Social Network Analysis and Mining (2018) 8:28
https://doi.org/10.1007/s13278-018-0505-2

ORIGINAL ARTICLE

Emotion detection from text and speech: a survey

Kashfia Sailunaz[1] · Manmeet Dhaliwal[1] · Jon Rokne[1] · Reda Alhajj[1]

Received: 7 December 2017 / Revised: 18 March 2018 / Accepted: 27 March 2018 / Published online: 7 April 2018

Abstract

Emotion recognition has emerged as an important research area which may reveal some valuable input to a variety of purposes. People express their emotions directly or indirectly through their speech, facial expressions, gestures or writings. Many different sources of information, such as speech, text and visual can be used to analyze emotions. Nowadays, writings take many forms of social media posts, micro-blogs, news articles, etc., and the content of these posts can be useful resource for text mining to discover and unhide various aspects, including emotions. Extracting emotions behind these postings is an immense and complicated task. To tackle this problem, researchers from diverse fields are trying to find an efficient way to more precisely detect human emotions from various sources, including text and speech. In this sense, different word-based and sentence-based techniques, machine learning, natural language processing methods, etc., have been used to achieve better accuracy. Analyzing emotions can be helpful in many different domains. One such domain is human computer interaction. With the help of emotion recognition, computers can make better decisions to help users. With the increase in popularity of robotic research, emotion recognition will also help making human–robot interaction more natural. This survey covers existing emotion detection research efforts, emotion models, emotion datasets, emotion detection techniques, their features, limitations and some possible future directions. We focus on reviewing research efforts analyzing emotions based on text and speech. We investigated different feature sets that have been used in existing methodologies. We summarize basic achievements in the field and highlight possible extensions for better outcome.

Keywords Emotion · Text · Emotion models · Emotion recognition · Emotion analysis · Speech · Classifiers

1 Introduction

Emotion recognition is the process of identifying human emotion by merely depending on personal skills and interpretation, by automating the process, or by a semi-automated approach. Many methodologies have been developed for automated emotion recognition, e.g., human computer interaction (HCI) (Cowie et al. 2001; Kjeldskov and Graham 2003; Kim et al. 2009) which targets making HCI more natural (Pane et al. 2002). Indeed, there are several intelligent assistants, such as Siri (Bellegarda 2013), Cortana (Maedche et al. 2016) which use many quirks to make the interactions with users more natural. Emotion recognition can further improve HCI by allowing the system to consider emotions before performing tasks.

Recognizing human emotions accurately is a laborious task due to their versatility and ambiguity. Same emotions may be expressed in multiple ways, and multiple emotions sometimes have the same expression. Emotions may vary based on personality, gender, location, ethnicity, culture, situation, in addition to many other psychological, social and individual parameters. Sometimes, detecting the actual emotion of a person from a piece of text, his/her speech or facial expressions is difficult even for an actual human. When it comes to automatic emotion detection by a computer, we can easily imagine the level of complication of the problem. A number of researchers have been working on emotion detection from various types of inputs, e.g., text, image, audio, video. In other words, emotion recognition

✉ Kashfia Sailunaz
kashfia.sailunaz@ucalgary.ca

Manmeet Dhaliwal
manmeet@ucalgary.ca

Jon Rokne
rokne@ucalgary.ca

Reda Alhajj
alhajj@ucalgary.ca

[1] Department of Computer Science, University of Calgary, Calgary, AB, Canada

can be performed based on different sources of information, such as speech (Lalitha et al. 2015; Kwon et al. 2003; Seehapoch and Wongthanavasu 2013), text (Calvo and Kim 2013; Perikos and Hatzilygeroudis 2016; Nahin et al. 2015; Mohammad and Kiritchenko 2015), or visually (Cohen et al. 2000; Rosenblum et al. 1994; Daga et al. 2016). There are several systems which use combination of information for emotion recognition, e.g., Kessous et al. (2010) and Schmitt et al. (1997). But in many cases, not all information is available. For example, some online retailers like Amazon[1] might perform emotion recognition through text, since they have text reviews data available. Call centers might use audio speech recognition for emotion detection (Neiberg et al. 2006). While emotion recognition for security cameras will have to rely on visual information only. In this work, we will focus on emotion detection from text and speech. For text-based detection, the input text can be any social media posts, comments, reviews, or any other form of text.

'Emotion' detection from text is comparatively a more complex task than 'Sentiment' detection. Although sometimes these two terms are used synonymously, they differ in definition when used in computer science (Munezero et al. 2014). According to Oxford Dictionary, 'Emotion' is "a strong feeling deriving from one's circumstances, mood, or relationships with others", whereas 'Sentiment' is "a view or opinion that is held or expressed". Cambridge Dictionary defines 'Emotion' as "a strong feeling such as love or anger, or strong feelings in general" and 'Sentiment' as "a thought, opinion, or idea based on a feeling about a situation, or a way of thinking about something". In general, 'Sentiment' is defined as the effect of 'Emotion' (Broad 1954). '*Happy*', '*Anger*', '*Love*' are examples of emotions and '*Positive*', '*Negative*', '*Positive*', respectively, are examples of corresponding sentiments of these emotions.

Sentiment analysis identifies subjective information from text to find out the polarity of attitude of a person toward another person, thing, event or task. On the other hand, emotion detection focuses on finding out how a person feels about some event, person or thing based on some predefined emotion models according to psychological emotion theories. Emotion detection from text has its application in almost every aspect of our daily life, .e.g., making efficient e-learning systems based on student's emotion, improving human computer interactions, monitoring mental health of people, modifying or improving business strategies according to the emotion of customers, detecting public emotion on any national, international or political event, detecting potential criminals or terrorists from analyzing the emotions of people after a terrorist attack or crime, improving the performances of chatbots and other automatic feedback systems, and so on.

With the rise of social networks, people now capture and express their emotions more frequently and comfortably through their social media activities. Instead of the increasing popularity of audio and video components, text is still the most common form of communication in social media. People convey their emotions through social media posts like Facebook status, Tweets, comments on own or other people's posts, product reviews, micro-blogs. Analyzing these texts and detecting emotion from their words and semantics is quite challenging. Emotion detection from text has been a promising research topic over years and considerable efforts try to build a perfect automated system capable of detecting correct human emotion from text.

Emotion detection from speech should not be neglected and should receive attention similar to emotion recognition from text. The way a person speaks, volume, tone, speed, words, etc., all contribute considerably to emotion detection from speech. However, culture, age, gender, etc., are other factors which may also affect emotion detection from speech. Fortunately, researchers have developed several approach for emotion detection from speech. These approaches are characterized by their specific advantages and attractions which turn them more preferred compared to others. In this paper, we have tried to cover from the literature recent works which handle emotion detection in text and speech. Existing works have been described with their basic features, approaches, limitations and scopes for improvements.

The rest of the paper is structured as follows—Sect. 2 describes emotion detection from different inputs. Section 3 discusses existing works on emotion detection from text. Section 4 covers emotion detection from speech. Section 5 is conclusions.

