

Generating Semantic Orientation Lexicon using Large Data and Thesaurus

Amit Goyal and Hal Daumé III

Dept. of Computer Science

University of Maryland

College Park, MD 20742

{amit,hal}@umiacs.umd.edu

Abstract

We propose a novel method to construct semantic orientation lexicons using large data and a thesaurus. To deal with large data, we use Count-Min sketch to store the approximate counts of all word pairs in a bounded space of 8GB. We use a thesaurus (like Roget) to constrain near-synonymous words to have the same polarity. This framework can easily scale to any language with a thesaurus and a unzipped corpus size ≥ 50 GB (12 billion tokens). We evaluate these lexicons intrinsically and extrinsically, and they perform comparable when compared to other existing lexicons.

1 Introduction

In recent years, the field of natural language processing (NLP) has seen tremendous growth and interest in the computational analysis of emotions, sentiments, and opinions. This work has focused on many application areas, such as sentiment analysis of consumer reviews e.g., (Pang et al., 2002; Nasukawa and Yi, 2003), product reputation analysis e.g., (Morinaga et al., 2002; Nasukawa and Yi, 2003), tracking sentiments toward events e.g., (Das and Chen, 2001; Tong, 2001), and automatically producing plot unit representations e.g., (Goyal et al., 2010b). An important resource in accomplishing the above tasks is a list of words with semantic orientation (SO): positive or negative. The goal of this work is to automatically create such a list of words using large data and a thesaurus structure.

For this purpose, we store exact counts of all the words in a hash table and use Count-Min (CM) sketch (Cormode and Muthukrishnan, 2004; Goyal et al., 2010) to store the approximate counts of all word pairs for a large corpus in a bounded space of

8GB. (Storing the counts of all word pairs is computationally expensive and memory intensive on large data (Agirre et al., 2009; Pantel et al., 2009)). Storage space saving in CM sketch is achieved by *approximating* the frequency of word pairs in the corpus without explicitly storing the word pairs themselves. Both updating (adding a new word pair or increasing the frequency of existing word pair) and querying (finding the frequency of a given word pair) are constant time operations making it an efficient online storage data structure for large data.

Once we have these counts, we find semantic orientation (SO) (Turney and Littman, 2003) of a word using its association strength with positive (e.g. good, and nice) and negative (e.g., bad and nasty) seeds. Next, we make use of a thesaurus (like Roget) structure in which near-synonymous words appear in a single group. We compute the SO of the whole group by computing SO of each individual word in the group and assign that SO to all the words in the group. The hypothesis is that near synonym words should have similar polarity. However, similar words in a group can still have different connotations. For example, one group has “slender”, “slim”, “wiry” and “lanky”. One can argue that, first two words have positive connotation and last two have negative. To remove these ambiguous words errors from the lexicon, we discard those words which have conflicting SO compared to their group SO. The idea behind using thesaurus structure is motivated from the idea of using number of positive and negative seed words (Mohammad et al., 2009) in thesaurus group to determine the polarity of words in the group.

In our experiments, we show the effectiveness of the lexicons created using large data and freely avail-

able thesaurus both intrinsically and extrinsically.

2 Background

2.1 Related Work

The literature on sentiment lexicon induction can be broadly classified into three categories: (1) Corpora based, (2) using thesaurus structure, and (3) combination of (1) and (2). Pang and Lee (2008) provide an excellent survey on the literature of sentiment analysis. We briefly discuss some of the works which have motivated our research for this work. A web-derived lexicon (Velikovich et al., 2010) was constructed for all words and phrases using graph propagation algorithm which propagates polarity from seed words to all other words. The graph was constructed using distributional similarity between the words. The goal of their work was to create a high coverage lexicon. In a similar direction (Rao and Ravichandran, 2009), word-net was used to construct the graph for label propagation. Our work is most closely related to Mohammad et al. (2009) which exploits thesaurus structure to determine the polarity of words in the thesaurus group.

