



Guest Editorial Introduction

Affective neural networks and cognitive learning systems for big data analysis

Amir Hussain^a, Erik Cambria^b, Björn Schuller^{c,d}, Newton Howard^e^a University of Stirling, United Kingdom^b Nanyang Technological University, Singapore^c Technische Universität München, Germany^d Imperial College, London^e MIT Media Laboratory, USA

ARTICLE INFO

Article history:

Available online 14 August 2014

Keywords:

Big data analysis

Affective neural networks

Cognitive learning systems

Web users continue to evolve with the rapidly evolving Web. In an era of social connectedness, people are becoming increasingly enthusiastic about interacting, sharing, and collaborating through social networks, online communities, blogs, Wikis, as well as a range of other online collaborative media. In recent years, this collective intelligence has spread to many different areas, with a particular focus in 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 large amount of unstructured information, however, remains an extremely difficult task, since the contents of today's Web are perfectly suitable for human consumption, yet hardly accessible to machines.

The opportunity to capture the opinions of the general public on 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 expected from marketing and financial market prediction. Existing approaches to opinion mining 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.

In this light, this Special Issue focuses on the introduction, presentation, and discussion of novel techniques that further develop and apply big data analysis tools and techniques for sentiment analysis. A key motivation for this Special Issue, in particular, was to explore the adoption of novel affective neural networks 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 facilitate a more efficient passage from (unstructured) natural language to (structured) machine-processable data, in potentially any domain.

Despite significant progress, opinion mining and sentiment analysis are still finding their own voice as new interdisciplinary fields. In recent years, several workshops have been specifically organized on such topics, to name just a few: Workshop on Opinion Mining and Sentiment Analysis (WOMSA) in CAEPIA09, Topic-Sentiment Analysis for Mass Opinion Measurement (TSA) in CIKM09, Computational Approaches to Analysis and Generation of Emotion in Text (CAAGET) in NAACL10, Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA) in ECAI10 and ACL11/12, NAACL13, and ACL14, Sentiment Elicitation from Natural Text for Information Retrieval and Extraction (SENTIRE) in ICDM11/12/13/14, the Workshop series on Emotion, Sentiment and Social Signals (ES3LOD) at LREC 10/12/14, and the Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM) in KDD12/13 and ICML14.

Apart from our 2013 IEEE Intelligent Systems (vol. 28, no. 2 and 3) Special Issue on Concept-Level Opinion and Sentiment Analysis, the most recent journal special issue published entirely on such topics is dated 2006, specifically the Special Issue on

E-mail addresses: ahu@cs.stir.ac.uk (A. Hussain), cambria@ntu.edu.sg (E. Cambria), schuller@ieee.org (B. Schuller), nhmit@mit.edu (N. Howard).

Sentiment Analysis of Computational Intelligence (vol. 22, no. 2). More recent special issues in leading journals include sentiment analysis only as one of the (many) topics of interest: for example, the Speech Communication Special Issue on Sensing Emotion and Affect – Facing Realism in Speech Processing (2011), focused mainly on the acoustic speech signal analysis, yet not a single paper dealt exclusively with natural language processing; the Computer Speech and Language Special Issue on Affective Speech in real-life interactions (2011) again mainly focused on acoustic properties of speech; the IEEE Transactions on Systems, Man and Cybernetics (Part A) Special Issue on Social Media Analytics – Understanding the Pulse of the Society (2011) was of rather broad focus and only consisted of four papers; and finally, the ACM Transactions on Intelligent Systems and Technology Special Issue on Search and Mining User-generated Contents (2011) had sentiment analysis covered in only a quarter of the issue's scope.

The selected articles in this Issue span areas such as machine learning, novel neural and cognitive models, data mining, pattern recognition, knowledge-based systems, information retrieval, natural language processing, and big data computing. We believe the breadth of topics covered reflect the Issue's aims to address *Affective Neural Networks and Cognitive Learning Systems for Big Data Analysis*. Out of the thirty-two submissions received for this special issue, twelve were accepted. Five of the accepted papers underwent two rounds of revisions, six underwent three, and one four.

The article “*Discrete Particle Swarm Optimization for Identifying Community Structures in Signed Social Networks*” by Qing Cai, Maoguo Gong, Bo Shen, Lijia Ma, and Licheng Jiao proposes a novel discrete particle swarm optimization (PSO) algorithm for identifying community structures in signed networks. PSO is an artificial intelligent algorithm originated from social behavior such as birds flocking and fish schooling, which has proved to be an effective optimization technique. In the suggested method, particles' status have been redesigned in discrete form so as to make PSO more appropriate for discrete scenarios, and particles' updating rules have been reformulated by making use of the topology of the signed network. Extensive experiments compared with three state-of-the-art approaches on both synthetic and real-world signed networks demonstrate that the proposed method is effective and promising.

