Recursive Parameter Estimation of Non-Gaussian Hidden Markov Models for Occupancy Estimation in Smart Buildings

Fatemeh Rezapoor Nikroo

A Thesis

in

The Department

of

Concordia Institue for Information Systems Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Applied Science (Information Systems Security) at

Concordia University

Montréal, Québec, Canada

December 2022

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Fatemeh Rezapoor Nikroo**

Entitled: **Recursive Parameter Estimation of Non-Gaussian Hidden Markov Models for Occupancy Estimation in Smart Buildings**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Information Systems Security)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

________________________________ Supervisor
Dr. Nizar Bouguila

________________________________ Chair and Examiner
Dr. Mohsen Ghafouri

________________________________ Examiner
Dr. Jamal Bentahar

Approved by ________________________________
Dr. Abdessamad Ben Hamza,, Chair
Department of Concordia Institue for Information Systems Engineering

____________ 2022

____
Dr. Mourad Debbabi, Dean
Faculty of Engineering and Computer Science

Abstract

Recursive Parameter Estimation of Non-Gaussian Hidden Markov Models for Occupancy Estimation in Smart Buildings

Fatemeh Rezapoor Nikroo

A significant volume of data has been produced in this era. Therefore, accurately modeling these data for further analysis and extraction of meaningful patterns is becoming a major concern in a wide variety of real-life applications. Smart buildings are one of these areas urgently demanding analysis of data. Managing the intelligent systems in smart homes, will reduce energy consumption as well as enhance users' comfort. In this context, Hidden Markov Model (HMM) as a learnable finite stochastic model has consistently been a powerful tool for data modeling. Thus, we have been motivated to propose occupancy estimation frameworks for smart buildings through HMM due to the importance of indoor occupancy estimations in automating environmental settings. One of the key factors in modeling data with HMM is the choice of the emission probability. In this thesis, we have proposed novel HMMs extensions through Generalized Dirichlet (GD), Beta-Liouville (BL), Inverted Dirichlet (ID), Generalized Inverted Dirichlet (GID), and Inverted Beta-Liouville (IBL) distributions as emission probability distributions. These distributions have been investigated due to their capabilities in modeling a variety of non-Gaussian data, overcoming the limited covariance structures of other distributions such as the Dirichlet distribution. The next step after determining the emission probability is estimating an optimized parameter of the distribution. Therefore, we have developed a recursive parameter estimation based on maximum likelihood estimation approach (MLE). Due to the linear complexity of the proposed recursive algorithm, the developed models can successfully model real-time data, this allowed the models to be used in an extensive range of practical applications.

Acknowledgments

I would like to express my deepest gratitude to my supervisor Dr. Nizar Bouguila whose endless support enlighten my journey. Dr. Bouguila generously provided me with his knowledge and expertise in the field of my interest in computer science as well as giving me the brilliant opportunity of pursuing my master degree in an outstanding international university abroad. I have always appreciated my teachers who assisted me to go steps forward in my life and now I am deeply indebted to Dr. Bouguila for the academic point I am standing on, such a deep honor to have the chance of being his student in my journey in this world.

Words can not describe how thankful I am to my parents and my sister. No matter they were miles far away from me, I have always been under their pure and deep supports, encouragements, motivations and hope. I have no doubt at this endeavor would not have been possible without them.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Machine learning is a fast-growing field that has resulted in the development of tremendous applications. Leveraging machine learning in smart buildings has been studied extensively. Smart buildings generate massive data which needs to be analyzed for further recommendation to optimize and solve potential operational issues. The idea of smart buildings has been under considerable attention due to the concern about global energy consumption. Building energy efficiency demand intelligent solutions to overcome this problem. The goal of studies on smart buildings is both to reduce energy consumption as well as to enhance the comfort of occupants inside rooms D'Oca, Hong, and Langevin (2018). Based on a recent survey González-Torres, Pérez-Lombard, Coronel, Maestre, and Yan (2022) buildings are responsible for one-third of global energy consumption. Among all the factors affecting this ratio, HVAC systems (heating, ventilation, and air conditioning) play a key role of 38% in building consumption González-Torres et al. (2022). However, apart from the external factors of building equipment, occupants behavior which refers to the activities of users has proven a significant influence on energy usage in smart rooms Jia and Srinivasan (2015). Given the above explanations, a number of studies have been conducted to optimize energy consumption in buildings Jia and Srinivasan (2015); Zamzami, Amayri, Bouguila, and Ploix (2019).

In this context, machine learning models are needed to model data from smart homes that can be further integrated into energy simulation software Jia and Srinivasan (2015) towards substantial energy use reductions and improvements in users' comfort D'Oca et al. (2018). In this work, we have mainly focused on occupancy estimation and prediction in smart buildings and propose HMM-based occupancy models. We approach this problem using HMMs as a powerful statistical machine learning technique. The models have been developed to predict the number of occupants in a smart room according to the information collected from sensors so that the energy performance in smart buildings can be managed. The collected data are seen as observation data that are needed to be modeled and the number of occupants is seen as the hidden states in HMM. We have investigated distributions that are flexible in modeling non-Gaussian data and not limited to any restricted structures. Thus, 5 different non-Gaussian distributions namely GD, BL, ID, GID, and IBL have been explored as the emission probability of HMM which have been described in detail in the following chapters. Due to the importance of learning the parameters of probabilistic models, we have developed approaches for parameter estimation in our models to extract the optimum value for parameters to accurately model the characterization of the observation data. We have formulated recursive parameter estimation algorithms to overcome the difficulties of processing batch datasets due to limited computational resources, memory overload, and high demand for processing real-time data. Developed models have been evaluated based on real-life data of occupancy detection from Machine Learning Repository of University of California Irvine (UCI) and the other one is obtained from an experiment which its testbed was an office in Grenoble Institute of Technology in France. Details of these datasets are represented in the following chapters. For the evaluation of the performance of the frameworks, 4 classification metrics have been considered as 1) accuracy, 2) precision, 3) recall and 4) f-score. The definitions of these metrics are as follow Ali and Bouguila (2022):

$$accuracy = \frac{TP + FN}{TP + TN + FP + FN}$$

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$f - score = \frac{2.precision.recall}{precision + recall}$$

where TP, TN, FP and FN represent true-positive, true-negative, false-positive and false-negative, respectively. TP denotes number of times the classifier predict the positive class as positive, while TN shows the number of predictions where the classifier predict negative labels as negative. FN and FP follow a similar analogy whereby the classifier incorrectly predict positive class as negative and negative class as positive, respectively. Precision quantifies the fraction of positive class predictions to real positive while recall allows us to measures the missed positive predictions Ali and Bouguila (2022) and f-score is considered as a harmonic mean of precision and recall.

1.2 Contributions

The main objective of this thesis is recursive parameter estimation of HMMs-based on various underlying distributions which provide us with more flexibility in modeling a wide variety of data. In this context, the contributions are listed as follows:

- **Recursive Parameter Estimation of Generalized Dirichlet Hidden Markov Models: Application to Occupancy Estimation in Smart Buildings**

  We propose a novel framework of recursive parameter estimation of Generalized Dirichlet HMM. We have also presented the comparison of model performance between this method and Dirichlet HMM on real-life data concerning occupancy estimation in smart rooms. This contribution has been published in *2022 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology* Rezapoor Nikroo, Amayri, and Bouguila (2022b).

- **Recursive Parameter Estimation of Beta-Liouville Hidden Markov Models** We extend our previous work considering Beta-Liouville distribution for modeling observation data in HMM. In this contribution we have also compared the capability of Beta-Liouville HMM and Dirichlet HMM in modeling real-life data. This contribution has been published in *2022 International Conference on Electrical, Computer and Energy Technologies* Rezapoor Nikroo, Amayri, and Bouguila (2022a).

- **HMMs Recursive Parameters Estimation for Semi-Bounded Data Modeling: Application to Occupancy Estimation in Smart Buildings** We further extend our previous works to investigate the use of Inverted Dirichlet, Generalized Inverted Dirichlet, and Inverted Beta-Liouville distributions as HMM emission probabilities for parameter estimation and parameter optimization for semi-bounded data. This contribution has been submitted to *International Conference on Smart Cities and Green ICT Systems, 2023* Rezapoor Nikroo, Amayri, and Bouguila (n.d.).

1.3 Thesis Overview

All the models that we have developed are explained in detail in each chapter as follows:

- In chapter 2, we introduce Generalized Dirichlet HMM and its recursive parameters estimation. We demon- strate the effectiveness of our model in comparison with Dirichlet HMM through both synthetic data and real-life data of occupancy estimation and occupancy detection in smart buildings.
- In chapter 3, the Beta-Liouville HMM has been investigated and compared with Dirichlet HMM using real-life data of occupancy estimation and occupancy detection in smart buildings.
- In chapter 4, we explore the use of Inverted Dirichlet, Generalized Inverted Dirichlet and Inverted Beta-Liouville distributions as the HMM underlying probability distributions to investigate the modeling capabilities of these models for positive vectors.
- In chapter 5, we conclude all of our contributions along with a discussion regarding the future work propositions.

Chapter 2

Recursive Parameter Estimation of Generalized Dirichlet Hidden Markov Models: Application to Occupancy Estimation in Smart Buildings

Hidden Markov model (HMM) is a powerful generative machine learning technique to model sequences. Therefore, analysing the characteristics of this model has been the topic of extensive research. In this chapter we go through parameter estimation of HMM. We apply recursive technique in order to be able to handle real time data without suffering from extensive time complexity and memory usage in calculation. In this context we investigate recursive parameter estimation of generalized Dirichlet HMM via the expectation-maximization algorithm. The generalized Dirichlet HMM is shown to be an interesting alternative to the Dirichlet HMM. Extensive simulations based on synthetic and real data show the effectiveness of the recursive approach for parameter estimation.

2.1 Introduction

With the ever increasing amounts of data in electronic form, the need for automated methods for data analysis continues to grow. The goal of machine learning is to develop methods that can automatically detect patterns in data, and then to use the uncovered patterns to predict future data or other outcomes of interest P.Murphy (2012). Machine learning approaches for classification can be roughly grouped into two main categories : 1) discriminative and 2) generative. These methods depend primarily on the estimation criterion and/or structure of the classification method Jebara (2012). Discriminative methods such as support vector machine focuses only on the conditional relation Jebara (2012) and model the decision boundary between the classes. Generative methods such as Bayesian networks rely on full structured joint probability Jebara (2012) and explicitly model the actual distribution of each class. The HMM is a generative probabilistic model of the joint probability of a collection of random variables with both observations and states which can be seen as stationary Markov distribution Xuan, Zhang, and Chai (2001). HMMs are widely used in sequential analysis such as speech recognition, weather inference, natural language processing, biological sequences modeling, etc. Whilst talking about HMM, the underling structure of observable data has an important role in performance of the model. In most studies the Gaussian distribution or Gaussian mixtures have been pointed out while data in real-life applications, have often non-Gaussian structures and are therefore more appropriate to be modeled by non-Gaussian probability distribution Fan, Wang, and Bouguila (2021). Gaussian HMMs are used in computer vision Pyun, Lim, Won, and Gray (2007), speech recognition Cui and Gong (2003), networking Mu and Wu (2011) and so on. Dirichlet HMM has been proposed as an alternative to Gaussian HMM in several applications such as handwriting recognition Biswas, Bhattacharya, and Parui (2012), anomaly detection Dorj, Chen, and Pecht (2013), speech processing Zhang and Chan (2012), texture classification Epaillard, Bouguila, and Ziou (2014), etc. The Dirichlet distribution provides flexibility in modeling data, but it has some drawbacks such as its restrictive negative covariance structure. The generalized Dirichlet distribution has a more flexible covariance structure than Dirichlet distribution Ali and Bouguila (2019a); Wong (2010). Thus, it is chosen as an underlying distribution of HMM in this chapter.

Learning the parameters of a probabilistic model has been always a critical issue in machine learning. Traditionally learning is performed offline. However, if we have streaming data, we need to perform online learning, so we can update our estimates as each new data point arrives rather than waiting until "the end" Fan and Bouguila (2014); P.Murphy (2012). Therefore, in this chapter, with the use of maximum likelihood as a parameter estimation approach, a novel recursive method is proposed to estimate the parameters of generalized Dirichlet HMM. The method is evaluated with both synthetic and real data.

In section 2, Dirichlet HMM, generalized Dirichlet HMM and their formulas are explained in details, section 3 presents the estimation algorithm that we have chosen, section 4 describes the recursive model and the proposed algorithm, experimental results are shown in section 5 and in section 6 we make a conclusion and explain the possible future work.

2.2 Background

HMM was introduced in a series of statistical papers by Leonard E. Baum Baum and Petrie (1966) and other authors and one of its first applications was in the domain of speech recognition Baker (1975). However, it has been widely used in other fields such as information retrieval Miller, Leek, and Schwartz (1999), topic identification Schwartz, Ima, Kubala, Nguyen, and Makhoul (1997), etc. In the following section we discuss distribution of underling data, HMM and maximum likelihood estimation as an approach to estimate the parameters in details.

2.2.1 Dirichlet Distribution

Dirichlet Distribution is a multivariate generalization of the Beta distribution. With the use of Dirichlet distribution both symmetric and asymmetric data can be modeled. In fact, the Dirichlet distribution can be skewed to the right or to the left or symmetric Bouguila, Ziou, and Vaillancourt (2004).

Consider the random vector $X = (x_1, x_2, ..., x_K)$ which follows a Dirichlet distribution, the joint density function is given by:

$$p(x_1, x_2, ..., x_K) = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \prod_{i=1}^{K} x_i^{\alpha_i - 1},$$

where

$$\sum_{i=1}^{K-1} x_i < 1 \quad and \quad x_K = 1 - \sum_{i=1}^{K-1} x_i \quad where \quad 0 < x_i < 1,\ i = 1...K,$$

$$\alpha_0 = \sum_{i=1}^{K} \alpha_i \quad where \quad \alpha_i > 0,\ i = 1...K.$$

The mean and variance of the Dirichlet distribution and the covariance between x_i and x_j are as follows:

$$E(x_i) = \frac{\alpha_i}{\alpha_0},$$

$$Var(x_i) = \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0 + 1)},$$

$$Cov(x_i, x_j) = \frac{-\alpha_i \alpha_j}{\alpha_0^2(\alpha_0 + 1)}.$$

Dirichlet distribution with the parameter vector $\alpha = (\alpha_1, \alpha_2, ..., \alpha_K)$ can be represented both as a distribution on the hyper-plane $D_K = (x_1, x_2, ..., x_K), \sum_{i=1}^{K} x_i = 1$ in IR_+^K or inside the simplex $D_K = (x_1, x_2, ..., x_K), \sum_{i=1}^{K-1} x_i < 1$ in IR_+^K.

2.2.2 Generalized Dirichlet Distribution

Due to the negative covariance structure of Dirichlet distribution and needs for modeling data with more general covariance structure, generalization of Dirichlet distribution was proposed using the concept of neutrality in statistics. Connor and Mosimann Connor and Mosimann (1969) introduced the concept of neutrality as below:

Consider a random vector of proportions $(P_1, P_2, ..., P_K)$ and the random variables $Z_i,\ i = 1, 2, ..., K$ where $0 \leq Z_i \leq 1$, the proportion P_1 is said to be neutral if $Z_1 = P_1$ is independent of the vector $Z_2 = P_2/[1-P_1], Z_3 = P_3/[1-P_1], ..., Z_K = P_K/[1-P_1]$. Now suppose the random vector X in Dirichlet distribution is completely neutral, then the random variables Z_i are mutually independent. The density function of each Z_i, $i = 1, ...K-1$ is set to be a univariate beta distribution as follow:

$$[Beta(\alpha_i, \beta_i)]^{-1} z_i^{\alpha_i - 1}(1 - z_i)^{\beta_i - 1}$$

where $\alpha_i, \beta_i > 0$ and $Beta(\alpha_i, \beta_i)$ is the Beta function which is defined as follow:

$$Beta(\alpha_i, \beta_i) = \frac{\Gamma(\alpha_i)\Gamma(\beta_i)}{\Gamma(\alpha_i + \beta_i)}$$

Now by the transformation below the x_i can be derived from Z_i:

$$x_i = Z_i[\prod_{m=1}^{i-1}(1 - Z_m)], \quad i = 1, 2, ..., K-1.$$

In this generalization covariance structure is changed to be more flexible due to the fact that the vector of random variables following the generalized Dirichlet distribution is not completely neutral. Suppose $X = (x_1, x_2, ..., x_K)$ is a random vector following a generalized Dirichlet distribution with parameters α and β: $GD(\alpha; \beta)$, where $\alpha = (\alpha_1, \alpha_2, ..., \alpha_K)$ and $\beta = (\beta_1, \beta_2, ..., \beta_K)$. The generalized Dirichlet distribution is defined by the following Boutemedjet, Ziou, and Bouguila (2007);

Epaillard and Bouguila (2014):

$$GD(X|\alpha,\beta) = \prod_{i=1}^{K} \frac{\Gamma(\alpha_i+\beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i)} x_i^{\alpha_i-1}(1-\sum_{r=1}^{i} x_r)^{\eta_i}, \quad (1)$$

where

$$\alpha \in \mathbb{R}_+^K, \quad \beta \in \mathbb{R}_+^K, \quad x \in \mathbb{R}_+^K,$$

$$\sum_{i=1}^{K} x_i < 1,$$

$$\eta_i = \beta_i - (\alpha_{i+1} + \beta_{i+1}) \quad i \in [1, K-1],$$

$$\eta_K = \beta_K - 1.$$

In case $\beta_{i-1} = \alpha_i + \beta_i$ for $2 \leq i \leq K$ we obtain the standard Dirichlet distribution.

