A Survey of Dual Sentiment Analysis Considering Two Sides of One Review

Ashwini Damodar Kanherkar[1], Sarika M. Chavan[2]

Department of Computer Science & Engg, Deogiri Institute of Engg & Mgmt Studies, Aurangabad[1,2]

Abstract: With flourishing of social media such as blogs, forum, social network, etc... Sentiment analysis is an excellent source for gathering the public opinions. This paper fully focuses on an exhaustive study of dual sentiment analysis considering two sides of one review. It is used to analyze and compare the public opinion for the product review. Sentiment analysis is a kind of natural language processing for tracking the mood of the public about a particular product or topic. Sentiment analysis, which is also called opinion mining, involves in building a system to collect and examine opinions about the product made in comments, reviews, blog posts, or tweets. Sentiment analysis can be useful in several ways. Polarity classification is the most classical sentiment analysis task which aims at classifying reviews into either positive or negative. However, in many cases, in addition to the positive and negative reviews, there still exist many neutral reviews. This survey paper focuses a comprehensive overview of the last updates in this era. Many recently proposed enhancements and various sentiment analysis applications are entitled and presented briefly in this survey.

Keywords: machine learning, sentiment analysis, natural language processing, opinion mining.

1. INTRODUCTION

Sentiment Analysis is an area of focus over last decade. Increase in user-generated content provide an important aspect for the researchers, industries and government to mine this information and is truly differentiating and valuable to today's corporations. The user-generated content is an important source for various organizations to know/learn/identify the general expression/sentiment of different users on the product. The Social Web has changed the ways people communicate, collaborate, and express their opinions. The potential for the sharing of opinions today is unmatched in history.

So many knowledgeable people been connected by such a time and cost efficient and effective network. Due to the vast growth and emergence of the consumer generated media (CGM) on internet such as blogs, forums ,websites and news articles, ecosystem of corporations has changed significantly[1]. Customers, retailers are tremendously interested in and about reviews and insight of companies, their products and services, brands offered on the web. The reviews of customers are really important to attract huge number of customers. In particular, an important form of insights can be derived from sentiment analysis from the web contents[2]. Recently, sentiment analysis of online customer reviews has emerged as a very important research topic.

In text mining, Sentiment Analysis and Opinion Mining consists study of sentiments, attitudes, reactions, their emotions and evaluation of the content of the text. Many times while analyzing people's opinions, attitudes, reactions, sentiments, evaluations towards entities, such as services, products, organizations, individuals , events, topics, issues and their attributes. Sentiment Analysis is also known as Opinion Mining. Sentiment Analysis (SA)

or Opinion Mining (OM) is nothing but the computational study of people's opinions, attitudes and emotions towards an entity. The entity can represent events ,topics or individuals. These topics are likely to be covered by reviews. The two expressions SA or OM are actually interchangeable and they express a mutual meaning. Some researchers also said that OM and SA have slight different notions. Opinion Mining extracts and analyzes people's opinion about an entity while Sentiment Analysis identifies sentiment expressed in a text then analyzes it.

Therefore, the aim of SA is to find opinions, identify the sentiments they express and then classify their polarity. Basically sentiment analysis have two types of polarity: i) Positive polarity and ii) Negative Polarity[3]. An object which holds the positive opinion comes under the positive polarity. (e.g., awesome, happy, nice, joy, fun, excellent).An object which holds the negative opinion comes under the negative polarity. (e.g., bad, worst, rubbish, terrible).

Feature based Sentiment classification,Sentiment classification and opinion summarization are some main fields of research where Sentiment analysis is predominant [4]. Main exploration in the realm of Sentiment Analysis and opinion mining are: sentiment classification, feature based Sentiment classification and opinion summarization. Sentiment classification bestows with classifying entire documents or text or review in conformity with the opinions towards certain objects.

Feature-based Sentiment classification on the other side deems the opinions on features of certain objects. For example, in reviews related to laptops classifying the sentiments only on the basis speaker quality.

2. PROCESS OF SENTIMENT ANALYSIS

Radically the sentiment analysis is done with the process as shown in the figure 1. There are total four prime steps:

- Text Extraction -This step implicates extracting words from text that influence the outcome of the result.
- Text Refinement -This step has involvement of refining text in form of relevant phrases, words etc.

