Opinion Mining in Legal Blogs

Jack G. Conrad
Research & Development
Thomson Legal & Regulatory
St. Paul, Minnesota 55123 USA
Jack.G.Conrad@Thomson.com

Frank Schilder
Research & Development
Thomson Legal & Regulatory
St. Paul, Minnesota 55123 USA
Frank.Schilder@Thomson.com

ABSTRACT

We perform a survey into the scope and utility of opinion mining in legal Weblogs (a.k.a. *blawgs*). The number of 'blogs' in the legal domain is growing at a rapid pace and many potential applications for opinion detection and monitoring are arising as a result. We summarize current approaches to opinion mining before describing different categories of blawgs and their potential impact on the law and the legal profession. In addition to educating the community on recent developments in the legal blog space, we also conduct some introductory opinion mining trials. We first construct a Weblog test collection containing blog entries that discuss legal search tools. We subsequently examine the performance of a language modeling approach deployed for both *subjectivity analysis* (i.e., is the text subjective or objective?) and *polarity analysis* (i.e., is the text affirmative or negative towards its subject?). This work may thus help establish early baselines for these core opinion mining tasks.

Categories and Subject Descriptors

H.3.0.a [**Information Storage and Retrieval**]: General—*Web Search*; I.2.7.i [**Information Storage and Retrieval**]: Natural Language Processing—*Web Text Analysis*; I.2.7.c [**Information Storage and Retrieval**]: Natural Language Processing—*Language Models*; H.3.m [**Information Storage and Retrieval**]: Miscellaneous—*Test Collections*

General Terms

Modeling, Experimentation, Measurement

Keywords

opinion mining, sentiment analysis, language modeling, legal blogs, test collections

1. INTRODUCTION

The mining of opinions in textual materials such as Weblogs adds another dimension to technologies that facilitate search and summarization. Opinion mining actually identifies the author's viewpoint about a subject, rather than simply identifying the subject itself. Current approaches tend to divide the problem space into sub-problems, for example, creating a lexicon of useful features that can help classify sentences (or portions of sentences) into categories of positive, negative or neutral. Existing techniques often try to identify words, phrases and patterns that indicate viewpoints. This has proven difficult, however, since it is not just the presence of a keyword that matters, but its context. For instance, *This is a great decision* conveys clear sentiment, but *The announcement of this decision produced a great amount of media attention* is neutral.

Analyzing text regarding its opinions can be extremely valuable to a legal researcher who is looking for a perspective on a legal issue or even information about a product or a service. Organizations may also benefit from automatic opinion mining by obtaining a timely picture of how their products or services, or more generally their names, are viewed by their clients. The Web is an expanding environment where customers go to find or submit opinions that may be ripe for mining. An increasing number of customer viewpoints are posted on Blogs (short for Weblogs). The content of these Blogs can range from short product reviews by consumers to elaborate essays on legal issues by law professors.

This paper provides an overview of recent approaches to opinion mining that harness advanced techniques such as language models. Moreover, we view the legal blog domain as ideal for applying opinion mining techniques and we propose several possible use cases that would leverage this new technology. We also describe some preliminary results from our own experiments that apply a sentiment analysis toolkit to a collection of blawg entries that we constructed.

The remainder of this paper is organized as follows: Section 2 reviews relevant work performed in the field. Section 3 defines legal blogs, characterizing the nature and scope of this application space. Section 4 describes our experimental framework and the test collection we developed for blawg-related applications. In Section 5, we draw our conclusions, assessing the strengths and weaknesses of our approach; we also discuss future prospects.

2. APPROACHES TO OPINION MINING

Academic and commercial interest in opinion mining research and applications has been steadily growing since at least the late 90s. This increase in attention seems to have culminated in the AAAI's 2004 Spring Symposium on "Exploring Attitude and Affect in Text: Theories and Applications" (AAAI-EAAT 2004),[1] at which some three dozen re-

ICAIL '07, June 4-8, 2007, Palo Alto, California USA
Copyright 2007 ACM 978-1-59593-680-6/07/0006 ...$5.00.

