International Journal of Innovative and Emerging Research in Engineering
Volume 3, Special Issue 1, ICSTSD 2016

Text-Based Sentiment Analysis

Swati M. Chandurkar
Dept. of Computer Science & Engg.
PRMIT&R , Badnera .
swatichandurkar24@gmail.com

Dr. M. A. Pund
Dept. of Computer Science & Engg.
PRMIT&R , Badnera .
mapund@mitra.ac.in

Abstract— Sentiments are nothing but the feelings like attitude, emotion, or opinion and sentiment analysis is the kind of text classification according to sentiments present in text. Sentiments can be expressed in positive and negative polarity. Sentiment analysis helps in finding how sentiments can be expressed in text. Text-based sentiment analysis is becoming popular in text mining and computational linguistic research. In these days, increase in the use of social media like blogs and social networks has increased the use of sentiment analysis. One of the wide use of text-based sentiment analysis is in taking decision about product at the time of both purchasing and manufacturing. This paper focuses on problem of sentiment polarity categorization , which is basic problem of sentiment analysis. In this paper , process of of determining the polarity and features of the product based on the considered reviews is proposed with detailed process description.

Keywords— sentiment analysis; text classification; text-to-speech synthesis.

I. INTRODUCTION

Sentiment Analysis comes under natural language processing and task of information extraction which can obtain feelings of writers expressed in positive or negative comments, questions and requests, by analyzing a large numbers of documents. In these days , increase in use of Internet and exchange of public opinion has increased the demand of Sentiment Analysis. Text-Based Sentiment Analysis includes finding correct entity from the text towards which the sentiment is directed. Text-Based Sntiment Analysis classifies text based on the sentimental orientation of opinion they contain. The sentiments can further be given a score based on their degree of positivity, negativity or objectivity.

II. RELATED WORK

Lina Zhou et al, investigated movie review mining using machine learning and semantic orientation [20] and in the proposed machine learning approach, Supervised classification and text classification techniques are used for classifying the movie review and the study says that the supervised machine learning is more efficient but requires a considerable amount of time to train the model.

Bo Pang et al., used machine learning techniques for investigating the effectiveness of classification of documents by overall sentiment [21].

Zhu et al., proposed aspect based opinion polling from free form textual customers reviews [22].

Jeonghee Yi et al., proposed a Sentiment Analyzer for extracting opinions about a subject from online data documents [23]. Sentiment analyzer uses natural language processing techniques.

Alekh Agarwal et al., proposed a machine learning method incorporating linguistic knowledge gathered through synonymy graphs, for effective opinion classification [24].

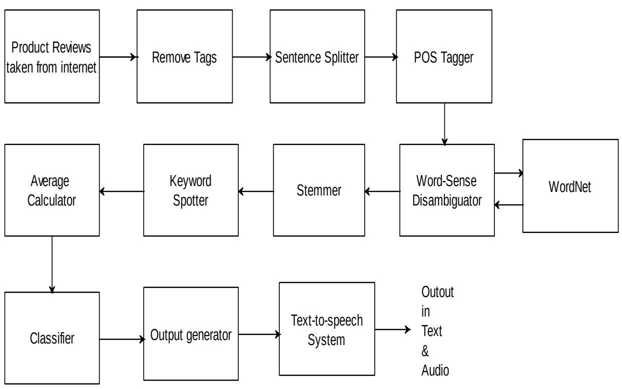
Michael et al., presented, a prototype system for mining topics and sentiment orientation from free text customer feedback [25].

Qui et al., analyzed the problems related to opinion mining such as opinion lexicon expansion and opinion target extraction [26].

III. TEXT-BASED SENTIMENT ANALYSIS

A. Framework

Fig. 1. Fig 1: Framework for text-based sentiment analysis



1) Reviews of product

Reviews of the product are collected from related websites using Internet. These reviews are used as input to the framework of text-based sentiment analysis.

2) Remove Tags

Reviews of the product which are used as an input are real time reviews. Such real time reviews may contain non-textual data and markup tags for html pages which are not required for sentiment analysis and thus should be removed.

3) POS Tagger

Part-of-speech (POS) information is useful in sentiment analysis. It is used to explain the significance of the words in the document. The adverbs, adjectives and verbs have significant effect on the sentence. Thus this module helps in identifying the most affective word in the given document.

4) Word-Sense Disambiguator

Human language is ambiguous, that is a word with same spelling can be used with different meanings depending on the context in which it occurs. For instance, consider the following sentences:

(a) The class learned that information.

