International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 3 Issue 11, November-2014

A Survey on Multimodal Sentiment Analysis

Sujit J. Fulse
MITCOE
Pune, India

Prof Rekha Sugandhi
MITCOE
Pune, India

Dr. Anjali Mahajan
P. I. E. T., RTM Nagpur University
Nagpur, India

***Abstract*— Presently large amount of data is available on social networking sites, product review sites, blogs, forums etc. This data holds expressed opinions and sentiments. The volume, variety, velocity are properties of data, whether it comes from the Internet or an enterprise resource planning system, sentiment analysis system should get the data and analyze it. Due to the large volume of opinion rich web resources such as discussion forum, review sites, blogs and news corpora available in digital form, much of the current research is focusing on the area of sentiment analysis and opinion mining. Expression of any sentiment is a mixture of text, prosody, facial expression, body posture etc. Thus only text input cannot fully represent a sentiment. A multimodal system uses a combination of input modes e.g. text and audio or text and video or all of these three. This paper analyses the techniques used for multimodal sentiment data and also demonstrate that how individual model works. Extracting the sentiments from different input modes is achieved by different classifying techniques. In this paper, firstly we will discuss the different input modes. In the end we analyse challenges of our proposed system. Also we have discussion on integration of different modes and its affect on emotion reorganization system**

Keywords- Opinion Mining, Sentiment Analysis, Multimodal Sentiment Analysis, Big Data, Data Mining.

I. INTRODUCTION

Sentiment analysis is a method, which classifies the given data due to having positive or negative opinion and being subjective or objective in general. In this case, the concept of emotions and opinions is comes in focus, the opinionated data is deeply analysed to determine the strength of opinions, which is closely related to the intensity of emotions such as happy, fear, sad, anger, surprise etc. It is claimed that people's sentiments can be identified by examining their language expressions, and can be classified, according to the level of their strength. It is mostly used in advertisement placement, product benchmarking and market intelligence, and detection of company reputation or brand popularity and identification of fake or misinforming comments [1].

Multimodal sentiment analysis is computational study of mood, sentiments, views, affective state etc. from the text and audio, video data.

Opinion mining is used to analyze the attitude of a speaker or a writer with respect to some topic Opinion mining is a type of NLP for tracking the mood of the public about a particular product. Collect and examine opinions about the product made in discussion board, review sites, blogs, tweets/comments etc.

Section 2 presents the different data types used for opinion mining i.e. sentiment analysis on different input modes such as text, audio, video. Section 3 introduces Data sources, as the data is coming from different sources. Section 4 presents semantic orientation approaches for sentiment classification. For each type of data. Section 5 presents Multi-modal sentiment analysis, Section 6 presents Application of Multi-modal Sentiment analysis over Big data.

II. SENTIMENT ANALYSIS BASED ON DIFFERENT INPUT MODES

A. Text

A basic task in sentiment analysis is classifying the given text on the basis of polarity. Polarity can be positive, negative or neutral at different level such as phrase, feature/aspect, sentence, document level. Apart from finding polarity, sentiment analysis looks forward for emotional state such as happy, sad, angry etc.(Boiy etal. 2007) [1]. Identify if the text contains a positive, negative or neutral opinion. Find the opinion of a person or organization (opinion holder) on a particular object or a feature of the object. Identify and extract subjective information in source materials. Classifies an evaluative text as being positive or negative, no details are discovered about what people liked or disliked. Sentiment Classification broadly refers to binary, multi-class distribution, regression and ranking. Sentiment Classification mainly consists of two important tasks, first one is assignment sentiment polarity and sentiment intensity i.e. scores assignment. Sentiment polarity assignment deals with analyzing, whether a text has a positive, negative, or neutral semantic orientation. Sentiment intensity assignment deals with analyzing, whether the positive or negative sentiments are mild or strong. There are several tasks in order to achieve the goals of Sentiment Analysis.

