A Survey of Approaches for Sentiment Analysis and Applications of OMSA Beyond Product Evaluation

Mrs. Mukta Raut[1], Mrs. Mayura Kulkarni [2], Dr. Sunita Barve[3]

Student, Department of Computer Engineering, MIT Academy of Engineering, Pune, India.[1]
Sr. Assistant Professor Department of Computer Engineering, MIT Academy of Engineering, Pune, India.[2]
Assistant Professor Department of Computer Engineering, MIT Academy of Engineering, Pune, India. [3]

Abstract-Sentiment analysis has been in the forefront in research in machine learning for a couple of decades. The need for sentiment classification arises from the upsurge of online trading, where customer satisfaction is crucial. Yet, as there is no face-to-face interaction between producer and consumer, feedback in form of text review, star ratings, comments, discussions on the blogs, and so on play an important part in product or service evaluation. This is where opinion mining and sentiment analysis (OMSA) comes into picture. Sentiment analysis of online reviews and other user generated content is an important research problem for its wide range of applications. In this paper we present a survey of different approaches for sentiment analysis and combining them to form a system with best features from several approaches. between concept-level and aspect level sentiment analysis. We further discuss different techniques used to perform OMSA and the applications of OMSA in the stand alone systems and as embedded systems in human-agent interaction and embodied conversational agents.

Keywords: machine learning, sentiment analysis, human-agent interactions, embedded conversational agents.

I. Introduction

The number of online users is escalating with the increase in availability and speed of internet. The online producer-consumer industry is flourishing at a constant speed. This upgrowth results in thegeneration of tons of data which is a type of big data, which compromises of various aspects of online trading, online social interactions, discussions and so on. Although this data is of great value, very little of the available data is utilized in real time. This is due to multiple reasons; most of this data is gathered in heterogeneous forms such as text reviews, feedbacks, star ratings, numerical ratings, different grading systems, and so on. This is called as multimodal data. Thus, not being structured, the multimodaldata is not directly machine usable. On the other hand, the tremendous volume and the speed at which it is generated makes it impossible to perform manual analysis of the data for knowledge discovery. Thus the use of machine learning techniques become inevitable. These include using statistics and natural language processing to extract, identify, or otherwise characterize the sentiment content of a text unit. This process is collectively called as opinion mining or sentiment analysis (OMSA). By developing automated techniques developed for sentiment analysis we can save the overhead of manual analysis of this big data. The main advantage of sentiment analysis is speeding up the decision making process of the consumers without compromising on the evaluation parameters.

II. Terminology

In sentiment analysis, there are several terms such as sentiment, view, belief, conviction, affect, opinion, etc. according to English linguistic rules, differences in these terms are subtle. However, in OMSA domain, these differences are significant, and each term has a distinct definition. For example, ‘opinion’ and ‘sentiment’ do not count as the same terms; Sentiment indicates involvement of emotions while opinion refers to personal interpretation of information and has no emotional value. Furthermore, emotions differ from sentiments in the sense that emotions are briefer in duration than sentiments. Unlike sentiments, emotions are not always targeted towards an object. The

terms, are in fact used as bynames in the machine learning context.

1.Sentiment- It is a subjective response to a person, thing, or situation, which tends to change according to the situation.
2.View- It is a close look at or over someone or something in order to judge condition.
3.Belief- belief is an idea that is believed to be true or valid without positive knowledge.
4.Conviction- a state of mind in which one is free from doubt.
5.Affect- affect is the conscious subjective aspect of an emotion considered apart from bodily changes. It can also be described as to act upon (a person or a person's feelings) so as to cause a response.
6.Opinion- Opinion can be seen as belief which is stronger than impression and less strong than positive knowledge. It can also be defined as a position arrived at after consideration.
7.Judgment- It is the process of forming an opinion or evaluation by discerning and comparing.
8.Appraisal- It is the act of placing a value on the nature, character, or quality of something.
9.Affect/emotion- personal reaction referring to an emotional state.

The research area of opinion mining or sentiment analysis is sometimes also referred as subjectivity analysis. It studies the phenomena of opinion, sentiment, evaluation, appraisal, attitude, and emotion.

