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 loosing 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 modeless. 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-production, 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. *Yihong Gong*
May 2007 *Wei Xu*

Springer

http://www.springer.com/978-0-387-69938-7

Machine Learning for Multimedia Content Analysis
Gong, Y.; Xu, W.
2007, XVI, 277 p. 20 illus., Hardcover
ISBN: 978-0-387-69938-7