# Opinion Learning from Medical Forums

**Tanveer Ali**
University of Ottawa
tali028@uottawa.ca

**Marina Sokolova**
University of Ottawa & CHEO
sokolova@uottawa.ca

**David Schramm**
University of Ottawa & CHEO
dschramm@ottawahospital.on.ca

**Diana Inkpen**
University of Ottawa
Diana.Inkpen@uottawa.ca

## Abstract

Our study focuses on opinion mining of several medical forums dedicated to Hearing Loss (HL). Surgeries related to HL are the most common surgeries in North America; thus, they affect many patients and their families. We have extracted the opinions of people from these forums related to stigma of HL, consequences of HL surgeries, living with HL, failures of HL loss treatments, etc. We performed a manual annotation first with two annotators and have 93% overall agreement with kappa 0.78 and then applied Machine Learning methods to classify the data into opinionated and non-opinionated messages. Using our feature set, we achieved best F-score 0.577 and 0.585 with SVM and logistic-R classifier respectively.

## 1 Introduction

The development of the Internet and of the user-friendly Web technologies profoundly changed the ways the general public can express their opinions on a multitude of topics. In order to make informed decisions, there is a necessity to develop methods that adequately – efficiently and effectively – extract new knowledge from the online messages (Bobicev et al., 2012). Opinions depend on individual's personality, culture and expectations of the society. Thus, opinions are challenging for independent external evaluation and categorization.

Natural language statements can be divided into two categories: facts and opinions. Facts can be expressed with topic keywords, while opinions are more difficult to express with a few keywords. They are the words of mouth on the web, e.g.,

Factual Sentence:
*Most things come in somewhere between 40 and 105, depending on the frequency.*

Opinionated Sentence:
*I don't think you will find anyone who this level of amplification is undamaging, but the option is to not hear.*

In this work, we have performed opinion mining of message posted on medical forums dedicated to Hearing Loss. Surgeries related to HL are the most common surgeries in North America; thus, they affect many patients and their families. Our current work aims to provide a tool that can extract opinions expressed by the general public. Understanding of what people think about the surgeries and their consequences helps health care providers to develop better health care policies and the general public outreach.

We collected data from web forums and we invited two annotators to manually annotate texts gathered from medical forums. We obtained the overall agreement of 93% and kappa was 0.78. Then we used a subjectivity lexicon and machine learning algorithms to automatically classify the posts. Our experiments with different combinations of features using different classifiers, i.e., Naïve Bayes, SVM and Logistics-R have shown significant improvement in F-score performance (55.7%, 56.8% and 57.8%, respectively) over the majority class baseline, which was 47.6%.

## 2 Related Work

A very limited work has been done on opinion mining on health related forums. Sokolova and Bobicev (2011) analyzed opinions posted on a general medical forum (i.e., the forum where the users discussed different health problems). The messages discussed health-related topics: medications, treatment, illness and cure, etc. The authors constructed a set of sentences manually labeled as positive, negative and neutral opinions. Among the three opinion categories, better results were obtained for the negative category (kappa = 0.365). For external evaluation of the labeling results, Machine Learning methods were applied on the annotated data. The best F-score = 0.839 was achieved by SVM. However, the authors used a small and imbalanced dataset, i.e., 169 positive and 74 negative sentences. Thus, the data had an inheritably high major class baseline of Accuracy = 70% and F-score = 57%. In our case, we used a considerably bigger and completely balanced data set having 93% overall



*Proceedings of Recent Advances in Natural Language Processing*, pages 18–24,
Hissar, Bulgaria, 7-13 September 2013.

agreement and 0.78 kappa between two annotators, with the majority class baseline of accuracy = 50% and F-score = 47.6%.

In (Goeuriotet al., 2012), the authors have built a medical domain lexicon in order to perform classification on a dataset that they collected from a website called Drug Expert. The dataset contains user reviews on drugs with ratings from 0 to 10 (negative to positive). The authors have performed the polarity detection on this dataset which already contains subjective information (opinions) about users' experience with particular drugs. However, in our case, we have extracted messages from health forums which publish both opinionated and non-opinionated posts.

