SICS Technical Report T2009:08
ISSN 1100-3154

A Literature Survey of Methods for Analysis of Subjective Language

Oscar Täckström
Swedish Institute of Computer Science
Box 1263, SE-164 29, Kista, Sweden
oscar@sics.se

April 16, 2009

Abstract

Subjective language is used to express attitudes and opinions towards things, ideas and people. While content and topic centred natural language processing is now part of everyday life, analysis of subjective aspects of natural language have until recently been largely neglected by the research community. The explosive growth of personal blogs, consumer opinion sites and social network applications in the last years, have however created increased interest in subjective language analysis. This paper provides an overview of recent research conducted in the area.

Keywords: Subjective language, attitude analysis, sentiment analysis, opinion analysis, survey.

1 Introduction

The analysis and processing of *subjective language*, as manifested for example in opinions, beliefs and judgements, is a growing area within the field of natural language processing. Though some interest in this area can be traced to the early 1990's and the work of Wiebe (1990), there has been a surge of interest in the field in the last five to ten years; presumably due to the increasing significance of informal information sources, such as blogs and wikis, and the booming growth of social network sites and people search facilities.

While one could argue that *topic*-level information retrieval, classification and clustering are more or less solved tasks, with results at a level of inter-annotator agreement, related tasks on the level of subjective language still remain largely unsolved. Though much insight has been gained on shallow tasks such as sentiment identification and classification or opinion extraction, the potential of analysis of the subjective aspects of language is to a large extent unexplored.

In this paper I provide a survey of use cases, methods and resources in the domain of subjective language analysis, focusing on written language. Analysis of subjective and emotional aspects specific to spoken language is a whole field of research of its own. The aim of this paper is to give a brief introduction to subjective language analysis, to give a comprehensible summary of recent research in the field and to pinpoint the main obstacles that should be addressed by future research.

1.1 Subjective Language

By *subjective language* we refer to aspects of language use related to the expression of *private states*, such as opinions, evaluations, emotions or speculations (Wiebe et al., 2004). A private state is categorised by an *attitude*, possibly having a *valence* of a certain *degree*, a *holder* and a *topic*.[1] Let the simple sentence "John likes apples a lot" serve as an example. Here an attitude, "liking", is held by a holder,

[1]There is little agreement on terminology in the literature on subjective language. Common notions used for what I have termed *attitude*, include *opinion*, *sentiment* and *affect*, with *sentiment* often encompassing aspects of both attitude and valence. The term *polarity* is often used for what I have termed *valence*.

”John”, towards a topic, ”apples”. ”Liking” further has a positive valence, with a degree indicated by ”a lot”. The aim of subjective language analysis is to be able to, with automated methods, uncover aspects such as these in free text.

Most research on subjective language has focused on attitude in isolation, or on attitude in combination with valence. Interest has commonly been limited to the identification of attitude, without any further distinction between different types of attitudes; and to classification of attitudinal valence into the categories of *positive*, *negative* and *neutral* (Mulder et al., 2004; Bai et al., 2005; Sahlgren et al., 2007). Thus, even when ignoring the directional aspects of holder and topic, most work has been rather coarse-grained in the characterisation of private states. Some recent notable exceptions are the work by Bethard et al. (2004), Choi et al. (2005), Kim and Hovy (2006a) and Kim and Hovy (2006b), in which holder and topic identification is also attempted.

Though most approaches to attitude and valence analysis have been coarse-grained, there has been some attempts on more fine-grained analysis. Liu et al. (2003) for example characterise attitude in terms of Ekman’s six fundamental categories of emotion: *happy, sad, angry, fearful, disgusted and surprised* (Ekman, 1993), while Subasic and Huettner (2001) use what they call ”affect sets”, which comprise a set of attitudinal categories with attached centralities and intensities. Other characterisations are presented in the work of Kim and Hovy (2006b), in which attitudes are characterised as *judgements* and *beliefs*, and in Thomas et al. (2006) and Kwon et al. (2007), wherein *claims* are identified and classified as to whether they are *supporting* or *opposing* an idea, or whether they are *proposing a new idea*, in the context of political discussions.

