Introduction to the Special Section: Mining Social Media

José Carlos Cortizo, Francisco Manuel Carrero and José María Gomez, Guest Editors

Social media are technological tools that allow users to share and discuss infor- mation. Most social media are Internet-based applications that manage textual information. These include blogs (Blogger, Wordpress), microblogging (Twitter, Pownce), wikis (Wikipedia), forums, and social networks (Facebook, MySpace, LinkedIn). But there also exist other social media applications through which users share more than text, such as photographs (Flickr, Picasa), videos (YouTube, Vimeo), livecasts (Ustream), and audio and music tracks (last.fm, ccMixter, FreeSound). More recent social media include virtual worlds (Second Life), online gaming (World of Warcraft, WarHammer Online), game sharing (Miniclip.com), and mobile social media, such as Nomad Social Networks, through which users share their current position in the real world.

Social media have been able to shift the way information is generated and consumed. At first, information was generated by one person and "consumed" by many people, but now the information is generated by many people and consumed by many people, changing the needs in information access and management. It is also noticeable that social media applications manage huge quantities of users and data: Facebook manages more than 500 million users; it is estimated that 1 million blog posts are generated each day; the microblog- ging service Twitter generates several million messages each day; YouTube has managed more than 15 billion videos views; and so forth. All this makes clear that social media are an excellent application field for data miners.

In a Forrester Research report, "The Future of the Social Web," [1] the development of the social Web has been divided into five eras. The first era witnessed the explosion of social relationships among users. Then, in the social functionality era, some Web sites started to add social functions in order to help users interact with their peers. <<PleaSe brIefly MentIon the other two eraS: The 2 other eras are mentioned later in the text: era of social context and era of social e-commerce>>We are now in the era of social colonization, in which technologies such as Facebook Connect or Google Friend Connect have standardized social functionalities among Web sites, and a vast majority of Web sites now include several social functionalities. Soon these federated identities will empower people to enter the era of social context with personalized and social content, and with the development of tools for personalization of social content, auguring the era of social e-commerce. Social media mining is a crucial aspect of social e-commerce. The five papers in this Special Issue address techniques for mining and exploiting social media data through strong theoretical grounding and empirical testing.

In the first paper, "Automatic Moderation of Online Discussion Sites," Jean-Yves Delort, Bavani Arunasalam, and Cecile Paris discuss a partially labeled data-based learning method for the automatic moderation of online forums. Online discussion sites are plagued with various types of unwanted content, such as spam and obscene or malicious content. Although there ex- ist widely adopted prevention techniques, detection of inappropriate content remains mostly a manual task. The authors found that their partially labeled learning method is able to automatically moderate inappropriate content in online discussion sites effectively. The method is easily adaptable to other domains as well.

In the second paper, "Fusing Recommendations for Social Bookmark- ing Web Sites," Toine Bogers and Antal van den Bosch present an empirical comparison of a

number of item recommendation approaches for social bookmarking systems. These methods explode several sources of informa- tion, such as user tags and item metadata, and are fused/merged/combined following well-known score aggregation methods. The paper addresses a problem that is gaining momentum with a renewed and growing interest in recommender systems, which in recent years have changed from novelties used by a few e-commerce sites to serious business tools that are changing the world of e-commerce.

The third paper, "Expert Stock Picker: The Wisdom of (the Experts in the) Crowds," by Shawndra Hill and Noah Ready-Campbell, describes a study of stock picks generated by users of a financial voting system. The authors show that the aggregated stock picks of the crowd outperform the S&P 500 index, and those of a group of experts identified by their past performance give even better results. The main contribution of the paper is a genetic algo- rithm approach that can be used to identify the appropriate vote weights for users based on the set of users' prior individual voting success rankings, as well as the number of most recent vote contributions to the user-generated content site.

The fourth paper, "Learning to Identify Internet Sexual Predation," writ- ten by India McGhee, Jennifer Bayzik, April Kontostathis, Lynne Edwards, Alexandra McBride, and Emma Jakubowski, focuses on an important problem fairly poorly studied by other applied researchers: an approach to detecting online predators in chat communication, through a mix of techniques. The identification of sexual predators is a difficult task, because it needs to process much textual content, the interactions between users, and also other attributes, such as the actual context.

In the fifth paper, "Internet Auction Fraud Detection Using Social Network Analysis and Classification Tree Approaches," Chaochang Chiu, Yungchang Ku, Ting Lie, and Yuchi Chen present a methodology to detect online auction fraud. In particular, they use social network analysis and data mining techniques to classify auction transactions into three types: normal, suspicious, and fraudulent. Internet fraud detection is a very interesting problem, specific to e-commerce. In fact, detecting and avoiding fraud in e-commerce would contain one of the most important barriers to online shopping.

In summary, the papers in this issue illustrate the leading research in the field of mining social media. This Special Issue also illustrates how social media information can be mined to improve e-commerce and related applications. This is especially so as social e-commerce begins to rise, using mined social information to improve the entire process of e-commerce.

JOSE CARLOS CORTIZO (josecarlos.cortizo@brainsins.com) is associate professor at Universidad Europea de Madrid, and CTO of BrainSINS, a company that develops several social information management systems like social recommender systems and social search engines. His research has focused on Social Intelligent Systems, including Personalization, Machine Learning and Recommender Systems in the Social Web. He has co-organized several workshops related to mining social media information like the 1st International Workshop on Mining Social Media, the 2nd International Workshop on Search and Mining User-generated Content celebrated as a CIKM 2010 workshop or the 1st International Workshop on Adaptation, Personalization and REcommendation in the Social-semantic Web co-located with the 7th Extended Semantic Web Conference. He has served as guest editor on several international journals and as PC member in several international conferences and workshops.

FRANCISCO MANUEL CARRERO (francisco.carrero@brainsins.com) is associate professor at Universidad Europea de Madrid, and CEO of BrainSINS, a company that develops several social information management systems like social recommender

systems and social search engines. His main research interests are in the field of Information Retrieval, and have evolved to User Modeling and Social Recommender Systems. He has co-organized several workshops related to mining social media information like the 1st International Workshop on Mining Social Media, the 2nd International Workshop on Search and Mining User-generated Content celebrated as a CIKM 2010 workshop or the 1st International Workshop on Adaptation, Personalization and REcommendation in the Social-semantic Web co-located with the 7th Extended Semantic Web Conference. He has served as guest editor on several international journals and as PC member in several international conferences and workshops.

JOSE MARIA GOMEZ (jgomez@optenet.com) has been professor at Complutense University of Madrid and Universidad Europea de Madrid, for more than ten years, being head director of the department of Computer Science Engineering. Actually he is research director at Optenet, a company that develops information filters and security systems. His main research lines are focused on Natural Language Processing and Machine Learning, with applications in news and biomedicine, and Adversarial Information Retrieval with applications at spam and pornographic content detection. He has been PC member at CEAS 2007, Spam Symposium 2007 and several other conferences like JASIST or ECIR. He has also been reviewing R&D projects for the European Commission.

[1] Owyang, J. K.; Bernoff, J.; Pflaum, C.; Bowen, E. The Future of the Social Web, http://www.forrester.com/rb/Research/future_of_social_web/q/id/46970/t/2 Accessed on October 2010.