2.2 Semantic Orientation

We use (Turney and Littman, 2003) framework to infer the Semantic Orientation (SO) of a word. We take the seven positive words (good, nice, excellent, positive, fortunate, correct, and superior) and the seven negative words (bad, nasty, poor, negative, unfortunate, wrong, and inferior) used in (Turney and Littman, 2003) work. The *SO* of a given word is calculated based on the strength of its association with the seven positive words, and the strength of its association with the seven negative words using pointwise mutual information (PMI). We compute the *SO* of a word "w" as follows:

$$SO(w) = \sum_{p \in Pwords} PMI(p, w) - \sum_{n \in Nwords} PMI(n, w)$$

where, Pwords and Nwords denote the seven positive and seven negative prototype words respectively. If this score is negative, the word is predicted as negative. Otherwise, it is predicted as positive.

2.3 CM sketch

The Count-Min sketch (Cormode and Muthukrishnan, 2004) with user chosen parameters (ϵ, δ) is

represented by a two-dimensional array with width w and depth d . Parameters ϵ and δ control the amount of tolerable error in the returned count (ϵ) and the probability with which the returned count is not within this acceptable error (δ) respectively. These parameters determine the width and depth of the two-dimensional array. We set $w = \frac{2}{\epsilon}$, and $d = \log(\frac{1}{\delta})$. The depth d denotes the number of pairwise-independent hash functions and there exists an one-to-one mapping between the rows and the set of hash functions. Each of these hash functions $h_k: \{x_1 \dots x_N\} \rightarrow \{1 \dots w\}$, $1 \leq k \leq d$, takes an item from the input stream and maps it into a counter indexed by the corresponding hash function. For example, $h_3(x) = 8$ indicates that the item "x" is mapped to the 8th position in the third row of the sketch.

Update Procedure: When a new item "x" with count c , the sketch is updated by:

$$\text{sketch}[k, h_k(x)] \leftarrow \text{sketch}[k, h_k(x)] + c, \quad \forall 1 \leq k \leq d$$

Query Procedure: Since multiple items can be hashed to the same position, the stored frequency in any one row is guaranteed to *overestimate* the true count. Thus, to answer the point query, we return the minimum over all the positions indexed by the k hash functions. The answer to Query(x) is: $\hat{c} = \min_k \text{sketch}[k, h_k(x)]$.

2.4 CU sketch

The Count-Min sketch with conservative update (CU sketch) (Goyal et al., 2010) is similar to CM sketch except the update operation. It is based on the idea of conservative update (Estan and Varghese, 2002) introduced in the context of networking. It is used with CM sketch to further improve the estimate of a point query. To update an item, x with frequency c , we first compute the frequency \hat{c} of this item from the existing data structure and the counts are updated according to:

$$\hat{c} = \min_k \text{sketch}[k, h_k(x)], \quad \forall 1 \leq k \leq d$$

$$\text{sketch}[k, h_k(x)] \leftarrow \max\{\text{sketch}[k, h_k(x)], \hat{c} + c\}$$

The intuition is that, since the point query returns the minimum of all the d values, we will update a counter only if it is necessary as indicated by the above equation.

3 Generating Polarity Lexicon

Our framework to generate lexicon has three main steps: First, we compute Semantic Orientation (SO) of words using a formula defined in Section 2.2 using a large corpus. Second, we use a thesaurus (like Roget) to constrain all synonym words in a group to have the same polarity. Third, we discard words which do not follow the above constraints. The three steps are discussed in the following subsections.

3.1 Computing SO of a word

We use CM sketch to store counts of word pairs (except word pairs involving stop words and numbers) within a sliding window of size¹ 7 using a large corpus: GWB66 of size 64GB (see Section 4.3). We fix the number of counters of the sketch to 2 billion (2B) (8GB of memory) with conservative update (CU) as it performs the best for (Goyal et al., 2010) with $d = 5$ (see Section 2.3) hash functions. We store exact counts of words in hash table.