[1] www.clairvoyancecorp.com/research/workshops/AAAI-EAAT-2004/home.html

search works were presented. An expanded version of these works was published by Springer last year as a book [19].

Alternatively known as sentiment analysis or sentiment classification, Opinion Mining focuses not on the topic of a text, but an author's attitude towards that topic. In recent years, opinion mining has been applied to movie reviews, reviews of other commercial products and services, to Weblogs, and to News.

Following the cumulative advances spearheaded by Pang and Lee [16, 14, 15], the sub-tasks of opinion mining have evolved over the course of the last few years. These sub-tasks presently include:

1. *Subjectivity Analysis* – involves the determination of whether a given text is objective (neutral in sentiment) or subjective (expressing a positive or negative sentiment or opinion); this can be viewed as a binary classification task [6, 7, 23, 14, 27].

2. *Polarity Analysis* – encompasses the prediction of whether a text that has been established as subjective is positive or negative in its polarity [21, 14, 20, 4].

3. *Polarity Degree* – measures the *degree* of polarity, positive or negative, in a subjective text [30, 15].

Initial work in the field has consisted largely of a combination of tasks 2 and 3, namely, NLP-based identification of keywords, their polarity and, as much as possible, their degree [6, 22, 10, 11, 1]. That is, is a given term a reliable indicator of opinionated content? One of the most basic approaches to Task 2 and also one of the earliest was proposed by Turney, which determined the polarity of a text by calculating the numeric sum of the “orientations” of the individual words it contains [21]. Other more advanced approaches have subsequently been developed [25, 2, 18, 30, 24, 5].

Until Pang and Lee's seminal work on subjectivity, much less research has focused on task 1 [14]. The motivation for this initial step is simple: why invest computational and linguistic resources in portions of a text that hold no substantive subjective content? Rather, it makes more sense to apply a binary subjectivity/objectivity classifier to the text to identify those segments (sentences) that merit exclusion from further consideration. Few others have researched this fundamental dimension of opinion mining [18, 23, 3].

Another branch of subjectivity research that has developed more recently is the summarization of single or multiple consumer reviews [9, 8, 13].

3. THE LEGAL BLOGOSPHERE

3.1 Definitions

The definition for a general blog, as provided by Wikipedia, captures the basic features:

> A blog is a user-generated Website where entries are made in journal style and displayed in reverse chronological order.

Starting mainly as personal commentaries written in short entries, blogs quickly became a phenomenon that gained influence in political opinion shaping. In addition, specialized blogging communities, apart from journalistic and political blogs, have been developing over the last few years. Today, blogs can be on any possible topic (e.g., entertainment, education, science & technology, etc.).

Our interest is in legal blogs, and we define them as follows:

> A legal blog (a.k.a. *blawg*) is any Weblog that focuses on substantive discussions of the law, the legal profession, including law schools, and the process by which judicial decisions are made.

Blawgs can also be further sub-divided into different areas. A comprehensive taxonomy covering the different topics found in Blawgs can be accessed at *3L Epiphany*,[2] which is a Blawg maintained by a recent law school graduate. The top-level of his taxonomy shows a variety of topics for this blogging sub-community:

1. General Legal Blogs
2. Blogs Categorized by Legal Specialty
3. Blogs Categorized by Law or Legal Event
4. Blogs Categorized by Jurisdictional Scope
5. Blogs Categorized by Author/Publisher
6. Blogs Categorized by Number of Contributors
7. Miscellaneous Blogs Categorized by Topic
8. Collections of Legal Blogs

This taxonomy also shows that Blawgs are not only written by law students discussing their daily experiences in law school but also by legal professionals who can provide in-depth analysis of recent court decisions. *Scotusblog*, for example, provides commentary and analysis of U.S. Supreme Court decisions.[3] Another popular Blawg is the The *Volokh Conspiracy*[4] which is hosted by a variety of different law professors, including Eugene Volokh and Orin Kerr. A blog magazine maintained by Stephen Bainbridge, a law professor at UCLA, includes his personal blog on topics such as non-business law, political correctness and photography as well as his professional blog which targets lawyers, judges, law students and legal academics.[5]

