Proceedings of the 2013 IEEE 17th International Conference on Computer Supported Cooperative Work in Design

Sentiment Analysis on Tweets for Social Events

Xujuan Zhou
Faculty of Science and Engineering
Queensland University of Technology
Brisbane, Australia
x.zhou@qut.edu.au

Xiaohui Tao, Jianming Yong
Centre for Systems Biology,
School of Information Systems
University of Southern Queensland, Australia
{xtao, jianming.yong}@usq.edu.au

Zhenyu Yang
Department of Energy Technology
Aalborg University
Denmark
yang@et.aau.dk

***Abstract*—Sentiment analysis or opinion mining is an important type of text analysis that aims to support decision making by extracting and analyzing opinion oriented text, identifying positive and negative opinions, and measuring how positively or negatively an entity (i.e., people, organization, event, location, product, topic, etc.) is regarded. As more and more users express their political and religious views on Twitter, tweets become valuable sources of people's opinions. Tweets data can be efficiently used to infer people's opinions for marketing or social studies. This paper proposes a Tweets Sentiment Analysis Model (TSAM) that can spot the societal interest and general people's opinions in regard to a social event. In this paper, Australian federal election 2010 event was taken as an example for sentiment analysis experiments. We are primarily interested in the sentiment of the specific political candidates, i.e., two primary minister candidates - Julia Gillard and Tony Abbot. Our experimental results demonstrate the effectiveness of the system.**

Keywords—Sentiment analysis; Tweets; Text analysis; Social network

I. Introduction

On Web 2.0, user-generated content, which is material submitted by users who interact with social network sites, is a major theme. Twitter is a social networking and micro-blogging service where users send messages (a.k.a., tweets) to a network of associates from a variety of devices. A tweet is a text-based post and only has 140 characters, which is approximately the length of a typical newspaper headline and subhead [11]. The short massages are very easy and convenience to both sender and reader to share things of interest and communicate their thoughts anywhere and anytime in the world. Twitter is a "what's-happening-right-now" social network [4] hence it can offer immediate sentiment.

Twitter's user base has grown rapidly and the volumes of messages produced by Twitter everyday is vast. According to [7], in April 2010, Twitter had 106 million registered users,180 million unique visitors every month, 3 billion requests per day based on its API, and 300,000 new users were signing up per day. Tweetrush (tweetrush.com) estimates traffic at approximately a million tweets a day. As more and more users post reviews about products and services they use, or express their political and religious views on Twitter, tweets become valuable sources of peoples opinions and sentiments. Tweets data can be efficiently used to infer people's opinions for marketing or social studies. Given its popularity, Twitter is seen as a potential new form of eWOM (electronic word-of-mouth) marketing by the businesses and organizations concerned with reputation management [6].

Sentiment analysis (or opinion mining) is stated as "the computational study of opinions, sentiments and emotions expressed in text" by Liu [10]. It is an exciting new research field with the potential for a number of real world applications where discovered opinion information can be used to help people or companies or organiztions to make better decisions. Currently, many of sentiment analysis works are focus on product reviews or movie review [18], [20], [14], [3], [2] on blogs, customer review sites, and WebPages. As the largest, most well-known, and most popular of the micro-blogging sites, Twitter is an ideal sources for spotting the information about societal interest and general peopler's opinions. However, there has been little prior opinion mining work in the microblogging area since Twitter is relative new technology. Opinion mining in Twitter is different from the opinion mining from the blogs, review sites or other Webpages. Reviews tend to be longer and more verbose than tweets which may only be a few words long and often contain significant spelling errors. Reviews usually focus on a specific product or entity and contains little irrelevant information. However, tweets tend to be much more diverse in terms of topics with issues ranging from politics and recent news to religion.

In this study, we focus on the tweets sentiment analysis that is to automatically identify whether a piece of text expresses a positive or negative opinion about an entity (i.e., politician) for an election event in politic domain.

The primary objective of this study is to develop a new Tweets Sentiment Analysis Model (TSAM) that can:

- provide early indications of topics and entities for which societal interests are emerging or may emerge.

