Intensified Sentiment Analysis of Customer Product Reviews Using Acoustic and Textual Features

Sureshkumar Govindaraj and Kumaravelan Gopalakrishnan

Sentiment analysis incorporates natural language processing and artificial intelligence and has evolved as an important research area. Sentiment analysis on product reviews has been used in widespread applications to improve customer retention and business processes. In this paper, we propose a method for performing an intensified sentiment analysis on customer product reviews. The method involves the extraction of two feature sets from each of the given customer product reviews, a set of acoustic features (representing emotions) and a set of lexical features (representing sentiments). These sets are then combined and used in a supervised classifier to predict the sentiments of customers. We use an audio speech dataset prepared from Amazon product reviews and downloaded from the YouTube portal for the purposes of our experimental evaluations.

Keywords: Intensified sentiment analysis, acoustic features, lexical features, sentiment classification, customer satisfaction categorization, emotion, sentiment.

Manuscript received July 28, 2015; revised Dec. 31, 2015; accepted Jan. 18, 2016.

Sureshkumar Govindaraj (corresponding author, mgsureshkumar_in@hotmail.com) and Kumaravelan Gopalakrishnan (gkumaravelanpu@gmail.com) are with the Department of Computer Science, Pondicherry University, Puducherry, India.

I. Introduction

Much of the work carried out in relation to sentiment analysis utilizes textual data [1]–[3]. However, when it comes to expressing human emotions, one could argue that this is more easily attainable through the use of the human voice (audio) as opposed to the human hand (writing) [4], [5]. Researchers interested in devising approaches to a sentiment analysis of audio data obtainable from call centers have vast audio datasets available to them. This motivated the researchers of [4] and [5] to propose multi-model sentiment analysis techniques that utilize videos, audio reviews, and text for the purposes of opinion mining and sentiment analysis. Sentiment analysis of audio data can assist the business community in better understanding the true emotions or satisfaction levels of their customers. In simplistic terms, when speaking of customer satisfaction, one can categorize customers into one of two categories: satisfied or unsatisfied. However, the challenge lies in determining to what extent they are satisfied or unsatisfied; that is, the level of customer satisfaction. This has motivated us to develop a method for categorizing a customer's satisfaction level using acoustic features (representing various emotions) and lexical features (representing various sentiments). In general, a sentiment can be categorized into one of three categories: positive, neutral, or negative. We propose that this can be further extended to five categories: *highly positive*, positive, neutral, negative, or *highly negative*. These extended categories of sentiment can then be used to represent five levels of customer satisfaction, very satisfied, satisfied, neutral, disappointed, and very disappointed.

http://dx.doi.org/10.4218/etrij.16.0115.0684

In our proposed approach to intensified sentiment analysis, we seek to investigate the effects of acoustic features (emotions) (AFs) when they are combined with lexical features (sentiments) (LFs).

In our study, AFs of customer product reviews are extracted using Munich OpenEAR, which is a tool for automatic emotion recognition and feature extraction [6]. We use the following AFs, among others: voice intensity, loudness, energy, fundamental frequency (F0), and Mel-frequency cepstral coefficients (MFCCs); these are widely used by researchers for emotion recognition processes [7]–[9]. The audio data (customer product reviews) in our study is transcribed into text data using the automatic speech recognition tool Kaldi [10]. We then use SentiWordNet [2] for extracting features such as SWN_positive, SWN_negative, and SWN_objectivity scores.

In this paper, we contribute to the field of sentiment analysis by proposing a novel model for combining AFs and LFs in a supervised classifier; here, the AFs are expected to enhance the indications of sentiment polarity obtained from the LFs. For example, the difference between saying "it is really good" in a normal tone and saying "it is reallllly good" with an excitement in positive sentiment, or the difference between saying "it is very bad" in a normal tone and saying "it is verrrry bad" in an aggressive tone with negative sentiment, is significant in predicting an actual deep emotion or intensified sentiment. We argue that LFs alone are insufficient to predict such intensified sentiments. The popular support vector machine (SVM) classifier is used for the automatic prediction of intensified customer sentiments about products. SVM is a simple, efficient machine learning algorithm, and is widely used for pattern recognition and classification problems. Under the conditions of limited training data, it can deliver a better classification performance compared to other classifiers [11]. Thus, in this paper we use an SVM classifier to classify intensified sentiments.

