

Machine Learning for Multimedia Content Analysis

MULTIMEDIA SYSTEMS AND APPLICATIONS SERIES

Consulting Editor

Borko Furht

Florida Atlantic University

Recently Published Titles:

DISTRIBUTED MULTIMEDIA RETRIEVAL STRATEGIES FOR LARGE SCALE NETWORKED SYSTEMS by Bharadwaj Veeravalli and Gerassimos Barlas; ISBN: 978-0-387-28873-4

MULTIMEDIA ENCRYPTION AND WATERMARKING by Borko Furht, Edin Muharemagic, Daniel Socek; ISBN: 0-387-24425-5

SIGNAL PROCESSING FOR TELECOMMUNICATIONS AND MULTIMEDIA edited by T.A. Wysocki, B. Honary, B.J. Wysocki; ISBN 0-387-22847-0

ADVANCED WIRED AND WIRELESS NETWORKS by T.A. Wysocki, A. Dadej, B.J. Wysocki; ISBN 0-387-22781-4

CONTENT-BASED VIDEO RETRIEVAL: A Database Perspective by Milan Petkovic and Willem Jonker; ISBN: 1-4020-7617-7

MASTERING E-BUSINESS INFRASTRUCTURE edited by Veljko Milutinović, Frédéric Patricelli; ISBN: 1-4020-7413-1

SHAPE ANALYSIS AND RETRIEVAL OF MULTIMEDIA OBJECTS by Maytham H. Safar and Cyrus Shahabi; ISBN: 1-4020-7252-X

MULTIMEDIA MINING: A Highway to Intelligent Multimedia Documents edited by Chabane Djeraba; ISBN: 1-4020-7247-3

CONTENT-BASED IMAGE AND VIDEO RETRIEVAL by Oge Marques and Borko Furht; ISBN: 1-4020-7004-7

ELECTRONIC BUSINESS AND EDUCATION: Recent Advances in Internet Infrastructures edited by Wendy Chin, Frédéric Patricelli, Veljko Milutinović; ISBN: 0-7923-7508-4

INFRASTRUCTURE FOR ELECTRONIC BUSINESS ON THE INTERNET by Veljko Milutinović; ISBN: 0-7923-7384-7

DELIVERING MPEG-4 BASED AUDIO-VISUAL SERVICES by Hari Kalva; ISBN: 0-7923-7255-7

CODING AND MODULATION FOR DIGITAL TELEVISION by Gordon Drury, Garegin Markarian, Keith Pickavance; ISBN: 0-7923-7969-1

CELLULAR AUTOMATA TRANSFORMS: Theory and Applications in Multimedia Compression, Encryption, and Modeling by Olu Lafe; ISBN: 0-7923-7857-1

COMPUTED SYNCHRONIZATION FOR MULTIMEDIA APPLICATIONS by Charles B. Owen and Fillia Makedon; ISBN: 0-7923-8565-9

Visit the series on our website: www.springer.com

Machine Learning for Multimedia Content Analysis

by

Yihong Gong

and

Wei Xu

*NEC Laboratories America, Inc.
USA*



Springer

Yihong Gong
NEC
Laboratories of America, Inc.
10080 N. Wolfe Road SW3-350
Cupertino CA 95014
ygong@sv.nec-labs.com

Wei Xu
NEC
Laboratories of America, Inc.
10080 N. Wolfe Road SW3-350
Cupertino CA 95014
xw@sv.nec-labs.com

Library of Congress Control Number: 2007927060

Machine Learning for Multimedia Content Analysis by Yihong Gong and Wei Xu

ISBN 978-0-387-69938-7

e-ISBN 978-0-387-69942-4

Printed on acid-free paper.

© 2007 Springer Science+Business Media, LLC

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

9 8 7 6 5 4 3 2 1

springer.com

Preface

Nowadays, huge amount of multimedia data are being constantly generated in various forms from various places around the world. With ever increasing complexity and variability of multimedia data, traditional rule-based approaches where humans have to discover the domain knowledge and encode it into a set of programming rules are too costly and incompetent for analyzing the contents, and gaining the intelligence of this glut of multimedia data.

