Large-Scale Sentiment Analysis for News and Blogs

Namrata Godbole[⋆] namratagodbole@gmail.com

Manjunath Srinivasaiah[⋆] manj.blr@gmail.com

Steven Skiena[⋄] skiena@cs.sunysb.edu

[⋆]Google Inc., New York NY, USA

[⋄]Dept. of Computer Science, Stony Brook University, Stony Brook, NY 11794-4400, USA

Abstract

Newspapers and blogs express opinion of news entities (people, places, things) while reporting on recent events. We present a system that assigns scores indicating positive or negative opinion to each distinct entity in the text corpus. Our system consists of a sentiment identification phase, which associates expressed opinions with each relevant entity, and a sentiment aggregation and scoring phase, which scores each entity relative to others in the same class. Finally, we evaluate the significance of our scoring techniques over large corpus of news and blogs.

1. Introduction

News can be good or bad, but it is seldom neutral. Although full comprehension of natural language text remains well beyond the power of machines, the statistical analysis of relatively simple sentiment cues can provide a surprisingly meaningful sense of how the latest news impacts important entities.

In this paper, we report on our development of a large-scale sentiment analysis system for news and blog entities built on top of the *Lydia* text analysis system [1, 2, 3, 4, 5]. We determine the public sentiment on each of the hundreds of thousands of entities that we track, and how this sentiment varies with time. We encourage the reader to study our historical sentiment analysis for your favorite news entities at *http://www.textmap.com* and view our daily sentiment analysis report at *http://www.textmap.com/sentiment*. We give several examples of our analysis in the demonstration paper of our system, which appears in this volume [6].

In this paper, we discuss several aspects of our sentiment analysis system, including:

- *Algorithmic Construction of Sentiment Dictionaries* – Our sentiment index relies critically on tracking the reference frequencies of adjectives with positive and negative connotations. We present a method for expanding small candidate seed lists of positive and negative words into full sentiment lexicons using path-based analysis of synonym and antonym sets in WordNet. We use sentiment-alternation hop counts to determine the polarity strength of the candidate terms and eliminate the ambiguous terms. We present the detailed algorithm and performance results.

- *Sentiment Index Formulation* – There is considerable subtlety in constructing a statistical index which meaningfully reflects the significance of sentiment term juxtaposition. We present our technique of using juxtaposition of sentiment terms and entities and a frequency-weighted interpolation with world happiness levels to score entity sentiment.

- *Evaluation of Significance* – We provide statistical evidence of the validity of our sentiment evaluation by correlating our index with several classes of real-world events, including (1) results of professional baseball and basketball games, (2) performance of stock-market indices, and (3) seasonal effects [6]. Positive correlations prove that our sentiment analyzer can accurately measure public sentiment. We also present additional anecdotal evidence corroborating our analysis.

Finally, we discuss possible applications and implications of our work.

2. Related work

Sentiment analysis of natural language texts is a large and growing field. Previous work particularly relevant to our task falls naturally in two groups. The first relates to techniques to automatically generate sentiment lexicons. The second relates to systems that analyze sentiment (on a global or local basis) for entire documents.

2.1 Determining semantic orientation of words

Hatzivassiloglou and McKeown[7] hypothesize that adjectives separated by "and" have the same polarity, while those separated by "but" have opposite polarity. Starting with small seed lists, this information is used to group adjectives into two clusters such that maximum constraints are satisfied.

Wiebe [8] evaluates adjectives for polarity as well as gradation classification. A statistical model groups adjectives into clusters, corresponding to their tone/orientation. The use of such gradable adjectives is an important factor in determining subjectivity. Statistical models are used to predict the gradability of adjectives.

Kim and Hovy[9] evaluate the sentiment of an opinion holder (entity) using WordNet to generate lists of positive and negative words by expanding seed lists. They assume that synonyms (antonyms) of a word have the same (opposite) polarity. The percentage of a word's synonyms belonging to lists of either polarity was used as a measure of its polarity strength, while those below a threshold were deemed neutral or ambiguous. Their best results were achieved when the

ICWSM'2007 Boulder, Colorado, USA

topic neighborhood consisted of words between the topic up to the end of the sentence.