Based on what Connor and Mosimann Connor and Mosimann (1969) described, the moments of the generalized Dirichlet distribution are obtained as bellows:

$$M_{j-1} = \prod_{m=1}^{j-1} [(\beta_m + 1)/(\alpha_m + \beta_m + 1)]$$

$$E(x_j) = (\alpha_j/[\alpha_j + \beta_j]) \prod_{m=1}^{j-1} [\beta_m/(\alpha_m + \beta_m)], \quad j = 1, ..., K;$$

$$Var(x_j) = E(x_j)[(\alpha_j + 1)/(\alpha_j + \beta_j + 1)]M_{j-1} - E(x_i), \quad j = 1, ..., K;$$

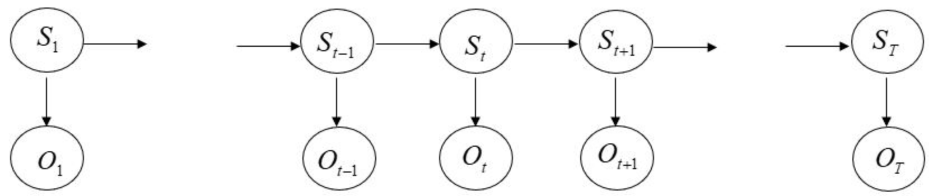

Figure 2.1: Graphical model representation of first-order HMM. The states with S_i representing the Markov hidden state and the states with O_i representing the observation according to each time step $t_i,\ i = 1, 2, ..., T$
.

$$Cov(x_i, x_j) = E(x_j)[(\alpha_i)/[\alpha_i + \beta_i + 1]M_{i-1} - E(x_i)], \quad i = 1, ..., K-1;\ j = i+1, ..., K.$$

2.2.3 Hidden Markov Models

HMM is based on augmenting the Markov model Jurafsky and Martin (2009). Markov model is a stochastic process to represent the sequential data following a Markov property meaning that giving the current knowledge and information, historical information has no impact on the future Jurafsky and Martin (2009). The mathematical formulation of Markov property is as follows:

$$p(X_t + 1 = s | X_t = s_t, X_t - 1 = s_t - 1, ..., X_0 = s_0) = p(X_t + 1 = s | X_t = s) \tag{2}$$

HMM is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved states. Both Markov chain and HMM are based on the idea of random walk in a directed graph, where probability of next step is defined by edge weight. However, in HMM at each step a symbol from set of observations is generated based on the probability distribution of that state. The output of Markov model is a sequence of visited states while in HMM the visited states are hidden and just the sequence of symbols is observable. There are two stochastic processes in HMM: 1) a Markov chain $S_1, S_2, ..., S_T$ which represents hidden states and 2) random variables of $O_1, O_2, ..., O_T$ which are distributed according to the parametric distribution and are observable. These random variables are independent but the parameters of their distribution depend on the state of the Markov chain at that time. The graphical model of HMM is illustrated in Fig. 2.1.

Parameters of HMM are defined as follows Vaičiulytė and Sakalauskas (2020):

(1) N : number of states in the model

(2) $s = (S_1, S_2, ..., S_N)$: set of hidden states

(3) $\pi = (\pi_1, \pi_2, ..., \pi_N)$: set of initial probability distribution of states at time $t = 1$

(4) $v = (V_1, V_2, ..., V_M)$: set of observations where M is number of symbols

(5) $o = (O_1, O_2, ..., O_T)$: set of observed sequence

(6) A_{N*N} : matrix of transition probability in which $a_{i,j}$ indicates the probability of moving from state i to state j at one time step

(7) B: $b_i(k)$ emission probability which is a probability function of observing symbol V_k in state i

HMM is represented by three main parameters as $\lambda = (\pi, A, B)$. The probabilities associated to transition and observation are $a_{i,j} = p(S_j^{t+1}|S_i^t)$ and $b_j(k) = p(V_k^t|S_j^t)$.

2.3 Maximum Likelihood Estimation

Maximum likelihood estimation is a frequentist approach for estimating the parameters of a model given observed data. Expectation-maximization (EM) is an iterative procedure to calculate the parameters until the algorithm meets the convergence criterion. In this chapter, we propose an EM algorithm to maximize the probability density function with respect to the set of parameters to be estimated. In this approach as we consider a recursive estimation, the EM algorithm updates the model after each new observation is introduced without loosing the previous knowledge. With the use of this method we estimate the parameters in order to maximize the generalized Dirichlet probability density function of each observation Nasfi, Amayri, and Bouguila (2020).

2.4 Recursive Model

In this section, we present the approach to recursively estimate the generalized Dirichlet HMM parameters by maximizing the log-likelihood function given as follows:

$$\log GD(X|\alpha,\beta) = \sum_{i=1}^{N} \log\Gamma\left(\alpha_i+\beta_i\right) - \sum_{i=1}^{N}\log\Gamma\left(\alpha_i\right) - \sum_{i=1}^{N}\log\Gamma\left(\beta_i\right) + \sum_{i=1}^{N}(\alpha_i - 1)\log x_i + \sum_{i=1}^{N}\eta_i \log(1-\sum_{r=1}^{i} x_r) \tag{3}$$

Now considering the HMM, log probability density of the observation according to each state of the HMM is:

$$L(\pi;\alpha,\beta) = \log[\sum_{q=1}^{N}\pi_q GD(X|\alpha,\beta)]$$

where π_q is the probability of being in state q.

As mentioned above, the maximum of the log-likelihood function (3) should be calculated, then the partial derivatives of the expression with respect to both α_i and $\beta_i,\ 1\leq i\leq K$ should be derived as below:

$$\frac{\partial \log GD(X|\alpha,\beta)}{\partial\alpha_i} = \psi\left(\alpha_i+\beta_i\right) - \psi\left(\alpha_i\right) + \log(x_i) \tag{4}$$

$$\frac{\partial \log GD(X|\alpha,\beta)}{\partial\beta_i} = \psi\left(\alpha_i+\beta_i\right) - \psi\left(\beta_i\right) + \log(1-\sum_{r=1}^{i}x_r). \tag{5}$$

$\psi(x)$ is digamma function which is defined as the logarithmic derivative of the Gamma function.

Now the probability of the system being in specific state q at time frame t given the observation x is:

$$\frac{\pi_t^{<q>}\log GD(x_t|\alpha^{<q>},\beta^{<q>})}{\sum_{j=1}^{N}\pi_t^{<j>}\log GD(x_t|\alpha^{<j>},\beta^{<j>})},$$

where

$$1 \leq q \leq N \quad and\ N\ is\ the\ number\ of\ states.$$

Therefore the total number of times there is a transition from state q is:

$$\sum_{t=1}^{T} \frac{\pi_t^{<q>} \log GD(x_t|\alpha^{<q>}, \beta^{<q>})}{\sum_{j=1}^{N} \pi_t^{<j>} \log GD(x_t|\alpha^{<j>}, \beta^{<j>})},$$

where

$$1 \leq q \leq N \quad and\ N\ is\ the\ number\ of\ states.$$

Based on the explanation of full likelihood of each observation sequence at Vaičiulytė and Sakalauskas (2020), the weighted averages is applied, therefore, the batch formula of α and β considering specific state q and according to Eqs. (4) and (5) are:

$$\alpha_i = \psi^{-1}[\psi(\alpha_i^q + \beta_i^q) + \frac{\frac{1}{T}\sum_{t=1}^{T} \log(x_t^s) \frac{\pi_t^{<q>} \log GD(x_t|\alpha^{<q>}, \beta^{<q>})}{\sum_{j=1}^{N} \pi_t^{<j>} \log GD(x_t|\alpha^{<j>}, \beta^{<j>})}}{\frac{1}{T}\sum_{t=1}^{T} \frac{\pi_t^{<q>} \log GD(x_t|\alpha^{<q>}, \beta^{<q>})}{\sum_{j=1}^{N} \pi_t^{<j>} \log GD(x_t|\alpha^{<j>}, \beta^{<j>})}}],$$

$$\beta_i = \psi^{-1}[\psi(\alpha_i^q + \beta_i^q) + \frac{\frac{1}{T}\sum_{t=1}^{T} \log(1 - \sum_{r=1}^{i} x_t^r) \frac{\pi_t^{<q>} \log GD(x_t|\alpha^{<q>}, \beta^{<q>})}{\sum_{j=1}^{N} \pi_t^{<j>} \log GD(x_t|\alpha^{<j>}, \beta^{<j>})}}{\frac{1}{T}\sum_{t=1}^{T} \frac{\pi_t^{<q>} \log GD(x_t|\alpha^{<q>}, \beta^{<q>})}{\sum_{j=1}^{N} \pi_t^{<j>} \log GD(x_t|\alpha^{<j>}, \beta^{<j>})}}].$$

where

$$1 \leq q \leq N\ , \quad 1 \leq s, i \leq K.$$

In the proposed recursive method, the parameters α and β are updated based on their estimated values in the previous observation at time $t-1$. The updated equations we derived are as follows:

$$\theta_t^q = \pi_t^q \log GD(x_t|\alpha^q, \beta^q) \ , \quad 1 \le q \le N, \tag{6}$$

$$\omega_t^{<q,s>} = \omega_{t-1}^{<q,s>} + \frac{1}{t}(\frac{\log(x_t^s)\theta_t^q}{\sum_{j=1}^N \theta_t^j} - \omega_{t-1}^{<q,s>}) \ , \quad 1 \le q \le N \ , \ 1 \le s \le K, \tag{7}$$

$$\gamma_t^q = \gamma_{t-1}^q + \frac{1}{t}(\frac{\theta_t^q}{\sum_{j=1}^N \theta_t^j} - \gamma_{t-1}^q) \ , \quad 1 \le q \le N, \tag{8}$$

$$\alpha_t^{<q,s>} = \psi^{-1}[\psi(\alpha_i^q + \beta_i^q) + \frac{\omega_t^{<q,s>}}{\gamma_t^q}] \ , \quad 1 \le s \le K, \tag{9}$$

$$\beta_t^{<q,s>} = \psi^{-1}[\psi(\alpha_i^q + \beta_i^q) + \log(1 - \sum_{r=1}^{s}(exp(\frac{\omega_t^{q,r}}{\gamma_t^{q,r}}))] \ , \quad 1 \le q \le N \ , \quad 1 \le s \le K. \tag{10}$$

Here the proof of the formula of how retrieving ω and γ are provided respectively.

$$\begin{aligned}
\omega_t^{<q,s>} &= \omega_{t-1}^{<q,s>} + \frac{1}{t}(\frac{\log(x_t^s)\theta_t^q}{\sum_{j=1}^N \theta_i^j} - \omega_{t-1}^{<q,s>}) \\
&= \frac{1}{t-1}\sum_{i=1}^{t-1}\frac{\log(x_t^s)\theta_i^q}{\sum_{j=1}^N \theta_i^j} + \frac{1}{t}(\frac{\log(x_t^s)\theta_t^q}{\sum_{j=1}^N \theta_t^j} - \frac{1}{t-1}\sum_{i=1}^{t-1}\frac{\log(x_t^s)\theta_i^q}{\sum_{j=1}^N \theta_i^j}) \\
&= \frac{1}{t}\sum_{i=1}^{t}\frac{\log(x_i^s)\theta_i^q}{\sum_{j=1}^N \theta_i^j},
\end{aligned}$$

$$
\begin{aligned}
\gamma_t^q &= \gamma_{t-1}^q + \frac{1}{t}\left(\frac{\theta_t^q}{\sum_{j=1}^N(\theta_t^j)} - \gamma_{t-1}^q\right) \\
&= \frac{1}{t-1}\sum_{i=1}^{t-1}\frac{\theta_i^q}{\sum_{j=1}^N\theta_i^j} + \frac{1}{t}\left(\frac{\theta_t^q}{\sum_{j=1}^N\theta_t^j} - \frac{1}{t-1}\sum_{i=1}^{t-1}\frac{\theta_i^q}{\sum_{j=1}^N\theta_i^j}\right) \\
&= \frac{1}{t}\sum_{i=1}^{t}\frac{\theta_i^q}{\sum_{j=1}^N\theta_i^j}.
\end{aligned}
$$

The approach used to estimate the parameters of generalized Dirichlet HMM is presented in Algorithm 1. The algorithm includes two main parts: 1) calculating the values of variables θ, ω and γ according to the Eqs. (6)-(8) respectively and 2) updating the parameters $\alpha_t^{<q,s>}$ and $\beta_t^{<q,s>}$ based on Eqs. (9)-(10). At the beginning, we initialize the parameters α, β, transition probability and initial probability of being in each state at time step $t = 1$, by random numbers and the goal was to reach the best value for α and β such that our proposed model can well define the hidden states. In this recursive classification algorithm, we receive the observation data in real time and update the parameters α and β and repeat the process in an iteration loop until we meet the termination criterion (the difference between the parameter values in iteration t and $t-1$ is less than a threshold $\epsilon = 0.5$). For this purpose, we split the data sets to training and testing data, for the training part at each iteration, do the E-step and M-step explained in Algorithm 1 until the termination condition is met. At this step we have the updated α and β ready to use for test data set in order to find the hidden states which are considered as the labels. In the testing part, we use the Bayes rule to find the closest state of each row of data. In fact, using the Bayes rule allows to calculate the probability of assigning each state to the specific row of data and consider the highest probability as the most probable hidden state for that observed data.

2.5 Experimental Results

In this section, validation of the proposed recursive parameter estimation of generalized Dirichlet HMM using real and synthetic data sets is done. In addition, a comparison between Dirichlet HMM

Algorithm 1 Recursive expectation maximization algorithm for generalized Dirichlet HMM parameter estimation

Result: $\alpha_t^{<q,s>}$ and $\beta_t^{<q,s>}$, $\quad 1 \leq t \leq T, 1 \leq s \leq K, 1 \leq q \leq N$
Initialization: initial parameters α, β, first-probability of each state and transition-probability between states
Input: each row of data set; x_t, $1 \leq t \leq T$

```
while ε > α_t − α_{t−1} and ε > β_t − β_{t−1} do
    Normalize input vector x_t;
    E-step
    Calculate values of θ_t^q, ω_t^{<q,s>} γ_t^q;
    M-step:
    Update values of α_t^{<q,s>}, β_t^{<q,s>}
end while
```

and generalized Dirichlet HMM is performed. In all the experiments mentioned bellow, the initial probability of the states at time $t = 1$ is set equally the same for all the states, which means having n states the probability of being in each state $s_i,\ i = 1, 2, ..., n$ is $\frac{1}{n}$. The transition matrix indicating the probability of moving from state s_i to state $s_j,\ i, j = 1, 2, ..., n$ are filled by random numbers. In addition, the parameters α and β are assigned by random numbers for each states for each dimension of the data. The termination criterion is set to $\epsilon = 0.5$.

2.5.1 Synthetic Data

To evaluate the capability of our model we have generated data set arising from generalized Dirichlet distribution. In this section we also provide a comparison between Dirichlet HMM and generalized Dirichlet HMM. This data set has 5000 row of data with 5 dimensions in which the last dimension indicates the hidden states. The first 4 features of each row of this data is considered as our observation data, then we apply the same method explained in Algorithm 1 on it. Based on the definition of the Dirichlet in section 2.1 and generalized Dirichlet distribution in section 2.2, the input data should fall in the range of $(0, 1)$. To meet this condition for Dirichlet HMM we apply softmax normalization formula as follows:

$$softmax_i(X) = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}}$$

However; in case of generalized Dirichlet HMM, the sum of input data should be less than 1, therefore, we first check if the summation of all the features is 1 or above, then add a random number in range of $(0, 1)$ then apply softmax and at the end omit the added dimension. In this data set we consider the same probability for the initial probability distribution of states and complete the transition matrix randomly, but the same random values are considered for both Dirichlet HMM and generalized Dirichlet HMM. The final estimated parameters are mentioned in Table 2.1. In addition, in Table 2.2 we show the evaluation of both HMM models using this data set. The results shown in Table 2.2 are the average of results over 20 runs using different initial values which were assigned randomly.

Table 2.1: Parameter values for Dirichlet HMM and generalized Dirichlet HMM parameter estimation algorithm for synthetic data (2-state and 4-dimensional).

Model	Number of States	Parameters
GD-HMM	2	$\alpha = \begin{bmatrix} 7.23828619 & 0.45192005 & 0.2373394 & 0.14765902 \\ 7.21223305 & 0.46147783 & 0.23963648 & 0.14796613 \end{bmatrix}$
		$\beta = \begin{bmatrix} 15.64232241 & 0.95157966 & 0.55731075 & 0.40486111 \\ 15.99296918 & 0.98423038 & 0.56608161 & 0.40332316 \end{bmatrix}$
D-HMM	2	$\alpha = \begin{bmatrix} 1.88495201 & 0.90703827 & 0.226631 & 0.69773528 \\ 0.58775399 & 0.62743067 & 0.58916982 & 0.46667091 \end{bmatrix}$

Table 2.2: Accuracy, f-score, precision and recall in percent for Dirichlet HMM and generalized Dirichlet HMM for synthetic data (2-state and 4-dimensional).

Model	Accuracy	F-score	precision	recall
GD-HMM	87.2	87.2	87.34	87.19
D-HMM	73.33	76.53	82.05	73.33

Based on the results shown in Table 2.2, the experimental results with synthetic data shows the great capabilities of our model in estimating recursively the generalized Dirichlet HMM. Contrary to the Dirichlet HMM for which 1100 of observations out of 1500 ones were correctly classified, generalized Dirichlet HMM assigned correct labels (hidden state) to 1300 new observations out of 1500 ones.

