Affective Reasoning for Big Social Data Analysis

Erik Cambria, Amir Hussain, and Alessandro Vinciarelli

As the Web rapidly evolves, Web users are evolving with it. In an era of social connectedness, people are becoming increasingly enthusiastic about interacting, sharing, and collaborating through social networks, online communities, blogs, Wikis, and other 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 Web to expand exponentially.

The distillation of knowledge from such a big 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. The opportunity to capture the opinions of the general public about social events, political movements, company strategies, marketing campaigns, and product preferences has raised growing interest both within the scientific community, leading to many exciting open challenges, as well as in the business world, due to the remarkable benefits to be had from marketing and financial market prediction.

Existing approaches to big social data analysis mainly rely on parts of text in which sentiment is explicitly expressed, e.g., through polarity terms or affect words (and their co-occurrence frequencies). However, opinions and sentiments are often conveyed implicitly through latent semantics, which make purely syntactical approaches ineffective. By spanning from computer science to psychology, and from social science to cognitive science, the emerging interdisciplinary field of research of affective computing can aid in enhancing such approaches by enabling intelligent systems to recognize, feel, infer and interpret human emotions.

To this end, this special section focuses on the introduction, presentation, and discussion of novel techniques that further develop and apply affective reasoning tools and techniques for big social data analysis. A key motivation for this special section, in particular, is to explore the adoption of novel affective reasoning frameworks and cognitive learning systems to go beyond a mere word-level analysis of natural language text and provide novel concept-level tools and techniques that allow a more efficient passage from (unstructured) natural language to (structured) machine-processable affective data, in potentially any domain. The selected papers aim to address the wide spectrum of issues related to affective computing research and, hence, better grasp the current limitations and opportunities related to this fast-evolving branch of artificial intelligence. Out of the 29 submissions received, 5 were accepted to appear in the special section. One of the accepted papers underwent 3 rounds of revisions, the rest were revised twice.

The article *"Bootstrapping Social Emotion Classification with Semantically Rich Hybrid Neural Networks"* by Xiangsheng Li, Yanghui Rao, Haoran Xie, Raymond Y. K. Lau, Jian Yin, and Fu Lee Wang describes the development of a semantically-rich hybrid neural network. This novel model leverages unsupervised teaching models to incorporate semantic domain knowledge into a neural network to bootstrap its inference power and interpretability. According to the authors, this is the first successful work of incorporating semantics into neural networks to enhance social emotion classification and network interpretability.

In *"Inferring Affective Meanings of Words from Word Embedding"*, Minglei Li, Qin Lu, Yunfei Long, and Lin Gui propose a regression-based method to automatically infer multi-dimensional affective representation of words via their word embedding based on a set of seed words. This method can make use of the rich semantic meanings obtained from word embedding to extract meanings in some specific semantic space. This is based on the assumption that different features in word embedding contribute differently to a particular affective dimension and a particular feature in word embedding contributes differently to different affective dimensions. Experiments on a sentiment analysis task show that the lexicons extended by this method achieve better results than publicly available sentiment lexicons on eight different sentiment corpora.

Next, the article *"A Neural Word Embeddings Approach For Multi-Domain Sentiment Analysis"* by Mauro Dragoni and Giulio Petrucci proposes a new approach for multi-domain sentiment analysis. One of the main issues related to sentiment analysis algorithms is their poor capabilities to adapt to domains that are different from those used for training the opinion model. In this paper, the authors present an approach exploiting the linguistic overlap between different domains to build sentiment models supporting polarity inference for documents belonging to every domain. The proposed technique is validated by following the Dranziera protocol in order to ease the repeatability of the experiments and the comparison of the results.

The article *"Coarse-grained +/-Effect Word Sense Disambiguation for Implicit Sentiment Analysis"* by Yoonjung Choi,

- *E. Cambria is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798. E-mail: cambria@ntu.edu.sg.*
- *A. Hussain is with the School of Natural Sciences, University of Stirling, Stirling FK9 4LA, United Kingdom. E-mail: ahu@cs.stir.ac.uk.*
- *A. Vinciarelli is with the School of Computing Science, University of Glasgow, Glasgow G12 8QQ, United Kingdom. E-mail: vincia@dcs.gla.ac.uk.*

For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TAFFC.2017.2763218

Janyce Wiebe, and Rada Mihalcea presents a knowledge-based +/− effect coarse-grained word sense disambiguation (WSD) method based on selectional preferences modeled via topic models. Since their purpose is to determine whether an instance of a word in the corpus is being used with a +effect, −effect, or null sense, authors do not need to perform fine-grained WSD to pinpoint the exact sense. In particular, they propose a WSD method that does not require any sense-tagged training data as it exploits WordNet relations and glosses, rather than supervised WSD.