(b) She is very learned individual.

Here the word learned in the two sentences has different meanings:

Acquired knowledge and a highly educated or knowledgeable, respectively. Word-sense disambiguator is used to identify the meaning for words in context.

5) Stemmer

It is used for the purpose of indexing. Words which are related semantically get mapped to the same stem, base or root form. Stemming is useful for reducing inflected words to their word stem, base or root form. Stemming programs is also called as stemming algorithms or stemmers.

6) Keyword Spotter

This module is used to develop monograms and bigrams. Keyword spotter takes up three features where initially the word features are identified by confirming whether or not the word exists in the document. It is like that of the next feature to identify the unigrams in the word and hence check for their availability. It provides the emotional dimensions to the emotional words.

7) Average calculator

It is used to compute the averaged emotional dimensions for the text of analysis. In the current work, this is the arithmetic mean of the dimensions at the sentence level [9], [11].

8) Classifier

Sentiments can be expressed on positive and negative polarity. Classifier helps to identify the class of input text that is whether it is positive or negative.

9) Processing by Text-to-speech System

In these days, speech researchers are taking interest to focus on full range and variation of speech for signaling the social and psychological aspects of a message. The new generation of Text-To-Speech (TTS) systems should automatically deliver expressive cues when synthesizing an affective message [5], [6]. Finding out the affect in text is difficult task. There are hurdles in translating human affect into explicit representations. Natural Language Processing (NLP) research community says that sentiment analysis uses positive and negative polarity for problem classification. Some authors also consider a neutral sentiment at the same level of the hierarchy [19]. TTS applications make use of this neutral sentiment for generating usual messages [15], [16].

B. Term selection and weighting

In sentence level sentiment analysis, each sentence denotes some features related to product to decide whether that product is good or bad. But we cannot take all the features in consideration as it increase the feature space and does not provide satisfactory result. Therefore term selection and term weighting are used.

1. Term Selection:

This process helps in reducing the dimensionality of features space. It removes the feature which does not contribute for classification task. This helps in maximizing the effectiveness and computational performance of classifier.

2. Term weighting:

This process is used to increase the discriminating power of certain features. It does not reduce the dimensionality of features space.

Leung et al. suggest a method to find the sentiment orientation and opinion strength of a word with respect to a sentiment class as its relative frequency of appearance in that class. For example, if the word “best” appeared 8 times in *Positive* reviews and 2 times in *Negative* reviews, its strength with respect to *Positive* sentiment orientation is then 8/(8+2) = 0.8.

C. Machine Learning

Machine learning approach is related to topic classification, that is classifying sentiments in positive and negative reviews.

Pang et al. uses three classifiers which includes Naïve Bayes, Support Vector Machines (SVM) and Maximum

Entropy, for a movie review corpus to check whether binary sentiment classification can be addressed using standard topic classification techniques [27]. Pang and Lee concluded that sentiment classification is more difficult than topic classification.

Goldberg and Zhu then extended work of Pang and Lee using transductive semi-supervised learning. They demonstrated that unlabeled reviews can help improve classification accuracy [29].

Zhu and Goldberg proposed a kernel regression algorithm utilizing order preferences of unlabeled data, and successfully applied the algorithm to sentiment classification.

IV. Experimental Result Analysis

The parameters like precision and recall helps in computing the result of sentiment analysis.

Precision and recall play a major role in evaluation of search strategies. Precision can be considered as one of the measure of calculating the effectiveness of some computer applications to determine search words, candidate terms, and other items. Precision is a measure of the proportion of the results of a computer application that are supposed to be correct.

Recall can also be considered as one of the measure of finding the effectiveness of some computer applications to find search words, candidate terms, and other items. Recall is a measure of the proportion of all possible correct results of a computer application that the application actually produces. For example, suppose you are using a computer application to search for terms in a document that has 100 terms in it. If the application finds 90 of these terms, then the recall of the application is 90 out of 100, or 90%. Here, these two parameters are evaluated for checking the accuracy of the system.

The rule based approach of the system is tested with reviews of the three products and the following results are calculated Table 1 shows the results of review calculation for three products.

Table 1: Reviews calculation

Sr No	Product Name	Positive Word	Negative Word	Neutral Word	Positive Review %
1	Samsung Laptop	160	34	160	82%
2	Nokia Mobiles	140	32	132	81%
3	Sony Laptop	125	50	149	71%

The table given below will provide us the number of adjectives and adverbs found for the product by the system and also the number of adjectives and adverbs which were not found, it is counted by us. The Table shows recall value calculated for adjective and adverbs.

