ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 6, June 2015

A Survey on Sentiment Analysis in NLP

Divya Bohra[1], Sanjay Deshmukh[2]

M.Tech Student, Computer Engineering Dept., MPSTME SVKM'S NMIMS University, Mumbai, India[1]

Assistant Professor, Computer Engineering Dept., MPSTME SVKM'S NMIMS University, Mumbai, India [2]

Abstract: The ever growing usage of Web articles have resulted in change of web from read only to read write. This has led to the evolution of Sentiment Analysis, which is nothing but analysing the emotions of the given text. Sentiment analysis is a branch of Natural Language Processing (NLP). Sentiment analysis is done on the content such as online news, reviews, blogs, and tweets for social events, political movements, company strategies, marketing campaigns, product preference etc. This paper, will give the overall view of sentiment analysis. It also discusses about Common Knowledge Base and Common Sense Knowledge Base. Hour Glass Model is also discussed in the paper which is effectively being used for sentiment analysis. Finally, various approaches for sentiment analysis have been discussed and compared based on their advantages and disadvantages. It has been found that concept based approach is best suited for sentiment analysis, as it is applicable to multi-word expression.

Keywords: Natural Language Processing; Sentiment analysis; Hour Glass Model; Opinion Mining; Knowledge Base; Concept based approach.

I. INTRODUCTION

Earlier people used to get the information from magazines, News-paper or other Web articles. Now the idea of sharing and getting information has changed to capturing and sharing opinion through World Wide Web. When dealing with the opinion and sentiment of web articles domain dependent corpus is not sufficient as the data is very large and noisy [5] [6]. Thus, sentiment analysis is the emerging field to deal with this type of data. Sometimes when evaluating a product reviews it becomes difficult to categorize whether a person has liked or disliked the product. Such problem is also solved through Sentiment Analysis.

This paper, gives the analysis of the latest approaches used for opinion mining and sentiment analysis [1]. An Hour Glass of Emotions is discussed in the paper which describes the complex range of emotions. Sentiment Analysis have be efficiently done using this Hour Glass Model [2].

In the II Section different approaches for sentiment analysis are described. Common knowledge base and Common Sense Knowledge Base are also discussed [3][4]. Section III will describe the Hour Glass of emotions which is being used for sentiment analysis. Section IV will give advantages and Dis- Advantages of the approaches discussed. Section V will have the conclusion/discussion about the approaches and hour glass model used for sentiment analysis.

II. RELATED WORK

A. Common Sentiment Analysis Tasks

Polarity classification is the basic task in Sentiment Analysis. It is done on a piece of text which will have opinion on a single issue. Like versus Dislike is one of the examples of polarity classification. Evaluation of a product is usually done using polarity classification.

Agreement Detection is another task in sentiment analysis which is an enhanced version of polarity classification. It helps in deciding the degree of positivity or negativity in the given article. But the problem arises when the given text document have multiple sentiments for multiple topics or does not have any strong opinion. Such documents are separated on the basis of given topics and then analysed [1].

B. Evolution of Opinion Mining

Vector Extraction which is currently used for opinion mining and sentiment analysis. It represents the most important text features. The commonly used feature vector in sentiment analysis is Presence. It is a binary valued feature vector, which just indicates whether a term has occurred "1" or have not occurred "0". It is widely used for polarity classification. Another term based feature can also be added to the feature vector, Position. The Position of the term will help to get better sentiment of a given text. N-Gram models, bi-gram and tri- gram can also be used for sentiment analysis.

Parts-of-speech tagging (example: noun, verb, adjectives) is another important task for analyzing the sentiments. Adjectives are good indicator for feature selection. Pre-classifies POS Tagging is also done using adverbs and adjectives [8].Authors have also used some of the important approaches for sentiment analysis [13][14][15]. First is the Key Word Spotting, here words such happy, bored, sad, afraid are used to analyze the text. These words are put into different category so that a better view is given.

Another approach is Lexical Affinity; it not only detects the words but also assigns a probable affinity to particular emotions. Example: the word "Accident" is assigned 75 percent negative effect. As car accident or hurt by accident. It uses probability form linguistics copra [12].

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 6, June 2015

Now, suppose there is a sentence "He is a pretty good liar", in such sentences above two approaches might fail. To overcome this problem, 'Concept-Based Approaches' have been discussed in the next part which is the latest approach for sentiment analysis.