III. SENTIMENT ANALYSIS

Sentiment classification is the main task in OMSA. This classification is performed at three levels, namely, document-level, sentence-level and aspect-level. We cannot separate completely the sentence-level and document-level as each sentence can also be viewed as a short document. The document-level classification assumes that the whole document reflects a single sentiment. This may or may not be true as we often come across discussions or comments which evaluates more than one aspect of an entity. In sentence-level, we assume that each sentence contain sentiment about only one topic. If we look at the comments such as, "How could anyone sit through this movie, even with comfy seats and fab food?!"; we know that this isn't entirely true. Alexandre and Trilla and FrancescAlías, in their paper Sentence-Based Sentiment Analysis for Expressive Text-to-Speech, present a sentence based sentiment classification method which is utilized for converting classic text-to-speech systems into expressive speech.

The aspect level sentiment classification, however takes place at finer layers. The overall sentiment of the document is considered referring to an entity. The motivation behind aspect level sentiment analysis is to go beyond the mere sentence-level sentiment classification for performing more fine-grained analysis and transition from unstructured data to machine usable data.

a. Basic tasks in sentiment analysis

The two basic tasks of sentient analysis are polarity detection and emotion detection. The polarity detection focuses on the orientation of polarity of an opiniated text being either positive or negative. Some advanced polarity classification methods present a 3-class polarity classification as positive, negative and neutral. The emotion recognition is extracting the set of emotion labels from given document.

The figure (1) shows the general process of sentiment analysis. In each of the above steps, there can be one or more sub-steps, depending upon the context. For example, prior to data collection, we may need to include a step 'multimodal fusion', which refers to collecting data of varied type such as text, speech, video and transforming it into a single text type. Sentiment detection can be sub-divided into emotion detection and polarity detection.

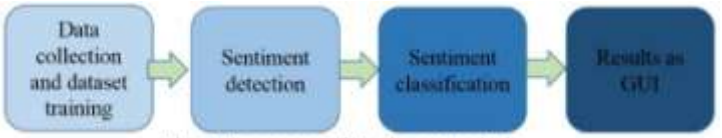

Fig. 1 The process of Sentiment Analysis.

Thus, the different tasks in sentiment classification are highly interdependent.

b. Approaches of sentiment analysis

The approaches of sentiment analysis can be categorized into three types: knowledge based approaches, statistical approaches and hybrid approaches.

1. Knowledge based approaches: These approaches are fairly accessible and economic. In knowledge based sentiment analysis, the opiniated text is classified into certain unambiguous affects such as 'happy', 'sad', 'afraid', etc. This is very straightforward method in which knowledge-bases are used for comparing with opiniated texts, to extract sentiment information. The major drawback of this approach is that these systems are only as accurate as your knowledge-base. Also knowledge based approach can only classify sentences which are straight forward. Such comments as," the location is good, I like this resort" can easily be classified into their proper class(positive). But when a customer quotes," there is no way I can appreciate this resort in any possible way!", the knowledge based approach faultier to deliver. It is possible that with Knowledge-based approach, this sentence will get (wrongly) classified as positive, due to presence of positive-sentiment words 'appreciate' and 'possible'. Thus it is very difficult to deal with linguistic features such as sarcasm, pun, hyperbole and such other figures of speech within the user comments, using Knowledge-based approaches.

2. Statistical Approaches: In these approaches, we use a previously trained dataset to train the classifier (machine learning). We then feed the opiniated text to these previously trained classifiers for statistical analysis. These classifiers then calculate the positivity and negativity 'scores' of the given text and classify it likewise. Naïve Bayes, Support Vector Mechanism, KNN are some commonly used classifiers. The main advantage of this approach is we can train the classifier according to our requirement so as to suit a particular domain. Also as classifier can be trained according to user requirement, we can use a dataset in language of our choice. Thus we can customize the sentiment classification task to a certain extent. However, training the classifier is can be an affliction if not done carefully and accurately, because entire system may malfunction due to shortcomings in classify training. Moreover, the accuracy of such classification systems is directly proportional to number of entries in training dataset. Most of the classifiers require large datasets to gain desired accuracy. Also many classifiers do no deal well with data redundancy issue.