Table 2: Recall calculation

Sr. No.	**Product**	**No.of Adj Foud**	**No. of Adj Not foud**	**Recall (adj)**	**No of Adv Foud**	**No. Adv Not Foud**	**Recall (Adv)**	**Avg Recall**
1	Samsung Laptop	65	34	66%	67	41	62%	64%
2	Nokia Mobiles	85	46	65%	68	32	68%	66.5%
3	Sony Laptop	95	34	74%	79	34	70%	72%
Overall Average Recall								67.5%

The precision ratio of the correct reviews and the total reviews are calculated and given in the table no.3.

Table 3: Polarity precision calculation

Sr. No	Product	Correct review	Total review	Precision
1	Samsung laptop	241	341	71%
2	Nokia mobile	190	280	68%
3	Sony laptop	249	345	72%
Overall Average Precision				70%

The precision ratio of correct reviews with total reviews is considered for the three products and the precision value is calculated in the Table 3.

In this way the results are calculated and get the actual performance of the system.

V. Conclusion

Sentiment Analysis is a task of information extraction which can obtain feelings expressed in positive and negative comments. The proposed framework of text-based sentiment analysis can find the features of product based on considered review and also determine whether those features denotes positive or negative polarity and provides the user accurate result generated by proposed system.

References

[1] N. Campbell, “Conversational speech synthesis and the need for some laughter,” IEEE Trans. Audio, Speech, Lang. Process., vol. 14, no. 4, pp. 1171–1178, Jul. 2006.

[2] M. Skowron, H. Pirker, S. Rank, G. Paltoglou, J. Ahn, and S. Gobron, “No peanuts! Affective cues for the virtual bartender,” in Proc. FLAIRS ’11, May 2011.

[3] R. Calix, S. Mallepudi, B. Chen, and G. Knapp, “Emotion recognition in text for 3-D facial expression rendering,” IEEE Trans. Multimedia, vol. 12, no. 6, pp. 544–551, Oct. 2010.

[4] T. Wilson and G. Hofer, “Using linguistic and vocal expressiveness in social role recognition,” , pp. 419–422, 2011.

[5] F. Alías, X. Sevillano, J. C. Socoró, and X. Gonzalvo, “Towards high-quality next-generation text to-speech synthesis: A multidomain approach by automatic domain classification,” IEEE Trans. Audio, Speech, Lang. Process., vol. 16, no. 7, pp. 1340–1354, Sep. 2008.

[6] J. Bellegarda, “A data-driven affective analysis framework toward naturally expressive speech synthesis,” IEEE Trans. Audio, Speech, Lang. Process., vol. 19, no. 5, pp. 1113–1122, Jul. 2011.

[7] J. Pitrelli, R. Bakis, E. Eide, R. Fernandez,W. Hamza, andM. Picheny, “The IBM expressive text-to-speech synthesis system for American English,” IEEE Trans. Audio, Speech, Lang. Process., vol. 14, no. 4, pp. 1099–1108, Jul. 2006.

[8] C. O. Alm, Affect in Text and Speech. Saarbrücken, Germany: VDM Verlag, 2009.

[9] V. Francisco, R. Hervás, F. Peinado, and P. Gervs, “EmoTales: Creating a corpus of folk tales with emotional annotations,” Lang. Res. Eval., vol. 45, pp. 1–41, Feb. 2011.

[10] M. Schröder, H. Pirker, M. Lamolle, F. Burkhardt, C. Peter, and E. Zovato, “Representing emotions and related states in technological systems,” Emotion-Oriented Systems—The Humaine Handbook, pp. 367–386, 2011.

[11] G. O. Hofer, K. Richmond, and R. A. J. Clark, “Informed blending of databases for emotional speech synthesis,” in Proc. Interspeech ’05, Sep. 2005.

[12] C. O. Alm, D. Roth, and R. Sproat, “Emotions from text: Machine learning for text-based emotion prediction,” in Proc. HLT’05, 2005, pp. 579–586, .

[13] A. R. F. Rebordao, M. A. M. Shaikh, K. Hirose, and N. Minematsu, “How to improve TTS systems for emotional expressivity,” in Proc. Interspeech’09, Sep. 2009, pp. 524–527.

[14] M. A. M. Shaikh, A. R. F. Rebordao, K. Hirose, and M. Ishizuka,“Emotional speech synthesis by sensing affective information from text,” in Proc. ACII, Sep. 2009, pp. 1–6.

