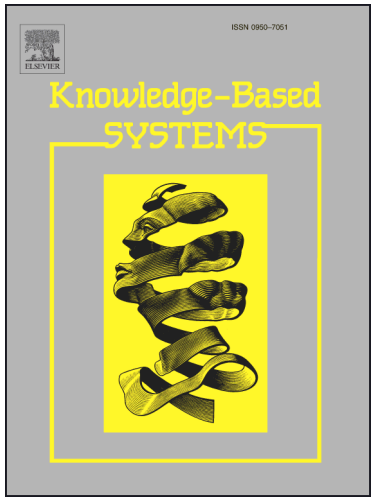

Sentic patterns: Dependency-based rules for concept-level sentiment analysis

Soujanya Poria, Erik Cambria, Grégoire Winterstein, Guang-Bin Huang

PII: S0950-7051(14)00183-X
DOI: http://dx.doi.org/10.1016/j.knosys.2014.05.005
Reference: KNOSYS 2846

To appear in: *Knowledge-Based Systems*

Received Date: 15 December 2013
Revised Date: 30 April 2014
Accepted Date: 2 May 2014

Please cite this article as: S. Poria, E. Cambria, G. Winterstein, G-B. Huang, Sentic patterns: Dependency-based rules for concept-level sentiment analysis, *Knowledge-Based Systems* (2014), doi: http://dx.doi.org/10.1016/j.knosys.2014.05.005

ELSEVIER

Knowledge-Based Systems (2014) 1–32

KBS

Sentic Patterns: Dependency-Based Rules for Concept-Level Sentiment Analysis

Soujanya Poria[a], Erik Cambria[b], Grégoire Winterstein[c], Guang-Bin Huang[a]

[a] *School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore*
[b] *School of Computer Engineering, Nanyang Technological University, Singapore*
[c] *Department of Linguistics, University of Paris Diderot, France*

Abstract

The Web is evolving through an era where the opinions of users are getting increasingly important and valuable. The distillation of knowledge from the huge amount of unstructured information on the Web can be a key factor for tasks such as social media marketing, branding, product positioning, and corporate reputation management. These online social data, however, remain hardly accessible to computers, as they are specifically meant for human consumption. The automatic analysis of online opinions involves a deep understanding of natural language text by machines, from which we are still very far. To this end, concept-level sentiment analysis aims to go beyond a mere word-level analysis of text and provide novel approaches to opinion mining and sentiment analysis that enable a more efficient passage from (unstructured) textual information to (structured) machine-processable data. A recent knowledge-based technology in this context is sentic computing, which relies on the ensemble application of common-sense computing and the psychology of emotions to infer the conceptual and affective information associated with natural language. Sentic computing, however, is limited by the richness of the knowledge base and by the fact that the bag-of-concepts model, despite more sophisticated than bag-of-words, misses out important discourse structure information that is key for properly detecting the polarity conveyed by natural language opinions. In this work, we 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, in particular, we 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.

Keywords: Artificial intelligence, Natural language processing, Opinion mining, Sentic computing, Linguistic rules

1. Introduction

Between the dawn of the Internet through year 2003, there were just a few dozens exabytes of information on the Web. Today, that much information is created weekly. The opportunity to capture the opinions of the general public about social events, political movements, company strategies, marketing campaigns, and product preferences has raised increasing interest both in the scientific community, for the exciting open challenges, and in the business world, for the remarkable fallouts in marketing and financial prediction. Keeping up with the ever-growing amount of unstructured information on the Web, however, is a formidable task and requires fast and efficient models for opinion mining. Hitherto, natural language processing (NLP) and online information retrieval have been mainly based on

Email addresses: SPoria@ntu.edu.sg (Soujanya Poria), cambria@ntu.edu.sg (Erik Cambria), gregoire.winterstein@linguist.univ-paris-diderot.fr (Grégoire Winterstein), EGBHuang@ntu.edu.sg (Guang-Bin Huang)