Emotion	Observed Facial Cues
Surprise	Brows raised (curved and high) Skin below brow stretched Horizontal wrinkles across forehead Eyelids opened and more of the white of the eye is visible Jaw drops open without tension or stretching of the mouth
Fear	Brows raised and drawn together Forehead wrinkles drawn to the center Upper eyelid is raised and lower eyelid is drawn up Mouth is open Lips are slightly tense or stretched and drawn back
Disgust	Upper lip is raised Lower lip is raised and pushed up to upper lip or it is lowered Nose is wrinkled Cheeks are raised Lines below the lower lid, lid is pushed up but not tense Brows are lowered
Anger	Brows lowered and drawn together Vertical lines appear between brows Lower lid is tensed and may or may not be raised Upper lid is tense and may or may not be lowered due to brows' action Eyes have a hard stare and may have a bulging appearance Lips are either pressed firmly together with corners straight or down or open, tensed in a squarish shape Nostrils may be dilated (could occur in sadness too) unambiguous only if registered in all three facial areas
Happiness	Corners of lips are drawn back and up Mouth may or may not be parted with teeth exposed or not A wrinkle runs down from the nose to the outer edge beyond lip corners Cheeks are raised Lower eyelid shows wrinkles below it, and may be raised but not tense Crow's-feet wrinkles go outward from the outer corners of the eyes
Sadness	Inner corners of eyebrows are drawn up Skin below the eyebrow is triangulated, with inner corner up Upper lid inner corner is raised Corners of the lips are drawn or lip is trembling

Table 1: Facial Cues and Emotion [3]

Linguistic content also plays a vital role in identifying emotions. Linguistic cues of the sentiment present in the text are acknowledged by lexical dictionaries. Renowned one being MPQA database (Wiebe 2005). The research is highly language-dependent and generalizing the work is a difficult task [2].

B. Image

Facial expression is nothing but position or movement of muscles beneath the skin of face such movements of skin convey the mood or emotion state of a person. Research study focuses on face to detect basic emotions, which is inspired by "emotions as expression". The expression is given for a short duration when the emotion is experienced, so detecting an emotion is simply a matter of detecting its prototypical facial movement. Emotions consist of multiple components that may include appraisals, intentions, action tendencies, other cognitions, central and peripheral changes in physiology, and feelings. Emotions are not directly observable, but are inferred from expressive behaviour, physiological indicators, self-report, and context. Here focus is on expressive behaviour because of its coherence with other indicators and the depth of research on the facial expression of emotion in behavioural and computer science.

In [3], Scott Brave and Clifford Nass stated that, Facial expression provides a fundamental means by which humans detect emotion. Figure 1 describes characteristic facial features of six basic emotions

C. *Video*

Facial expression, Body movement, posture, gestures etc. are analysed to extract the sentiment features. Facial expressions are among the most universal forms of body language which are studied in emotion analysis [14].

D. Audio

Emotion analysis of speech signals aims to identify the emotional or physical states of a person by analyzing his or her voice. Prosody features. These include intensity, loudness, and pitch that describe the speech signal in terms of amplitude and frequency. Energy features describe the human loudness perception. Voice probabilities are probabilities that represent an estimate of the percentage of voiced and unvoiced energy in the speech.

Spectral features are based on the characteristics of the human ear, which uses a nonlinear frequency unit to simulate the human auditory system. These features describe the speech formants, which model spoken content and represent speaker characteristics [3].

Table 2 describes characteristic prosodic features of six basic emotions.

	Fear	Anger	Sadness	Happiness	Disgust
Speech rate	Much faster	Slightly faster	Slightly slower	Faster or slower	Very much slower
Pitch average	Very much higher	Very much higher	Slightly lower	Much higher	Very much lower
Pitch range	Much wider	Much wider	Slightly narrower	Much wider	Slightly wider
Intensity	Normal	Higher	Lower	Higher	Lower
Voice quality	Irregular	Breathy chest tone	Resonant	Breathy blaring	Grumbled chest tone
Pitch changes	Voicing Normal	Abrupt on stressed syllables	Downward inflections	Smooth upward inflections	Wide downward terminal inflections
Articulation	Precise	Tense	Slurring	Normal	Normal

Table 2: Speech and Emotion [3]

III. DATA SOURCES

User's opinion is a major criterion for the improvement of the quality of services rendered and enhancement of the deliverables. Blogs, product review sites, micro blogs provide a good understanding of the reception level of the products and services.

A. Blogs

Blogs contains reviews of many product and issue. These blogs contains large amount of opinionated text. Blogger opens a topic in discussion and records daily events for opinions, feelings, emotions etc, so it is very important to apply sentiment features over blog data [15].