Before moving on it is worth pointing out the perhaps obvious fact that *subjectivity* in language is a notion orthogonal to the notion of *truth conditions*. A sentence expressed in subjective language might as well be *true*, while an objective sentence might as well be *false*. As Wiebe et al. (2004) puts it: ”[objective language] suggests that facts are being presented”. Thus we cannot hope to use analysis of subjective language to separate facts from fiction, since these are notions beyond the scope of language use.

1.2 Relation to Other Areas of NLP

The methods used for processing subjective aspects of natural language have to a large extent been carried over from traditional areas of natural language processing, such as information retrieval, topic-based text classification, clustering and information extraction. Further, more refined processing, such as dependency parsing and semantic role labelling have been used, most notably for analysing relations between attitude, holder and topic, as mentioned previously.

Current research suggest that analysing subjectivity in language is a more difficult task than standard problems related to content or topicality (Lee, 2004). Whether this is due to the immature nature of the field or an inherent aspect of the problem is however not settled. At least it has been argued that the standard representation in information retrieval, the *bag-of-words* representation, fails to capture the information necessary for successful processing of subjective aspects of language (Pang et al., 2002; Lee, 2004; Bai et al., 2005). These issues are discussed in more length in sections 2 to 4.

1.3 Paper Outline

At first glance a natural and pedagogical way of presenting this survey seemed to be to focus on some prototypical use cases or problem formulations and present methods and resources that have proved useful in solving these. However on a closer inspection, the immature and disparate nature of the area shows itself in that most of the work seem concerned with a rather unique use case.

Though different use cases often have special requirements, there are still quite large commonalities between them. Instead, what cuts across these disparate set of use cases are the levels of analysis on which the processing is performed. The remainder of this paper is therefore structured in the following sections: *word level analysis* (section 2), *document level analysis* (section 3) and *sentence/phrase level analysis* (section 4). The reason for starting with describing word level analysis, is that information extracted from that level is often used on the document and sentence levels. I conclude with a discussion of some open problems and directions for future research in section 5.

2 Word Level Analysis

The idea of words carrying attitudinal loading is usually attributed to Osgood's theory of semantic differentiation (Osgood et al., 1967). According to this theory, meaning is defined in a multidimensional semantic space, in which dimensions are defined through pairs of antonymous adjectives and direction and distance corresponds to valence and degree, respectively. A similar, somewhat influential model is Gärdenfors' theory of conceptual spaces (Gärdenfors, 1996).

The most comprehensive report on word level identification of subjective language and attitudinal valence is the article by Wiebe et al. (2004). They identify three clues signalling subjective language use: *hapax legomena* (words occurring only once in the corpus), collocations in the form of generalised n-grams containing words usually on stop lists, and verb and adjective features extracted using a clustering technique. These features indicate that the constructions that carry attitudinal information, differ from those carrying topicality.

Most approaches to word level analysis have ignored domain aspects and either used hand-crafted lexicons such as the General Inquirer (Stone et al., 1966) or hand-crafted lexicons expanded using semi-supervised learning such as SentiWordNet (Esuli and Sebastiani, 2006). Others have used unsupervised methods, together with very small sets of seed words, to infer the semantic orientation of other words. Sahlgren et al. (2007) for example hypothesise that attitudinally loaded terms are syntagmatically related to a set of seed words, including prototypical words such as "good" and "bad". They use a vector space model to represent co-occurrence patterns, and the similarity of a given word's vector to the centroids of the positive and negative seed words' vectors, to infer the valence of the word. A similar idea is exploited by Turney and Littman (2003) who use the point-wise mutual information between a word and the words from two seed sets, to infer valence. They also report experiments conducted with Latent Semantic Indexing. Other unsupervised approaches include clustering (Subasic and Huettner, 2001) and bootstrapping of extraction patterns based on a set of seed words (Riloff et al., 2003).

