DOI: http://dx.doi.org/10.26483/ijarcs.v8i7.4229

ISSN No. 0976-5697

Volume 8, No. 7, July – August 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

RESEARCH PAPER ON SENTIMENTAL ANALYSIS OF ONLINE CUSTOMER REVIEWS AND RATING

Er.Ramanpreet Kaur
Department of Computer Engineering
Punjabi University
Patiala, India

Dr.Gaurav Gupta
Department of Computer Engineering
Punjabi University
Patiala, India

Abstract: Sentiment Analysis is the process of determining whether a part of writing is positive, negative or neutral. It is also known as opinion mining or emotion AI (Artificial Intelligence) and refers to the use of natural language processing, text analysis, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information. It is widely applied to voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine. In order to determine the sentiment of the overall document, we can use our own scoring algorithms – using the weighted phrases from the previous side, and then using our proprietary way of adding them up. By taking a set of content we can build a document-level sentiment classifier on it. In this thesis we are going to see how frequent item set can be used for mining reviews from online reviews posted by the customers. Our main aim is to create a system for analyzing sentiments or opinions which implies judgment of different consumer products.

Keywords: Sentimental Analysis, Opinion Mining, Frequent Words, Online Reviews, Rating.

1. INTRODUCTION

Sentiment analysis means to classify the polarity of a given text at the document level, sentence level, or feature/aspect level whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral.

In Advanced Sentimental analysis we look beyond polarity classification, for instance, at emotional states such as angry, sad, and happy.

A different method for determining sentiment is the use of a scaling system whereby words commonly associated with having a negative, neutral, or positive sentiment with them are given an associated number simply from 0 to a positive upper limit such as +4 [1]. This makes it possible to adjust the sentiment of a given term relative to the level of the sentence.

When a piece of unstructured text is analyzed using natural language processing, each concept in the specified environment is given a score based on the way sentiment words relate to the concept and its associated score[9].

The rise of social media such as blogs and social networks has fuelled interest in sentiment analysis [10,11]. With the proliferation of reviews, ratings, recommendations and other forms of online expression, online opinion has turned into a kind of virtual currency for businesses looking to market their products, identify new opportunities and manage their reputations.

As businesses look to automate the process of filtering out the noise, understanding the conversations[13,14], identifying the relevant content and auctioning it appropriately, many are now looking to the field of sentiment analysis. One step towards this aim is accomplished in research.

Several research teams in universities around the world currently focus on understanding the dynamics of sentiment in e-communities through sentiment analysis.

Even though short text strings might be a problem, sentiment analysis within micro-blogging has shown that Twitter can be seen as a valid online indicator of political sentiment [2].

This paper is structured as follows. Section II describes the research methodology used in this study. Section III gives the steps in Sentimental Analysis. Section IV gives the classification of Sentimental Analysis. Section V presents the resulting reports. Section VI presents the conclusion and future directions.

2. RESEARCH METHODOLOGY

The research methodology is composed of three stages. The first stage involves the Extraction, Transformation, Loading of data in Data ware house [12]. We transform the flat file into relational table as it is easy to perform the SQL operation on relation table than on a flat file.

The second stage involves functions and procedures which are used to update the tables and check the type of review and rating given by the customer.

The third stage involves generating reports and creating dashboard with the help of IBM Cognos BI Software. The snapshots of three stages are given as follows.