Our experiments are conducted with audio clips extracted from videos downloaded from YouTube, which contain voice reviews of products purchased from Amazon.com. Results are obtained using an SVM classifier that has been trained with different types of features.

II. Literature Review

Early approaches to sentiment analysis have tended to use "bag-of-words" features extracted from text documents. Due to the lack of domain-specific training data, these approaches have poor records of accuracy [12]. Extended methods for sentiment analysis have been proposed that utilize adverb and adjective features, to enhance the sentiment polarity of LFs [12]–[14]. Bag-of-words features such as adverbs and adjectives have been proven to be suitable for use with rule-based classifiers [15]. Sentiment analysis has a variety of applications, such as text-to-speech synthesis [16], opinion mining in on-line forums and electronic news media [17], [18], question answering [19], text summarization [20], and citation analysis [21]. Machine learning techniques for sentiment analysis are considered a better alternative to existing techniques, such as the use of bag-of-words, adjectives, adverbs, and so on, due to their ability to handle large volumes of data with better accuracy [22]. Supervised classifiers, such as SVM and naive Bayes (NB), are more accurate at sentiment prediction than earlier text-based approaches. The first proposal to use supervised learning in sentiment analysis was made by Pang and others [22]. Emotional signals extracted from audio data can be used as a cue for efficient sentiment mining processes. Ezzat [23] proposed a method to predict speakers' emotions using AFs and used multiple classifiers, such as decision tree, NB, SVM, and *k*-nearest, for validation. LFs, such as adverbs, can be combined with AFs for improving the efficiency of classifiers; Dragut and Fellbaum [24] proposed a method to measure the effect of adverbs on strengthening sentiment scores. Large-scale, opinion-rich data provided by social media applications can be used for the purposes of sentiment analysis [25]. Charfuelan and Schröder [26] investigated the possible correlation between sentiment scores (attributed to LFs) and emotion scores (attributed to AFs) obtained from sentences. They found that the average energy and mean fundamental frequency were the best features for sentiment analysis as these features achieved a higher correlation compared to other features.

Moghaddam and Ester proposed a method called Opinion Digger [27]. Their method uses part-of-speech tags as features in an unsupervised machine learning algorithm to determine a set of aspects.

Our proposed method for performing intensified sentiment analysis on customer voice responses focuses on studying the effect of an AFs, particularly when these features are combined with LFs.

Research on the recognition of AFs, in general, focuses on voice pitch (fundamental frequency) and voice energy. There are alternative acoustic attributes (as opposed to voice pitch and voice energy) that can be used as a cue to recognize voice emotions, such as MFCCs, Mel-based speech signal power coefficients, formants, and temporal features (such as speech rate and pausing) [28].

In the methods proposed in [29] and [30], prosodic features are also used for recognition of AFs.

From the above investigations, it is observed that emotion-based sentiment analysis methods, in general, aim at recognizing both the mood and sentiment of a person through

the use of both conceptual semantics (sentiment words) and the mental (emotional) state of the person [31].

III. Proposed Method

We propose a method for performing intensified sentiment analysis; the method utilizes AFs extracted from customer product reviews (voice data) and LFs extracted from textual transcriptions of the same customer product reviews.

The objective of this paper is to clearly identify customer opinions about products in a more refined, categorical sense as opposed to just stating that an opinion is either "positive" or "negative." Customer opinions are classified into five types, through which one can then infer the overall sentiment and emotion of the customers.

The architecture of the proposed method is depicted in Fig. 1. The method accepts voice data as an input and will output an intensified sentiment (opinion). The central part of the architecture consists of a set of classifier models, which are trained using different feature sets. The trained classifier models can be used to predict intensified sentiments from new voice data through the use of an effective feature set. The preferred feature set is the combination of both AFs and LFs. The AFs are extracted using OpenEAR [6], a natural-language processing toolkit. The toolkit is capable of extracting multiple AFs from wave-format audio clips.