C. Concept Based Approaches

To accomplish sentiment text analysis, these methods uses semantic networks. Such approaches rely on large semantic knowledge bases, rather than just keywords or word occurrence counts. Concept-based approaches can analyse multi-word approaches.

The concept Based approaches depends widely upon the breadth and depth of the knowledge base. Knowledge Base (KB) is the compressive resource that encompasses human knowledge. So, without KB it will be difficult to perform opinion mining tasks on the text. Authors have discussed [4] [9] how use of concepts will result into concept level sentiment analysis. In this approach, just a single word is not considered for analysing the sentiments. In fact, multiple words or small phrases of text are considered. Concept based approaches completely depends upon Common sense knowledge base for sentiment analysis. Common Sense Knowledge Base is discussed in the next part.

D. Common Knowledge Base and Common Sense Knowledge Base

Common knowledge is knowledge that is known by everyone or nearly everyone. Common knowledge need not concern one specific subject, e.g., science or history. Rather, common knowledge can be about a broad range of subjects, including science, literature, history, entertainment etc. On the other hand common sense knowledge is the collection of facts and information. It is something which is gained by a person's day to day experience. So there is a little difference between both types of knowledge. Machines does not contain such kind of knowledge as humans. But with the course of time such types of knowledge base have been built.

Attempts to build both the knowledge bases are countless. WordNet (25,000 synsets), Freebases (a social database), YAGO [11] (a semantic database derived for Wikipedia) are examples of common knowledge bases. Another widely used KB is ProBase. ProBase contains approximately 12 million concepts learned iteratively from 1.68 billion webpages in the Bing repository.

Cyc is the biggest compressive common sense knowledge base. But Cyc requires involvement of experts to work on specific language, it contains just 120,000 concepts, as the knowledge engineering is labor intensive and time-consuming. A more recent and scalable project is Open Mind Common Sense (OMCS), which has been collecting pieces of knowledge from volunteers on the Internet since 1999 by enabling the general public to enter common sense into the system. OMCS exploits these pieces of common-sense knowledge to automatically build the Concept Net, a semantic network of 173,398 nodes [9]. Concept Net, is built from a corpus of common sense knowledge collected and rated by volunteers on the Internet. In Concept Net, the nodes are the concepts, which are nothing but words or short phrases of natural language. These concepts are linked together in the form of graph. The labelled edges shows the relationship between these concepts.

ProBase, which provides more concepts, includes pieces of knowledge that match the general distribution of human knowledge. ConceptNet, in turn, contains implicit knowledge that people rarely mention on the Web, which acts as complementary material to ProBase. Thus, the largest existing taxonomy of common knowledge is blended with a natural- language-based semantic network of common sense knowledge, in order to achieve opinion mining and sentiment analysis. This core taxonomy consists of the IsA relationships extracted by using syntactic patterns. For example, the segment "artists such as Pablo Picasso" can be considered a piece of evidence for the claim that "pablo picasso" is an instance of the concept "artist" [5].

This blending of common knowledge base and common sense knowledge base shows how knowledge is organized into human mind and how natural language tasks like sentiment analysis and opinion mining can be efficiently performed [10] [17].

III.HOUR GLASS MODEL

The study of emotions is one of the most confused (and still open) chapters in the history of psychology. Till now the emotions were described using a list of tables or two dimension models. Such approaches are insufficient to describe wide range of emotions and multi word concepts. To avoid this problem Hour Glass of Emotions was introduced.

The Hourglass of Emotions (Figure 1) is a biologically-inspired and psychologically-motivated model goes beyond the categorical and dimensional approaches. It can potentially describe any human emotion in terms of four independent but concomitant dimensions [2].The four dimensions are as follows: Pleasantness: the user is amused by interaction modalities, Attention: the user is interested in interaction contents, Sensitivity: the user is comfortable with interaction dynamics, Aptitude: the user is confident in interaction benefits.Each of these dimension contains Six sentic levels. These sentic levels gives the strength of the emotions. As seen in the figure, this model contains two types of dimensions: Vertical dimension and radial dimension. The vertical dimension represent the intensity of the different affective dimensions (Pleasantness, Attention, Sensitivity, Aptitude) while the radial dimension represents the strength of different emotional configurations. Since affective states are represented according to their strength (from strongly positive to null to strongly negative), the model assumes an hourglass shape.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 6, June 2015

