

# Knowledge-Based Approaches to Concept-Level Sentiment Analysis

Erik Cambria, *National University of Singapore*

Björn Schuller, *Technische Universität München*

Bing Liu, *University of Illinois at Chicago*

Haixun Wang, *Microsoft Research Asia*

Catherine Havasi, *Massachusetts Institute of Technology*

**T**he ways people express their opinions and sentiments have radically changed in the past few years thanks to the advent of social networks, Web communities, blogs, wikis, and other online collaborative media. The distillation of knowledge from the huge amount of unstructured information on

the Web can be a key factor for marketers who want to create an image or identity in the minds of their customers for their product, brand, or organization. These online social data, however, remain hardly accessible to computers, as they're specifically meant for human consumption. The automatic analysis of online opinions, in fact, involves a deep understanding of natural language text by machines, from which we're still very far.

## **Traditional Methods**

Until recently, online information retrieval has been mainly based on algorithms relying on the textual representation of webpages. Such algorithms are quite good at retrieving texts, splitting them into parts, checking the spelling, and counting their words. But when it comes to interpreting sentences and extracting meaningful information, their capabilities are limited.

Early works aimed to classify entire documents as containing overall positive or negative polarity, or rating scores of reviews. Such systems were mainly based on supervised approaches relying on manually labeled samples, such as movie or product reviews where the reviewer's overall positive or negative attitude was explicitly indicated. However, opinions and sentiments occur at more than the document level, and they aren't limited to a single valence or target. Contrary or complementary attitudes toward the same topic or multiple topics can be present across the span of a document.

Later works adopted a segment-level opinion analysis aiming to distinguish sentimental from non-sentimental sections—for example, by using graph-based techniques for segmenting sections of a document on the basis of their subjectivity, or by performing a classification based on some fixed syntactic phrases that are likely to be used to express opinions. In more recent works, text-analysis granularity has been taken down to the sentence level, for example, by using the presence of opinion-bearing lexical items (single words or  $n$ -grams) to detect subjective sentences, or by exploiting association rule mining for a feature-based analysis of product reviews. These approaches, however, are still far from being able to infer the cognitive and affective information associated with natural language, because they mainly rely on knowledge bases that are still too limited to efficiently process text at the sentence level. Moreover, such text analysis granularity might still not be enough, because a single sentence could contain different opinions about different facets of the same product or service.

### **Concept-Level Approaches**

In light of these considerations, this special issue focuses on the introduction,

presentation, and discussion of novel approaches to opinion mining and sentiment analysis that are based on domain-dependent corpora as well as general-purpose semantic knowledge bases. The main motivation for the special issue is to go beyond a mere word-level analysis of text and provide novel concept-level approaches to opinion mining and sentiment analysis that allow a more efficient passage from (unstructured) textual information to (structured) machine-processable data, in potentially any domain.

For this special issue, we received 30 articles, of which six were carefully selected. The first article, “New Avenues in Opinion Mining and Sentiment Analysis” by Erik Cambria and his colleagues—which was handled independently during the review process—introduces this special issue with a short survey. It reviews past, present, and future trends of sentiment analysis. It covers common tasks for sentiment analysis along with the main approaches, and discusses the evolution of different tools and techniques—from heuristics to discourse structure, from coarse- to fine-grained analysis, and from keyword to concept-level opinion mining. The article also discusses the emergence of multimodal sentiment analysis and considers future tendencies.

In “Building a Concept-Level Sentiment Dictionary Based on Commonsense Knowledge” by Angela Charng-Rung Tsai and her colleagues, a two-step method combining iterative regression and random walk with in-link normalization is proposed to build a concept-level sentiment dictionary. ConceptNet is exploited for propagating sentiment values based on the assumption that semantically related concepts share common sentiment. Another peculiarity of the article is that it uses polarity accuracy, Kendall distance, and average-maximum ratio, instead of mean error, to better evaluate sentiment dictionaries.

The next contribution, “Enhanced SenticNet with Affective Labels for Concept-Based Opinion Mining” by Soujanya Poria and his colleagues, presents a methodology for enriching SenticNet concepts with affective information by assigning to them an emotion label. The authors used various features extracted from the International Survey of Emotion Antecedents and Reactions (ISEAR, an emotion-related dataset), as well as similarity measures that rely on the polarity data provided in SenticNet, those based on WordNet, and ISEAR distance-based measures, including point-wise mutual information, and emotional affinity.

Then, in “Extracting and Grounding Contextualized Sentiment Lexicons,” Albert Weichselbraun, Stefan Gindl, and Arno Scharl suggest a hybrid approach that combines lexical analysis and machine learning to cope with ambiguity and integrate the context of sentiment terms. The method identifies ambiguous terms that vary in polarity (depending on the context) and stores them in contextualized sentiment lexicons. In conjunction with semantic knowledge bases, these lexicons help ground ambiguous sentiment terms to concepts that correspond to their polarity.

