Web-Scale Multimedia Processing and Applications

By EDWARD CHANG
Guest Editor

SHIH-FU CHANG, *Fellow IEEE*
Guest Editor

ALEXANDER G. HAUPTMANN
Guest Editor

THOMAS S. HUANG, *Life Fellow IEEE*
Guest Editor

MALCOLM SLANEY, *Fellow IEEE*
Guest Editor

This special issue provides a timely and comprehensive review of recent progress, applications, and challenges spurred by the introduction of web-scale data resources, involving text, image, video, speech, user recommendation, and their combinations. The first group of papers covers fundamental crosscutting techniques critical for developing web-scale systems, including large-scale linear classification, optimal hashing, robust graph-based semisupervised learning, and classifier adaptation. The second group reviews techniques for discovering knowledge, extracting information, and combining knowledge with multimedia content networks. The third group surveys representative work demonstrating the unique impact of web-scale data in practical applications, such as image annotation, media recommendation, enterprise social networks, and social media networks. This special issue is intended for broad audience interested in gaining comprehensive understanding of technical issues and opportunities in this important field.

This special issue provides a timely and comprehensive review of recent progress, applications, and challenges spurred by web-scale data resources.

I. INTRODUCTION

The scale of the Web brings not only a wealth of opportunity, but also a whole set of new and challenging problems. This special issue addresses many of the problems, solutions, and opportunities inherent in processing and providing web-scale multimedia solutions.

Web-scale multimedia data require an entirely different style of processing. This happens not only because the scale of the problem changes the nature of the solutions, but also because different kinds of algorithms and architectures are needed to return answers about billions of data instances in a fraction of a second. The good news is that in many domains—speech recognition is most notable—people have observed that the best way to improve their algorithm's performance is to add more data. Starting with hidden Markov models (HMMs) and support vector machines (SVMs), people have

Digital Object Identifier: 10.1109/JPROC.2012.2204110

applied ever greater amounts of data to their problems and been rewarded with new levels of performance. In other cases, entirely different approaches often yield superior results. State-of-the-art recommendation systems ignore the contents of a movie, a reasonable signal when considering whether a user who likes one movie will like another, because of the wealth of rating data available from users.

In addition, the demands placed on multimedia systems are even greater than before. Multimedia search has always been plagued by the semantic gap, the big difference between a multimedia object and its meaning. The problem only gets harder with user-generated content on the Web since multimedia content often has low production values, and its meaning is quite dependent on the user (a picture of a car is labeled Tokyo because the photographer took the photo on his once-per-decade trip to the Tokyo Auto Show). Additionally, in some applications, web-scale multimedia systems literally must find the needle in the haystack in tens of milliseconds or the user will look elsewhere for his/her information and entertainment.

The availability of web-scale media data and the associated rich metadata information provides unprecedented opportunities for developing new formulations, approaches, and solutions. Underlying the emerging efforts, a common theme is the shift from the traditional emphasis on building sophisticated models to the new one on efficient processing of the massive amount of data and innovative ways of using rich information associated with the data. New technical challenges also emerge, such as large-scale data matching and indexing, large-scale online machine learning, massive/parallel media computing algorithms and architectures, robustness to information uncertainty of web data, integration of hybrid types of data/users and relations, and processing of data sets in the streaming mode.

The impact of web-scale data is also found in fields such as social networking. Web-scale media is a novel "tool" to sample, observe, and analyze the activities of real-world users and study how they interact with the physical world (e.g., where they visit and what they do) and with the community (e.g., how they interact with each other) in specific domains like enterprise and social media forums. Social networking studies may further lead to enhanced understanding of the context about how media content is used in the network, helping solve the semantic gap mentioned above.

This special issue provides a timely and comprehensive review of recent progresses, applications, and challenges spurred by the web-scale data resources. It is intended for broad audience interested in gaining comprehensive understanding of technical issues and opportunities in this important field.

II. TOPICS AND PAPERS

The papers in this special issue cover three broad areas: fundamental technologies, knowledge discovery and information extraction, and applications and social networks.

The first group of papers is devoted to fundamental techniques critical for developing web-scale systems, including large-scale linear classification, optimal hashing, robust graph-based semisupervised learning, and classifier adaptation. Each paper focuses on the unique impact brought about by the exploitation of web-scale data, with clear explanations of how the technical approaches and performance are fundamentally changed due to the introduction of new web-scale resources. The second group reviews techniques for discovering knowledge, extracting information, and combining knowledge with multimedia content networks. The last group of papers surveys representative work that demonstrates the unique impact of web-scale data in practical applications, such as image annotation, media recommendation, enterprise social networks, and social media networks. We present a brief summary of each paper in the following.