The challenges in data complexity and variability have led to revolutions in machine learning techniques. In the past decade, we have seen many new developments in machine learning theories and algorithms, such as boosting, regressions, Support Vector Machines, graphical models, etc. These developments have achieved great successes in a variety of applications in terms of the improvement of data classification accuracies, and the modeling of complex, structured data sets. Such notable successes in a wide range of areas have aroused people's enthusiasms in machine learning, and have led to a spate of new machine learning text books. Noteworthily, among the ever growing list of machine learning books, many of them attempt to encompass most parts of the entire spectrum of machine learning techniques, resulting in a shallow, incomplete coverage of many important topics, whereas many others choose to dig deeply into a specific branch of machine learning in all aspects, resulting in excessive theoretical analysis and mathematical rigor at the expense of losing the overall picture and the usability of the books. Furthermore, despite a large number of machine learning books, there is yet a text book dedicated to the audience of the multimedia community to address unique problems and interesting applications of machine learning techniques in this area.

The objectives we set for this book are two-fold: (1) bring together those important machine learning techniques that are particularly powerful and effective for modeling multimedia data; and (2) showcase their applications to common tasks of multimedia content analysis. Multimedia data, such as digital images, audio streams, motion video programs, etc, exhibit much richer structures than simple, isolated data items. For example, a digital image is composed of a number of pixels that collectively convey certain visual content to viewers. A TV video program consists of both audio and image streams that complementally unfold the underlying story and information. To recognize the visual content of a digital image, or to understand the underlying story of a video program, we may need to label sets of pixels or groups of image and audio frames jointly because the label of each element is strongly correlated with the labels of the neighboring elements. In machine learning field, there are certain techniques that are able to explicitly exploit the spatial, temporal structures, and to model the correlations among different elements of the target problems. In this book, we strive to provide a systematic coverage on this class of machine learning techniques in an intuitive fashion, and demonstrate their applications through various case studies.

There are different ways to categorize machine learning techniques. Chapter 1 presents an overview of machine learning methods through four different categorizations: (1) Unsupervised versus supervised; (2) Generative versus discriminative; (3) Models for i.i.d. data versus models for structured data; and (4) Model-based versus modelless. Each of the above four categorizations represents a specific branch of machine learning methodologies that stem from different assumptions/philosophies and aim at different problems. These categorizations are not mutually exclusive, and many machine learning techniques can be labeled with multiple categories simultaneously. In describing these categorizations, we strive to incorporate some of the latest developments in machine learning philosophies and paradigms.

The main body of this book is composed of three parts: I. unsupervised learning, II. Generative models, and III. Discriminative models. In Part I, we present two important branches of unsupervised learning techniques: dimension reduction and data clustering, which are generic enabling tools for many multimedia content analysis tasks. Dimension reduction techniques are commonly used for exploratory data analysis, visualization, pattern recognition, etc. Such techniques are particularly useful for multimedia content analysis because multimedia data are usually represented by feature vectors of extremely

high dimensions. The curse of dimensionality usually results in deteriorated performances for content analysis and classification tasks. Dimension reduction techniques are able to transform the high dimensional raw feature space into a new space with much lower dimensions where noise and irrelevant information are diminished. In Chapter 2, we describe three representative techniques: Singular Value Decomposition (SVD), Independent Component Analysis (ICA), and Dimension Reduction by Locally Linear Embedding (LLE). We also apply the three techniques to a subset of handwritten digits, and reveal their characteristics by comparing the subspaces generated by these techniques.

Data clustering can be considered as unsupervised data classification that is able to partition a given data set into a predefined number of clusters based on the intrinsic distribution of the data set. There exist a variety of data clustering techniques in the literature. In Chapter 3, instead of providing a comprehensive coverage on all kinds of data clustering methods, we focus on two state-of-the-art methodologies in this field: spectral clustering, and clustering based on non-negative matrix factorization (NMF). Spectral clustering evolves from the spectral graph partitioning theory that aims to find the best cuts of the graph that optimize certain predefined objective functions. The solution is usually obtained by computing the eigenvectors of a graph affinity matrix defined on the given problem, which possess many interesting and preferable algebraic properties. On the other hand, NMF-based data clustering strives to generate semantically meaningful data partitions by exploring the desirable properties of the non-negative matrix factorization. Theoretically speaking, because the non-negative matrix factorization does not require the derived factor-space to be orthogonal, it is more likely to generate the set of factor vectors that capture the main distributions of the given data set.

In the first half of Chapter 3, we provide a systematic coverage on four representative spectral clustering techniques from the aspects of problem formulation, objective functions, and solution computations. We also reveal the characteristics of these spectral clustering techniques through analytical examinations of their objective functions. In the second half of Chapter 3, we describe two NMF-based data clustering techniques, which stem from our original works in recent years. At the end of this chapter, we provide a case study where the spectral and NMF clustering techniques are applied to the text clustering task, and their performance comparisons are conducted through experimental evaluations.

