

# A Survey on Sentiment Analysis Applied in Opinion Mining

S. Vasantharaj

PG Student, M.Tech in Computer science, Sri ManakulaVinayagar Engineering College, Puducherry.  
rajvasanth49@gmail.com

A. Martin

Associate Professor, Dept. of MCA, Sri ManakulaVinayagar Engineering College, Puducherry.  
cudmartin@gmail.com

A. Meiappane

Associate Professor, Dept. of IT, ManakulaVinayagar Institute of Technology, Puducherry.  
auromei@gmail.com

J. Madhusudnan

Associate Professor, Dept. of CSE, Sri ManakulaVinayagar Engineering College, Puducherry.  
contactmadhu@gmail.com

**Abstract – Sentiment analysis is the process to analysis the people emotions, characteristics and opinion, about the particular product and its problem. This survey paper tackles a comprehensive summary about the methods and its corresponding techniques in various fields. This article mainly focuses on opinion mining techniques such as CBR, Supervised, Unsupervised, Machine learning techniques. We provided complete details about the data set, techniques, and its result such as positive, negative and neutral feedback for each method. The literature survey will give the complete information about the sentiment analysis with the recent research work and its task to identify the opinion mining.**

**Index Terms – opinion mining-task identify-machine learning techniques – CBR-Supervised machine language.**

## 1. INTRODUCTION

Sentiment Analysis (SA) or Opinion Mining (OM) is the study of people's opinions, attitudes and its emotions toward an entity. The entity will represent people, products and domain. The SA or OM area unit interchangeable and categorize mutual which means some researchers declared that OM and SA have slightly completely different notions. Opinion Mining extracts to analyzes people's opinion regarding an entity whereas Sentiment Analysis identifies the sentiment expressed in an exceedingly text then analyzes it. Therefore, the target of SA is to search out the opinions to determine the feelings they categorical classify their polarity.

Document-level SA aims to classify the opinion document as expressing a positive or negative opinion or sentiment. It considers the complete document of a basic info unit (talking

regarding one topic). Sentence-level SA aims to classify the sentiment expressed in every sentence. The primary step is to identify whether or not the sentence is subjective or objective. If the sentence is subjective, Sentence-level SA can verify whether or not the sentence express positive or negative opinions have seen that sentiment expressions are not essentially subjective in nature. However, there's no fundamental distinction between document and sentence level classifications as a result of sentences square measure. Classifying text at the document level or at the sentence level does not offer the required detail required opinions on all aspects of the entity that is required in several applications, to obtain these details We apply the `sentiment analysis in various field to find the accuracy about the feedback. It will increase the product level. For example we apply the sentiment analysis in military field .The people in this field they work in different location they need to know details about the working location. By using the sentiment analysis it will specify the aspects area based on the expert knowledge in this field. The primary step is to spot the entities and their aspects. The opinion holders will provide completely different opinions for various aspects of constant entity like this sentence. "The voice quality is not sensible, but the battery life is long". This survey tackles the primary 2 styles of SA. The data sets utilized in militia square measure a crucial issue during this field. The reviews sources square measure chiefly review sites. Militia is not solely applied on product reviews however it will also be applied on stock markets [5] news

articles, [6] or political debates [7] and also applied in various field.

The rest of paper is organized as follows. Section 2 describes about literature review and section 3 describes about research issues and challenges and section 4 concludes the paper.

## 2. LITERATURE REVIEW

In this survey of opinion mining BR class may be classified to lexica, Corpora or dictionaries. To solve the Sentiment classification downside as SC. different articles that solve a general Sentiment Analysis downside are classified offer contribution within the feature choice section are classified as FS. The connected fields like feeling Detection (ED), Building Resource (BR) and Transfer Learning (TL). The survey methodology is as follows: transient clarification to the celebrated FS and SC algorithms representing some connected fields are mentioned. Then the contribution of those articles to these algorithms is presented illustrating how they use these algorithms to solve special problems in SA. The main target of this survey is to present a unique categorization for these SA related articles. It consists of lexicon based approach and machine learning approach. The lexicon based approach is classified into corpus based and dictionary approaches. In this classification the corpus is further classified into semantic and statistical approaches.

