BOOK REVIEW

Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval

Reviewed by Dhiraj Joshi, Kodak Research Labs, Eastman Kodak Company, Rochester, New York
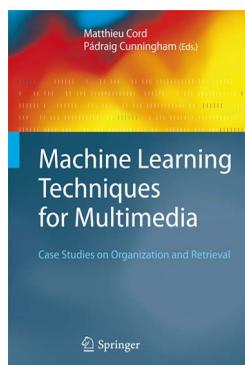
Multimedia pervades our lives today in totality. For the research community, understanding and interpreting multimedia has always been a challenge. Over the years, multimedia research has broadened to account for the evolving scope of the term "multimedia." Today, research under this umbrella encompasses image, video, and audio domains. Machine learning is an integral component of multimedia research. In fact, it would not be inappropriate to say that every multimedia-analysis problem is a machine-learning problem at heart. The contributions in the field of multimedia research are tough to summarize in a single book. This book is divided into two sections with four and seven chapters, respectively. The first section builds a background in machine learning, and the second section presents some featured multimedia applications. With contributions from a diverse set of authors, I believe that it is a good resource for advanced undergraduates and early graduate students. At the same time, the book is a pleasant read for experienced researchers in the field.

Chapter 1 introduces Bayesian methods and decision theory to the readers. Theory governing Bayesian learning is briefly presented, followed by a discussion of simulation methods like Markov Chain Monte Carlo, which is popular in multimedia analysis. A classical dichotomy in machine learning is supervised versus unsupervised learning, discussed in Chapters 2 and 3, respectively. Chapter 2 develops a theoretical background for maximum margin classification and discusses linear support vector machines (SVMs), kernel classification, and K-nearest neighbor (K-NN) classification. This is followed by a discussion of distance metrics especially within the multimedia domain. The authors conclude with a discussion of ensemble methods such as bagging and boosting. The chapter is well written, coherent, and complete, and the content is well placed with respect to current multimedia research. SVMs, kernel methods, and boosting are all popular and successful learning methodologies in multimedia. K-NN classification presents a fresh nonparametric data-driven approach to learning, as opposed to sophisticated methods such as SVMs, and is a growing trend in image understanding nowadays.

Unsupervised learning is the focus of Chapter 3. The chapter begins with a discussion of classical k-means, fuzzy, and hierarchical clustering methods, followed by newer ideas such as kernel k-means and spectral clustering. Self-organizing maps (SOMs) are also discussed with respect to clustering. The chapter concludes with a discussion of cluster validation methods. Chapter 4 deals with dimension reduction methods, specifically principal component analysis (PCA) and linear discriminant analysis (LDA). Feature selection, an important step in supervised and unsupervised learning alike, is also discussed here.

The second section of the book on multimedia applications begins with Chapter 5. The chapter experimentally compares and contrasts several approaches to feature representation, image similarity, and classification-based retrieval. However, the focus of the chapter is active-learning-based semisupervised image retrieval, an important topic in its own right. The discussion is detailed and a pleasant read. However, in addition to the content presented, it would have been interesting to see experiments on large-scale Web image data. Chapter 6 discusses object detection in a video-analysis task with the goal to minimize manual effort during learning. The chapter presents an interesting synergy between a discriminative and a generative classifier. Chapter 7 deals with face analysis, a very current topic in multimedia. A good background on face detection, facial-feature detection, and emotion recognition is provided, followed by discussions of machine learning with respect to the above topics. The chapter concludes with experiments on the aforementioned to make the reader appreciate the challenges involved.

Mental search, a relatively new search paradigm in retrieval, is described in Chapter 8. In contrast to the classical query by example image-search paradigm, the assumption here is that a user has a vague idea of what he or she is looking for at the beginning of the search. Successive interaction results in targeting the search toward the mental image of the user. In this chapter, mental search is presented with respect to image retrieval and object-recognition tasks. Chapter 9 delves into integration of textual and visual information for labeling of images and video. Tagging or labeling of images with appropriate words is a hot research problem in the multimedia community. The chapter presents a model for building correspondences between image regions and labels and discusses the problem of associating names with faces in news video.

The web abounds in semistructured multimedia documents where both content and structure can assist in understanding. Chapter 10 focuses on such documents and presents a generative modeling approach with respect to two applications: filtering pornographic content and classification of Wikipedia documents. The last chapter, Chapter 11, completes the multimedia discussion with focus on audio. It is dedicated to machine learning with respect to music analysis and retrieval. The chapter describes many popular audio features before discussing the problem of classification of music into genres. The chapter also discusses visualization of music with self-organizing maps and presents an interesting application for music retrieval on mobile devices.

Overall the book is well organized and an interesting read. Topics have been contributed by a diverse set of authors, and the chapters include detailed bibliographies. A

chapter to discuss other emerging trends in multimedia would have added even more value to the book. Nevertheless, I believe that the book will be appreciated by amateurs and experts alike.

Dhiraj Joshi graduated with a MSc in mathematics and scientific computing from the Indian Institute of Technology, Kanpur. He completed his PhD in computer science from Penn State University in 2007 and now works as a research scientist in the Intelligent Systems Group at Kodak Research Laboratories. He has been a research intern at IBM T. J. Watson Research Labs, USA, and the IDIAP Research Institute, Switzerland. His research interests include contextual-inference-based image understanding, large-scale image retrieval, content analysis in multimedia, statistical learning, and social network modeling.