B. Review sites

Review site contains views of people about product or topic. These views are in form of comments which are in unstructured format. S, people read those comments and makes purchasing decision. E-commerce websites[1, 2, 3, 4] hosts millions of product reviews given by customer such data is used in sentiment analysis. Other than these the available are professional review sites[5, 6] and consumer opinion sites on different topics.

C. Dataset

Most of the work in the field uses movie reviews data for classification. Movie review data is available as dataset[7]. Other dataset which is available online is multi-domain sentiment (MDS) dataset[8].

The MDS dataset includes four different types of product reviews which are summarized from Amazon.com including Books, cloths, Electronics and Decoration appliances, with 1000 positive and 1000 negative reviews for each domain.

1 www.amazon.com
2 www.yelp.com
3 www.CNETdownload.com
4 www.reviewcentre.com
5 www.dpreview.com
6 www.zdnet.com
7 www.cs.cornell.edu/People/pabo/movie-review-data
8 www.cs.jhu.edu/mdredze/datasets/sentiment

D. Micro-blogging

These are sites which allow user post small messages as a status e.g. tweets of tweeter. Sometimes these twits express opinions. So, Twitter messages are studied to classify sentiments.

IV. METHODS TO ANALYSE (SENTIMENT CLASSIFIER)

In [4], reviews or comments are classified into positive and negative. Traditionally the document

classification was performed on the topic basis but later research started working on opinion basis. Following machine learning methods Naive Bayes, Maximum Entropy Classification (MEC), and Support Vector Machine (SVM) are used for sentiment analysis. The conventional method of document classification based on topic is tried out for sentiment analysis. The major two classes are considered i.e. positive and negative and classify the reviews according to that.

In [5], Naïve Bayes is best suitable for textual classification, clustering for consumer services and Support Vector Machine for biological reading and interpretation. The four methods discussed in the paper are actually applicable in different areas like clustering is applied in movie reviews and Support Vector Machine (SVM) techniques is applied in biological reviews & analysis. Though field of opinion mining is latest technology, but still it provides diverse methods available to provide a way to implement these methods.

In [6], this paper has presented experiments on prosody-based Automatic Personality Perception, i.e., on automatic prediction of personality traits attributed by human listeners to unknown speakers. The APP results show an accuracy ranging between 60 and 72 percent (depending on the trait) in predicting whether a speaker is perceived to be high or low with respect to a given trait. The most probable reason is that the corpus includes two categories of speakers (professional and nonprofessional ones) that differ in terms of characteristics typically related to the trait e.g. efficiency, reliability etc..

Litman et al. [7] investigated the role of the context information (e.g. subject, gender, features representing local and global aspects of prior dialogue) on audio affective recognition. Still the natural interaction of human beings has subtle emotions and six basic emotions (described above) seldom occur. Thus, mostly researchers include acoustic features with linguistic features (language-dependent) to improve recognition process. Typical examples of linguistic-paralinguistic-fusion methods are of Litman et al.

Lee and Narayanan [8] used vocal features, spoken contents (words) for sentiment analysis over speech signals. This paper focuses on recognizing emotions from spoken language. The importance of emotion recognition from human speech has increased significantly with the need to improve both the naturalness and efficiency of spoken language human-machine interfaces

Another challenge is how to reliably extract these linguistic and paralinguistic feature from the audio channel. Several prosodic features are taken into account from the emotion detection within a speech signal but the research activities are not up to mark. The automatic extraction of high-level underlying semantic linguistic information such as dialogue act, corrections, repetitions, and syntactic information has to be further studied [9].

In [10], this paper shows that valence ratings and skin conductance were able to characterize the emotional nature of the content. A repeated measure ANOVA was run on valence and skin conductance (normalized separately for each participant using z-scores). Two participants were removed from the analysis due to a lack of variation in their raw skin conductance. The mean of valence and arousal successfully differentiated the emotional nature of different stimuli, especially negative as opposed to nonnegative, and detected cultural biases. The analysis of simultaneous agreement in both valence and arousal led to finding moments of universal experiences. The moments where a culture had high cohesion in their valence and arousal were moments were there were smaller differences between the emotional experiences between cultures.