2 Emotion analysis

Human beings express emotions in multidimensional ways. Some common ways are through their writing, their speech, their facial expression, their body language, their gesture, etc. These emotions can be categorized with different emotion models. To detect and analyze emotions from any content, an appropriate emotion model need to be selected. It should define the set of emotions applicable for a particular problem.

2.1 Evolution of emotion

In 1872, after completing some psychological experiments on facial expressions captured in different circumstances on both humans and animals, Charles Darwin claimed that human and other animals express emotions with similar

[1] https://www.amazon.ca/.

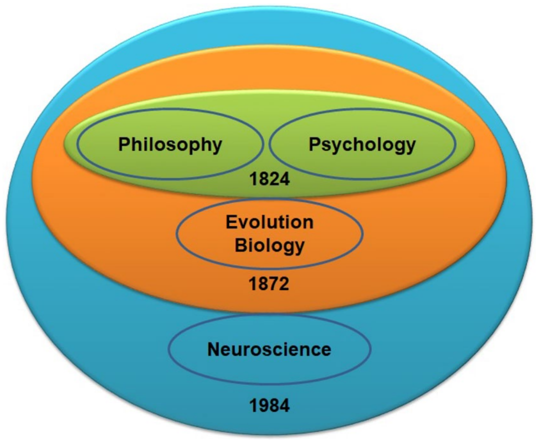

Fig. 1 Evolution of emotion in various fields

expressions and behavior under similar circumstances (Darwin 1998). His theory of emotion was also related to era and associated circumstances. According to his claims, emotions in human and other animals were realized over time. He discussed about general principles of emotions in detail; means of expressions of emotions in both human and animals; causes and effects of all possible emotions like anxiety, grief, dejection, despair, joy, love, devotion, etc.; explanation of emotions with images to show the expressions for particular emotions. Darwin claimed that some emotional expressions are universal for people all over the world. He also claimed that not only humans, but also animals of similar species react similarly to a situation. His experiments revealed that for some emotions can have similar expressions even in species which are not very similar. Some philosophical and spiritual categorization of emotions were present before that as described in Manser (1963) and Bell (1824).

The study of emotions started as a branch of philosophy and psychology theories. Darwin claimed that emotions and their expressions are also related to biological causes. Later, emotions were described as brain mechanisms that are outputs of functional properties of neural systems (LeDoux 1984). Figure 1 shows the evolution of emotion in different fields of studies.

According to the evolution theory, it was found that different human emotions were realized at different phases of human life in various eras. Human emotions were defined, explained, categorized and analyzed by psychologists, sociologists, neuroscientists, biologists and researchers from many other fields over time; hence, various emotion models emerged to cover all possible human emotions.

2.2 Emotion models

From psychological point of view, human emotions can be identified and grouped based on emotion type, emotion intensity, and many other parameters, which can be all combined and realized into emotion models. Emotion models are structured form of defining various human emotions according to some scores, ranks or dimensions. Existing emotion models define various kinds of emotions according to duration, intensity, synchronization, rapidity of change, event focus, appraisal elicitation, behavioral impact (Borod 2000).

Based on different emotion theories, existing emotion models can be divided into two classes—*Categorical* and *Dimensional* (Calvo and Kim 2013). *Categorical* emotion models define a list of categories of emotions which are discrete from each others. On the other hand, *Dimensional* emotion models define a few dimensions with some parameters and specify emotions according to those dimensions. Two or three dimensions are used in most dimensional emotion models—'*valence*' (indicates the positivity or negativity of an emotion), '*arousal*' (indicates the excitement level of an emotion) and '*dominance*' (indicates the level of control over an emotion) (Sreeja and Mahalakshmi 2017; Canales and Martinez-Barco 2014).

Table 1 summarizes few basic emotion models used frequently in the literature and express almost all possible human emotions. The table lists the emotions for each model and the dimensions for dimensional emotion models. Figure 2 shows some emotion models that are mostly used in emotion-based research.

2.3 Emotion detection

Emotion extraction from different types of social network components is a research topic which is being investigated for a long time now. Various kinds of contents posted by people on social networking platforms has been analyzed to detect emotions behind the posts. In some works, any combination of voice tone, speech, facial expressions, gestures, EEG signals, different types of bio signals and texts were used to detect emotion from multimodal data (Poria et al. 2016; Kudiri et al. 2016; Soleymani et al. 2016). On the other hand, most of the works on emotion detection focus on one specific type of content.

Speech, voice tone and frequency from audios were used as inputs for emotion detection in Semwal et al. (2017), Anagnostopoulos et al. (2015), whereas facial expressions and gestures from videos were used to extract emotions in Kahou et al. (2016). But as we mentioned earlier, this paper will focus on emotion detection from text and speech, i.e., emotion detection from images will be left out as future work. In Sect. 3, we will present a survey on the research works based on emotion recognition from textual contents; and in Sect. 4 we will cover emotion detection from speech.

Table 1 Emotion models

Model	Emotions	Approach	Structure
Ekman (1992)	Anger, disgust, fear, joy, sadness, surprise	Categorical	–
Shaver et al. (1987)	Anger, fear, joy, love, sadness, surprise	Categorical	Tree
Oatley and Johnson-Laird (1987)	Anger, anxiety, disgust, happiness, sadness	Categorical	–
Plutchik (1980)	Acceptance, admiration, aggressiveness, amazement, anger, annoyance, anticipation, apprehension, awe, boredom, contempt, disapproval, disgust, distraction, ecstasy, fear, grief, interest, joy, loathing, love, optimism, pensiveness, rage, remorse, sadness, serenity, submission, surprise, terror, trust, vigilance	Dimensional	Wheel
Circumplex Russell (1980)	Afraid, alarmed, angry, annoyed, aroused, astonished, at ease, bored, calm, content, delighted, depressed, distressed, droopy, excited, frustrated, glad, gloomy, happy, miserable, pleased, relaxed, sad, satisfied, serene, sleepy, tense, tired	Dimensional	Valence, arousal
OCC Ortony et al. (1988)	Admiration, anger, appreciation, disappointment, disliking, fear, fears-confirmed, gloating, gratification, gratitude, happy-for, hope, liking, pity, pride, sorry-for, relief, remorse, reproach, resentment, self-reproach, shame	Dimensional	Tree
Lovheim (2012)	Anger/rage, contempt/disgust, distress/anguish, enjoyment/joy, fear/terror, interest/excitement, shame/humiliation, surprise/startle	Dimensional	Cube

3 Emotion analysis from text

In the present era of smart phones and social networking platforms, a number of people prefer sharing their emotions, sentiments, achievements, etc. via audio or video files because of the real time features of these components. However, even if the popularity of these interactive contents is increasing, most people still turn to text for their communication and interaction in daily life as well as on social networking platforms.

