Opinion Mining and Sentiment Analysis

Opinion Mining and Sentiment Analysis

Bo Pang

Yahoo! Research
Sunnyvale, CA 94089
USA

bopang@yahoo-inc.com

Lillian Lee

Computer Science Department
Cornell University
Ithaca, NY 14853
USA

llee@cs.cornell.edu

the essence of knowledge

Boston – Delft

Foundations and Trends® in Information Retrieval

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is B. Pang and L. Lee, Opinion Mining and Sentiment Analysis, Foundations and Trends® in Information Retrieval, vol 2, nos 1–2, pp 1–135, 2008

ISBN: 978-1-60198-150-9

Foundations and Trends® in Information Retrieval
Volume 2 Issue 1–2, 2008

Editorial Board

Editorial Scope

Foundations and Trends® in Information Retrieval will publish survey and tutorial articles in the following topics:

- Applications of IR
- Architectures for IR
- Collaborative filtering and recommender systems
- Cross-lingual and multilingual IR
- Distributed IR and federated search
- Evaluation issues and test collections for IR
- Formal models and language models for IR
- IR on mobile platforms
- Indexing and retrieval of structured documents
- Information categorization and clustering
- Information extraction
- Information filtering and routing
- Metasearch, rank aggregation and data fusion
- Natural language processing for IR
- Performance issues for IR systems, including algorithms, data structures, optimization techniques, and scalability
- Question answering
- Summarization of single documents, multiple documents, and corpora
- Text mining
- Topic detection and tracking
- Usability, interactivity, and visualization issues in IR
- User modelling and user studies for IR
- Web search

Information for Librarians
Foundations and Trends® in Information Retrieval, 2008, Volume 2, 4 issues. ISSN paper version 1554-0669. ISSN online version 1554-0677. Also available as a combined paper and online subscription.

Foundations and Trends® in
Information Retrieval
Vol. 2, Nos. 1–2 (2008) 1–135

DOI: 10.1561/1500000001

now
the essence of knowledge

Opinion Mining and Sentiment Analysis

Bo Pang[1] and Lillian Lee[2]

[1] *Yahoo! Research,Sunnyvale, CA 94089, USA, bopang@yahoo-inc.com*
[2] *Computer Science Department, Cornell University, Ithaca, NY 14853, USA, llee@cs.cornell.edu*

Abstract

An important part of our information-gathering behavior has always been to find out what other people think. With the growing availability and popularity of opinion-rich resources such as online review sites and personal blogs, new opportunities and challenges arise as people now can, and do, actively use information technologies to seek out and understand the opinions of others. The sudden eruption of activity in the area of opinion mining and sentiment analysis, which deals with the computational treatment of opinion, sentiment, and subjectivity in text, has thus occurred at least in part as a direct response to the surge of interest in new systems that deal directly with opinions as a first-class object.

This survey covers techniques and approaches that promise to directly enable opinion-oriented information-seeking systems. Our focus is on methods that seek to address the new challenges raised by sentiment-aware applications, as compared to those that are already present in more traditional fact-based analysis. We include material

on summarization of evaluative text and on broader issues regarding privacy, manipulation, and economic impact that the development of opinion-oriented information-access services gives rise to. To facilitate future work, a discussion of available resources, benchmark datasets, and evaluation campaigns is also provided.

Contents

1

Introduction

> Romance should never begin with sentiment. It should begin with science and end with a settlement.
>
> — Oscar Wilde, *An Ideal Husband*

1.1 The Demand for Information on Opinions and Sentiment

"What other people think" has always been an important piece of information for most of us during the decision-making process. Long before awareness of the World Wide Web became widespread, many of us asked our friends to recommend an auto mechanic or to explain who they were planning to vote for in local elections, requested reference letters regarding job applicants from colleagues, or consulted *Consumer Reports* to decide what dishwasher to buy. But the Internet and the Web have now (among other things) made it possible to find out about the opinions and experiences of those in the vast pool of people that are neither our personal acquaintances nor well-known professional critics — that is, people we have never heard of. And conversely, more and more people are making their opinions available to strangers via the Internet.

Indeed, according to two surveys of more than 2000 American adults each [63, 127],

- 81% of Internet users (or 60% of Americans) have done online research on a product at least once;
- 20% (15% of all Americans) do so on a typical day;
- among readers of online reviews of restaurants, hotels, and various services (e.g., travel agencies or doctors), between 73% and 87% report that reviews had a significant influence on their purchase;[1]
- consumers report being willing to pay from 20% to 99% more for a 5-star-rated item than a 4-star-rated item (the variance stems from what type of item or service is considered);
- 32% have provided a rating on a product, service, or person via an online ratings system, and 30% (including 18% of online senior citizens) have posted an online comment or review regarding a product or service.[2]

We hasten to point out that consumption of goods and services is not the only motivation behind people's seeking out or expressing opinions online. A need for political information is another important factor. For example, in a survey of over 2500 American adults, Rainie and Horrigan [248] studied the 31% of Americans — over 60 million people — that were 2006 *campaign internet users*, defined as those who gathered information about the 2006 elections online and exchanged views via email. Of these,

- 28% said that a major reason for these online activities was to get perspectives from within their community, and 34% said that a major reason was to get perspectives from outside their community;
- 27% had looked online for the endorsements or ratings of external organizations;

[1] Section 6.1 discusses quantitative analyses of actual economic impact, as opposed to consumer perception.

[2] Interestingly, Hitlin and Rainie [123] report that "Individuals who have rated something online are also more skeptical of the information that is available on the Web."

- 28% said that most of the sites they use share their point of view, but 29% said that most of the sites they use challenge their point of view, indicating that many people are not simply looking for validations of their pre-existing opinions; and
- 8% posted their own political commentary online.

The user hunger for and reliance upon online advice and recommendations that the data above reveals is merely one reason behind the surge of interest in new systems that deal directly with opinions as a first-class object. But, Horrigan [127] reports that while a majority of American internet users report positive experiences during online product research, at the same time, 58% also report that online information was missing, impossible to find, confusing, and/or overwhelming. Thus, there is a clear need to aid consumers of products and of information by building better information-access systems than are currently in existence.

The interest that individual users show in online opinions about products and services, and the potential influence such opinions wield, is something that vendors of these items are paying more and more attention to [124]. The following excerpt from a whitepaper is illustrative of the envisioned possibilities, or at the least the rhetoric surrounding the possibilities:

> With the explosion of Web 2.0 platforms such as blogs, discussion forums, peer-to-peer networks, and various other types of social media ... consumers have at their disposal a soapbox of unprecedented reach and power by which to share their brand experiences and opinions, positive or negative, regarding any product or service. As major companies are increasingly coming to realize, these consumer voices can wield enormous influence in shaping the opinions of other consumers — and, ultimately, their brand loyalties, their purchase decisions, and their own brand advocacy. ... Companies can respond to the consumer insights they generate through social media monitoring and analysis by modifying their

> marketing messages, brand positioning, product development, and other activities accordingly.
>
> — Zabin and Jefferies [327]

But industry analysts note that the leveraging of new media for the purpose of tracking product image requires new technologies; here is a representative snippet describing their concerns:

> Marketers have always needed to monitor media for information related to their brands — whether it's for public relations activities, fraud violations,[3] or competitive intelligence. But fragmenting media and changing consumer behavior have crippled traditional monitoring methods. Technorati estimates that 75,000 new blogs are created daily, along with 1.2 million new posts each day, many discussing consumer opinions on products and services. Tactics [of the traditional sort] such as clipping services, field agents, and ad hoc research simply can't keep pace.
>
> — Kim [154]

Thus, aside from individuals, an additional audience for systems capable of automatically analyzing consumer sentiment, as expressed in no small part in online venues, are companies anxious to understand how their products and services are perceived.

1.2 What Might be Involved? An Example Examination of the Construction of an Opinion/Review Search Engine

Creating systems that can process subjective information effectively requires overcoming a number of novel challenges. To illustrate some of these challenges, let us consider the concrete example of what building an *opinion- or review-search* application could involve. As we have discussed, such an application would fill an important and prevalent

[3] Presumably, the author means "*the detection or prevention of* fraud violations," as opposed to the *commission* thereof.

information need, whether one restricts attention to blog search [213] or considers the more general types of search that have been described above.

The development of a complete review- or opinion-search application might involve attacking each of the following problems.

(1) If the application is integrated into a general-purpose search engine, then one would need to determine whether the user is in fact looking for subjective material. This may or may not be a difficult problem in and of itself: perhaps queries of this type will tend to contain indicator terms like "review," "reviews," or "opinions," or perhaps the application would provide a "checkbox" to the user so that he or she could indicate directly that reviews are what is desired; but in general, query classification is a difficult problem — indeed, it was the subject of the 2005 KDD Cup challenge [185].
(2) Besides the still-open problem of determining which documents are topically relevant to an opinion-oriented query, an additional challenge we face in our new setting is simultaneously or subsequently determining which documents or portions of documents contain review-like or opinionated material. Sometimes this is relatively easy, as in texts fetched from review-aggregation sites in which review-oriented information is presented in relatively stereotyped format: examples include Epinions.com and Amazon.com. However, blogs also notoriously contain quite a bit of subjective content and thus are another obvious place to look (and are more relevant than shopping sites for queries that concern politics, people, or other non-products), but the desired material within blogs can vary quite widely in content, style, presentation, and even level of grammaticality.
(3) Once one has target documents in hand, one is still faced with the problem of identifying the overall sentiment expressed by these documents and/or the specific opinions regarding particular features or aspects of the items or topics in question, as necessary. Again, while some sites make this

kind of extraction easier — for instance, user reviews posted to Yahoo! Movies must specify grades for pre-defined sets of characteristics of films — more free-form text can be much harder for computers to analyze, and indeed can pose additional challenges; for example, if quotations are included in a newspaper article, care must be taken to attribute the views expressed in each quotation to the correct entity.

(4) Finally, the system needs to present the sentiment information it has garnered in some reasonable summary fashion. This can involve some or all of the following actions:
    (a) Aggregation of "votes" that may be registered on different scales (e.g., one reviewer uses a star system, but another uses letter grades).
    (b) Selective highlighting of some opinions.
    (c) Representation of points of disagreement and points of consensus.
    (d) Identification of communities of opinion holders.
    (e) Accounting for different levels of authority among opinion holders.