[15] A. Trilla, F. Alías, and I. Lozano, “Text classification of domain-styled text and sentiment-styled text for expressive speech synthesis,” in Proc. VI Jornadas en Tecnología del Habla (FALA2010), Nov. 2010, pp. 75–78.

[16] A. Trilla and F. Alías, “Sentiment classification in english from sentence-level annotations of emotions regarding models of affect,” in Proc. Interspeech’09, Sep. 2009, pp. 516–519.

[17] S. Tan and X. Cheng, “Improving SCL model for sentiment-transfer learning,” in Proc. NAACL-HLT, 2009, pp. 181–184 .

[18] B. Pang and L. Lee, “opinionmining and sentiment analysis,” Foundat. Trends in Inf. Retrieval, vol. 2, no. 1–2, pp. 1–135, 2008.

[19] P. Shaver, J. Schwartz, D. Kirson, and C. O’Connor, “Emotion knowledge: Further exploration of a prototype approach,” J. Pers. Soc. Psychol., vol. 52, no. 6, pp. 1061–1086, 1987

[20] Lina Zhou, Pimwadee Chaovalit, “Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches”, Proceedings of the 38th Hawaii International Conference on system sciences, 2005.

[21] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan, “Thumbs up? Sentiment classification using machine learning techniques”, In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 79–86, 2002.

[22] Yi, J., T. Nasukawa, R. Bunescu, and W. Niblack: 2003, “Sentiment Analyzer: Extracting Sentiments about a Given Topic using Natural Language Processing Techniques”, In: Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM-2003). Melbourne, Florida.

[23] Alekh Agarwal and Pushpak Bhattacharyya, “Sentiment analysis: A new approach for effective use of linguistic knowledge and exploiting similarities in a set of documents to be classified”, In Proceedings of the International Conference on Natural Language Processing (ICON), 2005.

[24] Michael Gamon, Anthony Aue, Simon Corston-Oliver, and Eric Ringger, “Pulse: Mining Customer Opinions from Free Text Natural Language Processing”, Microsoft Research, Redmond, WA 98052, USA.

[25] Guang Qiu, Bing Liu, Jiajun Bu and Chun Chen. "Opinion Word Expansion and Target Extraction through Double Propagation." Computational Linguistics, March 2011, Vol. 37, No. 1: 9.27.

[26] Lei Zhang and Bing Liu. "Identifying Noun Product Features that Imply Opinions." ACL-2011 (short paper), Portland, Oregon, USA, June 19-24, 2011.

[27] Sanjiv Das and Mike Chen. Yahoo! for Amazon: Extracting market sentiment from stock message boards. In Proceedings of the Asia Pacific Finance Association Annual Conference (APFA), 2001.

[28] Kushal Dave, Steve Lawrence, and David M. Pennock. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In Proceedings of WWW, pages 519–528, 2003.

[29] Tetsuya Nasukawa and Jeonghee Yi. Sentiment analysis: Capturing favorability using natural language processing. In Proceedings of the Conference on Knowledge Capture (K-CAP), 2003.

[30] Erik Cambria, Robert Speer, Catherine Havasi, and Amir Hussain. Senticnet: A publicly available semantic resource for opinion mining. Artificial Intelligence, pages 14–18, 2010.

[31] Mostafa Karamibekr, Ali A. Ghorbani, Verb Oriented Sentiment Classification, 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology.

[32] Costin-Gabriel Chiru, Asmelash Teka Hadgu, Sentiment-Based Text Segmentation, 2013 2nd International Conference on Systems and Computer Science (ICSCS) Villeneuve d'Ascq, France, August 26-27, 2013

[33] Ms.K.Mouthami, Ms.K.Nirmala Devi, Dr.V.Murali Bhaskaran, Sentiment Analysis and Classification Based On Textual Reviews,

[34] Rabia Batool , Asad Masood Khattak , Jahanzeb Maqbool and Sungyoung Lee, Precise Tweet Classification and Sentiment Analysis

[35] Ellen Spertus. Smokey: Automatic recognition of hostile messages. In Proceedings of Innovative Applications of Artificial Intelligence (IAAI), pages 1058–1065, 1997.

[36] Xin Jin, Ying Li, Teresa Mah, and Jie Tong. Sensitive webpage classification for content advertising. In Proceedings of the International Workshop on Data Mining and Audience Intelligence for Advertising, 2007.