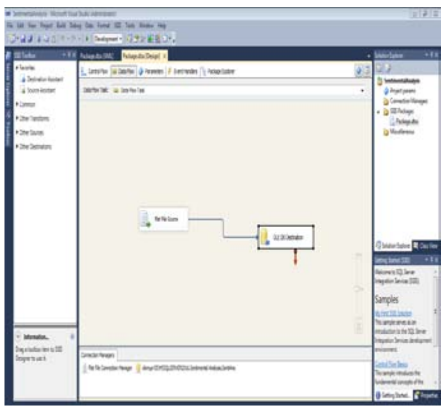

Figure 1 ETL STAGE

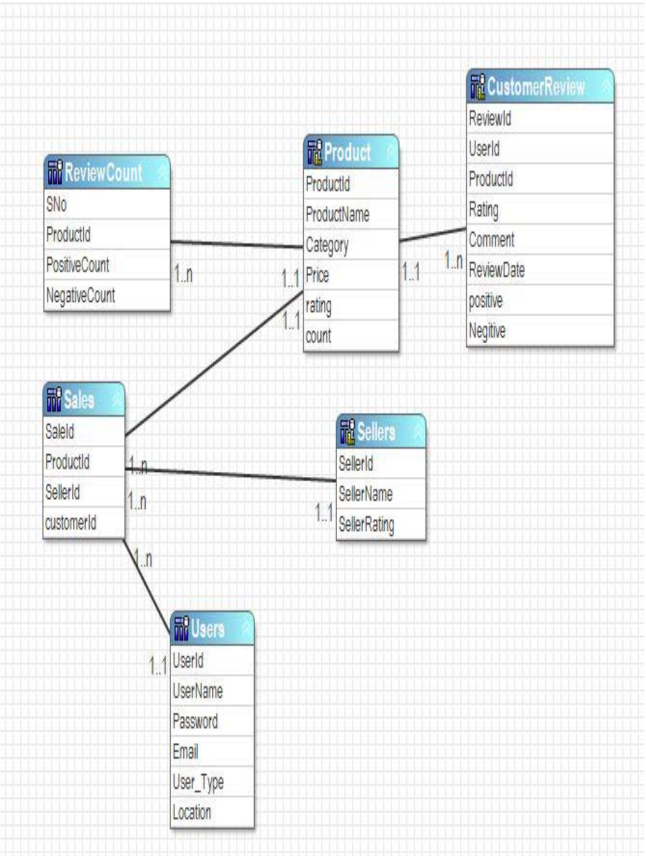

Figure 2 Dataware House Diagram

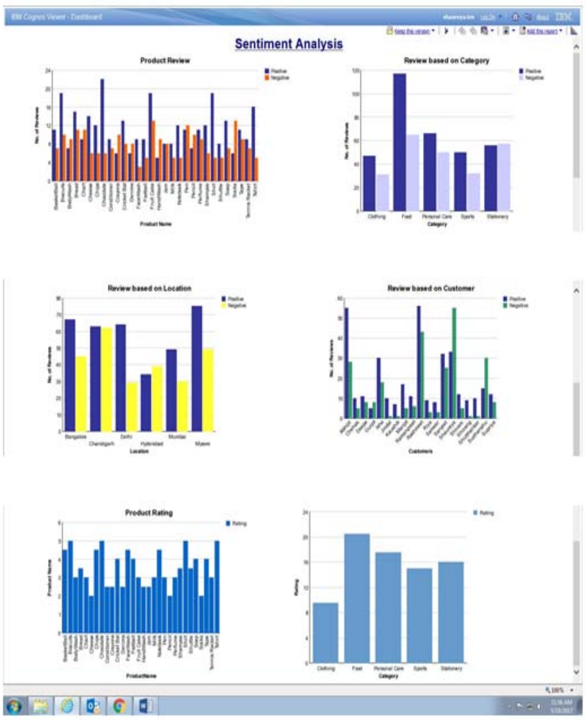

Figure 3Dashboard in IBM Cognos

3. STEPS IN SENTIMENT ANALYSIS (OPINION MINING)

There are basic six steps in sentimental analysis[8].

- First is information database creation,
- POS tagging is done on review,
- Features are extracted using grammar rules such as adjective + nouns so on, as noun are features and adjectives are sentiment words,
- Opinion word extraction
- Identify polarity of opinion word
- Identify polarity of sentence

There are many algorithms which can be used in sentimental analysis such as Navie Bayes Classification, Probabilistic Machine Learning approach to classify the reviews as positive or negative.