Text is still the primary choice for people to express their feelings toward other persons, events or things. Extracting emotion from text is still more difficult because of the nature of the data. Detecting emotions from text is easier if words representing a particular emotion were explicit in the text. But most of the time, emotion is expressed in a subtle way. Again, sometimes, there may exist multiple emotions in a single piece of text. Then, some text has emotions and words which are ambiguous, some words have multiple meanings, and multiple words mean the same emotion. Some text represents sarcasm, or use slangs. Multilingual text, spelling mistakes, acronyms, grammatically incorrect sentences are some common characteristics of texts available online. These limitations of textual data sometimes make automatic emotion detection nearly impossible. Research on emotion extraction from text is a very popular topic. Modifications, improvements and new approaches to address challenges of this task is attracting the attention of researchers all over the world. In Table 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12, we have listed some of the existing works and their contributions in this field. We have included the works from the last ten years (2008–2017).

3.1 Emotion detection methods

Several approaches exist for textual emotion detection. Actually, emotion recognition task is a part of *Affective Computing* and the computational methods used in this area have been classified into several categories by various researchers. The methods defined by these researchers can be generalized into four categories—*Keyword-based Method*, *Lexicon-based Method*, *Machine learning Method* and *Hybrid Method*. Existing works utilized unigrams (one word), n-grams (multiple words), emoticons, hashtag words, punctuations, negations as features for the emotion detection task. Approaches for textual emotion detection are defined as described in the following subsections.

3.1.1 Keyword-based method

Keyword-based emotion detection is the most intuitive and straightforward approach. The idea is to find out patterns similar to emotion keywords and match them. The first task is to find out the word which expresses the emotion in a sentence. This is normally done by tagging the words of a sentence with Parts-Of-Speech tagger and then extracting the Noun, Verb, Adjective and Adverb (NAVA) words. Most linguistic and emotion-based researches proved that these are the most probable emotion carrying words. Then these words are matched against a list of words representing emotions according to a specific emotion model. Whichever emotion matches with the keyword is considered as the emotion of the specific sentence. Different approaches can be applied when the word matches with multiple emotions from the list. In some keyword-dictionaries, each word has a probability score for each emotion and the emotion with

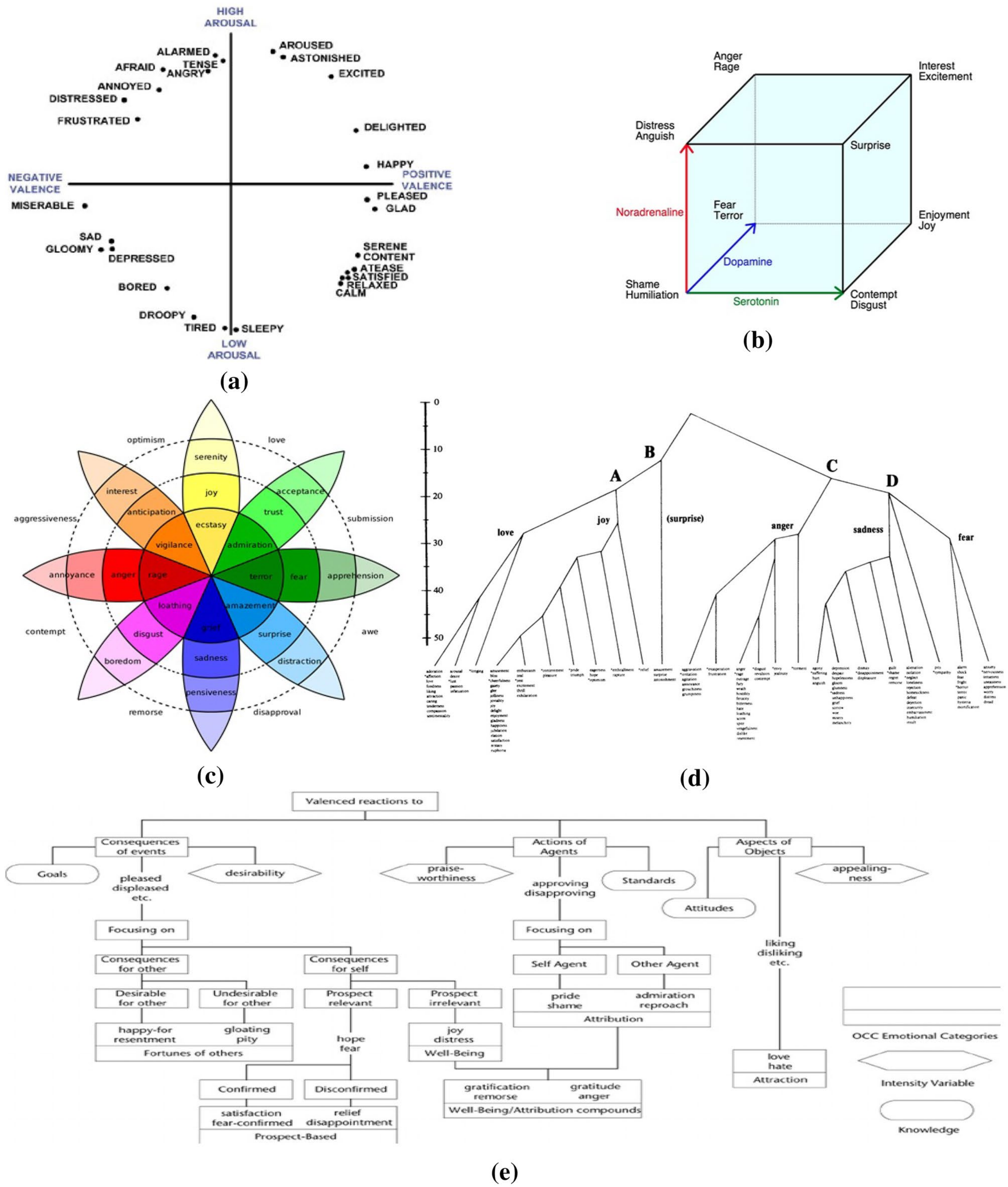

Fig. 2 Emotion models. **a** Circumplex emotion model. **b** Lovheim's emotion cube. **c** Plutchik's emotion wheel. **d** Shaver's emotion model. **e** OCC emotion model

the highest score is picked as the emotion of the word. In some other works, the first emotion matched with the word is picked as the primary emotion of the word. The reference list of keywords or the keyword dictionary differs depending on the researcher. The keyword dictionary is normally constructed by researchers based on the emotions and the

Table 2 Emotion analysis from text

Authors	Type	Approach	Method	Emotions	Features	Limitations
Yadollahi et al. (2017)	Survey	–	Existing emotion detection methods	Basic Emotion Theories	Existing Lexicons, Datasets, Mining methods for Twitter, Mining for English and other languages; Well structured investigation and categorization of emotion mining	–
Binali et al. (2010)	Survey and a proposed approach	Hybrid of keyword and learning-based method	Keyword and learning-based methods	Basic Emotion Theories	Description of computational approaches for emotion detection from text; Improved emotion detection by combining semantic and syntactic information	–
Canales and Martinez-Barco (2014)	Survey	–	Lexicon and machine-learning-based methods	Basic Emotion Theories	Collection of Lexical-based, Supervised-learning-based and Unsupervised-learning-based existing works; Comparisons between lexical and machine-learning-based approach and limitations of existing systems	Proposed a new direction on deep analysis but didn't explain
Chopade (2015)	Survey	–	Keyword Spotting, Lexical Affinity, Learning-based and Hybrid method	Basic Emotion Theories	Applications of textual emotion detection, limitations of existing approaches, text normalization techniques	–