In Part II and III, we focus on various graphical models that are aimed to explicitly model the spatial, temporal structures of the given data set, and therefore are particularly effective for modeling multimedia data. Graphical models can be further categorized as either generative or discriminative. In Part II, we provide a comprehensive coverage on generative graphical models. We start by introducing basic concepts, frameworks, and terminologies of graphical models in Chapter 4, followed by in-depth coverages of the most basic graphical models: Markov Chains and Markov Random Fields in Chapter 5 and 6, respectively. In these two chapters, we also describe two important applications of Markov Chains and Markov Random Fields, namely Markov Chain Monte Carlo Simulation (MCMC) and Gibbs Sampling. MCMC and Gibbs Sampling are the two powerful data sampling techniques that enable us to conduct inferences for complex problems for which one can not obtain closed-form descriptions of their probability distributions. In Chapter 7, we present the Hidden Markov Model (HMM), one of the most commonly used graphical models in speech and video content analysis, with detailed descriptions of the forward-backward and the Viterbi algorithms for training and finding solutions of the HMM. In Chapter 8, we introduce more general graphical models and the popular algorithms such as sum-product, max-product, etc. that can effectively carry out inference and training on graphical models.

In recent years, there have been research works that strive to overcome the drawbacks of generative graphical models by extending the models into discriminative ones. In Part III, we begin with the introduction of the Conditional Random Field (CRF) in Chapter 9, a pioneer work in this field. In the last chapter of this book, we present an innovative work, Max-Margin Markov Networks (M^3 -nets), which strives to combine the advantages of both the graphical models and the Support Vector Machines (SVMs). SVMs are known for their abilities to use high-dimensional feature spaces, and for their strong theoretical generalization guarantees, while graphical models have the advantages of effectively exploiting problem structures and modeling correlations among inter-dependent variables. By implanting the kernels, and introducing a margin-based objective function, which are the core ingredients of SVMs, M^3 -nets successfully inherit the advantages of the two frameworks. In Chapter 10, we first describe the concepts and algorithms of SVMs and Kernel methods, and then provide an in-depth coverage of the M^3 -nets. At the end of the chapter, we also provide our insights into why discriminative

graphical models generally outperform generative models, and M^3 -nets are generally better than discriminative models.

This book is devoted to students and researchers who want to apply machine learning techniques to multimedia content analysis. We assume that the reader has basic knowledge in statistics, linear algebra, and calculus. We do not attempt to write a comprehensive catalog covering the entire spectrum of machine learning techniques, but rather to focus on the learning methods that are powerful and effective for modeling multimedia data. We strive to write this book in an intuitive fashion, emphasizing concepts and algorithms rather than mathematical completeness. We also provide comments and discussions on characteristics of various methods described in this book to help the reader to get insights and essences of the methods. To further increase the usability of this book, we include case studies in many chapters to demonstrate example applications of respective techniques to real multimedia problems, and to illustrate factors to be considered in real implementations.

California, U.S.A.
May 2007

Yihong Gong
Wei Xu

Contents

1	Introduction	1
1.1	Basic Statistical Learning Problems	2
1.2	Categorizations of Machine Learning Techniques	4
1.2.1	Unsupervised vs. Supervised	4
1.2.2	Generative Models vs. Discriminative Models	4
1.2.3	Models for Simple Data vs. Models for Complex Data	6
1.2.4	Model Identification vs. Model Prediction	7
1.3	Multimedia Content Analysis	8

Part I Unsupervised Learning

2	Dimension Reduction	15
2.1	Objectives	15
2.2	Singular Value Decomposition	16
2.3	Independent Component Analysis	20
2.3.1	Preprocessing	23
2.3.2	Why Gaussian is Forbidden	24
2.4	Dimension Reduction by Locally Linear Embedding	26
2.5	Case Study	30
	Problems	34
3	Data Clustering Techniques	37
3.1	Introduction	37
3.2	Spectral Clustering	39
3.2.1	Problem Formulation and Criterion Functions	39
3.2.2	Solution Computation	42