2.5.2 Real Data: Occupancy Estimation in Smart Buildings

In this section results of implementation of generalized Dirichlet HMM on two different data sets and its comparison with Dirichlet HMM are provided. Both data sets are related to occupancy detection and estimation. The main motivation is that occupancy detection and estimation can be formalized using HMMs since states are not visible and only the outputs including environmental manifestations such as CO2 level, temperature of the area are provided Nasfi et al. (2020). The first data is Occupancy detection data set Candanedo and Feldheim (2016) which is used for binary classification to compare the two models namely Dirichlet HMM and generalized Dirichlet HMM. Estimated parameters of the Dirichlet HMM and generalized Dirichlet HMM using this data set are shown in Table 2.3. Calculated classification measures are shown in Table 2.4. This data set has been collected from data from sensors of light, temperature humidity and CO_2 as a means to detect occupancy Candanedo and Feldheim (2016). Here the hidden states are status of the room from the aspect of being occupied or not and the observable states are the featureS explained below which are presented as a time series:

- Temperature, in Celsius
- Relative Humidity,
- Light, in Lux
- CO2, in ppm
- Humidity Ratio, Derived quantity from temperature and relative humidity, in kgwater-vapor/kg-air Occupancy, 0 or 1, 0 for not occupied, 1 for occupied status

To have a high-level view of the data set, its distribution according to each feature is displayed in Fig. 2.2.

Table 2.3: Estimated parameters for Dirichlet HMM and generalized Dirichlet HMM for occupancy detection data set.

Data set	Model	Number of States	Parameters
occupancy detection	GD-HMM	2	$\alpha = \begin{bmatrix} 0.18 & 0.18 & 0.51 & 3.02 & 0.08 \\ 0.20 & 0.94 & 0.10 & 0.92 & 0.19 \end{bmatrix}$
			$\beta = \begin{bmatrix} 4.37 & 4.10 & 6.70 & 0.39 & 0.10 \\ 1.00 & 0.93 & 0.58 & 0.40 & 0.84 \end{bmatrix}$
occupancy detection	D-HMM	2	$\alpha = \begin{bmatrix} 0.17 & 0.17 & 0.47 & 3.26 & 0.16 \\ 0.19 & 0.20 & 0.76 & 6.86 & 0.18 \end{bmatrix}$

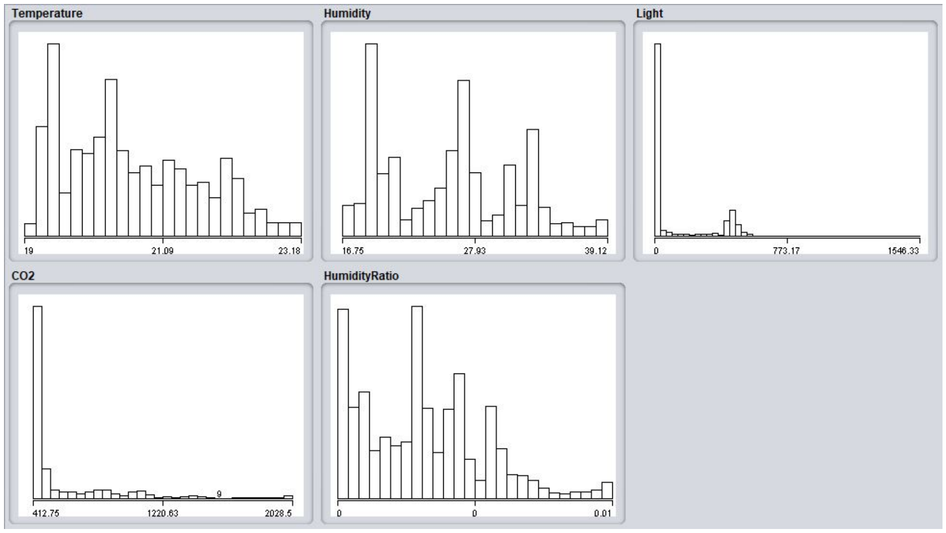

Figure 2.2: Visualization of occupancy detection data according to different features.

Table 2.4: Accuracy, f-score, precision and recall in percent for Dirichlet HMM and generalized Dirichlet HMM applied for occupancy detection data set.

Data set	Model	Accuracy	F-score	precision	recall
occupancy detection	GD-HMM	87.52	85.55	88.56	87.52
occupancy detection	D-HMM	85.74	83.10	86.62	85.74

The second data set that has been used to evaluate the developed model is office-occupancy data set related to smart building energy management. Indoor occupancy estimation plays a key role in the matter of automating environmental settings such as heating, ventilation, air conditioning Ebadat, Bottegal, Varagnolo, Wahlberg, and Johansson (2015); Nasfi et al. (2020); Oldewurtel, Sturzenegger, and Morari (2013) and lightning Candanedo and Feldheim (2016); Nasfi et al. (2020) for smart buildings. Studies showed that this occupancy-based control can in turn save one-third of major energy consumption in buildings Brooks and Barooah (2014); Erickson, Carreira-Perpiñán, and Cerpa (2011); Fan and Bouguila (2012). Unlike the previous application, it is considered as a multi-class classification as the test-bed was an office in Grenoble institute of Technology hosting four people Amayri et al. (2016); Nasfi et al. (2020). This data set is described by four attributes which are motion, power consumption, acoustic-pressure and door-opening. Table 2.5 provides the estimated parameters of the Dirichlet HMM and generalized Dirichlet HMM. In this data set, the hidden states

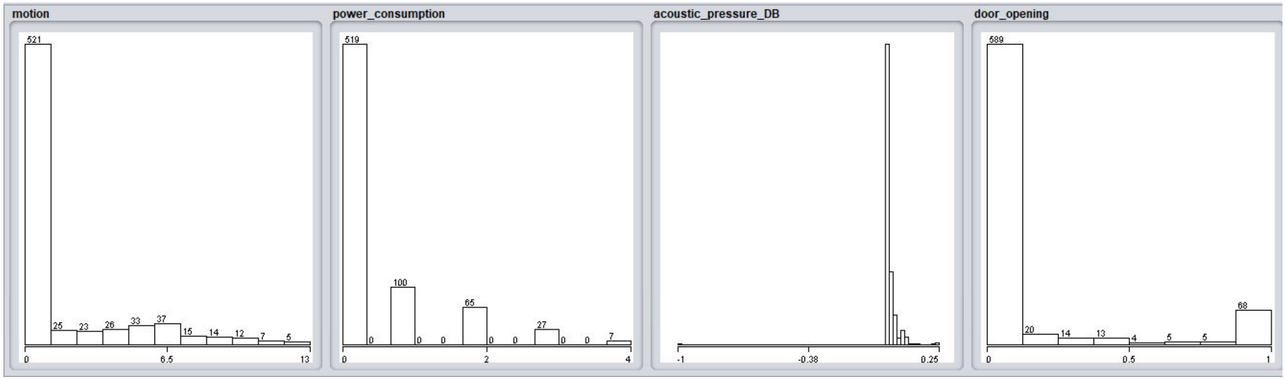

Figure 2.3: Visualization of office-occupancy data according to different features.

are the number of attendants in a room and the observable outputs are the attributes values which are considered as environmental factors. The evaluation results are given in Table 2.6. The data distributions with respect to the different attributes are provided in Fig. 2.3.

Table 2.5: Parameter values for Dirichlet HMM and generalized Dirichlet HMM parameters for office occupancy data.

Data set	Model	Number of States	Parameters
office-occupancy	GD-HMM	5	$\alpha = \begin{bmatrix} 2.10944859 & 0.4042412 & 0.26285416 & 0.25798753 \\ 0.65930223 & 0.2819966 & 0.23687576 & 0.21738184 \\ 0.34578291 & 0.29602738 & 0.37051746 & 0.37653626 \\ 1.0549046 & 0.34915824 & 0.25426651 & 0.24120343 \\ 0.79072965 & 0.49744349 & 0.44195965 & 0.93721427 \end{bmatrix}$ $\beta = \begin{bmatrix} 4.28646559 & 0.86270771 & 0.56043227 & 0.49709727 \\ 0.79759607 & 0.66045755 & 0.75406526 & 0.56017985 \\ 0.60264184 & 0.42642526 & 0.48270544 & 0.37656921 \\ 1.82794749 & 0.73762174 & 0.57707716 & 0.47968072 \\ 0.98236563 & 0.96256723 & 0.25564419 & 0.35323547 \end{bmatrix}$
office-occupancy	D-HMM	5	$\alpha = \begin{bmatrix} 0.16214643 & 0.60979381 & 0.43810455 & 0.27603032 \\ 0.05219396 & 0.80801507 & 0.46464707 & 0.24227922 \\ 1.23935713 & 0.75480023 & 0.64659841 & 0.66507963 \\ 1.57103448 & 1.021154 & 0.84586044 & 0.8520257 \\ 1.76158057 & 1.05946864 & 0.91643249 & 0.96383339 \end{bmatrix}$

Table 2.6: Accuracy, f-score, precision and recall in percent for Dirichlet HMM and generalized Dirichlet HMM for office-occupancy data set.

data set	Model	Accuracy	F-score	precision	recall
office-occupancy	GD-HMM	75.92	73.26	78.69	75.92
office-occupancy	D-HMM	71.76	60.94	52.96	71.76

According to Table 2.4 and Table 2.6, it is clear that the generalized Dirichlet HMM provides better estimation results which in turn provide more accurate hidden states for each observation. The flexibility of generalized Dirichlet over Dirichlet distribution to model the emission probability of HMM is also proven. In the occupancy detection data set, 216 observations analyzed in testing part which 162 of them were classified correctly with generalized Dirichlet HMM, while we have 153 well classified observations using Dirichlet HMM.

2.6 Conclusion

In this chapter we introduced a novel approach for parameter estimation of the generalized Dirichlet HMM and discuss its advantages over Dirichlet HMM. Due to the importance of real-time data processing, real-time analytic has been more in demand, therefore, in the provided model a recursive approach is applied to analyse the data in an online manner. In that case the memory overload and the latency have decreased dramatically. Since the complexity of this recursive model is linear and there is no need to extensive computational resources, it can be widely used in classifications tasks based on a generalized Dirichlet HMM model. Adopting the generalized Dirichlet distribution as an underlying data modeling for emission probability in HMM is encouraged by its capabilities in terms of covariance which is considered as a limit while modeling data using Dirichlet distribution. To learn the parameters, a learning approach based on the generalized Dirichlet distribution combined with HMM concepts was developed. Due to the nature of supervised problem in the context of defining the hidden states for each observation, in the training task, estimated values of parameters are updated as new data arrives during the process of model learning without a need to store the previous results. In the testing part these estimated values are used to calculate the most probable hidden states which can be the best match for each new observation of data using Bays rule. We have proved the capacity of the generalized Dirichlet distribution over Dirichlet distribution to model HMM-related tasks. The efficiency of generalized Dirichlet HMM in comparison with Dirichlet HMM has been validated using synthetic data and real data. It is noteworthy that our model is shown to be very effective in real-life application; estimating the occupancy of office in the context of smart buildings. Future work will be devoted to investigate different initialization approaches in order to improve the model to be less biased to initialization step. As well as, comparing the performance of generalized Dirichlet with other Dirichlet related distributions such as Beta-Liouville. In addition, when it comes to real data in case of smart buildings, experiments can be done to investigate the effectiveness of different features on the final classification result in order to reach more accurate results.

Chapter 3

Recursive Parameter Estimation of Beta-Liouville Hidden Markov Models

This chapter proposes a novel approach to recursively estimate the parameters of Beta-Liouville Hidden Markov Model (HMM). With the rapidly increasing volume of data nowadays, the need for analyzing such data is becoming more urgent. Classification is a a machine learning technique for analyzing data. Therefore, in this chapter, we assume that a given data set can be described as HMM sequences, then apply Beta-Liouville distribution as an emission probability of the HMM. By estimating the parameters of the considered Beta-Liouville HMM using expectation maximization algorithm, we build a machine learning model using Bayes rule to perform classification. Noting that classical learning methods are computationally extensive, we propose an online learning framework for real-time analysis using recursive parameter estimation approach. Both Dirichlet and Beta-Liouville distributions are studied and compared in this research. The Beta-Liouville distribution has proven to be more flexible in terms of data modeling. The effectiveness of the developed model is shown by evaluating it on real data that concern occupancy estimation in smart buildings.

3.1 Introduction

Statistical models are becoming increasingly important as they describe meaningful data pattern Bdiri, Bouguila, and Ziou (2016); Bouguila, Almakadmeh, and Boutemedjet (2012); Boutemedjet et al. (2007); Boutemedjet, Ziou, and Bouguila (2010); Fan and Bouguila (2013b, 2013c, 2015b); Mashrgy, Bdiri, and Bouguila (2014); Oboh and Bouguila (2017). HMM is a statistical model that considers Markov process and unobservable states. With the help of known parameters, hidden states can be identified. HMMs can be viewed as dynamic probabilistic models that have been used in numerous applications such as speech recognition Cui and Gong (2003), texture classification Epaillard et al. (2014), speech processing Zhang and Chan (2012), and occupancy estimation in smart buildings Ai, Fan, and Gao (2014); Guo, Amayri, Bouguila, and Fan (2021); Nasfi et al. (2020). In this chapter we mainly focus on occupancy detection/estimation problem using HMMs. In fact, the accurate determination of occupancy in buildings has been recently estimated to save energy in the order of 30 to 42% Candanedo and Feldheim (2016). We developed a model to first estimate the unknown parameters and second apply the learned model on new data for the classification task. There are two types of learning methods; 1) batch learning and 2) recursive learning. Limited computation resources such as memory and accessing just part of the data set in some cases motivate us to derive recursive algorithm to be able to handle data. Recursive method refers to a situation of continuous model re-estimation and adaption based on a continuous data stream Gepperth and Hammer (2016); Losing, Hammer, and Wersing (2018); Vaičiulytė and Sakalauskas (2020). Recursive learning methods enable us to estimate the model parameters in real time Vaičiulytė and Sakalauskas (2020).

One of the key parameters in HMM is the underlying distribution of observed data. Studies have shown inaccurate data modeling when considering the Gaussian in case the data is clearly non-Gaussian. Thus, we investigate two other distributions which have more flexible structures to model compositional data (multivariate observations confined in a simplex) Bouguila and Ziou (2010) using HMMs; Dirichlet and Beta-Liouville distributions Ali and Bouguila (2019b); Fan and Bouguila (2013d). In analyzing compositional data, scientists have been handicapped by the lack of known distributions to describe various patterns of variability. Aitchison Navarro-Lopez, Linares-Mustaros, and Mulet-Forteza (2022) gives an excellent account of the difficulties in dealing with compositional data and of the inadequacy of the Dirichlet distribution as a model Fang, Kotz, and Wangng (2018). Therefore, we applied Beta-Liouville as an emission probability in our model which provides a more general covariance in comparison with Dirichlet HMM which has restrictive negative covariance.

The chapter proceeds as follows. In section 2, we present the fundamental concepts of Dirichlet distribution then Beta-Liouville distribution and the definition of HMM. In section 3, the proposed estimation algorithm based on maximum likelihood is explained. In the next section, the recursive algorithm of Beta-Liouville HMM is developed. In section 5, experimental results of applying the model on occupancy estimation for smart buildings are presented. Finally, in section 6, the conclusion is given.

3.2 Background

Hidden Markov models (HMMs), are used for statistical modelling of nonstationary stochastic processes Jurafsky and Martin (2009). An HMM is essentially a Bayesian finite state process, with a Markovian prior for modelling the transitions between the states, and a set of state pdfs for the modelling of the random variations of the stochastic process within each state Fan and Bouguila (2013a); Jurafsky and Martin (2009). In this section we provide the definition of two pdfs as the emission probability of HMM and definition of HMM in details.

3.2.1 Dirichlet Distribution

The Dirichlet distribution Bouguila and Ziou (2005); Fan and Bouguila (2015a) is a continuous multivariate probability distribution that is commonly used as a prior to the multinomial in Bayesian statistics Bouguila (2011). Dirichlet distribution can be represented either as a distribution on the hyperplane $B_n = (y_1, ..., y_n) : \sum_{i=1}^{n} y_i = 1$ in IR_+^n, or as a distribution inside the simplex $A_n = (y_1, ..., y_{n-1}) : \sum_{i=1}^{n} y_i \leq 1$ in IR_+^{n-1} Fang et al. (2018). Consider $X = (x_1, x_2, ..., x_K)$ as a vector following a Dirichlet distribution $Dir(\alpha)$ with parameter $\alpha = (\alpha_1, \alpha_2, ..., \alpha_K)$, the probability density function has the following form:

$$p(x_1, x_2, ..., x_K) = \frac{\Gamma(\delta)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \prod_{i=1}^{K} x_i^{\alpha_i - 1} \tag{11}$$

where $\sum_{i=1}^{K-1} x_i < 1, x_K = 1 - \sum_{i=1}^{K-1} x_i$, where $0 < x_i < 1,\ i = 1...K$, and $\delta = \sum_{i=1}^{K} \alpha_i, \alpha_i > 0,\ i = 1...K$.

The mean and variance of the Dirichlet distribution and the covariance between x_i and x_j are as follows:

$$E(x_i) = \frac{\alpha_i}{\delta} \tag{12}$$

$$Var(x_i) = \frac{\alpha_i(\delta - \alpha_i)}{\delta^2(\delta + 1)} \tag{13}$$

$$Cov(x_i, x_j) = \frac{-\alpha_i \alpha_j}{\delta^2(\delta + 1)} \tag{14}$$

3.2.2 Beta-Liouville Distribution

The Beta-Liouville is a parametric generalization of Dirichlet distribution which is a natural model for analysing compositional data Fang et al. (2018); Navarro-Lopez et al. (2022). The Beta-Liouville is a Liouville distribution in which Beta distribution is chosen as a generating density as it has flexible shape and can approximate nearly any arbitrary distribution Fan and Bouguila (2015b). We first present the Liouville distribution and its general form, then describe how Beta-Liouville is extracted from it.