Table 3 Emotion analysis from text

Authors	Type	Approach	Method	Emotions	Features	Limitations
Tripathi et al. (2016)	Survey	–	Keyword spotting, lexical affinity, statistical NLP	Basic emotions, their properties representations and models	Emotive potential of text, Definition and generation of Datasets and Lexicons, List of existing textual emotion detection works, emotion analysis applications, possible future directions	–
Kao et al. (2009)	Survey and a proposed approach	Case-based reasoning method	Keyword-based, machine-learning-based and hybrid methods	22 OCC emotions	Formally defined 'emotion detection' problem, existing approaches and their limitations, possible solutions, a new approach with semantic analysis and case-based reasoning	Proposed method was not implemented
Shivhare and Khethawat (2012)	Survey and a proposed approach	Emotion ontology-based emotion detector algorithm	Keyword Spotting, Lexical Affinity, Learning-based and Hybrid method	–	Discussions on different emotion detection techniques, their limitations, design of a new ontology-based algorithm	Proposed method was not implemented
Gupta et al. (2017)	Novel approach	LSTM-based deep learning model	Machine learning (SVM, decision trees, Naive Bayes)	Angry, happy, sad, others	Combines sentiment and semantic-based embeddings, outperforms most baseline machine learning approaches	Unable to handle context
Desmet and Hoste (2013)	Novel approach	NLP and sentiment mining	Machine learning—SVM	Abuse, anger, blame, fear, forgiveness, guilt, happiness, hopefulness, hopelessness, information, instructions, love, pride, sorrow, thankfulness	Uses machine learning for emotion detection in suicide notes, uses both semantic and lexical features like bags-of-words, POS tags and trigrams	Lack of data, unable to recognize rare emotions, did not consider negations

Table 4 Emotion analysis from text

Authors	Type	Approach	Method	Emotions	Features	Limitations
Dini and Bittar (2016)	Novel approach	Symbolic and machine learning	Lexical-based and machine learning	Anger, disgust, fear, joy, sadness, surprise	Creates emotion tweet corpus for classification (ETCC) and emotion tweet corpus for relevance (ETCR), uses word, lemma, noun phrase, dependencies between POS as feature set	Quality verification needed for two new corpora
Mohammad and Bravo-Marquez (2017)	Novel approach	Best worst scaling (BWS)	Regression	Anger, fear, joy, sadness	Detects emotion intensity, creates four tweet datasets, shows correlations between emotion pairs, uses word n-grams, character n-grams, word embeddings and affect lexicons for regression	Accuracy not verified
Summa et al. (2016)	Novel approach	Interdisciplinary method	Graph-based semi-supervised learning, NLP	Anger, happiness, disgust, fear, sadness, surprise, none	Combines linguistic, temporal, and spatial information; computes similarity between tweet nodes	Accuracy was not very high, parameters selection was random
Sen et al. (2017)	Novel approach	Jointly learnt model for emotion and sentiment	Machine learning—SVM, CNN	Emotion—anger, amusement, excitement, happiness, hope, love, sadness; sentiment—positive, negative, neutral, none	Uses sentiments as auxiliary input for textual emotion detection, uses multi-tasking NN for embedding learning	–
Jain et al. (2017)	Survey and a proposed approach	Machine learning (SVM, Naive Bayes)	Intelligent text processing	Basic emotion theories, Ekman (for framework)	Compilation of different emotion detection methods in existing researches, detail survey on election prediction tweets, combines corpus-based features and emotion related features for multilingual text emotion detection	Proposed framework focuses on only political, health and sports domain

Table 5 Emotion analysis from text

Authors	Type	Approach	Method	Emotions	Features	Limitations
Kang et al. (2017)	Novel approach	Bayesian model	Bayesian inference method	Anger, anxiety, expect, hate, joy, love, sorrow, surprise	Works with contextual information to get the latent semantic dimensions, predicts emotion for both word and document text, Two Bayesian models DWET and HDWET are applied, DWET outperforms all baseline methods.	Bayesian model will never converge if the semantic dimensions increase much, works with Chinese language only
Perikos and Hatzilygeroudis (2016)	Ensemble approach	Machine learning	Ensemble of Naive Bayes, maximum entropy learner, knowledge-based tool	Circumplex model	A sentence-level emotion and polarity detector, analyzes each sentence for feature extraction and inserts them into three classifiers, output is decided by a voting system	Lack of lexical resources, accuracy is not better than all baseline methods
Nahin et al. (2014)	Novel approach	Machine learning (VSM with Jaccard Similarity)	Combination of keystroke and text pattern analysis	Anger, disgust, fear, guilt, joy, sadness, shame	Collects user keystrokes and emotions with and without some specific texts, uses this data to train and test the classifiers	Lack of proper user participation, time consuming data collection, lack of adequate words in dataset
Mohammad and Kiritchenko (2015)	Novel approach	Machine learning (majority classifier, SVM)	Machine learning with combinations of 11 feature sets containing different features	Anger, disgust, fear, joy, sadness, surprise	Collection of emotion tweets labeled with hashtag emotion words, generation of a large word-emotion association lexicon, detection of both emotion and personality, calculation of extroversion, neuroticism, agreeably, conscientiousness and openness	–

Table 6 Emotion analysis from text

Authors	Type	Approach	Method	Emotions	Features	Limitations
Agrawal and An (2012)	Novel approach	Machine learning	Unsupervised Context-based method	Ekman's model	Unsupervised approach classifies emotions based on the semantic relationships of Noun, Verb, Adjective and Adverb, generates emotion vectors with NAVA, uses a few emotion representative words to start, Do not need any annotated dataset or lexicon, Accurate than other unsupervised and some supervised method, Can work with any emotion model	Semantic relatedness is dependent on the corpus
Tiwari et al. (2016)	Novel approach	Hybrid approach	Hybrid of machine learning and rule-based methods	Ekman's model	Detail explanations about emotion, emotion models, emotion detection methods and their limitations; uses a combination of machine learning and rule-based method; Uses emotion vectors with NAVA extracted by POS tagger	Uses a small dataset which effects the generalization and standards of the results
Yat et al. (2010)	Novel approach	Emotion-cause detection	Rule-based emotion-cause detection	Anger, fear, happiness, sadness, surprise	Generated a Chinese annotated emotion—cause corpus, linguistic analysis was done to find out the correlation between emotion and cause, linguistic rules used for cause detection	Works with Chinese language only
Li and Xu (2014)	Novel approach	Emotion and cause detection and analysis	Rule-based emotion-cause detection and machine-learning-based emotion detection	Ekman's model	Works with Chinese micro-blog data, emotion-cause extraction system was used to extract the features, System was trained with that data, SVM was used to detect emotion	Works with Chinese language only, do not work well with complicated linguistic patterns