The idea of words carrying attitude in isolation is rather crude, since context most certainly plays a part in conveying attitudes. I discuss models dealing with attitudinal aspects in context in section 4.

3 Document Level Analysis

As noted in the introduction, most work on subjective language analysis has been performed on the document level. This commonly consist in identification and/or classification of documents according to subjective content or attitudinal valence. Some notable examples are Pang et al. (2002), Turney (2002) and Bai et al. (2005), in which valence classification is applied to movie reviews according to a *thumbs up/thumbs down* classification scheme. Pang et al. (2002) and Lee (2004) further suggest using valence classification in business intelligence applications by analysing free-form survey responses and for use in recommendation systems. Other examples are Pang and Lee (2005) and Goldberg and Zhu (2006) who also classify movie reviews, but use a multi-point rating scale instead of a bipolar classification. Dave et al. (2003) perform classification of online product reviews, in addition to mining attitudes towards specific product features, while Wiebe et al. (2004) propose using identification of subjective language and attitude and valence classification for relevance ranking and for separating factual and non-factual information for information extraction purposes.[2] Yih et al. (2004) present a method for finding "hot deals" in online deal forums. An interesting approach is that of Thomas et al. (2006) who classify political speeches in the form of U.S. Congressional floor debate transcripts according to *supporting* or *opposing* a legislation.

Some research has also been done at the borderline between document level and sentence level analysis, most notably in automatic summarisation of movie and product reviews Pang and Lee (2004), Hu and Liu (2004a), Hu and Liu (2004b) and Kim and Hovy (2006b); this line of research is discussed at more length in section 4.

The tasks one needs to address in these use cases are related to standard text classification, clustering

[2] As noted in the introduction, the latter might be a futile endeavour, though it may of course serve its purpose in separating obvious non-factual information from potentially factual.

and information retrieval as well as to analysis of genre and style (Pang et al., 2002; Dave et al., 2003; Karlgren and Cutting, 1994). Most approaches for attitude and valence analysis at the document level divide the process into two separate stages: an attitude identification stage, in which documents carrying attitude are filtered from those note carrying attitude, and a valence classification stage in which the valences of the attitudinal documents are inferred.

3.1 Attitude Identification

Though seemingly only a simplification of the classification case, attitude identification has proved challenging in its own right, and the features useful for identification of subjective language to some extent seem to differ from those useful for classification of valence (Karlgren et al., 2008). In fact, judging from the literature, reliable attitude identification might be an inherently difficult problem.

One problem seems to be that attitudinal words are often used in objective language, and conversely, neutral words might attain attitudinal meaning depending on context. The phrase ”a great deal” might for example signal a good buy in a product review, while being used synonymous with ”much” in a formal objective context (Lee, 2004). The fact that opinionated and factual documents tend to be composed of a mixture of subjective and objective language, makes document level attitude identification a difficult, perhaps even ill-posed, problem (Wiebe et al., 2004). To make things worse, depending on the domain there is also often a skew towards either opinionated or non-opinionated texts.

There have been some attempts on generally applicable attitude identification. Wiebe et al. (2004) use the normalised count of attitudinally loaded elements (words, phrases and generalised n-grams) to characterise a document in a nearest neighbour classifier. Another approach based on machine learning using extracted substrings as features is presented by Dave et al. (2003).

Related to these approaches are genre analysis, which has been used to separate editorials from other genres of news text (Karlgren and Cutting, 1994). The approach by Hatzivassiloglou and Wiebe (2000) is in this vein; they use presence of certain types of adjectives to separate objective and subjective language. While the proportion of subjective language use might be higher in certain genres and styles and the expressions of attitude are shaped and constrained by the conventions established in a particular genre, the notion of subjective language use cuts across these levels.