2.2 Sentiment analysis systems

Several systems have been built which attempt to quantify opinion from product reviews. Pang, Lee and Vaithyanathan[10] perform sentiment analysis of movie reviews. Their results show that the machine learning techniques perform better than simple counting methods. They achieve an accuracy of polarity classification of roughly 83%. In [11], they identify which sentences in a review are of subjective character to improve sentiment analysis. We do not make this distinction in our system, because we feel that that both fact and opinion contribute to the public sentiment about news entities.

Nasukawa and Yi[12] identify local sentiment as being more reliable than global document sentiment, since human evaluators often fail to agree on the global sentiment of a document. They focus on identifying the orientation of sentiment expressions and determining the target of these sentiments. Shallow parsing identifies the target and the sentiment expression; the latter is evaluated and associated with the target. Our system also analyzes local sentiments but aims to be quicker and cruder: we charge sentiment to all entities juxtaposed in the same sentence as instead of a specific target. In [13], they follow up by employing a feature-term extractor. For a given item, the feature extractor identifies parts or attributes of that item. e.g., *battery* and *lens* are features of a *camera*.

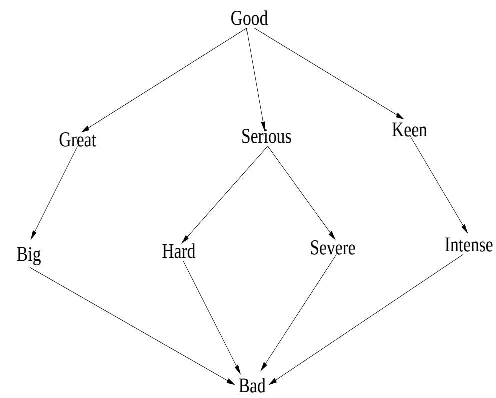

Fig. 1: *Four ways to get from bad to good in three hops*

Dimension	Seeds		Algorithmic		Hand-curated	
	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.
Business	11	12	167	167	223	180
Crime	12	18	337	337	51	224
Health	12	16	532	532	108	349
Media	16	10	310	310	295	133
Politics	14	11	327	327	216	236
Sports	13	7	180	180	106	53

Table 1: *Sentiment dictionary composition for adjectives*

3. Sentiment lexicon generation

Sentiment analysis depends on our ability to identify the sentimental terms in a corpus and their orientation. We defined separate lexicons for each of seven sentiment dimensions (general, health, crime, sports, business, politics, media). We selected these dimensions based on our identification of distinct news spheres with distinct standards of opinion and sentiment. Enlarging the number of sentiment lexicons permits greater focus in analyzing topic-specific phenomena, but potentially at a substantial cost in human curation. To avoid this, we developed an algorithm for expanding small dimension sets of seed sentiment words into full lexicons.

3.1 Lexicon expansion through path analysis

Previous systems detailed in Section 2 have expanded seed lists into lexicons by recursively querying for synonyms using the computer dictionary WordNet [14]. The pitfall of such methods that that synonym set coherence weakens with distance. Figure 1 shows four separate ways to get from *good* to *bad* using chains of WordNet synonyms.

To counteract such problems, our sentiment word generation algorithm expands a set of seed words using synonym and antonym queries as follows:

- We associate a *polarity* (positive or negative) to each word and query both the synonyms and antonyms, akin to [15, 16] Synonyms inherit the polarity from the parent, whereas antonyms get the opposite polarity.

- The significance of a path decreases as a function of its length or *depth* from a seed word, akin to [9, 17, 18]. The significance of a word W at depth d decreases exponentially as $score(W) = 1/c^d$ for some constant $c > 1$. The final score of each word is the summation of the scores received over all paths.

- Paths which alternate between positive and negative terms are likely spurious. Thus our algorithm runs in two iterations. The first iteration calculates a preliminary score estimate for each word as described above. The second iteration re-enumerates the paths while calculating the number of apparent sentiment alternations, or *flips*. The fewer the flips, the more trustworthy the path is. The final score is calculated taking into account only those paths whose flip value is within our threshold.

- WordNet [14] orders the synonyms/antonyms by sense, with the more common senses listed first. We improve accuracy by limiting our notion of synonym/antonym to only the top senses returned for a given word.

- This algorithm generates more than 18,000 words as being within five hops from our small set of seed words. Since the assigned scores followed a normal distribution, they are naturally converted to z-scores. Most words lying in the middle of this distribution are ambiguous, meaning they cannot be consistently classified as positive or negative. Such ambiguous words are discarded by taking only the top X% words from either extremes of the curve.