Let's assume that a random variable u is a generating variate, the density function $f(.)$ of u is called the generating density Fan and Bouguila (2013e); Fang et al. (2018). Therefore, we invariably write $X = (x_1, x_2, ..., x_K) \sim L_K[f_.; \alpha_1, \alpha_2, ..., \alpha_K]$ whenever $X = (x_1, x_2, ..., x_K)$ has a Liouville distribution Gupta and Richards (1987). The Liouville distribution has 2 kinds; if $f(.)$ has a noncompact support, we say that $X = (x_1, x_2, ..., x_K)$ has the Liouville distribution of first kind Gupta and Richards (1987). If the support of $f(.)$ is compact, then after scaling $f(.)$ to be supported on $(0, 1)$ the variable range over the simplex is defined as $(x_1, x_2, ..., x_K) : x_i > 0, i = 1, 2, ..., K; \sum_{i=1}^{K} x_i < 1$, then we say $X = (x_1, x_2, ..., x_K)$ has a Liouville distribution of the second type Gupta and Richards (1987).

Now the Liouville distribution with positive parameters $(\alpha_1, \alpha_2, ..., \alpha_K)$ and generating density of $f(.)$ with parameter δ is defined as below Bouguila (2012); Epaillard and Bouguila (2016); Fan and Bouguila (2013c); Fang et al. (2018):

$$p(\vec{X}|\alpha_1, ..., \alpha_K, \theta) = f(u|\delta)\frac{\Gamma(\sum_{i=1}^{K} \alpha_i)}{u^{\sum_{i=1}^{K} \alpha_i - 1}} \prod_{i=1}^{K} \frac{x_i^{\alpha_i - 1}}{\Gamma(\alpha_i)} \tag{15}$$

As defined in second kind Liouville distribution, the variables are described as follows: $\vec{X} = (x_1, ..., x_K)$, $u = \sum_{i=1}^{K} x_i < 1$, $x_i < 0$, $i = 1, ..., K$. After a brief overview of what is Liouville distribution, we go through Beta-Liouville. We choose Beta distribution Bouguila and Elguebaly (2012); Manouchehri, Nguyen, and Bouguila (2019) as the generating density $f(.)$ which means that the generating variate u is distributed as a Beta Fang et al. (2018) with α and β as below:

$$f(u|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} u^{\alpha - 1}(1 - u)^{\beta - 1} \tag{16}$$

By substituting the Eq. (6) into Eq. (5) we obtain the Beta-Liouville distribution as follows:

$$\begin{aligned} BL(\vec{X}|\vec{\theta}) = \frac{\Gamma(\sum_{i=1}^{K} \alpha_i)\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \prod_{i=1}^{K} \frac{x_i^{\alpha_i - 1}}{\Gamma(\alpha_i)} \\ \times (\sum_{i=1}^{K} x_i)^{\alpha - \sum_{i=1}^{K} \alpha_i} (1 - \sum_{i=1}^{K} x_i)^{\beta - 1} \end{aligned} \tag{17}$$

where $\vec{\theta} = (\alpha_1, ..., \alpha_K, \alpha, \beta)$ represents the distribution parameters which are all real and positive. The mean and variance of the Beta-Liouville distribution and the covariance between x_n and x_m are as follows Bouguila (2012):

$$E(x_i) = \frac{\alpha \alpha_i}{(\alpha + \beta)(\sum_{i-1}^{K} \alpha_i)} \tag{18}$$

$$Var(x_i) = \frac{\alpha \alpha_i (\alpha + 1)(\alpha_i + 1)}{(\alpha + \beta)(\alpha + \beta + 1) \sum_{i=1}^{K} \alpha_i + 1} - E(x_i)^2 (\frac{\alpha_i^2}{(\sum_{i=1}^{K} \alpha_i)^2}) \tag{19}$$

$$Cov(x_n, x_m) = \frac{\alpha_n \alpha_m}{\sum_{n=1}^{i} \alpha_n}\left(\frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)(\sum_{i=1}^{K}\alpha_i+1)} - \frac{\alpha^2}{(\alpha+\beta)^2(\sum_{i=1}^{K}\alpha_i)}\right) \quad (20)$$

3.2.3 Hidden Markov Models

In probability theory, Markov models are used to model sequences of stochastic variables following the Markov property. According to the Markov property, Markov models assume that we can predict the probability of some future unit without looking too far into the past Jurafsky and Martin (2009). HMM is a doubly embedded stochastic process with an underlying stochastic process that is not observable (it is hidden), but can only be observed through another set of stochastic processes that produce the sequence of observations Rabiner (1989). N hidden states $s = (S_1, S_2, ..., S_N)$ incarnate the underlying stochastic process that characterize an HMM Ali and Bouguila (2021). The initial probability π indicates the probability of the system following HMM being in each state at time step $t = 0$. There is also a transition matrix which shows the probability of transferring between each two states at specific time t. The observable layer which is considered as the output layer is a set of random variables $o = (O_1, O_2, ..., O_T)$ which are distributed according to the parametric distribution. As a first-order HMM, there will be two assumptions. First, giving the current knowledge and information, historical information has no impact on the future, so the probability of a particular state is only depends on the previous state Jurafsky and Martin (2009) which can be mathematically described as below:

$$p(S_t|S_{t-1}, S_{t-2}, ..., S_1) = p(S_t|S_{t-1})$$

Second, the probability of an output observation $O_t, t = 1, 2, ..., T$ depends only on the current state $S_t, i = 1, 2, ..., T$, not on any other states or any other observations, which can be described as below:

$$p(O_t|S_1, ..., S_T, O_1, ..., O_t, ..., O_T) = p(O_t|S_t)$$

The other important parameter related to HMM is emission probability, which indicates the probability of choosing one of the possible observations in current state as an output sequence. As described in Vaičiulytė and Sakalauskas (2020), parameters of HMM are briefly defined as follows:

(1) N : number of states in the model

(2) $s = (S_1, S_2, ..., S_N)$: set of hidden states

(3) $\pi = (\pi_1, \pi_2, ..., \pi_N)$: set of initial probability distribution of states at time $t = 1$

(4) $v = (V_1, V_2, ..., V_M)$: set of observations where M is number of symbols

(5) $o = (O_1, O_2, ..., O_T)$: set of observed sequence

(6) A_{N*N} : matrix of transition probability in which $a_{i,j}$ indicates the probability of moving from state i to state j at one time step

(7) B: $b_i(k)$ emission probability which is a probability function of observing symbol V_k in state i

HMM is represented by three main parameters as $\lambda = (\pi, A, B)$ Vaičiulytė and Sakalauskas (2020). The probabilities associated to transition and observation are $a_{i,j} = p(S_j^{t+1}|S_i^t)$ and $b_j(k) = p(V_k^t|S_j^t)$ respectively.

3.3 Maximum Likelihood Estimation

There are several methods to estimate the unknown parameters from generative models and one of them is maximum likelihood estimation. It calculates the likelihood of conditional probability of observed data given a probability distribution. So in this section, the model parameters are adjusted to maximize the observation sequence given the model. Expectation Maximization (EM) is an iterative approach to perform maximum likelihood estimation. The classic EM needs to receive all data in advance for each iteration, so in the context of streaming data it will not work, therefore we develop the recursive EM which estimates the parameters in real time Vaičiulytė and Sakalauskas (2020). This method involves 2 main steps. The first one is called E-step and the second one is M-step, these steps are repeated until the algorithm reach the convergence criteria. We will explain these steps in detail in the next section after explanation of our recursive model.

3.3.1 Recursive Model

To derive our recursive parameter estimation algorithm to maximize the likelihood of the Beta-Liouville HMM, we need to consider the log likelihood of the Beta-Liouville function:

$$\log BL(\vec{X}|\vec{\theta}) = \log\Gamma(\sum_{i=1}^{K}\alpha_i) + \log\Gamma(\alpha+\beta) - \log\Gamma(\alpha) - \log\Gamma(\beta) + (\beta-1)\log(1-\sum_{i=1}^{K}x_i)$$
$$+(\alpha - \sum_{i=1}^{K}\alpha_i)\log(\sum_{i}^{K}x_i) + \sum_{i=1}^{K}(\alpha-1)\log x_i - \sum_{i=1}^{K}\log\Gamma(\alpha_i) \tag{21}$$

Now we calculate the log probability of observing each data according to each state of HMM as follows:

$$L(\pi;\vec{\theta}) = \log[\sum_{q=1}^{N}\pi_q BL(\vec{X}|\vec{\theta})] \tag{22}$$

π_q is the probability of being in state q at specific time stamp. Our recursive equations to find the parameters are derived using the batch mode, so we first go through directly finding the parameters based on the fact that we access all the data in the beginning. Thus we need to take the partial derivative of the Eq. (21) with respect to parameters $\vec{\theta} = (\alpha_1, ..., \alpha_K, \alpha, \beta)$ as bellow:

$$\frac{\partial \log BL(\vec{X}|\vec{\theta})}{\partial \alpha_j} = \psi(\sum_{i=1}^{K}\alpha_i) - \log(\sum_{i=1}^{K}x_i) + \log x_j - \psi(\alpha_j) \tag{23}$$

$$\frac{\partial \log BL(\vec{X}|\vec{\theta})}{\partial \alpha} = -\psi(\alpha) + \log(\sum_{i=1}^{K}x_i) + \psi(\alpha+\beta) \tag{24}$$

$$\frac{\partial \log BL(\vec{X}|\vec{\theta})}{\partial \beta} = -\psi(\beta) + \log(1-\sum_{i=1}^{K}x_i) + \psi(\alpha+\beta) \tag{25}$$

ψ is the digamma function to describe the logarithmic derivative of the Gamma function.

Now the probability of the system following Beta-Liouville HMM being in specific state q at time frame t given the observation x is:

$$\frac{\pi_t^{<q>} \log BL(x_t|\alpha< q >)}{\sum_{j=1}^{N} \pi_t^{<j>} \log BL(x_t|\alpha< j >)}$$

where $1 \leq q \leq N$, and N is the number of states. Therefore the total number of times there is a transition from state q is:

$$\sum_{t=1}^{T} \frac{\pi_t^{<q>} \log BL(x_t|\alpha< q >)}{\sum_{j=1}^{N} \pi_t^{<j>} \log BL(x_t|\alpha< j >)}$$

where $1 \leq q \leq N$, and N is the number of states. Based on the explanation of full likelihood of each observation sequence in Vaičiulytė and Sakalauskas (2020), the weighted averages is applied, therefore, the batch formulas of $\vec{\theta} = (\alpha_1, ..., \alpha_K, \alpha, \beta)$ considering specific state q and according to Eqs. (24), (25) and (26) are:

$$\alpha_i = \psi^{-1}[\psi(\sum_{i=1}^{K}(\alpha_i^q)) - \frac{1}{T}\sum_{t=1}^{T} \frac{\pi_t^{<q>} \log BL(x_t|\vec{\theta}^{<q>})}{\sum_{j=1}^{N} \pi_t^{<j>} \log BL(x_t|\vec{\theta}^{<j>})} + \frac{\frac{1}{T}\sum_{t=1}^{T} \log(x_t^s) \frac{\pi_t^{<q>} \log BL(x_t|\vec{\theta}^{<q>})}{\sum_{j=1}^{N} \pi_t^{<j>} \log BL(x_t|\vec{\theta}^{<j>})}}{\frac{1}{T}\sum_{t=1}^{T} \frac{\pi_t^{<q>} \log BL(x_t|\vec{\theta}^{<q>})}{\sum_{j=1}^{N} \pi_t^{<j>} \log BL(x_t|\vec{\theta}^{<j>})}}] \tag{26}$$

$$\alpha = \psi^{-1}[\psi(\alpha+\beta) + \frac{1}{T}\sum_{t=1}^{T} \frac{\pi_t^{<q>} \log BL(x_t|\vec{\theta}^{<q>})}{\sum_{j=1}^{N} \pi_t^{<j>} \log BL(x_t|\vec{\theta}^{<j>})}] \tag{27}$$

$$\beta = \psi^{-1}[\psi(\alpha+\beta) + \frac{\frac{1}{T}\sum_{t=1}^{T} \log(x_t^s) \frac{\pi_t^{<q>} \log BL(x_t|\vec{\theta}^{<q>})}{\sum_{j=1}^{N} \pi_t^{<j>} \log BL(x_t|\vec{\theta}^{<j>})}}{\frac{1}{T}\sum_{t=1}^{T} \frac{\pi_t^{<q>} \log BL(x_t|\vec{\theta}^{<q>})}{\sum_{j=1}^{N} \pi_t^{<j>} \log BL(x_t|\vec{\theta}^{<j>})}}] \tag{28}$$

where $1 \leq q \leq N$, $1 \leq s, i \leq K$. In the proposed recursive method, the parameters α_i α and β are updated based on their estimated values in the previous observation at time $t-1$. The updated equations we derived are as follows:

$$\theta_t^q = \pi_t^q \log BL(x_t|\vec{\theta}^{<q>}) \ , \quad 1 \leq q \leq N \tag{29}$$

$$\omega_t^{<q,s>} = \omega_{t-1}^{<q,s>} + \frac{1}{t}(\frac{\log(x_t^s)\theta_t^q}{\sum_{j=1}^N \theta_t^j} - \omega_{t-1}^{<q,s>}) \ , \quad 1 \leq q \leq N \ , \ 1 \leq s \leq K \tag{30}$$

$$\nu_t^{<q>} = \nu_{t-1}^{<q,s>} + \frac{1}{t}(\frac{\log(1-\sum_{i=1}^K x_i)\theta_t^q}{\sum_{j=1}^N \theta_t^j} - \nu_{t-1}^{<q,s>}) \ , \quad 1 \leq q \leq N \ , \ 1 \leq s \leq K \tag{31}$$

$$\gamma_t^q = \gamma_{t-1}^q + \frac{1}{t}(\frac{\theta_t^q}{\sum_{j=1}^N \theta_t^j} - \gamma_{t-1}^q) \ , \quad 1 \leq q \leq N \tag{32}$$

$$\vec{\alpha}_t^{<q,s>} = \psi^{-1}[\psi(\sum_{i=1}^K(\alpha_i^q) - \gamma_t^q + \frac{\omega_t^{<q,s>}}{\gamma_t^q}] \ , \quad 1 \leq s \leq K \tag{33}$$

$$\alpha_t^{<q>} = \psi^{-1}[\psi(\sum_{i=1}^K(\alpha^q + \beta^q) + \gamma_t^q] \ , \quad 1 \leq q \leq N \tag{34}$$

$$\beta_t^{<q>} = \psi^{-1}[\psi(\sum_{i=1}^K(\alpha^q + \beta^q)) + \frac{\nu_t^{<q>}}{\omega_t^q}] \ , \quad 1 \leq q \leq N \ , 1 \leq s \leq K \tag{35}$$

In the following, the proofs of the formulas of how retrieving ω, ν and γ are provided respectively.

$$\begin{aligned}
&\omega_t^{<q,s>} = \omega_{t-1}^{<q,s>} + \frac{1}{t}(\frac{\log(x_t^s)\theta_t^q}{\sum_{j=1}^N \theta_i^j} - \omega_{t-1}^{<q,s>}) \\
&= \frac{1}{t-1}\sum_{i=1}^{t-1}\frac{\log(x_t^s)\theta_i^q}{\sum_{j=1}^N \theta_i^j} + \frac{1}{t}(\frac{\log(x_t^s)\theta_t^q}{\sum_{j=1}^N \theta_t^j} - \frac{1}{t-1}\sum_{i=1}^{t-1}\frac{\log(x_t^s)\theta_i^q}{\sum_{j=1}^N \theta_i^j}) \\
&= \frac{1}{t}\sum_{i=1}^{t}\frac{\log(x_i^s)\theta_i^q}{\sum_{j=1}^N \theta_i^j}
\end{aligned}$$

$$\begin{aligned}
&\nu t^{<q>} = \nu_{t-1}^{<q,s>} + \frac{1}{t}(\frac{\log(1-\sum_{i=1}^K x_i)\theta_t^q}{\sum_{j=1}^N \theta_i^j} - \nu_{t-1}^{<q,s>}) \\
&= \frac{1}{t-1}\sum_{i=1}^{t-1}\frac{\log(1-\sum_{i=1}^K x_i)\theta_i^q}{\sum_{j=1}^N \theta_i^j} + \frac{1}{t}(\frac{\log(1-\sum_{i=1}^K x_i)\theta_t^q}{\sum_{j=1}^N \theta_t^j} \\
&- \frac{1}{t-1}\sum_{i=1}^{t-1}\frac{\log(1-\sum_{i=1}^K x_i)\theta_i^q}{\sum_{j=1}^N \theta_i^j}) = \frac{1}{t}\sum_{i=1}^{t}\frac{\log(1-\sum_{i=1}^K x_i)\theta_i^q}{\sum_{j=1}^N \theta_i^j}
\end{aligned}$$

$$
\begin{aligned}
\gamma_t^q &= \gamma_{t-1}^q + \frac{1}{t}\left(\frac{\theta_t^q}{\sum_{j=1}^N(\theta_t^j)} - \gamma_{t-1}^q\right) \\
&= \frac{1}{t-1}\sum_{i=1}^{t-1}\frac{\theta_i^q}{\sum_{j=1}^N\theta_i^j} + \frac{1}{t}\left(\frac{\theta_t^q}{\sum_{j=1}^N\theta_t^j} - \frac{1}{t-1}\sum_{i=1}^{t-1}\frac{\theta_i^q}{\sum_{j=1}^N\theta_i^j}\right) \\
&= \frac{1}{t}\sum_{i=1}^{t}\frac{\theta_i^q}{\sum_{j=1}^N\theta_i^j}
\end{aligned}
$$

The approach used to estimate the parameters of Beta-Liouville HMM is presented in Algorithm 2. The algorithm includes two main parts: 1) calculating the values of variables θ, ω, ν and γ according to the Eqs. (29)-(32) respectively, and 2) updating the parameters $\vec{\alpha}_t^{<q,s>}$, $\alpha_t^{<q>}$ and $\beta_t^{<q>}$ based on Eqs. (33)-(35). At the beginning, we initialize the parameters $\vec{\alpha}$, α, β, transition probability and initial probability of being in each state at time step $t = 1$, by random numbers and the goal was to reach the best value for $\vec{\alpha}$, α and β such that our proposed model can well define the hidden states. In this recursive classification algorithm, we receive the observation data in real time and update the parameters $\vec{\alpha}$, α and β and repeat the process in an iteration loop until we meet the termination criterion (the difference between the parameter values in iteration t and $t-1$ is less than a threshold $\epsilon = 0.66$). For this purpose, we split the data sets into training and testing data, for the training part at each iteration, do the E-step and M-step explained in Algorithm 2 until the termination condition is met. At this step we have the updated $\vec{\alpha}$, α and β ready to use for test data set in order to find the hidden states which are considered as the labels. In the testing part, we use the Bayes rule to find the closest state of each row of data. In fact, using the Bayes rule allows to calculate the probability of assigning each state to the specific row of data and consider the highest probability as the most probable hidden state for that observed data.