Most research on attitude and valence classification seem to be ignoring the problem of attitude identification altogether, concentrating on bipolar or multi-polar scale classification (Pang et al., 2002; Turney, 2002; Pang and Lee, 2005; Pang and Lee, 2004; Thomas et al., 2006; Goldberg and Zhu, 2006). Except from cases where a ”neutral” category is included in the classification problem, these methods of course presuppose that a prior identification step has been performed. In the case of product reviews this might not be a major obstacle in practice, since there are a plethora of online forums and other sites specialising in product reviews. However, if one wants to leverage the ever growing amounts of data generated by personal blogs and social networks, a more principal solution to the identification step seems necessary. A further problem might then be that the use of rhetorical devices such as irony and sarcasm are more common in these domains (Lee, 2004).

3.2 Attitude and Valence Classification

Early attempts at attitude and valence classification on the document level, were based on word level or phrase level analysis. Variations of this approach include Turney (2002) who classify movie reviews as *thumbs up* or *thumbs down* based on the average semantic orientation of adjectives; Subasic and Huettner (2001) who use fuzzy conjunction of fuzzy representations of individual words' attitudinal loading; and Dave et al. (2003) who average over individual word scores computed from a labelled training set. The work of Wiebe et al. (2004) also fall into this paradigm. They use the average frequency of positive and negative elements, determined using generalised n-grams, co-occurrence statistics and a manually constructed lexicon, to classify documents as having ”positive” or ”negative” valence.

More recent work use principled machine learning approaches, instead of first classifying individual words and sentences. Most notably this include the work by Pang et al. (2002), Pang and Lee (2004) and Bai et al. (2005). The experiments by Pang and Lee

Pang et al. (2002) Pang and Lee (2004), indicate that using the standard *bag-of-words* model in conjunction with standard machine learning methods such as Support Vector Machines and Naive Bayes, yields significantly worse performance for attitude and valence classification compared to topic based classification. The fact that high-frequency words, which are often filtered out in content based text classification, seem to be useful for attitude classification, suggests that different approaches might be applicable.[3]

In order to overcome this problem, there has been some attempts to utilise non-surface features such as parts-of-speech and syntactic structure, to improve results on document level classification. The intuition seems to be that the valence of a document is determined by the valence of the words and sentences *in context*. Though this intuition might be valid for sentence and word level analysis, as with topic based document level text classification, such attempts have not proved fruitful for document level valence classification (Moschitti and Basili, 2004; Bai et al., 2005). The only exception seems to be the work of Pang et al. (2002), wherein a local negation heuristic, in which a prefix was prepended to each word following a negation until the following sentence boundary, seemed to marginally improve results.

Bai et al. (2005) take a different direction. Instead of "throwing more features at the problem", they relax the assumption of conditional independence between surface form features (i.e. words), which underlies other methods such as Naive Bayes and linear models such as the Perceptron, and they improve on the simple form of conditional dependence modelled by kernel methods such as Support Vector Machines. Linear models can only capture linear dependencies between features, since the inner-product is bilinear. By using kernel functions, non-linear relationships can be captured as well. A polynomial kernel of degree two for example can be seen as implicitly representing all pair-wise combinations of the features of the input space. A problem with these methods is that it is hard to control exactly which features are independent and which are dependent, though it can to some degree be accomplished using cross-validation and feature selection; According to Bai et al. (2005) this can readily be done using graphical models. See also Lewis (1998) for a discussion on the role of conditional independence assumptions in information retrieval and machine learning applications.

By using a *Markov blanket* classifier (Pearl, 1988) in conjunction with a feature pruning heuristic based on *Tabu search*, Bai et al. manage to significantly improve results over Support Vector Machines, Naive Bayes and the Voted Perceptron, using only 22 surface features. The success of this method is attributed to the intuition that attitudinal valence is carried by high frequency words used to express lexical patterns, together with mid- and low frequency words indicating valence (Bai et al., 2005).

Note that this is the same intuition underlying the attempts to incorporate structural features as discussed previously. I believe that the failure of the latter attempt comes from the fact that simply conjoining structural and surface features results in too specific features, while the relaxation of conditional independence allows the induction of those structural constraints which carry sufficiently generalisable information. I conjecture that another reason for the failure of methods based on structural features might be the often relatively high rate of errors introduced in the structure analysis.