3. Hybrid approaches: These approaches combine the best features of Knowledge-based approach and Statistical approach, to meet the system performance requirements. We dcan give an example of such system as stated below.

Consider a system with objective to classify product reviews into 3-class classification (positive, negative, neutral) and rate the system accordingly. Now we can design the system in three parts:

1. Training of classifier
2. Statistical analysis
3. Product evaluation

A. Training the classifier: for this part, we first collect the reviews related to desired product or service into a dataset from different online sources. To train the dataset, we label each review using 3-class classification. To perform this classification, we maintain two knowledge-bases in the system, ne with positive keywords, one containing negative keywords as well as all the negative words, and one containing all the stopword (stop words are the words in the sentence which has no significance in sentiment such as of, that, the, and, and so on.)

***B. Statistical analysis*:** The statistical analyzer will use the training module and knowledge-bases jointly to calculate the score of the given review. Here we can use any statistic based classifier such as logistic regression, Naïve Bayes, support vector, etc. to find the class of the review at hand, we can compare the score with the scale 0 to 5 or 0 to 10 and set the three limits. For example, if we consider comparing the score on scale of 0 to 5, we can set classification limits as follows:

0-2.7 negative class
2.8-3.5 neutral class
3.6-5.0 positive class

C. Product Evaluation: This module will calculate the total numbers of reviews in each class and present it in some form of visual graphics such as pie chart, bar graph, etc.

Thus, this system combines knowledge-based approach and statistical approach and presents a hybridapproach to sentiment classification. Figure 2 gives a general architecture of such a system.

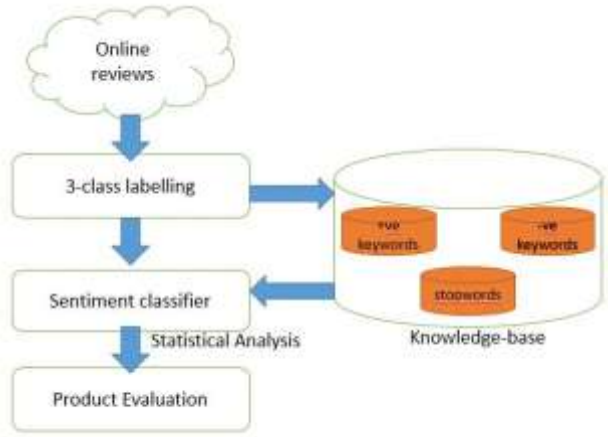

Fig. 2 Hybrid Approach to Sentiment Analysis

IV. OBSERVATIONS

The hybrid approach combines all the desirable features of both knowledge-based approach and Statistical approach. In above system, we use knowledge-base as well as statistical analysis to classify the reviews. We get the following observations from the above architecture:

1. By using knowledge-base, we ensure that the training of the sentiment classifier is according to our requirements and accurate. Thus, knowledge-based helps to train the dataset for classifier.
2. We can customize this knowledge-base according to the domain of system (e.g. Whether we want to classify reviews about a hotel, a hospital or about any social even, will decide which words will we included inn keyword databases and stopword database.). thus, our sentiment classification system can be domain adaptive and language independent.
3. Instead of including "all possible" reviews, can focus on adding "all types" of reviews in the training dataset. By this we can exclude the assumption that a statistical based system requires very large datasets. Also, while classifying the new reviews, we can save those reviews into our knowledge-base again (machine learning). By this, we can keep our dataset updated and the learning process of the system will never stop and be outdated.
4. By controlling training dataset, we can also perform redundancy checks for better performance by using discriminative and generative methods as per our requirement.

So, we observe that by combining both the approaches, we can minimize the shortcomings of both approaches. We also note here that, instead of fully automated or fully manual sentiment analysis systems, semi-automated systems as one proposed above are more yielding.

V. APPLICATIONS OF SENTIMENT ANALYSIS BEYOND PRODUCT EVALUATION

When thinking 'sentiment Analysis', one immediately thinks of movie reviews or election polls. But as of today's age, the applications of OMSA go far beyond Socializing and shopping on the world wide web. In this section, we intend to focus on all such research areas in sentiment analysis which carry great potential for various organizations.