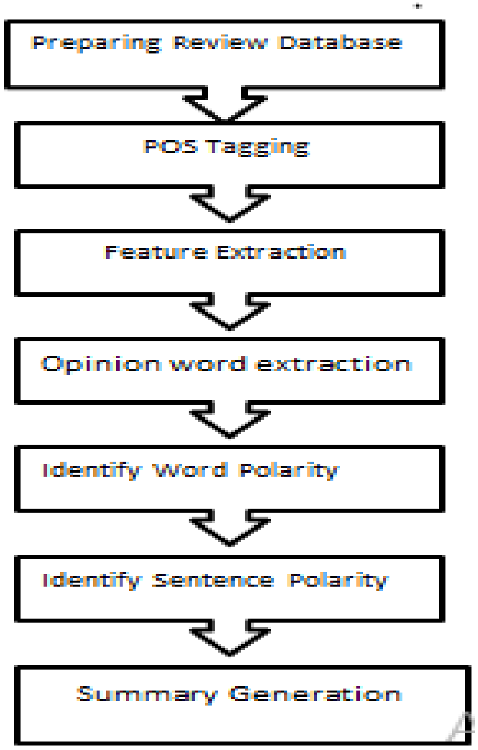

Figure 4 Basic Step of Sentimental Analysis

4. CLASSIFICATION OF SENTIMENTAL ANALYSIS

Sentimental Analysis of text can be mainly classified into four types which are given below.

- Document level Sentimental Analysis
- Sentence level Sentimental Analysis
- Phrase Level Sentimental Analysis
- Feature Level Sentimental Analysis

*A. **Document level Sentimental Analysis[3]***

Document level tasks primarily concerns with classification issues where the available document has to be arranged into a set of predefined classes. In subjectivity analysis a document is divided as subjective or objective. In sentiment analysis, a document can be classified as positive or negative or neutral depending upon the polarity of subjective information that is present in the document[15]. Opinion quality and support evaluation makes decision whether an opinion is useful or not and opinion spam identification groups and divides opinions as spam and not spam.

*B. **Sentence level Sentimental Analysis[4]***

The problem at this measure everything refers to sentences. In opinion data extraction and recovery and opinion question answering, sentences are generally placed and positioned focused around some criteria. Opinion rundown intends to select a set of sentences which outline the opinion all the more exactly[16]. At long last, opinion mining in relative sentences incorporates recognizing similar sentences and concentrating data from them.

*C. **Phrase level Sentimental Analysis***

Phrase level mining [5] came into picture because document level mining and sentence level mining approaches can not find accurately what actually users likes and they does not like. Phrase level opinion mining looks for sentiments on features of products.

*D. **Feature level Sentimental Analysis***

Feature level sentimental analysis [6] comes into picture when a customer or user looking for feedback of certain feature or attribute of a product rather than total feedback of the product. We see many customers interested in only certain features of some products rather than the whole product like some people look for a mobile that has excellent battery life and they are not concerned with other features like camera clarity, music clarity and so on[7]. In situations like mentioned in this section feature level opinion mining helps a lot for extracting polarity information for a particular feature or attribute from a product.

5. RESULTS

After doing ETL, creating views, modifying properties, publishing package and we plot the facts and dimensions to generate the reports to visualize the results. Firstly, to have an overview of all the reviews about the products we have a dashboard containing six different reports.

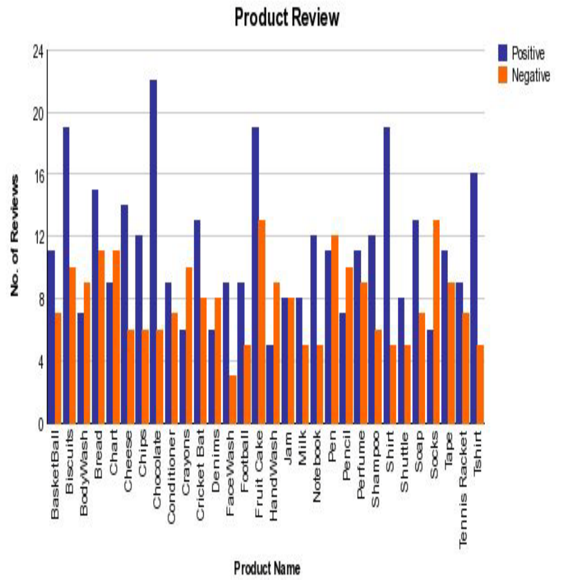

Figure 5 Positive and Negative reviews for products

- The Report above shows number of reviews given to each product.In this report number of reviews is used as measure and product name is the other dimension.
- Positive review and the negative review by the customer for each product can be seen in different color.