4 Phrase and Sentence Level Analysis

The issues raised in the previous section regarding contextual factors and the mixture of subjective and objective language at the document level, suggest that a more suitable level of analysis is on the sentence level. Furthermore, when turning to sentence level, identification of holders and topics become seemingly more tractable, though inter-sentence relationships has been shown to be exploitable in the analysis of individual sentences. Use cases on this level of analysis include product and movie review mining and summarisation, possibly applied towards specific product features (Dave et al., 2003; Pang and Lee, 2004; Hu and Liu, 2004a; Hu and Liu, 2004b; Popescu and Etzioni, 2005; Kim and Hovy,

[3] It should be noted, however, that there is a possibility that the inferior experimental results for attitude identification and valence classification are an artefact of the usually much lower inter-annotator agreement on the data sets compared to that on the data sets used in evaluating topical classification.

2006b), classification of claims made in political discourse (Kwon et al., 2007), classification of attitudinal valence of news headlines (Sahlgren et al., 2007), information extraction separating facts from non-facts and multi-perspective question answering (Wiebe et al., 2004) and identification and analysis of judgements and beliefs (Kim and Hovy, 2006b).

4.1 Identification, Classification and Summarisation

As with document level classification, baseline approaches for sentence level identification and classification use simple averaging of word level attitude and valence scores (Sahlgren et al., 2007; Wiebe et al., 2004; Dave et al., 2003). In some cases contextual heuristics such as negating the valence of words in the vicinity of negations (Hu and Liu, 2004a) or prepending them with a prefix (Wiebe et al., 2004; Dave et al., 2003) have been applied. A more linguistically motivated heuristic based on a lexical grammatical model is presented in Mulder et al. (2004). These heuristics seem to improve results in some cases, however the evidence as to their usefulness should not be considered conclusive.

The intuition that local argument-modifier relationships and structural properties are crucial in determining attitudinal valence in context has however proved useful when used for engineering features for use with machine learning methods. This work include that by Wilson et al. (2005), who use structural features extracted from dependency parse trees, to encode for example local and longer range modifications and passive constructions, for use with a boosting meta-algorithm. By using structural features they hope to be able to handle issues such as local negation, long-distance relationships and valence scaling by for example adjectives. Similarly Kwon et al. (2007) apply machine learning using structural features to identify claims in a document by comparing each sentence to the full document. They use separate features for identifying and classifying claims, in both cases making use of some structural features extracted from constituent parse trees. Another use of structural features in a machine learning approach to separating subjective from objective sentences, is presented in the work of Riloff et al. (2003). Use of more linguistically motivated structure is also reported in Yi et al. (2003), where for example subject/object relationships are exploited for identifying the topics of online reviews.

The cited work indicate that the problem of identifying subjective from objective sentences is an easier problem than the corresponding problem at the document level. This might be due to the fact that while documents often consist of a mixture of subjective and objective language, sentences tend to be either subjective or objective to a larger extent. These approaches has, as of yet, not provided groundbreaking results, though significant improvements compared to the baseline are reported. As with most approaches to document level classification, all reported experiments on sentence level analysis assume conditional independence between surface features. I suspect that relaxing this assumption might provide significant improvements on sentence level as well, however this hypothesis remains to be validated in future work.

As well as leveraging intra-sentence structure, successful attempts on making use of inter-sentence structure have been reported. Pang and Lee (2004) combine a Support Vector Machine classifier which assigns valence scores to individual sentences, with a graph representation for encoding adjacency weights between sentences, to create coherent summaries of movie reviews, in which the attitudinal content is maintained. They then apply a minimal-cut formulation to cluster sentences such that nearby sentences are optimally likely to assume the same attitudinal valence. Thomas et al. (2006) use the same method for classifying statements in political speeches. The intuition that valence should exhibit local coherence is also exploited in a heuristic for finding chunks of subjective sentences in Wiebe et al. (2004).