Accuracy of Different Method-

Based on the survey, by considering N-gram features we can find the accuracy of different classifier such as Naïve Bayes (NB), Multi-Layer Perceptron (MLP), Support Vector Machine (SVM) in different data set shown in table.
According to the survey, accuracy of SVM is better than other three methods when N-gram feature was used.

	Movie Reviews			Product Reviews		
N-Gram Features	NB	MLP	SVM	NB	MLP	SVM
	75.5	81.05	**81.15**	62.50	79.27	**79.40**

Table 1 Accuracy of Different Method

Naive Bayes- It is probabilistic and supervised classification algorithm. Its accuracy is 79.93 % for linguistic. The advantage of Naive Bayes is, it's simple to implement. It performs efficient computation. This is simple counting based method. It does not consider dependency among lexicons.

Multi-Layer Perceptron- It is feed forward neural network with one or more N-layer among input and output. This model takes more time for training. This model is used where the complex input is classified into different groups.
Support Vector Machine- This method is based on decision plane that defines decision boundary. Multiclass SVM is extension to SVM, which is the most widely used classifier. Multiclass SVM classifies data into multiple classes. It uses one vs one or one vs all method to distinguish. It usually performs well.

V. DESIGN ISSUES

The previous research work has already revealed that two new trends are studied i.e. analysis of ambiguous emotion detection and multimodal analysis of human affective states, which is comprised of audio-visual analysis, linguistic and paralinguistic analysis. The following paragraph addresses the problems that occur at various levels in developing emotion detection.

The present schemes are dealing with only facial image which holds good quality of resolution. But to be practical this is not the case indeed existing system may fail due to several reasons such as low resolution, fast movement of subject etc.

Another issue with respect to present system in how to reliable extract there linguistic and paralinguistic features from the speech signal plenty of prosodic features has been from the emotion detection within a certain speech signal, but such research activity were not satisfied to overcome the traditional challenges. There is need of an automatic extraction of linguistic information from the spoken contents. Spoken contents can be extracted from speech signals with the help of speech to text converter.
In order to implement an effective multimodal, the fusion of different modes must be done with suitable joint feature vectors which is composed of various features of different modalities with differ in time scales.

Furthermore to make multimodal effective and realistic other parameters can be considered with respect to human entity such as age, gender.

VI. MULTIMODAL SENTIMENT ANALYSIS

One mode doesn't give sufficient solution. There is need to consider other modes also such as audio/video. Social media are a huge untapped source of user opinion for various products and services. Multi Modality entails the use of multiple media such as audio and video in addition to text to enhance the accuracy of sentiment analyzers.Textual emotional classification is done on basis of polarity, intensity of lexicons. Audio emotional Classification is done on basis of prosodic features. Video emotional Classification is done on basis postures, gestures etc. In integration, we can integrate the results of all these modes.

Facial expressions helps to understand the present emotion of human thus various methods have been proposed to identify typical part of face and movement of specific points associated with different emotional states. Often techniques used are sign judgement in which method describes the appearance in spite the meaning of shown behaviour, message judgement which addresses the interpretation of shown behaviour. One more technique is known as FACS (Facial Action Coding System) propped by Ekman which does manual labelling of facial behaviour, with respect to sign judgement. The movement of certain parts of facial muscles are encoded by FACS.

