

Computing Attitude and Affect in Text: Theory and Applications

James G. Shanahan, Yan Qu, and Janyce Wiebe (editors)
(Clairvoyance Corporation and University of Pittsburgh)

Dordrecht: Springer (The information retrieval series, edited by W. Bruce Croft),
2006, xv+341 pp; hardbound, ISBN 1-4020-4026-1, \$99.00, €74.95

Reviewed by
Michael Gamon
Microsoft Research

The detection of attitude and subjectivity in text has received increased attention over the past several years. Whereas standard text classification typically deals with the identification of the topic of a text, it has become clear that attitudes expressed in the text constitute another very useful dimension. With the rapid rise of customer-produced content such as product reviews on the Web, for example, opinion mining has become an important and viable technology in business intelligence. Similarly, question answering and information retrieval have been shown to benefit from the detection of attitude and subjectivity, and—as this new collection shows—the range of other possible applications is large.

The volume contains 24 extended versions of papers that were originally presented at the American Association for Artificial Intelligence (AAAI) Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications at Stanford University in 2004, organized by the editors of this volume. This symposium was the first such event to focus exclusively on this topic. Since then, there has been a surge of work in the area; today one is hard pressed to find a major conference in the field that does not have a workshop or a dedicated session on the topic. Publicly available annotated corpora such as the MPQA corpus (Multi-Perspective Question Answering) and the movie review corpus from Cornell University have contributed greatly to a focused research effort.

The chapters in the book are arranged in three consecutive categories: Linguistic and Cognitive Models, Lexical Resources and Attitude/Affect Recognition and Generation, and Applications. For brevity, I will discuss only a sample of the papers, to give an impression of the breadth of the research. The first eight papers explore the linguistic complexity of the expression of attitude and the cognitive background of these expressions. Polanyi and Zaenen, and Karlgren, Eriksson, and Franzén investigate how context word choices and syntactic structure can influence and change the attitudinal dimension of an expression. Bucci and Maskit approach the problem from a psychological angle; with multiple annotated data sets, they use statistical methods to identify a weighted set of words corresponding to the psycholinguistic dimension of Referential Activity (roughly equivalent to the capability of a text to evoke nonverbal experience). Stoyanov et al. discuss the annotation scheme used in the MPQA corpus.

The next seven papers of the collection touch on an even wider range of areas, focusing on lexical resources and the automatic detection and generation of affect. Bethard et al. use semantic parsing to identify opinion clauses and their holders, pointing out that doing so can be beneficial for tasks where it is important to distinguish opinion from fact, such as summarization, information extraction, and question answering. Teufel explores the role of affect structure for improved citation indexing of academic papers, which often contain indications of whether a citation is used in support or in contrast

to the proposed hypothesis or experimental results. Inkpen, Feiguina, and Hirst explore the possibility of changing affect in language generation, presenting the same content in positive or negative terms.

The Applications section of the book starts with two investigations of the use of Systemic Functional Grammar, one for assessing interpersonal distance in text and the other for identifying genre distinctions. The next two articles deal with the argumentative structure of text. Argumentative structure is used here for critiquing scientific abstracts by novice writers. DiMarco et al. examine the role of hedging in citation contexts in scientific papers, work that is closely related to Teufel's paper in the previous section. Two articles in this section examine machine-learning techniques: for feature selection and for feature generalization. The latter is performed in the joint extraction of topics and their associated opinions, using WordNet as a resource to identify synonymy and hypernymy. Tong and Yager's work addresses the summarization of temporal trends in opinions using Internet data sources such as blogs and newsgroups. The remaining papers deal with the summarization of viewpoints, and with creating labeled data for good versus bad news from the reaction of financial markets to company news.

It should be clear from this quick overview that there is no shortage of perspectives on the topic of affect and attitude in text in this collection. Although this breadth may seem at first glance to pull in too many diverse directions, the reader should keep in mind that it is exactly this diversity of perspectives that is the hallmark of a successful first meeting dedicated to the topic. No one will come away from a reading of this volume with a narrow and simplistic view of the problem, and it would be surprising if the papers did not spark additional ideas in the interested reader.

The volume could have benefitted, however, from some more thorough editorial organization to bring more structure into the collection. The categorization into three main topics is not indicated in the table of contents; it seems to be more of a hasty afterthought. For many papers, it is not clear at all why they are part of one section rather than another. A more thorough introduction could have helped to categorize the papers along various possible dimensions. These dimensions could include features and resources used (words, part-of-speech tags, deeper linguistic analysis, semantic resources), type of attitude/affect computed (simple polarity, trending, interpersonal distance, argumentative structure, certainty, multiple viewpoints, hedging, etc.), and type of data (product review texts, news, blogs/newsgroups, scientific writing, transcribed dialogues).

In summary, this volume should become an indispensable resource for anyone interested in this area. Whether the reader is more interested in the computational or the linguistic aspects of the problem—or even just the range of possible applications—this collection will broaden the perspective on the issue. For readers with no background in sentiment detection the volume can serve as an initial overview of the field, with the caveat that they would have to do some additional reading of both earlier seminal work such as Hatzivassiloglou and McKeown (1997), Wiebe (2000), Turney (2002), and Pang, Lee, and Vaithyanathan (2002), and some more recent developments (as can be found in recent proceedings of COLING, ACL, and EMNLP).

References

Hatzivassiloglou, Vasileios and Kathleen McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th Annual Meeting of the Association for Computational*

Linguistics and 8th conference of the European Chapter of the Association for Computational Linguistics, pages 174–181, Madrid.
Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?

- Sentiment classification using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*, pages 79–86, Philadelphia.
- Turney, Peter. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 417–424, Philadelphia.
- Wiebe, Janyce. 2000. Learning subjective adjectives from corpora. In *Proceedings of the 17th National Conference on Artificial Intelligence and 12th Conference on Innovative Application of Artificial Intelligence*, pages 735–740, Austin.

Michael Gamon has been a member of the Natural Language Processing group of Microsoft Research since 1996. His publications include research on sentiment detection and mining for sentiment terminology. He organized the Workshop on Sentiment and Subjectivity in Text at COLING/ACL 2006 in Sydney. Gamon's e-mail address is: mgamon@microsoft.com; URL: <http://research.microsoft.com/users/mgamon/mgamon.aspx>.