- predict the developing trend of an event within a specified time period.
- provide a fast and less expensive alternative to traditional polls (e.g., telephone poll) for mining public opinion.

The remainder of the paper is organized as follows. Section 2 provides a brief review of related work. Section 3 illustrates the sentiment analysis processing framework and technique. The experimental results and limitation discussions are reported in Section 4. Finally, concluding remarks are sketched in Section 5.

II. RELATED WORK

A. Sentiment Classification

One of important topics in sentiment analysis is sentiment classification that classifies the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative or neutral. However, sentiment classification is different from classic topic-based text classification. For the traditional topic-based classification, topics are often identifiable by keywords alone whereas sentiment can be expressed in a more subtle manner. For example, the sentence "How could anyone sit through this movie?" contains no single word that is obviously negative. Thus, sentiment seems to require more understanding than the usual topic-based classification [12].

Sentiment classification at sentence-level is different from that at document-level. A document can be more or less opinionated, whereas a sentence can be only subjective or objective. Thus, comparing with the document-level sentiment classification techniques, sentence-level sentiment classification has one more task to do - it needs to filter out the sentences containing no opinion before classifying the contained opinions on the objects and their features to positive and negative [10] . Sentence-level sentiment classification needs to mine opinions from both subjective and objective sentences, because objective sentences may also imply opinions [10] .

Many supervised and unsupervised classification methods have been used in sentiment classification. Wiebe and Riloff [17] train sentence level subjectivity and objectivity classifiers using a subjective term list and subjective/objective extraction patterns as features. They bootstrap the patterns in an extraction pattern learning and supervised sentence classification cycle. Turney [15] used an unsupervised classification method to classify opinionated text based on fixed syntactic opinion phrases . The method first extracts opinionated phrases (e.g. those containing adjectives or adverbs) from an opinionated text, and then estimates the orientation of these extracted phrases. Finally, by using the phrases with their opinion orientation values, the method classifies the text into positive and negative.

The lexicon-based classifier is a typical example of an unsupervised approach, because it can function without any reference corpus and doesn't require any training (i.e. can be applied "off-the-shelf"). Dictionary/lexicon-based sentiment analysis is typically based on lists of words with some sort of pre-determined emotional weight. Examples of such dictionaries include the General Inquirer (GI) dictionary [19]. The lexicons are build with the aid of "experts" that classify certain tokens in terms of their affective content (e.g. positive or negative). The "Affective Norms for English Words" (ANEW) lexicon [1] contains ratings of terms on a nine-point scale in regard to three individual dimensions: valence, arousal and dominance. The ratings were produced manually by psychology class students.

B. Entity Discovery and Extraction

People's opinions, sentiments and emotions are expressed against an object. Such an object is also the target that a sentiment analysis and opinion mining technique serves. The problem of sentiment analysis and opinion mining were formalized by Liu in [9] : "Given a set of evaluative text documents *D* that contain opinions (or sentiments) about an object, sentiment analysis and opinion mining aim to extract attributes and components of the object that have been commented on in each document *d* in *D* and to determine whether the comments are positive, negative or neutral." He [10] defined an object as : "An object *o* is an entity which can be a product, person, event, organization, or topic. It is associated with a pair, o : (T,A), where T is a hierarchy of components (or parts), sub-components, and so on, and A is a set of attributes of o. Each component has its own set of sub-components and attributes".

For sentiment analysis, besides distinguishing between positive and negative opinions, identifying which entities correlated with which opinions are also important because without knowing which entity each sentence talks about the opinion mined from the sentence is meaningless. There are some researches on entity discovery and extraction. Open Calais is one of well-known entity extraction tools. Open Calais extracts entities from textual (natural language) input and returns an XML document contains meta-information about entities in RDF format, including name and type. Detailed information can be found at http://www.opencalais.com.