3.2 Recent Developments

A good starting point for exploring the ‘Blawgosphere’ is one of the well-maintained Web portals — there now exist Web portals exclusively for legal blogs[6,7] as well as taxonomies of legal blogs.[8]

A significant indication of how important Blawgs may become for the legal profession is given by the fact that blog entries have been used in court cases. Entries have either been cited by the court or used as evidence by one of the parties in a judicial proceeding (e.g., *United States v. Booker, 543 U.S. 220, 278 (2005) (Stevens, J., dissenting)*).

Moreover, Blawgs are now also maintained by law firms.[2] There are several reasons why law firms use this new forum. First of all, they can use it to advertise the firm's authority in different areas. Secondly, it may be useful for recruiting law

[2] 3lepiphany.typepad.com/3l_epiphany/a_taxonomy_of_legal_blogs/index.html

[3] www.scotusblog.com/movabletype/

[4] www.volokh.com

[5] www.professorbainbridge.com

[6] www.blawg.com

[7] law-library.rutgers.edu/resources/lawblogs.php

[8] www.3lepiphany.typepad.com/

school graduates and, finally, it helps retain younger attorneys, allowing them to continue blogging activities started in law school.

As a general trend, lawyers seem increasingly to appreciate blogging as a discussion vehicle that can accelerate legal analysis. This medium can be advantageous for legal topics on which there have not been any law reports published yet.

3.3 Prospective Applications

There exist a number of compelling potential applications regarding opinion mining of legal blogs. Some of these include:

- *profiling* — reactions to high-level court decisions;
- *alerting* — subscribers to unfavorable news and disclosures that may impact a firm's clients;
- *monitoring* — e.g., what communities are saying about commercial legal research services;
- *tracking* — reputations of law firms based on client feedback over time;
- *hosting & surveying* — blog space for practitioners to comment on legal topics and decisions that can subsequently be mined for trends.

The degree to which these will become intrinsic to the field of law will depend on technological, economic, and domain-related factors, not to mention parallel developments in other professional fields.

4. EXPERIMENTS WITH LEGAL BLOGS

Our experiments were conducted in three distinct phases. The first phase exclusively targets polarity analysis. The second phase emphasizes potential benefits from introducing a subjectivity component as an initial step. Finally, the third phase is designed to harness larger corpora from outside the legal domain to provide additional training data for the subjectivity component.

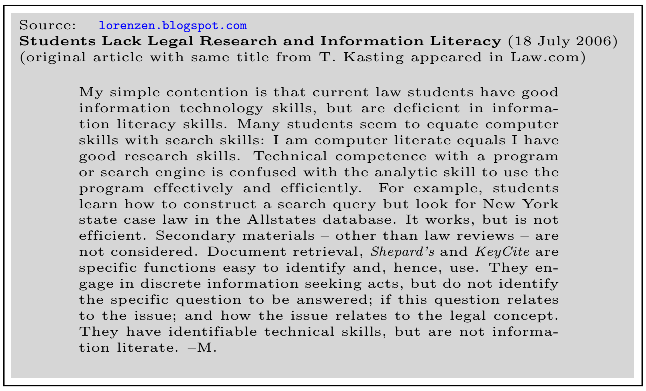
4.1 Test Corpus Generation

A principal use case supplied by one of our business units focused on the views of legal practitioners towards legal research tools such as *Westlaw*, *LexisNexis*, or others. In order to perform initial opinion mining trials on such data, we needed to produce our own working test corpus. We thus proceeded to construct such a collection by running a set of directed queries against an assortment of Web search engines that focus on blogs.

We ran these queries against a number of general as well as blog-specific search engines, for instance, Technorati, BlogPulse, IceRocket, Blog-Search, and Google's BlogSearch.[9] These queries produced a wide range of entries which had to be selectively pruned, in our case, by paralegal reviewers. In the end, this task produced roughly 200 blawg entries consisting of approximately 1,000 sentences. This collection focused on the original blawg entries, not on subsequent responses to them. Figure 1 presents a sample entry.

