Pivotal Sentiment Tree Classifier

Vijayan Nagarajan, Punitha Chandrasekar

Abstract: Sentiment Analysis, also known as Opinion Mining, plays a vital role in social media analytics, call center data etc. There are many existing algorithms and methods to approachSentiment Analysis. Though these algorithms produce reasonable results, they fail to give near optimal (close to 100%) results. This paper aims to obtain maximum accuracy in Sentiment Analysis in-comparisonwiththe other existing algorithms and approaches.We devised a new algorithm called "Sentiment pivotal" tree to achieveresults with maximum accuracy by taking into account on few key factors like identifying the expectations of the customers along with the inclusion of neutral words for analysis.In order to attribute the above factor we introduce a new term called "Expective"along with the existing terms"Positive", "Negative"and"Neutral".

Index Terms: Sentiment Analysis, Opinion Mining, Expective Sentiment, Social Media Analytics, Text Analytics, Brand Specific Analysis, Natural Language Processing.

1 Introduction

Sentiment Analysis is a process of identifying the sentiment of the content in a text unit. There are several methods defined in the books as standards for performing sentiment analysis such as Natural language processing, Statistics, Machine learning, etc. Now-a-days Social Media content is one of the most challenging data set (by nature of its non grammatical structure) that requires much in-depth analysis in order to harness the abundant potential it boresuch as the comments from Facebook, Twitter, etc. There has been a lot of works in Sentiment Analysis field since 2002 (Semantic Orientation Applied to Unsupervised Classification of Reviews Peter D. Turney (National Research Council of Canada) (Submitted on 11 Dec 2002)). There aremuch software'sfor semantic separation (SentiBank WordNet-Affect SentiWordNet and SenticNet), word tagging, and sentiment tree for complex sentence (Recursive Deep Models for Semantic Compositionality over a Sentiment Treebank). Even though the usage of these algorithms (Ontology, Semantic_network) comes handy for analyzing sentences which are already in proper grammatical format, but thechallenge comes when we analyze raw data from social media sites as said earlier which doesn't pertain to any grammar.There are many steps to analyze these social media contents such as preprocessing, stop word removal, bag of words, Latent_semantic_analysis, Support_vector_machine, though a human intervention is required for the manual analysis to get more accurate results, as automated systems are unable to deliver more accurate results. To overcome this limitation, we are introducing a **Pivotal Sentiment Tree Classifier (PSTC).**PSTCuses the binary tree approach to give the more accurate results by classifying thecritical sentences correctly. The general terms which are used in Sentiment Analysis are – Positive, Negative, Neutral. Earlier neutral class is usually ignoredas per following thesis (Ding and Liu, 2008, Taboada et al, 2010, (Bo Pang and Lillian Lee, 2002, Wilson et al, 2005).However, Koppel and Schler (2006)proved in their research how neutral sentiment is important. They suggested that, in every polarity problem, three categories must be identified (positive, negative and neutral) and introduction of the neutral category can improve the overall accuracy. Business userswill generally focuson positive and negative sentimentsfor processing the customer reviews leaving out the significant factor "Neutral" in judging the results. Only countable numbers of Neutral comments are considered for devising the results. But, we think the most critical requirementfor a successful business is its focusto capture its potential customers. So far, there is no major formula to generate the customer's expectation meter. To target and grow the business keeping in mid these potential customers, we are introducing a new term called as EXPECTIVE sentiment**.EXPECTIVE** sentiment will find the customer expectation for a particular product. So our paper uses following terms POSITIVE, NEGATIVE, NEUTRAL and EXPECTIVE to classify the sentiments of the users.

2 Data Collection

We took data for Sony's product (Music Unlimited (MU)) mainly used in Android phones, PlayStation and other Sony devices fromsocial media's like Twitter, Facebook. This product is used in 51 countries, and in few other countries this app is in extreme demand.So, we are in an urge to collect these raw data and turn into useful information. Along with this, location details data arealso collected in terms of author name, published time, title, media type etc. As the product mainly deals with music applications, the majority of the data pertainsto the album/music share.