III. CONCEPTUAL FRAMEWORK AND SENTIMENT ANALYSIS TECHNIQUE

The Tweets Sentiment Analysis Model (TSAM) automatically analyses tweets data. It can identify the positive, negative or neutral opinions and measure intensity (or strength) of positive/negative opinions in regard to an entity (people, organization, location, product, etc.). The conceptual framework of the TSAM consists of three modules:

- Feature selection module that extracts the opinionated words from each sentence.
- Sentiment identification module that associates expressed opinions with each relevant entity in each sentence level.
- Sentiment aggregation and scoring module that calculates the sentiment scores for each entity.

Fig. 1 illustrates the TSAM's conceptual framework. The details of each module are discussed in the following sections.

A. Features Extraction

Most work on sentiment analysis has relied on traditional "bag-of-words" method, which attempts to learn a positive or negative document classifier based on occurrence frequencies of the various words in the document. In our sentiment analysis model, instead of using all the words appearing in the news articles or tweets, we only extract the opinion-bearing words as the features to input into opinion mining algorithm. Opinion words that are primarily used to express subjective opinions in the opinion sentence are identified and extracted. Words that encode a desirable state (e.g., beautiful, awesome) have a positive orientation, while words that represent undesirable states have a negative orientation (e.g., disappointing).

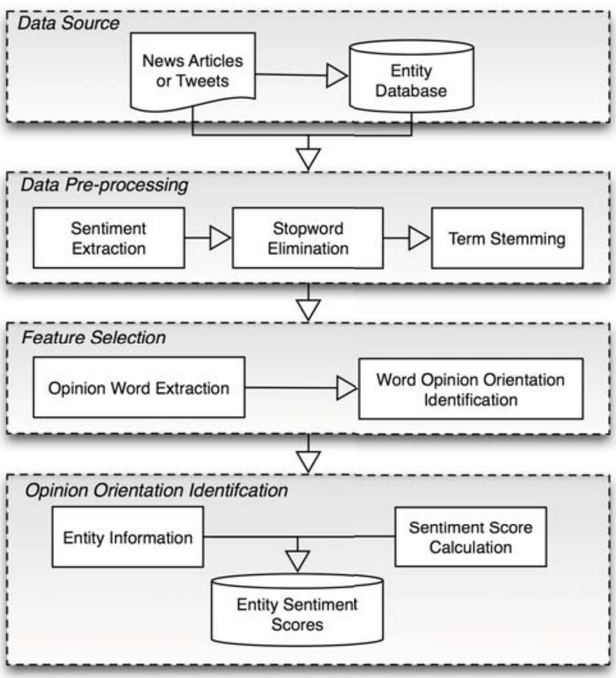

Fig. 1. The conceptual framework

There are those words that have no orientation (e.g., external, digital). The semantic orientation of a word will be used to predict the semantic orientation of each opinion sentence. Opinion sentence is a sentence that contains one or more entities and one or more opinion words.

To identify the opinionated words, we use Wilson opinion lexicon list [19] to decide the words' semantic orientations. This list is a lexical resource of sentiment information for words, where each word is associated with positive, negative and neutral sentiment information. Wilson lexicon consists of three lists of subjectivity clues: (i) the prior polarity lexicon, (ii) the intensifier lexicon, and (iii) the valence shifter lexicon. All parts contain unigram as well as n-gram entries with POS and stemming attributes. The POS attribute indicates the POS of the subjectivity term. The stemming attribute indicates whether the look-up should be performed with lemmas or tokens. For instance, the look-up for the lexicon entry (word1 = abuse pos1 = verb stemmed1 = y) should be performed with lemmas and match all the verb instances of the entry like "abused" (verb), "abusing" (verb), but not "abuse" (noun) or "abuses" (noun). Entries of the prior polarity lexicon additionally have the prior polarity and reliability attributes. Prior polarity represents the polarity of an entry out of context with the possible values of positive, negative, both or neutral. The reliability attribute indicates whether the entry has a subjective usage most of the time (strongsubj), or whether it has only certain subjective usages (weaksubj). The intensifier lexicon contains a list of intensifier words such as "fierce, enormous, more, most". The valence shifter lexicon contains entries which shift the polarity of an existing opinion towards negative or positive including negation words.