3.4 Experimental Results

In this section, we validate the proposed recursive parameter estimation of Beta-Liouville HMM using real data sets. In addition, a comparison between Dirichlet HMM and Beta-Liouville HMM is performed. In all the experiments mentioned bellow, the initial probability of the states at time $t = 1$ is set equally the same for all the states, which means having n states the probability of being in each state $s_i,\ i = 1, 2, ..., n$ is $\frac{1}{n}$. The transition matrix indicating the probability of moving from

Algorithm 2 Recursive expectation maximization algorithm for Beta-Liouville HMM parameter estimation

Result:$\vec{\alpha}_t^{<q,s>}$, $\alpha_t^{<q>}$ and $\beta_t^{<q>}$, $\quad 1 \le t \le T, 1 \le s \le K, 1 \le q \le N$
Initialization:initial parameters $\vec{\alpha}$, α, β, first-probability of each state and transition-probability between states
Input: each row of data set; x_t, $1 \le t \le T$

while $\epsilon >= \vec{\alpha}_t - \vec{\alpha}_{t-1}, \epsilon >= \alpha_t - \alpha_{t-1}$ and $\epsilon >= \beta_t - \beta_{t-1}$ **do**
Normalize input vector x_t;
    E-step
    Calculate values of $\theta_t^q, \omega_t^{<q,s>}, \nu_t^{<q,s>}, \gamma_t^q$;
    M-step:
    Update values of $\vec{\alpha}_t^{<q,s>}$, $\alpha_t^{<q>}$, $\beta_t^{<q>}$
end while

state s_i to state s_j, $i, j = 1, 2, ..., n$ are filled by random numbers. In addition, the parameters α and β are assigned by random numbers for each states for each dimension of the data. The termination criterion is set to $\epsilon = 0.66$.

3.4.1 Real Data: Occupancy Estimation in Smart Buildings

In this section results of implementation of Beta-Liouville HMM on two different data sets and its comparison with Dirichlet HMM are provided. Both data sets are related to occupancy detection and estimation. The main motivation is that occupancy detection and estimation can be formalized using HMMs since states are hidden and only the outputs including environmental manifestations such as CO2 level, temperature of the area are provided Nasfi et al. (2020). The first data is Occupancy detection data set Candanedo and Feldheim (2016) which is used for binary classification to compare the two models namely Dirichlet HMM and Beta-Liouville HMM. This data set is obtained from Machine Learning Repository of University of California Irvine (UCI). Each row of this data is considered as an online input data for our recursive model. With the use of Beta-Liouville HMM, we first estimate the hidden parameters according to the data then applying the Bayes rules to classify the room as occupied or not. The features of this data set are environmental variables of a room which are explained in details below. Estimated parameters of the Dirichlet HMM and Beta-Liouville HMM using this data set are shown in Table 3.2. Classification measures are shown in Table 3.1. This data set has been collected from sensors of light, temperature humidity and CO_2 to detect occupancy Candanedo and Feldheim (2016). Here the hidden states are status of the room from the aspect of being or not being occupied and the observable states are the features as below which are indicated as a time series: 1) Temperature, in Celsius, 2) Relative Humidity, 3) Light, in Lux, 4) CO2, in ppm, 5) Humidity Ratio, Derived quantity from temperature and relative humidity, in kgwater-vapor/kg-air. This data set has 2 hidden states and 5 dimensions considered as 5 features as explained above. To have a high-level view of the data set, its distribution according to each feature is displayed in Fig. 2.2.

Table 3.1: Accuracy, f-score, precision and recall in percent for Dirichlet HMM and Beta-Liouville HMM applied for occupancy detection data set.

Model	Accuracy	F-score	Precision	Recall
BL-HMM	78.98	69.71	62.39	78.98
D-HMM	75.74	63.10	66.62	75.74

Table 3.2: Estimated parameters for Dirichlet HMM and Beta-Liouville HMM for occupancy detection data set.

Model	Number of States	Parameters
BL-HMM	2	$\alpha = \begin{bmatrix} 0.8245 & 1.5745 \end{bmatrix}$ $\beta = \begin{bmatrix} 0.4709 & 0.1912 \end{bmatrix}$ $\delta = \begin{bmatrix} 0.1503 & 0.0017 & 0.5473 & 0.8021 & 0.3163 \\ 0.1450 & 0.1459 & 0.3444 & 1.4569 & 0.1403 \end{bmatrix}$
D-HMM	2	$\alpha = \begin{bmatrix} 0.5064 & 0.51725 & 0.6047 & 75.7309 & 0.4605 \\ 0.2574 & 0.2601 & 1.5703 & 18.0087 & 0.2432 \end{bmatrix}$

As shown in Table 3.2, in terms of Beta-Liouville HMM, α and β have 2 different values according to the number of states which is 2 as well, while parameter δ is described as a 2×4 matrix. Each row indicates the values for each states respectively and each column is derived based on the features of the data sets.

The second data set that has been used to evaluate the developed model is an office-occupancy data set. Unlike the previous application, it is considered as a multi-class classification as the test-bed was an office in Grenoble institute of Technology hosting four people Nasfi et al. (2020). This data set is described by four attributes which are motion, power consumption, acoustic-pressure and door-opening. Table 3.3 provides the estimated parameters of the Dirichlet HMM and Beta-Liouville HMM. In this data set, the hidden states are the number of attendants in a room and the observable outputs are the attributes values which are considered as environmental factors. The evaluation results are given in Table 3.4. This data set has 5 hidden states and 4 dimensions. The data distributions with respect to the different attributes are provided in Fig. 2.3.

Table 3.3: Parameter values for Dirichlet HMM and Beta-Liouville HMM parameters for office occupancy data.

Model	Number of States	Parameters
BL-HMM	5	$\alpha = \begin{bmatrix} 17.9097 & 0.9136 & 0.2120 & 0.1412 & 0.3679 \end{bmatrix}$ $\beta = \begin{bmatrix} 3.8418 & 0.3897 & 0.5251 & 0.7228 & 0.8409 \end{bmatrix}$ $\delta = \begin{bmatrix} 1.8642 & 0.8634 & 0.6992 & 0.7066 \\ 0.9331 & 0.7037 & 0.6461 & 0.6487 \\ 0.6620 & 0.8044 & 0.1265 & 0.7309 \\ 0.2593 & 0.8164 & 0.8471 & 0.8665 \\ 0.9548 & 0.2715 & 0.8045 & 0.04926 \end{bmatrix}$
D-HMM	5	$\alpha = \begin{bmatrix} 2.7746 & 0.8220 & 0.8797 & 1.0617 \\ 2.5784 & 0.3424 & 0.8706 & 0.8711 \\ 0.4348 & 0.8180 & 0.0873 & 0.0590 \\ 1.8384 & 0.9840 & 0.4037 & 0.3731 \\ 0.3054 & 0.9850 & 0.4771 & 0.5966 \end{bmatrix}$

As shown in Table 3, in this data set due to the fact that the number of hidden states are 5 the matrix indicating the parameter α has 5 rows and 4 columns as we have considered 4 features of the data in our evaluation in terms of Dirichlet HMM. However, in Beta-Liouville HMM we have 3 parameters which are demonstrated in a matrix based on the number of states and number of features. According to Table 3.1 and Table 3.4, it is clear that the Beta-Liouville HMM provides better estimation results which in turn provide more accurate hidden states for each observation. The

Table 3.4: Accuracy, f-score, precision and recall in percent for Dirichlet HMM and Beta-Liouville HMM for office-occupancy data set.

Model	Accuracy	F-score	Precision	Recall
BL-HMM	80.55	74.20	69.27	80.55
D-HMM	75.00	67.66	73.85	75.00

flexibility of Beta-Liouville over Dirichlet distribution to model the emission probability of HMM is also proven. In the occupancy detection data set, 216 observations analyzed in testing part which 172 of them were classified correctly with Beta-Liouville HMM, while we have 162 well classified observations using Dirichlet HMM.

3.5 Conclusion

The goal of this work was to formulate recursive parameter estimation of Beta-Liouville HMM to tackle the modeling of online data. In this chapter, we specifically focus on occupancy states of rooms in smart buildings using real-world data. This model can be widely used in real-life applications due to its 2 key characteristics; 1) recursive algorithm: the complexity of our model is reduced to linear as compared with offline batch learning, thus, this model can be applied on even large data sets without facing limited time and memory computations, 2) Beta-Liouville as an underlying distribution: this distribution provides flexibility and accurate results for a wide range of data sets. Future work could be devoted to applying different initialization methods in order to investigate to what extent the initial values may affect the accuracy of parameter estimation at the end. Furthermore, in the specific context of smart buildings occupancy estimation, experiments can be done to evaluate the effectiveness of each features used by the integration of a feature selection methodology within the proposed model.

Chapter 4

HMMs Recursive Parameters Estimation for Occupancy Estimation in Smart Buildings

Optimizing energy consumption is one of the key factors in smart buildings developments. It is crucial to estimate the number of occupants and detect their presence when it comes to energy saving in smart buildings. In this chapter, we propose a Hidden Markov Models (HMM)-based approach to estimate and detect the occupancy status in smart buildings. In order to dynamically estimate the occupancy level, we develop a recursive estimation algorithm. The developed models are evaluated using two different real data sets.

4.1 Introduction

The daily increase in energy consumption has led to global warming. Global energy demand has continuously increased while building sector has had a major effect in this rapid growth in energy consumption Kim et al. (2022). In this context, smart buildings promise automated systems to control energy consumption. Environmental control systems have been proved as a crucial factor in smart buildings. Another decisive factor in energy consumption concerns the occupants themselves Gaetani, Hoes, and Hensen (2016). Automatic occupancy detection and estimation approaches allow building energy systems to manage the energy consumed. While 35% of USA's energy consumption is attributed to the heating, cooling and ventilation (HVAC) systems Ali and Bouguila (2022); Erickson, Carreira-Perpi, and Cerpa (2014), occupants' behavior has also a major influence on building energy consumption Jiang, Liu, and Yang (2004); Ploix, Amayri, and Bouguila (2021); Sarkar, S.N., and Prasad (2016); Scott et al. (2011); Soltanaghaei and Whitehouse (2016).

Studies have shown that machine learning algorithms are crucial for smart buildings applications Alawadi et al. (2020). For instance, machine learning models have been used to measure HVAC actuation levels Ebadat, Bottegal, Varagnolo, Wahlberg, and Johansson (2013); Ploix et al. (2021). Furthermore, a supervised learning model has been developed in Amayri et al. (2016) to estimate the number of occupants based on sensorial data (e.g. motion detection, power consumption, CO2 concentration sensors, microphone, or door/window positions) Ploix et al. (2021).

Machine learning approaches can be grouped into 3 main categories: 1) generative models such as mixture models and HMMs, 2) Discriminative models such as support vector machine (SVM) and 3) Heuristic-based models which combine the 2 previous families with heuristic information Ploix et al. (2021). In this chapter we propose HMM-based occupancy models. A first crucial factor when deploying a HMM model is the choice of probability density function Nguyen et al. (2019). Thus, we investigate 3 distributions dedicated for semi-bounded data (i.e. positive vectors) which are detailed in section 2. The second important factor is to estimate the unknown parameters which is generally done using maximum likelihood estimation (MLE) within the expectation maximization (EM) framework. Handling real-time data of smart buildings requires continuous processing which is challenging Bouhamed, Amayri, and Bouguila (2022); Elkhoukhi, Bakhouya, Ouadghiri, and Hanifi (2022). Therefore, one of the motivations of this chapter is to propose a novel architecture to cope with real-time data. Online learning techniques provide solutions addressing real-time occupancy estimation to build models that can be continuously updated Amayri, Ploix, Bouguil, and Wurtz (2020). We introduce a recursive algorithm with linear time complexity to further detect and/or estimate the number of occupants.

The rest of this chapter is organized as follows. Section 2, introduces the inverted Dirichlet HMM, generalized inverted Dirichlet HMM and inverted Beta-Liovuille HMM, section 3 investigates the recursive model and the proposed algorithm. Experimental results are presented in section 4 and in section 5 we concludes the chapter.

4.2 Background

In this section, we first discuss HMM in details, then describe the 3 probability density functions dedicated to positive vectors which have been investigated in this chapter.

4.2.1 Hidden Markov Models

HMMs are powerful statistical models. The idea of HMM comes from a limit of Markov model in modeling problems which output is the probabilistic function of the states Rabiner (1989). This function called emission probability is described in details in the following sections. Indeed, we assume more general distributions dedicated to positive vectors to model the output observable data from HMM to have more flexible models. It is noteworthy to mention the denomination of Hidden in HMM refers to the states not the parameters of the model Dymarski (2011). Markov chain is referred to a time-varying random phenomenon Dymarski (2011) meeting the Markov property which indicates that the conditional probability of the forthcoming state is just based on the current state and not on historical information, which can be mathematically formulated as bellow:

$$p(X_{t+1}|X_t, X_{t-1}, ..., X_1) = p(X_{t+1}|X_t) \tag{36}$$

HMM elements are completely defined as follow, however, it is mainly represented by three parameters $\lambda = (\pi, A, B)$ Vaičiulytė and Sakalauskas (2020).

1) The number of states N, each state is defined by S_i such that $S = 1, 2, ..., N$.

2) Vector $\pi = \pi_1, \pi_2, ..., \pi_N$ indicate the probability of being in each state.

3) The number of possible observations M for each state as $V = V_1, V_2, ..., V_M$. In case of our work that observations come from the distributions that we will define later, they are continuous thus, M is infinite.

4) A is the state transition probability matrix such that $a_{i,j}$, $1 \leq i, j \leq N$ is the probability of moving to state j at time $t + 1$ while the model was in state i at time t. The constraints for transition

matrix should be met as bellow Dymarski (2011):

$$a_{i,j} >= 0, \quad \sum_{j=1}^{N} a_{i,j} = 1, \quad 1 \leq i,j \leq N$$

5) B is the matrix showing the observation probability where $b_j(k)$ is the probability of observing V_k in state S_j.

$$b_j(k) = p(V_k^t | S_j^t), \quad 1 \leq i,j \leq N$$

The constraints for continuous observations are defined based on the specific probability distribution considered:

$$b_j(k) = \sum_{m=1}^{M} c_{jm} p(x|\theta)$$

where θ represents the parameters of the defined distribution, c_{jm} is weighting coefficient with $\sum_{m=1}^{M} c_{jm} = 1$

$$b_j(k) >= 0, \quad 1 \leq j \leq N, \quad 1 \leq k \leq M, \quad \sum_{j=1}^{M} b_j(k) = 1$$

4.2.2 Inverted Dirichlet Distribution

Consider $X = (x_1, x_2, ..., x_K)$ as a vector following the inverted Dirichlet distribution $ID(\alpha)$ with parameter $\alpha = (\alpha_1, \alpha_2, ..., \alpha_K, \alpha_{K+1})$, the probability density function has the following form Bdiri and Bouguila (2013):

$$p(x_1, x_2, ..., x_K) = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^{K+1} \Gamma(\alpha_i)} \prod_{i=1}^{K} x_i^{\alpha_i - 1} (1 + \sum_{i=1}^{K} X_i)^{-\alpha_0},$$

where

$$\alpha_0 = \sum_{i=1}^{K+1} \alpha_i,$$

$$x_i > 0, \ i = 1...K$$

The mean and variance of the inverted Dirichlet distribution and the covariance between x_i and x_j are as given in Bdiri and Bouguila (2012); Tiao and Cuttman (1965):

$$E(x_i) = \frac{\alpha_i}{\alpha_{K+1} - 1} \tag{37}$$

$$Var(x_i) = \frac{\alpha_i(\alpha_i + \alpha_{K+1} - 1)}{(\alpha_{K+1} - 1)^2(\alpha_{k+1} - 2)} \tag{38}$$

$$Cov(x_i, x_j) = \frac{\alpha_i \alpha_j}{(\alpha_{K+1} - 1)^2(\alpha_{K+1} - 2)} \tag{39}$$

4.2.3 Generalized Inverted Dirichlet Distribution

Inverted Dirichlet distribution assumes a positive correlation, therefore a generalization of it is introduced to cope with this limitation in order to have the capability of modeling wider range of real-life data Bourouis, Mashrgy, and Bouguila (2014). Consider $X = (x_1, x_2, ..., x_K)$ as a vector following the Generalized Inverted Dirichlet distribution $GID(\alpha; \beta)$ with parameter $\alpha = (\alpha_1, \alpha_2, ..., \alpha_K, \alpha_K)$ and $\beta = (\beta_1, \beta_2, ..., \beta_K)$. The probability density function has the following form Bourouis et al. (2014); lingmnwah (1976):

$$p(x_1, x_2, ..., x_K) = \prod_{i=1}^{K} \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i)} \frac{x_i^{\alpha_i - 1}}{(1 + \sum_{m=1}^{i} x_m)^{\eta_i}} \tag{40}$$

where

$$\eta_i = \beta_i + \alpha_i - \beta_{i+1} \quad i = 1, 2, ..., K,$$

$$\beta_{K+1} = 0,$$

$$x_i > 0, \; i = 1...K.$$

By substituting $\beta_1 = \beta_2 = ... = \beta_{K-1} = 0$ in Eq. (40), inverted Dirichlet distribution with parameter $\alpha = (\alpha_1, \alpha_2, ..., \alpha_K, \alpha_K, \beta_K)$ is obtained Bourouis et al. (2014); lingmnwah (1976).