We select only a small number of features [7]–[9], each of which is widely used in research related to emotion recognition. The LFs are extracted by converting voice data into speech text using a speech recognition toolkit called Kaldi [10]. However, not all the words present within a speech text are expected to carry indications of sentiment. Thus, we proposed a method to automatically extract LFs from textual documents containing customer product reviews using a piece of software called Stanford Parser [32], which generates a parse tree for each individual sentence. Then, feature selection is performed by using a binary decision tree–based rule engine that applies a set of hand-coded rules to the generated lexical parsing patterns. The acoustic and lexical features are validated using well-performing classifier models to identify a suitable feature set for the more accurate prediction of intensified sentiments from voice data.

In this work, we use an SVM classifier to analyze the effectiveness of AFs and LFs for intensified sentiment prediction.

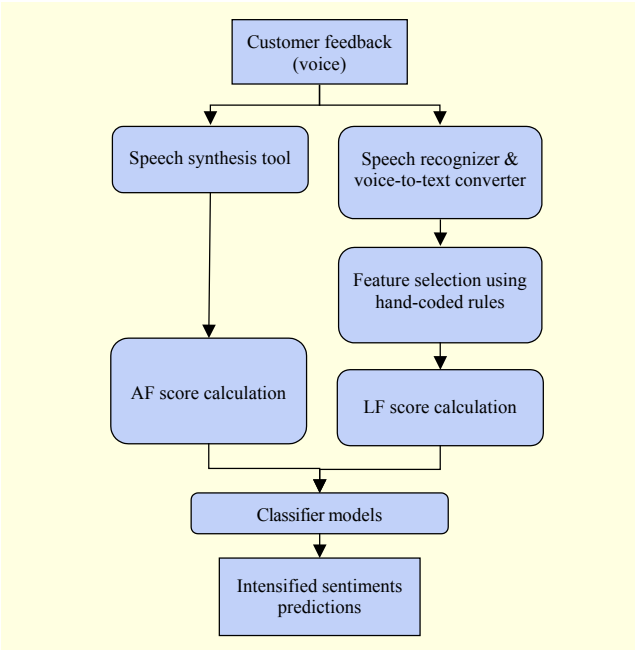

Fig. 1. Architecture of proposed method.

1. Feature Selection Using Hand-Coded Rules

In this paper, we propose a rule-based feature selection method to extract from an input sentence only those words that carry sentiment. The process of extracting such words from a lexical parsing pattern is performed using a set of hand-coded rules. The rules are formulated by an empirical analysis performed on an experimental dataset and are shown in Table 1.

A pattern-matching engine checks a lexical parse tree for the presence of aspect terms and words that carry sentiment.

The formulated hand-coded rules are specifically suited to lexical parsing patterns generated by Stanford Parser.

Extracting sentiment-carrying lexicons from individual sentences is a basic pattern matching process. Thus, the rules

Table 1. Hand-coded rules for lexical parsing patterns.

Rule no.	Rule components		Example
	Attribute lex.-pattern	RHS of parent VP	
1	VBZ	NP, S	"is"
2	VBZ + DT	NP, S	"is a"
3	VBD	NP, S	"used"
4	VBZ + IN	PP	"used in"
5	VBD + IN	PP	"used in"
6	VBG + IN	PP	"using in"
7	VBN + TO	PP	"used to"
8	VBN + IN	PP	"used for"
9	VB + RP	NP	"carry out"
10	VBP	ADJP	"are"
11	VBP	NP	"are"
12	VBP	ADVP	"play song"

are treated in the rule engine as If-Then-Else Normal Form patterns. A binary decision tree–based rule engine is used for this purpose. The outcome of the rule engine is a triplet comprising three components: aspect (representing an attribute of a product), predicate (linking an emotion with the aspect), and target sentiment-carrying words. For example, from the input source sentence "the sound quality of the mobile is very good," the following features are selected: "sound," "is," and "very good." A set of such triplets is used in the calculation of sentiment scores for the purposes of training the classifier models; ultimately, the scores are used to predict intensified sentiments.

2. LF Score Calculation

The lexical tool SentiWordNet [2] is used for calculating scores attributed to the set of lexical features (sentiment scores). SentiWordNet consists of annotated synsets of WordNet according to the notions of "positivity," "negativity," and "neutrality." SentiWordNet calculates three numerical scores, Pos(s), Neg(s), and Obj(s), for each word of an input datum. It does so by using the synsets associated with each word. The three scores serve to indicate the positive, negative, and objective nature of the words contained within the input datum. The scores are then used to construct a feature vector (representing sentiment).