“*An Incremental Community Detection Method for Social Tagging Systems Using Locality-Sensitive Hashing*” is presented by Zhenyu Wu, and Ming Zou who model the dynamic unstructured data generated from social networks as a stream. Tag assignments stream clustering (TASC), an incremental scalable community detection method, is proposed based on locality-sensitive hashing. Both tags and latent interactions among users are incorporated in the method. In the experiments, the social dynamic behaviors of users are first analyzed. The proposed TASC method is then compared with state-of-the-art clustering methods such as StreamKmeans and incremental k-clique; results indicate that TASC can detect communities more efficiently and effectively.

In “*Affective Topic Model for Social Emotion Detection*”, Yanghui Rao, Qing Li, Liu Wenyin, Qingyuan Wu, and Xiaojun Quan focus on the detection of emotions evoked in a reader by social media. Compared to classical sentiment analysis conducted from the writer's perspective, analysis from the reader's perspective can be more meaningful when applied to social media. The authors propose an affective topic model with the intention to bridge the gap between social media materials and a reader's emotions by introducing an intermediate layer. The proposed model can be used to classify the social emotions of unlabeled documents and to generate a social emotion lexicon. Extensive evaluations using real-world data validate the effectiveness of the proposed model for both these applications.

Next, the article “*Modeling Virtual Organizations with Latent Dirichlet Allocation: a Case for Natural Language Processing*” by Alexander Gross, and Dhiraj Murthy explores a variety of methods for applying the Latent Dirichlet Allocation (LDA) automated topic modeling algorithm to modeling of the structure and behavior of virtual organizations found within modern social media and social networking environments. The authors introduce variants of LDA and ultimately make the argument that natural language processing is a critical interdisciplinary methodology to make better sense of social 'Big Data' and they were able to successfully model nested discussion topics from forums and blog posts using LDA. Importantly, they found that LDA is key to go beyond the state-of-the-art in conventional Social Network Analysis techniques.

Following, “*Semi-Supervised Word Polarity Identification in Resource-Less Languages*” is elaborated upon by Iman Dehdar-behbahani, Azadeh Shakery, and Hesham Faili who present a method that exploits a language with rich subjectivity analysis resources (English) to identify the polarity of words in a resource-less foreign language. The English WordNet and a sparse foreign WordNet infrastructure are used to create a heterogeneous, multilingual and weighted semantic network. To identify the semantic orientation of foreign words, a random walk based method is applied to the semantic network along with a set of automatically weighted English positive and negative seeds. In a post-processing phase, synonym and antonym relations in the foreign WordNet are used to filter the random walk results. The experiments on English and Persian languages show that the proposed method can outperform state-of-the-art word polarity identification methods in both languages.

The possibility of “*Incorporating Conditional Random Fields and Active Learning to Improve Sentiment Identification*” is analyzed by Kunpeng Zhang, Yusheng Xie, Yi Yang, Aaron Sun, Hengchang Liu, and Alok Choudhary who propose a method based on conditional random fields to incorporate sentence structure and context information in addition to syntactic information for improving sentiment identification. The authors also investigate how human interaction affects the accuracy of sentiment labeling using limited training data. They propose and evaluate two different active learning strategies for labeling sentiment data. Their experiments with the proposed approach demonstrate a 5–15% improvement in accuracy on Amazon customer reviews compared to existing supervised learning and rule-based methods.

“*A Classification of User-Generated Content into Consumer Decision Journey Stages*” is subsequently suggested by Silvia Vázquez, Oscar Muñoz-García, Inés Campanella, Marc Poch, Beatriz Fisas, Nuria Bel, and Gloria Andreu who present a novel analysis and classification of user-generated content in terms of its belonging to one of the four stages of the Consumer Decision Journey (i.e., the purchase process from the moment when a customer is aware of the existence of the product to the moment when he or she buys, experiences and talks about it). Using a corpus of short texts written in English and Spanish and extracted from different social media, the authors identify a set of linguistic patterns for each purchase stage that will be then used in a rule-based classifier. Additionally, they use machine learning algorithms to automatically identify business indicators such as the Marketing Mix elements. The classification of the purchase stages achieves an average precision of 74%. The proposed classification of texts depending on the Marketing Mix elements expressed achieved an average precision of 75% for all the elements analyzed.

The contribution “*Sentiments Analysis at Conceptual Level Making Use of the Narrative Knowledge Representation Language*” by Gian Piero Zarrilli illustrates some of the knowledge representation structures and inference procedures proper to a high-level, fully

implemented conceptual language, NKRL (Narrative Knowledge Representation Language). The aim is to show how these tools can be used to deal, in a sentiment analysis/opinion mining context, with some common types of human (and non-human) 'behaviors'. These behaviors correspond, in particular, to the concrete, mutual relationships among human and non-human characters that can be expressed under the form of non-fictional and real-time 'narratives' (i.e., as logically and temporally structured sequences of 'elementary events').