4.2 Topic and Holder Identification

Compared to the methods used for sentence level classification, topic and holder identification are more closely related to information extraction and sequence tagging methods used in for example syntactic chunking or semantic role labelling. Some current attempts actually make use of some variant of semantic role labelling to solve the problem, either as a pre-processing step as in the work of Kim and Hovy (2006a) or directly by using a custom labelling scheme as reported by Bethard et al. (2004).

Others have attempted to model the problem directly as a sequence labelling problem using a Maximum Entropy model (Kim and Hovy, 2006b) or Conditional Random Fields in conjunction with pattern matching (Choi et al., 2005). The latter two only address the problem of holder identification. The systems described by Hu and Liu (2004a) and Popescu and Etzioni (2005) could also be considered as identifying topic. However, their approach is too specifically targeted at the domain of product reviews to be considered general solutions to the problem of topic identification.

As with sentence level classification and identification, it is still an open question whether structural features are necessary for successful topic and holder identification or not. While all the systems reported above use some kind of structural features, it has not been evaluated to what extent these features improve performance. The question as to whether one should employ semantic role labelling or some other type of semantic dependency framework, also remain unanswered; two of the systems reported do so, while two of them do not. Unfortunately the approaches are not readily comparable, since they are not evaluated on comparable data sets.

The use of semantic role labelling has still been quite naive. Kim and Hovy (2006a) use manually created mappings from FrameNet frame elements to holders and topic, while Bethard et al. (2004) train a semantic role labelling system to instead output categories indicating holders and topics. However, their approach is also limited in that it only works with explicitly attitudinal verbs such as "accuse" and "suggest", which make holder identification rather trivial, since the holder of the attitude almost always assumes the role of *agent* for these verbs. Topic identification in general seems to be a more difficult problem than holder identification as noted by Kim and Hovy (2006a). However, if nested attitude attributions such as "John thought that Lisa did not like him" are considered, topic identification is a non-trivial problem as well.

5 Conclusions

In this paper I have presented a survey of recent research in the field of subjective language analysis. This turned out to be an highly active field, with a great deal of interesting results. The large body of research in the field has been performed in the last five to ten years, which can be seen in its somewhat disparate and immature character.

Most of the methods being employed have been carried over from nearby fields such as topic based text classification, information retrieval and information extraction. As is the case in most areas of natural language processing these days, machine learning, especially in the form of supervised learning, are being used extensively. On the other hand a surprisingly large body of research is based on handcrafted resources and linguistically inspired heuristics. I conjecture that these approaches will be more rare as the field matures, since data driven methods are more flexible and allows adaptation to new domains, languages and use cases.

5.1 Standardisation

Though research in the field have evolved quickly since the late 1990's and an impressive array of successful approaches have been brought about, I believe that more work is needed on standardising vocabulary and theory on subjective aspects of language use. Possibly it would also be beneficial if research could be focused on a more standardised set of use cases. In fields such as information retrieval, information extraction and text classification, standardised data sets have played an important part in focusing research, allowing for principled comparisons of different methods. In order to construct such resources, work would be required on standardising annotation schemes, use cases and theory. I further believe that moving beyond the simplified notion of attitudinal valence as being either "positive" and "negative" could prove beneficial. This could go hand in hand with exploring new use case that would benefit from aspects more related to emotions rather than attitude.

5.2 Unstructured Domains

The most common and in my view most interesting manifestation of subjective language is that used in personal blogs and on social forums. Language used in these domains are often rather unstructured, multi-lingual and often rapidly evolving. While full syntactic parsing and semantic role labelling seem promising directions for extraction of elements of subjective language in structured domains such as news text, their dependence on structure and well-formed language might be detrimental to their adoption for analysis in more unstructured domains.

Given the inter-textual nature of blogs and social network sites, combining attitude analysis with link analysis or other tools from mathematical sociology as proposed by Lee (2004), could be a fruitful direction for future research.

References

Xue Bai, Rema Padman, and Edoardo Airoldi. 2005. On learning parsimonious models for extracting consumer opinions. In *Proceedings of HICSS-05,the 38th Annual Hawaii International Conference on System Sciences*, page 75b, Washington, DC, USA. IEEE Computer Society.