## 3 Building the Dataset

We wanted our data be specific to the problem at hand. This is why we concentrated only a few health forums dedicated to Hearing Loss (HL). Although the very specific topic prevented us to have access to a high volume of data, at the same time, focusing on relevant forums only helped us to reduce the volume of unrelated messages. Also, we wanted to analyze the forum discussions, i.e., threads, which consist of more opinionated messages rather than questions and answers about the medical problems.

For the opinion mining, we have chosen a critical domain of HL problems: opinions about Hearing Aids. To the best of our knowledge, no relevant previous work was done in this area. For our dataset; we have collected individual posts from 26 different threads on three health forums[1].

### 3.1 Data Description

The initial collection of data contains about 893 individual posts from 34 threads. They were extracted using the XPath query by using the Google Chrome extension "XPathHelper".

This data was filtered and reduced to 26 threads by removing the threads in which people did not discuss Hearing Aids. The threads contained 607 posts in them. Table 1 lists the forum web sites, the number of threads collected from each forum, the number of posts gathered from each forum, and an average number of posts written by each author.

[1] http://www.medhelp.org,
http://www.alldeaf.com,
http://www.hearingaidforums.com

| Forums | Threads | Posts | Avg. posts per person |
|---|---|---|---|
| www.hearingaidforums.com | 7 | 185 | 2.9 |
| www.medhelp.org | 9 | 105 | 2.77 |
| www.alldeaf.com | 10 | 317 | 1.93 |
| Total | 26 | 607 | 2.53 |

**Table 1. Filtered dataset collection statistics**

We split the data from individual threads into sentences using our version of a regular expression based sentence splitter. We partly removed noise from the text by removing sentences containing very few words (4 in our case) as they did not convey well-formed opinions, for example:

Sentence: *No, educate me.*
Sentence: *Max AVERAGE SPL.*
Sentence: *Am I right ?*
Sentence: *It is permanent.*

The remaining sentences from the 26 threads were manually annotated by two independent annotators into two classes (opinionated and non-opinionated). There were several categories of opinionated and non-opinionated sentences. We provide the examples below.

**Non-opinionated about Hearing Aids:**

**Factual on Hearing Aids:**

*So a doubling of 'power' equates to a 3dB rise in measured output.*

**Not relevant to Hearing Aids:**

*Lots of jobs in that field and I was pleased that I have met all of the qualifications.*

**Opinionated about Hearing Aids:**

**Positive**

*The aids you see discussed on this forum are designed with limiting factors intended to keep sound from being amplified to damaging levels.*

**Neutral/Unknown**

*I have yet to see an ENT indicate that properly adjusted hearing aids will either cause or not cause ear damage.*

**Negative**

*"I was referring to perception and in my understanding, even a duration of a few minutes can damage the ears."*

In this paper, however, we work only with two broad message categories: opinionated about Hearing Aids and non-opinionated about them.

### 3.2 Subjectivity Lexicon

For our experiments, we used the Subjectivity Lexicon (SL) built by Wilson, Wiebe, and Hoffman (2005). The lexicon contains 8221 subjective expressions manually annotated as strongly or weakly subjective, and as positive, negative, neutral or both. We have chosen this lexicon over other large automatically generated dictionaries like SentiWordNet (Baccianella, Esuli, and Sebastiani, 2010), as it has been manually annotated and provides rich information with the subjectivity strength and prior polarity for each word considering the context of the word in the form of part of speech information.

The quality of this Subjectivity Lexicon is higher than the quality of other large automatically generated dictionaries; for example, SentiWordNet (Baccianella, Esuli, and Sebastiani 2010) includes more than 65,000 entries. Some papers (Taboada et al., 2011) have shown that larger dictionaries contain information which is not detailed and include more words which may lead to more noise.