Finally, *"Distantly Supervised Lifelong Learning for Large-scale Social Media Sentiment Analysis"* is presented by Rui Xia, Jie Jiang, and Huihui, who propose a distantly supervised lifelong learning framework for large-scale social media sentiment analysis. The key characteristic of their approach is continuous sentiment learning in social media. It learns on past tasks sequentially, retains the knowledge obtained from past learning and uses the past knowledge to help future learning. The proposed classifier is trained on two large-scale distantly supervised social media datasets respectively, and evaluated on nine benchmark datasets.

ACKNOWLEDGMENTS

The guest editors are grateful to the Editor-in-Chief, Björn Schuller, and to the many reviewers for their timely and insightful reviews of the submissions.

Erik Cambria received the PhD in computing science and mathematics, in 2012 following the completion of an EPSRC project in collaboration with MIT Media Lab, which was selected as impact case study by the University of Stirling for the UK Research Excellence Framework (REF2014). After working at HP Labs India, Microsoft Research Asia, and NUS Temasek Labs, in 2014 he joined NTU SCSE as an assistant professor. His current affiliations include Rolls Royce, A*STAR, and MIT Synthetic Intelligence Lab. He is associate editor of several journals edited by Elsevier, e.g., INFFUS and KBS, Springer, e.g., AIRE and Cognitive Computation, and IEEE, e.g., CIM and the *Intelligent Systems*, where he manages the Department of Affective Computing and Sentiment Analysis. He is also recipient of many awards, e.g., Temasek Research Fellowship and Emerald Citations of Excellence, founder of SenticNet, a Singapore-based university spin-off offering B2B sentiment analysis services, and is involved in several international conferences as PC member, e.g., AAAI, UAI, ACL, and EMNLP, workshop organizer, e.g., ICDM SENTIRE, and program chair, e.g., ELM. He is regularly serves as invited speaker at both international and local events, e.g., IEEE SSCI and OpenGov, and his work is often featured in the news, e.g, Datanami.

Amir Hussain received the BEng and PhD degrees from the University of Strathclyde in Glasgow, Scotland, United Kingdom, in 1992 and 1997, respectively. He is currently professor of computing science, and founding director of the COSIPRA Laboratory, University of Stirling in Scotland, United Kingdom. His research interests include inter-disciplinary and industry-focused, and include multimodal big data cognitive and sentic computing techniques and applications. He has conducted and led collaborative research with industry (attracting over US$3m in research grants, as Principal Investigator); partnered in major European research programs, and supervised more than 20 PhDs. He has published more than 270 papers, including over a dozen books and 80 journal papers. He is founding editor-in-chief of the journals the *Cognitive Computation* (Springer Neuroscience, USA), and the *Big Data Analytics* (BioMed Central), and chief-editor of the Springer Book Series on Socio-Affective Computing, and SpringerBriefs on Cognitive Computation. He is associate editor of the *IEEE Transactions on Neural Networks and Learning Systems*, a member of several Technical Committees of the IEEE Computational Intelligence Society (CIS), founding publications co-Chair of the IINNS Big Data Section and its annual INNS Conference on Big Data, and chapter chair of the IEEE UK & RI Industry Applications Society.

Alessandro Vinciarelli is with the University of Glasgow, where he is a full professor (associate professor) with the School of Computing Science and an associate academic with the Institute of Neuroscience and Psychology. His main research interest is social signal processing, the domain aimed at modeling analysis, and synthesis of nonverbal behavior in social interactions. In particular, he has investigated approaches for role recognition in multiparty conversations, automatic personality perception from speech, and conflict analysis and measurement in competitive discussions. Overall, he has published more than 120 works, including one authored book, six edited volumes, and 35 journal papers. He has participated in the organization of the IEEE International Conference on social computing as a program chair in 2011 and as a general chair in 2012, he has initiated and chaired a large number of international workshops, including the Social Signal Processing Workshop, the International Workshop on Socially Intelligence Surveillance and Monitoring, the International Workshop on Human Behavior Understanding, the Workshop on Political Speech, and the Workshop on Foundations of Social Signals. Furthermore, he is or has been the principal investigator of several national and international projects, including a European Network of Excellence (the SSPNet). He is the cofounder of Klewel, a knowledge management company recognized with several awards.