Steven Bethard, Hong Yu, Ashley Thornton, Vasileios Hatzivassiloglou, and Dan Jurafsky. 2004. Automatic extraction of opinion propositions and their holders. In James G. Shanahan, Janyce Wiebe, and Yan Qu, editors, *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, Stanford, US.

Yejin Choi, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. 2005. Identifying sources of opinions with conditional random fields and extraction patterns. In *Proceedings of HLT-EMNLP-05, the Human Language Technology Conference/Conference on Empirical Methods in Natural Language Processing*, pages 355–362, Vancouver, CA.

Kushal Dave, Steve Lawrence, and David M. Pennock. 2003. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of WWW-03, 12th International Conference on the World Wide Web*, pages 519–528, Budapest, HU. ACM Press.

Paul Ekman. 1993. Facial expression and emotion. *American Psychologist*, 48:384–392.

Andrea Esuli and Fabrizio Sebastiani. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC-06, the 5th Conference on Language Resources and Evaluation*, Genova, IT.

Peter Gärdenfors. 1996. Conceptual spaces as a framework for cognitive semantics. In A. Clark, editor, *Philosophy and Cognitive Science*, pages 159–180. Kluwer, Dordrecht.

Andrew B. Goldberg and Xiaojin Zhu. 2006. Seeing stars when there aren't many stars: Graph-based semi-supervised learning for sentiment categorization. In *HLT-NAACL 2006 Workshop on Textgraphs: Graph-based Algorithms for Natural Language Processing*, New York, NY.

Vasileios Hatzivassiloglou and Janyce M. Wiebe. 2000. Effects of adjective orientation and gradability on sentence subjectivity. In *Proceedings of COLING-00, 18th International Conference on Computational Linguistics*, pages 299–305, Saarbrücken, GE. Morgan Kaufmann.

Minqing Hu and Bing Liu. 2004a. Mining and summarizing customer reviews. In *Proceedings of KDD '04, the ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177, Seattle, US. ACM Press.

Minqing Hu and Bing Liu. 2004b. Mining opinion features in customer reviews. In *Proceedings of AAAI-04, the 19th National Conference on Artificial Intellgience*, San Jose, US.

Jussi Karlgren and Douglass Cutting. 1994. Recognizing text genres with simple metrics using discriminant analysis. In *Proceedings of the 15th. International Conference on Computational Linguistics (*COLING *94)*, volume II, pages 1071 – 1075, Kyoto, Japan.

Jussi Karlgren, Gunnar Eriksson, and Oscar Täckström. 2008. Sics at ntcir-7 moat: constructions represented in parallel with lexical items. In *Proceedings of The 7th NTCIR Workshop (2007/2008) - Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access*, page 4, Tokyo, Japan.

Soo-Min Kim and Eduard Hovy. 2006a. Extracting opinions, opinion holders, and topics expressed in online news media text. In *Proceedings of ACL/COLING Workshop on Sentiment and Subjectivity in Text*, Sidney, AUS.

Soo-Min Kim and Eduard Hovy. 2006b. Identifying and analyzing judgment opinions. In *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pages 200–207, Morristown, NJ, USA. Association for Computational Linguistics.

Namhee Kwon, Liang Zhou, Eduard Hovy, and Stuart W. Shulman. 2007. Identifying and classifying subjective claims. In *dg.o '07: Proceedings of the 8th annual international conference on Digital government research*, pages 76–81. Digital Government Research Center.

Lillian Lee. 2004. A matter of opinion: Sentiment analysis and business intelligence (position paper). Written for and presented at the IBM Faculty Summit on the Architecture of On-Demand Business, May.

David D. Lewis. 1998. Naive (Bayes) at forty: The independence assumption in information retrieval. In

Claire Nédellec and Céline Rouveirol, editors, *Proceedings of ECML-98, 10th European Conference on Machine Learning*, pages 4–15, Chemnitz, DE. Springer Verlag, Heidelberg, DE.