    Note that it might be more appropriate to produce a visualization of sentiment data rather than a textual summary of it, whereas textual summaries are what is usually created in standard topic-based multi-document summarization.

1.3 Our Charge and Approach

Challenges (2), (3), and (4) in the above list are very active areas of research, and the bulk of this survey is devoted to reviewing work in these three sub-fields. However, due to space limitations and the focus of the journal series in which this survey appears, we do not and cannot aim to be completely comprehensive.

In particular, when we began to write this survey, we were directly charged to focus on information-access applications, as opposed to work of more purely linguistic interest. We stress that the importance of work in the latter vein is absolutely not in question.

Given our mandate, the reader will not be surprised that we describe the applications that sentiment-analysis systems can facilitate and review many kinds of approaches to a variety of opinion-oriented classification problems. We have also chosen to attempt to draw attention to single- and multi-document summarization of evaluative text, especially since interesting considerations regarding graphical visualization arise. Finally, we move beyond just the technical issues, devoting significant attention to the broader implications that the development of opinion-oriented information-access services have: we look at questions of privacy, manipulation, and whether or not reviews can have measurable economic impact.

1.4 Early History

Although the area of sentiment analysis and opinion mining has recently enjoyed a huge burst of research activity, there has been a steady undercurrent of interest for quite a while. One could count early projects on beliefs as forerunners of the area [48, 317]. Later work focused mostly on interpretation of metaphor, narrative, point of view, affect, evidentiality in text, and related areas [121, 133, 149, 262, 306, 310, 311, 312, 313].

The year 2001 or so seems to mark the beginning of widespread awareness of the research problems and opportunities that sentiment analysis and opinion mining raise [51, 66, 69, 79, 192, 215, 221, 235, 291, 296, 298, 305, 326], and subsequently there have been literally hundreds of papers published on the subject.

Factors behind this "land rush" include:

- the rise of machine learning methods in natural language processing and information retrieval;
- the availability of datasets for machine learning algorithms to be trained on, due to the blossoming of the World Wide Web and, specifically, the development of review-aggregation web-sites; and, of course
- realization of the fascinating intellectual challenges and commercial and intelligence applications that the area offers.

1.5 A Note on Terminology: Opinion Mining, Sentiment Analysis, Subjectivity, and All that

> 'The beginning of wisdom is the definition of terms,' wrote Socrates. The aphorism is highly applicable when it comes to the world of social media monitoring and analysis, where any semblance of universal agreement on terminology is altogether lacking.
>
> Today, vendors, practitioners, and the media alike call this still-nascent arena everything from 'brand monitoring,' 'buzz monitoring' and 'online anthropology,' to 'market influence analytics,' 'conversation mining' and 'online consumer intelligence'. ... In the end, the term 'social media monitoring and analysis' is itself a verbal crutch. It is placeholder [sic], to be used until something better (and shorter) takes hold in the English language to describe the topic of this report.
>
> — Zabin and Jefferies [327]

The above quotation highlights the problems that have arisen in trying to name a new area. The quotation is particularly apt in the context of this survey because the field of "social media monitoring and analysis" (or however one chooses to refer to it) is precisely one that the body of work we review is very relevant to. And indeed, there has been to date no uniform terminology established for the relatively young field we discuss in this survey. In this section, we simply mention some of the terms that are currently in vogue, and attempt to indicate what these terms tend to mean in research papers that the interested reader may encounter.

The body of work we review is that which deals with the computational treatment of (in alphabetical order) *opinion*, *sentiment*, and *subjectivity* in text. Such work has come to be known as *opinion mining*, *sentiment analysis*, and/or *subjectivity analysis*. The phrases *review mining* and *appraisal extraction* have been used, too, and there are some connections to *affective computing*, where the goals include enabling computers to recognize and express emotions [239]. This proliferation of terms reflects differences in the connotations that these terms carry,

both in their original general-discourse usages[4] and in the usages that have evolved in the technical literature of several communities.

In 1994, Wiebe [311], influenced by the writings of the literary theorist Banfield [26], centered the idea of *subjectivity* around that of *private states*, defined by Quirk et al. [245] as states that are not open to objective observation or verification. Opinions, evaluations, emotions, and speculations all fall into this category; but a canonical example of research typically described as a type of subjectivity analysis is the recognition of opinion-oriented language in order to distinguish it from objective language. While there has been some research self-identified as subjectivity analysis on the particular application area of determining the value judgments (e.g., "four stars" or "C+") expressed in the evaluative opinions that are found, this application has not tended to be a major focus of such work.

The term *opinion mining* appears in a paper by Dave et al. [69] that was published in the proceedings of the 2003 WWW conference; the publication venue may explain the popularity of the term within communities strongly associated with Web search or information retrieval. According to Dave et al. [69], the ideal opinion-mining tool would "process a set of search results for a given item, generating a list of product attributes (quality, features, etc.) and aggregating opinions

[4] To see that the distinctions in common usage can be subtle, consider how interrelated the following set of definitions given in *Merriam-Webster's Online Dictionary* are:

> Synonyms: opinion, view, belief, conviction, persuasion, sentiment mean a judgment one holds as true.
>
> - Opinion implies a conclusion thought out yet open to dispute ⟨each expert seemed to have a different opinion⟩.
> - View suggests a subjective opinion ⟨very assertive in stating his views⟩.
> - Belief implies often deliberate acceptance and intellectual assent ⟨a firm belief in her party's platform⟩.
> - Conviction applies to a firmly and seriously held belief ⟨the conviction that animal life is as sacred as human⟩.
> - Persuasion suggests a belief grounded on assurance (as by evidence) of its truth ⟨was of the persuasion that everything changes⟩.
> - Sentiment suggests a settled opinion reflective of one's feelings ⟨her feminist sentiments are well-known⟩.

about each of them (poor, mixed, good)." Much of the subsequent research self-identified as opinion mining fits this description in its emphasis on extracting and analyzing judgments on various aspects of given items. However, the term has recently also been interpreted more broadly to include many different types of analysis of evaluative text [190].

The history of the phrase *sentiment analysis* parallels that of "opinion mining" in certain respects. The term "sentiment" used in reference to the automatic analysis of evaluative text and tracking of the predictive judgments therein appears in 2001 papers by Das and Chen [66] and Tong [296], due to these authors' interest in analyzing market sentiment. It subsequently occurred within 2002 papers by Turney [298] and Pang et al. [235], which were published in the proceedings of the annual meeting of the Association for Computational Linguistics (ACL) and the annual conference on Empirical Methods in Natural Language Processing (EMNLP). Moreover, Nasukawa and Yi [221] entitled their 2003 paper, "Sentiment analysis: Capturing favorability using natural language processing", and a paper in the same year by Yi et al. [323] was named "Sentiment Analyzer: Extracting sentiments about a given topic using natural language processing techniques." These events together may explain the popularity of "sentiment analysis" among communities self-identified as focused on NLP. A sizeable number of papers mentioning "sentiment analysis" focus on the specific application of classifying reviews as to their polarity (either positive or negative), a fact that appears to have caused some authors to suggest that the phrase refers specifically to this narrowly defined task. However, nowadays many construe the term more broadly to mean the computational treatment of opinion, sentiment, and subjectivity in text.

Thus, when broad interpretations are applied, "sentiment analysis" and "opinion mining" denote the same field of study (which itself can be considered a sub-area of subjectivity analysis). We have attempted to use these terms more or less interchangeably in this survey. This is in no small part because we view the field as representing a unified body of work, and would thus like to encourage researchers in the area to share terminology regardless of the publication venues at which their papers might appear.

References

[1] A. Abbasi, "Affect intensity analysis of dark web forums," in *Proceedings of Intelligence and Security Informatics* (*ISI*), pp. 282–288, 2007.

[2] L. A. Adamic and N. Glance, "The political blogosphere and the 2004 U.S. election: Divided they blog," in *Proceedings of LinkKDD*, 2005.

[3] A. Agarwal and P. Bhattacharyya, "Sentiment analysis: A new approach for effective use of linguistic knowledge and exploiting similarities in a set of documents to be classified," in *Proceedings of the International Conference on Natural Language Processing (ICON)*, 2005.

[4] R. Agrawal, S. Rajagopalan, R. Srikant, and Y. Xu, "Mining newsgroups using networks arising from social behavior," in *Proceedings of WWW*, pp. 529–535, 2003.

[5] E. M. Airoldi, X. Bai, and R. Padman, "Markov blankets and meta-heuristic search: Sentiment extraction from unstructured text," *Lecture Notes in Computer Science*, vol. 3932 (Advances in Web Mining and Web Usage Analysis), pp. 167–187, 2006.

[6] G. A. Akerlof, "The market for "Lemons": Quality uncertainty and the market mechanism," *The Quarterly Journal of Economics*, vol. 84, pp. 488–500, 1970.

[7] S. M. Al Masum, H. Prendinger, and M. Ishizuka, "SenseNet: A linguistic tool to visualize numerical-valence based sentiment of textual data," in *Proceedings of the International Conference on Natural Language Processing (ICON)*, pp. 147–152, 2007. (Poster paper).

[8] J. Allan, "Introduction to topic detection and tracking," in *Topic Detection and Tracking: Event-based Information Organization*, (J. Allan, ed.), pp. 1–16, Norwell, MA, USA: Kluwer Academic Publishers, ISBN 0-7923-7664-1, 2002.

[9] C. O. Alm, D. Roth, and R. Sproat, "Emotions from text: Machine learning for text-based emotion prediction," in *Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, 2005.

[10] A. Anagnostopoulos, A. Z. Broder, and D. Carmel, "Sampling search-engine results," *World Wide Web*, vol. 9, pp. 397–429, 2006.

[11] R. K. Ando and T. Zhang, "A framework for learning predictive structures from multiple tasks and unlabeled data," *Journal of Machine Learning Research*, vol. 6, pp. 1817–1853, 2005.

[12] A. Andreevskaia and S. Bergler, "Mining WordNet for a fuzzy sentiment: Sentiment tag extraction from WordNet glosses," in *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL)*, 2006.