In this project, only the prior polarity lexicon subjectivity clue is used. We quantify the semantic orientation of words by given each type of word a numeric score. Therefore, a positive and strong subjectivity words is assigned the semantic orientation score of +1, a positive and weak subjectivity word is assigned the semantic orientation score of +0.5, and a negative and strong subjectivity word is assigned the semantic orientation score of −1, a negative and weak subjectivity word is assigned the semantic orientation score of −0.5, and a neutral word is given the semantic orientation score of 0. These text strings can be placed into categories (positive, negative, neutral) and one can differentiate their strength or impact by assigning different weights. For example, the word bankruptcy can carry a stronger weight value than lawsuit even though they both might fall under the category Negative.

B. Sentiment Analysis Technique

Given a set of tweets, T , that contains a set of sentences, s, $T = \{s_1, s_2, \ldots, s_i\}$; and each sentence s_k describes something on a subset of entities $e = \{\mathrm{e_i}, \ldots, \mathrm{e_j} | \mathrm{e_i}, \mathrm{e_j} \in \mathrm{E}\}$, where E is the set of all entities. An entity can be a person, an organisation, a location, a product, etc. Each sentence also contains a set of opinion word, w, $s = \{w_1, w_2, \ldots, w_j\}$. At first, a Sentence Sentiment Scoring Function (SSSF) is used to determine the orientation of sentiment expressed on each entity e_i in s (i.e., the pair of (e_i , s)). Then an Entity Sentiment Aggregation Function (ESAF) is used to obtain the total sentiment scores for an given entity e_i.

1) *Sentence Sentiment Scoring Function*: in this stage, the classification algorithm detects all words that belong to Wilson lexicon list and extracts their polarity. Adjectives are good indicators of sentiment and have been used as features for sentiment classification by a number of researchers [8], [16], [5]. However, it does not necessarily imply that other parts of speech do not contribute to expressions of opinion or sentiment. In fact, nouns (e.g., "gem") and verbs (e.g., "love") can be strong indicators for sentiment. Therefore, in this study, we use all the parts of speech. We summed up the semantic orientation score of the opinion words in the sentence to determine the orientation of the opinion sentence. The score function for a sentence is as follow:

$$score(s) = \sum_{w_j : w_j \in s \wedge w_j \in WL} \frac{w_j \cdot sentOri}{dis(w_j, e_i)} \qquad (1)$$

where w_j is an opinion word, WL is the set of all opinion words from Wilson lexicon list and s is the sentence that contains the entity e_i, and $dis(w_j, e_i)$ is the distance between entity e_i and opinion word w_j in the sentence s, and $w_j.sentOri$ is the semantic orientation of the word w_j (i.e., +1, or +0.5, or 0, or −1, or −0.5). If a sentence contains more than one entity then the opinion word close to the entity has smaller value of $dis(w_j, e_i)$ and indicates this word makes more contribution to that entity's sentiment scores.

The *scores*(*s*) is normalized by the number of the opinion words, *n,* in the sentence to reflect the sentiment scores distributions of opinion words. So, normalized sentiment score will be:

$$score(s)_N = {score(s)}/{n} \qquad (2)$$

2) *Entity Sentiment Aggregation Function*: in the given set of tweets, an entity appears in the set of sentences $s = \{s_1, s_2, ..., s_i\}$. 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.

For a given entity e_i, which may appear in multiple sentences $\{s_1, s_2, ..., s_i\}$, the normalized sentiment score for this entity in a sentence s_k is $score(e_i, s_k)_N$. The total sentiment scores of this entity will be aggregated by Entity Sentiment Aggregation Function that is depicted as below:

$$score(e_i) = \sum_{(s_k : s_k \in s)} score(s_k)_N \qquad (3)$$

This score is normalized by the number of the sentences, *m*, and then the final sentiment score for an entity will ranges in the interval [+1, −1].

$$score(e_i)_N = {score(e_i)}/{m} \qquad (4)$$

In regard to sentiment intensity (or strength) for a given entity, e_i, appears in the sentences, the following heuristic rule is applied:

$$intensity(e_i) = \begin{cases} \textbf{SP} & \text{if } (+0.5 < score(e_i)_N < +1) \\ \textbf{P} & \text{if } (0 < score(e_i)_N < +0.5) \\ \textbf{Neu} & \text{if } (score(e_i)_N = 0) \\ \textbf{Neg} & \text{if } (-0.5 < score(e_i)_N < 0) \\ \textbf{SN} & \text{if } (-1 < score(e_i)_N < -0.5) \end{cases}$$