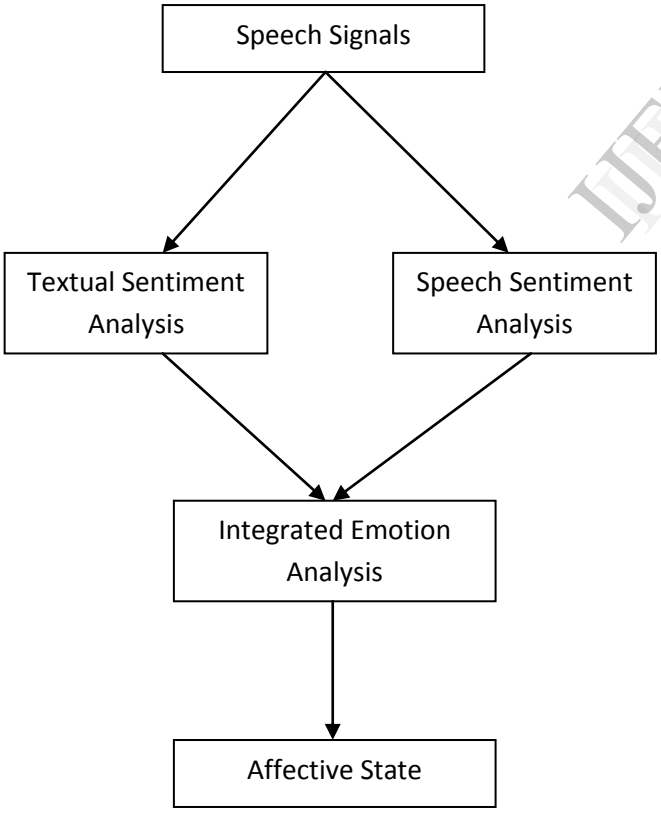

Figure 1: Integrated Speech and Text Emotional Analyser

Speech is another non trivial part which serves purpose of emotion reorganization. Speech helps to understand the emotion information via explicit (linguistic) message and implicit (paralinguistic) message most of the researches made use of vocal features for acoustic analysis e.g. mean, standard deviation, speech rate, energy in utterances minimum and maximum of pitch contour.

Another important component to identify emotion is linguistic contains the existing sentiment in the text identified by lexical dictionary e.g. sentiwordnet

Here in figure 1 given speech signal is converted into textual format and fed to textual analyzer. From same speech signal vocal features are extracted and fed to audio analyzer. Output of these two is taken under consideration for final decision of detected affective state. The textual part will provide the state and audio analysis will focus on speech intensity & pitch through amplitude & frequency respectively. Thus we will be in a position to provide the degree of sentiments.

Speech signals convey not only words but also emotions. Various analysis and models had been submitted and explored for Textual Analysis but this analysis is incomplete due to ignorance of Sentiments involved and result may not be reliable and in addition Textual Analysis only on focus is on word content and thereby ignores the acoustic features of speech . Thus it needs analysis of Sentiments as well as text simultaneously. This project explores such methodology for analysis of textual and sentimental aspect of speech, thereby providing new effective techniques for understanding speech (Audio) signals.

VII. APPLICATION OF MULTI-MODAL SENTIMENT ANALYSIS OVER BIG DATA

In [12], the volume, variety, velocity are properties of data, whether it comes from the Internet or an enterprise resource planning system, sentiment analysis system should get the data and analyse it. Due to the large volume of opinion rich web resources such as discussion forum, review sites, blogs and news corpora available in digital form much of the current research is focusing on the area of sentiment analysis and opinion mining

The Opinion mining provides functions over to data, no matter how great its volume, velocity or variability or where it lives. That data is often out-of-date, missing or stored in disparate systems, so analyser relies on their talent alone to make strategic decisions.

Variety- Data is available in different formats e.g. Text, audio, video etc. Opinion mining on unstructured data
Volume- Opinions are given in audio, video files which have large size. Opinion mining on large amount of data.
Velocity- Opinion mining on high rate incoming data. E.g. Billions of people are posting their opinions.
Some of the applications Product Reviews, Lie Detector, Prognosis of physician disorder, Voting advice system, Automated content analysis, Marketing managers, firms ,equity investors.

In order to identify potential risks, it is important for companies to collect and analyze information about their competitors' products and plans. Sentiment analysis find a major role in competitive intelligence (Kaiquan Xu , 2011) to extract and visualize comparative relations between products from customer reviews, this information can be used o improve product, marketing strategy and potential risk can be identified in early days.

CONCLUSION

Multi-modal Sentiment Analysis problem is a machine learning problem that has been a research interest for recent years. Though lot of work is done till date on sentiment analysis, there are many difficulties to sentiment analyser since Cultural influence, linguistic variation and differing contexts make it highly difficult to derive sentiment. The reason behind this is unstructured nature of natural language.

The main challenging aspects exist in use of other modes; dealing with Multi Modality entails the use of multiple media such as audio and video in addition to text to enhance the accuracy of sentiment analyzers. Textual emotional classification is done on basis of polarity, intensity of lexicons. Audio emotional Classification is done on basis of prosodic features. Video emotional Classification is done on basis postures, gestures etc. In fusion, we can integrate the results of all these modes; to get more accuracy. Future research could be dedicated to these challenges. So we are moving from uni-modal to multi-modal.

