Social Media Analytics and Intelligence
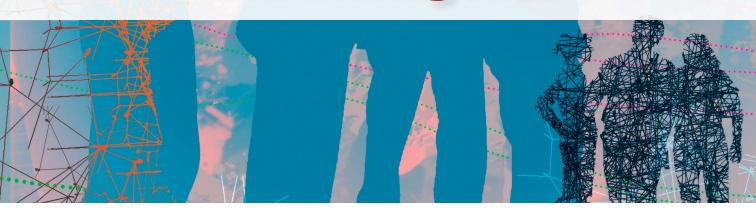
Daniel Zeng, *Chinese Academy of Sciences and University of Arizona*

Hsinchun Chen and Robert Lusch, *University of Arizona*

Shu-Hsing Li, *National Taiwan University*

I n a broad sense, social media refers to a conversational, distributed mode of content generation, dissemination, and communication among communities. Different from the broadcast-based traditional and industrial media, social media has torn down the boundaries between authorship and readership, while the information consumption and dissemination process is becoming intrinsically intertwined with the process of generating and sharing information.

Application and Impact

The Internet and mobile technologies have been the primary force behind the rise of social media, providing technological platforms for information dissemination, content generation, and interactive communications. In fact, major components of social media such as user-generated content or consumer-generated media have been viewed as the defining characteristics of Web 2.0. From a tool perspective, an array of Web-based applications define the way social media functions. Examples include weblogs, microblogs, online forums, wikis, podcasts, life streams, social bookmarks, Web communities, social networking, and avatar-based virtual reality.

From an application perspective, many websites dedicated to social media are among

the most popular—Wikipedia (collective knowledge generation), MySpace and Facebook (social networking), YouTube (social networking and multimedia content sharing), Digg and Delicious (social browsing, news ranking, and bookmarking), Second Life (virtual reality), and Twitter (social networking and microblogging), to name just a few.

Because social media is already a critical part of the information ecosystem and as social media platforms and applications gain widespread adoption with unprecedented reach to users, consumers, voters, businesses, governments, and nonprofit organizations alike, interest in social media from all walks of life has been skyrocketing from both application and research perspectives. For-profit businesses are tapping into social media as both a rich source of information and a business-execution platform for product design and innovation, consumer and stakeholder relations management, and marketing. For them, social media is an essential component of the next-generation business intelligence platform. For politicians, political parties, and governments, social media represents the ideal vehicle and information base to gauge public opinion on policies and political positions as well as to build community support for candidates running for public offices. Public-health officials could potentially use social media as valuable, early clues about disease outbreaks and to provide feedback on public-health policies and response measures. For homeland security and intelligence analysis communities, social media presents immense opportunities to study terrorist group behavior, including their recruiting and public relation schemes and the grounding social and cultural contexts. Even think

tanks and social science and business researchers are conceptually using social media as an unbiased sensor network and a laboratory for natural experimentation, providing valuable indicators and helping test hypotheses about social production and interactions as well as their economic, political, and societal implications.

For many individuals, social media has become a unique information source to deal with information- and cognitive-overload problems, find answers to specific questions, and discover more valuable opportunities for social and economic exchange. In addition, it has become a platform for them to network and contribute to all kinds of dynamic dialogues by sharing their expertise and opinions. It is safe to claim that social media has already penetrated a spectrum of applications with remarkable impact. Given the continued interest and the ever-growing information and meta-information generated through social media, it is expected to continue enabling new exciting applications and revolutionizing many existing ones.

Social Media Analytics and Intelligence Research

Research on social media has greatly intensified in the past few years given the significant interest from the application's perspective and the associated unique technical and social science challenges and opportunities. This research agenda is multidisciplinary in nature and has drawn attention from research communities in all major disciplines. From an information technology standpoint, social media research has primarily focused on social media analytics and, more recently, social media intelligence.

Social media analytics is concerned with developing and evaluating

informatics tools and frameworks to collect, monitor, analyze, summarize, and visualize social media data, usually driven by specific requirements from a target application. Social media analytics research serves several purposes:

- facilitating conversations and interaction between online communities and
- extracting useful patterns and intelligence to serve entities that include, but are not limited to, active contributors in ongoing dialogues.