Once, we have stored the counts for all words and word pairs, we can compute the SO of a word using a formula defined in Section 2.2. Moreover, a word can have multiple senses, hence it can belong to multiple paragraphs. To assign a single label to a word, we combine all its SO scores. We use positive SO scores to label words as positive and negative SO to label words as negative. We discard words with SO equal to zero. We apply this strategy to all the words in a thesaurus (like Roget) (refer to Section 3.2), we call the lexicon constructed using SO scores using thesaurus words as “SO” lexicon.

3.2 Using Thesaurus structure

Thesaurus like Roget², Macquarie are available in several languages. We use freely available version of Roget thesaurus which has 1046 categories, each containing on average 64 words and phrases. Terms within a category are closely related to one another, and they are further grouped into near-synonymous words and phrases called paragraphs. There are about 3117 paragraphs in Roget thesaurus. One of the examples of paragraphs from the Roget thesaurus is shown in Table 1. All the words appears to be near-synonymous with positive polarity.

¹Window size 7 is chosen from intuition and not tuned.

²<http://www.nzdl.org/ELKB/>

pure undefiled modest delicate decent decorous cherry chaste
continent virtuous honest platonic virgin unsullied simonpure

Table 1: A paragraph from the Roget thesaurus

We assign semantic orientation (SO) score to a thesaurus paragraph³ ($SO(TP)$) by averaging over SO scores over all the words in it. The $SO(TP)$ score constrains all the words in a paragraph to have same polarity. If $SO(TP) > 0$, all the words in a paragraph are marked as positive. If $SO(TP) < 0$, all the words in a group are marked as negative. For $SO(TP) = 0$, we discard all the words of a paragraph. For the paragraph in Table 1, the $SO(TP)$ for the paragraph is 8.72. Therefore, all the words in this paragraph are labeled as positive. However, the SO scores for “virgin” and “decorous” are negative, therefore they are marked as negative by previous lexicon “SO”, however they seem to be more positive than negative. Therefore, using the structure of the lexicon helps us in correcting the polarity of these words to negative. We apply this strategy to all the 3117 Roget thesaurus paragraphs and construct “SO-TP” lexicon using $SO(TP)$ scores.

3.3 Words and Thesaurus Consensus

Since near-synonymous words could have different connotation or polarity. Hence, here we use both SO of word and $SO(TP)$ of its paragraph to assign polarity to a word. If $SO(w) > 0$ and $SO(TP) > 0$, then we mark that word as positive. If $SO(w) < 0$ and $SO(TP) < 0$, then we mark that word as negative. In other cases, we discard the word.

We refer to the lexicon constructed using this strategy on Roget thesaurus paragraphs as “SO-WTP” lexicon. The motivation behind this is to generate precision orientated lexicon by having consensus over both individual and paragraph scores. For the paragraph in Table 1, we discard words “virgin” and “decorous” from the lexicon, as they have conflicting $SO(w)$ and $SO(TP)$ scores. In experiments in Section 5.2.1, we also examine existing lexicons to constrain the polarity of thesaurus paragraphs.

4 Evaluating SO computed using sketch

We compare the accuracy of computed SO using different sized corpora. We also compare exact counts with approximate counts using sketch.

³We do not assign polarity to phrases and stop words.

4.1 Data

We use Gigaword corpus (Graff, 2003) and a 66% portion of a copy of web crawled by (Ravichandran et al., 2005). For both the corpora, we split the text into sentences, tokenize and convert into lower-case. We generate words and word pairs over a sliding window of size 7. We use four different sized corpora: Gigaword (GW), GigaWord + 16% of web data (GWB16), GigaWord + 50% of web data (GWB50), and GigaWord + 66% of web data (GWB66). Corpus Statistics are shown in Table 2. We store exact counts of words in a hash table and store approximate counts of word pairs in the sketch.