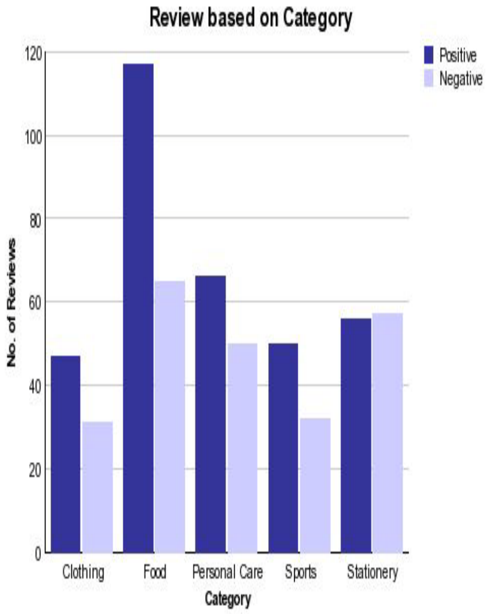

Figure 6 Reviews on the basis of Product Category

- The above report shows the positive and negative reviews of different categories of products.
- In this report number of review is fact and Product Category is the dimension.
- The report below will give number of positive and negative reviews for selected category.

- In report which is based on the location of the customer, the number of positive and negative reviews are shown on the bases of the location.
- As we can see the number of positive reviews are more in each city.