Table 7 Emotion analysis from text

Authors	Type	Approach	Method	Emotions	Features	Limitations
An et al. (2017)	Novel approach	Machine learning-based approach for emotion detection in song lyrics	Machine learning—Naive Bayes	Comfortable, happy, inspirational, joyful, lonely, miss, nostalgic, passionate, quiet, relaxed, romantic, sad, soulful, sweet, yearning	Naive Bayes classifier was used for emotion classifications from music lyrics, the positivity and negativity of emotions were also analyzed, four different dataset of English and Chinese music lyrics were used for classification	The result for English lyrics were not as accurate as Chinese lyrics
Hajar (2016)	Novel approach	Machine learning	Unsupervised machine learning approach	Ekman's model	You tube comments texts are used for emotion detection, NAVA words are extracted from the sentences, PMI for each emotion word against NAVA words were calculated for measuring the correlations	–
Li et al. (2016)	Novel approach	Machine learning	Hybrid of neural networks (biterm topic model and convolution neural network)	Ekman's model	Uses biterm topic model (BTM) which uses probabilistic latent semantic analysis on the words of a text to find out the co-occurrence patterns; BTM is used to learn the layers of a neural network; Back-propagation method is used to fine-tune the network	User context was not considered
Strapparava and Mihalcea (2008)	Comparative study	Existing methods	Keyword-based, lexicon-based, supervised and unsupervised machine learning-based methods	Ekman's model	A comparative study to find out which method is better for annotating datasets	No decision was made after the comparisons; Not a complete survey on emotion detection; No new approach was proposed

Table 8 Emotion analysis from text

Authors	Type	Approach	Method	Emotions	Features	Limitations
Ghazi et al. (2014)	Novel approach	Hybrid method	Hybrid of keyword-based, lexicon-based and machine learning-based approaches with logical regression	Ekman's Model	Uses keywords to find to match with the sentences; Extracts syntactic and semantic features from the sentences; Applies machine learning to classify the sentences according to their emotions by using the extracted features	Lack of generalization of features; Indirect dependencies between emotion word and other words of the sentence was ignored
Hasan et al. (2014)	Novel approach	Machine learning	Naive Bayes, support vector machine, decision tree, K-nearest Neighbor	Circumplex Model	Uses hashtag words, punctuations, emoticons and negations as features; Uses hashtags to label the tweets automatically	–
Hasan et al. (2017)	Novel approach	Machine learning	Emotex Classifier	Angry, Happy Sad	Filters tweets with a binary classifier to find out the tweets with emotion; Applies Emotex to extract features; Classifies tweets about a public event; Generates decision about public mood on an event; Tracks temporal changes in public emotions	Considers only three emotions
Kanger and Bathla (2017)	Novel approach	Machine learning	Combination of fuzzy logic and Neural Network	Anger, Happy, Sad	Applies hybrid of fuzzy logic and neural network; Creates fuzzy rules and extracts features from the input data	Absence of results on benchmark data; Lack of generalization for any domain; Lack of detail of some steps of the approach; Works with three emotions only

Table 9 Emotion analysis from text

Authors	Type	Approach	Method	Emotions	Features	Limitations
Grover and Verma (2016)	Hybrid approach	Hybrid of keyword-based and machine learning	Keyword-based and Machine learning (Naive Bayes and support vector machine)	Ekman's Model	Tokenizes the data; After removing stop words applies stemming steps; Extracts features and applies rule-based engine to filter data with emotions; Uses classifiers to classify the filtered data	Works with Punjabi language only
Joshi et al. (2016)	Novel approach	Lexicon-based approach	Lexicon (LIWC and Emo-Lex)-based Emotion Tracker	Angry, Anxious, Happy, Sad	Develops an emotion tracker; Downloads tweets, predicts emotion, predicts overall emotion in a time-period; Generates and shows emotion time-sequence graphs; Computes the excitement levels of a cricket match, emotions between characters of a play and emotions toward a product accurately	Lack of detection of subjectivity and emotion interactions

words they are associated with. There are online tools and programs like WordNet[2] which can find synonyms and antonyms of words that can be used to make the dictionary.

3.1.2 Lexicon-based method

Lexicon-based emotion detection approach classifies text using a lexicon (a knowledge-base with text labeled according to emotions) appropriate for the input dataset (Dini and Bittar 2016; Mohammad and Bravo-Marquez 2017). Here emotion detection is similar to the previous method, but in this case, an emotion lexicon is used instead of the word list. National Research Council of Canada (NRC) and EmoSenticNet (ESN) are some of the commonly used emotion and sentiment lexicons (Abak and Evrim 2016).

3.1.3 Machine learning method

Both supervised and unsupervised machine learning methods are used for textual emotion detection in which a model is designed to train a classifier with a part of the dataset and then test the classifier with the rest of the data. For supervised method, an annotated emotion dataset is used for training and testing the supervised classifier. Naive Bayes, support vector machine, and decision tree may be named as the most commonly used classifiers. By definition of 'Unsupervised Classification', the data used are not labeled with the classes. The classifier starts with several seed words for each emotion which are then cross-referenced to the sentences. This way, sentences are classified to corresponding emotions. This trains the classifier model which is then further used to label the testing data. Unsupervised method is a more generalized one but in most cases, supervised classification achieves better accuracy.

3.1.4 Hybrid method

Hybrid approach for emotion detection in text combines any two or all three methods defined to achieve the benefit of multiple methods and reach the maximum level of accuracy. In some previous works, it was proved that applying a combination of multiple emotion detection methods gives

[2] https://wordnet.princeton.edu/.

Table 10 Emotion analysis from text—features: 1

Features	Binali et al. (2010)	Kao et al. (2009)	Shivhare and Khethawat (2012)	Gupta et al. (2017)	Desmet and Hoste (2013)	Dini and Bittar (2016)	Mohammad and Bravo-Marquez (2017)	Summa et al. (2016)	Sen et al. (2017)	Jain et al. (2017)
Keyword-based	X									
Lexicon-based						X	X			
Learning-based	X			X	X	X		X	X	X
Supervised machine learning-based				X	X				X	X
Support vector machine					X				X	X
Naive Bayes				X						X
Bayesian models										
Decision tree				X						
K-nearest neighbor										
Neural networks									X	
Maximum entropy										
Majority classifier										
Unsupervised machine learning-based										
Semi-supervised machine learning-based								X		
Hybrid method	X									
Ensemble method										
Fuzzy logic										
Rule-based method										
Reasoning method		X								
Semantic-based method		X		X	X					
Ontology-based method			X							
Regression-based method							X			
Uses sentiment			X	X					X	
Generation of new data corpus						X	X			
Generation of new lexicon										
Works with English language	X	X	X	X	X	X	X	X	X	X
Works with Other languages										X
Detects single emotion	X	X	X	X	X	X	X	X	X	X
Detects multiple emotions										
Word-level detection	X		X	X	X	X	X	X	X	X
Sentence-level detection		X								
Document-level detection										
Emotion intensity detection							X			
Personality detection										

Table 10 (continued)

Features	Binali et al. (2010)	Kao et al. (2009)	Shivhare and Khethawat (2012)	Gupta et al. (2017)	Desmet and Hoste (2013)	Dini and Bittar (2016)	Mohammad and Bravo-Marquez (2017)	Summa et al. (2016)	Sen et al. (2017)	Jain et al. (2017)
Emotion-cause detection										
Public mood detection										
Temporal effect										

better results than individual methods (Binali et al. 2010; Tiwari et al. 2016).

