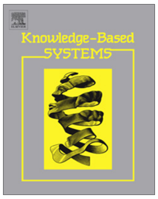
Knowledge-Based Systems 69 (2014) 1–2

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys



Guest Editorial: Big Social Data Analysis

In the era of social connectedness, Web users 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 accessible to machines.

Big social data analysis grows out of this need and combines 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 limiting 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 these 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. Accordingly, the contained articles in this issue cover variegated topics including stock market prediction, political forecasting, time-evolving opinion mining, social network analysis, and human–robot interaction. Out of the forty submissions received for this special issue, twelve were accepted. Three of the accepted papers underwent four rounds of revisions, four papers underwent three, and the rest underwent two revisions.

The article *"Time Corpora: Epochs, Opinions and Changes"* by Octavian Popescu and Carlo Strapparava proposes to explore diachronic phenomena by using large corpora of chronologically ordered language and, hence, identify previously unknown correlations between language usage and time periods, or epochs. Authors focus on a statistical approach to epoch delimitation and introduce the task of epoch characterization. They investigate the significant changes in the distribution of terms in the Google N-gram corpus and their relationships with emotion words.

In *"News Impact on Stock Price Return via Sentiment Analysis"*, Xiaodong Li, Haoran Xie, Li Chen, Jianping Wang, and Xiaotie Deng first implement a generic stock price prediction framework and plug in six different models with different analyzing approaches. They use Harvard psychological dictionary and Loughran–McDonald financial sentiment dictionary to construct a sentiment space. Textual news articles are then quantitatively measured and projected onto such a sentiment space. Instance labeling method is rigorously discussed and tested. Authors evaluate the models' prediction accuracy and empirically compare their performance at different market classification levels. Experiments are conducted on five years historical Hong Kong Stock Exchange prices and news articles.

Next, the article *"PoliTwi: Early Detection of Emerging Political Topics on Twitter and the Impact on Concept-Level Sentiment Analysis"* by Sven Rill, Dirk Reinel, Jorg Scheidt, and Roberto Zicari presents a system designed to detect emerging political topics in Twitter sooner than other standard information channels. For the analysis, authors have collected about 4 million tweets before and during the parliamentary election 2013 in Germany, from April until September 2013. It is shown, that new topics appearing in Twitter can be detected right after their occurrence. Moreover, authors have compared their results to Google Trends, observing that the topics emerged earlier in Twitter than in Google Trends.

"Analyzing Future Communities in Growing Citation Networks" is presented by Sukhwan Jung and Aviv Segev who view the research community as a Social Web where the communication happens through academic papers. The paper proposes methods to analyze how communities change over time in the citation network graph without additional external information and based on node and link prediction and community detection. Different combinations of the proposed methods are also analyzed. The identified communities are classified using key term labeling. Experiments show that the proposed methods can identify the changes in citation communities multiple years in the future with performance differing according to the analyzed time span.

Following, *"Sentic Patterns: Dependency-Based Rules for Concept-Level Sentiment Analysis"* – paper handled independently during review process – is elaborated upon by Soujanya Poria, Erik Cambria, Gregoire Winterstein, and Guang-Bin Huang who introduce a novel paradigm to concept-level sentiment analysis that merges linguistics, common-sense computing, and machine learning for improving the accuracy of tasks such as polarity detection. By allowing sentiments to flow from concept to concept based on the dependency relation of the input sentence, authors achieve a better understanding of the contextual role of each concept within the sentence and, hence, obtain a polarity detection engine that outperforms state-of-the-art statistical methods.

"Microblogging as a Mechanism for Human-Robot Interaction" is presented by David Bell, Theodora Koulouri, Stanislao Lauria, Robert D. Macredie and James Sutton who propose a novel approach to social data analysis, exploring the use of microblogging to manage

http://dx.doi.org/10.1016/j.knosys.2014.07.002
0950-7051/© 2014 Published by Elsevier B.V.

interaction between humans and robots, and evaluating an architecture that extends the use of social networks to connect humans and devices. The approach uses natural language processing techniques to extract features of interest from textual data retrieved from a microblogging platform in real-time and, hence, generate appropriate executable code for the robot. The simple rule-based solution exploits some of the 'natural' constraints imposed by microblogging platforms to manage the potential complexity of the interactions and create bi-directional communication.

The possibility of *"Enriching Semantic Knowledge Bases for Opinion Mining in Big Data Applications"* is analyzed by Albert Weichselbraun, Stefan Gindl, and Arno Scharl who present a novel method for contextualizing and enriching large semantic knowledge bases for opinion mining with a focus on Web intelligence platforms and other high-throughput big data applications. The method is not only applicable to traditional sentiment lexicons, but also to more comprehensive, multi-dimensional affective resources such as SenticNet. It comprises the following steps: (i) identify ambiguous sentiment terms, (ii) provide context information extracted from a domain-specific training corpus, and (iii) ground this contextual information to structured background knowledge sources such as ConceptNet and WordNet.