Similarly the machine learning approach is classified into unsupervised learning and supervised learning. The supervised learning is classified into probabilistic classifier, linear classifier, rule based classifiers and decision tree classifiers. The rule based classifier further classified into maximum entropy, Bayesian and Navie bayes. Finally the decision tree classifier is classified into neural network and SVM.

An opinion is that the personal state of a personal, and intrinsically, it represents the individual's concepts, beliefs, assessments, judgments and evaluations a couple of specific subject/topic/item.[6]conclude that others' opinions have an excellent impact on and supply steerage for people, government organizations and social communities throughout the decision-making method. This method, groups of people need quick, correct and info in order that they will build fast and correct selections. Through opinions, humans will integrate the various approaches, experiences, knowledge and data of the many individuals once creating selection. To raise their friends, relations, and field specialists for info throughout the decision-making method, and their opinions and views area unit supported experiences, observations, concepts, and beliefs. One's perspective a couple of subject will either be positive or negative, that is remarked because the polarity of the opinion. It will be expressed in numerous ways in which the area unit samples of opinion statements.

- He isn't a decent actor.

- The breakfast was quite smart.
- The edifice was big-ticket.
- Terrorists merit no mercy!
- Hotel A is dearer than edifice B.
- Coffee is dear, however tea is affordable.

An opinion has 3 main parts, i.e., the opinion holder or supply of the opinion, the survey regarding that the opinion is expressed and therefore the analysis, read or appraisal, that is, the opinion. For opinion identification, all of those parts area unit vital.

While opinions will be collected from totally different sources, e.g., individual interactions, newspapers, television, net etc., the web has become the richest supply of opinion assortment. Before the globe Wide internet (www), individuals collected opinions manually. If a personal was to create a call, he/she usually asked for opinions from friends and relations to accumulate popular opinion, organizations usually conducted surveys through centered of survey, however, was big-ticket and effortful. Now, the web provides this info with one click and at little or no value. With the arrival of internet two.0, the web permits internet users to get website on-line and post their info severally.

This side of the web permits internet users to participate in cooperative world environments. Hence, the web has become an expensive supply for social networks, client feedback, on-line searching etc. in line with a survey, quite forty five, new blogs area unit created daily in conjunction with one.2 million new posts day after day [8]. the data collected through these services is employed for varied sorts of deciding for instance, social networks will be used for political, religious, and security problems still as for affairs of state, whereas client feedback will be used for product sales, purchases, and producing. Not solely is that the trend of on-line searching increasing daily, however vendors collect client feedback for future trend predictions and merchandise improvement through these portals. The key part that has provided the inspiration for this work is opinion.

Though the web may be a wealthy supply of opinions with immeasurable blogs, forums and social websites giving an oversized volume of updated info, the net information, sadly, area unit usually unstructured text that can't be directly used for data illustration. Moreover, such an enormous volume of information cannot be processed manually. Hence, efficient tools and potential techniques area unit required to extract and summarize the opinions contained in that analysis communities area unit checking out Associate in Nursing efficient thanks to rework this internet info into data requisition and so gift the data to the user during a laconic and understandable manner. The emergence of internet two has created the task of posting and aggregation opinions via the

net abundant easier, the standard management, processing, compilation, and summarization of those opinions became potential analysis issues.

OM may be a procedure accustomed extract opinion from text. “OM may be a recent discipline at the crossroads of knowledge retrieval, text mining and linguistics that tries to discover the opinions expressed in language texts” [6]. OM may be a field of information discovery and data processing (KDD) that uses NLP and applied math machine learning techniques to differentiate narrow text from factual text. As such, OM tasks involve opinion identification, (positive, negative, and neutral), target identification, supply identification and opinion summarization. Hence, OM tasks need techniques from the field of NLP, info retrieval (IR), and text mining. The most concern is a way to mechanically establish opinion parts from unstructured text Associate in Nursing summarize the opinion regarding an entity from an enormous volume of unstructured text.

Textual info will be classified as either objective or subjective. Objective statements represent facts, whereas subjective statements represent perceptions, views or opinions. The NLP analysis preliminary centered on mining factual info from a text, that is a crucial space with varied applications; but, with the arrival of internet two.0, that permits the user to get website, some new and attention-grabbing concepts are developed for the extraction of information from user-generated discourse. The Web 2.0 facility provides the chance to accumulate needed info from internet users and apply IR and KD techniques for varied applications. User feedback on the net is collected through social networks, blogs, industrial organizations, promoting etc. immeasurable reviews and comments area unit collected through promoting and repair websites (Amazon, Trip

authority etc.), social networks (Facebook, Flickr, YouTube etc.), industrial and social media (Voice of America, BBC, CNN, Yahoo etc.), and plenty of alternative blogs and forum websites. The mining of those reviews will offer answers to varied analysis queries.