4.2.4 Inverted Beta-Liouville Distribution

Inverted Beta-Liouville distribution (IBL) has been proved as an efficient way of modeling positive vectors Bouguila (2014); Bouguila, Fan, and Amayri (2022); Bourouis et al. (2021); Fan and Bouguila (2019). It overcomes the limit of the Inverted Dirichlet in the aspect of positive covariance and presents less parameters as compared with generalized Inverted Dirichlet Bouguila et al. (2022); Bourouis et al. (2021). IBL is in in the family of the Beta-Liouville distribution which is a natural model for analyzing compositional data Fang et al. (2018). Consider $r = \lambda w/(1 - w)$ which w is a Beta distribution with parameters α and β as the generating variate, thus r follows an inverted Beta distribution with parameters β and λ. The generating density function is described as bellow Fang et al. (2018):

$$f(u|\theta) = \frac{1}{B(\alpha, \beta)} \frac{\lambda^{\beta} u^{\alpha-1}}{(\lambda + u)^{\alpha+\beta}} \tag{41}$$

where $u > 0$ and $\theta = (\alpha, \beta, \lambda)$. The Liouville distribution with positive parameters $(\alpha_1, \alpha_2, ..., \alpha_K)$ and generating density of $f(.)$ with parameter δ is defined as below Bouguila (2012); Fang et al. (2018):

$$p(\vec{X}|\alpha_1, ..., \alpha_K, \delta) = f(u|\delta) \frac{\Gamma(\sum_{i=1}^{K} \alpha_i)}{u \sum_{i=1}^{K} \alpha_i - 1} \prod_{i=1}^{K} \frac{x_i^{\alpha_i - 1}}{\Gamma(\alpha_i)} \tag{42}$$

Now by substituting Eq. (41) into Eq. (42) we have the probability density function for $\vec{X}$ as follows:

$$p(x_1, x_2, ..., x_K) = \frac{\Gamma(\alpha_0)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\lambda^{\beta}(\sum_{i=1}^{K} x_i)^{\alpha-\alpha_0}}{(\lambda + \sum_{i=1}^{K} x_i)^{\alpha+\beta}} \prod_{i=1}^{K} \frac{x_i^{\alpha_i-1}}{\Gamma(\alpha_i)} \tag{43}$$

where $\theta = (\alpha_1, \alpha_2, ..., \alpha_K, \alpha, \beta, \lambda)$ is the vector of parameters and

$$\alpha_0 = \sum_{i=1}^{K} \alpha_i,$$

$$x_i > 0, \; i = 1...K$$

This distribution can be converted to inverted Dirichlet distribution by equalizing $\alpha = \alpha_0$ Fang et al. (2018).

The mean and variance of the IBL distribution and the covariance between x_n and x_m are as follows Hu, Fan, Du, and Bouguila (2019); Koochemeshkian, Zamzami, and Bouguila (2020):

$$E(x_i) = \frac{\lambda\alpha\alpha_i}{(\beta-1)(\sum_{i-1}^{K} \alpha_i)} \tag{44}$$

$$Var(x_i) = \frac{\lambda^2\alpha\alpha_i(\alpha+1)^2}{(\beta-1)(\beta-2)\sum_{i=1}^{K}\alpha_i\sum_{i=1}^{K}\alpha_i + 1} - \frac{\lambda^2\alpha^2\alpha_i^4}{(\beta-1)^2(\sum_{i=1}^{K}\alpha_i)^4} \tag{45}$$

$$Cov(x_n, x_m) = \frac{\alpha_1\alpha_m}{\sum_{i=1}^{K}\alpha_i}\left[\frac{\lambda^2\alpha\alpha+1}{(\beta-1)^2(\beta-2)(\sum_{i=1}^{K} x_i)} - \frac{\lambda^2\alpha^2}{(\beta-1)^2(\sum_{i=1}^{K}\alpha_i)}\right] \tag{46}$$

4.3 Recursive Model

In this section, the novel proposed models are described mathematically. As the first step to learn the HMM parameters, we need to estimate the distribution parameters of the HMM Vaičiulytė and Sakalauskas (2020). To achieve this goal we calculate the log likelihood of each probability density function then maximize it. Each probability density function is explained in a different sub-section. However, the whole process is the same. It is noteworthy that ψ used in equations below is the Digamma function which is the logarithmic derivative of the Gamma function. The probability that sequence x is observed while the system is at specific time t considering state q, which can be any in range of $q = 1, , 2, ..., N$, is:

$$\frac{\pi_t^{<q>} \log distribution(x_t|\alpha< q >)}{\sum_{j=1}^{N} \pi_t^{<j>} \log distribution(x_t|\alpha< j >)} \tag{47}$$

Now the summation over t in Eq. (47) indicates the probability of the model being in specific state considering specific output sequence x is observed:

$$\sum_{t=1}^{T} \frac{\pi_t^{<q>} \log distribution(x_t|\alpha< q >)}{\sum_{j=1}^{N} \pi_t^{<j>} \log distribution(x_t|\alpha< j >)} \tag{48}$$

where "$distribution$" refers to each distribution we have used in this chapter.

Here, we define some same variables which are used in the rest of the chapter for all the 3 distributions.

$$\theta_t^q = \pi_t^q \log distribution(x_t|\vec{\theta}^{<q>}) \tag{49}$$

$$\omega_t^{<q,s>} = \omega_{t-1}^{<q,s>} + \frac{1}{t}(\frac{\log(x_t^s)\theta_t^q}{\sum_{j=1}^{N} \theta_t^j} - \omega_{t-1}^{<q,s>}) \tag{50}$$

$$\gamma_t^q = \gamma_{t-1}^q + \frac{1}{t}\left(\frac{\theta_t^q}{\sum_{j=1}^N \theta_t^j} - \gamma_{t-1}^q\right) \tag{51}$$

where

$$1 \le q \le N\ , \quad 1 \le s, i \le K$$

4.3.1 Inverted Dirichlet Distribution

The log likelihood function of inverted Dirichlet (ID) distribution is as follows:

$$\log ID(\vec{X}|\vec{\theta}) = \log\Gamma(\sum_{i=1}^{K+1}\alpha_i) + \sum_{i=1}^{K}(\alpha_i - 1)\log x_i - (\sum_{i=1}^{K+1}\alpha_i)\log(1 + \sum_{i=1}^{K}x_i) - \sum_{i=1}^{K+1}\log\Gamma(\alpha_i) \tag{52}$$

To maximize the above equation, derivative of it with respect to each parameter is shown below:

$$\frac{\partial \log ID(\vec{X}|\vec{\theta})}{\partial \alpha_j} = \begin{cases} \psi(\sum_{i=1}^{K+1}\alpha_i) + \log x_j - \log(1 + \sum_{i=1}^{K} x_j) - \psi(\alpha_j) & \text{if } j = 1, 2, ..., K \\ \psi(\sum_{i=1}^{K+1}\alpha_i) - \log(1 + \sum_{i=1}^{K} x_j) - \psi(\alpha_j) & \text{if j = K+1} \end{cases}$$

Now the batch formula to derive α_i is as below:

$$\alpha_i = \psi^{-1}[\psi(\sum_{i=1}^{K+1}(\alpha_i^q)) + \frac{\frac{1}{T}\sum_{t=1}^{T}\log(x_t^s)\frac{\pi_t^{<q>}\log ID(x_t|\vec{\theta}^{<q>})}{\sum_{j=1}^{N}\pi_t^{<j>}\log ID(x_t|\vec{\theta}^{<j>})}}{\frac{1}{T}\sum_{t=1}^{T}\frac{\pi_t^{<q>}\log ID(x_t|\vec{\theta}^{<q>})}{\sum_{j=1}^{N}\pi_t^{<j>}\log ID(x_t|\vec{\theta}^{<j>})}} - \frac{1}{T}\sum_{t=1}^{T}\frac{\pi_t^{<q>}\log ID(x_t|\vec{\theta}^{<q>})}{\sum_{j=1}^{N}\pi_t^{<j>}\log ID(x_t|\vec{\theta}^{<j>})}], \quad i = 1, 2, ...K \tag{53}$$

$$\alpha_{K+1} = \psi^{-1}[\psi(\sum_{i=1}^{K+1}(\alpha_i^q)) - \frac{1}{T}\sum_{t=1}^{T}\frac{\pi_t^{<q>}\log ID(x_t|\vec{\theta}^{<q>})}{\sum_{j=1}^{N}\pi_t^{<j>}\log ID(x_t|\vec{\theta}^{<j>})}] \tag{54}$$

where

$$1 \leq q \leq N \ , \quad 1 \leq s \leq K$$

By defining middle variables in Eqs. (49), (50) and (51), we derive the recursive formulas to calculate values of the parameters. We assume each observation occurs in specific time step t. Therefor, in our recursive model, the parameters are updated as time goes until they meet the condition of $|\theta_{new} - \theta_{old}| <= \epsilon$. Here the middle variables and the main parameters are explained:

$$\alpha_t^{<q,s>} = \psi^{-1}[\psi(\sum_{i=1}^{K+1}(\alpha_i^q)) + \frac{\omega_t^{<q>}}{\gamma_t^q} - \gamma_t^q] \tag{55}$$

$$\alpha_t^{<q,K+1>} = \psi^{-1}[\psi(\sum_{i=1}^{K+1}(\alpha_i^q)) + \frac{\omega_t^{<q>}}{\gamma_t^q} - \gamma_t^q] \tag{56}$$

where

$$1 \leq q \leq N \ , \quad 1 \leq s, i \leq K$$

4.3.2 Generalized Inverted Dirichlet Distribution

The log likelihood function of generalized inverted Dirichlet (GID) distribution is as follows:

$$\log GID(\vec{X}|\vec{\theta}) = \sum_{i=1}^{K} \log(\alpha_i + \beta_i) + (\alpha_i - 1)\log x_i - \log\Gamma(\alpha_i) - \log\Gamma(\beta_i) - \eta \log(1 + \sum_{m=1}^{i} x_m) \tag{57}$$

To maximize the above equation, derivative of it with respect to each parameter is shown below:

$$\frac{\partial \log GID(\vec{X}|\vec{\theta})}{\partial \alpha_j} = \psi(\alpha_j + \beta_j) + \log(x_j) - \psi(\alpha_j) - \log(1 + \sum_{m=1}^{j} x_m) \tag{58}$$

$$\frac{\partial \log GID(\vec{X}|\vec{\theta})}{\partial \beta_j} = \psi(\alpha_j + \beta_j) - \psi(\beta_j) - \log(1 + \sum_{m=1}^{j} x_m) \tag{59}$$

Now the batch formulas to derive α_i and β_i are as below:

$$\alpha_i = \psi^{-1}[\psi(\alpha_i + \beta_i) + \frac{\frac{1}{T}\sum_{t=1}^{T} \log(x_t^s) \frac{\pi_t^{<q>} \log GID(x_t|\vec{\theta}^{<q>})}{\sum_{j=1}^{N} \pi_t^{<j>} \log GID(x_t|\vec{\theta}^{<j>})}}{\frac{1}{T}\sum_{t=1}^{T} \frac{\pi_t^{<q>} \log GID(x_t|\vec{\theta}^{<q>})}{\sum_{j=1}^{N} \pi_t^{<j>} \log GID(x_t|\vec{\theta}^{<j>})}} - \frac{1}{T}\sum_{t=1}^{T} \frac{\pi_t^{<q>} \log GID(x_t|\vec{\theta}^{<q>})}{\sum_{j=1}^{N} \pi_t^{<j>} \log GID(x_t|\vec{\theta}^{<j>})}] \quad (60)$$

$$\beta_i = \psi^{-1}[\psi(\alpha_i + \beta_i) - \frac{1}{T}\sum_{t=1}^{T} \frac{\pi_t^{<q>} \log GID(x_t|\vec{\theta}^{<q>})}{\sum_{j=1}^{N} \pi_t^{<j>} \log GID(x_t|\vec{\theta}^{<j>})}] \quad (61)$$

where

$$1 \le q \le N\ , \quad 1 \le s, i \le K$$

Based on the logic we explained in previous section, we obtain the following:

$$\alpha_t^{<q,s>} = \psi^{-1}[\psi(\alpha + \beta) + \frac{\omega_t^{<q>}}{\gamma_t^q} - \gamma_t^q] \quad (62)$$

$$\beta_t^{<q,s>} = \psi^{-1}[\psi(\alpha + \beta) - \gamma_t^q]\ , \quad 1 \le q \le N \quad (63)$$

where

$$1 \le q \le N\ , \quad 1 \le s \le K$$

4.3.3 Inverted Beta-Liouville Distribution

The log likelihood function of inverted Beta-Liouville (IBL) distribution is as follows:

$$\log IBL(\vec{X}|\vec{\theta}) = \log \Gamma(\sum_{i=1}^{K} \alpha_i) + \log \Gamma(\alpha + \beta) - \log \Gamma(\alpha) - \log \Gamma(\beta) + \beta \log \lambda + (\alpha - \sum_{i=1}^{K} \alpha_i)$$
$$\log(\sum_{i=1}^{K} x_i) - (\alpha + \beta) \log(\lambda + \sum_{i=1}^{K} x_i) + \sum_{i=1}^{K} (\alpha_i - 1) \log x_i - \sum_{i=1}^{K} \log \Gamma(\alpha_i) \tag{64}$$

To maximize the above equation, derivative of it with respect to each parameter is shown below:

$$\frac{\partial \log IBL(\vec{X}|\vec{\theta})}{\partial \alpha_j} = \psi(\sum_{i=1}^{K} \alpha_i) + \log x_j - \psi(\alpha_j) - \log(\sum_{i=1}^{K} x_i) \tag{65}$$

$$\frac{\partial \log IBL(\vec{X}|\vec{\theta})}{\partial \alpha} = \psi(\alpha + \beta) - \psi(\alpha) + \log(\sum_{i=1}^{K} x_i) - \log(\lambda + \sum_{i=1}^{K} x_i) \tag{66}$$

$$\frac{\partial \log IBL(\vec{X}|\vec{\theta})}{\partial \beta} = \psi(\alpha + \beta) - \psi(\beta) + \log \lambda - \log(\lambda + \sum_{i=1}^{K} x_i) \tag{67}$$

$$\frac{\partial \log IBL(\vec{X}|\vec{\theta})}{\partial \lambda} = \frac{\beta}{\lambda} - \frac{\alpha + \beta}{\lambda + \sum_{i=1}^{K} x_i} \tag{68}$$