A. Fundamental Technologies

In the paper entitled "Recent advances of large-scale linear classification," by Yuan *et al.*, a comprehensive survey is presented on the recent development of efficient optimization methods to construct linear classifiers suitable for large-scale applications. The authors show that for some data in high-dimensional spaces, the accuracy of linear classifiers is close to that of nonlinear classifiers such as kernel methods.

In the paper entitled "Optimal parameters for locality-sensitive hashing," Slaney *et al.* describe an algorithm to optimize the parameters needed when applying the locality-sensitive hashing (LSH) technique that is the basis of many promising nearest neighbor retrieval techniques for web-scale search. Given a desired performance level and a simple computational cost model, the algorithm returns the LSH parameters that allow an LSH index to meet the performance goal and have the minimum computational cost.

To exploit the information from available labeled data and the massive amount of unlabeled samples, Liu *et al.* review graph-based semisupervised learning methods and new techniques for handling contaminated noisy labels and gigantic data sizes for web applications in the paper "Robust and scalable graph-based semisupervised learning." Popular data sets and software tools are also described to help new researchers. Yang *et al.*, in the paper entitled "A framework for classifier adaptation for large-scale multimedia data," tackle the heterogeneity challenge in large-scale multimedia data using cross-domain model adaptation for better performance and reduced human cost. Specifically, they investigate the problem of adapting supervised classifiers trained from one or more source domains to a new classifier for a target domain that has only limited labeled examples.

B. Knowledge Discovery and Information Extraction

In the paper entitled "Uncertainty reduction for knowledge discovery and information extraction on the World Wide Web," Ji *et al.* give an overview of knowledge discovery (KD) and information extraction (IE) techniques for web-based applications. They focus on the new techniques that can be used to handle the additional uncertainty challenge in KD and IE incurred in the web setting.

In their paper "Statistical entity extraction from the Web," Nie *et al.* present recent work on statistical extraction of structured entities, named entities, entity facts, and relations from the Web. They also describe an interactive knowledge mining framework, iKnoweb, for entity information integration, together with two working examples of web applications: Microsoft Academic Search (aka Libra) and EntityCube.

To combine the information in the multimedia content network, cross-media links, and the associated ontological network, Qi *et al.* present a paper entitled "Web-scale multimedia information networks." A unified structured representation called multimedia information networks (MINets) is proposed together with advanced methods for network construction and utilizing MINets for information propagation across media and categories.

C. Applications and Social Networks

The first paper on applications, entitled "Duplicate-search-based image annotation using web-scale data," by Wang *et al.*, focuses on annotation of images on the Web based on label propagation over similar images and social information found on the Web. A system, called Arista, is used to demonstrate scalable implementations over 2 billion web images and interesting applications such as celebrity face databases and text-to-image translation.

In the paper entitled "Web-scale media recommendation systems," Dror *et al.* present a design process useful for building web-scale media recommendation systems that direct consumers to media products such as music, movie, and books. They introduce a music rating data set that is the largest of its kind and formulate a detailed recommendation model, specifically designed to account for the data set properties.

The last two papers of the issue extend the scope to rapidly growing social media networks. Sundaram *et al.*, in the paper entitled "Multimedia semantics: Interactions between content and community," review the state of the art and emerging issues in research related to pattern analysis and monitoring of Web-based social communities. Specifically, they study how media-rich social networks provide additional insight into familiar multimedia research problems, including tagging and video ranking.

Lin *et al.* describe a comprehensive study in the paper entitled "Social network analysis in enterprise" with a focus on the challenges and solutions of mining and analyzing existing social networks in enterprises. Their study is based on a social network analysis tool (SmallBlue) that was designed to collect data in a global organization of more than 400 000 employees in more than 100 countries. ■

ABOUT THE GUEST EDITORS

Edward Chang received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1999.

He has been Head of Google Research in China since March 2006. He joined the Department of Electrical & Computer Engineering, University of California at Santa Barbara, Santa Barbara, in 1999. He received his tenure in 2003, and was promoted to Full Professor of Electrical Engineering in 2006. His recent research activities are in the areas of distributed data mining and their applications to rich-media data management and social network collaborative filtering. His research group (which consists of members from Google, University of California, Massachusetts Institute of Technology, Tsinghua University, Peking University, and Zheda) recently parallelized support vector machines (NIPS 07), probabilistic latent semantic analysis (KDD 08), association mining (ACM RS 08), spectral clustering (ECML 08), and latent Dirichlet allocation (WWW 09) to run on thousands of machines for mining large-scale data sets. His team at Google developed and launched Google Confucius (a Q&A system) at China, Russia, South East Asia, Africa, and Arabic countries.