Below is the sample entry from the lexicon:

*type=strongsubj len=1 word1=boundless pos1=adj stemmed1=n priorpolarity=positive*

This entry contains the term *boundless*, which is an adjective. Its length is 1 (single term), it is not stemmed; it is strongly subjective and positive. Similarly following are other entries from lexicon:

*type=weaksubj len=1 word1=buckle pos1=verb stemmed1=y priorpolarity=negative*

*type=strongsubj len=1 word1=desiccated pos1=adj stemmed1=n priorpolarity=negative*

Table 2 shows the relation between strong and weak subjectivity with the polarity lexicon.

| | Strong Subj | Weak Subj | Total | Percent |
|---|---|---|---|---|
| Positive | 1717 (30.8%) | 1001 (37.74%) | 2718 | 33.06 |
| Negative | 3621 (65%) | 1291 (48.6%) | 4912 | 59.75 |
| Neutral | 231 (4.14%) | 360 (13.57%) | 591 | 7.18 |
| Total | 5569 | 2652 | 8221 | 100 |
| Percent | 67.74 | 32.26 | 100 | |

**Table 2. Distribution among subjectivity and polarity in the lexicon**

## 4 Methodology

In this work, we have used several different features for the opinion mining of the sentences. Section 4.1 discussed the use of parts of speech in opinion mining. Section 4.2 lists all these features. These features are computed and presented for each sentence in a data file format used by the WEKA suite (Hall et al., 2009). Classification is performed based on the computed features and accuracy is measured using for different combinations of features in order to improve the classification performance.

### 4.1 Lemmatization

For all nouns and verbs, we have used the lemmatization using the GATE [2] morphological plugin which provides the root word. In case of noun the root word is the singular form of the plural noun, e.g., bottles becomes bottle, etc. In the case of verbs, the plugin provides the base form for infinitive, e.g., helping becomes help, and watches become watch. After performing lemmatization, we found 158 more words that were detected with same part of speech considered as the original. There were still 175 words which were found with the root word in the lexicon, but with different part of speech, e.g., *senses* was used as nouns in the data, after lemmatization it becomes *sense*, which exists as verb in the lexicon. Therefore it cannot be matched as the context and meaning of the word is different.

### 4.2 Features

All the features considered for the experiment are based on sentence level. Table 3 shows the final features selected for the experiments. The most common features were pronouns, followed by weak subjective clues, adjectives and adverbs.

| | |
|---|---|
| STRONGSUBJ | # of words found as strong subjective in current sentence |
| WEAKSUBJ | # of words found as weak subjective in current sentence |
| ADJECTIVE | # of adjectives |
| ADVERBS | # of adverbs |
| PRONOUN | # of pronouns |
| POSITIVE | # of words found having prior polarity as positive |
| NEGATIVE | # of words found having prior polarity as negative |
| NEUTRAL | # of words found having prior polarity as neutral |
| PRP_PHRASE | # of phrases containing pronouns found in current sentence |

**Table 3. Final features considered for the experiments**

[2] http://gate.ac.uk/sale/tao/splitch21.html#x26-52600021.11

## 5 Experiments

### 5.1 Manual Annotation

The dataset of 3515 sentences from 26 threads were manually annotated by two annotators. The annotators were asked to tag a sentence as opinionated if it conveys positive, negative or mixed opinions on hearing aids. All the sentences which do not contain any opinions are left blank and they are considered as non-opinionated. According to Table 4, annotator1 and annotator 2 did not put the opinionated label a large number of sentences, i.e., 2939 and 2728 respectively. We further considered them as non-opinionated.

| | Annotator 1 | | |
|---|---|---|---|
| Annotator 2 | Opinionated | Non-opinionated | Total |
| Opinionated | 557 | | 787 |
| Non- | | 557 | 2728 |
| Total | 576 | 2939 | 3515 |

**Table 4. Annotations statistics of Sentences between the two annotators**

To evaluate the annotator agreement, we calculated *kappa* as in (Sokolova & Bobicev, 2011):

$$\text{kappa} = \frac{\frac{a+d}{N} - \frac{f1g1+f2g2}{N^2}}{1 - \frac{f1g1+f2g2}{N^2}}$$

The overall percentage agreement between the annotators for the dataset was 93% and kappa was 0.78. This indicates a substantial agreement between the taggers in both the cases.