[13] W. Antweiler and M. Z. Frank, "Is all that talk just noise? The information content of internet stock message boards," *Journal of Finance*, vol. 59, pp. 1259–1294, 2004.

[14] N. Archak, A. Ghose, and P. Ipeirotis, "Show me the money! Deriving the pricing power of product features by mining consumer reviews," in *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2007.

[15] S. Argamon, ed., *Proceedings of the IJCAI Workshop on DOING IT WITH STYLE: Computational Approaches to Style Analysis and Synthesis*. 2003.

[16] S. Argamon, J. Karlgren, and J. G. Shanahan, eds., *Proceedings of the SIGIR Workshop on Stylistic Analysis of Text For Information Access*. ACM, 2005.

[17] S. Argamon, J. Karlgren, and O. Uzuner, eds., *Proceedings of the SIGIR Workshop on Stylistics for Text Retrieval in Practice*. ACM, 2006.

[18] S. Argamon-Engelson, M. Koppel, and G. Avneri, "Style-based text categorization: What newspaper am I reading?" in *Proceedings of the AAAI Workshop on Text Categorization*, pp. 1–4, 1998.

[19] Y. Attali and J. Burstein, "Automated essay scoring with e-rater v.2," *Journal of Technology, Learning, and Assessment*, vol. 26, February 2006.

[20] A. Aue and M. Gamon, "Automatic identification of sentiment vocabulary: Exploiting low association with known sentiment terms," in *Proceedings of the ACL Workshop on Feature Engineering for Machine Learning in Natural Language Processing*, 2005.

[21] A. Aue and M. Gamon, "Customizing sentiment classifiers to new domains: A case study," in *Proceedings of Recent Advances in Natural Language Processing (RANLP)*, 2005.

[22] B. Awerbuch and R. Kleinberg, "Competitive collaborative learning," in *Proceedings of the Conference on Learning Theory (COLT)*, pp. 233–248, 2005. (Journal version to appear in *Journal of Computer and System Sciences*, special issue on computational learning theory).

[23] P. Bajari and A. Hortaçsu, "The winner's curse, reserve prices, and endogenous entry: Empirical insights from eBay auctions," *RAND Journal of Economics*, vol. 34, pp. 329–355, 2003.

[24] P. Bajari and A. Hortaçsu, "Economic insights from internet auctions," *Journal of Economic Literature*, vol. 42, pp. 457–486, 2004.

[25] C. F. Baker, C. J. Fillmore, and J. B. Lowe, "The Berkeley Framenet Project," in *Proceedings of COLING/ACL*, 1998.

[26] A. Banfield, *Unspeakable Sentences: Narration and Representation in the Language of Fiction*. Routledge and Kegan Paul, 1982.

[27] M. Bansal, C. Cardie, and L. Lee, "The power of negative thinking: Exploiting label disagreement in the min-cut classification framework," in *Proceedings of the International Conference on Computational Linguistics (COLING)*, 2008. (Poster paper).

[28] R. Bar-Haim, I. Dagan, B. Dolan, L. Ferro, D. Giampiccolo, B. Magnini, and I. Szpektor, "The second PASCAL recognising textual entailment challenge," in *Proceedings of the Second PASCAL Challenges Workshop on Recognising Textual Entailment*, 2006.

[29] R. Barzilay and L. Lee, "Learning to paraphrase: An unsupervised approach using multiple-sequence alignment," in *Proceedings of the Joint Human Language Technology/North American Chapter of the ACL Conference (HLT-NAACL)*, pp. 16–23, 2003.

[30] R. Barzilay and K. McKeown, "Extracting paraphrases from a parallel corpus," in *Proceedings of the Association for Computational Linguistics (ACL)*, pp. 50–57, 2001.

[31] S. Basuroy, S. Chatterjee, and S. A. Ravid, "How critical are critical reviews? The box office effects of film critics, star power and budgets," *Journal of Marketing*, vol. 67, pp. 103–117, 2003.

[32] M. Bautin, L. Vijayarenu, and S. Skiena, "International sentiment analysis for news and blogs," in *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*, 2008.

[33] P. Beineke, T. Hastie, C. Manning, and S. Vaithyanathan, "Exploring sentiment summarization," in *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text*, AAAI technical report SS-04-07, 2004.

[34] F. Benamara, C. Cesarano, A. Picariello, D. Reforgiato, and V. S. Subrahmanian, "Sentiment analysis: Adjectives and adverbs are better than adjectives alone," in *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*, 2007. (Short paper).

[35] J. Berger, A. T. Sorensen, and S. J. Rasmussen, "Negative publicity: When is negative a positive?," Manuscript. PDF file's last modification date: October 16, 2007, *URL*: http://www.stanford.edu/~asorense/papers/Negative_Publicity.pdf, 2007.

[36] Y. Bestgen, C. Fairon, and L. Kerves, "Un baromètre affectif effectif: Corpus de référence et méthode pour déterminer la valence affective de phrases," in *Journées internationales d'analyse statistique des donnés textuelles (JADT)*, pp. 182–191, 2004.

[37] S. Bethard, H. Yu, A. Thornton, V. Hatzivassiloglou, and D. Jurafsky, "Automatic extraction of opinion propositions and their holders," in *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text*, 2004.

[38] D. Biber, *Variation Across Speech and Writing*. Cambridge University Press, 1988.

[39] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[40] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, Bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification," in *Proceedings of the Association for Computational Linguistics (ACL)*, 2007.

[41] S. R. K. Branavan, H. Chen, J. Eisenstein, and R. Barzilay, "Learning document-level semantic properties from free-text annotations," in *Proceedings of the Association for Computational Linguistics (ACL)*, 2008.

[42] E. Breck and C. Cardie, "Playing the telephone game: Determining the hierarchical structure of perspective and speech expressions," in *Proceedings of the International Conference on Computational Linguistics (COLING)*, 2004.

[43] E. Breck, Y. Choi, and C. Cardie, "Identifying expressions of opinion in context," in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, Hyderabad, India, 2007.

[44] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," in *Proceedings of the 7th International World Wide Web Conference*, pp. 107–117, 1998.

[45] R. F. Bruce and J. M. Wiebe, "Recognizing subjectivity: A case study in manual tagging," *Natural Language Engineering*, vol. 5, 1999.

[46] J. K. Burgoon, J. P. Blair, T. Qin, and J. F. Nunamaker, Jr., "Detecting deception through linguistic analysis," in *Proceedings of Intelligence and Security Informatics (ISI),* number 2665 in *Lecture Notes in Computer Science*, p. 958, 2008.

[47] L. Cabral and A. Hortaçsu, "The dynamics of seller reputation: Theory and evidence from eBay," Working Paper, downloaded version revised in March, 2006, URL http://pages.stern.nyu.edu/~lcabral/workingpapers/Cabral Hortacsu_Mar06.pdf, 2006.

[48] J. Carbonell, *Subjective Understanding: Computer Models of Belief Systems*. PhD thesis, Yale, 1979.

[49] C. Cardie, "Empirical methods in information extraction," *AI Magazine*, vol. 18, pp. 65–79, 1997.

[50] C. Cardie, C. Farina, T. Bruce, and E. Wagner, "Using natural language processing to improve eRulemaking," in *Proceedings of Digital Government Research (dg.o)*, 2006.

[51] C. Cardie, J. Wiebe, T. Wilson, and D. Litman, "Combining low-level and summary representations of opinions for multi-perspective question answering," in *Proceedings of the AAAI Spring Symposium on New Directions in Question Answering*, pp. 20–27, 2003.

[52] G. Carenini, R. Ng, and A. Pauls, "Multi-document summarization of evaluative text," in *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL)*, pp. 305–312, 2006.

[53] G. Carenini, R. T. Ng, and A. Pauls, "Interactive multimedia summaries of evaluative text," in *Proceedings of Intelligent User Interfaces (IUI)*, pp. 124–131, ACM Press, 2006.

[54] D. Cartwright and F. Harary, "Structural balance: A generalization of Heider's theory," *Psychological Review*, vol. 63, pp. 277–293, 1956.

[55] P. Chaovalit and L. Zhou, "Movie review mining: A comparison between supervised and unsupervised classification approaches," in *Proceedings of the Hawaii International Conference on System Sciences (HICSS)*, 2005.

[56] P.-Y. S. Chen, S.-Y. Wu, and J. Yoon, "The impact of online recommendations and consumer feedback on sales," in *International Conference on Information Systems (ICIS)*, pp. 711–724, 2004.

[57] Y. Chen and J. Xie, "Online consumer review: Word-of-mouth as a new element of marketing communication mix," *Management Science*, vol. 54, pp. 477–491, 2008.

[58] P. Chesley, B. Vincent, L. Xu, and R. Srihari, "Using verbs and adjectives to automatically classify blog sentiment," in *AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW)*, pp. 27–29, 2006.

[59] J. A. Chevalier and D. Mayzlin, "The effect of word of mouth on sales: Online book reviews," *Journal of Marketing Research*, vol. 43, pp. 345–354, August 2006.

[60] Y. Choi, E. Breck, and C. Cardie, "Joint extraction of entities and relations for opinion recognition," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2006.

[61] Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan, "Identifying sources of opinions with conditional random fields and extraction patterns," in *Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, 2005.

[62] E. K. Clemons, G. Gao, and L. M. Hitt, "When online reviews meet hyperdifferentiation: A study of the craft beer industry," *Journal of Management Information Systems*, vol. 23, pp. 149–171, 2006.

[63] comScore/the Kelsey group, "Online consumer-generated reviews have significant impact on offline purchase behavior," Press Release, http://www.comscore.com/press/release.asp?press=1928, November 2007.

[64] J. G. Conrad and F. Schilder, "Opinion mining in legal blogs," in *Proceedings of the International Conference on Artificial Intelligence and Law (ICAIL)*, pp. 231–236, New York, NY, USA: ACM, 2007.

[65] W. B. Croft and J. Lafferty, eds., *Language modeling for information retrieval*. Number 13 in the Information Retrieval Series. Kluwer/Springer, 2003.

[66] S. Das and M. Chen, "Yahoo! for Amazon: Extracting market sentiment from stock message boards," in *Proceedings of the Asia Pacific Finance Association Annual Conference (APFA)*, 2001.

