g., sentiment analysis, credit approval, etc.). In Chap. 10, Maanicshah et al. tackle the problem of positive vector clustering by developing a variational Bayesian algorithm to learn finite inverted Beta-Liouville mixture models. Applications such as image clustering and software defect detection are used to validate the model. In Chap. 11, Kalra et al. examine and analyze multimodal medical images by developing an unsupervised learning algorithm based on online variational inference for finite inverted Dirichlet mixture models. The algorithm is validated using challenging applications from the medical domain. Kalsi et al. tackle in Chap. 12 image segmentation problem by integrating spatial information within three mixture models based on inverted Dirichlet, inverted generalized Dirichlet, and inverted Beta-Liouville distributions. The same problem is approached in Chap. 13 by Chen et al. by developing a spatially constrained inverted Beta-Liouville mixture model applied to both simulated and real brain magnetic resonance imaging data. Finally, Chap. 14 by Channoufi et al. presents a flexible statistical model for unsupervised image modeling and

segmentation. The model is based on bounded generalized Gaussian mixtures learned using maximum likelihood estimation and minimum description length principle.

Montreal, QC, Canada  
Xiamen, China

Nizar Bouguila  
Wentao Fan

## References

1. McLachlan, G.J., Peel, D.: *Finite Mixture Models*. Wiley, New York (2000)
2. McNicholas, P.D.: *Mixture Model-Based Classification*. Chapman and Hall/CRC, Boca Raton (2016)
3. Schlattmann, P.: *Medical Applications of Finite Mixture Models*. Springer, Berlin (2009)

# Contents

## Part I Gaussian-Based Models

- 1 **A Gaussian Mixture Model Approach to Classifying Response Types** ..... 3  
Owen E. Parsons
- 2 **Interactive Generation of Calligraphic Trajectories from Gaussian Mixtures** ..... 23  
Daniel Berio, Frederic Fol Leymarie, and Sylvain Calinon
- 3 **Mixture Models for the Analysis, Edition, and Synthesis of Continuous Time Series** ..... 39  
Sylvain Calinon

## Part II Generalized Gaussian-Based Models

- 4 **Multivariate Bounded Asymmetric Gaussian Mixture Model** ..... 61  
Muhammad Azam, Basim Alghabashi, and Nizar Bouguila
- 5 **Online Recognition via a Finite Mixture of Multivariate Generalized Gaussian Distributions** ..... 81  
Fatma Najar, Sami Bourouis, Rula Al-Azawi, and Ali Al-Badi

## Part III Spherical and Count Data Clustering

- 6  **$L_2$  Normalized Data Clustering Through the Dirichlet Process Mixture Model of von Mises Distributions with Localized Feature Selection** ..... 109  
Wentao Fan, Nizar Bouguila, Yewang Chen, and Ziyi Chen
- 7 **Deriving Probabilistic SVM Kernels from Exponential Family Approximations to Multivariate Distributions for Count Data** ..... 125  
Nuha Zamzami and Nizar Bouguila



**8 Toward an Efficient Computation of Log-Likelihood Functions in Statistical Inference: Overdispersed Count Data Clustering** ..... 155  
 Masoud Daghyani, Nuha Zamzami, and Nizar Bouguila

**Part IV Bounded and Semi-bounded Data Clustering**

**9 A Frequentist Inference Method Based on Finite Bivariate and Multivariate Beta Mixture Models** ..... 179  
 Narges Manouchehri and Nizar Bouguila

**10 Finite Inverted Beta-Liouville Mixture Models with Variational Component Splitting** ..... 209  
 Kamal Maanicshah, Muhammad Azam, Hieu Nguyen, Nizar Bouguila, and Wentao Fan

**11 Online Variational Learning for Medical Image Data Clustering** .... 235  
 Meeta Kalra, Michael Osadebey, Nizar Bouguila, Marius Pedersen, and Wentao Fan

**Part V Image Modeling and Segmentation**

**12 Color Image Segmentation Using Semi-bounded Finite Mixture Models by Incorporating Mean Templates** ..... 273  
 Jaspreet Singh Kalsi, Muhammad Azam, and Nizar Bouguila

**13 Medical Image Segmentation Based on Spatially Constrained Inverted Beta-Liouville Mixture Models** ..... 307  
 Wenmin Chen, Wentao Fan, Nizar Bouguila, and Bineng Zhong

**14 Flexible Statistical Learning Model for Unsupervised Image Modeling and Segmentation** ..... 325  
 Ines Channoufi, Fatma Najar, Sami Bourouis, Muhammad Azam, Alrence S. Halibas, Roobaea Alroobaea, and Ali Al-Badi

**Index** ..... 349

# Contributors

**Rula Al-Azawi** Gulf College, Al Maabelah, Muscat, Oman

**Ali Al-Badi** Gulf College, Al Maabelah, Muscat, Oman

**Basim Alghabashi** Concordia Institute for Information Systems Engineering (CIISE), Concordia University, Montreal, QC, Canada

**Roobaea Alroobaea** Taif University, Taif, Saudi Arabia

**Muhammad Azam** Department of Electrical and Computer Engineering (ECE), Concordia University, Montreal, QC, Canada

**Daniel Berio** Goldsmiths, University of London, London, UK

**Nizar Bouguila** Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC, Canada

**Sami Bourouis** Taif University, Taif, Saudi Arabia

Université de Tunis El Manar, LR-SITI Laboratoire Signal, Image et Technologies de l'Information, Tunis, Tunisie

**Sylvain Calinon** Idiap Research Institute, Martigny, Switzerland

**Ines Channoufi** Université de Tunis El Manar, Ecole Nationale d'Ingénieurs de Tunis, LR-SITI Laboratoire Signal, Image et Technologies de l'Information, Tunis, Tunisie

**Wenmin Chen** Department of Computer Science and Technology, Huaqiao University, Xiamen, China

**Yewang Chen** Department of Computer Science and Technology, Huaqiao University, Xiamen, China

**Ziyi Chen** Department of Computer Science and Technology, Huaqiao University, Xiamen, China

**Masoud Daghyani** Department of Electrical and Computer Engineering (ECE), Concordia University, Montreal, QC, Canada

**Wentao Fan** Department of Computer Science and Technology, Huaqiao University, Xiamen, China

**Alrence S. Halibas** Gulf College, Al Maabelah, Muscat, Oman

**Meeta Kalra** Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC, Canada

**Jaspreet Singh Kalsi** Department of Electrical and Computer Engineering (ECE), Concordia University, Montreal, QC, Canada

**Frederic Fol Leymarie** Goldsmiths, University of London, London, UK

**Kamal Maanicshah** Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC, Canada

**Narges Manouchehri** Department of Electrical and Computer Engineering (ECE), Concordia University, Montreal, QC, Canada

**Fatma Najjar** Laboratoire RISC Robotique Informatique et Systèmes Complexes, Université de Tunis El Manar, ENIT, Tunis, Tunisie

**Hieu Nguyen** Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC, Canada

**Michael Osadebey** Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway

**Owen E. Parsons** University of Cambridge, Cambridge, UK

**Marius Pedersen** Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway

**Nuha Zamzami** Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC, Canada

Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

**Bineng Zhong** Department of Computer Science and Technology, Huaqiao University, Xiamen, China