Table 1 presents the composition of our algorithmically-generated and curated sentiment dictionaries for each class of adjectives.

3.2 Performance evaluation

We evaluated our sentiment lexicon generation in two different ways. The first we call the *un-test*. The prefixes *un-* and *im-* generally negate the sentiment of a term. Thus the terms of form X and unX should appear on different ends of the sentiment spectrum, such as *competent* and *incompetent*. Table 2 reports the fraction of (term, negated term) pairs with same

Flips/%	100%	75%	50%
0	88/961 (0.092)	43/667 (0.064)	26/465 (0.056)
1	92/977 (0.094)	58/717 (0.081)	41/526 (0.078)
2	94/977 (0.096)	58/725 (0.080)	47/543 (0.087)
3	94/977 (0.096)	58/725 (0.080)	47/544 (0.086)

Table 2: *Precision/recall tradeoffs for lexicon expansion as a function of flip threshholds and prunning less polar terms*

Reference file		Intersection Polarity			
Name	Words	Diff.	Same	Recall	Precision
PolPMan	657	21	468	0.712	0.957
PolMMan	679	5	549	0.809	0.991
PolPauto	344	42	221	0.642	0.840
PolMauto	386	56	268	0.694	0.827

Table 3: *Comparison of algorithmically-generated and human-curated lexicons*

polarity. Thus the lower this ratio, the better. Our results show that precision increases at the expense of recall as we (1) restrict the number of path sentiment alternations and (2) prune increasing fractions of less polar terms.

We also compared our sentiment lexicons against those obtained by Wiebe [19], as reported in Table 3. There is a high degree of agreement between our algorithmically-generated lexicon and the manually curated lexicons. Further, we find our algorithmically-generated polarity is often sound even when it differs from [19]. For example, the negative lexicon *PolMauto* contained such clearly positive words like *bullish*, *agile*, and *compassionate*, while the positive lexicon *PolPman* contained words like *strenuous*, *uneventful*, and *adamant*.

4. Interpretation and scoring of sentiment data

We use our sentiment lexicons to mark up all sentiment words and associated entities in our corpus. We reverse the polarity of a sentiment word whenever it is preceded by a negation. We increase/decrease the polarity strength when a word is preceded by a modifier. Thus *not good* = -1; *good* = +1; *very good* = +2.

Our sentiment analyzer ignores articles which are detected as being a duplicate of another [1]. This prevents news syndicate articles from having a larger impact on the sentiment than other articles. Since our system processes vast quantities of text on a daily basis, speed considerations prevent us from doing careful parsing. Instead, we use co-occurrence of an entity and a sentiment word in the same sentence to mean that the sentiment is associated with that entity. This is not always accurate, particularly in complex sentences. Still the volume of text we process enables us to generate accurate sentiment scores.

We take several steps to aggregate entity references under different names. By employing techniques for pronoun resolution, we can identify more entity/sentiment co-occurrences than occur in the original news text. Further, *Lydia*'s system for identifying co-reference sets [4] associates alternate references such as *George W. Bush* and *George Bush* under the single synonym set header *George W. Bush*. This consolidates sentiment pertaining to a single entity.

4.1 Polarity scores

DIMENSION	BUS	CRIME	GEN	HEAL	MED	POL	SPT
BUSINESS	-	-.004	.278	.187	.189	.416	.414
CRIME	-.004	-	.317	.182	-.117	-.033	-.125
GENERAL	.278	.317	-	.327	.253	.428	.245
HEALTH	.187	.182	.327	-	.003	.128	.051
MEDIA	.189	-.117	.253	.003	-	.243	.241
POLITICS	.416	-.033	.428	.128	.243	-	.542
SPORTS	.414	-.125	.245	.051	.241	.542	-

Table 4: *Dimension correlation using monthly data*

We use the raw sentiment scores to track two trends over time:

- *Polarity:* Is the sentiment associated with the entity positive or negative?
- *Subjectivity:* How much sentiment (of any polarity) does the entity garner?

Subjectivity indicates proportion of sentiment to frequency of occurrence, while *polarity* indicates percentage of positive sentiment references among total sentiment references.