Now the batch formulas to derive α_i, α, β, λ are as below:

$$\alpha_i = \psi^{-1}[\psi(\sum_{i=1}^{K} (\alpha_i^q)) + \frac{\frac{1}{T} \sum_{t=1}^{T} \log(x_t^s) \frac{\pi_t^{<q>} \log IBL(x_t|\vec{\theta}^{<q>})}{\sum_{j=1}^{N} \pi_t^{<j>} \log IBL(x_t|\vec{\theta}^{<j>})}}{\frac{1}{T} \sum_{t=1}^{T} \frac{\pi_t^{<q>} \log IBL(x_t|\vec{\theta}^{<q>})}{\sum_{j=1}^{N} \pi_t^{<j>} \log IBL(x_t|\vec{\theta}^{<j>})}}$$
$$- \frac{1}{T} \sum_{t=1}^{T} \frac{\pi_t^{<q>} \log IBL(x_t|\vec{\theta}^{<q>})}{\sum_{j=1}^{N} \pi_t^{<j>} \log IBL(x_t|\vec{\theta}^{<j>})}] \tag{69}$$

$$\alpha = \psi^{-1}[\psi(\alpha + \beta) + \frac{1}{T} \sum_{t=1}^{T} \frac{\pi_t^{<q>} \log IBL(x_t|\vec{\theta}^{<q>})}{\sum_{j=1}^{N} \pi_t^{<j>} \log IBL(x_t|\vec{\theta}^{<j>})}] \tag{70}$$

$$\beta = \psi^{-1}[\psi(\alpha + \beta) + \log \lambda - \frac{1}{T} \sum_{t=1}^{T} \frac{\pi_t^{<q>} \log IBL(x_t|\vec{\theta}^{<q>})}{\sum_{j=1}^{N} \pi_t^{<j>} \log IBL(x_t|\vec{\theta}^{<j>})}] \tag{71}$$

$$\lambda = \frac{\beta}{\alpha}\frac{1}{T}\sum_{t=1}^{T}\frac{\pi_t^{<q>}\log IBL(x_t|\vec{\theta}^{<q>})}{\sum_{j=1}^{N}\pi_t^{<j>}\log IBL(x_t|\vec{\theta}^{<j>})} \tag{72}$$

$$1 \leq q \leq N \;, \quad 1 \leq s, i \leq K$$

Based on the middle variables defined earlier, above explanation of recursive logic and the batch formula of IBL, the main parameters are described as below:

$$\alpha_t^{<q,s>} = \psi^{-1}[\psi(\sum_{i=1}^{K+1}(\alpha_i^q)) + \frac{\omega_t^{<q>}}{\gamma_t^q} - \gamma_t^q] \tag{73}$$

$$\alpha_t^{<q,s>} = \psi^{-1}[\psi(\alpha+\beta) + \gamma_t^q] \tag{74}$$

$$\beta_t^{<q,s>} = \psi^{-1}[\psi(\alpha+\beta) + \log\gamma - \gamma_t^q] \tag{75}$$

$$\lambda_t^{<q,s>} = \frac{\beta}{\alpha}\gamma \tag{76}$$

where $1 \leq q \leq N \;, \quad 1 \leq s \leq K$. We present the recursive proof of the model equations:

$$\begin{aligned}
\omega_t^{<q,s>} &= \omega_{t-1}^{<q,s>} + \frac{1}{t}(\frac{\log(x_t^s)\theta_t^q}{\sum_{j=1}^{N}\theta_i^j} - \omega_{t-1}^{<q,s>}) \\
&= \frac{1}{t-1}\sum_{i=1}^{t-1}\frac{\log(x_t^s)\theta_i^q}{\sum_{j=1}^{N}\theta_i^j} + \frac{1}{t}(\frac{\log(x_t^s)\theta_t^q}{\sum_{j=1}^{N}\theta_t^j} \\
&- \frac{1}{t-1}\sum_{i=1}^{t-1}\frac{\log(x_t^s)\theta_i^q}{\sum_{j=1}^{N}\theta_i^j}) = \frac{1}{t}\sum_{i=1}^{t}\frac{\log(x_i^s)\theta_i^q}{\sum_{j=1}^{N}\theta_i^j}
\end{aligned}$$

$$
\begin{aligned}
\gamma_t^q &= \gamma_{t-1}^q + \frac{1}{t}\left(\frac{\theta_t^q}{\sum_{j=1}^{N}(\theta_t^j)} - \gamma_{t-1}^q\right) \\
&= \frac{1}{t-1}\sum_{i=1}^{t-1}\frac{\theta_i^q}{\sum_{j=1}^{N}\theta_i^j} + \frac{1}{t}\left(\frac{\theta_t^q}{\sum_{j=1}^{N}\theta_t^j} - \frac{1}{t-1}\sum_{i=1}^{t-1}\frac{\theta_i^q}{\sum_{j=1}^{N}\theta_i^j}\right) \\
&= \frac{1}{t}\sum_{i=1}^{t}\frac{\theta_i^q}{\sum_{j=1}^{N}\theta_i^j}
\end{aligned}
$$

We briefly explain the algorithm for all of the models in Algorithm 3 which is based on the expectation-maximization framework.

Please note that in Algorithm 3. the 3 following variables $\vec{\mu}_1$=$(\alpha_1, \alpha_2, ..., \alpha_K, \alpha_{K+1})$, $\vec{\mu}_2$=$(\alpha_1, \alpha_2, ..., \alpha_K, \alpha_K, \beta_1, \beta_2, ..., \beta_K)$, $\vec{\mu}_3$=$(\alpha_1, \alpha_2, ..., \alpha_K, \alpha, \beta, \lambda)$ are representative of ID-HMM, GID-HMM and IBL-HMM parameters respectively.

Algorithm 3 Recursive expectation maximization algorithm for ID-HMM, GID-HMM and IBL-HMM parameter estimation

Result:$\vec{\mu}_1$, $\vec{\mu}_2$ and $\vec{\mu}_3$ with respect to this conditions, $1 \leq t \leq T, 1 \leq s \leq K, 1 \leq q \leq N$
Initialization: $\vec{\mu}_1, \vec{\mu}_2, \vec{\mu}_3$, initial-probability of each state and transition-probability between states
Input: each row of dataset; x_t, $1 \leq t \leq T$

while $\vec{\mu}_{1t} - \vec{\mu}_{1t-1} < \epsilon$, $\vec{\mu}_{2t} - \vec{\mu}_{2t-1} < \epsilon$, $\vec{\mu}_{3t} - \vec{\mu}_{3t-1} < \epsilon$ **do**
    E-step
    Calculate values of $\theta_t^q, \omega_t^{<q,s>}, \gamma_t^q$;
    M-step:
    Update values of $\vec{\mu}_1$, $\vec{\mu}_2$, $\vec{\mu}_3$
end while

The algorithm starts with random initialization of the parameters of both HMM and distributions we consider for data. Then, the algorithm goes through the loop for E-step and M-step until the termination criterion is met which is based on monitoring the difference between the previous value and update one of each parameter after each loop. This difference value is shown as ϵ and set to 0.6. As we have mentioned before, in each loop the values are updated based on the new rows of data feed to the model. We assume that each row of dataset occurs in a specific time, so that each time the new data is obtained, parameters are updated.

4.4 Experimental Results

In this section we present the validation of our proposed models using 2 different real datasets for both binary and multiclass classification of occupants in smart buildings. Data that we have investigated, is collected during time. Thus, based on the logic of its application to estimate occupants in a room, each row of data shows the conditions of the room with the number of people in specific time, so we were able to train the model as each row of the data comes, to update the parameters. In that case we have successfully overcame the problem of batch learning for intensive data. The main motivation to evaluate our model on occupancy datasets is the fact that in smart buildings the goal is to automate the systems related to HVAC (heating, ventilation, air condition) which offers less energy wasting along with better comfort of residents for facilities management strategic decisions. In smart buildings the sensors collecting environmental factors like the amount of CO_2 concentration, temperature, relative humidity, etc are integrated in automation settings. Thus, we have the collected information of sensors in hand allowing us to estimate the number of occupants. This is the idea in HMM, by mapping the collected data as our observation and the number of occupants as the hidden states. In all of the experiments, we define the initial probability of each states (matrix π explained earlier) as $\frac{1}{n}$ assuming n is the total number of states. Transition matrix along with each distributions' parameters are assigned randomly considering their limit according to the distribution definition. The termination criterion of the algorithm to avoid endless recursion is set to $\epsilon = 0.6$. To analyze our datasets collected over time, we assume each record occurred in new time step. Therefore, in our recursive model, the parameters are updated as new record of data feed into the model.

Binary classification dataset (occupancy detection dataset): The first data used to evaluate our models is related to occupancy detection. This dataset is obtained from Machine Learning Repository of University of California Irvine (UCI). Fig. 2.2 describes the distribution of the features. Dataset consists of 5 different features as below which are indicated in a time series:

- Temperature, in Celsius
- Relative Humidity,

- Light, in Lux
- CO2, in ppm
- Humidity Ratio, Derived quantity from temperature and relative humidity, in kgwater-vapor/kg-air

Table 4.1: Estimated parameters for ID-HMM, GID-HMM and IBL-HMM for occupancy detection dataset.

Model	Number of States	Parameters
ID-HMM	2	$\alpha = \begin{bmatrix} 0.2389 & 0.2414 & 0.4244 & 12.47 & 0.2266 & 13.03 \\ 0.2346 & 0.2369 & 1.196 & 12.42 & 0.2227 & 13.34 \end{bmatrix}$
GID-HMM	2	$\alpha = \begin{bmatrix} 8.723 & 8.706 & 8.708 & 8.876 & 8.633 \\ 8.786 & 8.603 & 8.610 & 8.778 & 8.586 \end{bmatrix}$ $\beta = \begin{bmatrix} 8.281 & 8.149 & 8.164 & 9.564 & 7.607 \\ 9.329 & 9.339 & 9.329 & 9.330 & 9.328 \end{bmatrix}$
IBL-HMM	2	$\vec{\alpha} = \begin{bmatrix} 0.9105 & 0.5656 & 0.0679 & 0.0322 & 0.5916 \\ 0.3517 & 0.3470 & 0.1467 & 0.1037 & 0.3807 \end{bmatrix}$ $\alpha = \begin{bmatrix} 0.7223 & 2.478 \end{bmatrix}$ $\beta = \begin{bmatrix} 0.6733 & 0.5181 \end{bmatrix}$ $\lambda = \begin{bmatrix} 2.89 & 2.03 \end{bmatrix}$

This dataset has two different parts for training and testing. We have used the training data to train the model to estimate its parameters, then using the test data to evaluate the models.

Table 4.1 shows the number of hidden states for the occupancy detection dataset along with the values of the model parameters. The values computed through the EM algorithm are explained earlier. Model parameters, based on the distribution considered for emission probability in HMM, follow different dimensions. Table 4.2 indicates the evaluation results of our model based on the 3 distributions we discussed which are computed on the testing dataset.

Multiclass classification dataset (occupancy estimation dataset): This dataset is used for occupancy estimation in smart building as the goal was to estimate the number of occupants from 0 to 4. Thus, the number of hidden states in this sample will be 5. This dataset is obtained from an experiment which testbed was an office in Grenoble Institute of Technology in France Nasfi et al. (2020). This data was obtained from 30 sensors of motions, power consumption, acoustic pressure,

Table 4.2: Accuracy, F-score, precision and recall in percent for Inverted Dirichlet HMM, Generalized Inverted Dirichlet HMM and Inverted Beta-Liouville HMM applied for occupancy detection dataset.

Model	Accuracy	F-score	precision	recall
ID-HMM	86.81	85.65	82.86	85.65
GID-HMM	86.90	87.55	89.22	86.90
IBL-HMM	84.00	79.85	86.58	84.00

Table 4.3: Estimated parameters for ID-HMM, GID-HMM and IBL-HMM for occupancy estimation dataset.

Model	Number of States	Parameters
ID-HMM	5	$\alpha = \begin{bmatrix} 0.5860 & 0.0859 & 0.9971 & 0.5545 & 0.2836 \\ 0.5054 & 0.9837 & 0.2110 & 0.7645 & 0.9573 \\ 0.3307 & 0.1928 & 0.3457 & 0.2385 & 0.7199 \\ 0.4826 & 0.3005 & 0.1633 & 0.8344 & 0.9434 \\ 0.5469 & 0.4305 & 0.3090 & 0.0707 & 0.6270 \end{bmatrix}$
GID-HMM	5	$\alpha = \begin{bmatrix} 7.361 & 6.149 & 4.917 & 4.447 \\ 2.146 & 2.493 & 2.591 & 2.592 \\ 4.498 & 5.912 & 6.943 & 6.747 \\ 1.057 & 1.062 & 1.369 & 1.134 \\ 8.013 & 6.171 & 1.003 & 5.850 \end{bmatrix}$ $\beta = \begin{bmatrix} 5.294 & 1.663 & 1.329 & 1.200 \\ 2.720 & 3.288 & 3.456 & 3.457 \\ 1.365 & 1.426 & 1.543 & 1.490 \\ 2.365 & 4.393 & 3.264 & 2.781 \\ 1.211 & 9.174 & 4.215 & 2.368 \end{bmatrix}$
IBL-HMM	5	$\vec{\alpha} = \begin{bmatrix} 15.25 & 40.16 & 39.95 & 39.33 \\ 0.1666 & 0.4702 & 0.7526 & 0.6555 \\ 0.3187 & 0.4525 & 0.718 & 0.6868 \\ 0.1185 & 0.1837 & 0.3916 & 0.2882 \\ 0.1212 & 0.1379 & 0.3014 & 0.2345 \end{bmatrix}$ $\alpha = \begin{bmatrix} 0.6309 & 15.27 & 13.84 & 18.46 & 8.989 \end{bmatrix}$ $\beta = \begin{bmatrix} 0.1706 & 2.971 & 73.36 & 33.32 & 38.40 \end{bmatrix}$ $\lambda = \begin{bmatrix} 6.469 & 2.001 & 0.3127 & 2.343 & 1.790 \end{bmatrix}$

and door position Amayri et al. (2020). In order to investigate the model we split the dataset into training and testing data. We use the training data to estimate the parameters of our model then apply the trained model on testing data for classification to estimate the number of occupants. Table

Table 4.4: Accuracy, F-score, precision and recall in percent for Inverted Dirichlet HMM, Generalized Inverted Dirichlet HMM and Inverted Beta-Liouville HMM applied for occupancy detection dataset.

Model	Accuracy	F-score	precision	recall
ID-HMM	80.59	79.25	79.36	84.51
GID-HMM	78.24	85.27	82.87	88.02
IBL-HMM	74.53	86.86	87.93	89.94

4.3 represents the values of parameters for each model learned with occupancy estimation dataset, while the evaluation metrics are shown in Table 4.4. The visualisation of the data is presented in Fig. 2.3.

4.5 Conclusion

In this chapter, we introduce a novel approach for occupancy estimation in smart buildings through HMMs. Considering the inverted Dirichlet, generalized inverted Dirichlet and inverted Beta-Liouville distributions as underlying distributions describing the observation data in HMM. The models have been successfully evaluated on real-data of occupancy estimation. The goal was to reach an accurate prediction of number of occupants in a room which plays a key role in smart buildings technology to reduce energy consumption. One of the main motivations to apply these distributions is their flexibility. Therefore, they can be used in extensive range of applications. In addition, the developed models are based on a recursive approach, thus the time complexity of the algorithm reduces to linear time as compared with batch processing which in turn causes a substantial decrease in memory overload and computational resources. Future aspirations could be devoted to improve the initialization methods. Additionally, due to the importance of prepossessing data in machine learning models, to further enhance the model accuracy, feature selection and data quality assessment methods could be investigated.

Chapter 5

Conclusion

In this thesis, we have studied HMMs and developed a number of models as novel methods of supervised learning for the task of occupancy estimation in smart buildings. We tackle the challenges of modeling positive vectors by applying several distributions as the emission probability of the HMMs. Although HMMs have been studied in numerous research works but still the choice of appropriate emission probabilities need to be addressed. In each chapter, we have considered various underlying distributions of HMM. Generally, we have developed statistical machine learning models to formulate recursive parameter estimation of non-Gaussian HMMs to tackle the modeling of data obtained from smart buildings. We focused on estimation of the state of occupancy of smart building's rooms which is considered as the hidden states in our models. The goal of smart buildings relies on estimating the number of occupants within it which corresponds to the nature of HMM while satisfying two objectives; the users comfort and energy saving. The models can be widely used in a variety of real-life applications according to their main characteristics of 1) recursive algorithms and 2) flexibility of chosen emission probabilities. It is noteworthy to mention the advantage of recursive algorithms in our models over iterative ones is the linear complexity which in turn allows us to apply these models to analyse even larger datasets without facing memory overload or computational resources.

In chapter 2, we introduced Generalized Dirichlet HMM and discuss its advantages over Dirichlet HMM. Due to the lack of Dirichlet distribution in modeling various patterns of variability, we have 5 started by applying Generalized Dirichlet as an underlying distribution of HMM which overcomes the restrictive negative covariance of Dirichlet distribution.

In chapter 3, we formulated recursive parameter estimation of Beta-Liouville HMM and compare our newly proposed model's performance to Dirichlet HMM based on the more general covariance of Beta-Liouville distribution in modeling data.

In chapter 4, we proposed 3 novel models of Inverted Dirichlet HMM, Generalized Inverted Dirichlet HMM and Inverted Beta-Liouville HMM for recursive parameter estimation in HMM. These models have been successfully evaluated using real-life datasets of smart buildings.

Future aspirations can be devoted to contributing to this work to further improve the accuracy of each of the models by applying feature selection techniques on datasets. Another interesting venue for future work is investigating the models' capabilities for infinite real-life applications and modifying them in order to be able to adopt infinite hidden states.

References

Ai, B., Fan, Z., & Gao, R. X. (2014). Occupancy estimation for smart buildings by an auto-regressive hidden markov model. In *2014 american control conference* (p. 2234-2239).

Alawadi, S., Mera, D., Fernandez-Delgado, M., Alkhabbas, F., Olsson, C. M., & Davidsson, P. (2020). A comparison of machine learning algorithms for forecasting indoor temperature in smart buildings. *Energy Systems*, 1–17.

Ali, S., & Bouguila, N. (2019a). Hybrid generative-discriminative generalized dirichlet-based hidden markov models with support vector machines. In *2019 ieee international symposium on multimedia (ism)* (p. 231-2311).

Ali, S., & Bouguila, N. (2019b). Variational learning of beta-liouville hidden markov models for infrared action recognition. In *2019 ieee/cvf conference on computer vision and pattern recognition workshops (cvprw)* (p. 898-906).

Ali, S., & Bouguila, N. (2021). Maximum a posteriori approximation of hidden markov models for proportional sequential data modeling with simultaneous feature selection. *IEEE Transactions on Neural Networks and Learning Systems*, 1-12.

Ali, S., & Bouguila, N. (2022). Towards scalable deployment of hidden markov models in occupancy estimation: A novel methodology applied to the study case of occupancy detection. *Energy and Buildings*, *254*, 111594.

Amayri, M., Arora, A., Ploix, S., Bandhyopadyay, S., Ngo, Q.-D., & Badarla, V. R. (2016). Estimating occupancy in heterogeneous sensor environment. *Energy and Buildings*, *129*, 46–58.

Amayri, M., Ploix, S., Bouguil, N., & Wurtz, F. (2020). Database quality assessment for interactive learning: Application to occupancy estimation. *Energy and Buildings*, *209*, 109578.