According to the previous surveys (Yadollahi et al. 2017; Binali et al. 2010; Canales and Martinez-Barco 2014; Chopade 2015; Tripathi et al. 2016) conducted to compare the accuracy of emotion detection using *Keyword-based Method*, *Lexicon-based Method*, *Machine learning Method* and *Hybrid Method*, it was mentioned that *Keyword-based Method* and *Lexicon-based Method* worked better. *Machine learning Method* worked better with larger datasets, while *Hybrid Method* gives similar accuracy to *Machine learning Method*. Although these methods are the benchmark approaches for textual emotion detection, some researchers tried natural language processing (Desmet and Hoste 2013), linguistic rule-based methods (Kanger and Bathla 2017), ensemble of multiple methods cite37 or some novel methods almost completely unique and achieved some good results. Modifications to those novel methods may lead to an improved and highly accurate automatic emotion detection system from text input.

3.2 Datasets

In the literature, different existing or customized datasets have used for emotion detection according to the types of the experiments of different researchers. The annotated datasets are annotated with particular emotion models, hence bearing only those emotions that exist in the specific emotion model. A small number of emotion labeled datasets are available. Some researchers contributed some data corpus that they built to utilized in their experiments. But still the number of available datasets for emotion analysis is not enough. Further, there is no generalized dataset which can be used for any emotion model. Researchers are either bound to use the emotion model that exists in the datasets or to generate a data corpus and then manually annotate the data according to emotion labels. The latter task is time consuming. Some of the most commonly used datasets in recent works are listed in Table 13 together with a brief description of the amount of data that exists in each dataset.

3.3 Discussions

Emotion detection is a research field with lots of unturned stones and possibilities. Due to the diverse nature of human emotions, there are still lots of scopes for developing unique systems or improving and enriching existing systems with effective modifications and capabilities. Some possible future directions in the research of emotion detection and analysis from texts are listed in the following sections.

Table 11 Emotion analysis from text—features: 2

Features	Kang et al. (2017)	Perikos and Hatzilygeroudis (2016)	Nahin et al. (2014)	Mohammad and Kiritchenko (2015)	Agrawal and An (2012)	Tiwari et al. (2016)	Yat et al. (2010)	Li and Xu (2014)	An et al. (2017)
Keyword-based									
Lexicon-based									
Learning-based	X	X	X	X	X	X		X	X
Supervised machine learning-based	X	X	X	X				X	X
Support vector machine				X				X	
Naive Bayes		X							X
Bayesian models	X								
Decision tree									
K-nearest neighbor									
Neural networks									
Maximum entropy		X							
Majority classifier				X					
Unsupervised machine learning-based					X				
Semi-supervised machine learning-based									
Hybrid method						X			
Ensemble method		X							
Fuzzy logic									
Rule-based method						X	X	X	
Reasoning method									
Semantic-based method	X				X				
Ontology-based method									
Regression-based method									
Uses sentiment		X							
Generation of new data corpus							X		
Generation of new lexicon				X					
Works with English language			X	X	X	X			X
Works with other languages	X	X					X	X	X
Detects single emotion	X	X	X	X	X	X	X	X	X
Detects multiple emotions									
Word-level detection	X		X	X	X	X	X	X	X
Sentence-level detection		X							
Document-level detection	X								

Table 11 (continued)

Features	Kang et al. (2017)	Perikos and Hatzilygeroudis (2016)	Nahin et al. (2014)	Mohammad and Kiritchenko (2015)	Agrawal and An (2012)	Tiwari et al. (2016)	Yat et al. (2010)	Li and Xu (2014)	An et al. (2017)
Emotion intensity detection									
Personality detection				X					
Emotion-cause detection							X	X	
Public mood detection									
Temporal effect									

3.3.1 Emotion intensity detection

Consider amount of research is going on about emotion detection, but a very small number of researchers are concentrating on intensity of the detected emotion. Each emotion may have different intensity level and detecting the intensity of an emotion can be beneficial for emotion analysis. For example, if a person writes “I am sad” or “I want to end my life, there is nothing left for me”, then the existing systems will label both of these declarations as ‘sad’ emotions. However, in reality, the intensity of sadness in both statements is different—the former represents normal sadness while the other reflects severe depression or suicidal emotion.

3.3.2 Sarcasm detection

Sarcasm detection from text is a very complex task and the maximum accuracy achieved so far is not very much. Correct sarcasm detection depends on correctly using sentence structure, detecting accurate emotion behind the sentence, understanding the context of the sentence, and many other parameters. For example, “I’m trying to imagine you with a personality”[3] represents a sarcasm which is very implicit and hard to detect for an existing automatic sarcasm detection system.

3.3.3 Multiple emotion detection

In most of the emotion detection efforts, researchers concentrate on the primary emotion in the text. Sentences expressing multiple emotions are either discarded or labeled with the first emotion that was detected. For example, if a person writes “I was sad this morning but now I am happy” or “This makes me happy and sad at the same time”, then the system should be able to recognize both emotions with a temporal and/or spatial dimension.

3.3.4 Emotion-cause detection

A limited number of works focus on the cause of the emotion that was expressed by a person. Sometimes, detecting the cause can increase the accuracy of the detection of the correct emotion in a text or even speech. For example, if a person expresses “I am so happy! It is raining!”, then the system should recognize that the reason of his/her happiness is rain. This can also be applied to personalized systems where the system has knowledge of the likes and dislikes of a person and can detect the cause of an emotion even if the cause is not directly stated in the same sentence.

3.3.5 Personality or mood detection

By detecting the emotion from text for a particular person, his/her personality or mood can be detected and analyzed. Personality or mood detection can be integrated into existing social networking platforms and other applications for personalized suggestions. For example, if a person is angry and is expressing anger in his/her text, then an application may suggest to him/her different things he/she likes (i.e., a nearby restaurant which serves his/her favorite food, a nearby movie theater where a new movie is playing which is of his/her favorite genre, and so on).

3.3.6 Emotion versus individual or social parameters

Human emotions can be related to their age, gender, time, location, ethnicity, political views, educational qualifications, in addition to some other individual, social, temporal and spatial parameters. The connection between emotion and these parameters can present some new dimensions for emotion analysis-based research. For example, adult people are in general more concerned about political issues than

[3] Your Dictionary. Examples of sarcasm.

