Guest Editorial

Erik Cambria
School of Computer Science and Engineering, Nanyang Technological University, SINGAPORE

Newton Howard
University of Oxford, UK

Yunqing Xia
Microsoft Research Asia, CHINA

Tat-Seng Chua
National University of Singapore, SINGAPORE

Computational Intelligence for Big Social Data Analysis

In the eras of social connectedness and social colonization, people are becoming increasingly enthusiastic about interacting, sharing, and collaborating through online collaborative media. In recent years, this collective intelligence has spread to many different areas, with particular focus on fields related to everyday life such as commerce, tourism, education, and health, causing the size of the Social Web to expand exponentially. The distillation of knowledge from such a large amount of unstructured information, however, is an extremely difficult task, as the contents of today's Web are perfectly suitable for human consumption, but remain hardly understandable to machines.

Big social data analysis grows out of this need and combines multiple disciplines such as social network analysis, multimedia management, social media analytics, trend discovery, and opinion mining. For example, studying the evolution of a social network merely as a graph is very limited as it does not take into account the information flowing between network nodes. Similarly, processing social interaction contents between network members without taking into account connections between them is limited by the fact that information flows cannot be properly weighted. Big social data analysis, instead, aims to study large-scale Web phenomena such as social networks from a holistic point of view, i.e., by concurrently taking into account all the socio-technical aspects involved in their dynamic evolution. Hence, big social data analysis is inherently interdisciplinary and spans areas such as machine learning, graph mining, information retrieval, knowledge-based systems, linguistics, common-sense reasoning, natural language processing, and big data computing.

Big social data analysis finds applications in several different scenarios. There is a good number of companies, large and small, that include the analysis of social data as part of their missions. Big social data analysis can be exploited for the creation and automated upkeep of review and opinion aggregation websites, in which opinionated text and videos are continuously gathered from the Web and not restricted only to product reviews, but also to wider topics such as political issues and brand perception. Big social data analysis has also a great potential as a sub-component technology for other systems. They can enhance the capabilities of customer relationship management and recommendation systems allowing, for example, to find out which features customers are particularly happy about or to exclude the recommendations items that have received very negative feedbacks. Similarly, they can be exploited for affective tutoring and affective entertainment or for troll filtering and spam detection in online social communication.

Business intelligence is also one of the main factors behind corporate interest in big social data analysis. Nowadays, companies invest an increasing amount of money in marketing strategies and they are constantly interested in both collecting and predicting the attitudes of the general public towards their products and brands. The design of automatic tools capable to mine sentiments over the Web in real-time and to create condensed versions of them represents one of the most active research and development areas. The development of such systems, moreover, is not only important for commercial purposes, but also for government intelligence applications able to monitor increases in hostile communications or to model cyber-issue diffusion.

Several commercial and academic tools track public viewpoints on a large-scale by offering graphical summarizations of trends and opinions in the blogosphere. Nevertheless, most commercial off-the-shelf (COTS) tools are limited to a polarity evaluation or a mood classification according to a very limited set of emotions. In addition, such methods mainly rely on parts of text in which emotional states are explicitly expressed and, hence, they are unable to capture opinions and sentiments that are expressed implicitly. Because they are mainly based on statistical properties associated with words, in fact, many COTS tools are easily tricked by linguistic operators such as negation and disjunction.

The main motivation for this Special Issue is to explore how computational intelligence can help overcome such hurdles through new forms of processing that can handle high volume, high

Digital Object Identifier 10.1109/MCI.2016.2572481
Date of publication: 18 July 2016

velocity, and high variety information assets and, hence, enable a more efficient passage from (unstructured) social information to (structured) machine-processable data, in potentially any domain. We received 34 submissions, which were reviewed by 63 prominent scholars, and selected 4 articles out of them.

The article "*Leveraging Cross-Domain Social Media Analytics to Understand TV Topics Popularity*" by Ruggero Pensa, Maria Luisa Sapino, Claudio Schifanella, and Luca Vignaroli proposes a concept-level integration framework in which users' activities on different social media are collectively represented, and possibly enriched with external knowledge, such as information extracted from the Electronic Program Guides, or available ontological domain knowledge. The integration framework has a knowledge graph as its core data model. It keeps track of active users, the television events they talk about, the concepts they mention in their activities, as well as different relationships existing among them. Temporal relationships are also captured to enable temporal analysis of the observed activity. The data model allows different types of analysis and the definition of global metrics in which the activity on different media concurs with the measure of success.

In "*An Efficient Memetic Algorithm for Influence Maximization in Social Networks*", Maoguo Gong, Chao Song, Chao Duan, Lijia Ma, and Bo Shen implement a memetic algorithm for community-based influence maximization in social networks. Such an algorithm optimizes the 2-hop influence spread to find the most influential nodes. Problem-specific population initialization and similarity-based local search are designed to accelerate the convergence of the algorithm. Experiments on three real-world datasets demonstrate that the proposed algorithm has competitive performance to the compared algorithms in terms of effectiveness and efficiency. For example, on a real-world network of 15233 nodes and 58891 edges, the influence spread of the proposed algorithm is 12.5%, 13.2% and 173.5% higher than the three frequently-used algorithms Degree, PageRank and Random, respectively.

Next, in "*Learning User and Product Distributed Representations Using a Sequence Model for Sentiment Analysis*", handled independently during review process, Tao Chen, Ruifeng Xu, Yulan He, Yunqing Xia, and Xuan Wang argue that the temporal relations of product reviews might be potentially useful for learning user and product embedding and thus propose employing a sequence model to embed these temporal relations into user and product representations so as to improve the performance of document-level sentiment analysis. Specifically, a distributed representation of each review is first learnt by a one-dimensional convolutional neural network. Then, taking these representations as pre-trained vectors, authors use a recurrent neural network with gated recurrent units to learn distributed representations of users and products. Finally, user, product and review representations are fed into a machine learning classifier for sentiment classification. Experimental results show that sequence modeling for the purposes of distributed user and product representation learning can improve the performance of document-level sentiment classification.

Finally, "*Statistical Learning Theory and ELM for Big Social Data Analysis*" is presented by Luca Oneto, Federica Bisio, Erik Cambria, and Davide Anguita. The paper, handled independently during review process, illustrates how to exploit the most recent technological tools and advances in Statistical Learning Theory in order to efficiently build an Extreme Learning Machine (ELM) for big social data analysis. ELM represents a powerful learning tool, developed to overcome some issues in back-propagation networks. The main problem with ELM is training it to work in the event of a large number of available samples, where the generalization performance has to be carefully assessed. For this reason, the authors propose an ELM implementation that exploits the Spark distributed in memory technology in order to address the issue of selecting ELM hyperparameters that give the best generalization performance.

President's Message

(continued from page 3)

Within the CIS organization, there is an Emergent Technologies Technical Committee (ETTC), led by Dr. Mengjie Zhang. The ETTC identifies, promotes, nurtures, and tracks new and emergent approaches, concepts, and areas that relate to or are within the scope of the CIS. Their activities include organizing special sessions in conferences, symposia, special issues in publications, and other technical activities. The ETTC is one of the 11 technical committees under the Technical Activities Committee. And let me take this opportunity to welcome Dr. Hussein Abbass, a professor with the University of New South Wales, Australia, who was recently elected IEEE CIS Vice-President of Technical Activities for the term 2016-2017, and is now chairing the Technical Activities Committee.

I invite all of you who are interested in emerging topics in CI to contact Dr. Hussein Abbass, or me. I encourage you to move the frontiers of knowledge forward with your research, and to submit original articles to the new TETCI journal, with your contribution, making it highly successful.

Pablo A. Estevez