3 Proposed System

Our proposed system consists of four major steps in generating the Sentiment Analysis results.

1. Smiley'sExtraction
2. Text Cracking
3. Word Predictor
4. Stop Word Removal
5. Pivotal Sentiment Tree Classifier (PSTC)

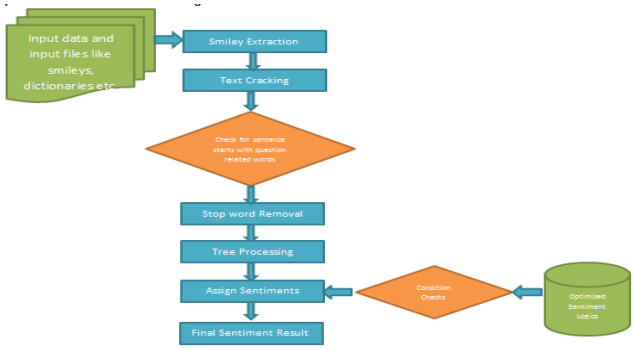

Fig. 1. *High Level process flow diagram of Pivotal Sentiment Tree Classifier Algorithm.*

The first three steps deal with preprocessing of the data collected from the social media, as they are not expected to have any grammar. It is of major focus to inputproper data for Pivotal Sentiment Tree Classifier (PSTC).

3.1 Smiley'sExtraction

Most of the existing algorithms will not includesmileys for its sentence analysis. But we believe the smileysalso aid in our journey for attainingan efficient analytic algorithm.As the very old cliché quote goes, `Pictures speak thousand words`;smileysdepict the exact opinion of the product from user's view. As an initial step, we give sentiment ratings for smileys in the comments which we get from MU.For example: "Oh is it good: P". It should be classified as negative sentiment, as the smiley depicts sarcasm. But most of the typical algorithms including Stanfordsentiment tree, will classify it as a positive sentiment.For easier analysis we segregate aset of positive and negative smileysin order to match and distinguish a smiley and a special character. Algorithm removes the special characters but keeps the smiley's and classifiesthem as POSITIVE, NEGATIVE, NEUTRAL or EXPECTIVE sentiment based on mode of the speech.

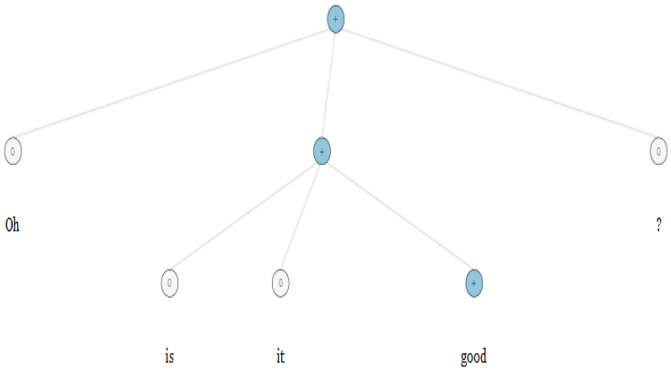

Fig. 2. *Stanford sentiment Treebank output of the comment "Oh is it good: P".*

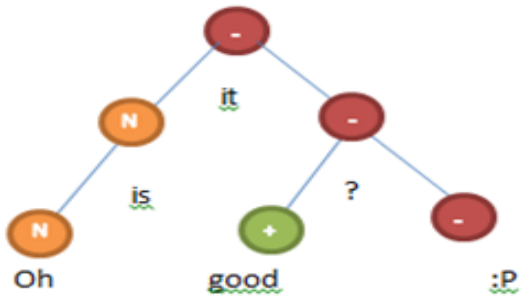

Fig. 3. *Pivotal Sentiment Tree Classifier output of comment "Oh is it good: P".*

3.2 Text Cracking

Comments can contain more than one sentence, and it is very important to split the sentences separately and process them accordingly. So, in the second step, the sentences are split based on predefined splitters such as dot (.), comma (,), semicolon (;), etc. A sentence splitters list is maintained to achieve the desired processing.For example, consider the below comment from Twitter about Sony's MU app: "@AskPlayStation just got an email about Music unlimited charging me 9.99 for another month, when I purchased I assumed it was for" The above comment will be divided into two parts, "@AskPlayStation just got an email about Music unlimited charging me 9.99 for another month" and "when I purchased I assumed it was for" as per aforementioned predefined splitters.