### 3. OPINION MINING TECHNIQUES

The sentimental analysis or opinion mining techniques work with the collection of the opinion get from the people comments. The techniques of opinion mining are listed in Figure 1.

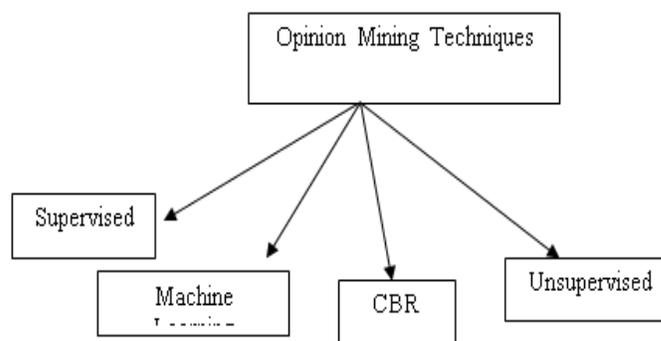


Figure 1 Opinion mining techniques

#### 3.1. Supervised Machine Learning

The supervised machine learning is that one in all the techniques Classification is most often used and well-liked data processing technique. Classification wont to predict the doable outcome from given knowledge assault the premise of outlined set of attributes and a given prophetic attributes. The given dataset is termed coaching dataset consist on freelance variables and a dependent attribute (predicted attribute) is detail in Table 1.

Model	Dataset	Techniques	Results
A Study of Opinion Mining and Visualization of Hotel Reviews	booking.com, TripAdvisor.com	Double Propagation Algorithm. (DPLR-R) state-of-the-art product feature extraction algorithm	Trip advisor: naïve Bayes: 57%, Dyn. LM Classifier: 57%, Booking: naïve Bayes: 44%, Dyn. LM Classifier: 48%
Mining Product Features from Online Reviews	reviews for 5 product classes from Hu and Liu data	Senti Word Net based algorithm Fuzzy matching	Recall: 0.903 Precision: 0.927
A Kernel-based Sentiment Classification Approach for Chinese Sentences	China cars reviews	Word Kernel Path Kernel N-gram Kernel	Precision: 80.93% Recall: 62.03% F-score: 70.23%
Web Opinion Mining: How to extract opinions from blogs?	Movie Reviews	AMOD approach	positive: 82.6% negative: 52.4%

Table 1. List of techniques applied in opining mining

3.2. Unsupervised Learning:

In distinction of supervised learning, the unsupervised learning has no express targeted output related to input. Category label for any instance is unknown therefore unsupervised learning is near to learn by observation rather

than learn by example Clustering could be a technique utilized in unsupervised learning. The process of gathering objects of similar characteristics into a group is called clustering. Objects in one cluster are dissimilar to the objects in other clusters. In opinion mining, different clustering techniques are used described in Table 2.

Models	Technique	Results
A Framework to Answer Questions Of Opinion Type	Bayes classifier k-means clustering	Accuracy: 71%
An Upgrading Feature-Based Opinion Mining Model On Vietnamese Product reviews	HAC clustering semi-SVM-KNN classification	Purity = 0.7 Accuracy = 0.72% Entropy = 0.77
A Novel Product Features Categorize Method Based On Twice-Clustering	Means COP K-Means	Accuracy: 66%
Enriching Sentinel Polarity Scores Through Semi-Supervised Fuzzy Clustering	Word net-Affect (WNA) scores Fuzzy-C clustering	Accuracy: 92.15%
Mining Web Videos For Video Quality Assessment	K-Means cluster Mean Opinion score	Accuracy: 82%
Sentiment Clustering Of Product Object Based On Feature Reduction	Feature dimension reduction algorithm K-Means clustering technique	runtime reduced: 41.38%, scarcity reduced:15.69%,
Opinion Mining Based On Feature-Level	K-means Cop-K means	Precision:71% Recall:61%
Product Feature Mining With Nominal Semantic Structure	Open NLP tagger Malt parser factorization method	More clusters with lower entropies
An Unsupervised Method For Joint Information Extraction And Feature Mining Across Different Web Sites	undirected graphical model	Recall: 81.9% Precision: 81.2%