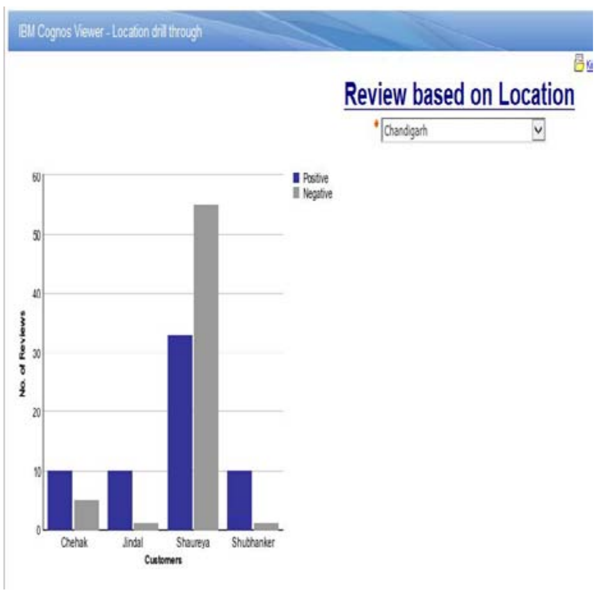

Figure 8 Reviews based on selected Location

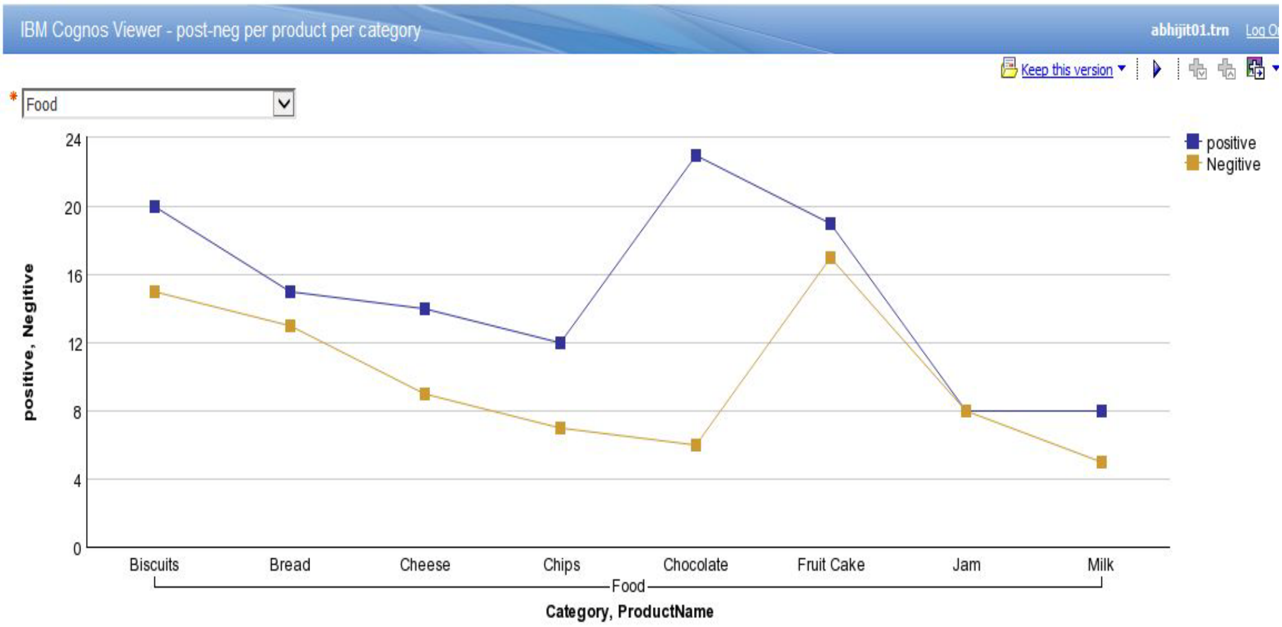

Figure7 Reviews for selected category of Product

- Rating of all Products.
- Rating is on the scale from 1 to 5.

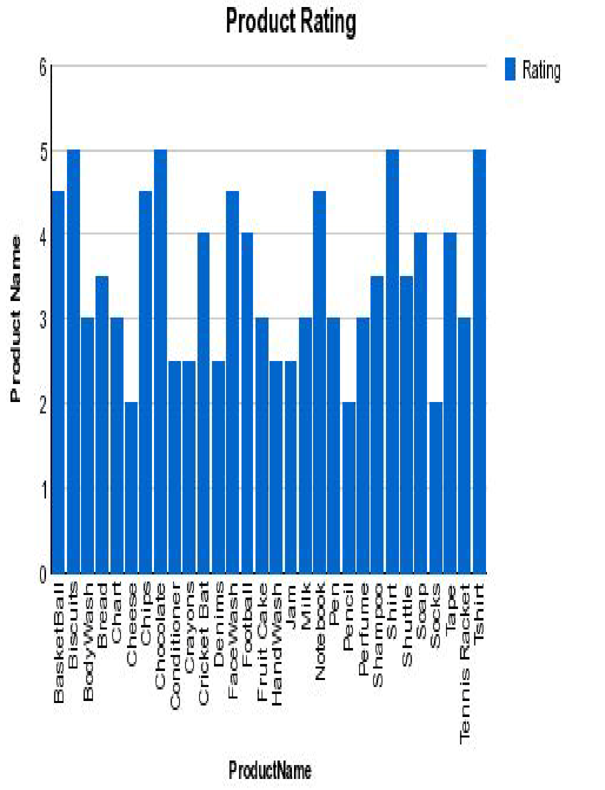

Figure 9 Rating of Products

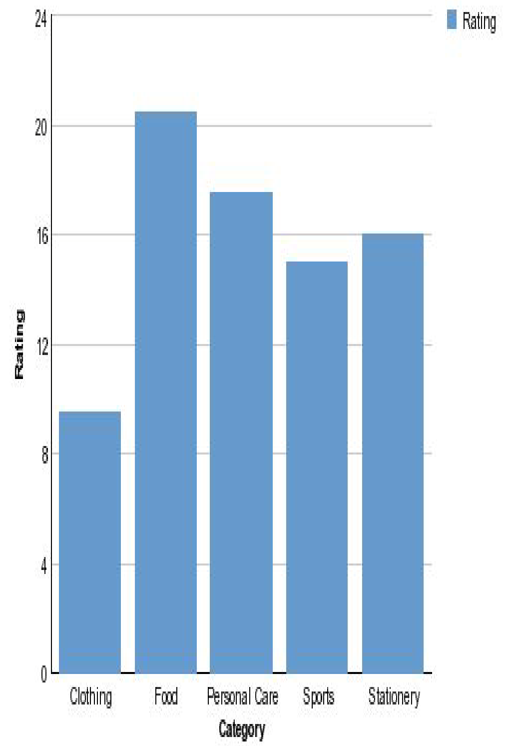

Figure 10 Rating Category

- Category wise rating given by the customer.
- List view of products with average of rating.