### 5.2 Dataset preprocessing

Due to the large number of irrelevant sentences, the dataset is very much imbalanced. A balanced dataset is necessary for accurate classification, as in the case of imbalanced dataset as this, if all sentences are considered as non-opinionated, the accuracy of the system is very high (83%), as the non-opinionated class dominates the opinionated class in the dataset. To be exact, there are 557 opinionated sentences and 2728 non-opinionated sentences. For this purpose, we reduce the non-opinionated sentences by applying a version of the under-sampling technique (Barandela et al., 2004).

In contrast with a commonly applied random under-sampling, our under-sampling method selects only certain sentences to keep them in the data set. For each occurrence of an opinionated sentence, the next non-opinionated sentence is chosen to be kept, and the rest are discarded. The final dataset contains 1152 total sentences with 576 opinionated and non-opinionated sentences each.

### 5.3 Classification results

The output files generated by the system for both the datasets are classified using the WEKA (Hall et al., 2009). For our evaluation, we used 10-fold cross validation which is a standard classifier selection for classification purpose. Experiments were performed using three different classifiers: Naïve Bayes, support vector machine (SVM) and logistic regression (logistic-R). Performance was evaluated using the F1-measure between the three classifiers on the given datasets. The best performance for Naïve Bayes and support vector machine were 55.7% and 56.7% respectively with (strongsubj, weaksubj) feature. With Logistics-R the best performance was 57.8% with (strongsubj, weaksubj, pronoun) feature. It was found that the performance of logistic regression was the best on the features selected for our evaluation.

For the baseline, we considered the majority class baseline having 50% accuracy and achieved F-score 47.6%. For the gold classification standard, the feature vector of bag of words is considered. We have not considered the unique words for the bag of words because eliminating the words that appeared only once reduces the size of the vectors to half, and it makes it easier for the classifier to handle them. Also, these words do not contribute much to the post classification since they appear only once, i.e., in one post, and cannot be used to analyze other posts. From experiments, it was found that the gold standard result for our dataset was rather high for each classifier. Still, all the classifiers improved the results over the majority class baseline.

| Opinionated vs. non-opinionated classification | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Naive Bayes** | | | **SVM** | | | **Logistic-R** | | |
| | P | R | F-1 | P | R | F-1 | P | R | F-1 |
| strongsubj,weaksubj | 0.599 | 0.579 | **0.557** | 0.602 | 0.585 | **0.567** | 0.573 | 0.572 | 0.57 |
| strongsubj,weaksubj,neutral | 0.593 | 0.573 | 0.548 | 0.603 | 0.586 | 0.568 | 0.568 | 0.567 | 0.566 |
| **strongsubj,weaksubj,pron** | 0.583 | 0.565 | 0.539 | 0.586 | 0.574 | 0.557 | 0.585 | 0.582 | **0.578** |
| all features | 0.600 | 0.578 | 0.554 | 0.584 | 0.571 | 0.554 | 0.574 | 0.571 | 0.566 |
| Gold Standard | 0.628 | 0.626 | 0.624 | 0.628 | 0.626 | 0.624 | 0.590 | 0.590 | **0.589** |

**Table 5. Comparison of performance between different features among three classifiers**

Table 5 shows that the improvement was 8.1% for Naïve Bayes, 9.2% for SVM and 10.2% for logistic-R. We evaluated different sets of features for the classification performance. Table 5 shows that the best performance of all classifiers was with different feature sets, as for Naïve Bayes it was with (strongsubj, weaksubj) at 55.7%, for SVM it was with (strongsubj, weaksubj, neutral) at 56.8% and for logistic-R it was with (strongsubj, weaksubj, pron) at 57.8%. It was assumed that neutral word clues should indicate non-subjectivity, as they are neutral in polarity; however, the results did not show improvement with neutral features. This may be due to very limited neutral words in the lexicon, i.e., only 7.18%. The best classifier was logistic regression with the feature set (strongsubj, weaksubj, pron) with F1-measure 57.8%, which is slightly lower than the gold standard of 58.9% with logistic-R.