"Meta-Level Sentiment Models for Big Social Data Analysis" is subsequently suggested by Felipe Bravo-Marquez, Marcelo Mendoza, and Barbara Poblete who study how different dimensions of opinions, such as subjectivity, polarity, intensity and emotion, complement each other in specific scenarios. To this end, authors propose a novel approach for sentiment classification based on meta-level features. This supervised approach boosts existing sentiment classification of subjectivity and polarity detection on Twitter. Results show that the combination of meta-level features provides significant improvements in performance. However, authors observe that there are important differences that rely on the type of lexical resource, the dataset used to build the model, and the learning strategy.

The contribution *"Using Relation Selection to Improve Value Population in a ConceptNet-based Sentiment Dictionary"* by Chi-En Wu and Richard Tzong-Han Tsai illustrates how a common-sense knowledge base can be used as the foundation to build a larger dictionary by propagating sentiment values from concepts with known values to empty ones. Based on the assumption that concepts pass their sentiment values to their neighbors in different ways depending on the relations connecting them, authors use sequential forward search to find the best combination of relations. They also propose a bias correction method that guarantees that the average deviation and standard deviation of sentiment values in the whole sentiment dictionary remain unchanged.

"EmoSenticSpace: A Novel Framework for Affective Common-Sense Reasoning" – paper handled independently during review process – is then discussed by Soujanya Poria, Alexander Gelbukh, Erik Cambria, Amir Hussain, and Guang-Bin Huang who propose a new common-sense reasoning framework that extends WordNet-Affect and SenticNet by providing both emotion labels and polarity scores for a large set of natural language concepts. The framework is built by means of fuzzy c-means clustering and support-vector-machine classification, and takes into account different similarity measures, such as point-wise mutual information and emotional affinity. EmoSenticSpace was tested on three emotion-related natural language processing tasks, namely sentiment analysis, emotion recognition, and personality detection. In all cases, the proposed framework outperforms the state of the art.

In *"Extracting Relevant Knowledge for the Detection of Sarcasm and Nastiness in the Social Web"*, Raquel Justo, Thomas Corcoran, Stephanie Lukin, Marilyn Walker, Maria Ines Torres discuss the task of automatic detection of sarcasm or nastiness in online written conversation. It requires a system that can manage some kind of knowledge to interpret that emotional language is being used. In this work, authors try to provide this knowledge to the system by considering alternative sets of features obtained according to different criteria. They test a range of different feature sets using two different classifiers. Their results show that the sarcasm detection task benefits from the inclusion of linguistic and semantic information sources, while nasty language is more easily detected using only a set of surface patterns or indicators.

Finally, *"Crowd Explicit Sentiment Analysis"* is presented by Arturo Montejo-Raez, Manuel Diaz-Galiano, Fernando Martinez-Santiago, and Alfonso Urena-Lopez. This paper introduces an approach for sentiment analysis in social media environments. Similar to explicit semantic analysis, microblog posts are indexed by a predefined collection of documents. In the proposed approach, performed by means of latent semantic analysis, these documents are built up from common emotional expressions in social streams.

Acknowledgements

The guest editors are grateful to the Editors-in-Chief, Hamido Fujita and Jie Lu, and to the thirty-eight reviewers for their timely and insightful reviews of the submissions: Basant Agarwal, Aijun An, Ayoub Bagheri, Alexandra Balahur, Federica Bisio, Cristina Bosco, Paula Carvalho, Hsin-Hsi Chen, Ping Chen, Xueqi Cheng, Alok N. Choudhary, Rossana Damiano, Frantisek Darena, Amitava Das, Rafael Del Hoyo, Shi Feng, Flavius Frasincar, Paolo Gastaldo, Alexander Gelbukh, Marco Grassi, Catherine Havasi, Minlie Huang, Amir Hussain, Andrea Kleinsmith, Eugenio Martinez Camara, Thomas Mazzocco, Nir Ofek, Soujanya Poria, Paolo Rosso, Rafal Rzepka, Ulrich Schiel, Rui Xia, Yunqing Xia, Yves Xie, Hua Xu, Duo Zhang, Lei Zhang, Yongzheng Zhang.

Erik Cambria
Nanyang Technological University, Singapore

Haixun Wang
Google Research, USA

Bebo White
Stanford University, USA

Available online 11 July 2014