- SN (*Strong Negative*) Sentences about the entity e_i contain purely negative words or phrases or only allowed a slightly positive word.
- N (*Negative*) Sentences contain mainly negative phrases and words. There may be a few positive words, but the negative words or phrases outweigh the positive ones.
- Neu (*Neutral*) Sentences have a mediocre or balanced sentiment. The positive and negative words or phrases seem to balance each other, or it is neither positive nor negative overall. Even if there are more negative phrases, the positive ones use a stronger language than the negative ones.
- P (*Positive*) Sentences have mainly positive terms. There may be some negative ones; however, the positive ones are stronger and outweigh the negative ones.
- SP (*Strong Positive*) Sentences have purely positive words expressing strong affirmative feelings with no complaints. It may have the smallest negative words, but the sentence has mostly great-sounding words or phrases.

IV. EXPERIMENTS AND RESULTS

In order to test our Tweets Sentiment Analysis Model (TSAM), the preliminary experiments have been conducted on the tweets dataset associates with a special event - Australian federal election 2010.

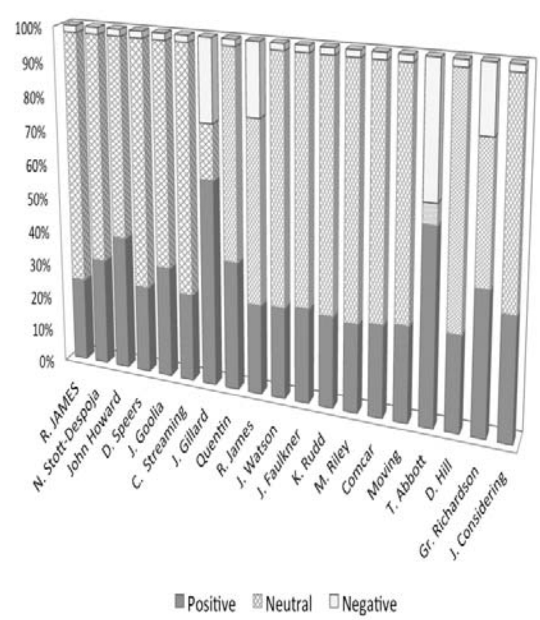

Fig. 2. The politicians' sentiment semantic orientation

A. Experiment Dataset

As discussed earlier, our motivating application is to automate the analysis of tweets as they relate to an event. For this purpose, Australian federal election 2010 event was taken as an example for sentiment analysis experiments. Announcement of election was made on Saturday 17 July and election was held on Saturday, 21 August 2010. Twitter has seen a big spike in messages during the elections. We downloaded the dataset that is comprised of 2 weeks of tweets (from Saturday 17 to Saturday 31 July). We only used all tweets that have hashtag "#ausvotes". All tweets data are split into 57 files. Each file contains about 1000 tweets on average and in total there are about 57000 tweets. Within the dataset, tweets are ordered by tweet *id*, which are chronologically ordered for our experiment.

One nontrivial task of tweets data collection for sentiment analysis is the extraction of the relevant entities from the tweets. To identify and extract the entities that appear in the sentences the Open Calais is used for TSAM currently. We measure the system performance with its accuracy as following:

$$accuracy = \frac{NTSCL}{TNTTS} \qquad (5)$$

where NTSCL is the number of tweets the system correctly labeled and TNTTS is the total number of tweets in a test set.

B. Experiment Results

Fig. 2 illustrates the politicians' sentiment semantic orientation. It shows the percentage of positive, negative or Neutral opinion in the test tweets data.

Fig. 3 indicates the sentiment of the specific political candidates, i.e., two primary minister candidates - Julia Gillard and Tony Abbot. If one is interested in knowing who is a better Prime Minister: Julia Gillard vs Tony Abbott then Fig. 3 will give an answer. In fact, it provides a fast and effective public opinion survey.