[67] S. R. Das and M. Y. Chen, "Yahoo! for Amazon: Sentiment extraction from small talk on the Web," *Management Science*, vol. 53, pp. 1375–1388, 2007.

[68] S. R. Das, P. Tufano, and F. de Asis Martinez-Jerez, "eInformation: A clinical study of investor discussion and sentiment," *Financial Management*, vol. 34, pp. 103–137, 2005.

[69] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews," in *Proceedings of WWW*, pp. 519–528, 2003.

[70] S. David and T. J. Pinch, "Six degrees of reputation: The use and abuse of online review and recommendation systems," *First Monday*, July 2006. (Special Issue on Commercial Applications of the Internet).

[71] C. Dellarocas, "The digitization of word-of-mouth: Promise and challenges of online reputation systems," *Management Science*, vol. 49, pp. 1407–1424, 2003. (Special issue on e-business and management science).

[72] C. Dellarocas, X. Zhang, and N. F. Awad, "Exploring the value of online product ratings in revenue forecasting: The case of motion pictures," *Journal of Interactive Marketing*, vol. 21, pp. 23–45, 2007.

[73] A. Devitt and K. Ahmad, "Sentiment analysis in financial news: A cohesion-based approach," in *Proceedings of the Association for Computational Linguistics (ACL)*, pp. 984–991, 2007.

[74] M. Dewally, "Internet investment advice: Investing with a rock of salt," *Financial Analysts Journal*, vol. 59, pp. 65–77, July/August 2003.

[75] M. Dewally and L. Ederington, "Reputation, certification, warranties, and information as remedies for seller-buyer information asymmetries: Lessons from the online comic book market," *Journal of Business*, vol. 79, pp. 693–730, March 2006.

[76] S. Dewan and V. Hsu, "Adverse selection in electronic markets: Evidence from online stamp auctions," *Journal of Industrial Economics*, vol. 52, pp. 497–516, December 2004.

[77] D. W. Diamond, "Reputation acquisition in debt markets," *Journal of Political Economy*, vol. 97, pp. 828–862, 1989.

[78] X. Ding, B. Liu, and P. S. Yu, "A holistic lexicon-based approach to opinion mining," in *Proceedings of the Conference on Web Search and Web Data Mining (WSDM)*, 2008.

[79] L. Dini and G. Mazzini, "Opinion classification through information extraction," in *Proceedings of the Conference on Data Mining Methods and Databases for Engineering, Finance and Other Fields (Data Mining)*, pp. 299–310, 2002.

[80] W. Duan, B. Gu, and A. B. Whinston, "Do online reviews matter? — An empirical investigation of panel data," Social Science Research Network (SSRN) Working Paper Series, http://ssrn.com/paper=616262, version as of January, 2005.

[81] D. H. Eaton, "Valuing information: Evidence from guitar auctions on eBay," *Journal of Applied Economics and Policy*, vol. 24, pp. 1–19, 2005.

[82] D. H. Eaton, "The impact of reputation timing and source on auction outcomes," *The B. E. Journal of Economic Analysis and Policy*, vol. 7, 2007.

[83] M. Efron, "Cultural orientation: Classifying subjective documents by cociation [sic] analysis," in *Proceedings of the AAAI Fall Symposium on Style and Meaning in Language, Art, Music, and Design*, pp. 41–48, 2004.

[84] K. Eguchi and V. Lavrenko, "Sentiment retrieval using generative models," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 345–354, 2006.

[85] K. Eguchi and C. Shah, "Opinion retrieval experiments using generative models: Experiments for the TREC 2006 blog track," in *Proceedings of TREC*, 2006.

[86] P. Ekman, *Emotion in the Human Face*. Cambridge University Press, Second ed., 1982.

[87] J. Eliashberg and S. M. Shugan, "Film critics: Influencers or predictors?," *Journal of Marketing*, vol. 61, pp. 68–78, April 1997.

[88] C. Engström, *Topic Dependence in Sentiment Classification*. Master's thesis, University of Cambridge, 2004.

[89] A. Esuli and F. Sebastiani, "Determining the semantic orientation of terms through gloss analysis," in *Proceedings of the ACM SIGIR Conference on Information and Knowledge Management (CIKM)*, 2005.

[90] A. Esuli and F. Sebastiani, "Determining term subjectivity and term orientation for opinion mining," in *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL)*, 2006.

[91] A. Esuli and F. Sebastiani, "SentiWordNet: A publicly available lexical resource for opinion mining," in *Proceedings of Language Resources and Evaluation (LREC)*, 2006.

[92] A. Esuli and F. Sebastiani, "PageRanking WordNet synsets: An application to opinion mining," in *Proceedings of the Association for Computational Linguistics (ACL)*, 2007.

[93] D. K. Evans, L.-W. Ku, Y. Seki, H.-H. Chen, and N. Kando, "Opinion analysis across languages: An overview of and observations from the NTCIR6 opinion analysis pilot task," in *Proceedings of the Workshop on Cross-Language Information Processing,* vol. 4578 (Applications of Fuzzy Sets Theory) *of Lecture Notes in Computer Science*, pp. 456–463, 2007.

[94] A. Fader, D. R. Radev, M. H. Crespin, B. L. Monroe, K. M. Quinn, and M. Colaresi, "MavenRank: Identifying influential members of the US senate using lexical centrality," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2007.

[95] C. Fellbaum, ed., *Wordnet: An Electronic Lexical Database*. MIT Press, 1998.

[96] D. Feng, E. Shaw, J. Kim, and E. Hovy, "Learning to detect conversation focus of threaded discussions," in *Proceedings of the Joint Human Language Technology/North American Chapter of the ACL Conference (HLT-NAACL)*, pp. 208–215, 2006.

[97] A. Finn and N. Kushmerick, "Learning to classify documents according to genre," *Journal of the American Society for Information Science and Technology (JASIST)*, vol. 7, 2006. (Special issue on computational analysis of style).

[98] A. Finn, N. Kushmerick, and B. Smyth, "Genre classification and domain transfer for information filtering," in *Proceedings of the 24th BCS-IRSG European Colloquium on IR Research: Advances in Information Retrieval,* number 2291 in *Lecture Notes in Computer Science*, pp. 353–362, Glasgow, 2002.

[99] P. W. Foltz, D. Laham, and T. K. Landauer, "Automated essay scoring: Applications to education technology," in *Proceedings of ED-MEDIA*, pp. 939–944, 1999.

[100] C. Forman, A. Ghose, and B. Wiesenfeld, "Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets," *Information Systems Research*, vol. 19, 2008. (Special issue on the interplay between digital and social networks).

[101] G. Forman, "An extensive empirical study of feature selection metrics for text classification," *Journal of Machine Learning Research*, vol. 3, pp. 1289–1305, 2003.

[102] T. Fukuhara, H. Nakagawa, and T. Nishida, "Understanding sentiment of people from news articles: Temporal sentiment analysis of social events," in *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*, 2007.

[103] M. Gamon, "Sentiment classification on customer feedback data: Noisy data, large feature vectors, and the role of linguistic analysis," in *Proceedings of the International Conference on Computational Linguistics (COLING)*, 2004.

[104] M. Gamon, A. Aue, S. Corston-Oliver, and E. Ringger, "Pulse: Mining customer opinions from free text," in *Proceedings of the International Symposium on Intelligent Data Analysis (IDA),* number 3646 in *Lecture Notes in Computer Science*, pp. 121–132, 2005.

[105] R. Ghani, K. Probst, Y. Liu, M. Krema, and A. Fano, "Text mining for product attribute extraction," *SIGKDD Explorations Newsletter*, vol. 8, pp. 41–48, 2006.

[106] A. Ghose and P. G. Ipeirotis, "Designing novel review ranking systems: Predicting usefulness and impact of reviews," in *Proceedings of the International Conference on Electronic Commerce (ICEC)*, 2007. (Invited paper).

[107] A. Ghose, P. G. Ipeirotis, and A. Sundararajan, "Opinion mining using econometrics: A case study on reputation systems," in *Proceedings of the Association for Computational Linguistics (ACL)*, 2007.

[108] N. Godbole, M. Srinivasaiah, and S. Skiena, "Large-scale sentiment analysis for news and blogs," in *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*, 2007.

[109] A. B. Goldberg and X. Zhu, "Seeing stars when there aren't many stars: Graph-based semi-supervised learning for sentiment categorization," in *TextGraphs: HLT/NAACL Workshop on Graph-based Algorithms for Natural Language Processing*, 2006.

[110] A. B. Goldberg, X. Zhu, and S. Wright, "Dissimilarity in graph-based semi-supervised classification," in *Artificial Intelligence and Statistics (AISTATS)*, 2007.

[111] S. Greene, *Spin: Lexical Semantics, Transitivity, and the Identification of Implicit Sentiment*. PhD thesis, University of Maryland, 2007.

[112] G. Grefenstette, Y. Qu, J. G. Shanahan, and D. A. Evans, "Coupling niche browsers and affect analysis for an opinion mining application," in *Proceedings of Recherche d'Information Assistée par Ordinateur (RIAO)*, 2004.

[113] M. L. Gregory, N. Chinchor, P. Whitney, R. Carter, E. Hetzler, and A. Turner, "User-directed sentiment analysis: Visualizing the affective content of documents," in *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, pp. 23–30, Sydney, Australia, July 2006.

[114] B. Gu, P. Konana, A. Liu, B. Rajagopalan, and J. Ghosh, "Predictive value of stock message board sentiments," McCombs Research Paper No. IROM-11-06, version dated November, 2006.

[115] R. V. Guha, R. Kumar, P. Raghavan, and A. Tomkins, "Propagation of trust and distrust," in *Proceedings of WWW*, pp. 403–412, 2004.

[116] B. A. Hagedorn, M. Ciaramita, and J. Atserias, "World knowledge in broad-coverage information filtering," in *Proceedings of the ACM Special Interest Group on Information Retrieval (SIGIR)*, 2007. (Poster paper).

[117] J. T. Hancock, L. Curry, S. Goorha, and M. Woodworth, "Automated linguistic analysis of deceptive and truthful synchronous computer-mediated communication," in *Proceedings of the Hawaii International Conference on System Sciences (HICSS)*, p. 22c, 2005.

