Concept-Level Sentiment Analysis

A World Wide Web Conference 2014 Tutorial

http://sentic.net/tutorial

Erik Cambria
Temasek Laboratories
National University of Singapore
cambria@nus.edu.sg

Aims and Scope

The WWW'14 tutorial on Concept-Level Sentiment Analysis aims to provide its participants means to efficiently design models, techniques, tools, and services for concept-level sentiment analysis and their commercial realizations. The tutorial draws on insights resulting from the recent IEEE Intelligent Systems special issues on Concept-Level Opinion and Sentiment Analysis[1] and the IEEE CIM special issue on Computational Intelligence for Natural Language Processing[2]. The tutorial includes a hands-on session to illustrate how to build a concept-level opinion-mining engine step-by-step, from semantic parsing to concept-level reasoning.

Background and Motivation

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 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. 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.

[1] `http://sentic.net/closa`
[2] `http://sentic.net/cinlp`

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2577268

Mining opinions and sentiments from natural language, however, is an extremely difficult task as it involves a deep understanding of most of the explicit and implicit, regular and irregular, syntactical and semantic rules proper of a language. Existing approaches mainly rely on parts of text in which opinions and sentiments are explicitly expressed such as polarity terms, affect words and their co-occurrence frequencies. However, opinions and sentiments are often conveyed implicitly through latent semantics, which make purely syntactical approaches ineffective [2].

Concept-level sentiment analysis focuses on a semantic analysis of text through the use of web ontologies or semantic networks, which allow the aggregation of conceptual and affective information associated with natural language opinions [5, 6, 7, 8, 9, 10, 11]. By relying on external knowledge, such approaches step away from blind use of keywords and word co-occurrence count, but rather rely on the implicit features associated with natural language concepts. Unlike purely syntactical techniques, concept-based approaches are able to detect also sentiments that are expressed in a subtle manner, e.g., through the analysis of concepts that do not explicitly convey any emotion, but which are implicitly linked to other concepts that do so [4]. The bag-of-concepts model can represent semantics associated with natural language much better than bags-of-words. In the bag-of-words model, in fact, a concept such as `cloud computing` would be split into two separate words, disrupting the semantics of the input sentence (in which, for example, the word `cloud` could wrongly activate concepts related to `weather`) [1].

The analysis at concept-level allows for the inference of semantic and affective information associated with natural language text and, hence, enables comparative fine-grained feature-based sentiment analysis. Rather than gathering isolated opinions about a whole item (e.g., iPhone5), users are generally more interested in comparing different products according to specific features (e.g., iPhone5's vs Galaxy S3's touchscreen), or even sub-features (e.g., fragility of iPhone5's vs Galaxy S3's touchscreen).

In this context, the construction of comprehensive common and common-sense knowledge bases is key for feature-spotting and polarity detection, respectively. Common-sense, in particular, is necessary to properly deconstruct natural language text into sentiments – for example, to appraise the concept `small room` as negative for a hotel review and `small queue` as positive for a post office, or the concept `go read the book` as positive for a book review but negative for a movie review [3].

Tutorial Program

- Introduction *(5 mins)*
- New Avenues in Sentiment Analysis Research
  - From Heuristics to Discourse Structure *(5 mins)*
  - From Coarse- to Fine-Grained Analysis *(5 mins)*
  - From Keywords to Concepts *(10 mins)*
- Concept-Level Models
  - Knowledge acquisition models *(10 mins)*
  - Emotion categorization models *(10 mins)*
  - Vector space models *(10 mins)*
- Concept-Level Techniques
  - Analogical reasoning *(10 mins)*
  - Parallel analogy *(10 mins)*
  - Spreading activation *(10 mins)*
- Concept-Level Tools
  - Sentiment resources *(15 mins)*
  - Common knowledge repositories *(15 mins)*
  - Aspect mining and polarity detection *(10 mins)*
- Building a Concept-Level Opinion-Mining Engine
  - Semantic parsing *(15 mins)*
  - Sentic API *(15 mins)*
  - Application Samples *(20 mins)*
- Conclusion *(5 mins)*

Impact and Relevance

The World Wide Web Conference is a global event bringing together key researchers, innovators, decision-makers, technologists, and business experts trying to make meaning out of Web data. Within this research and business area, opinion mining and sentiment analysis have become increasingly important subtasks in recent years. However, there are still many challenges, including social information understanding and integration, that need to be addressed. For these reasons, a tutorial on concept-level sentiment analysis is strongly relevant to WWW'14.