Fig. 4 shows the sentiment semantic orientation about Julia Gillard based on a three days collection of tweets. The result shows how people felt about Julia Gillard, as well the changes in public opinions about her over time.

V. CONCLUSIONS AND FUTURE WORKS

In this paper, a straightforward Tweets Sentiment Analysis Model (TSAM) is proposed. This research work has demonstrated that building a lexicon-based sentiment analysis intelligent system is doable and can be very beneficial. However, in its current form the opinion analysis tool is not yet reached full potential. In order to improve current TSAM, a number of research issues are required to be sorted out:

- Distinguishing between parts of speech: in current model, the opinion words are extracted as the features to input into sentiment score function. We did not have the POS processing. It was found that the accuracy of part of speech tagging influence overall sentiment scores. Therefore, the advance NLP technique must be applied to improve the current approach.
- Taking emotion analysis into account: sentiment analysis is the task of identifying positive and negative opinions, emotions, and evaluations [19]. It judges an entity in the dimension of positivity or negativity. On the other hand, emotion analysis of text goes beyond positive-negative dimension to discrete emotion categories like happiness, sadness etc. Text-based mood analysis that is a sub- problem of sentiment and opinion mining have many potential applications identified in [13]. However, text- based mood analysis poses a lot of challenges beyond standard text analysis such as text classification and clustering. Emotion analysis will be our future challenge and will be explored.
- Utilising more accurate entity recognition techniques: current entity extraction task is carried out by using open sources "Open Calais". There are still problems for aggregating entity references under different name. Fig. 2 reveals this problem. For example, Julia Gillard and Julia Goolia may be the same person during the Australian federal election 2010.

The TSAM model will yield much more accurate results with the above works implemented. We believe that the further researches can overcome the limitations and will improve the performances of the TSAM.

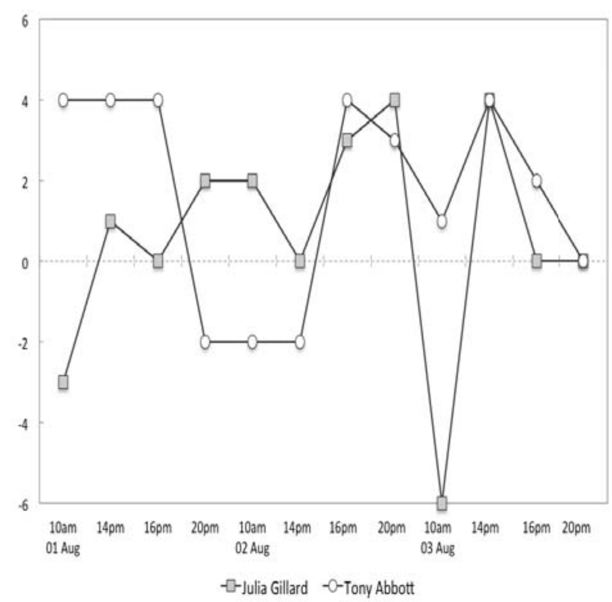

Fig. 3. Who is a better Prime Minister?