- Text Classification – This step includes classification of text into its class as positive, negative.
- Score Aggregation – This step assembles total scores from classifier and then aggregates it forward to produce the total sentiment score.
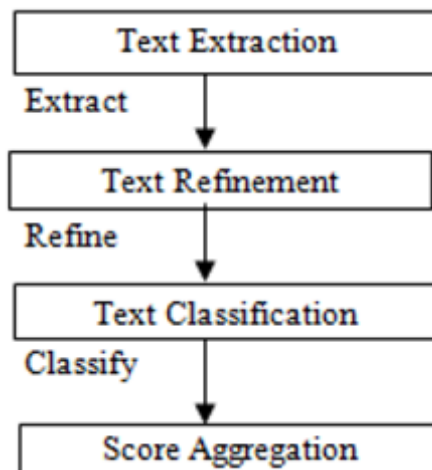

Figure 1: Steps of sentiment analysis

3. SENTIMENT ANALYSIS TECHNIQUES

a) Supervised approach
Supervised approach is called Machine learning. Machine learning technique utilizes training data to build predictive model. Predictive models such as decision trees, logistic regressions or neural networks are serviced to make prediction on documents which are present outside the training set. This approach has advantage as it is based on learning patterns that are beneficial in making automated and efficient predictions. Also the algorithms are able-bodied of discovering complex and unimagined patterns that would be beyond what a human could wean. However it has drawbacks as large training data is necessary to build the model and consolidating it is time consuming and challenging.

A rating is need to be provided for every document, and if there are attributes of documents it should provide a rating for each of these as well. Another complication unsprings if two different reviewers assign two different sentiment ratings to the same document, then this can involve unexpected errors in building and measuring the performance of model.

b) Unsupervised approach
Unsupervised approach is known as Natural language processing. Natural language processing (NLP) is an era of artificial intelligence that contributes with automatically extracting meaning from natural language text. It utilizes entities and syntactic patterns in the text to understand its meaning. It also services an amalgamation of language dictionaries, linguistic constructs like parts of speech, noun phrases along with a range of operators.

The major role of rule-based methods is that it supplies freedom for the rule developers to use their domain knowledge to devise rules for analysis purpose. Rule-based methods are totally unsupervised and they do not require any training data. This is a main advantage in real-life applications where training data is scant. Additionally it provides the amenity to refine the rules over a time based on the feedback from analysts or subject-matter experts to adjust the models. The major problem with NLP approach is that they need a lot of human involvement in developing the rules and it completely rely on the domain knowledge of rule developers.
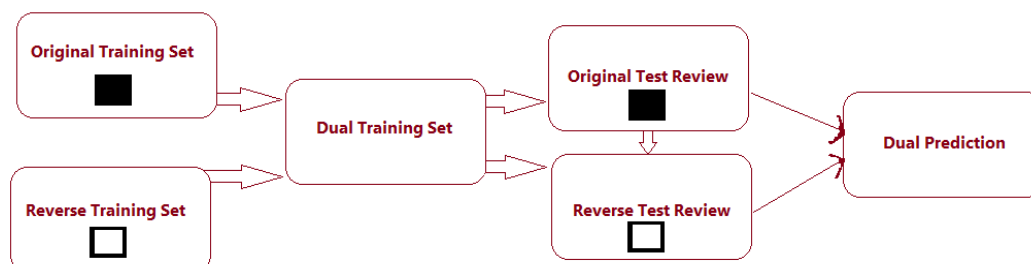

Figure 2: Dual sentiment analysis

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 6, June 2016

4. DUAL SENTIMENT ANALYSIS

Above figure illustrates the process of dual sentiment analysis (DSA). The Black rectangle denotes the original data, and the White filled rectangle denotes the reversed data. DSA contains two main stages: 1) dual training and 2) dual prediction[5].

4.1 Dual Training

In the training stage, all of the original training samples are reversed to their opposite ones and we refer to them as "original training set" and "reversed training set" respectively. In the data expansion technique, there is a one-to-one correspondence in between the original and reversed reviews. The classifier is trained by maximizing a combination of the feasibilities of the original and reversed training samples. This process is known as dual training.

4.2 Dual Prediction

In the prediction stage, for every test sample x, we create a reversed test sample $\tilde{x}$. Note that the target is not to predict the class o $\tilde{x}$. But rather, we use $\tilde{x}$ to assist the prediction of x. This process is entitled as dual prediction. A simple yet efficient model, known as dual sentiment analysis (DSA) addresses the polarity shift problem in sentiment classification.

4.3 DSA with Selective Data Expansion

In dual training, all of the training reviews are used in data expansion. However, in many cases, not all of the reviews have different sentiment polarity.

For example-

- Review (a). The house is very beautiful, and the price is very cheap. I like it.
- Review (b). The house is somehow beautiful, but the price is too expensive. I don't dislike it.

In both review, for review (a), the sentiment is strong and the polarity shift rate is low. In this case, the original review and the reversed review will also be a good one. for review (b), the sentiment polarity is less distinct. In such case, the sentiment polarity of the reversed review is also not distinct and confident. Therefore, creating reversed review for review (b) is not that essential in comparison with review (a) [6].