In "*Exploring Personalized Searches using Tag-based User Profiles and Resource Profiles in Folksonomy*", Yi Cai, Qing Li, Haoran Xie, and Huaqin Min examine the limitations of previous tag-based personalized searches. To handle these limitations, the authors propose a new method to model user profiles and resource profiles in collaborative tagging systems. They use a normalized term frequency to indicate the preference degree of a user on a tag. A novel search method using such profiles of users and resources is proposed to facilitate the desired personalization in resource searches. In the proposed framework, instead of the keyword matching or similarity measurement used in previous works, the relevance measurement between a resource and a user query (termed the query relevance) is treated as a fuzzy satisfaction problem of a user's query requirements. The authors implement a prototype system called the Folksonomy-based Multimedia Retrieval System (FMRS). Experiments using the FMRS data set and the MovieLens data set show that the proposed method outperforms baseline methods.

A "*Community-aware User Profile Enrichment in Folksonomy*" is then discussed by Haoran Xie, Qing Li, Xudong Mao, Xiaodong Li, Yi Cai, and Yanghui Rao. They propose a method of enriching user profiles based on latent user communities in folksonomy data. Specifically, the proposed approach contains four sub-processes: (i) tag-based user profiles are extracted from a folksonomy tripartite graph; (ii) a multi-faceted folksonomy graph is constructed by integrating tag and image affinity subgraphs with the folksonomy tripartite graph; (iii) random walk distance is used to unify various relationships and measure user similarities; (iv) a novel prototype-based clustering method based on user similarities is used to identify user communities, which are further used to enrich the extracted user profiles. To evaluate the proposed method, the authors conducted experiments using a public dataset, the results of which show that the approach outperforms previous ones in user profile enrichment.

"*A Multi-Label, Semi-Supervised Classification Approach Applied to Personality Prediction in Social Media*" is presented subsequently by Ana Carolina E. S. Lima, and Leandro Nunes de Castro. A

personality prediction system for social media data is introduced that differs from most approaches in the literature, in that it works with groups of texts, instead of single texts, and does not take users' profiles into account. Also, the proposed approach extracts meta-attributes from texts and does not work directly with the content of the messages. The set of possible personality traits is taken from the Big Five model and allows the problem to be characterized as a multi-label classification task. The problem is then transformed into a set of five binary classification problems and solved by means of a semi-supervised learning approach, due to the difficulty in annotating the massive amounts of data generated in social media. The system was applied to predict the personality of Tweets taken from three datasets available in the literature, and resulted in an approximately 83% accurate prediction.

Finally, in "*Semantically-Based Priors and Nuanced Knowledge Core For Big Data, Social AI, and Language Understanding*" Daniel Osher introduces COGBASE, a noise-resistant and nuanced framework that makes 10 million pieces of commonsense data and a host of novel reasoning algorithms available via a family of semantically-driven prior probability distributions. Machine learning, Big Data, natural language understanding/processing, and social AI can draw on COGBASE to determine lexical semantics, infer goals and interests, simulate emotion and affect, calculate document gists and topic models, and link commonsense knowledge to domain models and data from external domains (including social, spatial, cultural, psychological, and others). COGBASE is particularly ideal for social Big Data, which tends to involve highly implicit social contexts, cognitive artifacts, difficult-to-parse texts, and deep domain knowledge dependencies.

Acknowledgments

The guest editors are most grateful to the co-editors-in-chief, Kenji Doya and DeLiang Wang, for giving us the opportunity to guest edit this exciting issue. Finally, the guest editors are indebted to the following thirty-four reviewers for their timely and insightful reviews of the submissions which has assured the quality of this Special Issue: Hasan Abbasi Nozari, Haydar Akca, Sabri Arik, Waqas Bangyal, Cristina Bosco, Federica Bisio, Ping Chen, Rossana Damiano, Rafael Del Hoyo, Kamran Farooq, Tom Foulsham, Paolo Gastaldo, Marco Grassi, William Neil Howell, Chia-Nan Ko, Olcay Kursun, Kang Li, Qingshan Liu, Thomas Mazzocco, Ali Minai, Nir Ofek, Soujanya Poria, Roman Rosipal, Paolo Rosso, Rafal Rzepka, Mohammad Saif, Marcello Sanguineti, Richard Tsai, Rui Xia, Yunqing Xia, Yves Xie, Erfu Yang, Lei Zhang, Yongzheng Zhang.