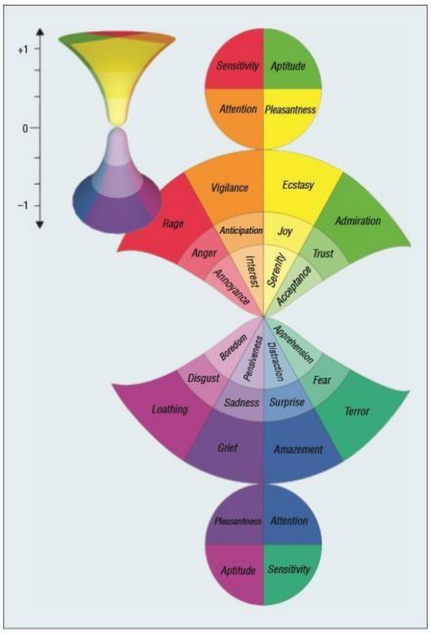

Fig. 1 Hourglass of Emotions [2]

Below is the tabular form of 24 emotions (Six emotions of each dimension) which are present in hour glass of model. Like, in Pleasantness, there are six emotions. Both the emotions ecstasy and grief are strong estimations. Next are joy and sadness with medium emotional strength. Finally, the emotions serenity and pensiveness have minimum strength. Thus, stronger emotions induce higher emotional sensitivity.

TABLE I.VARIOUS EMOTIONSOF HOUR GLASSMODEL

	4 Dimensions			
	Pleasantness	**Attention**	**Sensitivity**	**Aptitude**
24 EMOTIONS	Ecstasy	vigilance	Rage	admiration
	Joy	anticipation	Anger	trust
	serenity	interest	Annoyance	acceptance
	pensiveness	distraction	Apprehension	boredom
	sadness	surprise	Fear	disgust
	Grief	amazement	Terror	loathing

This model is now being used in variety of application for sentiment analysis. Hour Glass of Emotions is used in Open domain sentiment analysis and sentiment analysis in social media marketing [7] [16].

IV. ADVANTAGES AND DISADVANTAGES OF SENTIMENT ANALYSIS APPROACHES DISCUSSED

This section will give the advantages and disadvantages of various approaches of opinion mining and sentiment analysis discussed above. Table II shows the advantages and dis-advantages of various approaches in opinion mining. The first is the keyword spotting; its advantage over other approaches is the ease of accessibility and economy. Which makes it very popular? But its two weaknesses are that it can't reliably recognize affect negated words, and it relies on surface features. For example: the sentence "today was a happy day" as being affectively positive, it is likely to assign the same classification to a sentence like "today wasn't a happy day at all." Also some sentences like "My son just filed for a case and he wants to take custody of my property" evoke strong emotions, but use no affect keywords, and therefore is ineffective.

The second approach, Lexical Affinity has the advantage of outperforming pure keyword spotting but sentences like (I avoided an accident) and sentences with other meanings (I met my girl- friend by accident) trick lexical affinity, because they operate solely on the word level.

TABLEII.ADVANTAGESANDDIS-ADVANTAGESOFSENTIMENTANALYSIS APPROACHES

Approaches forSentiment Analysis	**Advantages**	**Dis-advantages**
Keyword spotting	Accessibility And Economy	It can't reliably Recognize affect negated words, and it relies on surface features.
Lexical affinity	Outperforms Pure keyword spotting	Negated sentences Because it operate solely on the word level
Concept-based approaches	Analysis multi-word expressions	Relyblindlyon Knowledge base's

The final approach Concept-Based Approach can analyse multi-word expressions that don't explicitly convey emotion, but are related to concepts that do. But the only problem with this approach is that it rely blindly on Knowledge bases. So, if the KB is not constructed properly or contains high levels of ambiguity concept-based approach will not give effective results.

V. CONCLUSION

The Web has changed from "read- only" to "read-write". Blending scientific theories of emotion with the practical engineering provides goal of analyzing sentiments in natural- language text. This will lead to more bio-inspired approaches like hour glass model, which is discussed in the paper. Hour glass model is being widely used for designing of intelligent opinion-mining systems which are capable of handling semantic knowledge. It is making easier to detect, perceive, and feel the complete range of emotions more accurately. The review on the various sentiment analysis approaches have also been done in the paper. It has been found that concept based approach is best for sentiment analysis. Also, the blend of both common and common sense knowledge bases will be a good choice as it will represent the best trade-off between common and common-sense knowledge.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 6, June 2015

This is particularly needed when aiming to infer both the conceptual and affective information associated with text. As concept based approaches rely completely on KB for their effective results, it is best suited for such blended knowledge bases.