3.3Word Predictor

Social media data can be of any formats. We cannot expect proper grammar insocial media comments. For example, comments such as "I like it soooooo much", "It is veryyyyy horrible" will be giving some serious sentiments but not in correct grammar. There is a very high chance to omit the sentimentof these sentences and decide it as NEUTRAL which will end up in accurate results. So In order to overcome these limitations, Word Predictor is introducedto predict the correct word by maintaining some predefined set of words.

3.4Stop Word Removal

Stop words are the words which should be filtered out before Pivotal Sentiment Tree Processing. Definite lists of product specific stop words aremaintained in our experiment, which we got from Sony MU app's stop word list. After dividing the sentences into words, algorithm removes the words, which are neither not required for analysis nor can't be concluded to POSITIVE, NEGATIVE, NEUTRALandEXPECTIVE. Note: if the sentence starts from can, if, etc. then such words will not be considered as stop words as they represent a question in the sentence. A sample original comment is shown below with the process of stop word removal, "@PlayStation you guys should work with Spotify I mean music unlimited is cool but need a little work" List of Stop words: you, with, I, is, a. Final sentence: @PlayStation guys should work Spotify mean music unlimited cool but need little work

3.5 Pivotal Sentiment Tree Classifier (PSTC)

PSTC is designed based on binary search tree and recursive algorithm. After the analysis,the words are processed in tree format toget the sentiment result for a sentence.

4 Pivotal Element

Initially, a pivotal element is taken for all the sentences. Pivotal element can be chosen using below formulas:

$$P = \frac{n+1}{2} \; if \; n = odd$$

n – Number of words

$$P = \frac{n}{2} \; if \; n = even$$

n – Number of words

Where,'n' is the number of words in the preprocessed sentence. Chosen pivotal element will be the root for the PSTC tree. Then the words which are in left and right side of the pivotal word willbe processed till the last pivotal element in their respective sides. If a particular word does not contain any other left or right word, then it will be the last pivotal element. Till then the tree is filled by repeating the similar steps. After the tree gets filled, itis processed to get the sentiment.

5 Binary PSTC Search Tree

The algorithm looks for leaf nodes and chooses the sentiment among the words defined in sentiment word list. Sentiment word list is nothing but the list of predefined words with assigned sentiments such as NEGATIVE, POSITIVE, NEUTRAL or EXPECTIVE. Based on this list, sentiment will be assigned to the leaf nodes. Then the algorithm moveson to their parent node to assign its sentiment based on the sentiments of its child nodes. Find the sentiment assignment for the nodes in below table:

TABLE 1
SENTIMENT TABLE

Child 1 Sentiment	Child 2 Sentiment	Parent Sentiment
Positive	Positive	Positive
Positive	Negative	Negative
Positive	Neutral	Positive
Positive	Expective	Expective (If parent/super parent element is also Expective)
Negative	Positive	Expective (If parent/super parent element is also Expective)
Negative	Negative	Positive
Negative	Neutral	Negative
Negative	Expective	Negative
Neutral	Positive	Positive
Neutral	Negative	Negative
Neutral	Neutral	Neutral
Neutral	Expective	Expective (If parent/super parent element is also Expective)
Expective	Positive	Expective (If parent/super parent element is also Expective)
Expective	Negative	Expective(If parent/super parent element is also Expective)
Expective	Neutral	Expective (If parent/super parent element is also Expective)
Expective	Expective	Expective