Table 12 Emotion analysis from text—features: 3

Features	Hajar (2016)	Li et al. (2016)	Strapparava and Mihalcea (2008)	Ghazi et al. (2014)	Hasan et al. (2014)	Hasan et al. (2017)	Kanger and Bathla (2017)	Grover and Verma (2016)	Joshi et al. (2016)
Keyword-based			X	X				X	
Lexicon-based			X	X	X	X			X
Learning-based	X	X	X	X	X	X	X	X	
Supervised machine learning-based		X	X	X	X	X	X		
Support vector machine				X	X	X		X	
Naive Bayes			X		X	X		X	
Bayesian models									
Decision tree					X	X			
K-nearest neighbor					X	X			
Neural networks		X					X		
Maximum entropy									
Majority classifier									
Unsupervised machine learning-based	X								
Semi-supervised machine learning-based									
Hybrid method		X		X				X	
Ensemble method									
Fuzzy logic							X		
Rule-based method							X	X	X
Reasoning method									
Semantic-based method		X		X					
Ontology-based method									
Regression-based method				X					
Uses sentiment									
Generation of new data corpus			X						
Generation of new lexicon									
Works with English language	X	X	X	X	X	X	X		X
Works with other languages								X	
Detects single emotion	X	X	X	X	X	X	X	X	X
Detects multiple emotions									
Word-level detection			X		X	X	X	X	X
Sentence-level detection	X	X		X					X
Document-level detection									
Emotion intensity detection									
Personality detection									

Table 12 (continued)

Features	Hajar (2016)	Li et al. (2016)	Strapparava and Mihalcea (2008)	Ghazi et al. (2014)	Hasan et al. (2014)	Hasan et al. (2017)	Kanger and Bathla (2017)	Grover and Verma (2016)	Joshi et al. (2016)
Emotion-cause detection									
Public mood detection						X			
Temporal effect					X	X			X

Table 13 Datasets

Dataset	Datasize	Description
EmoBank[a]	10K sentences	Double annotation with valence, arousal and dominance were used from the perspectives of both writer and reader
The valence and arousal Facebook posts (Preotiuc-Pietro et al. 2016)	2895 Facebook posts	Double annotation with valence and arousal values
The emotion in text data set[b]	40,000 Tweets	Annotated with Anger, Boredom, Empty, Enthusiasm, Fun, Happiness, Hate, Love, Relief, Sadness, Surprise, Worry, Neutral
EmoLex (Mohammad and Bravo-Marquez 2017)	14,182 Unigrams	Annotated with sentiments—Negative, Positive and Emotions—Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, Trust
ISEAR[c]	7666 sentences	Contains responses of questionnaires on seven emotions (joy, fear, anger, sadness, disgust, shame, and guilt) from 37 countries from 5 continents
Affective Text (Strapparava and Mihalcea 2008)	1200 News Headlines	Annotated with 6 basic emotions from Ekman's model and polarity

[a]JULIELab. Emobank

[b]CrowdFlower

[c]AAAC Emotion Research

teenagers. Issues in one country may effect people of neighbor countries and other countries of the world in different ways.

4 Emotion recognition from speech

Speech is an important source of information for emotion recognition. Through text only, some words might not help portray any sort of emotion. For example, the word 'Hey' will not help the system recognizing any sort of emotion. But through speech, we might be able to find out the state of emotion. Businesses like call center will not have access to visual information at all, making speech and text the only types of information they have available.

Speech emotion recognition has a wide range of applications. It can be used for in-car board system where monitoring the mental state of drivers could help improve the safety of drivers and passengers (Ayadi et al. 2011). In automatic remote call center, it can be used to timely detect customers' dissatisfaction (Lee et al. 2001). Intelligent assistant like Siri and Cortana can make better decision or change the way of interaction based on user's emotion.

There are different types of features available, such as prosodic, spectral. Many of the research papers covered in this survey use combination of both types of features to product better results (Lin and Wei 2005). Also, other sources of information such as linguistic and visual are also sometimes used to cover the weaknesses of classifiers based only on speech information.

There are both linear and nonlinear classifiers such as GMM, SVM, HMM, that are being used to classify utterances into emotions. There are different variations of how these are applied to this problem and the feature set that is being used to train these models. Also, hybrid and ensemble systems are also used to produce better results (Petrushin 2000).

4.1 Database

There are two main types of speech databases, Simulated and Natural (real world). Simulated emotional speech utterances are collected from experienced actors who can express sentences in different emotions. Recording is done in a fully setup environment without any noise. Sometimes databases can be speaker dependent, where speech

Table 14 List of speech datasets

No.	Database name	Database type	Number of Speakers	Emotions	References
1	Swedish company Voice Provider	Natural (8 kHz)	Telephone Service Customers	Neutral, emphatic, negative	Neiberg et al. (2006)
2	BerlinEmo DB	Simulated	10 (5 male and 5 female)	Anger, boredom, disgust, fear, happiness, sadness, neutral	Lalitha et al. (2014)
3	Recorded	Simulated (16 kHz)	8 native speakers (4 male and 4 female)	Anger, fear, happiness, sadness, neutral	Hu et al. (2007)
4	Danish Emotional Speech (DES)	Simulated (16 kHz)	4 (2 male and 2 female)	Anger, happiness, neutral, sadness, surprise	Lin and Wei (2005)
5	Berlin German Database	Simulated	10	Happy, angry, anxious, fearful, bored, disgusted, neutral	Pan et al. (2012)
6	SJTU Chinese Database	Simulated	Unknown	Happy, sad, neutral	Pan et al. (2012)
7	Using FERMUS 3 framework	Simulated	13 (10 male, 1 female)	Anger, disgust, fear, joy, neutral, sad, surprise	Schuller et al. (2004)
8	SUSAS	Simulated	9	Stress, noise, fear, anxiety, depression, angry, neutral	Kwon et al. (2003)
9	Recorded	Simulated	2	Anger, dislike, fear, happy, sad, surprise, neutral	Silva and Ng (2000)
10	Spanish and Sinhala	Simulated (Video/audio)	N/A	Happiness, sadness, anger, dislike, surprise, fear	Chen et al. (1998)
11	Recorded	Simulated (22 kHz)	30	Happiness, sadness, anger, fear, neutral	Petrushin (2000)
12	Interactive Emotional Dyadic Motion Capture	Simulated	10 (5 male and 5 female)	Excitement, frustration, happiness, neutral, surprise, anger, frustration, fear, neutral	Han et al. (2014)
13	Recorded	Simulated (48 kHz)	1	Sadness, happiness, anger, neutral	Busso et al. (2004)

recognition software is dependent on characteristic of the actor. While others are speaker independent where software is not trained to each speakers' characteristics. Most of the databases included in this survey fall in this category. Number of actors are usually many with male to female ration close to half. Actors express a linguistically neutral utterance in variety of emotions. Some databases include video source of information as well.

With simulated databases, complete range of emotions is available. Lots of data are available in various languages. Also, results can be compared easily, as every utterance is already labeled with the emotion an actor is portraying.

Though, there are some disadvantages with this sort of database. Actor speech tells how emotion should be portrayed rather than how they are portrayed. They are also full-blown emotions as the emotions tend to be more expressive than in a natural environment.

Natural emotional speech utterances are collected from sources like call centers, HCI environment, etc. There tends to be lots of noise present in the background. Speech software are speaker independent. This sort of databases is best since it shows how emotions are expressed in real scenarios. Though it is difficult to have access to these databases because of copyright and privacy issues. There is only one database from real-life scenario presented in this paper. It is from Swedish company voice provider which has recorded audio from telephone service customers. Not all types of emotions might be available in such databases (Table 14).