[118] L. Hankin, "The effects of user reviews on online purchasing behavior across multiple product categories," Master's final project report, UC Berkeley School of Information, http://www.ischool.berkeley.edu/files/lhankin_report.pdf, May 2007.

[119] V. Hatzivassiloglou and K. McKeown, "Predicting the semantic orientation of adjectives," in *Proceedings of the Joint ACL/EACL Conference*, pp. 174–181, 1997.

[120] V. Hatzivassiloglou and J. Wiebe, "Effects of adjective orientation and gradability on sentence subjectivity," in *Proceedings of the International Conference on Computational Linguistics (COLING)*, 2000.

[121] M. Hearst, "Direction-based text interpretation as an information access refinement," in *Text-Based Intelligent Systems*, (P. Jacobs, ed.), pp. 257–274, Lawrence Erlbaum Associates, 1992.

[122] R. Higashinaka, M. Walker, and R. Prasad, "Learning to generate naturalistic utterances using reviews in spoken dialogue systems," *ACM Transactions on Speech and Language Processing (TSLP)*, 2007.

[123] P. Hitlin and L. Rainie, "The use of online reputation and rating systems," Pew Internet & American Life Project Memo, October 2004.

[124] T. Hoffman, "Online reputation management is hot — but is it ethical?" Computerworld, February 2008.

[125] T. Hofmann, "Probabilistic latent semantic indexing," in *Proceedings of SIGIR*, pp. 50–57, 1999.

[126] D. Hopkins and G. King, "Extracting systematic social science meaning from text,". Manuscript available at http://gking.harvard.edu/files/words.pdf, 2007 version was the one most recently consulted, 2007.

[127] J. A. Horrigan, "Online shopping," Pew Internet & American Life Project Report, 2008.

[128] D. Houser and J. Wooders, "Reputation in auctions: Theory, and evidence from eBay," *Journal of Economics and Management Strategy*, vol. 15, pp. 252–369, 2006.

[129] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 168–177, 2004.

[130] M. Hu and B. Liu, "Mining opinion features in customer reviews," in *Proceedings of AAAI*, pp. 755–760, 2004.

[131] M. Hu, A. Sun, and E.-P. Lim, "Comments-oriented blog summarization by sentence extraction," in *Proceedings of the ACM SIGIR Conference on*

Information and Knowledge Management (CIKM), pp. 901–904, 2007. (Poster paper).

[132] N. Hu, P. A. Pavlou, and J. Zhang, "Can online reviews reveal a product's true quality?: Empirical findings and analytical modeling of online word-of-mouth communication," in *Proceedings of Electronic Commerce (EC)*, pp. 324–330, USA, New York, NY: ACM, 2006.

[133] A. Huettner and P. Subasic, "Fuzzy typing for document management," in *ACL 2000 Companion Volume: Tutorial Abstracts and Demonstration Notes*, pp. 26–27, 2000.

[134] M. Hurst and K. Nigam, "Retrieving topical sentiments from online document collections," in *Document Recognition and Retrieval XI*, pp. 27–34, 2004.

[135] C. Jacquemin, *Spotting and Discovering Terms through Natural Language Processing*. MIT Press, 2001.

[136] G. Jin and A. Kato, "Price, quality and reputation: Evidence from an online field experiment," *The RAND Journal of Economics*, vol. 37, 2006.

[137] X. Jin, Y. Li, T. Mah, and J. Tong, "Sensitive webpage classification for content advertising," in *Proceedings of the International Workshop on Data Mining and Audience Intelligence for Advertising*, 2007.

[138] N. Jindal and B. Liu, "Identifying comparative sentences in text documents," in *Proceedings of the ACM Special Interest Group on Information Retrieval (SIGIR)*, 2006.

[139] N. Jindal and B. Liu, "Mining comparative sentences and relations," in *Proceedings of AAAI*, 2006.

[140] N. Jindal and B. Liu, "Review spam detection," in *Proceedings of WWW*, 2007. (Poster paper).

[141] N. Jindal and B. Liu, "Opinion spam and analysis," in *Proceedings of the Conference on Web Search and Web Data Mining (WSDM)*, pp. 219–230, 2008.

[142] N. Kaji and M. Kitsuregawa, "Automatic construction of polarity-tagged corpus from HTML documents," in *Proceedings of the COLING/ACL Main Conference Poster Sessions*, 2006.

[143] N. Kaji and M. Kitsuregawa, "Building lexicon for sentiment analysis from massive collection of HTML documents," in *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pp. 1075–1083, 2007.

[144] A. Kale, A. Karandikar, P. Kolari, A. Java, T. Finin, and A. Joshi, "Modeling trust and influence in the blogosphere using link polarity," in *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*, 2007. (Short paper).

[145] K. Kalyanam and S. H. McIntyre, "The role of reputation in online auction markets," Santa Clara University Working Paper 02/03-10-WP, 2001, dated June 26.

[146] J. Kamps, M. Marx, R. J. Mokken, and M. de Rijke, "Using WordNet to measure semantic orientation of adjectives," in *Proceedings of LREC*, 2004.

[147] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina, "The Eigentrust algorithm for reputation management in P2P networks," in *Proceedings of WWW*, pp. 640–651, New York, NY, USA: ACM, ISBN 1-58113-680-3, 2003.

[148] H. Kanayama and T. Nasukawa, "Fully automatic lexicon expansion for domain-oriented sentiment analysis," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, (Sydney, Australia), pp. 355–363, July 2006.

[149] M. Kantrowitz, "Method and apparatus for analyzing affect and emotion in text," U.S. Patent 6622140, Patent filed in November 2000, 2003.

[150] J. Karlgren and D. Cutting, "Recognizing text genres with simple metrics using discriminant analysis," in *Proceedings of COLING*, pp. 1071–1075, 1994.

[151] Y. Kawai, T. Kumamoto, and K. Tanaka, "Fair news reader: Recommending news articles with different sentiments based on user preference," in *Proceedings of Knowledge-Based Intelligent Information and Engineering Systems (KES),* number 4692 in *Lecture Notes in Computer Science*, pp. 612–622, 2007.

[152] A. Kennedy and D. Inkpen, "Sentiment classification of movie reviews using contextual valence shifters," *Computational Intelligence*, vol. 22, pp. 110–125, 2006.

[153] B. Kessler, G. Nunberg, and H. Schütze, "Automatic detection of text genre," in *Proceedings of the Thirty-Fifth Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, pp. 32–38, 1997.

[154] P. Kim, "The forrester wave: Brand monitoring, Q3 2006," Forrester Wave (white paper), 2006.

[155] S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," in *Proceedings of the International Conference on Computational Linguistics (COLING)*, 2004.

[156] S.-M. Kim and E. Hovy, "Automatic detection of opinion bearing words and sentences," in *Companion Volume to the Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP)*, 2005.

[157] S.-M. Kim and E. Hovy, "Identifying opinion holders for question answering in opinion texts," in *Proceedings of the AAAI Workshop on Question Answering in Restricted Domains*, 2005.

[158] S.-M. Kim and E. Hovy, "Automatic identification of pro and con reasons in online reviews," in *Proceedings of the COLING/ACL Main Conference Poster Sessions*, pp. 483–490, 2006.

[159] S.-M. Kim and E. Hovy, "Identifying and analyzing judgment opinions," in *Proceedings of the Joint Human Language Technology/North American Chapter of the ACL Conference (HLT-NAACL)*, 2006.

[160] S.-M. Kim and E. Hovy, "Crystal: Analyzing predictive opinions on the web," in *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, 2007.

[161] S.-M. Kim, P. Pantel, T. Chklovski, and M. Pennacchiotti, "Automatically assessing review helpfulness," in *Proceedings of the Conference on Empirical*

Methods in Natural Language Processing (EMNLP), pp. 423–430, Sydney, Australia, July 2006.

[162] B. Klein and K. Leffler, "The role of market forces in assuring contractual performance," *Journal of Political Economy*, vol. 89, pp. 615–641, 1981.

[163] J. Kleinberg, "Authoritative sources in a hyperlinked environment," in *Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 668–677, 1998. (Extended version in *Journal of the ACM*, 46:604–632, 1999).

[164] J. Kleinberg and E. Tardos, "Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields," *Journal of the ACM*, vol. 49, pp. 616–639, ISSN 0004-5411, 2002.

[165] J. Kleinberg and E. Tardos, *Algorithm Design*. Addison Wesley, 2006.

[166] N. Kobayashi, K. Inui, Y. Matsumoto, K. Tateishi, and T. Fukushima, "Collecting evaluative expressions for opinion extraction," in *Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP)*, 2004.

[167] M. Koppel and J. Schler, "The importance of neutral examples for learning sentiment," in *Workshop on the Analysis of Informal and Formal Information Exchange During Negotiations (FINEXIN)*, 2005.

[168] M. Koppel and I. Shtrimberg, "Good news or bad news? Let the market decide," in *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, pp. 86–88, 2004.

[169] L.-W. Ku, L.-Y. Li, T.-H. Wu, and H.-H. Chen, "Major topic detection and its application to opinion summarization," in *Proceedings of the ACM Special Interest Group on Information Retrieval (SIGIR)*, pp. 627–628, 2005. (Poster paper).

[170] L.-W. Ku, Y.-T. Liang, and H.-H. Chen, "Opinion extraction, summarization and tracking in news and blog corpora," in *AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW)*, pp. 100–107, 2006.

[171] L.-W. Ku, Y.-T. Liang, and H.-H. Chen, "Tagging heterogeneous evaluation corpora for opinionated tasks," in *Conference on Language Resources and Evaluation (LREC)*, 2006.

[172] L.-W. Ku, Y.-S. Lo, and H.-H. Chen, "Test collection selection and gold standard generation for a multiply-annotated opinion corpus," in *Proceedings of the ACL Demo and Poster Sessions*, pp. 89–92, 2007.

[173] T. Kudo and Y. Matsumoto, "A boosting algorithm for classification of semi-structured text," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2004.

[174] S. Kurohashi, K. Inui, and Y. Kato, eds., *Workshop on Information Credibility on the Web*, 2007.

