

Preface

The 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 this huge amount of unstructured information 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 organisation. These online social data, however, remain hardly accessible to computers, as they are 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 are still very far.

Hitherto, online information retrieval has been mainly based on algorithms relying on the textual representation of Web pages. Such algorithms are very 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 known to be very limited. Existing approaches to opinion mining and sentiment analysis, in particular, can be grouped into three main categories: keyword spotting, in which text is classified into categories based on the presence of fairly unambiguous affect words; lexical affinity, which assigns arbitrary words a probabilistic affinity for a particular emotion; statistical methods, which calculate the valence of affective keywords and word co-occurrence frequencies on the base of a large training corpus.

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 opinionist's overall positive or negative attitude was explicitly indicated. However, opinions and sentiments do not occur only at document level, nor are they 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. In more recent works, text analysis granularity has been taken down to segment and sentence level, e.g., by using 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 as they mainly rely on knowledge bases that are still too limited to efficiently process text at sentence level. In this book, common sense computing techniques are further developed and applied to bridge the semantic gap between word-level natural language data and the concept-level opinions conveyed by these. In particular, the ensemble application of graph mining and multi-dimensionality reduction techniques on two common sense knowledge bases was exploited to develop a novel intelligent engine for opendomain opinion mining and sentiment analysis. The proposed approach, termed sentic computing, performs a clause-level semantic analysis of text, which allows the inference of both the conceptual and emotional information associated with natural language opinions and, hence, a more efficient passage from (unstructured) textual information to (structured) machine-processable data.

The engine was tested on three different resources, namely a Twitter hashtag repository, a LiveJournal database, and a PatientOpinion dataset, and its performance compared both with results obtained using standard sentiment analysis techniques and using different state-of-the-art knowledge bases such as Princeton's WordNet, MIT's ConceptNet, and Microsoft's Probase. Differently from most currently available opinion mining services, the developed engine does not base its analysis on a limited set of affect words and their co-occurrence frequencies, but rather on common sense concepts and the cognitive and affective valence conveyed by these. This allows the engine to be domain-independent and, hence, to be embedded in any opinion mining system for the development of intelligent applications in multiple fields such as Social Web, HCI, and e-health. Looking ahead, the combined novel use of different knowledge bases and of common sense reasoning techniques for opinion mining proposed in this work, will, eventually, pave the way for development of more bio-inspired approaches to the design of natural language processing systems capable of handling knowledge, retrieving it when necessary, making analogies, and learning from experience.

Sentic Computing

Techniques, Tools, and Applications

Cambria, E.; Hussain, A.

2012, XVIII, 153 p. 39 illus., 35 illus. in color., Softcover

ISBN: 978-94-007-5069-2