4.2 Test Set

We use General Inquirer lexicon⁴ (Stone et al., 1966) as a benchmark to evaluate the semantic orientation scores similar to (Turney and Littman, 2003) work. Our test set consists of 1597 positive and 1980 negative words. Accuracy is used as an evaluation metric.

Corpus	GW	GWB16	GWB50	GWB66
Unzipped Size (GB)	9.8	22.8	49	64
# of sentences (Million)	56.78	191.28	462.60	608.74
# of Tokens (Billion)	1.8	4.2	9.1	11.8
Stream Size (Billion)	2.67	6.05	13.20	17.31

Table 2: Corpus Description

4.3 Effect of Increasing Corpus Size

We evaluate SO of words on four different sized corpora (see Section 4.1): GW (9.8GB), GWB20 (22.8GB), GWB50 (49GB) and GWB66 (64GB). First, we will fix number of counters to 2 billion (2B) (CU-2B) as it performs the best for (Goyal et al., 2010). Second, we will compare the CU-2B model with the Exact over increasing corpus size.

We can make several observations from the Figure 1: • It shows that increasing the amount of data improves the accuracy of identifying the SO of a word. We get an absolute increase of 5.5 points in accuracy when we add 16% Web data to GigaWord (GW). Adding 34% more Web data (GWB50), gives a small increase of 1.3 points. Adding 16%

⁴The General Inquirer lexicon which is freely available at <http://www.wjh.harvard.edu/~inquirer/>

more Web data (GWB66), give an increase of 0.5 points. • Second, CU-2B performs as good as Exact. • These results are also comparable to Turney’s (2003) state-of-the-art work where they report an accuracy of 82.84%. Note, they use a 100 billion tokens corpus which is larger than GWB66 (12 billion tokens).

This experiments shows that using unzipped corpus size ≥ 50 GB (12 billion tokens), we get performance comparable to the state-of-the-art. Hence, this approach is applicable for any language which has large collection of monolingual data available in it. Note that these results compared to best results of (Goyal et al., 2010) that is 77.11 are 4.5 points better; however in their work their goal was to show their approach scales to large data. We suspect the difference in results is due to difference in pre-processing and choosing the window size. We used counts from GWB66 (64GB) to generate lexicons in Section 3.

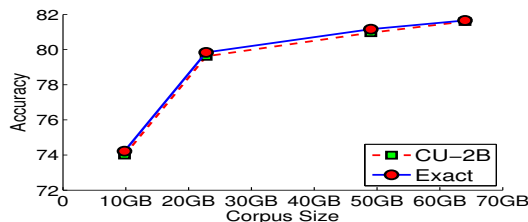


Figure 1: Evaluating Semantic Orientation of words with Exact and CU counts with increase in corpus size

5 Lexicon evaluation

We evaluate the lexicons proposed in Section 3 both intrinsically (by comparing their lexicon entries against General Inquirer (GI) lexicon) and extrinsically (by using them in a phrase polarity annotation task). We remove stop words and phrases for comparison from existing lexicons as our framework does not assign polarity to them.

5.1 Intrinsic evaluation

We compare the lexicon entries of “SO”, “SO-TP”, and “SO-WTP” against entries of GI Lexicon. This evaluation is similarly used by other authors (Turney and Littman, 2003; Mohammad et al., 2009) to evaluate sentiment lexicons.

Table 3 shows the percentage of GI positive (Pos), negative (Neg) and all (All) lexicon entries that

Lexicon (size)	Pos (1597)	Neg (1980)	All (3577)
SO (32.2K)	0.79	0.73	0.76
SO-TP (33.1K)	0.88	0.64	0.75
SO-WTP (22.6K)	0.78	0.65	0.71
Roget-ASL (27.8K)	0.79	0.40	0.57

Table 3: The percentage of GI entries (positive, negative, and all) that match those of the automatically generated lexicons

match the proposed lexicons. The recall of our precision orientated lexicon SO-WTP is only 5 and 4 % less compared to SO and SO-TP respectively which are more recall oriented. We evaluate these lexicons against Roget-ASL (discussed in Section 5.2.1). Even, Our SO-WTP precision oriented lexicon has more recall than Roget-ASL.