ACKNOWLEDGMENT

I am grateful to my H.O.D **Dr. Dhirendra Mishra** and faculty guide Shubha Puthran for sharing their pearls of wisdom with me during the course of this research.

REFERENCES

[1] Erik Cambria, Bjӧrn Schuller and Yunqing Xia, "New Avenues in Opinion Mining and Sentiment Analysis", 1541-1672/13 IEEE 15 Published by the IEEE Computer Society, 2013.
[2] E. Cambria, A. Livingstone, and A. Hussain, "The Hourglass of Emotions," Cognitive Behavioral Systems, A. Esposito et al., eds., LNCS 7403, Springer, pp. 144–157, 2012.
[3] Havasi, Catherine, Erik Cambria, Björn Schuller, BingLiu, and Haixun Wang."Knowledge-based approaches to concept-level sentiment analysis."IEEE Intelligent Systems 28,no.2(2013): 0012-14.
[4] Tsai, Angela Charng-Rurng, Chi-En Wu, Richard Tzong-HanTsai, and Jane Yung-jen Hsu. "Building a concept-level sentiment dictionary based on common sense knowledge."IEEEIntelligentSystems28, no.2 (2013):22-30.
[5] ErikCambria, Yangqiu Song and Haixun Wang, "Semantic Multidimensional Scaling for Open-Domain Sentiment Analysis", IEEE Published by the IEEE Computer Society2014.
[6] Jalaj S. Modha, Gayatri S. Pandi, Sandip, J. Modha, "Automatic Sentiment Analysis for Unstructured Data", IJARCSSE, Volume 3, Issue12,December 2013.
[7] Erik Cambria, Marco Grassi, Amir Hussain and Catherine Havasi," Sentic Computing for social media marketing", Springer Science+ Business Media, LLC 2011.
[8] Shichang Sun, Hongbo Liuand Ajith Abraham, "Twitter Part-Of-Speech Tagging Using Pre-classification Hidden Markov Model", IEEE International Conference on Systems, Man, and Cybernetics October 14-17,2012,COEX,Seoul,Korea 2012.
[9] Catherine Havasi, James Pustejovsky, Robert Speer and Henry Lieberman "Digital Intuition: Applying Common Sense Using Dimensionality Reduction", IEEE intelligent systems , July/August 2009.
[10] Cambria, Erik, Yangqiu Song, Haixun Wang, and Amir Hussain. "Isanette: A common and common sense knowledge base for opinion mining."In Data Mining Workshops (ICDMW), 2011IEEE 11th International Conference on, pp. 315-322.IEEE,2011.
[11] F.Suchanek, G.Kasneci,and G.Weikum, "YAGO:A Core of Semantic Knowledge, " Proc. Int'l World Wide Web Conf., 2007, pp. 697–706.
[12] Partha Sarkar and Bipul Syam Purkayastha, "A Study on Deep Linguistic Processing with Special Reference to Semantic and Syntactic Levels," International Journal of Computer Applications(0975– 8887) Volume88 –No.13,February2014.
[13] Im, TanLi, Phang Wai San, Chin Kim On, RaynerAlfred,and Philip Anthony. "Analyzing market sentiment in financial news using lexical approach. "In Open Systems (ICOS), 2013 IEEE Conference on, pp. 145-149.IEEE,2013.
[14] Maharani, Warih. "Microblogging sentiment analysis with lexical based and machine learning approaches."InInformation and Communication Technology (ICoICT), 2013 International Conference of,pp.439-443. IEEE, 2013.
[15] Solakidis, Georgios S., Konstantinos N. Vavliakis, and Pericles A. Mitkas."Multilingual Sentiment AnalysisUsing Emoticonsand Keywords."InProceedingsof the 2014IEEE/WIC/ACMInternational Joint ConferencesonWebIntelligence (WI) andIntelligentAgent Technologies (IAT)-Volume02, pp.102109.IEEComputerSociety, 2014.
[16] Cambria,Erik,CatherineHavasi,andAmirHussain."SenticNet2:A SemanticandAffective Resource for Opinion Mining andSentiment Analysis."InFLAIRS conference, pp. 202-207. 2012.
[17] Poria,Soujanya,AlexanderGelbukh, Erik Cambria, PeipeiYang,Amir Hussain,andTariqDurrani."Merging SenticNetandWordNet-Affect emotionlistsforsentimentanalysis."InSignalProcessing(ICSP),2012 IEEE11thInternationalConferenceon,vol.2,pp.1251-1255.IEEE,2012.