Table 2. List of clustering techniques applied in Opining Mining

3.3. Case Based Reasoning

The Case based reasoning is the artificial computing of supervised techniques accustomed notice the answer of a brand new drawback on the premise of past similar issues. In real time scenario CBR is act as a powerful tools. It is a recent problem solving technique in which knowledge is personified as past cases in library and it does not depend on classical

rules. The solution of the problem are stored in the knowledge base of the repository systems of the CBR. It will use the solution for similar problem is arrive in some process to form a new instances in the data. The case based reasoning CBR cycle consists of four R's. Now days it is the most emerging technique used in opinion mining systems which are described in Table 3.

Model	Dataset	Techniques	Products Data
A Survey In The Area Of Machine Learning And Its Application For Software Quality Prediction	University Student's Data	Case-Based Reasoning And Fuzzy Logic	Accuracy: 91.3% Within 10% Error

A Fuzzy CBR Technique For Generating Product ideas	1600 Products Data	Fuzzy CBR	Cell Phone Retrieving: 91.86% Cell Phone Retrieving+ Filtering:47.78% Ball Retrieving: 82.64% Ball Phone Retrieving+ Filtering:44.45%
Intelligent Project Approval Cycle For Local Government Case-Based Reasoning Approach	Health Care Project Data	Manhattan Or City Distance, CBR	Accuracy: 90%
Predictions Using Data Mining And Case-Based Reasoning: A Case Study For Retinopathy	Diabetes Patients data in Malaysia	C5.0 Algorithm And CBR	Accuracy :85%
Combining Case Based Reasoning And Data Mining – A Way Of Revealing And Reusing Rams Experience	Schlumberger And Norms Hydro Data	Data Mining + CBR	Accuracy: 92%
Research On CBR System Based On Data Mining	GHMOS, SOM data	GHSOM (Growing Hierarchical Self Organizing Map) CBR	TCBR Accuracy: 78% Ga-CBR : 84 % Proposed CBR: 94%
Improving User Experience With Case-Based Reasoning Systems Using Text Mining And Web 2.0	50 Special Education Related Files	Kite CBR	KITE CBR reduce effort time in blog site case retrieve
Extending jcolibri For Textual CBR	Restaurant Data	Information Extraction Information Retrieval	IR+IE Accuracy: 75%

Table 3. Classification of positive and negative opinion using CBR

#### 4. OPINION MINING TASKS

The OM downside and its sub-problems, every of that has its own relevant importance, are found throughout a range of topics. The most elements of associate degree OM downside are the supply of the opinion, the target of the opinion, and also the critical expressions or comments created by the opinion holder define the OM downside. ‘‘Given a group of critical text documents D that contain opinions (or sentiments) concerning associate degree object, opinion mining aims to extract attributes and elements of the thing that are commented on in every document d a pair of D and to see whether or not the comments are positive, negative or neutral.

Generally, associate degree opinion is expressed by someone (opinion holder) UN agency expresses a viewpoint (positive,

negative, or neutral) concerning associate degree entity (target object, e.g., person, item, organization, event, service, etc.). The subsequent subsections describe the key tasks and approaches to every sub-problem of opinion mining

##### 4.1. Subjectivity and polarity classification

The core task of opinion mining is that the automatic identification of narrow-minded text in documents [8][9]. The well-mined text is then categorized as objective and subjective. Most of the prevailing analysis concurs that objective text constitutes factual data whereas subjective text represents individual views, beliefs, opinions or sentiments. Hence, most opinion mining systems use the subjective text for opinion hood determination [9]. The varied approaches are adopted for this subtask of OM. The foremost common embody heuristics and discourse structure, coarse- and fine-grained analysis, key word and idea analysis [15].