5.2 Extrinsic evaluation

In this section, we compare the effectiveness of our lexicons on a task of phrase polarity identification. We use the MPQA corpus which contains news articles from a wide variety of news sources manually annotated for opinions and other private states (like beliefs, emotions, sentiments, speculations, etc.). Moreover, it has polarity annotations (positive/negative) at the phrase level. We use MPQA⁵ version 2.0 collection of 2789 positive and 6079 negative phrases. We perform an extrinsic evaluation of our automatic generated lexicons (using large data and thesaurus) against existing automated and manually generated lexicons by using them to automatically determine the phrase polarity. This experimental setup is similar to Mohammad et al. (2009). However, in their work, they used MPQA version 1.0.

We use a similar algorithm as used by Mohammad et al. (2009) to determine the polarity of the phrase. If any of the words in the target phrase is labeled in the lexicon as having negative SO, then the phrase is marked as negative. If there are no negative words in the target phrase and it contains one or more positive words, then the phrase is marked as positive. In all other cases, do not assign any tag.

The only difference with respect to Mohammad et al. (2009) is that we use a list of 58 negation words used in OpinionFinder⁶ (Wilson et al., 2005b) (Version 1.4) to flip the polarity of a phrase if it contains odd number of negation words. We can get better

⁵<http://www.cs.pitt.edu/mpqa/databaserelease/>

⁶www.cs.pitt.edu/mpqa/opinionfinderrelease

Lexicon	# of positives	# of negatives	# of all
GI	1597	1980	3577
MPQA	2666	4888	7554
ASL	2320	2616	4936
Roget (ASL)	21637	6161	27798
Roget (GI)	10804	16319	27123
Roget (ASL+GI)	16168	12530	28698
MSOL	22088	32712	54800
SO	16620	15582	32202
SO-TP	22959	10117	33076
SO-WTP	14357	8257	22614
SO+GI	8629	9936	18565
SO-TP+GI	12049	9317	21366

Table 4: Summarizes all lexicons size accuracies on phrase polarity identification using supervised classifiers (Wilson et al., 2005a). However, the goal of this work is only to show the effectiveness of large data and thesaurus learned lexicons.

5.2.1 Baselines

We compare our method against the following baselines: First, MPQA Lexicon⁷ ((Wilson et al., 2005a)). Second, we use Affix seed lexicon (ASL) seeds used by Mohammad et al. (2009) to assign labels to Roget thesaurus paragraphs. ASL was constructed using 11 affix patterns, e.g. honest-dishonest (X-disX pattern). If ASL matches more positive words than negative words in a paragraph then all the words in the paragraph are labeled as positive. However, if ASL matches more negative words than positive words in a paragraph, then all words in the paragraph are labeled as negative. For other cases, we do not assign any labels. The generated lexicon is referred as Roget (ASL). Third, we use GI Lexicon instead of ASL and generate Roget (GI) Lexicon. Fourth, we use ASL + GI, and generate Roget (ASL+GI) Lexicon. Fifth, MSOL⁸ generated by Mohammad et al. (2009) using ASL+GI lexicon on Macquarie Thesaurus. Note that Macquarie Thesaurus is not freely available and its size is larger than the freely available Roget’s thesaurus.