5. EXISTING RESEARCH IN SENTIMENT ANALYSIS

Existing studies found in the literature of sentiment analysis has significant highlights on sentiment classification, which is intended to differentiate user opinions and classify opinion comments into positive, negative and neutral categories. Following are some papers describe different kinds of researches in the era of sentiment analysis.

A] Effects of Adjective Orientation and Gradability on Sentence Subjectivity

To compute the subjectivity of a sentence, Hatzivassiloglou and Wiebe [7] presented supervised classification technique to forecast sentence subjectivity. Hatzivassiloglou and Wiebe enlightened the overall effects of semantically oriented adjectives, dynamic adjectives and also gradable adjectives on predicting subjectivity of the text document holding reviews. The sentences in a document are either subjective or objective to be visible to this Pang and Lee [8] proposed a sentence-level subjectivity detector. This explained technique retains subjective sentences and discards the objective sentences. After that they applied sentiment classifier. Major task of sentiment classifier is to contemplate resulted subjectivity with enhanced results.

B] Thumbs up: Sentiment Classification Using Machine Learning Techniques

Pang et al. [9] introduced machine learning model as maximum entropy, naive Bayes, and support vector machines to sort entire movie reviews into negative or positive sentiments,. They concluded results generated by standard machine learning methods are rover to result by human-generated baselines. Yet machine learning method performs well on only traditional topic based categorization and lack in functionality on sentiment classification.

C] Hidden Sentiment Association in Chinese Web Opinion Mining

To categorize review documents into positive or negative in which as thumbs up represented positivity of document and thumbs down represents negativity of document an unsupervised learning method was stated [10]. Average sentiment orientations of phrases and words are counted for each review document to compute sentiment of review document. Domain-dependent contextual information is serviced to anticipate sentiments of phrases in review document, but this technique has limitation as it relies on external search engine.

D] Sentiment Analysis of Chinese Documents: From Sentence to Document Level

Zhang et al. [11] presented a rule-based semantic analysis technique to distribute sentiments for text reviews. Word dependence structures are supplied to classify the sentiment of a sentence. The author has predicted document-level sentiments by aggregating sentiments of sentence. This technique has some limitation as rule-based methods experience poor exposure also they do not hold comprehensiveness in their rules.

E] Thumbs Up or Thumbs Down: Semantic Orientation Applied to Unsupervised Classification of Reviews

To avoid the above limitation, Maas et al. [12] introduced method for both document-level and sentence-level sentiment classification. The proposed method provides integration of unsupervised and supervised approaches to learn vectors and for learning process, they take semantic term-document information as well as rich sentiment content.

F] Effective Sentiment Analysis of Social Media Datasets using Naive Bayesian Classification

Dhiraj Gurkhe et al. (2014) [13] uses combination of different labeled dataset. Various Approaches like Machine Learning Unigram, Bigram, Unigram are used. It gives best results with Unigram detection without neutral labels. But the drawback is that it leads to less accuracy as

the size of training data is less also the sarcasm cannot be detected.

G] Twitter Sentiment Analysis using Machine Learning and Knowledge-based Approach Riya Such dev et al. (2014) [14] introduced the use of hybrid approach gives 100% of accuracy. Techniques used are machine Learning & knowledge-based approach using feature vector. The dataset used is Sanders Analytics dataset.

H] Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study
There has been a recent encourage of interest in sentiment analysis, Eman M.G. Younis et al. (2015) [15] proposed that Sentiment Analysis can be carried out by unsupervised technique without using pre classified training dataset. System gives a way to improve business competitive and customer relationship management in real-time. It has drawback that it leads to false sentiment classification as sarcasm cannot be detected.

I] TwiSent: A Multistage System for Analyzing Sentiment in Twitter
Subha-brata Mukher-jee et al. [17] expressed a hybrid approach that is machine learning approach & Rule based approach with extended module for each phase in architecture. TwiSent achieves higher negative precision improvement than positive precision improvement, it can capture negative sentiment strongly. But System cannot capture sarcasm or implicit sentiment.

6. APPLICATIONS

Sentiment Analysis has many applications in various fields. The application from a user's standpoint is the applications related to review websites. Application of Sentiment Analysis is important in the automatic summarization of user reviews. Automatic summarization is nothing but the creation of summary of the entire review using an automated program.

In case of user reviews, a new user find it difficult to look at all the reviews thoroughly and understand what aspect of the product is not get appreciated. Thus, there is a huge need of a summarizing application that will briefly inform the user about the polarity of the reviews. It has been observed that more and more people nowadays like to look upon reviews of products online before buying them. And for many businesses the online opinion can make evaluation of their product [17].