To enrich the experience of the content: Sentiment Analysis can be performed on the media transcripts and captions. In case of audience with impaired hearing, this analysis can mark up the emotions in the content and give more sentiment information than merely offering them plain Tran scripted text. This could make the media content more interesting to such people. It can also be used as teaching subjects like literature, politics to students with impaired hearing where emotions are as important as text.

Sentiment Analysis will also be equally helpful for developing educational curriculums for students with learning disabilities such as autism andAsperger's syndrome.

Emotion recognition: Sentiment Analysis can make possible emotion recognition. This proves useful in interactive games, e-learning programs. Better e-learning programs can be developed by studying the audience sentiment towards each section. It can be used for character design in interactive games. Emotions can be very important in the virtual interactions.

Expressive Speech Synthesis: it can prove helpful especially people with visual disability. In converting the content from text to speech, generation of expressive speech will not only enrich the overall experience of visually impaired audience, but will also aid in better understanding of the content for them.

Human-Agent Interaction: Such interacting systems can be used in tourism industry for automated robotic guidance. Such guidance systems can perform in extreme geographical conditions. For example, assisting a lost mountaineer to the

nearest civilization. It can be used in luxury hotels and resorts, corporate offices where people expect ambience along with prompt service.

Affective Monitoring: It has application in the designing wearable devices for customer satisfaction, developing systems that respond according to individual's mental state, monitoring emotional state of a mob through CCTV recording i.e. Sentiment Analysis of Audio-visual input.

VI. CONCLUSION

in this paper we study the process of sentiment analysis and its different approaches. While studying the different approaches, we can conclude that the task of sentiment analysis be performed better byusing combined approaches, instead of using a single approach. Also we note that the present OMSA systems are very far from age of fully automated systems. For this reason, one good practice is to train the systems in a supervised learning environment, and carry out the analysis part in automated environment. Also in this paper we presented an overview of the applications of OMSA which go far beyond customer satisfaction.

ACKNOWLEDGEMENT

It is our immense pleasure to work on this project. It is only the blessing of our divine guides Prof. Mayura Kulkarni and Co-guide Prof. Sunita Barve who has prompted and mentally equipped us to undergo the study of this project. We are also grateful to our entire staff of Computer Engineering Department, Family and Friends for their kind co-operation which helped us in successful implementation of project.

REFERENCES

[1] Affective Computing and Sentiment Analysis, Editor: Erik cambria, Nanyang Technological University, Singapore, cambria@ ntu.edu.sg, Published by the IEEE Computer Society

[2] Developing Corpora for Sentiment Analysis: The Case of Irony and Senti- TUT Cristina Bosco, University of Torino, Viviana Patti, University of Torino, Andrea Bolioli, CELI srl, Published by the IEEE Computer Society, 2013.

[3] Selecting Attributes for Sentiment Classification Using Feature Relation Networks Ahmed Abbasi, Member, IEEE, Stephen France, Member, IEEE,Zhu Zhang, and Hsinchun Chen, Fellow, IEEE. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 3, MARCH 2011.

[4] Sentiment Embeddings with Applications to Sentiment Analysis Duyu Tang, Furu Wei, Bing Qin, Nan Yang, Ting Liu, and Ming Zhou. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 28, NO. 2, FEBRUARY 2016.

[5] Dual Sentiment Analysis: Considering Two Sides of One Review Rui Xia, Feng Xu, Chengqing Zong, QianmuLi, Yong Qi, and Tao Li. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 8, AUGUST 2015.

[6] Scope of Negation Detection in Sentiment Analysis. MaralDadwar, Claudia Huff. Human Media Interaction Group University of Twente Enschede, Netherlands.

[7] Sentiment Analysis using Product Review Data. Xing Fang and Justin Zhan. Fang and Zhan Journal of Big Data (2015) 2:5

[8] Pang B, Lee L (2008) Opinion mining and sentiment analysis. Found Trends InfRetr 2(1-2):1135

[9] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Found. Trends Inf. Retrieval, vol. 2, no. 1-2, pp. 1–135, 2008.

[10] B. Liu, Sentiment Analysis and Opinion Mining (series Synthesis Lectures on Human Language Technologies). vol. 16. San Mateo, CA, USA: Morgan, 2012.