6 Optimized Sentiment Logics

Sentiment assignment for the words may not work perfectly in all cases, because of the complication in the sentence.For example, the review comments, "It is good, but responds slowly" and "Even though it is better, it responds slowly", should be classifiedSOMEWHAT NEGATIVE. However,the existing algorithms classify themas POSITIVEandSOMEWHAT POSITIVE respectively.To get accurate results for these kinds of sentences, we use CORE WORDS(SENTIMENT CHANGE WORDS), Competitor Analysis and Immediate Sentiment methods. A list of core words is maintained to predict the conjunctions such as but, even though, though,etc. If any of these words occur in the list and if one part of the tree (left or right of the core word) conflictsthe sentiment, then the sentiment is assigned based on the core words list. In above sentence, PSTC gives the result as Core word + {POSITIVE, NEGATIVE}. So, the final result will be NEGATIVE, because core word contains POSITIVE sentiment and second part contains NEGATIVE sentiment. Another limitation in other algorithms is the sentiment classification for consecutive NEGATIVE sentiments orNEGATIVE followed by POSITIVE sentiment. Consecutive NEGATIVE should be classified POSITIVE and NEGATIVE followed by POSITIVE should be classified NEGATIVE. PSTC overcomes this limitation using Immediate Sentiment method. After sentiment processing, all these special scenarios are handled for optimized sentiment logics. Then, the final sentiment for the sentence is found.

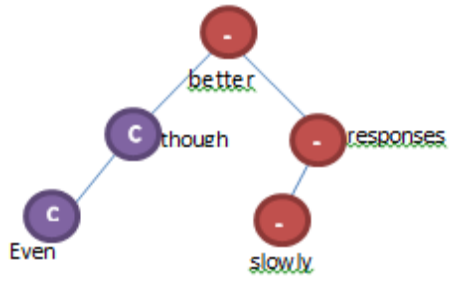

Fig. 4. *Pivotal Sentiment Tree Classifier output of sentence "Even though it is better, responses slowly"*

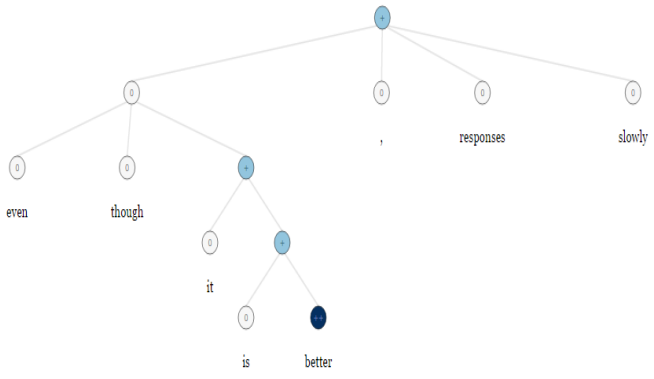

Fig. 5. *Stanford sentiment treebank output of sentence "Even though it is better, responses slowly"*

7 COMPETITIVE ANALYSIS

Any analyst formulating a rule for analysis will get struck ata point where they meet a dead end, as the taxonomy and word rules will help in finding POSITIVE or NEGATIVE sentiment to the sentences, it is very difficult to find whether it is POSITIVE to a given brand or the competitor's brand. Our algorithm bridges the gap between various algorithmsand classifiesthe brand specific sentiment as well.

E.g. 1: I love Pepsi not coke - Coke is brandand Pepsi is competitor

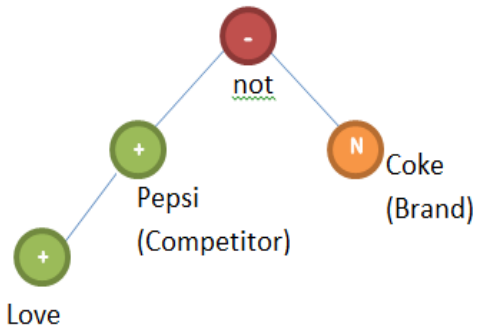

Fig. 6. *Pivotal Sentiment Tree Classifier output of sentence "Love pepsi not coke"*

Not is a pivot element here, love is POSITIVE, PepsiisNEUTRAL (as it is a brand name). As love is POSITIVE, Pepsi is classified as NEUTRAL +POSITIVE = POSITIVE. AsPepsi is the competitor, we'renot concerned here.Coke is NEUTRAL (as it is a brand name), but the pivot element not is NEGATIVE. So, the coke in the right is classifiedNEUTRAL + NEGATIVE = NEGATIVE. The results predict POSITIVE sentiment to the competitor and NEGATIVE to brand.And so, the final result is NEGATIVE for the taken brand.