[175] N. Kwon, S. Shulman, and E. Hovy, "Multidimensional text analysis for eRulemaking," in *Proceedings of Digital Government Research (dg.o)*, 2006.

[176] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proceedings of ICML*, pp. 282–289, 2001.

[177] J. D. Lafferty and C. Zhai, "Document language models, query models, and risk minimization for information retrieval," in *Proceedings of SIGIR*, pp. 111–119, 2001.

[178] M. Laver, K. Benoit, and J. Garry, "Extracting policy positions from political texts using words as data," *American Political Science Review*, vol. 97, pp. 311–331, 2003.

[179] V. Lavrenko and W. Bruce Croft, "Relevance-based language models," in *Proceedings of SIGIR*, pp. 120–127, 2001.

[180] C. G. Lawson and V. C. Slawson, "Reputation in an internet auction market," *Economic Inquiry*, vol. 40, pp. 533–650, 2002.

[181] L. Lee, ""I'm sorry Dave, I'm afraid I can't do that": Linguistics, statistics, and natural language processing circa 2001," in *Computer Science: Reflections on the Field, Reflections from the Field*, (Committee on the Fundamentals of Computer Science: Challenges and Opportunities, Computer Science and Telecommunications Board, National Research Council, ed.), pp. 111–118, The National Academies Press, 2004.

[182] Y.-B. Lee and S. H. Myaeng, "Text genre classification with genre-revealing and subject-revealing features," in *Proceedings of the ACM Special Interest Group on Information Retrieval (SIGIR)*, 2002.

[183] D. Leinweber and A. Madhavan, "Three hundred years of stock market manipulation," *Journal of Investing*, vol. 10, pp. 7–16, Summer 2001.

[184] H. Li and K. Yamanishi, "Mining from open answers in questionnaire data," in *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 443–449, 2001. (Journal version in *IEEE Intelligent Systems* vol. 17, no. 5, pp. 58–63, 2002).

[185] Y. Li, Z. Zheng, and H. Dai, "KDD CUP-2005 report: Facing a great challenge," *SIGKDD Explorations*, vol. 7, pp. 91–99, 2005.

[186] W.-H. Lin and A. Hauptmann, "Are these documents written from different perspectives? A test of different perspectives based on statistical distribution divergence," in *Proceedings of the International Conference on Computational Linguistics (COLING)/Proceedings of the Association for Computational Linguistics (ACL)*, pp. 1057–1064, Sydney, Australia: Association for Computational Linguistics, July 2006.

[187] W.-H. Lin, T. Wilson, J. Wiebe, and A. Hauptmann, "Which side are you on? Identifying perspectives at the document and sentence levels," in *Proceedings of the Conference on Natural Language Learning (CoNLL)*, 2006.

[188] J. Liscombe, G. Riccardi, and D. Hakkani-Tür, "Using context to improve emotion detection in spoken dialog systems," in *Interspeech*, pp. 1845–1848, 2005.

[189] L. V. Lita, A. H. Schlaikjer, W. Hong, and E. Nyberg, "Qualitative dimensions in question answering: Extending the definitional QA task," in *Proceedings of AAAI*, pp. 1616–1617, 2005. (Student abstract).

[190] B. Liu, "Web data mining; Exploring hyperlinks, contents, and usage data," *Opinion Mining*. Springer, 2006.

[191] B. Liu, M. Hu, and J. Cheng, "Opinion observer: Analyzing and comparing opinions on the web," in *Proceedings of WWW*, 2005.

[192] H. Liu, H. Lieberman, and T. Selker, "A model of textual affect sensing using real-world knowledge," in *Proceedings of Intelligent User Interfaces (IUI)*, pp. 125–132, 2003.

[193] J. Liu, Y. Cao, C.-Y. Lin, Y. Huang, and M. Zhou, "Low-quality product review detection in opinion summarization," in *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pp. 334–342, 2007. (Poster paper).

[194] Y. Liu, "Word-of-mouth for movies: Its dynamics and impact on box office revenue," *Journal of Marketing*, vol. 70, pp. 74–89, 2006.

[195] Y. Liu, J. Huang, A. An, and X. Yu, "ARSA: A sentiment-aware model for predicting sales performance using blogs," in *Proceedings of the ACM Special Interest Group on Information Retrieval (SIGIR)*, 2007.

[196] J. A. Livingston, "How valuable is a good reputation? A sample selection model of internet auctions," *The Review of Economics and Statistics*, vol. 87, pp. 453–465, August 2005.

[197] L. Lloyd, D. Kechagias, and S. Skiena, "Lydia: A system for large-scale news analysis," in *Proceedings of String Processing and Information Retrieval (SPIRE),* number 3772 in *Lecture Notes in Computer Science*, pp. 161–166, 2005.

[198] D. Lucking-Reiley, D. Bryan, N. Prasad, and D. Reeves, "Pennies from eBay: The determinants of price in online auctions," *Journal of Industrial Economics*, vol. 55, pp. 223–233, 2007.

[199] C. Macdonald and I. Ounis, "The TREC Blogs06 collection: Creating and analysing a blog test collection," Technical Report TR-2006-224, Department of Computer Science, University of Glasgow, 2006.

[200] Y. Mao and G. Lebanon, "Sequential models for sentiment prediction," in *ICML Workshop on Learning in Structured Output Spaces*, 2006.

[201] Y. Mao and G. Lebanon, "Isotonic conditional random fields and local sentiment flow," in *Advances in Neural Information Processing Systems*, 2007.

[202] L. W. Martin and G. Vanberg, "A robust transformation procedure for interpreting political text," *Political Analysis*, vol. 16, pp. 93–100, 2008.

[203] H. Masum and Y.-C. Zhang, "Manifesto for the reputation society," *First Monday*, vol. 9, 2004.

[204] S. Matsumoto, H. Takamura, and M. Okumura, "Sentiment classification using word sub-sequences and dependency sub-trees," in *Proceedings of PAKDD'05, the 9th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*, 2005.

[205] R. McDonald, K. Hannan, T. Neylon, M. Wells, and J. Reynar, "Structured models for fine-to-coarse sentiment analysis," in *Proceedings of the Association for Computational Linguistics (ACL)*, pp. 432–439, Prague, Czech Republic: Association for Computational Linguistics, June 2007.

[206] Q. Mei, X. Ling, M. Wondra, H. Su, and C. X. Zhai, "Topic sentiment mixture: Modeling facets and opinions in weblogs," in *Proceedings of WWW*, pp. 171–180, New York, NY, USA: ACM Press, 2007. (ISBN 978-1-59593-654-7).

[207] M. I. Melnik and J. Alm, "Does a seller's eCommerce reputation matter? Evidence from eBay auctions," *Journal of Industrial Economics*, vol. 50, pp. 337–349, 2002.

[208] M. I. Melnik and J. Alm, "Seller reputation, information signals, and prices for heterogeneous coins on eBay," *Southern Economic Journal*, vol. 72, pp. 305–328, 2005.

[209] R. Mihalcea, C. Banea, and J. Wiebe, "Learning multilingual subjective language via cross-lingual projections," in *Proceedings of the Association for Computational Linguistics (ACL)*, pp. 976–983, Prague, Czech Republic, June 2007.

[210] R. Mihalcea and C. Strapparava, "Learning to laugh (automatically): Computational models for humor recognition," *Journal of Computational Intelligence*, 2006.

[211] G. Mishne and M. de Rijke, "Capturing global mood levels using blog posts," in *AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW)*, pp. 145–152, 2006.

[212] G. Mishne and M. de Rijke, "Moodviews: Tools for blog mood analysis," in *AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW)*, pp. 153–154, 2006.

[213] G. Mishne and M. de Rijke, "A study of blog search," in *Proceedings of the European Conference on Information Retrieval Research (ECIR)*, 2006.

[214] G. Mishne and N. Glance, "Predicting movie sales from blogger sentiment," in *AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW)*, pp. 155–158, 2006.

[215] S. Morinaga, K. Yamanishi, K. Tateishi, and T. Fukushima, "Mining product reputations on the Web," in *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 341–349, 2002. (Industry track).

[216] F. Mosteller and D. L. Wallace, *Applied Bayesian and Classical Inference: The Case of the Federalist Papers*. Springer-Verlag, 1984.

[217] T. Mullen and N. Collier, "Sentiment analysis using support vector machines with diverse information sources," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 412–418, July 2004. (Poster paper).

[218] T. Mullen and R. Malouf, "Taking sides: User classification for informal online political discourse," *Internet Research*, vol. 18, pp. 177–190, 2008.

[219] T. Mullen and R. Malouf, "A preliminary investigation into sentiment analysis of informal political discourse," in *AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW)*, pp. 159–162, 2006.

[220] J.-C. Na, H. Sui, C. Khoo, S. Chan, and Y. Zhou, "Effectiveness of simple linguistic processing in automatic sentiment classification of product reviews," in *Conference of the International Society for Knowledge Organization (ISKO)*, pp. 49–54, 2004.

[221] T. Nasukawa and J. Yi, "Sentiment analysis: Capturing favorability using natural language processing," in *Proceedings of the Conference on Knowledge Capture (K-CAP)*, 2003.

[222] V. Ng, S. Dasgupta, and S. M. N. Arifin, "Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews," in *Proceedings of the COLING/ACL Main Conference Poster Sessions*, pp. 611–618, Sydney, Australia: Association for Computational Linguistics, July 2006.

[223] X. Ni, G.-R. Xue, X. Ling, Y. Yu, and Q. Yang, "Exploring in the weblog space by detecting informative and affective articles," in *Proceedings of WWW*, 2007. (Industrial practice and experience track).

[224] N. Nicolov, F. Salvetti, M. Liberman, and J. H. Martin, eds., *AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW)*. AAAI Press, 2006.

[225] K. Nigam and M. Hurst, "Towards a robust metric of polarity," in *Computing Attitude and Affect in Text: Theories and Applications,* number 20 in *The Information Retrieval Series*, (J. G. Shanahan, Y. Qu, and J. Wiebe, eds.), 2006.

[226] Y. Niu, X. Zhu, J. Li, and G. Hirst, "Analysis of polarity information in medical text," in *Proceedings of the American Medical Informatics Association 2005 Annual Symposium*, 2005.