Product List

ProductName	Category	Price	Rating
Denims	Clothing	1200	2.5
Shirt		800	5
Socks		80	2
Tshirt		500	5
Biscuits	Food	30	5
Bread		60	3.5
Cheese		98	2
Chips		20	4.5
Chocolate		60	5
Fruit Cake		50	3
Jam		25	2.5
Milk		13	3
BodyWash	Personal Care	300	3
Conditioner		250	2.5
FaceWash		100	4.5
HandWash		100	2.5
Perfume		350	3
Shampoo		200	3.5
Soap		50	4
BasketBall	Sports	500	4.5

Figure 11 List View of Products with Average of Rating

- List view of products with average of rating.
- Report which we get on drill trough given below.
- These reports get filtered on the basis of particular condition.

Figure 12 Rating of selected Category

6. CONCLUSION

The following conclusions are drawn on the basis of experimental observations and analysis. SA is typically used to analyse people's sentiments opinions, appraisals, attitudes, evaluations and emotions towards such entities as organisations, products, services, individuals, topics, issues, events and their attributes, as presented online via text, video and other means of communication. These communications can fall into three broad categories, namely positive, neutral and negative.

7. REFERENCES

[1] Taboada, Maite; Brooke, Julian (2011). "Lexicon-based methods for sentiment analysis". Computational Linguistics. **37** (2): 272–274. doi:10.1162/coli_a_00049

[2] CORDIS. "Collective emotions in cyberspace (CYBEREMOTIONS)", European Commission, 2009-02-03. Retrieved on 2010-12-13.

[3] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of the conference on human language technology and empirical methods in natural language processing, pages 347–354. Association for Computational Linguistics, 2005.

[4] Alexander Pak and Patrick Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In LREC,2010.

[5] Andrea Esuli. Automatic generation of lexical resources for opinion mining: models, algorithms and applications. In ACM SIGIR Forum, volume 42, pages 105–106. ACM, 2008.

[6] Richa Sharma, Shweta Nigam, and Rekha Jain. Supervised opinion mining techniques: A survey.

[7] Akshat Bakliwal. Fine-grained opinion mining from different genre of social media content. 2013.

[8] Samaneh Moghaddam and Martin Ester. Aspect-based opinion mining from online reviews. In Tutorial at SIGIR Conference, 2012.

[9] Bo Pang and Lillian Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of the 42nd annual meeting on Association for Computational Linguistics, page 271. Association for Computational Linguistics, 2004.

[10] Ellen Riloff, Janyce Wiebe, and William Phillips. Exploiting subjectivity classification to improve information extraction. In Proceedings of the National Conference On Artificial Intelligence, volume 20, page 1106. MenloPark, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2005.

[11] Stephan Raaijmakers and Wessel Kraaij. A shallow approach to subjectivity classification. In ICWSM, 2008.

[12] Janyce M Wiebe, Rebecca F Bruce, and Thomas P O'Hara. Development and use of a gold-standard data set for subjectivity classifications. In Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics, pages 246–253. Association for Computational Linguistics, 1999.

[13] Anindya Ghose and Panagiotis G Ipeirotis. Estimating the helpfulness and economic impact of product reviews:Mining text and reviewer characteristics. Knowledge and Data Engineering, IEEE Transactions on, 23(10):1498–1512, 2011.

[14] Nitin Jindal and Bing Liu. Opinion spam and analysis. In Proceedings of the 2008 International Conference on Web Search and Data Mining, pages 219–230. ACM, 2008.

[15] Nitin Jindal and Bing Liu. Review spam detection. In Proceedings of the 16th international conference on WorldWide Web, pages 1189–1190. ACM, 2007.

[16] Wei Zhang, Clement Yu, and Weiyi Meng. opinion retrieval from blogs. In Proceedings of the sixteenth ACM conference on Conference on information and knowledge management, pages 831–840. ACM, 2007.

[17] Lun-Wei Ku, Yu-Ting Liang, and Hsin-Hsi Chen. Opinion extraction, summarization and tracking in news and blog corpora. In AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs, volume 100107,2006.