E.g. 2:I love Pepsi not coke- Coke is competitor and Pepsi is brand

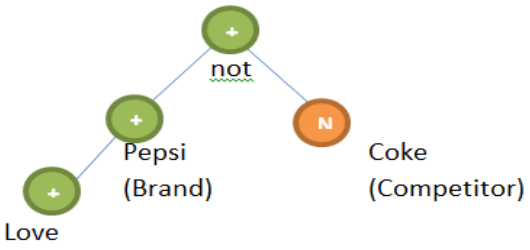

Fig. 7. *Pivotal Sentiment Tree Classifier output of sentence "Love pepsi not coke"*

Not is pivot element here, love is positive, Pepsi which is a brand name so its neutral but as love is positive, neutral + positive=positive. Coke which is a brand since it is neutral and having a negative pivotal element (not) and knowing coke brand as competitor brand here so the Pepsi brand will not be considered negative and immediate next to Pepsi is negative and immediate right is positive so the final result will be positive.

8 EXPECTIVE SENTIMENT PREDICTION AND EXPERIMENTS

Expective sentiment will give the expectation of customers. We will maintain some list of words and combinations to find the expectation of customer. The list may vary based on brand and competitor. We prepared expectation list for Sony MU app. On fly we will decide the words to the desired category based on the term it suits to. The Music Unlimited app on the Play station isn't bad but it would be cool if they put music from games on it.

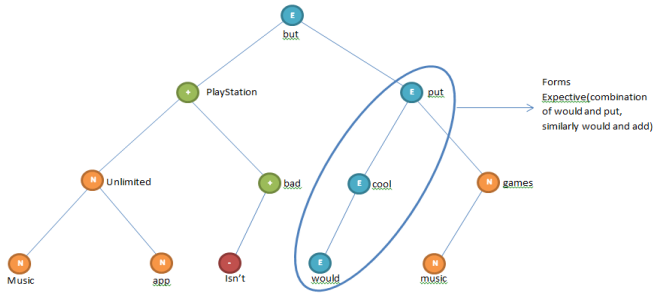

Fig. 8. *Pivotal Sentiment Tree Classifier output of sentence "The Music Unlimited app on the Playstation isn't bad but it would be cool if they put music from games on it"*

Example 2

Can a stop button be added in music unlimited?

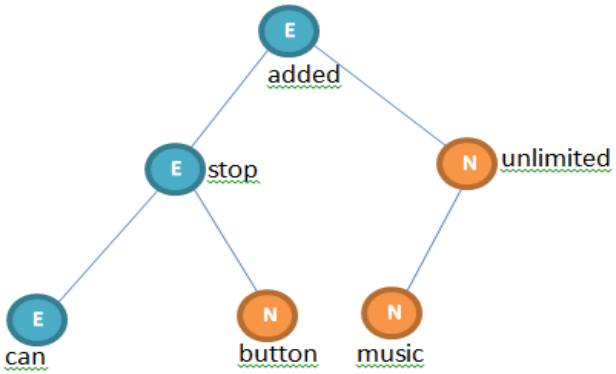

Fig. 9. *Pivotal Sentiment Tree Classifier output of sentence "can a stop button be added in music unlimited"*

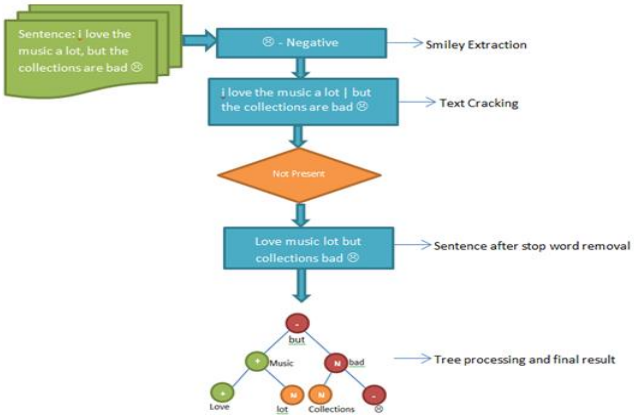

Fig. 10. *Sentence process flow of Pivotal Sentiment Tree Classifier.*

9 ACCURACY RESULT

Finally the most important thing is accuracy so we took the same Sony data and calculated with different algorithms such as Support Vector Machine (SVM), supervised Latent Dirichlet Allocation(sLDA), Naïve Bayes (NB), Maximum Entropy Classifiers (MAXENT) and Gazetter Classifier, PSTC algorithm gave more accurate results than other algorithms.