| Opinionated vs. non-opinionated classification with lemmatization | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Naive Bayes** | | | **SVM** | | | **Logistic-R** | | |
| | P | R | F-1 | P | Re | F-1 | P | R | F-1 |
| **Strongsubj,weaksubj,prp_phrase** | **0.596** | **0.58** | **0.562** | **0.604** | **0.591** | **0.577** | **0.586** | **0.58** | **0.57** |
| strongsubj,weaksubj | 0.604 | 0.58 | 0.554 | 0.605 | 0.591 | 0.576 | 0.584 | 0.58 | 0.57 |
| strongsubj,weaksubj,neutral | 0.600 | 0.582 | **0.562** | 0.597 | 0.583 | 0.568 | 0.584 | 0.58 | 0.58 |
| strongsubj,weaksubj,pron | 0.602 | 0.578 | 0.552 | 0.586 | 0.575 | 0.561 | 0.592 | 0.58 | **0.58** |
| all features | 0.602 | 0.58 | 0.556 | 0.593 | 0.582 | 0.569 | 0.582 | 0.57 | 0.57 |
| Gold standard | 0.628 | 0.626 | 0.624 | 0.628 | 0.626 | 0.624 | 0.590 | 0.59 | **0.58** |

**Table 6. Comparison of performance with lemmatization between different features among three classifiers**

As most opinions are expressed with the use of personal pronouns, we extracted the phrases that contain pronouns within sentences, e.g., I would assume, I feel as, I could sympathize. We consider the number of such phrases within sentences and evaluated the performance using combinations with other features. Also, to increase the number of matched words in the lexicon, all the nouns and verbs were lemmatized to see if the classification performance increases. The classification results show improvement for all the classifiers. It is interesting to note that Naïve Bayes and SVM both have shown their best performance with the feature combining subjectivity clues and phrases with pronouns, which indicate the significance of pronouns for subjectivity; however logistics-R performed best with subjectivity and phrases with pronoun features, but in this case pronoun phrase features show the 2nd best performance.

The classification performance in Table 6 increased with Naïve Bayes, SVM and logistic-R with 0.5%, 0.9% and 0.7%, respectively. Also note that the gold standard representation exceptionally performed better with Naïve Bayes and SVM, but with the logistic-R it was relatively comparable to our previous results and the performance with best features (strongsubj, weaksubj, pron) was just 0.4% less than the gold standard; so the results with (strongsubj, weaksubj, pron) are equivalent with the gold standard.

## 6 Analysis

The results from the experiments have provided various insights about opinion mining in health-related forums. For classification, the bag-of-words representation provided higher results than the other feature sets. We interpret this result an indication of the importance of the word meaning. The words were more important than their semantic orientation or polarity. We noticed that the subjectivity clues such as strong subjective or weak subjective labels from the lexicon have not increased the performance for identifying opinionated and non-opinionated sentences; they performed equivalently to the gold standard (i.e., bag-of-words). Also note that the bag-of-word representation (BOW) is a high gold standard that is hard to beat in many texts classification problems. In our case, a simple baseline of classifying every sentence into the most frequent class is outperformed by the BOW representation by 13.6% on average among all the three classifiers. This difference indicates how difficult the opinion mining task is. The personal pronouns such as *I, me, ours, yours*, etc. also play an important role, as these are commonly found in subjective sentences and the results have shown some improvement for features with pronouns. However, subjective clues and phrases that contain pronouns can lead to false prediction, e.g.:

**Sentence 1:**
***I can understand*** *that once the lost gain has been reapplied, techniques such as compression can reduce the additional amount of SPL DB that is required.*

**Sentence 2:**
***I understand*** *you will have to practice for some time with any type of hearing aid.*

Sentence 1 from our data is labeled by both annotators as non-opinionated but it contains ***understand*** which is strong subjective in lexicon; also ***I can understand*** contains a pronoun. At the same time, Sentence 2 contains the same strong subjective word and the same pronoun, but it is labeled by both annotators as opinionated in the data. It has been noted that **understand** has occurred more in non-opinionated sentences, which in part provides the reason for the high performance of the baseline.

Our results are comparative to other related studies. We achieved Precision = 0.604, Recall = 0.591 and F-score = 0.577 with (strongsubj,weaksubj,prp_phrase) feature set using the support vector machine classifier.