XII Contents

3.2.3	Example	46
3.2.4	Discussions	50
3.3	Data Clustering by Non-Negative Matrix Factorization	51
3.3.1	Single Linear NMF Model	52
3.3.2	Bilinear NMF Model	55
3.4	Spectral vs. NMF	59
3.5	Case Study: Document Clustering Using Spectral and NMF	
	Clustering Techniques	61
3.5.1	Document Clustering Basics	62
3.5.2	Document Corpora	64
3.5.3	Evaluation Metrics	64
3.5.4	Performance Evaluations and Comparisons	65
	Problems	68

Part II Generative Graphical Models

4	Introduction of Graphical Models	73
4.1	Directed Graphical Model	74
4.2	Undirected Graphical Model	77
4.3	Generative vs. Discriminative	79
4.4	Content of Part II	80
5	Markov Chains and Monte Carlo Simulation	81
5.1	Discrete-Time Markov Chain	81
5.2	Canonical Representation	84
5.3	Definitions and Terminologies	88
5.4	Stationary Distribution	91
5.5	Long Run Behavior and Convergence Rate	94
5.6	Markov Chain Monte Carlo Simulation	100
5.6.1	Objectives and Applications	100
5.6.2	Rejection Sampling	101
5.6.3	Markov Chain Monte Carlo	104
5.6.4	Rejection Sampling vs. MCMC	110
	Problems	112
6	Markov Random Fields and Gibbs Sampling	115
6.1	Markov Random Fields	115
6.2	Gibbs Distributions	117

6.3	Gibbs – Markov Equivalence	120
6.4	Gibbs Sampling	123
6.5	Simulated Annealing	126
6.6	Case Study: Video Foreground Object Segmentation by MRF	133
6.6.1	Objective	134
6.6.2	Related Works	134
6.6.3	Method Outline	135
6.6.4	Overview of Sparse Motion Layer Computation	136
6.6.5	Dense Motion Layer Computation Using MRF	138
6.6.6	Bayesian Inference	140
6.6.7	Solution Computation by Gibbs Sampling	141
6.6.8	Experimental Results	143
	Problems	146
7	Hidden Markov Models	149
7.1	Markov Chains vs. Hidden Markov Models	149
7.2	Three Basic Problems for HMMs	153
7.3	Solution to Likelihood Computation	154
7.4	Solution to Finding Likeliest State Sequence	158
7.5	Solution to HMM Training	160
7.6	Expectation-Maximization Algorithm and its Variances	162
7.6.1	Expectation-Maximization Algorithm	162
7.6.2	Baum-Welch Algorithm	164
7.7	Case Study: Baseball Highlight Detection Using HMMs	167
7.7.1	Objective	167
7.7.2	Overview	167
7.7.3	Camera Shot Classification	169
7.7.4	Feature Extraction	172
7.7.5	Highlight Detection	173
7.7.6	Experimental Evaluation	174
	Problems	175
8	Inference and Learning for General Graphical Models	179
8.1	Introduction	179
8.2	Sum-product algorithm	182
8.3	Max-product algorithm	188

8.4	Approximate inference	189
8.5	Learning	191
	Problems	196

Part III Discriminative Graphical Models

9	Maximum Entropy Model and Conditional Random Field	201
9.1	Overview of Maximum Entropy Model	202
9.2	Maximum Entropy Framework	204
9.2.1	Feature Function	204
9.2.2	Maximum Entropy Model Construction	205
9.2.3	Parameter Computation.....	208
9.3	Comparison to Generative Models	210
9.4	Relation to Conditional Random Field	213
9.5	Feature Selection	215
9.6	Case Study: Baseball Highlight Detection Using Maximum Entropy Model	217
9.6.1	System Overview	218
9.6.2	Highlight Detection Based on Maximum Entropy Model	220
9.6.3	Multimedia Feature Extraction.....	222
9.6.4	Multimedia Feature Vector Construction	226
9.6.5	Experiments	227
	Problems	232
10	Max-Margin Classifications	235
10.1	Support Vector Machines (SVMs)	236
10.1.1	Loss Function and Risk	237
10.1.2	Structural Risk Minimization	237
10.1.3	Support Vector Machines.....	239
10.1.4	Theoretical Justification.....	243
10.1.5	SVM Dual.....	244
10.1.6	Kernel Trick	245
10.1.7	SVM Training	248
10.1.8	Further Discussions	255
10.2	Maximum Margin Markov Networks	257
10.2.1	Primal and Dual Problems	257

10.2.2 Factorizing Dual Problem	259
10.2.3 General Graphs and Learning Algorithm	262
10.2.4 Max-Margin Networks vs. Other Graphical Models	262
Problems	264
A Appendix	267
References	269
Index	275