5.2.2 GI seeds information with SO Lexicon

We combine the GI seed lexicon with semantic orientation of word computed using large corpus to mark the words positive or negative in thesaurus paragraphs. We combine the information

⁷www.cs.pitt.edu/mpqa/lexiconrelease/collectinfo1.html

⁸<http://www.umiacs.umd.edu/~saif/Release/MSOL-June15-09.txt>

Polarity	+ (2789)			- (6079)			All (8868)		
	R	P	F	R	P	F	R	P	F
SO Lexicon									
MPQA	.48	.73	.58	.48	.95	.64	.48	.87	.62
Roget (ASL)	.64	.45	.53	.32	.90	.47	.42	.60	.49
Roget (GI)	.50	.60	.55	.55	.86	.67	.53	.76	.62
Roget (ASL+GI)	.62	.57	.59	.49	.91	.64	.53	.75	.62
MSOL	.51	.58	.54	.60	.84	.70	.57	.74	.64
SO	.63	.54	.58	.50	.90	.64	.54	.73	.62
SO-TP	.68	.51	.58	.44	.93	.60	.52	.69	.59
SO-WTP	.65	.54	.59	.44	.93	.60	.51	.72	.60
SO+GI	.60	.57	.58	.46	.93	.62	.50	.75	.60
SO-TP+GI	.62	.58	.60	.45	.93	.61	.51	.76	.61

Table 5: Results on marking polarity of phrases using various lexicons. The # in parentheses is the # of gold +/-all phrases.

from large corpus with GI in two forms: • **SO+GI**: If GI matches more number of positive words than negative words in a paragraph and SO of a word > 0 , then that word is labeled as positive. However, if GI matches more number of negative words than positive words in a paragraph and SO of a word < 0 , that word is labeled as negative. For other cases, we do not assign any labels to words. • **SO-TP+GI**: Here, we use $SO(TP)$ scores instead of SO scores and use the same strategy as in previous bullet to generate the lexicon.

Table 4 summarizes the size of all lexicons. MPQA has the largest size among manually created lexicons. It is build on top of GI Lexicon. Roget (ASL) has 78% positive entries. MSOL is the biggest lexicon and it is about 2.5 times bigger than our precision oriented SO-WTP lexicon.

5.2.3 Results

Table 5 demonstrates the performance of the algorithm (discussed in Section 5.2) when using different lexicons. The performance of existing lexicons is shown in the top part of the table. The performance of large data and thesaurus lexicons is shown in the middle of the table. The bottom of the table combines GI information with large data and thesaurus.

In the first part of the Table 5, our results demonstrate that MPQA in the first row of the table has the best precision on this task for both positive and negative phrases. Roget (ASL) in the second row has the best recall for positives which is double the recall for negatives. Hence, this indicates that ASL is biased towards positive words. Using GI with Roget gives more balanced recall for both positives and negatives in third row. Roget (ASL+GI) are more biased towards positive words. MSOL has the best

recall for negatives; however it comes at an expense of equal drop in precision with respect to MPQA.

In the second part of the Table using large data, “SO” lexicon has same F-score as MPQA with precision and recall trade-offs. Using thesaurus along with large data has comparable F-score; however it again gives some precision and recall trade-offs with noticeable 6 points drop in recall for negatives. The small decrease in F-score for SO-WTP precision-oriented lexicon (22, 614 entries) is due to its small size in comparison to SO lexicon (32, 202 entries). We are currently working with a small sized freely available thesaurus which is smaller than Macquarie, hence MSOL performs the best.

Using GI lexicon in bottom part of the Table, we incorporate another form of information, which provides overall better precision than SO, SO-TP, and SO-WTP approaches. Even for languages, where we have only large amounts of data available, “SO” can be beneficial. If we have thesaurus available for a language, it can be combined with large data to produce precision oriented lexicons.

6 Discussion and Conclusion

We constructed lexicons automatically using large data and a thesaurus and evaluated its quality both intrinsically and extrinsically. This framework can easily scale to any language with a thesaurus and a unzipped corpus size of ≥ 50 GB (12 billion tokens). However, if a language does not have thesaurus, word similarity between words can be used to generate word clusters. Currently we are exploring using word clusters instead of using thesaurus in our framework. Moreover, if a language does not have large collection of data, we like to explore bilingual lexicons to compute semantic orientation of a word in another language. Another promising direction would be to explore the idea of word similarity combined with CM sketch (stores the approximate counts of all word pairs in a bounded space of 8GB) in graph propagation setting without explicitly representing the graph structure between words.