As an example, the objectivity score of a word belonging to an input datum is calculated according to the following equation:

$$\text{ObjScore} = 1 - (\text{PosScore} + \text{NegScore}). \quad (1)$$

Given a datum (input sentence), we construct its corresponding feature vector (representing sentiment), which consists of a set of sentiment scores; each sentiment-carrying word within the input sentence is assigned a sentiment score during the feature extraction process.

In total, there are six lexical features (positive, negative, objectivity, positive mean, negative mean, and objectivity mean) used for each datum (input sentence). The arrangement of sentiment scores in each feature vector is as follows: <pos_score_1> <neg_score_1> <obj_score_1> <pos_score_2> <neg_score_2> <obj_score_2> ... <pos_score_*n*> <neg_score_*n*> <obj_score_*n*> <pos_mean> <neg_mean> <obj_mean>, where *n* represents the number of words within a datum. In addition, the unweighted mean sentiment score (that is, the mean of the three aforementioned means) is included as a part of each feature vector.

3. AF Score Calculation

The scores attributed to the set of acoustic features (emotion scores) are calculated from the voice data. We use the most important speech-related AFs currently used in research related to emotion recognition [7], [8].

The open-source natural-language processing software OpenEAR [6] is used in this work for extracting AFs from the voice data. The 23 most important voice-data measures calculated from prosodic, energy, voicing, spectrum, and cepstral features are used to form a set of AFs [6]. Prosodic features are extracted from both the frequency and the amplitude of the speech signal and represent its intensity, loudness, and pitch. Energy features describe the human loudness perception, and the presence of voice is identified using an energy feature. Cepstral features represent changes or periodicity in those spectrum features that are measured by MFCCs (calculated based on the Fourier transform of a speech frame). The complete set of AFs represents the emotional state of the speaker in question.

The following signal processing steps are performed by the OpenEAR tool for calculating the MFCCs and energy scores:

- Speech signal pre-processing
- Framing
- Windowing
- FFT or DFT conversion
- Mel filter bank and frequency wrapping
- Log conversion
- DCT calculation
- MFCCs calculation
- Energy coefficients calculations

Based on the above steps, we select 23 acoustic features (each of which represents emotion) and divide them into five categories (see the list below). During the feature extraction process, each emotion-carrying word within an input sentence (datum) is assigned an emotion score. The resulting emotion scores are then used in the construction of a corresponding feature vector (representing emotion).

- Pitch: minimum, maximum, mean, standard deviation, absolute value, quantile, ratio between "voiced" and "unvoiced" frames.
- Duration: time (ε_t) and height (ε_h).
- Intensity: Minimum, maximum, mean, standard deviation, quantile.
- Formant: first formant, second formant, third formant, fourth formant, fifth formant, second formant/first formant, third formant/first formant.
- Rhythm: speaking rate.
- In the above list, ε_t and ε_h are duration features, which are prominent measures of emotion.

Figure 2 depicts the measure of F0 for computing parameters ε_t and ε_h, which corresponds to the rising and lowering of intonation. The said duration features are a measure of the gaps

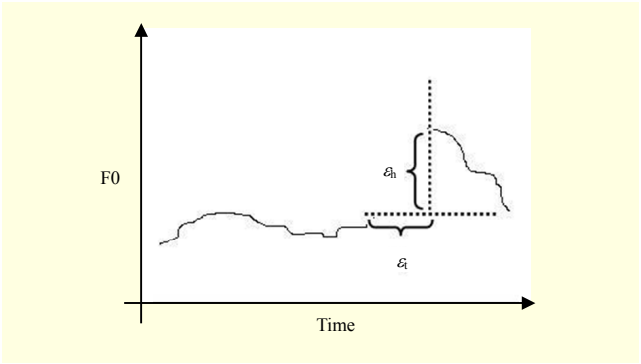

Fig. 2. Measure of parameters ε_t and ε_h relative to F0.

that appear in a wave form (that is, indications of silence between words). The notation ε_t refers to the pause time between two disjoint segments of F0 (known as "pitch"), whereas ε_h refers to the vertical distance between the two disjoint segments, which symbolizes a "voice break." The inclusion of the duration features accounts for possible low or high pitch accents.