[227] I. Ounis, M. de Rijke, C. Macdonald, G. Mishne, and I. Soboroff, "Overview of the TREC-2006 blog track," in *Proceedings of the 15th Text Retrieval Conference (TREC)*, 2006.

[228] I. Ounis, C. Macdonald, and I. Soboroff, "On the TREC blog track," in *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*, 2008.

[229] S. Owsley, S. Sood, and K. J. Hammond, "Domain specific affective classification of documents," in *AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW)*, pp. 181–183, 2006.

[230] M. Palmer, D. Gildea, and P. Kingsbury, "The proposition bank: A corpus annotated with semantic roles," *Computational Linguistics*, vol. 31, March 2005.

[231] B. Pang, K. Knight, and D. Marcu, "Syntax-based alignment of multiple translations: Extracting paraphrases and generating new sentences," in *Proceedings of HLT/NAACL*, 2003.

[232] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in *Proceedings of the Association for Computational Linguistics (ACL)*, pp. 271–278, 2004.

[233] B. Pang and L. Lee, "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales," in *Proceedings of the Association for Computational Linguistics (ACL)*, pp. 115–124, 2005.

[234] B. Pang and L. Lee, "Using very simple statistics for review search: An exploration," in *Proceedings of the International Conference on Computational Linguistics (COLING)*, 2008. (Poster paper).

[235] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 79–86, 2002.

[236] D.-H. Park, J. Lee, and I. Han, "The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement," *International Journal of Electronic Commerce*, vol. 11, pp. 125–148, (ISSN 1086-4415), 2007.

[237] P. A. Pavlou and A. Dimoka, "The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation," *Information Systems Research*, vol. 17, pp. 392–414, 2006.

[238] S. Piao, S. Ananiadou, Y. Tsuruoka, Y. Sasaki, and J. McNaught, "Mining opinion polarity relations of citations," in *International Workshop on Computational Semantics (IWCS)*, pp. 366–371, 2007. (Short paper).

[239] R. Picard, *Affective Computing*. MIT Press, 1997.

[240] T. Pinch and K. Athanasiades, "ACIDplanet: A study of users of an on-line music community," 2005. http://sts.nthu.edu.tw/sts_camp/files/ACIDplanet%20by%20Trevor%20Pinch.ppt, Presented at the 50th Society for Ethnomusicology (SEM) conference.

[241] G. Pinski and F. Narin, "Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics," *Information Processing and Management*, vol. 12, pp. 297–312, 1976.

[242] L. Polanyi and A. Zaenen, "Contextual lexical valence shifters," in *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text*, AAAI technical report SS-04-07, 2004.

[243] J. M. Ponte and W. Bruce Croft, "A language modeling approach to information retrieval," in *Proceedings of SIGIR*, pp. 275–281, 1998.

[244] A.-M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in *Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, 2005.

[245] R. Quirk, S. Greenbaum, G. Leech, and J. Svartvik, *A comprehensive grammar of the English language*. Longman, 1985.

[246] D. Radev, T. Allison, S. Blair-Goldensohn, J. Blitzer, A. Çelebi, S. Dimitrov, E. Drabek, A. Hakim, W. Lam, D. Liu, J. Otterbacher, H. Qi, H. Saggion, S. Teufel, M. Topper, A. Winkel, and Z. Zhang, "MEAD — A platform for multidocument multilingual text summarization," in *Conference on Language Resources and Evaluation (LREC)*, Lisbon, Portugal, May 2004.

[247] D. R. Radev, E. Hovy, and K. McKeown, "Introduction to the special issue on summarization," *Computational Linguistics*, vol. 28, pp. 399–408, (ISSN 0891-2017), 2002.

[248] L. Rainie and J. Horrigan, "Election 2006 online," Pew Internet & American Life Project Report, January 2007.

[249] J. Read, "Using emoticons to reduce dependency in machine learning techniques for sentiment classification," in *Proceedings of the ACL Student Research Workshop*, 2005.

[250] D. A. Reinstein and C. M. Snyder, "The influence of expert reviews on consumer demand for experience goods: A case study of movie critics," *Journal of Industrial Economics*, vol. 53, pp. 27–51, 2005.

[251] E. Reiter and R. Dale, *Building Natural Language Generation Systems*. Cambridge, 2000.

[252] P. Resnick, K. Kuwabara, R. Zeckhauser, and E. Friedman, "Reputation systems," *Communications of the Association for Computing Machinery (CACM)*, vol. 43, pp. 45–48, (ISSN 0001-0782), 2000.

[253] P. Resnick, R. Zeckhauser, J. Swanson, and K. Lockwood, "The value of reputation on eBay: A controlled experiment," *Experimental Economics*, vol. 9, pp. 79–101, 2006.

[254] E. Riloff, S. Patwardhan, and J. Wiebe, "Feature subsumption for opinion analysis," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2006.

[255] E. Riloff and J. Wiebe, "Learning extraction patterns for subjective expressions," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2003.

[256] E. Riloff, J. Wiebe, and W. Phillips, "Exploiting subjectivity classification to improve information extraction," in *Proceedings of AAAI*, pp. 1106–1111, 2005.

[257] E. Riloff, J. Wiebe, and T. Wilson, "Learning subjective nouns using extraction pattern bootstrapping," in *Proceedings of the Conference on Natural Language Learning (CoNLL)*, pp. 25–32, 2003.

[258] E. Rogers, *Diffusion of Innovations*. Free Press, New York, 1962. (ISBN 0743222091. Fifth edition dated 2003).

[259] S. Rosen, "Hedonic prices and implicit markets: Product differentiation in pure competition," *The Journal of Political Economy*, vol. 82, pp. 34–55, Jan–Feb 1974.

[260] D. Roth and W. Yih, "Probabilistic reasoning for entity and relation recognition," in *Proceedings of the International Conference on Computational Linguistics (COLING)*, 2004.

[261] V. L. Rubin and E. D. Liddy, "Assessing credibility of weblogs," in *AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW)*, pp. 187–190, 2006.

[262] W. Sack, "On the computation of point of view," in *Proceedings of AAAI*, p. 1488, 1994. (Student abstract).

[263] F. Sebastiani, "Machine learning in automated text categorization," *ACM Computing Surveys*, vol. 34, pp. 1–47, 2002.

[264] Y. Seki, K. Eguchi, and N. Kando, "Analysis of multi-document viewpoint summarization using multi-dimensional genres," in *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, pp. 142–145, 2004.

[265] Y. Seki, K. Eguchi, N. Kando, and M. Aono, "Multi-document summarization with subjectivity analysis at DUC 2005," in *Proceedings of the Document Understanding Conference (DUC)*, 2005.

[266] Y. Seki, K. Eguchi, N. Kando, and M. Aono, "Opinion-focused summarization and its analysis at DUC 2006," in *Proceedings of the Document Understanding Conference (DUC)*, pp. 122–130, 2006.

[267] Y. Seki, D. Kirk Evans, L.-W. Ku, H.-H. Chen, N. Kando, and C.-Y. Lin, "Overview of opinion analysis pilot task at NTCIR-6," in *Proceedings of the*

Workshop Meeting of the National Institute of Informatics (NII) Test Collection for Information Retrieval Systems (NTCIR), pp. 265–278, 2007.

[268] C. Shapiro, "Consumer information, product quality, and seller reputation," *Bell Journal of Economics*, vol. 13, pp. 20–35, 1982.

[269] C. Shapiro, "Premiums for high quality products as returns to reputations," *Quarterly Journal of Economics*, vol. 98, pp. 659–680, 1983.

[270] B. Shneiderman, "Tree visualization with tree-maps: 2-d space-filling approach," *ACM Transactions on Graphics*, vol. 11, pp. 92–99, 1992.

[271] S. Shulman, J. Callan, E. Hovy, and S. Zavestoski, "Language processing technologies for electronic rulemaking: A project highlight," in *Proceedings of Digital Government Research (dg.o)*, pp. 87–88, 2005.

[272] B. Snyder and R. Barzilay, "Multiple aspect ranking using the Good Grief algorithm," in *Proceedings of the Joint Human Language Technology/North American Chapter of the ACL Conference (HLT-NAACL)*, pp. 300–307, 2007.

[273] S. Somasundaran, J. Ruppenhofer, and J. Wiebe, "Detecting arguing and sentiment in meetings," in *Proceedings of the SIGdial Workshop on Discourse and Dialogue*, 2007.

[274] S. Somasundaran, T. Wilson, J. Wiebe, and V. Stoyanov, "QA with attitude: Exploiting opinion type analysis for improving question answering in on-line discussions and the news," in *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*, 2007.

[275] X. Song, Y. Chi, K. Hino, and B. Tseng, "Identifying opinion leaders in the blogosphere," in *Proceedings of the ACM SIGIR Conference on Information and Knowledge Management (CIKM)*, pp. 971–974, 2007.

[276] E. Spertus, "Smokey: Automatic recognition of hostile messages," in *Proceedings of Innovative Applications of Artificial Intelligence (IAAI)*, pp. 1058–1065, 1997.

[277] E. Stamatatos, N. Fakotakis, and G. Kokkinakis, "Text genre detection using common word frequencies," in *Proceedings of the International Conference on Computational Linguistics (COLING)*, 2000.

[278] S. S. Standifird, "Reputation and e-commerce: eBay auctions and the asymmetrical impact of positive and negative ratings," *Journal of Management*, vol. 27, pp. 279–295, 2001.

[279] A. Stepinski and V. Mittal, "A fact/opinion classifier for news articles," in *Proceedings of the ACM Special Interest Group on Information Retrieval (SIGIR)*, pp. 807–808, New York, NY, USA: ACM Press, 2007. (ISBN 978-1-59593-597-7).

[280] B. Stone and M. Richtel, "The hand that controls the sock puppet could get slapped," *The New York Times*, July 16 2007.

[281] P. J. Stone, *The General Inquirer: A Computer Approach to Content Analysis*. The MIT Press, 1966.

[282] V. Stoyanov and C. Cardie, "Partially supervised coreference resolution for opinion summarization through structured rule learning," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 336–344, Sydney, Australia: Association for Computational Linguistics, July 2006.