Hugo Liu, Henry Lieberman, and Ted Selker. 2003. A model of textual affect sensing using real-world knowledge. In *Proceedings of IUI-03, the 8th international conference on Intelligent user interfaces*, pages 125–132, Miami, US. ACM Press.

Alessandro Moschitti and Roberto Basili. 2004. Complex linguistic features for text classification: A comprehensive study. In Sharon McDonald and John Tait, editors, *ECIR*, volume 2997 of *Lecture Notes in Computer Science*, pages 181–196. Springer.

Matthijs Mulder, Anton Nijholt, Marten den Uyl, and Peter Terpstra. 2004. A lexical grammatical implementation of affect. In *Proceedings of TSD-04, the 7th International Conference Text, Speech and Dialogue*, volume 3206 of *Lecture Notes in Computer Science*, pages 171–178, Brno, CZ. Springer-Verlag.

Charles E. Osgood, George J. Suci, and Percy H. Tannenbaum. 1967. *The Measurement of Meaning*. University of Illinois Press.

Bo Pang and Lillian Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of ACL-04, 42nd Meeting of the Association for Computational Linguistics*, pages 271–278, Barcelona, ES. Association for Computational Linguistics.

Bo Pang and Lillian Lee. 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of ACL-05, 43nd Meeting of the Association for Computational Linguistics*, pages 115–124, Ann Arbor, US. Association for Computational Linguistics.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of EMNLP-02, the Conference on Empirical Methods in Natural Language Processing*, pages 79–86, Philadelphia, US. Association for Computational Linguistics.

J. Pearl. 1988. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann.

Ana-Maria Popescu and Oren Etzioni. 2005. Extracting product features and opinions from reviews. In *Proceedings of HLT-EMNLP-05, the Human Language Technology Conference/Conference on Empirical Methods in Natural Language Processing*, pages 339–346, Vancouver, CA.

Ellen Riloff, Janyce Wiebe, and Theresa Wilson. 2003. Learning subjective nouns using extraction pattern bootstrapping. In Walter Daelemans and Miles Osborne, editors, *Proceedings of CONLL-03, 7th Conference on Natural Language Learning*, pages 25–32, Edmonton, CA.

Magnus Sahlgren, Jussi Karlgren, and Gunnar Eriksson. 2007. Sics: Valence annotation based on seeds in word space. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 296–299, Prague, Czech Republic, June. Association for Computational Linguistics.

Philip J. Stone, Dexter C. Dunphy, Marshall S. Smith, and Daniel M. Ogilvie. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press.

Pero Subasic and Alison Huettner. 2001. Affect analysis of text using fuzzy semantic typing. *IEEE Transaction on Fuzzy Systems*, 9(4):483–496.

Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In *Proceedings of EMNLP*, pages 327–335.

Peter D. Turney and Michael L. Littman. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems*, 21(4):315–346.

Peter Turney. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of ACL-02, 40th Annual Meeting of the Association for Computational Linguistics*, pages 417–424, Philadelphia, US. Association for Computational Linguistics.

Janyce Wiebe, Theresa Wilson, Rebecca F. Bruce, Matthew Bell, and Melanie Martin. 2004. Learning subjective language. *Computational linguistics*, 30(3):277–308.

Janyce Wiebe. 1990. Identifying subjective characters in narrative. In *Proceedings of the 11th. International Conference on Computational Linguistics (*COLING *90)*, pages 401–408.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of Human Language Technologies Conference/Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005)*, Vancouver, CA.

Jeonghee Yi, Tetsuya Nasukawa, Razvan Bunescu, and Wayne Niblack. 2003. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Proceeding of ICDM-03, the 3ird IEEE International Conference on Data Mining*, pages 427– 434, Melbourne, US. IEEE Computer Society.

W. Yih, P. Chang, and W. Kim. 2004. Mining online deal forums for hot deals. In *Proceedings of WI-04, the IEEE/WIC/ACM International Conference on Web Intelligence*, pages 384–390. IEEE Computer Society.