4. Formation of LF + AF Feature Vector

We create a single combined feature vector by taking the six sentiment scores corresponding to the six lexical features of the feature vector representing sentiment and the 23 emotion scores corresponding to the 23 acoustic features of the feature vector representing emotion; thus, each datum (input sentence) is represented by a single combined feature vector comprising 29 scores.

IV. Data and Experiments

Our data set consists of 200 audio clips extracted from videos taken from the portal YouTube. The videos contain customer product reviews of products purchased from amazon.com. The same video dataset has already been used in other researches on sentiment analysis [33], [34].

We selected data from the customer reviews of Amazon products and categorized each datum into one of five different categories (scenarios representing sentiment); each categorization is carried out based on the speaker's emotion, which is inferred from the audio data at the time. These scenarios reflect a customer's opinion of a product in terms of a Likert [35] five-point scale, whereby the scale ranges from "very satisfied" to "very disappointed." The customer product reviews cover five different products; mobile phones, mp3 players, digital cameras, shoes, and watches. The audio clips are converted from speech to text for extracting lexical sentiment features. The audio clips used for training the classifiers are pre-labeled depending on the context of the speech; we used 80 data instances for training the classifier, 80 instances for validation, and 40 instances for testing.

For the purpose of feature evaluation, we prepared a separate training and test set for each feature using LFs, AFs, and a combination of the two (AFs + LFs). We separately tested each data set to find the effect of AFs on strengthening the sentiment classification process.

An SVM is considered by many researches to be a state-of-the-art technology for machine learning [36]–[38]. SVMs are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. We use LibSVM [39], a multiclass SVM model, for assigning five different class labels to the customer product review data. A multiclass SVM (*n*-SVM) defines a hyperplane by calculating the distances between feature values. The general approach adopted to achieve multiclass classification is the use of binary classification.

We use an *n*-SVM model with an RBF kernel, with a bias and gamma of 1.0 as kernel function parameters.

V. Results

For the purpose of a performance comparison, we manually evaluated the classification results obtained from the test data sets through use of a Gold Standard evaluation strategy. Based on the evaluation, "correct" and "incorrect" predictions are calculated to obtain a prediction accuracy. To avoid skewed or sparse results produced by the classifier, the classifier is trained using multiple training and test data sets. The overall training accuracy of the classifier is then calculated by taking the average of the accuracy values obtained from these training sets. This approach is known as a *k*-fold cross-validation, and it serves to optimize the prediction accuracy. In such an approach, the training is iteratively conducted *k* times (folds) by dividing the labeled data into *k* nearly equal–sized data subsets. In each fold, a training dataset is constructed using all but one of the *k* data subsets; the remaining data subset is used for validation of the fold. We used 5-fold cross-validation as an optimum limit

Table 2. Overall classification results with LF, AF, and LF + AF data sets.

Classifier models	Prediction accuracy (%)	Cross-validation accuracy (%)
LF	69.70	70.50
AF	64.56	65.62
LF + AF	83.12	83.33

based on our empirical evaluations.

The trained classifier has been tested using a separate unlabeled test set comprising 40 instances. The prediction accuracy of the classifier has been calculated as the ratio between the number of correct predictions and the total number of predictions. The cross-validation accuracy results and prediction accuracy results obtained using the three different classification models with three categories of features, LFs, AFs, and LFs + AFs, are depicted in Table 2.

VI. Discussion and Comparisons

From the accuracy graph shown in Fig. 3, it is evident that the prediction accuracy is closer to the cross-validation accuracy for all three classification models. Thus, the SVM classifier's performance is well above the acceptable level. The prediction performance of the model with AFs (64.56%) is not as good as the model with LFs (69.7%). The reason for this is the better feature selection provided by hand-coded rules and the semantic support provided by SentiWordNet. However, when the LF model was bootstrapped by AFs in the LFs + AFs model, it produced an overall prediction accuracy of 83.12%. From this, it is evident that the intensified sentiment analysis performance is improved by combining the AFs and LFs.

