

UNIVERSITY OF TECHNOLOGY, SYDNEY

DOCTORAL THESIS

**Bayesian Nonparametric Modeling and Its
Applications**

By
Minqi LI

*A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy*

in the

School of Computing and Communications
Faculty of Engineering and Information Technology

October 2016

Declaration of Authorship

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signed:

Date:

Abstract

Bayesian nonparametric methods (or nonparametric Bayesian methods) take the benefit of unlimited parameters and unbounded dimensions to reduce the constraints on the parameter assumption and avoid over-fitting implicitly. They have proven to be extremely useful due to their flexibility and applicability to a wide range of problems. In this thesis, we study the Bayesian nonparametric theory with Lévy process and completely random measures (CRM). Several Bayesian nonparametric techniques are presented for computer vision and pattern recognition problems. In particular, our research and contributions focus on the following problems.

Firstly, we propose a novel example-based face hallucination method, based on a nonparametric Bayesian model with the assumption that all human faces have similar local pixel structures. We use distance dependent Chinese restaurant process (ddCRP) to cluster the low-resolution (LR) face image patches and give a matrix-normal prior for learning the mapping dictionaries from LR to the corresponding high-resolution (HR) patches. The ddCRP is employed to assist in learning the clusters and mapping dictionaries without setting the number of clusters in advance, such that each dictionary can better reflect the details of the image patches. Experimental results show that our method is efficient and can achieve competitive performance for face hallucination problem.

Secondly, we address sparse nonnegative matrix factorization (NMF) problems by using a graph-regularized Beta process (BP) model. BP is a nonparametric method which lets itself naturally model sparse binary matrices with an infinite number of columns. In order to maintain the positivity of the factorized matrices, an exponential prior is proposed. The graph in our model regularizes the similar training samples having similar sparse coefficients. In this way, the structure of the data can be better represented. We demonstrate the effectiveness of our method on different databases.

Thirdly, we consider face recognition problem by a nonparametric Bayesian model combined with Sparse Coding Recognition (SCR) framework. In order to get an appropriate dictionary with sparse coefficients, we use a graph regularized Beta process prior for the dictionary learning. The graph in our model regularizes training samples in a same class to have similar sparse coefficients and share similar dictionary atoms. In this way, the proposed method is more robust to noise and occlusion of the testing images.

The models in this thesis can also find many other applications like super-resolution, image recognition, text analysis, image compressive sensing and so on.

Acknowledgements

Foremost, I wish to appreciate my principle supervisor Professor Xiangjian He for his patience, motivation, constant encouragement and help during my candidature. His guidance helped me in all the time of research. Without his support, I cannot finish my PhD study.

I would also like to express my gratitude to my Co-supervisor Dr. Richard Yida Xu for leading me into the wonderful world of Machine Learning, especially, Bayesian nonparametric models. Without his professional guidance and help, this thesis would not be completed.

I am grateful to Professor Kin-Man Lam in Hong Kong Polytechnic University for his insightful comments and valuable suggestions on my research. I am also grateful to Associate Professor Jian Zhang for his financial support and valuable advice on my PhD study in the first year. I wish to thank my colleagues and the staff in the Faculty of Engineering and Information Technology (FEIT): Qiang Wu, Wenjing Jia, Min Xu, Ruo Du, Sheng Wang, Chao Zen, Muhammad Abul Hasan, Mohammed Ambu Saidi, Khaled W. Aldebei, Ying Wan, Xuhui Fan, Ava Bargi, Shaukat Abidi, etc., for their invaluable help and support. I have the honour of studying and working with them in the past four years which is valuably stamped in my life.

Last but not least, I would like to thank my family for their unconditional support throughout my life.

To my parents for your love and support.

Contents

Declaration of Authorship	i
Abstract	ii
Acknowledgements	iv
Contents	vi
List of Figures	ix
List of Tables	xi
Abbreviations	xii
Symbols	xiii
1 Introduction	1
1.1 Background	1
1.2 Applications and Motivation	2
1.2.1 Image Segmentation	2
1.2.2 Latent Factor Analysis	3
1.2.3 Other Applications	4
1.2.4 Motivations	5
1.3 Thesis Organization	5
2 Preliminaries	8
2.1 Basic Concepts	8
2.2 Completely Random Measure	12
2.3 Lévy Process	13
2.3.1 Lévy Process and Infinite Divisible Distribution	14
2.3.2 Lévy-Khintchine Representation	16
2.3.3 Lévy-Itô Decomposition of Lévy processes	18
2.3.4 Lévy Measure	19
2.3.5 Examples of Lévy Processes	20
2.4 Markov Chain Monte Carlo (MCMC) methods	38