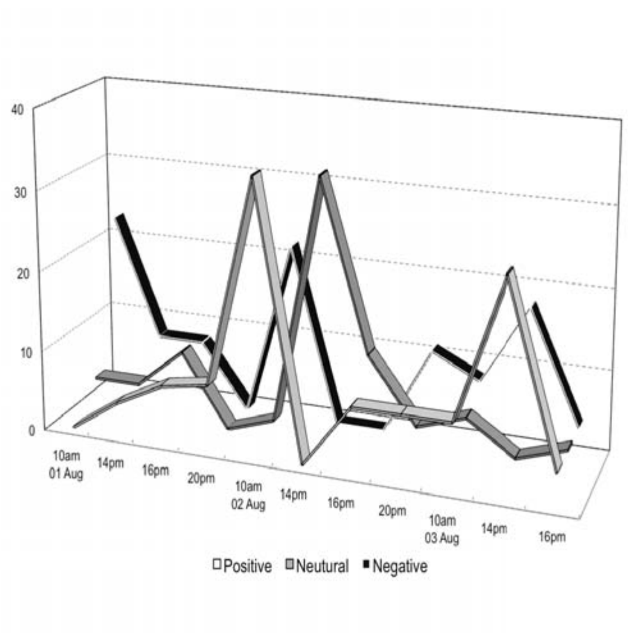

Fig. 4. The sentiment semantic orientation about Julia Gillard

REFERENCES

[1] M. M. Bradley and P. J. Lang. Affective norms for english words (anew): Stimuli, instruction manual, and affective ratings. Technical report, Center for Research in Psychophysiology, University of Florida, Gainesville, Florida, 1999.

[2] S. Brody and N. Elhadad. An unsupervised aspect-sentiment model for online reviews. In HLT '10: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pages 804–812, Morristown, NJ, USA, 2010. Association for Computational Linguistics.

[3] K. Dave, S. Lawrence, and D. Pennock. Opinion extraction and semantic classification of product reviews. In Proceedings of the 12th International World Wide Web Conference (WWW), pages 519–528, 2003.

[4] E.Schonfeld. Mining the thought stream. TechCrunch Weblog Article, heep://techcrunch.com/2009/02/15/mining-the-thought-stream/, 02 2009.

[5] V. Hatzivassiloglou and J. M. Wiebe. Effects of adjective orientation and gradability on sentence subjectivity. pages 299–305, 2000.

[6] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury. Twitter power: Tweets as electronic word of mouth. J. Am. Soc. Inf. Sci., 60(11):2169–2188, 2009.

[7] J.Yarow.Twitterfinallyrevealsallitssecretstats.BusinessInsiderWeblog Article, http://www.businessinsider.com/twitter-stats-2010-4/, 04 2010.

[8] J. Kamps, M. Marx, R. Mokken, and M. de Rijke. Using WordNet to measure semantic orientation of adjectives. In Proceedings of the 4th International Conference on Language Resources and Evaluation, pages 1115–1118, 2004.

[9] B.Liu.Opinionmining.InvitedcontributiontoEncyclopediaofDatabase Systems, 2008.

[10] B. Liu. Handbook of Natural Language Processing, chapter Sentiment Analysis and Subjectivity. Second edition edition, 2010.

[11] S. Milstein, A. Chowdhury, G. Hochmuth, B. Lorica, and R. Magoulas. Twitter and the micro-messaging revolution: Communication, connections, and immediacy140 characters at a time. An OReilly Radar Report . 54 pages, November 2008.

[12] B. Pang and L. Lee. Thumbs up? sentiment classification using machine learning techniques. In In Proceedings of EMNLP, pages 79–86, 2002.

[13] B. Pang and L. Lee. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2):1–135, Jan. 2008.

[14] A.-M. Popescu and O. Etzioni. Extracting product features and opinions from reviews. In HLT '05: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, pages 339 – 346, Morristown, NJ, USA, 2005. Association for Computational Linguistics.

[15] P. D. Turney and M. L. Littman. Unsupervised learning of semantic orientation from a hundred-billion-word corpus. Dec 2002.

[16] J.Wiebe.Learning subjective adjectives from corpora. In Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence, pages735–740. AAAI Press, 2000.

[17] J. Wiebe and E. Riloff. Creating subjective and objective sentence classifiers from unannotated texts. In In CICLing2005, pages 486–4972005.

[18] D. T. Wijaya and S. Bressan. A random walk on the red carpet: rating movies with user reviews and pagerank. In CIKM '08: Proceeding of the 17th ACM conference on Information and knowledge management, pages 951–960. ACM, 2008.

[19] T. Wilson, J. Wiebe, and P. Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In HLT '05: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, pages 347–354, Morristown, NJ, USA, 2005. Association for Computational Linguistics.

[20] W. Zhang, C. Yu, and W. Meng. Opinion retrieval from blogs. In Proceedings of the sixteenth ACM conference on Conference on information and knowledge management, CIKM '07, pages 831–840. ACM, 2007.