We compared our results with similar well-performing methods that used both Amazon product review data as source data and an SVM classifier for validation. P´erez-Rosas and others [33] proposed a method for sentiment classification from acoustic signals extracted from video data and achieved a recall accuracy of 74.66%. Ezzat [23] reported an overall accuracy of 74.4% using an SVM Key Graph method to analyze speech emotion recognition. Poria and others [34] proposed a method for sentiment analysis using audio data prepared using Amazon product reviews and an SVM for machine learning. They achieved an overall accuracy of 77.03%. In comparison to the aforementioned approaches, our proposed method achieves a much better accuracy.

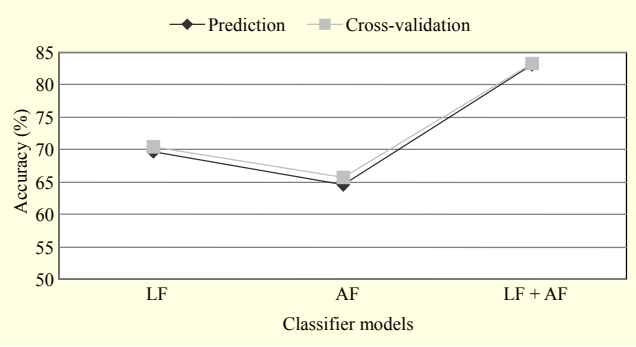

Fig. 3. Comparison between cross-validation accuracy and prediction accuracy of SVM classifier.

VII. Conclusion and Future Work

In this paper, we propose a method to extract 23 (AFs) and six (LFs) from a given input audio clip (WAVE file format). The extracted features are used to predict the intensified sentiments of a customer speaking to a call center agent. Three different feature sets (LF, AF, and LF + AF) are extracted and used to build three corresponding classifier models. We used an SVM classifier for developing a trained classifier model to predict the opinions of a customer. For this, the opinions of a customer speaking to a call center agent are is categorized into one of the following five types: very satisfied, satisfied, neutral, disappointed, or very disappointed. These types are mapped to five levels of sentiment score, highly positive, positive, neutral, negative, and highly negative.

The notable contribution of our work lies in the analysis of the role of human emotions expressed in the form of speech for improving the classification accuracy of an intensified sentiment analysis. From the experimental results, it is evident that the human emotion signal increases the classification accuracy of the sentiment categorization process. The findings of this qualitative study can be used in a variety of applications such as automated customer behavior analysis, customer redress systems, customized retail services, and business process quality improvement.

This work can be further extended by correlating emotion cues with the sentiment polarities of corresponding words to exploit any explicit relationship between sentiment and emotion. Furthermore, our approach can be tested with languages other than English to devise a multilingual intensified sentiment analysis system.

References

[1] J. Wiebe and E. Riloff, "Creating Subjective and Objective Sentence Classifiers from Unannotated Texts," *Int. Conf. Intell. Text Process. Computational Linguistics*, Mexico City, Mexico, 2005, pp. 486–497.

[2] A. Esuli and F. Sebastiani, "SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining," *Language Resources Evaluation Conf.*, Genoa, Italy, May 2006, pp. 417–422.

[3] A.L. Maas et al., "Learning Word Vectors for Sentiment Analysis," *Proc. Ann. Meeting Association Computational Linguistics: Human Language Technol.*, Portland, OR, USA, 2011, pp. 142–150.

[4] L.P. Morency, R. Mihalcea, and P. Doshi, "Towards Multimodal Sentiment Analysis: Harvesting Opinions from the Web," *Int. Conf. Multimodal Interfaces,* Alicante, Spain, 2011, pp. 169–176.

[5] J. Wagner et al., "Exploring Fusion Methods for Multimodal Emotion Recognition with Missing Data," *IEEE Trans. Affective*

Comput., vol. 2, no. 4, Dec. 2011, pp. 206–218.

[6] F. Eyben, M. Wöllmer, and B. Schuller, "OpenEAR - Introducing the Munich Open-Source Emotion and Affect Recognition Toolkit," *Int. Conf. Affective Comput. Intell. Interaction Workshop*, Amsterdam, Netherlands, Sept. 10–12, 2009, pp. 8–12.