TABLE 2
ACCURACY RESULT TABLE

Value	PSTC	SVM	sLDA	NB	MAXENT	Gazetteer
Precision Positive	0.909820	0.603175	0.520408	0.199441	0.361868	0.400709
Recall Positive	0.921340	0.16521 7	0.22173 9	0.93043 5	0.404348	0.491304
Precision Negative	0.972130	0.632576	0.625	0.57714 3	0.535897	0.582677
Recall Negative	0.872310	0.445333 3	0.44533 3	0.26933 3	0.557333	0.394667
Precision Neutral	0.910200	0.609011	0.610433	0.869565	0.655449	0.625 85
Recall Neutral	0.891343	0.863363	0.82582 6	0.03003	0.614114	0.690691
Accuracy	0.912875	0.613 69	0.60660 9	0.26357 2	0.559402	0.56727

10 CONCLUSION

We introduced Pivotal Sentiment Tree Classifier to predict the sentiment of social media comments. We included competitor analysis for a particular brand. Data was crawled from public site such twitter and Facebook. We compared the results with other algorithms for the sentiment ratings on the same data.We can see higher accuracy for our Algorithm and also had an added advantage of processing very complicated sentences in a very efficient manner. For instance, PSTC obtained 91.4% accuracy on non-grammatical social media reviews than any other previous models.

11 REFERENCES

[1] Suin Kim, Jianwen Zhang, Alice Oh, Shixia Liu. A Hierarchical Aspect-Sentiment Model for Online Reviews

[2] Tetsuji Nakagawa, Kentaro Inui , Sadao Kurohashi. Dependency Tree-based Sentiment Classification

[3] Wei Wei, Jon Atle Gulla. Sentiment Learning on Product Reviews via Sentiment Ontology Tree

[4] Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng and Christopher Potts. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank

[5] Peter Turney (2002). "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews". Proceedings of the Association for Computational Linguistics. pp. 417–424. arXiv:cs.LG/0212032.

[6] Bo Pang; Lillian Lee and Shivakumar Vaithyanathan (2002). "Thumbs up? Sentiment Classification using Machine Learning Techniques". Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 79–86.

[7] Bo Pang; Lillian Lee (2005). "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales". Proceedings of the Association for Computational Linguistics (ACL). pp. 115–124.

[8] Benjamin Snyder; Regina Barzilay (2007). "Multiple Aspect Ranking using the Good Grief Algorithm". Proceedings of the Joint Human Language Technology/North American Chapter of the ACL Conference (HLT-NAACL). pp. 300–307.

[9] Vasilis Vryniotis (2013). "The importance of Neutral Class in Sentiment Analysis".

[10] Moshe Koppel; Jonathan Schler (2006). "The Importance of Neutral Examples for Learning Sentiment". Computational Intelligence 22. pp. 100–109.

[11] Thelwall, Mike; Buckley, Kevan; Paltoglou, Georgios; Cai, Di; Kappas, Arvid (2010)."Sentiment strength detection in short informal text". Journal of the American Society for Information Science and Technology 61 (12): 2544–2558. doi:10.1002/asi.21416.

[12] Pang, Bo; Lee, Lillian (2008). "4.1.2 Subjectivity Detection and Opinion Identification".Opinion Mining and Sentiment Analysis. Now Publishers Inc.

[13] Rada Mihalcea; Carmen Banea and Janyce Wiebe (2007). "Learning Multilingual Subjective Language via Cross-Lingual Projections". Proceedings of the Association for Computational Linguistics (ACL). pp. 976–983.