4.2 Language Modeling Resources

Alias-i, a text analytics and data mining company, has implemented its own version of an opinion mining application within its *LingPipe* toolkit. It is based upon the Min-Cut Graph model of Pang and Lee [14] and uses a n-gram language model (LM) [17]. Such language models have recently been deployed in the legal domain by Moens [12]. We chose *LingPipe* for our trials for three reasons. First, because it was readily available and rapidly deployable; second, because our group has developed a set of productive expertise with the toolkit; and, third and most significantly, because its Sentiment Analysis module has been shown to produce polarity analysis results comparable to Pang and Lee.[10] Yet it relies upon a character-based rather than token-based language model. A character-based LM assigns a probability to a sequence of characters rather than words. The learned probability is based on the probability distribution of the characters found in the training corpus.


Source: lorenzen.blogspot.com
Students Lack Legal Research and Information Literacy (18 July 2006)
(original article with same title from T. Kasting appeared in Law.com)

> My simple contention is that current law students have good information technology skills, but are deficient in information literacy skills. Many students seem to equate computer skills with search skills: I am computer literate equals I have good research skills. Technical competence with a program or search engine is confused with the analytic skill to use the program effectively and efficiently. For example, students learn how to construct a search query but look for New York state case law in the Allstates database. It works, but is not efficient. Secondary materials – other than law reviews – are not considered. Document retrieval, *Shepard's* and *KeyCite* are specific functions easy to identify and, hence, use. They engage in discrete information seeking acts, but do not identify the specific question to be answered; if this question relates to the issue; and how the issue relates to the legal concept. They have identifiable technical skills, but are not information literate. –M.

Figure 1. Sample Blog Entry from Blawg Corpus

Some of the parameters associated with the toolkit include which classifier to use (Language Modeling vs. Naive Bayes with LM smoothing) and n, the number of characters used in the n-grams. The basic *LingPipe* application has both a subjectivity and polarity analysis component.

4.3 Results

4.3.1 Polarity Analysis

We first focus on a straight evaluation of the performance of our positive/negative polarity analysis system.

In order to study the impact that product, service, or company entity mentions may have on training and overall performance, we created three separate versions of our corpus:

1. *original* – product, service, company mentions left intact.
2. *masked* – product, service, company mentions replaced with a *'[product-company]'* token.
3. *removed* – product, service, company mentions deleted.

We also consider the impact of a balanced set of training data, where the number of positive and negative examples is equal, versus a set of training data based strictly on the ratios found in the blog acquisition process. It is not uncommon to encounter in these blogs more examples of negative comments about products or services than positive ones.

We also distinguish initially between two different levels of granularity in our training data, entry-level and sentence-level, which can be described as follows:

[9] www.technorati.com, www.blogpulse.com, www.icerocket.com, www.blog-search.com, www.blogsearchengine.com, blogsearch.google.com

[10] www.alias-i.com/lingpipe

1. *entry-level* – the relevant portions of the entire blog are treated as a composite entry. If applicable, a single entry is permitted to produce one segment for positive sentiment and another for negative sentiment.
2. *sentence-level* – each sentence in the blog is treated as a separate entity, which may possess positive sentiment, negative sentiment, or neither.

The *entry-level* approach, while more straightforward and less costly, is also less precise in terms of what is input for training (i.e., while a blog entry may clearly represent an overall affirmative review, it may also contain appreciable negative comments as well). The *sentence-level* approach, by contrast, is more precise in terms of just what one is able to tag and submit as positive and negative examples, but requires the additional overhead of sentence-level assessment and tagging. After attempting to allow for the existence of both positive and negative portions of any entry, in the end, we found that sentence-level training data was more effective and directly satisfied our training needs.