#### 4.2. Opinion target identification

The opinion target refers to the person, object, feature, event or topic concerning that the opinion is expressed a result of opinion target identification is a vital feature of OM, an intensive summary of approaches associated with opinion target extraction is critical. The in-depth analysis of each facet of a product supported client opinion is equally necessary for the general public, the merchants and also the makers [13]. To check reviews, it's necessary to mechanically determine and extract those options that are mentioned within the reviews. Hence, feature mining of product is vital for opinion mining and summarization particularly on condition that the task of feature mining provides the inspiration for opinion summarization [17]. However, there are issues associated with opinion target extraction. If a system is capable of characteristic target options in a very sentence or document, then the system should even be able to spot opinion targets at the sentence or document level, the system able to determine critical expressions. Moreover, some options don't seem to be expressly given, but rather, they're expected from term linguistics, conjointly noted as implicit options. A background study reveals that the method of opinion target extraction involves varied language process tasks and techniques like pre-processing, tokenization, part-of-speech tagging, noise removal, feature choice and classification.

#### 4.3. Opinion supply identification

An opinion holder or the supply of associate degree opinion is that the person or medium UN agency presents the opinion. [16]The opinion holder or opinion supply is vital once authenticating the opinion similarly because the strength, application and classification of the opinion, because the quality associate degree dependability of an opinion is greatly smitten by the supply of that opinion. An example, an announcement is also reliable if the holder or supply that produces it's authentic. Associate degree professional opinion has bigger strength than will the opinion of a normal person. Opinions may be classified supported the opinion holder. For an example, a doctor's opinion once creating choices associated with health and medical treatment whereas belief is vital for a political candidate. Thus, it's necessary to spot the supply or the holder of the opinion [6].

### 5. RESEARCH ISSUES AND CHALLENGES

As the web and internet technologies still grow and expand, the area and scope within the space of knowledge retrieval is additionally increasing. Hence, researchers take a keen interest in solving the issues related to OM, one in all the subareas associated with info retrieval and knowledge discovery from the net. OM is taken into account a noteworthy area of analysis as a result of its several applications in society. Over the past few years, the ever-present dependency on e-marketing, e-business, e-banking,

product recommendations, political reviews, and alternative social activities has attracted analysis communities worldwide. Special attention has been given to customer mining of reviews as they look for info from the Web a few products and/or the product's name. A number of sub-areas of this subject are explored and extensive research has been reportable on every of the sub-problems [17] [18].

Despite various analysis efforts, this OM studies and applications still have limitations and margins for improvement. Accordingly, OM suffers from variety of issues, such as accuracy; quantify ability, quality, normal of information, natural language understanding comprehension, among others. [2][8] Some of the main challenges associated with natural language processing, like context dependency, linguistics connectedness and ambiguity, have created OM tough. As sensible applications require high accuracy, a number of the work should still be performed manually due to the difficult issues with the NLP [5] [13] as an example, the matter of ambiguity, context dependency, and complicated and obscure sentences need more attention to enhance the accuracy of the info analyses. whereas non-public blogs square measure a very important supply of information for OM, the diary posts square measure generally written informally and square measure extremely numerous and therefore subject to inaccuracies and misinterpretations in analysis.

Also presenting an enormous challenge within the face of OM is that the availableness and accessibility of a typical dataset. Another main challenge during this space is that the quality of reviews. As a result of the net is overtly accessible to everybody, anyone will post a review, a state of affairs that brings into question the standard of a review or opinion. Once people square measure creating selections supported the reviews accessed from the net, it's necessary that the reviews be credible and of prime quality. However, solely restricted work has been conducted on opinion quality determination.

1. The recent research is process on Paragraph level polarity for positive and negative feedback.
2. Contents analysis to avoid the human effort.
3. Opinion detection will be increase by their algorithm.
4. Sentiment classification and semantic analysis based on lexical or corpus text.
5. The best data set will be involved to get the polarity level improvement.
6. In opinion mining based on expert knowledge domain will identify.

## 6. CONCLUSION

This survey paper given an outline of the recent updates in algorithms and applications in Sentiment analysis the recently printed and cited articles were categorized and summarized. Articles provide contributions to several militia connected fields that use militia techniques for numerous real-world applications. After analyzing these articles, it's clear that the enhancements of SC associate degreed FS algorithms square measure still an open field for analysis. Naïve Bayes and Support Vector Machines square measure the foremost frequently used algorithms for determination and drawback. They're considered a reference model wherever several planned algorithms are compared. The interest in languages apart from English during this field is growing as there's still an absence of resources and researches regarding these languages. The foremost common lexicon supply used is Word Net that exists in languages apart from English. Building resources, employed in militia tasks, continues to be required for several natural languages. This media info plays a good role in expressing people's feelings, or opinions a couple of sure topic or product.