[14] Fangzhong Su; Katja Markert (2008). "From Words to Senses: a Case Study in Subjectivity Recognition". Proceedings of Coling 2008,

Manchester, UK.

[15] Bo Pang; Lillian Lee (2004). "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts". Proceedings of the Association for Computational Linguistics (ACL). pp. 271–278.

[16] Minqing Hu; Bing Liu (2004). "Mining and Summarizing Customer Reviews".Proceedings of KDD 2004.

[17] Bing Liu; Minqing Hu and Junsheng Cheng (2005). "Opinion Observer: Analyzing and Comparing Opinions on the Web". Proceedings of WWW 2005.

[18] Bing Liu (2010). "Sentiment Analysis and Subjectivity". Handbook of Natural Language Processing, Second Edition, (editors: N. Indurkhya and F. J. Damerau), 2010.

[19] Cambria, Erik; Schuller, Björn; Xia, Yunqing; Havasi, Catherine (2013). "New Avenues in Opinion Mining and Sentiment Analysis". IEEE Intelligent Systems 28 (2): 15–21.doi:10.1109/MIS.2013.30.

[20] Ortony, Andrew; Clore, G; Collins, A (1988). The Cognitive Structure of Emotions. Cambridge Univ. Press.

[21] Stevenson, Ryan; Mikels, Joseph; James, Thomas (2007). "Characterization of the Affective Norms for English Words by Discrete Emotional Categories". Behavior Research Methods 39 (4): 1020–1024.

[22] Kim, S.M. & Hovy, E.H. (2006). "Identifying and Analyzing Judgment Opinions.".Proceedings of the Human Language Technology / North American Association of Computational Linguistics conference (HLT-NAACL 2006). New York, NY.

[23] Lipika Dey , S K Mirajul Haque (2008). "Opinion Mining from Noisy Text Data".Proceedings of the second workshop on Analytics for noisy unstructured text data, p.83-90.

[24] Cambria, Erik; Hussain, Amir (2012). Sentic Computing: Techniques, Tools, and Applications. Springer.

[25] Carlo Strapparava; Alessandro Valitutti (2004). "WordNet-Affect: An affective extension of WordNet". Proceedings of LREC. pp. 1083–1086.

[26] Stefano Baccianella; Andrea Esuli and Fabrizio Sebastiani (2010). "Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining". Proceedings of LREC. pp. 2200–2204. Retrieved 2014-04-05.

[27] Erik Cambria; Daniel Olsher; Dheeraj Rajagopal (2014). "SenticNet 3: A common and common-sense knowledge base for cognition-driven sentiment analysis". Proceedings of AAAI. pp. 1515–1521.

[28] Damian Borth; Rongrong Ji, Tao Chen, Thomas Breuel and Shih-Fu Chang (2013)."Large-scale Visual Sentiment Ontology and Detectors Using Adjective Noun Pairs".Proceedings of ACM Int. Conference on Multimedia. pp. 223–232.

[29] "Case Study: Advanced Sentiment Analysis". Retrieved 18 October 2013.

[30] Boris Galitsky, Eugene William McKenna. "Sentiment Extraction from Consumer Reviews for Providing Product Recommendations". Retrieved 18 November 2013.

[31] Boris Galitsky, Gabor Dobrocsi, Josep Lluís de la Rosa (2010). "Inverting Semantic Structure Under Open Domain Opinion Mining". FLAIRS Conference.

[32] Boris Galitsky, Huanjin Chen, Shaobin Du (2009). "Inversion of Forum Content Based on Authors' Sentiments on Product Usability". AAAI Spring Symposium: Social Semantic Web: Where Web 2.0 Meets Web 3.0: 33–38.

[33] Ogneva, M. "How Companies Can Use Sentiment Analysis to Improve Their Business". Retrieved 2012-12-13.

[34] Wright, Alex. "Mining the Web for Feelings, Not Facts", New York Times, 2009-08-23. Retrieved on 2009-10-01.

[35] Kirkpatrick, Marshall. ", ReadWriteWeb, 2009-04-15. Retrieved on 2009-10-01.