REFERENCES

1. Boiy, E. Hens, P., Deschacht, K. &Moens, M.F., "Automatic Sentiment Analysis in Online Text", In Proceedings of the Conference on Electronic Publishing(ELPUB-2007).
2. J. Wiebe, T. Wilson, and C. Cardie. "Annotating expressions of opinions and emotions in language", Language Resources and Evaluation, 2005.
3. Scott Brave and Clifford Nass, Emotion in Human-Computer Interaction. Retrieved from http://lrcm.com.umontreal.ca/dufresne/COM7162/EmotionHumanInteraction.pdf
4. S. V. Bo Pang, Lillian Lee, "Thumbs up? Sentiment classifcation using machine learning techniques", Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), ACL, pp. 79–86,July 2002.
5. Pravesh Kumar Singh and Mohd Shahid Husain "Methodological study of opinion mining and sentiment analysis techniques" International Journal on Soft Computing (IJSC) Vol. 5, No. 1, February 2014
6. Gelareh Mohammadi and Alessandro Vinciarelli "Automatic Personality Perception: Prediction of Trait Attribution Based on Prosodic Features", IEEE transactions on affective computing, vol. 3, no. 3, july-september 2012
7. Litman, D.J. and Forbes-Riley, K., "Predicting Student Emotions in Computer-Human Tutoring Dialogues". In Proc. of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL), July 2004
8. Lee C M Narayanan, S.S., "Toward detecting emotions in spoken dialogs". IEEE Tran. Speech and Audio Processing, Vol. 13 NO. 2, March 2005
9. Mozziconacci, S., "Prosody and Emotions". Int. Conf. on Speech Prosody. 2002
10. Danielle Lottridge, Mark Chignell, and Michiaki Yasumura," Identifying Emotion through Implicit and Explicit Measures: Cultural Differences, Cognitive Load, and Immersion", IEEE transactions on affective computing, vol. 3, no. 2, april-june 2012
11. Mohammad Soleymani, Maja Pantic, Thierry Pun," Multimodal Emotional Recognition in Response to Videos" , IEEE transactions Affective Computing , Vol.3,No.2, April-June 2012
12. Gaurav Vasmani and Anuradha Bhatia "A Real Time Approach with Big data- A Survey", International Journal of Engg Sciences & Research Technology, Vol 3, Issue 9, September 2013
13. LouisPhilippe, Morency Rada Mihalcea and Payal Doshi "Towards Multimodal Sentiment Analysis: Harvesting Opinions from the Web", ICMI'11, November 14–18, 2011, Alicante, Spain.
14. Andrea Kleinsmith and Nadia Bianchi-Berthouze "Affective Body Expression Perception And Recognition:A Survey", IEEE transactions Affective Computing , Vol.4,No.1, January-March 2013
15. Gérard Dray, Michel Plantie , Ali Harb, Pascal Poncelet, Mathieu Roche and François Trousset, "Opinion Mining From Blogs" International Journal of Computer Information Systems and Industrial Management Applications - IJCISIM Vol. 1 2009
16. Charles B. Ward, Yejin Choi, Steven Skiena, "Empath: A Framework for Evaluating Entity-Level Sentiment Analysis", IEEE 2011
17. Georgios Paltoglou and Michael Thelwall, "Seeing Stars of Valence and Arousal in Blog Posts" IEEE transactions on affective computing, vol. 4, no. 1, January-March 2013
18. Shangfei Wang, Zhilei Liu, Zhaoyu Wang, Guobing Wu, Peijia Shen, Shan He, and Xufa Wang "Analyses of a Multimodal Spontaneous Facial Expression Database" IEEE transactions on affective computing, vol. 4, no. 1, January-March 2013
19. Ashish Tawari and Mohan Manubhai Trivedi, "Speech Emotion Analysis: Exploring the Role of Context" IEEE transactions on multimedia, vol. 12, no. 6, October 2010
20. Tal Sobol-Shikler and Peter Robinson, "Classification of Complex Information: Inference of Co-Occurring Affective States from Their Expressions in Speech", IEEE transactions on pattern analysis and machine intelligence, vol. 32,no.7,July 2010
21. Felix Weninger, Jarek Krajewski, Anton Batliner, and Bjorn Schuller, "The Voice of Leadership:Models and Performances of Automatic Analysis in Online Speeches", IEEE transactions on affective computing, vol. 3, no. 4, October-December.2012