## REFERENCES

- [1] T. Khushboo "Mining of Sentence Level Opinion Using Supervised Term Weighted Approach of Naïve Bayesian Algorithm", Int. Journal. Computer Technology & Applications, Vol 3 IJCTA | MAY-JUNE 2012.
- [2] J. Han and M. Kamber, "Data Mining: Concepts and Techniques", third edition Publication Date: July 6, 2011.
- [3] N, Anwer and A, Rashid "Feature Based Opinion Mining of Online Free Format Customer Reviews Using Frequency Distribution and Bayesian Statistics" Networked Computing and Advanced Information Management (NCM), 2010 Sixth International Conference on 16-18 Aug. 2010.
- [4] B. Seerat and F.Azam "Opinion Mining: Issues and Challenges (A survey)" International Journal of Computer Applications (0975 – 8887) Volume 49– No.9, July 2012.
- [5] N. M. Shelke, S. Deshpande and V. Thakre "Survey of Techniques for Opinion Mining" International Journal of Computer Applications (0975 – 8887) Volume 57– No.13, November 2012.
- [6] S. Deshpande "A Survey on Web Data Mining Applications" Emerging Trends in Computer Science and Information Technology - 2012(ETCSIT2012) Proceedings published in International Journal of Computer Applications® (IJCA).
- [7] Aamodt, H. A. Sandtorv, O. M. Winnem "Combining case base Based Reasoning and Data Mining - A way of revealing and reusing RAMS experience" Safety and Reliability; Proceedings of ESREL '98, Trondheim, June 16-19, 1998.
- [8] Thain, T, Tannenbaum and M. Livny "Distributed Computing in Practice: The Condor Experience"
- [9] Concurrency and Computation: Practice & Experience - Grid Performance Volume 17 Issue 2-4, February 2005 Pages 323 – 356.
- [10] Wiebe, J., and Philips, W., "Exploiting Subjectivity Classification to Improve Information Extraction", Proceedings of the 20th National Conference on Artificial Intelligence in 2005.
- [11] P. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews," in Proceedings of the Association for Computational Linguistics (ACL), pp. 417–424, 2002.
- [12] D Wang, S Zhu, and T Li. "SumView: A Web-based engine for summarizing product reviews and customer opinions" Expert Syst. Appl. 40(1):27-33 (2013).
- [13] B. Liu "Sentiment Analysis and Opinion Mining", April 22, 2012. Ah-hwee Tan "Text Mining: The state of the art and the challenges", In Proceedings of the PAKDD Workshop on Knowledge Discovery from Advanced Databases in 1999.
- [14] Kim, Y. and Myaeng, S., "Opinion Analysis based on Lexical Clues and their Expansion", Proceedings of NII Test Collection for Information Retrieval in 2007.
- [15] K. Evans, L.-W. Ku, Y. Seki, H.-H. Chen, and N. Kando "Opinion analysis across languages: An overview of and observations from the NTCIR6 opinion analysis illotask," Applications of Fuzzy Sets Theory Lecture Notes in Computer Science Volume 4578, 2007, pp 456-463
- [16] L. Dini and G. Mazzini, "Opinion classification through information extraction," in Proceedings of Conference on Data Mining Methods and Databases for Engineering, Finance and Other Fields Data Mining in 2002.
- [17] Ana-Maria Popescu, Oren Etzioni, "Extracting Product Features and Opinions from Reviews", Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, p.339-34 in 2005.
- [18] M. Baroni and S. Vegnaduzzo "Identifying subjective adjectives through web-based mutual information". In E. Buchberger, editor, In Proceedings of the Conference for the Processing of Natural Language and Speech KONVENS), pages 17--24, in 2004.
- [19] T. Yao and L. Li "A Kernel-based Sentiment Classification Approach for Chinese Sentences" World Congress on Computer Science and Information Engineering in March 31 2009-April 2009
- [20] Q. Su, X. Xu, H. Guo, Z. Guo, X. Wu, X. Zhang, B. Swen and Z. Su "Hidden Sentiment Association in Chinese Web" Proceedings of the 17th international conference on World Wide Web Pages 959-968 in 2008.4.