We use accuracy (total correct/total count) and F-score, the evenly weighted harmonic mean of precision and recall, to measure our performance. The initial results for Language Modeling and Naive Bayes classification are shown in Table 1. Results reported are produced by 10-way cross validation and use the top performing n-grams.[11]

Corpus-type	Entity-type	Accuracy		F-score	
		LM,n=3	NB,n=2	LM,n=3	NB,n=2
Balanced	Original	0.6254	*0.6667*	0.6222	*0.6663*
	Masked	0.6159	0.6571	0.6108	0.6569
	Removed	0.6127	0.6571	0.6067	0.6569
Unbalanced	Original	0.6650	**0.6895**	0.5931	**0.6546**
	Masked	0.6601	0.6870	0.5785	0.6513
	Removed	0.6626	0.6846	0.5827	0.6479

Table 1: Accuracy, F-score Results for Initial Polarity Analysis (Sentence-level)

In the case of the balanced trials, where the amount of positive and negative training examples are the same, our models outperform a baseline of random assignment (0.666 vs. 0.500). In the case of the unbalanced trials, by contrast, where the amount of negative training examples exceeds the positive by two-thirds (250 vs. 150 sentences), our models again surpass this baseline in terms of accuracy and F-score, but less significantly (0.690/0.655 vs. 0.625). Furthermore, the Naive Bayes model slightly outperforms Language Modeling. A key observation about this initial step is that processing an entry without first identifying (and ignoring) neutral or non-subjective content can only dilute the effectiveness of a dedicated polarity analyzer. It is also worth mentioning that Hatzivassiloglou [6], Turney [22], and Pang [16] have all reported precision levels of over 80% for related polarity analysis tasks. Two of the distinctions to be made, however, lie in the focus of their measurements and the nature of their data sets. In the first two works, assessment is essentially based on the semantic orientation of proximate words, whereas in the third case, the data set (movie reviews) tends to have limited extraneous material, and has by its very nature a high subjectivity/sentiment "quotient." By contrast, we focus on sentence-level assessment using data that comes from the blogosphere, and, as such, is arguably more heterogeneous. Consequently, direct head to head comparisons of metrics like precision must be made with caution, since key underlying experimental parameters and targets are different.

4.3.2 Subjectivity Analysis

Our initial results suggest that we might improve our performance if able to add a filtering mechanism that would exclude portions of an entry that possess no opinion.[12] For this reason we introduced an initial subjectivity analysis component, which trained on sentence-level data that was tagged for subjectivity/objectivity. The results appear in Table 2.

Corpus-type	Entity-type	Accuracy	F-score
Balanced	Original	0.6220	0.6192
	Masked	*0.6341*	*0.6322*
	Removed	*0.6341*	*0.6322*
Unbalanced	Original	**0.5978**	**0.5896**
	Masked	0.5870	0.5798
	Removed	0.5870	0.5798

Table 2: Accuracy, F-score Results for Subjectivity Analysis

When compared to our initial polarity analysis results, these subjectivity results, with an accuracy and F-score averaging around ±0.60, remain below practical requirements. Again we see that our models surpass a random baseline in the case of the balanced trials, and surpass it marginally for the unbalanced trials (0.590-0.600 vs 0.444, given 400 of 900 examples being subjective). For subjectivity analysis, there has been less published work, but these results appear sub-optimal. For example, Wiebe and Riloff report a 72%-78% F-score using Naive Bayes [27]. Again one needs to be cautious in making such a comparison due to the distinct nature of two content sets examined.

Given the means by which language models learn their data distributions, the benefits of larger amounts of annotated training data became apparent. For this reason, we undertook a search for additional data tagged for sentiment.

4.3.3 Supplemental Training Data

We subsequently sought to improve our performance by increasing the size of our training data set. In order to do this, we harnessed three larger subjectivity corpora based on consumer reviews or news.[13]

Source DB for Training	Neutral/Subjective Sentences Used	Accuracy	F-score
Movie Reviews	5060/4090	0.5190	0.5164
Customer Reviews	2148/2396	**0.5736**	**0.5647**
MPQA Database	3439/4090	0.5418	0.5412

Table 3: Accuracy and F-score for Supp. Subjectivity — External Training Data with Blog Test Data

[11] Best performance for balanced corpora is shown in *italics* and for unbalanced corpora in **boldface**.