Hence, sentiment analysis plays a vital role in businesses. Businesses wish to understand the online reviews in order to improve quality of their products which in turns their reputation. In trend prediction system sentiment analysis can also be used. By tracking public opinion, important data regarding sales trends and customer satisfaction can be extracted. So far mentioned only applications pertaining to the business setting. But, Sentiment Analysis used various applications across the different domains. Studies in sociology and other fields have been aided by Sentiment Analysis systems which show trends in the human emotions especially on social networks.

CONCLUSION

Thus the basic knowledge required to do sentiment analysis is well explained in this review paper. What is Sentiment Analysis with respect to stages of dual sentiment analysis, what are the approaches to do sentiment analysis, process for doing sentiment analysis, existing research and recent updates in this era of sentiment analysis and the applications where it can be utilized is mentioned hierarchically.

REFERENCES

[1] Bhagyashri Ramesh Jadhav, Manjushri Mahajan, "Review of Dual Sentiment Analysis", International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064.

[2] Kishori K. Pawar, Pukhraj P Shrishrimal, R. R. Deshmukh," Twitter Sentiment Analysis: A Review" International Journal of Scientific & Engineering Research, Volume 6, Issue 4, April-2015 ISSN 2229-5518 IJSER © 2015

[3] S Revathi, N Rajkumar,"A Survey on Sentiment Analysis for Web-Based Data by Using Various Machine Learning Techniques", International Journal of Engineering Trends and Applications (IJETA) – Volume 1 Issue 2, Sep-Oct 2014

[4] A. Abbasi, S. France, Z. Zhang, and H. Chen, "Selecting attributes for sentiment classification using feature relation networks," IEEE Transactions on Knowledge and Data Engineering (TKDE), vol. 23, no. 3, pp. 447-462, 2011.

[5] Rui Xia, Feng Xu, Chengqing Zong, Qianmu Li, Yong Qi, and Tao Li," Dual Sentiment Analysis:Considering Two Sides of One Review", 1041-4347 (c) 2015 IEEE.

[6] Farheen Pathan, Anjali Phaltane, Prof.Sarika joshi," DUAL SENTIMENT ANALYSIS", International Journal of Innovation in Engineering, Research and Technology [IJIERT] ICITDCEME'15 Conference Proceedings ISSN No - 2394-3696

[7] V. Hatzivassiloglou and J.M. Wiebe, "Effects of Adjective Orientation and Gradability on Sentence Subjectivity," Proc. 18th Conf. Computational Linguistics, pp. 299-305, 2000.

[8] B. Pang and L. Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts," Proc. 42nd Ann. Meeting on Assoc. for Computational Linguistics, 2004.

[9] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: Sentiment Classification Using Machine Learning Techniques," Proc. Conf. Empirical Methods in Natural Language Processing, pp. 79-86, 2002.

[10] Q. Su, X. Xu, H. Guo, Z. Guo, X. Wu, X. Zhang, B. Swen, and Z. Su, "Hidden Sentiment Association in Chinese Web Opinion Mining," Proc. 17th Int'l Conf. World Wide Web, pp. 959-968, 2008.

[11] C. Zhang, D. Zeng, J. Li, F.-Y. Wang, and W. Zuo, "Sentiment Analysis of Chinese Documents: From Sentence to Document Level," J. Am. Soc. Information Science and Technology, vol. 60, no. 12, pp. 2474-2487, Dec. 2009.

[12] P.D. Turney, "Thumbs Up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews," Proc. 40th Ann. Meeting on Assoc. for Computational Linguistics, pp. 417- 424, 2002

[13] Dhiraj Gurkhe, Niraj Pal, Rishit Bhatia, "Effective Sentiment Analysis of Social Media Datasets using Naive Bayesian Classification", International Journal of Computer Applications (0975 8887), Volume 99 - No. 13, August 2014.

[14] Riya Suchdev, Pallavi Kotkar,Rahul Ravindran, "Twitter Sentiment Analysis using Machine Learning and Knowledge-based Approach", International Journal of Computer Applications (0975 – 8887),Volume 103 – No.4, October 2014

[15] 14 Eman M.G. Younis,"Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study", International Journal of Computer Applications (0975 – 8887), Volume 112 – No. 5, Febru-ary 2015.

[16] Subhabrata Mukherjee, Akshat Malu, Balamurali A.R., Pushpak Bhattacharyya, "TwiSent: A Multistage System for Analyzing Sentiment in Twitter"

[17] Michael Goul, Olivera Marjanovic, Susan Baxley, Susan Baxley," Managing the Enterprise Business Intelligence App Store: Sentiment Analysis Supported Requirements Engineering", 2012 45th Hawaii International Conference on System Sciences.