In general, for consumer reviews, opinion-bearing text segments are classified into positive and negative with Precision 56%−72% (Hu & Liu 2004). For online debates, the complete texts (i.e. posts) were classified as positive or negative stance with F-score 39%−67% (Somasundaran & Wiebe, 2009); when those posts were enriched with preferences learned from the Web, F-score increased to 53%−75%.

## 7 Conclusion and Future Work

In this work, we performed opinion mining of online messages related to Hearing Loss. We used several lexicon-based features together with the rule based features like pronoun phrases classification of opinionated and non-opinionated sentences. As categories, we considered sentences being opinionated if they contained opinions about Hearing Aids. Other sentences were considered as non-opinionated. Evaluations have been made using three different classifiers and it is shown that our proposed features outperformed the baseline classifier which uses only bag-of-word features.

In future work, we could use structural features, dialogue act features, and sentiment features (Biyani & Bhatia, 2012) for the subjectivity classification of sentences. The lexicon could be improved, as the domain lexicon created in (Goeuriot et al., 2012) has shown better results over other dictionaries for polarity detection.

## Acknowledgements

This work was supported by NSERC and by the CHEO Department of Surgery Research. We thank Brian Dewar for his assistance in finding medical forums.

## References

Baccianella, S., Esuli, A., &Sebastiani, F. (2010, May).*Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining.*In Proceedings of the 7th conference on International Language Resources and Evaluation (LREC'10), Valletta, Malta, May.

Barandela, R., Valdovinos, R. M., Sánchez, J. S., & Ferri, F. J. (2004). *The imbalanced training sample problem: Under or over sampling?*. In Structural, Syntactic, and Statistical Pattern Recognition (pp. 806-814). Springer Berlin Heidelberg.

Biyani, P., Caragea, S. B. C., & Mitra, P. (2012) *Thread specific features are helpful for identifying subjectivity orientation of online forum threads.*COLING.

Bobicev, V., Sokolova, M., Jafer, Y., & Schramm, D. (2012).*Learning sentiments from tweets with personal health information.*In Advances in Artificial Intelligence (pp. 37-48).Springer Berlin Heidelberg.

Eysenbach, G. (2009). *Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet.* Journal of medical Internet research, 11(1).

Gillick, D. (2009, May). *Sentence boundary detection and the problem with the US.* In Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers (pp. 241-244). Association for Computational Linguistics.

Goeuriot, L., Na, J. C., Min Kyaing, W. Y., Khoo, C., Chang, Y. K., Theng, Y. L., & Kim, J. J. (2012, January). *Sentiment lexicons for health-related opinion mining.*In Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium (pp. 219-226).ACM.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009).*The WEKA data mining software: an update.* ACM SIGKDD Explorations Newsletter, 11(1), 10-18.

Hu, M., & Liu, B. (2004, August).*Mining and summarizing customer reviews.*In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 168-177).ACM.

Kennedy, A., & Inkpen, D. (2006).*Sentiment classification of movie reviews using contextual valence shifters.* Computational Intelligence, 22(2), 110-125.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003).*Lying words: Predicting deception from linguistic styles.* Personality and Social Psychology Bulletin, 29(5), 665-675.

Rhodewalt, F., & Zone, J. B. (1989). *Appraisal of life change, depression, and illness in hardy and nonhardy women.* Journal of Personality and Social Psychology, 56(1), 81.

Sokolova, M., &Bobicev, V. (2011).*Sentiments and Opinions in Health-related Web messages.*In Recent Advances in Natural Language Processing (pp. 132-139).

Somasundaran, S., &Wiebe, J. (2009, August).*Recognizing stances in online debates.* In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1 (pp. 226-234). Association for Computational Linguistics.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., &Stede, M. (2011).*Lexicon-based methods for sentiment analysis.* Computational linguistics, 37(2), 267-307.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005, October).*Recognizing contextual polarity in phrase-level sentiment analysis.*In Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing (pp. 347-354).Association for Computational Linguistics.

Yi, J., Nasukawa, T., Bunescu, R., &Niblack, W. (2003, November). Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Data Mining, 2003.*ICDM 2003. Third IEEE International Conference on (pp. 427-434).