[12] When Pang and Lee added such a filter to their polarity analysis engine, their results increased from 82.8% to 86.4% using the database described in the next footnote (cf: [14]).

[13] The **movie review database** can be found at: www.cs.cornell.edu/people/pabo/movie-review-data/ . In addition to containing 1,000 positive and 1,000 negative reviews, it also contains 5,000 subjective and 5,000 objective sentences originating in reviews.
The **customer review data set** was constructed from Amazon data and reported on in [9, 8].
The **MPQA database** was constructed at the 2003 AAAI Spring Symposium on New Directions in Question Answering and was subsequently annotated during a corpus enhancement project [26, 28, 29].

Using a 90/10 split between training data and test data, we used the external data for our training set (90%) and blawg data for our test set (10%). Results are shown in Table 3. The ratios of neutral to subjective sentences were dependent on their availability in the training data. We note that the results from the Customer Reviews training set were significantly better than that from the Movie Reviews and MPQA databases. This makes intuitive sense since the Customer Reviews set included comments on interactions with customer support as well as with digital technology, both having analogous discussions in our blawg data set. Furthermore, one can observe that the threshold for subjectivity in the MPQA if not the Movie Review database is arguably lower than that for the Customer Reviews set [28].

The results above raise an important question about the source of such sub-optimal performance relative to published results for this task. Is the cause the experimental techniques and configuration used, or the nature and scope of the blawg data itself? To help answer this question, we repeated our subjectivity experiment, but using the external data sets for both training (90%) and testing (10%). The results can be seen in Table 4. As mentioned in Section 4.2, *LingPipe* performs as well against the Movie Review database as Pang and Lee [14].[14] It also performs comparably to Wiebe and Riloff against the MPQA database [27]. We observe, however, that as we apply these LM techniques to Customer Reviews, the data set arguably closest to our own, we see an appreciable drop in performance, to levels resembling what we witnessed in Table 3. These results suggest that our poorer performance on the blawg data is less due to the classifier and more due to the distinct nature of the domain and the data itself.

Corpus-type	Source DB for Training / Testing	No. Neutral / Subjective Sentences	Accuracy	F-score
Balanced	Movie Revs.	2500/2500	*0.8980*	*0.8979*
	Cust. Revs.	1933/1933	0.5825	0.5754
	MPQA DB	3095/3095	0.7387	0.7355
Unbalanced	Movie Revs.	5000/5000	**0.9210**	**0.9209**
	Cust. Revs.	1933/2156	0.6098	0.5966
	MPQA DB	3095/6670	0.7554	0.6666

Table 4: Accuracy and F-score for Subjectivity Analysis — External Only Training and Test Data

4.3.4 Sequential Subjectivity – Polarity Analysis

LingPipe also permits the sequential application of subjectivity and polarity analysis via LM. We subsequently used the subjectivity model constructed from the Customer Reviews data above and ran it on our unbalanced data set from Section 4.3.1. This produced an accuracy of 0.70 with P = 0.69, R = 0.64. Although this represents a modest improvement in performance over results in 4.3.1, it also suggests that training with enough suitable and sizable supplemental training data may boost performance. What is clear is that in order to benefit appreciably from such a first stage subjectivity filter, one would need to obtain accuracy levels substantially higher than we have witnessed using these three external resources. Also worth noting is that the higher precision requirements of operational environments at the expense of recall may be acceptable, depending on customer needs and use cases. Thus a system may miss some evidence, but still be able to perform well enough to make sound global sentiment assessments.

5. CONCLUSIONS

We conduct a survey on opinion mining techniques and discuss how this new field could be applied to the growing domain of legal blogs. Opinion mining promises to be highly beneficial to legal practitioners who must remain current in new developments about the law and the legal profession.

We also perform a series of preliminary experiments on opinion mining in *blawgs*, which may help answer some basic research questions about sentiment analysis in the legal space, such as appropriate applications of this technology and the strengths and weaknesses of Language Model vs. Naive Bayes classifiers. Introductory as this study is, it leaves open other questions about more sophisticated opinion mining applications, such as those that examine authoritative legal perspectives of current issues under scrutiny by the courts.