Baker, J. (1975). The dragon system–an overview. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *23*(1), 24-29.

Baum, L. E., & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state markov chains. *The annals of mathematical statistics*, *37*(6), 1554–1563.

Bdiri, T., & Bouguila, N. (2012). Positive vectors clustering using inverted dirichlet finite mixture models. *Expert Systems with Applications*, *39*(2), 1869–1882.

Bdiri, T., & Bouguila, N. (2013). Bayesian learning of inverted dirichlet mixtures for svm kernels generation. *Neural Computing and Applications*, *23*(5), 1443–1458.

Bdiri, T., Bouguila, N., & Ziou, D. (2016). Variational bayesian inference for infinite generalized inverted dirichlet mixtures with feature selection and its application to clustering. *Appl. Intell.*, *44*(3), 507–525.

Biswas, C., Bhattacharya, U., & Parui, S. K. (2012). Hmm based online handwritten bangla character recognition using dirichlet distributions. In *2012 international conference on frontiers in handwriting recognition* (pp. 600–605).

Bouguila, N. (2011). Count data modeling and classification using finite mixtures of distributions. *IEEE Transactions on Neural Networks*, *22*(2), 186-198.

Bouguila, N. (2012). Infinite liouville mixture models with application to text and texture categorization. *Pattern Recognition Letters*, *33*(2), 103-110.

Bouguila, N. (2014). A model-based discriminative framework for sets of positive vectors classification: Application to object categorization. In *2014 1st international conference on advanced technologies for signal and image processing (atsip)* (p. 277-282).

Bouguila, N., Almakadmeh, K., & Boutemedjet, S. (2012). A finite mixture model for simultaneous high-dimensional clustering, localized feature selection and outlier rejection. *Expert Syst. Appl.*, *39*(7), 6641–6656.

Bouguila, N., & Elguebaly, T. (2012). A fully bayesian model based on reversible jump MCMC and finite beta mixtures for clustering. *Expert Syst. Appl.*, *39*(5), 5946–5959.

Bouguila, N., Fan, W., & Amayri, M. (2022). *Hidden markov models and applications*. Springer.

Bouguila, N., & Ziou, D. (2005). On fitting finite dirichlet mixture using ECM and MML. In P. Wang, M. Singh, C. Apté, & P. Perner (Eds.), *Pattern recognition and data mining, third* *international conference on advances in pattern recognition, ICAPR 2005, bath, uk, august 22-25, 2005, proceedings, part I* (Vol. 3686, pp. 172–182). Springer.

Bouguila, N., & Ziou, D. (2010). A dirichlet process mixture of generalized dirichlet distributions for proportional data modeling. *IEEE Transactions on Neural Networks*, *21*(1), 107-122.

Bouguila, N., Ziou, D., & Vaillancourt, J. (2004). Unsupervised learning of a finite mixture model based on the dirichlet distribution and its application. *IEEE Transactions on Image Processing*, *13*(11), 1533-1543.

Bouhamed, O., Amayri, M., & Bouguila, N. (2022). Weakly supervised occupancy prediction using training data collected via interactive learning. *Sensors*, *22*(9), 3186.

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

Bourouis, S., Mashrgy, M. A., & Bouguila, N. (2014). Bayesian learning of finite generalized inverted dirichlet mixtures: Application to object classification and forgery detection. *Expert Systems with Applications*, *41*(5), 2329–2336.

Boutemedjet, S., Ziou, D., & Bouguila, N. (2007). Unsupervised feature selection for accurate recommendation of high-dimensional image data. In J. C. Platt, D. Koller, Y. Singer, & S. T. Roweis (Eds.), *Advances in neural information processing systems 20, proceedings of the twenty-first annual conference on neural information processing systems, vancouver, british columbia, canada, december 3-6, 2007* (pp. 177–184). Curran Associates, Inc.

Boutemedjet, S., Ziou, D., & Bouguila, N. (2010). Model-based subspace clustering of non-gaussian data. *Neurocomputing*, *73*(10-12), 1730–1739.

Brooks, J., & Barooah, P. (2014). Energy-efficient control of under-actuated hvac zones in buildings. In *2014 american control conference* (p. 424-429).

Candanedo, L. M., & Feldheim, V. (2016). Accurate occupancy detection of an office room from light, temperature, humidity and co2 measurements using statistical learning models. *Energy and Buildings*, *112*, 28–39.

Connor, R. J., & Mosimann, J. E. (1969). Concepts of independence for proportions with a generalization of the dirichlet distribution. *Journal of the American Statistical Association*, *64*(325), 194–206.

Cui, X., & Gong, Y. (2003). Variable parameter gaussian mixture hidden markov modeling for speech recognition. In *2003 ieee international conference on acoustics, speech, and signal processing, 2003. proceedings.(icassp'03).* (Vol. 1, pp. I–I).

Dorj, E., Chen, C., & Pecht, M. (2013). A bayesian hidden markov model-based approach for anomaly detection in electronic systems. In *2013 ieee aerospace conference* (pp. 1–10).

Dymarski, P. (2011). *Hidden markov models: Theory and applications*. BoD–Books on Demand.

D'Oca, S., Hong, T., & Langevin, J. (2018). The human dimensions of energy use in buildings: A review. *Renewable and Sustainable Energy Reviews*, *81*, 731–742.

Ebadat, A., Bottegal, G., Varagnolo, D., Wahlberg, B., & Johansson, K. H. (2013). Estimation of building occupancy levels through environmental signals deconvolution. In *Proceedings of the 5th acm workshop on embedded systems for energy-efficient buildings* (pp. 1–8).

Ebadat, A., Bottegal, G., Varagnolo, D., Wahlberg, B., & Johansson, K. H. (2015). Regularized deconvolution-based approaches for estimating room occupancies. *IEEE Transactions on Automation Science and Engineering*, *12*(4), 1157–1168.

Elkhoukhi, H., Bakhouya, M., Ouadghiri, D. E., & Hanifi, M. (2022, 03). Using stream data processing for real-time occupancy detection in smart buildings. *Sensors*, *22*, 2371.

Epaillard, E., & Bouguila, N. (2014). Hidden markov models based on generalized dirichlet mixtures for proportional data modeling. In *Artificial neural networks in pattern recognition* (pp. 71–82). Springer International Publishing.

Epaillard, E., & Bouguila, N. (2016). Proportional data modeling with hidden markov models based on generalized dirichlet and beta-liouville mixtures applied to anomaly detection in public areas. *Pattern Recognit.*, *55*, 125–136.

Epaillard, E., Bouguila, N., & Ziou, D. (2014). Classifying textures with only 10 visual-words using hidden markov models with dirichlet mixtures. In *Adaptive and intelligent systems* (pp. 20–28). Springer International Publishing.

Erickson, V. L., Carreira-Perpi, M. A., & Cerpa, A. E. (2014). Occupancy modeling and prediction for building energy management. *ACM Transactions on Sensor Networks (TOSN)*, *10*(3), 1–28.

Erickson, V. L., Carreira-Perpiñán, M. Á., & Cerpa, A. E. (2011). Observe: Occupancy-based system for efficient reduction of hvac energy. In *Proceedings of the 10th acm/ieee international conference on information processing in sensor networks* (pp. 258–269).

Fan, W., & Bouguila, N. (2012). Online variational finite dirichlet mixture model and its applications. In *2012 11th international conference on information science, signal processing and their applications (isspa)* (p. 448-453).

Fan, W., & Bouguila, N. (2013a). Infinite dirichlet mixture models learning via expectation propagation. *Adv. Data Anal. Classif.*, *7*(4), 465–489.

Fan, W., & Bouguila, N. (2013b). Learning finite beta-liouville mixture models via variational bayes for proportional data clustering. In F. Rossi (Ed.), *IJCAI 2013, proceedings of the 23rd international joint conference on artificial intelligence, beijing, china, august 3-9, 2013* (pp. 1323–1329). IJCAI/AAAI.

Fan, W., & Bouguila, N. (2013c). Online facial expression recognition based on finite beta-liouville mixture models. In *Tenth conference on computer and robot vision, CRV 2013, regina, saskatchewan, canada, may 28-31, 2013* (pp. 37–44). IEEE Computer Society.

Fan, W., & Bouguila, N. (2013d). Online learning of a dirichlet process mixture of beta-liouville distributions via variational inference. *IEEE Transactions on Neural Networks and Learning Systems*, *24*(11), 1850-1862.

Fan, W., & Bouguila, N. (2013e). Variational learning of finite beta-liouville mixture models using component splitting. In *The 2013 international joint conference on neural networks (ijcnn)* (p. 1-8).

Fan, W., & Bouguila, N. (2014). Online data clustering using variational learning of a hierarchical dirichlet process mixture of dirichlet distributions. In W. Han, M. Lee, A. Muliantara, N. A. Sanjaya, B. Thalheim, & S. Zhou (Eds.), *Database systems for advanced applications - 19th international conference, DASFAA 2014, international workshops: Bdma, damen, SIM - 3 - , uncrowd; bali, indonesia, april 21-24, 2014, revised selected papers* (Vol. 8505, pp. 18–32). Springer.

Fan, W., & Bouguila, N. (2015a). Dynamic textures clustering using a hierarchical pitman-yor process mixture of dirichlet distributions. In *2015 ieee international conference on image* *processing (icip)* (p. 296-300).

Fan, W., & Bouguila, N. (2015b, aug). Expectation propagation learning of a dirichlet process mixture of beta-liouville distributions for proportional data clustering. *Eng. Appl. Artif. Intell.*, *43*, 1-14.

Fan, W., & Bouguila, N. (2019). Nonparametric hierarchical bayesian models for positive data clustering based on inverted dirichlet-based distributions. *IEEE Access*, *7*, 83600-83614.

Fan, W., Wang, R., & Bouguila, N. (2021). Simultaneous positive sequential vectors modeling and unsupervised feature selection via continuous hidden markov models. *Pattern Recognition*, *119*, 108073.

Fang, K. T., Kotz, S., & Wangng, K. (2018). *Symmetric multivariate and related distributions*. Chapman and Hall/CRC.

Gaetani, I., Hoes, P.-J., & Hensen, J. L. (2016). Occupant behavior in building energy simulation: Towards a fit-for-purpose modeling strategy. *Energy and Buildings*, *121*, 188–204.

Gepperth, A., & Hammer, B. (2016). Incremental learning algorithms and applications. In *European symposium on artificial neural networks (esann).*

González-Torres, M., Pérez-Lombard, L., Coronel, J. F., Maestre, I. R., & Yan, D. (2022). A review on buildings energy information: Trends, end-uses, fuels and drivers. *Energy Reports*, *8*, 626–637.

Guo, J., Amayri, M., Bouguila, N., & Fan, W. (2021). A hybrid of interactive learning and predictive modeling for occupancy estimation in smart buildings. *IEEE Trans. Consumer Electron.*, *67*(4), 285–293.

Gupta, R. D., & Richards, D. S. (1987). Multivariate liouville distributions. *Journal of Multivariate Analysis*, *23*(2), 233-256.

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

Jebara, T. (2012). *Machine learning: discriminative and generative* (Vol. 755). Springer Science & Business Media.

Jia, M., & Srinivasan, R. S. (2015). Occupant behavior modeling for smart buildings: A critical review of data acquisition technologies and modeling methodologies. In *2015 winter simulation conference (wsc)* (pp. 3345–3355).

Jiang, L., Liu, D.-Y., & Yang, B. (2004). Smart home research. In *Proceedings of 2004 international conference on machine learning and cybernetics (ieee cat. no.04ex826)* (Vol. 2, p. 659-663 vol.2).

Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition, 2nd edition*. Prentice Hall, Pearson Education International.

Kim, D., Yoon, Y., Lee, J., Mago, P., Lee, K., & Cho, H. (2022, 06). Design and implementation of smart buildings: A review of current research trend. *Energies*, *15*, 4278.

Koochemeshkian, P., Zamzami, N., & Bouguila, N. (2020). Distribution-based regression models for semi-bounded data analysis. In *2020 ieee international conference on systems, man, and cybernetics (smc)* (pp. 4073–4080).

lingmnwah, G. (1976). On the generalised inverted dirichlet distribution. *Demonstratio Mathematica*, *9*(3), 119–130.

Losing, V., Hammer, B., & Wersing, H. (2018). Incremental on-line learning: A review and comparison of state of the art algorithms. *Neurocomputing*, *275*, 1261-1274.

Manouchehri, N., Nguyen, H., & Bouguila, N. (2019). Component splitting-based approach for multivariate beta mixture models learning. In *2019 ieee global conference on signal and information processing (globalsip)* (p. 1-5).

Mashrgy, M. A., Bdiri, T., & Bouguila, N. (2014). Robust simultaneous positive data clustering and unsupervised feature selection using generalized inverted dirichlet mixture models. *Knowl. Based Syst.*, *59*, 182–195.

Miller, D. R., Leek, T., & Schwartz, R. M. (1999). A hidden markov model information retrieval system. In *Proceedings of the 22nd annual international acm sigir conference on research and development in information retrieval* (pp. 214–221).

Mu, X., & Wu, W. (2011). A parallelized network traffic classification based on hidden markov model. In *2011 international conference on cyber-enabled distributed computing and knowledge discovery* (pp. 107–112).

Nasfi, R., Amayri, M., & Bouguila, N. (2020). A novel approach for modeling positive vectors with inverted dirichlet-based hidden markov models. *Knowledge-Based Systems*, *192*, 105335.

Navarro-Lopez, C., Linares-Mustaros, S., & Mulet-Forteza, C. (2022). ”the statistical analysis of compositional data” by john aitchison (1986): A bibliometric overview. *SAGE Open*, *12*(2), 21582440221093366.

Nguyen, H., Rahmanpour, M., Manouchehri, N., Maanicshah, K., Amayri, M., & Bouguila, N. (2019). A statistical approach for unsupervised occupancy detection and estimation in smart buildings. In *2019 ieee international smart cities conference (isc2)* (pp. 414–419).

Oboh, B. S., & Bouguila, N. (2017). Unsupervised learning of finite mixtures using scaled dirichlet distribution and its application to software modules categorization. In *2017 ieee international conference on industrial technology (icit)* (p. 1085-1090).

Oldewurtel, F., Sturzenegger, D., & Morari, M. (2013). Importance of occupancy information for building climate control. *Applied Energy*, *101*, 521-532.

Ploix, S., Amayri, M., & Bouguila, N. (2021). *Towards energy smart homes algorithms, technologies, and applications: Algorithms, technologies, and applications*.

P.Murphy, K. (2012). *Machine learning: a probabilistic perspective*. MIT press.

Pyun, K., Lim, J., Won, C. S., & Gray, R. M. (2007). Image segmentation using hidden markov gauss mixture models. *IEEE Transactions on Image Processing*, *16*(7), 1902–1911.

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257-286.

Rezapoor Nikroo, F., Amayri, M., & Bouguila, N. (n.d.). Hmms recursive parameters estimation for semi-bounded data modeling: Application to occupancy estimation in smart buildings. In *2023 international conference on smart cities and green ict systems (smartgreens).* Manuscript submitted.

Rezapoor Nikroo, F., Amayri, M., & Bouguila, N. (2022a). Recursive parameter estimation of beta-liouville hidden markov models. In *2022 international conference on electrical, computer and energy technologies (icecet)* (p. 1-7).

Rezapoor Nikroo, F., Amayri, M., & Bouguila, N. (2022b). Recursive parameter estimation of generalized dirichlet hidden markov models: Application to occupancy estimation in smart buildings. In *2022 ieee international conference on industry 4.0, artificial intelligence, and communications technology (iaict)* (pp. 90–96).

Sarkar, C., S.N., A. U. N., & Prasad, V. (2016, 02). iltc: Achieving individual comfort in shared spaces..

Schwartz, R., Ima, T., Kubala, F., Nguyen, L., & Makhoul, J. (1997). A maximum likelihood model for topic classification of broadcast news. In *Fifth european conference on speech communication and technology, EUROSPEECH 1997, rhodes, greece, september 22-25, 1997.* ISCA.

Scott, J., Brush, A., Krumm, J., Meyers, B., Hazas, M., Hodges, S., & Villar, N. (2011, 09). Preheat: Controlling home heating using occupancy prediction. In (p. 281-290).

Soltanaghaei, E., & Whitehouse, K. (2016). Walksense: Classifying home occupancy states using walkway sensing. In (p. 167–176). New York, NY, USA: Association for Computing Machinery.

Tiao, G. G., & Cuttman, I. (1965). The inverted dirichlet distribution with applications. *Journal of the American Statistical Association*, *60*(311), 793-805.

Vaičiulytė, J., & Sakalauskas, L. (2020). Recursive parameter estimation algorithm of the dirichlet hidden markov model. *Journal of Statistical Computation and Simulation*, *90*(2), 306-323.

Wong, T.-T. (2010, jul). Parameter estimation for generalized dirichlet distributions from the sample estimates of the first and the second moments of random variables. *Comput. Stat. Data Anal.*, *54*(7), 1756–1765.

Xuan, G., Zhang, W., & Chai, P. (2001). Em algorithms of gaussian mixture model and hidden markov model. In *Proceedings 2001 international conference on image processing (cat. no.01ch37205)* (Vol. 1, p. 145-148 vol.1).

Zamzami, N., Amayri, M., Bouguila, N., & Ploix, S. (2019). Online clustering for estimating occupancy in an office setting. In *2019 ieee 28th international symposium on industrial electronics (isie)* (p. 2195-2200).

Zhang, L., & Chan, K.-P. (2012). Adaptive bayesian HMM for fully unsupervised chinese part-of-speech induction. *ACM Trans. Asian Lang. Inf. Process.*, *11*(3), 9:1–9:22.