[7] C. Busso, S. Lee, and S. Narayanan, "Analysis of Emotionally Salient Aspects of Fundamental Frequency for Emotion Detection," *IEEE Trans. Audio, Speech Language Process*., vol. 17, no. 4, May 2009, pp. 582–596.

[8] D. Ververidis, C. Kotropoulos, and I. Pitas, "Automatic Emotional Speech Classification," *Int. Conf. Acoust., Speech Signal Process*., Montreal, Canada, May 2004, pp. 593–596.

[9] Y. Han, G. Wang, and Y. Yang, "Speech Emotion Recognition based on MFCC," *J. ChongQing University Posts Telecommun*., vol. 20, no. 15, 2008, pp. 1162–1181.

[10] D. Povey et al., "The Kaldi Speech Recognition Toolkit," *IEEE Workshop Automat. Speech Recog. Understanding*, Hawaii, USA, Dec. 2011, pp. 1–4.

[11] T.L. Pao et al., "Mandarin Emotional Speech Recognition Based on SVM and NN," *Int. Conf. Pattern Recogn.*, vol. 1, Hong Kong, China, Sept. 2006, pp. 1096–1100.

[12] P.D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," *Ann. Proc. Meeting Assoc. Computational Linguistics*, Philadelphia, PA, USA, 2002, pp. 417–424.

[13] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Seattle, WA, USA, 2004, pp. 168–177.

[14] M. Taboada et al., "Lexicon-Based Methods for Sentiment Analysis," *Computational Linguistics*, vol. 37, no. 2, May 2011, pp. 267–307.

[15] B. Yang and C. Cardie, "Extracting Opinion Expressions with Semi-markov Conditional Random Fields," *Conf. Empirical Methods Natural Language Process. Computational Natural Language Learn*., Jeju, Rep. of Korea, 2012, pp. 1335–1345.

[16] C.O. Alm, D. Roth, and R. Sproat, "Emotions from Text: Machine Learning for Text-Based Emotion Prediction," *Conf. Empirical Methods Natural Language Process*., Vancouver, Canada, 2005, pp. 347–354.

[17] K. Balog, G. Mishne, and M. de Rijke, "Why are They Excited? Identifying and Explaining Spikes in Blog Mood Levels," *Proc. Conf. European Chapter Assoc. Computational Linguistics*, Trento, Italy, 2006, pp. 207–210.

[18] P. Carvalho et al., "Liars and Saviors in a Sentiment Annotated Corpus of Comments to Political Debates," *Proc. Ann. Meeting Assoc. Computational Linguistics*, Portland, OR, USA, 2011, pp. 564–568.

[19] J.H. Oh et al., "Why Question Answering Using Sentiment Analysis and Word Classes," *Joint Conf. Empirical Methods Natural Language Process. Computational Natural Language Learn*., Jeju, Rep. of Korea, 2012, pp. 368–378.

[20] G. Carenini, R. Ng, and X. Zhou, "Summarizing Emails with Conversational Cohesion and Subjectivity," *Assoc. Computational Linguistics: Human Language Technol.*, Columbus, OH, USA, 2008, pp. 773–782.

[21] A. Athar and S. Teufel, "Context-Enhanced Citation Sentiment Detection," *Conf. North America Chapter Assoc. Computational Linguistics: Human Language Technol.*, Montreal, Canada, 2012, pp. 597–601.

[22] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs Up?: Sentiment Classification Using Machine Learning Techniques," *Assoc. Computational Linguistics Conf. Empirical Methods Natural Language Process.,* Morristown, NJ, USA, 2002, pp. 79–86.

[23] S. Ezzat, N. el Gayar, and M. Ghanem, "Sentiment Analysis of Call Centre Audio Conversations Using Text Classification," *Int. J. Comput. Inf. Syst. Ind. Manag. Appl*., vol. 4, no. 1, 2012, pp. 619–627.

[24] E. Dragut and C. Fellbaum, "The Role of Adverbs in Sentiment Analysis," *Frame Semantics NLP: Workshop Honor Chuck Fillmore* (*1929–2014*), Baltimore, MA, USA, 2004, pp. 38–41.