5.1 Observations on Findings

Opinion mining is a hard problem, in particular when done on real data. The results of our efforts to harness Language Modeling and Naive Bayes techniques have yet to meet operational requirements. We can, however, make some meaningful observations.

- Using a balanced corpus simply does not reflect how the data is actually distributed. Negative sentiment often occurs more frequently than positive sentiment in the blogosphere. This finding may call into question results from other studies which rely exclusively upon balanced training data [14].
- A Naive Bayes classifier with Language Model smoothing performs slightly better than Language Modeling alone, at least for reasonably small values of n.
- Masking or removing product names does not have any significant impact on our results.
- Adding subjectivity data from a different domain may help, provided that it is relevant to the application and of sufficient quantity.

5.2 Future Work

We are currently investigating other approaches to opinion mining which use alternatives to language modeling. These include other basic binary classifiers for the sentiment and polarity analysis tasks. But future mining tools should also be able to assess and *summarize* the sentiment of legal communities, at levels that may not yet require human-level understanding.

6. ACKNOWLEDGMENTS

We thank Breck Baldwin and Bob Carpenter for their insights into the application of language modeling in this domain. We are also grateful to Therese Peterson for her querying, editing and annotating services.

7. REFERENCES

[1] M. Baroni and S. Vegnaduzzo. Identifying subjective adjectives through web-based mutual information. In E. Buchberger, editor, *In Proceedings of the Conference for*

[14] To avoid giving the Movie Review database with its 10,000 sentences an unfair advantage over the other two in Table 4, we used half the set when running side-by-side comparisons with the 'balanced' collections.

the Processing of Natural Language and Speech (KONVENS), pages 17–24, Vienna, Austria, September 2004.

[2] K. Dave, S. Lawrence, and D. M. Pennock. Mining the peanut gallery: Opinon extraction and semantic classification of product reviews. In *Proceedings of the International World Wide Web Conference*, Budapest, Hungary, May 2003.

[3] A. Esuli and F. Sebastiani. Determining term subjectivity and term orientation for opinion mining. In *Proceedings EACL-06, the 11th Conference of the European Chapter of the Association for Computational Linguistics*, Trento, Italy, 2006.

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

[5] M. Gamon, A. Aue, S. Corston-Oliver, and E. Ringger. Pulse: Mining customer opinions from free text. In *Proceedings of IDA-05, the 6th International Symposium on Intelligent Data Analysis*, Lecture Notes in Computer Science, Madrid, Spain, 2005. Springer-Verlag.

[6] V. Hatzivassiloglou and K. R. McKeown. Predicting the semantic orientation of adjectives. In *Proceedings of ACL-97, 35th Annual Meeting of the Association for Computational Linguistics*, pages 174–181, Madrid, Spain, 1997. Association for Computational Linguistics.

[7] V. Hatzivassiloglou and J. M. Wiebe. Effects of adjective orientation and gradability on sentence subjectivity. In *Proceedings of COLING-00, 18th International Conference on Computational Linguistics*, pages 299–305, Saarbrücken, Germany, 2000. Morgan Kaufmann.

[8] M. Hu and B. Liu. Mining opinion features in customer reviews. In *Proceedings of AAAI-04, the 19th National Conference on Artificial Intellgience*, San Jose, California, 2004.

[9] M. Hu and B. Lu. Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 168–177, Seattle, Washington, August 2004. ACM Press.

[10] J. Kamps, M. Marx, R. J. Mokken, and M. de Rijke. Words with attitude. In *Proceedings of the 1st International Conference on Global WordNet*, pages 332–341, Mysore, India, January 2002.

[11] S.-M. Kim and E. Hovy. Determining the sentiment of opinions. In *Proceedings COLING-04, the Conference on Computational Linguistics*, Geneva, Switzerland, 2004.