[283] V. Stoyanov, C. Cardie, D. Litman, and J. Wiebe, "Evaluating an opinion annotation scheme using a new multi-perspective question and answer corpus," in *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text*, AAAI Technical Report SS-04-07.

[284] V. Stoyanov, C. Cardie, and J. Wiebe, "Multi-perspective question answering using the OpQA corpus," in *Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, pp. 923–930, Vancouver, British Columbia, Canada: Association for Computational Linguistics, October 2005.

[285] P. Subasic and A. Huettner, "Affect analysis of text using fuzzy semantic typing," *IEEE Transactions on Fuzzy Systems*, vol. 9, pp. 483–496, 2001.

[286] M. Taboada, C. Anthony, and K. Voll, "Methods for creating semantic orientation dictionaries," in *Conference on Language Resources and Evaluation (LREC)*, pp. 427–432, 2006.

[287] M. Taboada, M. A. Gillies, and P. McFetridge, "Sentiment classification techniques for tracking literary reputation," in *LREC Workshop: Towards Computational Models of Literary Analysis*, pp. 36–43, 2006.

[288] H. Takamura, T. Inui, and M. Okumura, "Extracting semantic orientation of words using spin model," in *Proceedings of the Association for Computational Linguistics (ACL)*, pp. 133–140, 2005.

[289] H. Takamura, T. Inui, and M. Okumura, "Latent variable models for semantic orientations of phrases," in *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL)*, 2006.

[290] H. Takamura, T. Inui, and M. Okumura, "Extracting semantic orientations of phrases from dictionary," in *Proceedings of the Joint Human Language Technology/North American Chapter of the ACL Conference (HLT-NAACL)*, 2007.

[291] K. Tateishi, Y. Ishiguro, and T. Fukushima, "Opinion information retrieval from the internet," *Information Processing Society of Japan (IPSJ) SIG Notes,* 2001, vol. 69, no. 7, pp. 75–82, 2001. (Also cited as "A reputation search engine that gathers people's opinions from the Internet", IPSJ Technical Report NL-14411. In Japanese).

[292] J. Tatemura, "Virtual reviewers for collaborative exploration of movie reviews," in *Proceedings of Intelligent User Interfaces (IUI)*, pp. 272–275, 2000.

[293] L. Terveen, W. Hill, B. Amento, D. McDonald, and J. Creter, "PHOAKS: A system for sharing recommendations," *Communications of the Association for Computing Machinery (CACM)*, vol. 40, pp. 59–62, 1997.

[294] M. Thomas, B. Pang, and L. Lee, "Get out the vote: Determining support or opposition from congressional floor-debate transcripts," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 327–335, 2006.

[295] R. Tokuhisa and R. Terashima, "Relationship between utterances and 'enthusiasm' in non-task-oriented conversational dialogue," in *Proceedings of the SIGdial Workshop on Discourse and Dialogue*, pp. 161–167, Sydney, Australia: Association for Computational Linguistics, July 2006.

[296] R. M. Tong, "An operational system for detecting and tracking opinions in on-line discussion," in *Proceedings of the Workshop on Operational Text Classification (OTC)*, 2001.

[297] R. Tumarkin and R. F. Whitelaw, "News or noise? Internet postings and stock prices," *Financial Analysts Journal*, vol. 57, pp. 41–51, May/June 2001.

[298] P. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews," in *Proceedings of the Association for Computational Linguistics (ACL)*, pp. 417–424, 2002.

[299] P. D. Turney and M. L. Littman, "Measuring praise and criticism: Inference of semantic orientation from association," *ACM Transactions on Information Systems (TOIS)*, vol. 21, pp. 315–346, 2003.

[300] S. Wan and K. McKeown, "Generating overview summaries of ongoing email thread discussions," in *Proceedings of the International Conference on Computational Linguistics (COLING)*, pp. 549–555, Geneva, Switzerland, 2004.

[301] M. White, C. Cardie, and V. Ng, "Detecting discrepancies in numeric estimates using multidocument hypertext summaries," in *Proceedings of the Conference on Human Language Technology*, pp. 336–341, 2002.

[302] M. White, C. Cardie, V. Ng, K. Wagstaff, and D. McCullough, "Detecting discrepancies and improving intelligibility: Two preliminary evaluations of RIPTIDES," in *Proceedings of the Document Understanding Conference (DUC)*, 2001.

[303] C. Whitelaw, N. Garg, and S. Argamon, "Using appraisal groups for sentiment analysis," in *Proceedings of the ACM SIGIR Conference on Information and Knowledge Management (CIKM)*, pp. 625–631, ACM, 2005.

[304] J. Wiebe, "Learning subjective adjectives from corpora," in *Proceedings of AAAI*, 2000.

[305] J. Wiebe, E. Breck, C. Buckley, C. Cardie, P. Davis, B. Fraser, D. Litman, D. Pierce, E. Riloff, T. Wilson, D. Day, and M. Maybury, "Recognizing and organizing opinions expressed in the world press," in *Proceedings of the AAAI Spring Symposium on New Directions in Question Answering*, 2003.

[306] J. Wiebe and R. Bruce, "Probabilistic classifiers for tracking point of view," in *Proceedings of the AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*, pp. 181–187, 1995.

[307] J. Wiebe and R. Mihalcea, "Word sense and subjectivity," in *Proceedings of the Conference on Computational Linguistics / Association for Computational Linguistics (COLING/ACL)*, 2006.

[308] J. Wiebe and T. Wilson, "Learning to disambiguate potentially subjective expressions," in *Proceedings of the Conference on Natural Language Learning (CoNLL)*, pp. 112–118, 2002.

[309] J. Wiebe, T. Wilson, and C. Cardie, "Annotating expressions of opinions and emotions in language," *Language Resources and Evaluation (formerly Computers and the Humanities)*, vol. 39, pp. 164–210, 2005.

[310] J. M. Wiebe, "Identifying subjective characters in narrative," in *Proceedings of the International Conference on Computational Linguistics (COLING)*, pp. 401–408, 1990.

[311] J. M. Wiebe, "Tracking point of view in narrative," *Computational Linguistics*, vol. 20, pp. 233–287, 1994.

[312] J. M. Wiebe, R. F. Bruce, and T. P. O'Hara, "Development and use of a gold standard data set for subjectivity classifications," in *Proceedings of the Association for Computational Linguistics (ACL)*, pp. 246–253, 1999.

[313] J. M. Wiebe and W. J. Rapaport, "A computational theory of perspective and reference in narrative," in *Proceedings of the Association for Computational Linguistics (ACL)*, pp. 131–138, 1988.

[314] J. M. Wiebe and E. Riloff, "Creating subjective and objective sentence classifiers from unannotated texts," in *Proceedings of the Conference on Computational Linguistics and Intelligent Text Processing (CICLing),* number 3406 in *Lecture Notes in Computer Science*, pp. 486–497, 2005.

[315] J. M. Wiebe, T. Wilson, and M. Bell, "Identifying collocations for recognizing opinions," in *Proceedings of the ACL/EACL Workshop on Collocation: Computational Extraction, Analysis, and Exploitation*, 2001.

[316] J. M. Wiebe, T. Wilson, R. Bruce, M. Bell, and M. Martin, "Learning subjective language," *Computational Linguistics*, vol. 30, pp. 277–308, September 2004.

[317] Y. Wilks and J. Bien, "Beliefs, points of view and multiple environments," in *Proceedings of the international NATO symposium on artificial and human intelligence*, pp. 147–171, USA, New York, NY: Elsevier North-Holland, Inc., 1984.

[318] Y. Wilks and M. Stevenson, "The grammar of sense: Using part-of-speech tags as a first step in semantic disambiguation," *Journal of Natural Language Engineering*, vol. 4, pp. 135–144, 1998.

[319] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," in *Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, pp. 347–354, 2005.

[320] T. Wilson, J. Wiebe, and R. Hwa, "Just how mad are you? Finding strong and weak opinion clauses," in *Proceedings of AAAI*, pp. 761–769, 2004. (Extended version in *Computational Intelligence*, vol. 22, no. 2, pp. 73–99, 2006).

[321] H. Yang, L. Si, and J. Callan, "Knowledge transfer and opinion detection in the TREC2006 blog track," in *Proceedings of TREC*, 2006.

[322] K. Yang, N. Yu, A. Valerio, and H. Zhang, "WIDIT in TREC-2006 blog track," in *Proceedings of TREC*, 2006.

[323] J. Yi, T. Nasukawa, R. Bunescu, and W. Niblack, "Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques," in *Proceedings of the IEEE International Conference on Data Mining (ICDM)*, 2003.

[324] J. Yi and W. Niblack, "Sentiment mining in WebFountain," in *Proceedings of the International Conference on Data Engineering (ICDE)*, 2005.

[325] P.-L. Yin, "Information dispersion and auction prices," Social Science Research Network (SSRN) Working Paper Series, Version dated March 2005.

[326] H. Yu and V. Hatzivassiloglou, "Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences,"

in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2003.

[327] J. Zabin and A. Jefferies, "Social media monitoring and analysis: Generating consumer insights from online conversation," Aberdeen Group Benchmark Report, January 2008.

[328] Z. Zhang and B. Varadarajan, "Utility scoring of product reviews," in *Proceedings of the ACM SIGIR Conference on Information and Knowledge Management (CIKM)*, pp. 51–57, 2006.

[329] L. Zhou, J. K. Burgeon, and D. P. Twitchell, "A longitudinal analysis of language behavior of deception in e-mail," in *Proceedings of Intelligence and Security Informatics (ISI),* number 2665 in *Lecture Notes in Computer Science*, p. 959, 2008.

[330] L. Zhou and E. Hovy, "On the summarization of dynamically introduced information: Online discussions and blogs," in *AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW)*, pp. 237–242, 2006.

[331] F. Zhu and X. Zhang, "The influence of online consumer reviews on the demand for experience goods: The case of video games," in *International Conference on Information Systems (ICIS)*, 2006.

[332] L. Zhuang, F. Jing, X.-Y. Zhu, and L. Zhang, "Movie review mining and summarization," in *Proceedings of the ACM SIGIR Conference on Information and Knowledge Management (CIKM)*, 2006.