[25] X. Hu et al., "Unsupervised Sentiment Analysis with Emotional Signals," *Proc. Int. Conf. World Wide Web*, Rio de Janeiro, Brazil, 2013, pp. 607–618.

[26] M. Charfuelan and M. Schröder, "Correlation Analysis of Sentiment Analysis Scores and Acoustic Features in Audiobook Narratives," *Int. Workshop Corpora Res. Emotion Sentiment Social Signals*, Istanbul, Turkey, 2012, pp. 1–5.

[27] S. Moghaddam and M. Ester, "Opinion Digger: An Unsupervised Opinion Miner from Unstructured Product Reviews," *Proc. ACM Int. Conf. Inf. Knowl. Manag.*, Toronto, Canada, Oct. 2010, pp. 1825–1828.

[28] C. Busso et al., "Analysis of Emotion Recognition Using Facial Expressions, Speech and Multimodal Information," *Proc. Int. Conf. Multimodal Interfaces*, State College, PA, USA, Oct. 13, 2004, pp. 205–211.

[29] R. Tato et al., "Emotional Space Improves Emotion Recognition," *Int. Conf. Spoken Language Process*., CO, USA, 2002, pp. 2029–2032.

[30] M. El Ayadi, M. Kamel, and F. Karray, "Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases," *Pattern Recogn.*, vol. 44, no. 3, 2011, pp. 572–587.

[31] D. Ververidis and C. Kotropoulos, "Emotional Speech Recognition: Resources, Features, and Methods," *Speech Commun.*, vol. 48, no. 9, Sept. 2006, pp. 1162–1181.

[32] M.C. de Marneffe and C.D. Manning, "The Stanford Typed Dependencies Representation," *Proc. Workshop Cross-framework Cross-Domain Parser Evaluation*, Manchester, UK, Aug. 23, 2008, pp. 1–8.

[33] V. P´erez-Rosas, R. Mihalcea, and L. Morency, "Utterance-Level Multimodal Sentiment Analysis," *Ann. Meeting Association*

Computational Linguistics, Sofia, Bulgaria, Aug. 4, 2013, pp. 973–982.

[34] S. Poria et al., "Fusing Audio, Visual and Textual Clues for Sentiment Analysis from Multimodal Content," *Neurocomputing*, vol. 174, Jan. 2016, pp. 50–59.

[35] K.L. Wuensch, "*What is a Likert Scale? and How Do You Pronounce 'Likert?'*," Ph.D. dissertation, East Carolina University, Greenville, North Carolina , USA, Apr. 30, 2009.

[36] H. Drucker, D. Wu, and V. Vapnik, "Support Vector Machines for Spam Categorization," *IEEE Trans*. *Neural Netw.*, Sept. 1999, pp. 1048–1054.

[37] S. Dumais et al., "Inductive Learning Algorithms and Representations for Text Categorization," *Int*. *Conf*. *Inf*. *Knowl*. *Manag*., Washington, DC, USA, Nov. 1998, pp. 148–155.

[38] T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," *European Conf*. *Mach*. *Learn*., Chemnitz, Germany, Apr. 21–23, 1998, pp. 137–142.

[39] C. Chang and C. Lin, "LIBSVM: A Library for Support Vector Machines," *ACM Trans*. *Intell*. *Syst*. *Technol*., vol. 2, no. 3, Apr. 2011, pp. 27–54.

Sureshkumar Govindaraj is working as an assistant professor with the Department of Computer Science, Pondicherry University, Puducherry, India. He received his PhD degree in computer science and engineering from Pondicherry University, in 2014. He received his MS degree in computer science and engineering from the College of Engineering Guindy, Anna University, Chennai, India, in 2006. Currently, he teaches postgraduate degree courses. His research areas of interest include knowledge engineering and information retrieval systems.

Kumaravelan Gopalakrishnan is working as an assistant professor with the Department of Computer Science, Pondicherry University, Puducherry, India. He received his PhD degree in computer science from Bharathidasan University, Tiruchirappalli, India in 2013. He received his MS degree in information technology from Bharathidasan University in 2009. Currently, he teaches postgraduate degree courses. His research areas of interest include human–computer interaction and spoken dialogue systems.