[12] M.-F. Moens. A retrieval model for accessing legislation. In *Proceedings of the Tenth International Conference on Artificial Intelligence and Law (ICAIL 2005)*, pages 141–145, Bologna, Italy, June 2005. ACM Press.

[13] S. Morinaga, K. Yamanishi, K. Tateishi, and T. Fukushima. Mining product reputations on the web. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Management Discovery and Data Mining*, pages 341–349, Edmonton, Alberta Canada, July 2002. ACM Press.

[14] B. Pang and L. Lee. Sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the ACL 2004*, pages 271–278, Madrid, Spain, 2004. Association for Computational Linguistics, ACL Press.

[15] B. Pang and L. Lee. Seeing stars: Exploting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, 2005.

[16] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP-02)*, pages 79–86, Philadelphia, Pennsylvania, July 2002. ACL Press.

[17] J. M. Ponte and W. B. Croft. A language modeling approach to information retrieval. In *Proceedings of the 21st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR98)*, pages 275–281, Melbourne, Australia, August 1998. ACM Press.

[18] E. Riloff and J. Wiebe. Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP-03)*, pages 105–112, Sapporo, Japan, July 2003. ACL Press.

[19] J. Shanahan, Y. Qu, and J. Wiebe, editors. *Computing Attitude and Affect in Text.* Springer, 2006. Collected Works.

[20] H. Takamura, T. Inui, and M. Okumura. Extracting emotional polarity of words using spin model. In *Proceedings of ACL-05, 43rd Annual Meeting of the Association for Computational Linguistics*, Ann Arbor, Michigan, 2005. Association for Computational Linguistics.

[21] P. D. Turney. Thumbs up or thumbs down? In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 417–424, Philadelphia, Pennsylvania, July 2002. ACL Press.

[22] P. D. Turney and M. L. Littman. Measuring praise and criticism: Inference of semantic orientation from association. In G. Marchionini, editor, *Transactions on Information Systems (TOIS)*, volume 21, no. 4, pages 315–346, New York, October 2003. ACM Press.

[23] S. Vegnaduzzo. Acquisition of subjective adjectives with limited resources. In J. G. Shanahan, J. Wiebe, and Y. Qu, editors, *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, Stanford, US, 2004.

[24] C. Whitelaw, N. Garg, and S. Argamon. Using appraisal groups for sentiment analysis. In O. Herzog, H.-J. Schek, N. Fuhr, A. Chowdhury, and W. Teiken, editors, *Proceedings of the 14th ACM International Conference on Information and Knowledge Management (CIKM05)*, pages 625–631, Bremen, Germany, November 2005. ACM Press.

[25] J. Wiebe. Learning subjective adjectives from corpora. In *Proceedings of AAAI-00, 17th Conference of the American Association for Artificial Intelligence*, pages 735–740, Austin, Texas, 2000. AAAI Press / The MIT Press.

[26] 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 2003 AAAI Spring Symposium on New Directions in Question Answering*. AAAI Press, 2003.

[27] J. Wiebe and E. Riloff. Creating subjective and objective sentence classifiers from unannotated texts. In *Proceeding of CICLing-05, International Conference on Intelligent Text Processing and Computational Linguistics.*, volume 3406 of *Lecture Notes in Computer Science*, pages 475–486, Mexico City, Mexico, 2005. Springer-Verlag.

[28] J. Wiebe, T. Wilson, and C. Cardie. Annotating expressions of opinions and emotions in language. In *Language Resources Evaluation*, volume 29, pages 165–210, the Netherlands, 2005. Kluwer Academic Publishers.

[29] T. Wilson, J. Wiebe, and P. Hoffmann. 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)*, pages 347–354, Vancouver, BC Canada, October 2005. ACL Press.

[30] T. Wilson, J. Wiebe, and R. Hwa. Just how mad are you? finding strong and weak opinion clauses. In D. L. McGuinness and G. Ferguson, editors, *Proceedings of the 19th National Conference on Artificial Intelligence, 16th Conference on Innovative Applications of AI*, pages 761–769. American Association of Artificial Intelligence (AAAI), AAAI Press / The MIT Press, July 2004.