Next, in “Using Objective Words in SentiWordNet to Improve Word-of-Mouth Sentiment Classification,” Chihli Hung and Hao-Kai Lin propose the re-evaluation of objective words in SentiWordNet by assessing the sentimental relevance of such words and their associated sentiment sentences. Two sampling strategies are proposed and integrated with support vector machines for sentiment classification. According to the experiments, the proposed approach significantly outperforms the traditional sentiment-mining approach that ignores the importance of objective words in SentiWordNet.

## THE AUTHORS

**Erik Cambria** is a research scientist in the Cognitive Science Programme, Temasek Laboratories, National University of Singapore. His research interests include AI, the Semantic Web, natural language processing, and big social data analysis. Cambria has a PhD in computing science and mathematics from the University of Stirling. He is on the editorial board of Springer's *Cognitive Computation* and is the chair of many international conferences such as Brain-Inspired Cognitive Systems (BICS) and Extreme Learning Machines (ELM). Contact him at cambria@nus.edu.sg.

**Björn Schuller** leads the Machine Intelligence and Signal Processing group at the Institute for Human-Machine Communication at the Technical University of Munich. His research interests include machine learning, affective computing, and automatic speech recognition. Schuller has a PhD in electrical engineering and information technology from the Technical University of Munich. Contact him at schuller@tum.de.

**Bing Liu** is a professor of Computer Science at the University of Illinois at Chicago (UIC). His research interests include opinion mining and sentiment analysis, Web mining, and data mining. Liu has a PhD in artificial intelligence from the University of Edinburgh. He has served as the associate editor of *IEEE Transactions on Knowledge and Data Engineering* (TKDE), the *Journal of Data Mining and Knowledge Discovery* (DMKD), and *KDD Explorations*. Contact him at liub@cs.uic.edu.

**Haixun Wang** is a researcher in the Natural Language Processing Department at Google Research. His research interests include data management, graph systems, data mining, semantic networks, and text analytics. Wang has a PhD in computer science from the University of California, Los Angeles. He is the associate editor of *IEEE Transactions on Knowledge and Data Engineering* (TKDE) and the *Journal of Computer Science and Technology* (JCST). Contact him at haixun@gmail.com.

**Catherine Havasi** is a cofounder of the Open Mind Common Sense project at the Massachusetts Institute of Technology (MIT) Media Lab, where she works as a postdoctoral associate. Her research interests include commonsense reasoning, dimensionality reduction, machine learning, language acquisition, cognitive modeling, and intelligent user interfaces. Havasi has a PhD in computer science from Brandeis University. Contact her at havasi@media.mit.edu.

The last article, “Developing Corpora for Sentiment Analysis: The Case of Irony and Senti-TUT” by Cristina Bosco, Viviana Patti, and Andrea Bolioli focuses on the main issues related to the development of a corpus for opinion mining and sentiment analysis both by surveying existing works in this area and presenting, as a case study, an ongoing project for Italian, called Senti-TUT, where a corpus for the investigation of irony about politics in social media is developed.

**T**hese articles are a solid and varied representation of some of the exciting challenges and solutions emerging in this field. We hope that you enjoy the special issue and that this research fosters future innovations.

### Acknowledgments

We would like to thank the Editor in Chief, Daniel Zeng, for his help with this special issue and the 68 reviewers that not only helped with the decision process, but contributed with excellent reviews to make this issue special: Abhishek Jaientilal, Alexandra Balahur, Alexey Solovyev, Amac Herdagdelen, Andreas Hotho, Antonio Reyes, Appavu Balamurugan, A.R. Balamurali, Carlo Strapparava, Carmen Banea, Cristina Bosco, Cyril Joder, Daniel Olsher, Danushka Bollegala, Dennis Clark, Efstratios Kontopoulos, Eugene Bann, Felix Burkhardt, Felix Weninger, Fernando Fernandez-Martinez, Florian Eyben, Giovanni Acampora, Girgori Sidorov, Henry Anaya-Sanchez, Hidenao Abe, Huan Huang, Jane Malin, Jorge Carrillo de Albornoz, Jose Antonio Troyano, Jose Barranquero, Juan Augusto, Jun Deng, Kalina Bontcheva, Karthik Dinakar, Ken Arnold, Khiet Truong, Laura Plaza, Laurence Devillers, Leslie Fife, Ling Chen, Maarten van der Heijden, Mandy Dang, Marcelo Armentano, Marchi Erik, Mariel Ale, Matthew Aitkenhead, Mehdi Adda, Mitsuru Ishizuka, Mohd Helmy Abd Wahab, Nikolaos Engonopoulos, Paolo Gastaldo, Paolo Rosso, Rada Mihalcea, Richard Crowder, Rodrigo Agerri, Sameera Abar, Serge Sharoff, Stefan Siersdorfer, Stefan Steidl, Stephen Poteet, Vered Aharonson, Wen Xiong, Wen-Han Chao, Wenjing Han, Yair Neuman, Yongzheng Zhang, Zixing Zhang, and Zornitsa Kozareva.