From a technical perspective, social media analytics research faces several unique challenges. First, social media contains an enriched set of data or metadata, which have not been treated systematically in data- and text-mining literature. Examples include tags (annotations or labels using free-form keywords); user-expressed subjective opinions, insights, evaluation, and perspectives; ratings; user profiles; and both explicit and implicit social networks. Second, social media applications are a prominent example of human-centered computing with their own unique emphasis on social interactions among users. Hence, issues such as context-dependent user profiling and needs elicitation as well as various kinds of human-computer interaction considerations must be reexamined. Third, although social media promises a new approach to tackling the noise and information-overload problem with Web-based information processing, issues such as semantic inconsistency, conflicting evidence, lack of structure, inaccuracies, and difficulty in integrating different kinds of signals abound in social media. Fourth, social media data are dynamic streams, with their volume

rapidly increasing. The dynamic nature of such data and their sheer size pose significant challenges to computing in general and to semantic computing in particular.

Making use of technology, solution frameworks, and toolsets from social media analytics, *social media intelligence* aims to derive actionable information from social media in context-rich application settings, develop corresponding decision-making or decision-aiding frameworks, and provide architectural designs and solution frameworks for existing and new applications that can benefit from the "wisdom of crowds" through the Web. Compared to social media analytics, social media intelligence research is still in an early stage of development, despite increasing attention from businesses and other communities that could benefit from such studies. From a research perspective, there have been discussions about various conceptual dimensions of social media intelligence, related technical challenges, and reference disciplines that could potentially bring about useful tools to help meet these challenges. However, systematic research and concrete, well-evaluated results are still lacking.

Nevertheless, social media intelligence presents great potential with important practical relevance as a rich, new area of inquiry, potentially drawing on disciplines from within AI as well as other fields. Several key challenges currently face social media intelligence research. First, social media intelligence research calls for highly integrated multidisciplinary research. Although this need has been reiterated often in this growing field, the level of integration in the existing research tends to be low. In many cases, the informatics research methodology and research questions are taking a dominant role, whereas research focusing on methods and issues from other equally relevant disciplines, such as social psychology, media theory, political science, and social sciences, among others, has been scattered; there have been few highly integrated research programs that consider both informatics and domain sciences.

Second, social media intelligence research requires well-articulated and clearly defined performance measures because much of it must be conducted in application settings with an aim to support decisions. However, in a broad spectrum of applications in which social medial intelligence could be relevant, it is challenging to quantify these measures. This measurement problem makes it especially hard to judge social media intelligence's return on investment (ROI), and it leads to modeling difficulties.

Third, from a pure modeling and decision-making perspective, social media intelligence represents a unique class of problems with the need for efficient data-driven, dynamic decision making; uncertainty and subjective risk analysis; and modeling and optimization over large dynamic networks. As social media intelligence research matures and finds real-time applications, researchers will likely need to develop new analytical and computational frameworks and methods.

In this Issue

This special issue samples the state of the art in social media analytics and intelligence research that has direct relevance to AI from either a methodological or domain perspective.

In "Searching Social Media Streams on the Web," Jonghun Park, Yongwook Shin, Kwanho Kim, and Beom-Suk Chung observe that much of social media content takes the form of streams—dynamically arriving and updated series of texts with structured and unstructured components. They develop a search approach called FeedMil with topic-driven search and retrieval capabilities from a variety of live social media streams. Their technical research mainly focuses on identifying potentially relevant streams based on user queries and ranking these streams based on relevance scores and query-independent quality measures, such as the streams' popularity, authority, and activity.

Davide Barbieri, Daniele Braga, Stefano Ceri, Emanuele Della Valle, Yi Huang, Volker Tresp, Achim Rettinger, and Hendrik Wermser also focus on a stream-based view of social media in "Deductive and Inductive Stream Reasoning for Semantic Social Media Analytics," but with a distinctively different goal. They have developed reasoning methods to support management of changing knowledge drawing from stream-based content and structured background knowledge bases. Their approach is based on the notion of RDF streams as time-stamped RDF triples, which provide a representational framework to enable integration of inductive and deductive reasoning techniques.