We focus first on polarity. We evaluate *world_polarity* using sentiment data for all entities for the entire time period:

$$world\_polarity = \frac{positive\_sentiment\_references}{total\_sentiment\_references}$$

We evaluate $entity\_polarity_i$ using sentiment data for that day (day_i) only:

$$entity\_polarity_i = \frac{positive\_sentiment\_references_i}{total\_sentiment\_references_i}$$

Table 4 shows the correlation coefficient between the various sentiment indices. In general, pairs of indices are positively correlated but not very strongly. This is good, as it shows each subindex measures different things. The *General* index is the union of all the indices and hence is positively correlated with each individual index.

4.2 Subjectivity scores

The subjectivity time series reflects the amount of sentiment an entity is associated with, regardless of whether the sentiment is positive or negative. Reading all news text over a period of time and counting sentiment in it gives a measure of the average subjectivity levels of the world. We evaluate *world_subjectivity* using sentiment data for all entities for the entire time period:

$$world\_subjectivity = \frac{total\_sentiment\_references}{total\_references}$$

We evaluate $entity\_subjectivity_i$ using sentiment data for that day (day_i) only:

$$entity\_subjectivity_i = \frac{total\_sentiment\_references_i}{total\_references_i}$$

5. News vs. blogs

The issues and the people discussed in blogs varies considerably from newspapers [2]. Table 5 lists the people that were the most positive in newspapers and blogs, respectively, as of July 2006. American investor *Warren Buffet* and F-1 driver *Fernando Alonso*, driver are regarded positively both in blogs and newspapers. Other sportsmen (*Rafael Nadal*,

	Net sentiment			Net sentiment	
Actor	News	Blog	Actor	Blog	News
Felicity Huffman	1.337	0.774	Joe Paterno	1.527	0.881
Fernando Alonso	0.977	0.702	Phil Mickelson	1.104	0.652
Dan Rather	0.906	-0.040	Tom Brokaw	1.042	0.359
Warren Buffett	0.882	0.704	Sasha Cohen	1.000	0.107
Joe Paterno	0.881	1.527	Ted Stevens	0.820	0.118
Ray Charles	0.843	0.138	Rafael Nadal	0.787	0.642
Bill Frist	0.819	0.307	Felicity Huffman	0.774	1.337
Ben Wallace	0.778	0.570	Warren Buffett	0.704	0.882
John Negroponte	0.775	0.059	Fernando Alonso	0.702	0.977
George Clooney	0.724	0.288	Chauncey Billups	0.685	0.580
Alicia Keys	0.724	0.147	Maria Sharapova	0.680	0.133
Roy Moore	0.720	0.349	Earl Woods	0.672	0.410
Jay Leno	0.710	0.107	Kasey Kahne	0.609	0.556
Roger Federer	0.702	0.512	Tom Brady	0.603	0.657
John Roberts	0.698	-0.372	Ben Wallace	0.570	0.778

Table 5: *Top positive entities in news (left) and blogs (right)*

	Net sentiment			Net sentiment	
Actor	News	Blog	Actor	Blog	News
Slobodan Milosevic	-1.674	-0.964	John Muhammad	-3.076	-0.979
John Ashcroft	-1.294	-0.266	Sammy Sosa	-1.702	0.074
Zacarias Moussaoui	-1.239	-0.908	George Ryan	-1.511	-0.789
John Muhammad	-0.979	-3.076	Lionel Tate	-1.112	-0.962
Lionel Tate	-0.962	-1.112	Esteban Loaiza	-1.108	0.019
Charles Taylor	-0.818	-0.302	Slobodan Milosevic	-0.964	-1.674
George Ryan	-0.789	-1.511	Charles Schumer	-0.949	0.351
Al Sharpton	-0.782	0.043	Scott Peterson	-0.937	-0.340
Peter Jennings	-0.781	-0.372	Zacarias Moussaoui	-0.908	-1.239
Saddam Hussein	-0.652	-0.240	William Jefferson	-0.720	-0.101
Jose Padilla	-0.576	-0.534	King Gyanendra	-0.626	-0.502
Abdul Rahman	-0.570	-0.500	Ricky Williams	-0.603	-0.470
Adolf Hitler	-0.549	-0.159	Ernie Fletcher	-0.580	-0.245
Harriet Miers	-0.511	0.113	Edward Kennedy	-0.575	0.330
King Gyanendra	-0.502	-0.626	John Gotti	-0.554	-0.253

Table 6: *Top negative entities in news (left) and blogs (right)*

Maria Sharapova) are also among the top positive people in blogs. Because the percentile ratings of news and blogs are not directly comparable, we report our results here in terms of net positive and negative sentiment.