Target Audience and Prerequisites

The target audience includes researchers and professionals in the fields of sentiment analysis, Web data mining, and related areas. The tutorial also aims to attract researchers from industry community as it covers research efforts for the development of applications in fields such as commerce, tourism, education, and health. The audience is expected to have basic computer science skills, but psychologists and sociologists are also very welcome. The tutorial not only covers state-of-the-art approaches to concept-level sentiment analysis, but also provides information about techniques and tools to be used for practical opinion mining.

About the Tutor

Erik Cambria received his BEng and MEng with honors in Electronic Engineering from the University of Genova in 2005 and 2008, respectively. In 2011, he has been awarded a PhD in Computing Science and Mathematics, following the completion of an industrial Cooperative Awards in Science and Engineering (CASE) research project born from the collaboration between the University of Stirling and the MIT Media Laboratory. Today, Erik is the lead investigator of a MINDEF-funded project on Commonsense Knowledge Representation & Reasoning at the National University of Singapore (Temasek Laboratories) and an associate researcher at the MIT Media Laboratory (Synthetic Intelligence Project). His interests include AI, Semantic Web, KR, NLP, opinion mining and sentiment analysis, affective and cognitive modeling, intention awareness, HCI, and e-health. Erik is also chair of several international conferences, e.g., Extreme Learning Machines (ELM), and workshop series, e.g., ICDM SENTIRE. He is on the editorial board of Springer Cognitive Computation and he is a guest editor of many other leading AI journals. Erik is also a fellow of the Brain Sciences Foundation, the Chinese Academy of Sciences, National Taiwan University, Microsoft Research Asia, and HP Labs India.

1. REFERENCES

[1] Erik Cambria, Paolo Gastaldo, Federica Bisio, and Rodolfo Zunino. An ELM-based model for affective analogical reasoning. *Neurocomputing*, 2014.

[2] Erik Cambria and Amir Hussain. *Sentic Computing: Techniques, Tools, and Applications*. Springer, Dordrecht, Netherlands, 2012.

[3] Erik Cambria, Bjoern Schuller, Yunqing Xia, and Catherine Havasi. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 28(2):15–21, 2013.

[4] Erik Cambria and Bebo White. Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine*, 9(2), 2014.

[5] Aldo Gangemi, Valentina Presutti, and Diego Reforgiato. Frame-based detection of opinion holders and topics: a model and a tool. *IEEE Computational Intelligence Magazine*, 9(1):20–30, 2014.

[6] Raymond Lau, Yunqing Xia, and Yunming Ye. A probabilistic generative model for mining cybercriminal networks from online social media. *IEEE Computational Intelligence Magazine*, 9(1):31–43, 2014.

[7] Veronica Perez-Rosas, Rada Mihalcea, and Louis-Philippe Morency. Multimodal sentiment analysis of Spanish online videos. *IEEE Intelligent Systems*, 28(3):38–45, 2013.

[8] Soujanya Poria, Alexander Gelbukh, Amir Hussain, Dipankar Das, and Sivaji Bandyopadhyay. Enhanced SenticNet with affective labels for concept-based opinion mining. *IEEE Intelligent Systems*, 28(2):31–30, 2013.

[9] Angela Tsai, Richard Tsai, and Jane Hsu. Building a concept-level sentiment dictionary based on commonsense knowledge. *IEEE Intelligent Systems*, 28(2):22–30, 2013.

[10] Albert Weichselbraun, Stefan Gindl, and Arno Scharl. Extracting and grounding context-aware sentiment lexicons. *IEEE Intelligent Systems*, 28(2):39–46, 2013.

[11] Rui Xia, Chengqing Zong, Xuelei Hu, and Erik Cambria. Feature ensemble plus sample selection: A comprehensive approach to domain adaptation for sentiment classification. *IEEE Intelligent Systems*, 28(3):10–18, 2013.