Collaborative tagging and social search are among the most successful social media and "wisdom of crowds" applications in which users annotate webpages or other resources using tags. Such tags are shared among users and can be explored to enable a range of information retrieval and recommendation capabilities. One major obstacle hindering the adoption of tagging-based systems or services is the presence

THE AUTHORS

Daniel Zeng is a research professor in the Institute of Automation at the Chinese Academy of Sciences and a professor and Honeywell Fellow in the Department of Management Information Systems at the University of Arizona. His research interests include social computing, intelligence and security informatics, infectious disease informatics, recommender systems, and software agents. Zeng has a PhD in industrial administration from Carnegie Mellon University. He is the vice president for technical activities for the IEEE Intelligent Transportation Systems Society and chair of the INFORMS College on Artificial Intelligence. Contact him at zengdaniel@gmail.com.

Hsinchun Chen is the McClelland Professor of Management Information Systems at the University of Arizona and the founding director of the Artificial Intelligence Lab and the Hoffman E-Commerce Lab. His research interests include Web computing; search engines; digital libraries; intelligence analysis; biomedical informatics; data, text, and web mining; and knowledge management. Chen has a PhD in information systems from New York University. He is editor in chief of the *ACM Transactions on Management Information Systems* (ACM TMIS) and Springer's *Security Informatics* journal. He is a fellow of IEEE and the American Association for the Advancement of Science (AAAS). Contact him at hchen@eller.arizona.edu.

Robert F. Lusch is a professor of marketing and the Jim and Pamela Muzzy Chair in Entrepreneurship in the Eller College of Management at the University of Arizona. His expertise is in service quality and strategy, organizational performance, and supply-chain management with a focus on supply-chain collaborations. Lusch has a PhD in business administration from the University of Wisconsin. Lusch has served as editor of the *Journal of Marketing*, chairperson of the American Marketing Association, and trustee of the American Marketing Association Foundation. Contact him at rlusch@email.arizona.edu.

Shu-Hsing Li is a professor of accounting and dean of the College of Management at the National Taiwan University. His research interests include business intelligence and enterprise risk management, supply-chain management, field study of management accounting in high-technology companies, text mining and capital market research, and deposit-receipt pricing and market behaviors. Li has a PhD in accounting from New York University. Contact him at shuhsingli@yahoo.com.tw.

of noises and ambiguities in user-provided tags. In their article "Modeling Social Annotations via Latent Reason Identification," Xiance Si, Zhiyuan Liu, and Maosong Sun propose the Tag Allocation Model to tackle this challenge. They show their generative model delivers good performance in tag recommendations and tag-hierarchy discovery.

Three-dimensional virtual worlds such as Second Life have drawn a lot of attention from research communities and the industry. User-generated social media content through avatars abounds in virtual worlds. Yet, systematic data collection from this relatively new social media channel and behavioral analysis of avatars have been underexplored. Yulei Zhang, Ximing Yu, Yan Dang, and Hsinchun Chen's article "An Integrated Framework for Avatar Data Collection from the Virtual World" proposes an integrated approach that combines bot- and spider-based techniques to collect avatar behavioral and profile data. They also report empirical findings examining differences in avatar behavior based on avatar gender and age.

"Using Social Media to Predict Future Events with Agent-Based Markets" by Efthimios Bothos, Dimitris Apostolou, and Gregoris Mentzas proposes a prediction market approach using computational agents as opposed to human participants. Such agents embody human-user sentiments—as well as their knowledge, beliefs, and assessments, all extracted from social media—and participate in prediction markets to predict future events. This type of automated approach could potentially overcome implementation difficulties associated with standard "wisdom of crowds" approaches such as prediction markets with human participants.

Social media analytics and intelligence research continues to flourish. As special issue guest editors, we are pleased about the large number of manuscripts (35) submitted in response to our special issue call for papers, which is indicative of the research community's strong interest in this area. The five articles accepted for this issue present diverse, yet synergetic viewpoints and technical approaches. Social media analytics research is showing clear signs of outgrowing its reference disciplines, such as text mining and Web computing, with its unique problem characteristics and solution frameworks. Social media intelligence research is also growing with an expanding set of applications and interesting mixing with other academic disciplines.

We express our sincere gratitude to all contributing authors and reviewers for their time and effort. We hope that the perspectives, models, technological development, research findings, and case studies presented in this special issue will help encourage exciting new, synergetic research in this important field of great practical impact.

cn *Selected CS articles and columns are also available for free at http://ComputingNow.computer.org.*