Table 6 lists the most negative people appearing in newspapers and blogs. International (*Slobodan Milosevic*, *Zacarias Moussaoui*) and domestic criminal figures (*John A. Muhammed*, *Lionel Tate*, *George Ryan*) are regarded as losers in both blogs and newspapers. The blogs of angry fans reveal their extreme displeasure at certain sports figures (*Sammy Sosa*, *Esteban Loaiza*, *Ricky Williams*).

Most interesting are the distinct fates of certain controversial American political figures. Some (e.g. *Harriet Miers*, *Al Sharpton*) are regarded negatively in newspapers but positively in blogs, while others (e.g. *Charles Schumer*, *Edward Kennedy*) are thought of negatively only by bloggers. These clearly reflect political biases among bloggers, and perhaps the mainstream press.

6. Conclusion

There are many interesting directions that can be explored. We are interested in how sentiment can vary by demographic group, news source or geographic location. By expanding our spatial analysis of news entities [1] to sentiment maps, we can identify geographical regions of favorable or adverse opinions for given entities. We are also studying in analyzing the degree to which our sentiment indices predict future changes in popularity or market behavior.

References

[1] Mehler, A., Bao, Y., Li, X., Wang, Y., Skiena, S.: Spatial analysis of news sources. IEEE Trans. Visualization and Computer Graphics **12** (2006) 765–772

[2] Lloyd, L., Kaulgud, P., Skiena, S.: Newspapers vs. blogs: Who gets the scoop? In: Computational Approaches to Analyzing Weblogs (AAAI-CAAW 2006). Volume AAAI Press, Technical Report SS-06-03. (2006) 117–124

[3] Lloyd, L., Kechagias, D., Skiena, S.: Lydia: A system for large-scale news analysis. In: String Processing and Information Retrieval (SPIRE 2005). Volume Lecture Notes in Computer Science, 3772. (2005) 161–166

[4] Lloyd, L., Mehler, A., Skiena., S.: Identifying co-referential names across large corpra. In: Proc. Combinatorial Pattern Matching (CPM 2006). Volume Lecture Notes in Computer Science, 4009. (2006) 12–23

[5] Kil, J., Lloyd, L., Skiena, S.: Question answering with lydia. Proc. 14th Text Retrieval Conference (TREC 2005) (2005)

[6] Godbole, N., Srinivasaiah, M., Skiena, S.: Large-scale sentiment analysis for news and blogs (demonstration). In: Proc. Int. Conf. Weblogs and Social Media (ICWSM 07). (2007)

[7] Hatzivassiloglou, V., McKeown, K.R.: Predicting the semantic orientation of adjectives. In: Proc. 8th Conf. on European chapter of the Association for Computational Linguistics, Morristown, NJ, USA, Association for Computational Linguistics (1997) 174–181

[8] Wiebe, J.: Learning subjective adjectives from corpora. In: Proc. 17th Nat. Conf. on Artificial Intelligence and 12th Conf. on Innovative Applications of Artificial Intelligence, AAAI Press / The MIT Press (2000) 735–740

[9] Kim, S.M., Hovy, E.: Determining the sentiment of opinions. In: Proceedings of the Coling Conference. (2004)

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

[11] Pang, B., Lee, L.: A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the ACL. (2004) 271–278

[12] Nasukawa, T., Yi, J.: Sentiment analysis: Capturing favorability using natural language processing. In: The Second International Conferences on Knowledge Capture. (2003) 70–77

[13] J. Yi, T. Nasukawa, R.B., Niblack, W.: Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In: 3rd IEEE Conf. on Data Mining (ICDM'03). (2003) 423–434

[14] Miller, G.A.: Wordnet: A lexical database. Communications of the ACM **38**(11) (1995) 39–41

[15] Andreevskaia, A., Bergler, S.: Mining wordnet for a fuzzy sentiment: Sentiment tag extraction from wordnet glosses. In: EACL. (2006)

[16] Esuli, A., Sabastiani, F.: Determining term subjectivity and term orientation for opinion mining. In: EACL. (2006)

[17] Ide, N.: Making senses: Bootstrapping sense-tagged lists of semantically-related words. In: CICLING. (2006)

[18] Wiebe, J., Mihalcea, R.: Word sense and subjectivity. In: ACL. (2006)

[19] Wiebe, J.: Learning subjective adjectives from corpora. In: AAAI/IAAI. (2000) 735–740