2.4.1	MH Sampler	40
2.4.2	Gibbs Sampler	42
3	Face Hallucination Based on Spatial-Distance-Dependent Nonparametric Bayesian Learning	44
3.1	Introduction	44
3.2	Bayesian Image Super-resolution Model	47
3.3	Proposed method	48
3.3.1	Training Stage	49
3.3.2	Testing Stage	53
3.4	Experiments	53
3.4.1	Face Hallucination on ORL Database	54
3.4.2	Face Hallucination on Yale Database	57
3.4.3	Performance Analysis	58
3.5	Conclusion	59
4	A Graph-regularized Nonparametric Bayesian Approach to Sparse Nonnegative Matrix Factorization	60
4.1	Introduction	60
4.2	Bayesian Modeling	62
4.2.1	Background of Graph Construction	63
4.2.2	NMF with Beta-Bernoulli Process Prior	64
4.3	Hierarchical Model	65
4.4	Gibbs Sampling Inference	67
4.5	Experimental Results	69
4.5.1	A Simple Image Example	69
4.5.2	Face Dataset Example	71
4.5.3	Digit Dataset Example	72
4.5.4	Performance Analysis	73
4.6	Discussion and Conclusion	73
5	A Graph-regularized Nonparametric Bayesian method for Face Recognition	74
5.1	Introduction	74
5.2	Proposed Method	76
5.2.1	Graph Construction	77
5.2.2	Bayesian Nonparametric Model	77
5.2.3	Gibbs Sampling Inference	80
5.3	Experimental Results	84
5.3.1	ORL Face Database	84
5.3.2	ORL Face Database with Occlusion	85
5.3.3	AR Face Database	86
5.3.4	AR Face Database with Real Disguise	88
5.4	Conclusion	89
6	Conclusions and Future work	91
6.1	Conclusions	91
6.2	Future Work	93
6.2.1	Advanced Inference on Large-Scale Data	93

6.2.2	Graph on Random Measures	94
6.2.3	Hierarchical Factor Analysis for Deep Learning	95
A	Several Distributions Used in the Thesis	97
A.1	Gamma Distribution	97
A.2	Dirichlet Distribution	97
A.3	Beta Distribution	98
A.4	Bernoulli Distribution	98
A.5	Poisson Distribution	98
A.6	Normal-inverse-Wishart Distribution	98
A.7	Matrix Normal Distribution	99
B	Bayesian Nonparametric Non-negative Matrix Factorization Model and Inference	100
B.1	Hierarchical Model	100
B.2	Gibbs Sampling Inference	101
	Bibliography	104

List of Figures

2.1	Stick breaking for constructing random measure	11
2.2	An example of Lévy processes: standard Brownian motion	21
2.3	A sample of compound Poisson process.	23
2.4	A sample of Gamma process	25
2.5	A sample of Dirichlet process	28
2.6	Example of Chinese restaurant process	30
2.7	Example of distance dependent Chinese restaurant process	32
2.8	Example of Indian buffet process	37
3.1	The framework of our proposed method.	49
3.2	Patches learned in an example cluster.	55
3.3	Patches positions constrained in the example cluster.	55
3.4	Face hallucination results on the ORL database with different methods: (a) the input LR faces, (b) Bicubic interpolation, (c) Chang’s method, (d) Sun’s method, (e) Yang’s method, (f) our proposed method, and (g) the original HR faces.	56
3.5	Face hallucination results on the Yale database with different methods: (a) the input LR faces, (b) Bicubic interpolation, (c) Chang’s method, (d) Sun’s method, (e) Yang’s method, (f) our proposed method, and (g) the original HR faces.	58
4.1	Generated Markov chains from Algorithm 2. Top: log posterior probability of \mathbf{X} ; Middle: the residual computed by Equation 4.17; Bottom: average sparseness of factor \mathbf{B}	70
4.2	A facial image NMF by Algorithm 2. (a) original image, and (b) reconstructed image by the NMF factors from the proposed method.	70
4.3	NMF on the ORL face image dataset. (a) ORL face image dataset; Factors computed by (b) Lee’s method, (c) Lin’s method, (d) Schmidt’s method, (e) Hoyer’s method, and (f) proposed method.	71
4.4	NMF on MNIST image dataset. (a) MNIST image dataset; Factors computed by (b) Lin’s method, (c) Schmidt’s method (d), Hoyer’s method, and (e) proposed method.	72
5.1	Graph constructed by the training samples. (a) The true graph constructed by the training labels, and (b) the graph constructed by the heat kernel. Five samples in each class are used here for graph construction in this example.	85
5.2	Recognition results on ORL database	86
5.3	ORL face images with random occlusions	87
5.4	Recognition results on ORL database with occlusion	87
5.5	Recognition results on AR database	88
5.6	Example images in AR database	88

5.7 Recognition results on AR database with real disguise.	89
--	----

List of Tables

2.1	Relationship between Lévy process and infinitely divisible distribution.	16
2.2	Stick-breaking construction of Dirichlet process.	28
2.3	Conclusion of different Lévy processes.	37
3.1	PSNR performance of different algorithms on ORL database.	54
3.2	SSIM performance of different algorithms on ORL database.	55
3.3	PSNR performance of different algorithms on Yale database.	57
3.4	SSIM performance of different algorithms on Yale database.	57
4.1	Results based on the ORL face database.	71
4.2	Results based on the MNIST digit image database.	72

Abbreviations

GaP	Gamma Process
DP	Dirichlet Process
CRP	Chinese Restaurant Process
ddCRP	distance dependent Chinese Restaurant Process
BeP	Bernoulli Process
BP	Beta Process
IBP	Indian Buffet Process
SCR	Sparse Coding Recognition

Symbols

\mathbb{R}	the set of reals
\mathbb{N}	the set of natural numbers
$\mathbb{E}[\cdot]$	expectation of a random variable
δ_θ	measure concentrated at θ
\mathbf{c}_i^-	the assignment set excluding c_i
$DP(\alpha, H)$	Dirichlet process with concentration parameter α and base measure H
$GaP(c, H)$	Gamma process with concentration parameter c and base measure H
$BP(\alpha, H)$	Beta process with positive scalar α and base measure H
$\mathcal{IW}(\cdot)$	normal-inverse-Wishart distribution
$\mathcal{MN}(\cdot)$	matrix-normal distribution
$\mathcal{MT}(\cdot)$	matrix-t distribution
$\ \cdot\ _{l_p}$	l_p norm
$G_{i,n}$	graph weight between data i and n