Prof. Chang has served on ACM (SIGMOD, KDD, MM, CIKM), VLDB, IEEE, WWW, and SIAM conference program committees, and co-chaired several conferences including MMM, ACM MM, ICDE, and WWW. He is a recipient of the IBM Faculty Partnership Award and the National Science Foundation (NSF) Career Award.

Shih-Fu Chang (Fellow, IEEE) received the B.S. degree in electrical engineering from the National Taiwan University, Taipei, Taiwan, in 1985 and the M.S. and Ph.D. degrees in electrical engineering and computer sciences from the University of California at Berkeley, Berkeley, in 1991 and 1993, respectively.

He is the Senior Vice Dean of School of Engineering and Applied Science, the Richard Dicker Professor in the Department of Electrical Engineering and the Department of Computer Science, and the Director of Digital Video and Multimedia Lab at Columbia University, New York, NY. He has made significant contributions to multimedia search, visual communication, media forensics, and international standards. He has worked in different advising/consulting capacities for industry research labs and international institutions.

Prof. Chang has been recognized with ACM SIGMM Technical Achievement Award, IEEE Kiyo Tomiyasu Award, Navy ONR Young Investigator Award, IBM Faculty Award, Recognition of Service Awards from ACM and the IEEE Signal Processing Society and NSF CAREER Award. He and his students have received many Best Paper Awards, including the Most Cited Paper of the Decade Award from the *Journal of Visual Communication and Image Representation*. He is a Fellow of the American Association for the Advancement of Science. He served as Editor-in-Chief for the IEEE Signal Processing Magazine (2006–2008), and Chair of Columbia's Electrical Engineering Department (2007–2010).

Alex Hauptmann studied computer science at Technische Universität Berlin, Berlin, Germany, from 1982 to 1984. He received the B.A. and M.A. degrees in psychology from Johns Hopkins University, Baltimore, MD and the Ph.D. degree in computer science from Carnegie Mellon University (CMU), Pittsburgh, PA, in 1991.

He is a Senior Systems Scientist in the CMU Computer Science Department and also a faculty member with CMU's Language Technologies Institute. His research interests have led him to pursue and combine several different areas: man–machine communication, natural language processing, speech understanding and synthesis, and machine learning. He worked on speech and machine translation at CMU from 1984 to 1994, when he joined the Informedia project where he developed the News-on-Demand application. Since then he has conducted research projects on video analysis and retrieval on broadcast news as well as observational video with success documented by outstanding performance in the annual NIST TRECVID video retrieval evaluations.

Thomas S. Huang (Life Fellow, IEEE) received the Sc.D. degree in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 1963.

He was on the faculty of MIT and Purdue University, West Lafayette, IN. He joined University of Illinois at Urbana-Champaign, Urbana, in 1980 and is currently William L. Everitt Distinguished Professor of Electrical and Computer Engineering, Research Professor of Coordinated Science Laboratory, Professor of the Center for Advanced Study, and Co-Chair of the Human Computer Intelligent Interaction major research theme of the Beckman Institute for Advanced Science and Technology. He has published 21 books and more than 600 technical papers in network theory, digital holography, image and video compression, multimodal human–computer interfaces, and multimedia databases.

Prof. Huang is a member of the National Academy of Engineering and has received numerous honors and awards, including the IEEE Jack S. Kilby Signal Processing Medal (with Ar. Netravali) and the King-Sun Fu Prize of the International Association of Pattern Recognition.

Malcolm Slaney (Fellow, IEEE) received the B.S, M.S., and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN.

He was a Principal Scientist at Yahoo! Research, Sunnyvale, CA and now holds the same position in the Conversational Systems Laboratory at Microsoft Corporation, Mountain View, CA. He is a (consulting) Professor at Stanford University's Center for Computer Research in Music and Acoustics (CCRMA), Stanford, CA, where he has led the Hearing Seminar for the last 20 years. Before Yahoo!, he has worked at Bell Laboratory, Schlumberger Palo Alto Research, Apple Computer, Interval Research, and IBM's Almaden Research Center. For the last several years he has helped lead the auditory group at the Telluride Neuromorphic Cognition Workshop. Lately, he has been working on multimedia analysis and music- and image-retrieval algorithms in databases with billions of items. He is a coauthor, with A. C. Kak, of the IEEE book *Principles of Computerized Tomographic Imaging*. This book was republished by SIAM in their "Classics in Applied Mathematics" series. He is coeditor, with S. Greenberg, of the book *Computational Models of Auditory Function.*

Prof. Slaney has served as an Associate Editor of the IEEE Transactions on Audio, Speech, and Signal Processing, IEEE Multimedia Magazine, and the Proceedings of the IEEE.