Acknowledgments

We thank the anonymous reviewers for helpful comments. This work is partially funded by NSF grant IIS-0712764 and Google Research Grant Grant for Large-Data NLP.

References

- Eneko Agirre, Enrique Alfonseca, Keith Hall, Jana Kravalova, Marius Paşca, and Aitor Soroa. 2009. A study on similarity and relatedness using distributional and wordnet-based approaches. In *NAACL '09: Proceedings of HLT-NAACL*.
- Graham Cormode and S. Muthukrishnan. 2004. An improved data stream summary: The count-min sketch and its applications. *J. Algorithms*.
- S. R. Das and M. Y. Chen. 2001. Yahoo! for Amazon: Opinion extraction from small talk on the Web. In *Proceedings of the 8th Asia Pacific Finance Association Annual Conference (APFA)*, Bangkok, Thailand.
- Cristian Estan and George Varghese. 2002. New directions in traffic measurement and accounting. *SIGCOMM Comput. Commun. Rev.*, 32(4).
- Amit Goyal, Jagadeesh Jagarlamudi, Hal Daumé III, and Suresh Venkatasubramanian. 2010. Sketching techniques for Large Scale NLP. In *6th WAC Workshop at NAACL-HLT*.
- Amit Goyal, Ellen Riloff, and Hal Daume III. 2010b. Automatically producing plot unit representations for narrative text. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 77–86. Association for Computational Linguistics, October.
- D. Graff. 2003. English Gigaword. Linguistic Data Consortium, Philadelphia, PA, January.
- Saif Mohammad, Cody Dunne, and Bonnie Dorr. 2009. Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 599–608. Association for Computational Linguistics.
- Satoshi Morinaga, Kenji Yamanishi, Kenji Tateishi, and Toshikazu Fukushima. 2002. Mining product reputations on the Web. In *Proceedings of the 8th Association for Computing Machinery SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2002)*, pages 341–349, Edmonton, Canada.
- Tetsuya Nasukawa and Jeonghee Yi. 2003. Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the 2nd International Conference on Knowledge Capture (K-CAP 2003)*, pages 70–77, Sanibel Island, Florida.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, Vol. 2(1-2):pp. 1–135.
- Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. pages 79–86, Philadelphia, Pennsylvania.
- Patrick Pantel, Eric Crestan, Arkady Borkovsky, Ana-Maria Popescu, and Vishnu Vyas. 2009. Web-scale distributional similarity and entity set expansion. In *Proceedings of EMNLP*.
- Delip Rao and Deepak Ravichandran. 2009. Semi-supervised polarity lexicon induction. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 675–682, Athens, Greece, March. Association for Computational Linguistics.
- Deepak Ravichandran, Patrick Pantel, and Eduard Hovy. 2005. Randomized algorithms and nlp: using locality sensitive hash function for high speed noun clustering. In *Proceedings of ACL*.
- 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.
- Richard Tong. 2001. An operational system for detecting and tracking opinions in on-line discussions. In *Working Notes of the Special Interest Group on Information Retrieval (SIGIR) Workshop on Operational Text Classification*, pages 1–6, New Orleans, Louisiana.
- Peter D. Turney and Michael L. Littman. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Trans. Inf. Syst.*, 21:315–346, October.
- Leonid Velikovich, Sasha Blair-Goldensohn, Kerry Hannan, and Ryan McDonald. 2010. The viability of web-derived polarity lexicons. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 777–785, Los Angeles, California, June. Association for Computational Linguistics.
- Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005a. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 347–354. Association for Computational Linguistics.
- T. Wilson, P. Hoffmann, S. Somasundaran, J. Kessler, J. Wiebe, Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan. 2005b. OpinionFinder: A system for subjectivity analysis. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing Interactive Demonstrations*, pages 34–35.