4.2 Features

An important part of designing speech emotion recognition systems is the selection of suitable features that efficiently categorize different emotions. As described in the literature, there are few different types of features that can be extracted, such as prosodic and spectral. Many times, a combination of spectral and prosodic features is used to train the model (Table 15).

Prosodic features are aspects of speech which go beyond phonemes and deal with auditory qualities of sound. We use these in our daily life to understand the emotions behind the speech of someone. There are many different variables of prosodic features, such as pitch, loudness, energy, timbre, pause. There are many variables

Table 15 List of features

No.	Features	Feature selection	References
1	MFCCs, MFCC-low using filters from 20 to 300 Hz, pitch	All features were modeled using GMMs	Neiberg et al. (2006)
2	Pitch, Entropy, Auto Correlation, Energy, Jitter and Shimmer, HNR, ZCR, Statistics	All features were used for SVM	Lalitha et al. (2014)
3	Spectral Features	GMM supervector was constructed using spectral features for each utterance	Hu et al. (2007)
4	39 features including energy, F0–F4, MFCC1, MFCC2, MBE1-MBE5	SFS feature selection was employed to select the best features	Stearns (1976) and Lin and Wei (2005)
5	Speech rate, energy, pitch, formant, spectrum features, LPC, LPCC, MFCC, MEDC, and related features	Different combination of features was used to model different SVM	Pan et al. (2012)
6	33-dimensional feature vector including pitch, energy, silence duration, signal, spectral	All features were used for Linear classifier, GMM, MLP	Schuller et al. (2004)
7	Pitch, log energy, formant, band energies, MFCCs and many more	Forward selection and backward elimination (Devijver and Kittler 1982) to rank and extract the key features	Kwon et al. (2003)
8	Pitch contours	Pitch contours are fed into a HMM using a left-right model	Silva and Ng (2000)
9	Prosodic features including pitch contours, statistic, energy, and their derivatives	All features were used for Nearest-mean criteria, model each class with Gaussian distribution	Chen et al. (1998)
10	F0, energy, speaking rate, F1–F3, BW1-BW3	RELIEF-F (Kononenko 1994) algorithm was used for feature selection	Petrushin (2000)
11	Segment level features including MFCC ad pitch related features	Only segments with highest energy in an utterance were chosen for training sample	Han et al. (2014)
12	Prosodic features as well as duration of voiced and unvoiced segments	Sequential backward features selection technique was used to get 11-dimensional feature vector	Busso et al. (2004)

of prosodic features available. Spectral features are frequency-based features, which are obtained by converting the time-based signal into frequency domain using Fourier Transform, like fundamental frequency, frequency component, etc.

When features extracted from speech sources are not enough, linguistic content of spoke utterance is used as well to classify emotions. Integration of speech and linguistic features produced better results compared to just using speech information (Schuller et al. 2004). The basic idea is that, for some emotions, models that are based on speech source only produce worse result in comparison with a model which is only based on linguistic information, and vice-versa. Therefore, to improve the result, a hybrid method which uses both sources of information is used. This is also the case for using visual information in combination with speech information. Some emotions could be easier to recognize through visual source than audio. Therefore, some research papers used a hybrid system that uses both visual and audio information to categorize emotions and produce better results (Busso et al. 2004). Finally, there exist some research efforts which categorize emotions based on all three types of information, namely visual, audio, and text (Poria et al. 2016).

4.3 Classification

Once the features are extracted from speech data, models are trained on the extracted feature set so that new instances can be classified based on the emotions they portray. There are many different classifiers used as models to analyze emotions from data, e.g., GMM, SVM, HMM, Belief network Architecture, LDA, and QDA. There are variations of each classifier in different research papers and they are usually trained on a different feature set. There is also hybrid systems which use multiple methods to classify speech data into emotion categories. We also included in this survey a paper which uses an ensemble approach to produce better results. Some of the common models are GMM, SVM, and HMM.

4.3.1 SVM

SVM is a statistical classifier which classifies data into binary classes based on training data. SVM constructs a hyperplane or a set of hyperplanes in a high-dimensional or infinite dimensional space, which can be used for classification, regression or other tasks. The utterances in speech can be categorized in many different emotions. To classify these

Table 16 Methodologies and results

No.	Methods	Methodology overview	Results	References
1	GMMs	All features were used to model GMMs on the frame level. Emotions were classified in three categories. Method has been tested on two different datasets	85% accuracy	Neiberg et al. (2006)
2	SVM	Emotions were classified into 7 different categories. Majority of features that were used are in time domain. Methodology was testing using one dataset	81% recognition rate	Lalitha et al. (2014)
3	GMM Supervector-based SVM vs GMM	GMM supervectors for calculated for each utterance, which were further used as input for SVM. Utterances were classified to 5 emotions	GMM Supervector-based SVM significantly outperforms standard GMM system	Hu et al. (2007)
4	HMM and SVM	Utterances were classified in 5 categories. Feature selection was performed using SFS. Both HMM and SVM were used for classification separately to compare	Recognition rate of 99.5% through HMM, 88.9% through SVM	Lin and Wei (2005)
5	SVM	Different combination of features was used to develop different SVM models. Best one was chosen based on accuracy rate	Accuracy rate of 91.3% for Chinese database, 95.1% for Berlin databases	Pan et al. (2012)
6	Hybrid SVM-Belief Network Architecture	Utterances were classifieds into 7 emotions. Hybrid system was built using SVM and Belief Network. Results were integrated in a soft decision fusion using MLP	Error rate of 8.0%	Schuller et al. (2004)
7	SVM, LDA, QDA, HMM	Important features were selected, multiple methods were used	Accuracy rate of 70.1% (4 emotions), 96.3% (2 emotions) using GSVM	Kwon et al. (2003)
8	HMM (bimodal, integrating audio and video)	Hybrid method, both video and audio sources were used to classify emotions into 4 categories	Approximate accuracy of 70% through video source, 30% through audio, and 72% through bimodal	Silva and Ng (2000)
9	Nearest-mean criterion, model each class with Gaussian distribution and classify test samples	Features that give highest recognition rates are selected. Both video and audio sources were used to classify emotions into 6 categories	Best accuracy of 77.8% through audio source, 97.2% through audio and video source	Chen et al. (1998)
10	k-nearest neighbor, neural network, ensemble of neural network	Emotions were classified into 5 categories. Recognition rate of each emotion was calculated. Accuracy rate of each emotion was determined to find out which emotions are being categorized more accurately	Accuracy of 55% through K-nearest neighbors, 65% through neural network, 70% through ensemble of neural network. Accuracy of classifying fear was worst, while anger and sadness was best	Petrushin (2000)
11	DNN, then ELM	DNN have been used to produce emotion state probability distribution for each segment, which was to construct utterance-level features. These features were fed into ELM to identify utterance-level emotions (5 categories)	Accuracy rate of 45% through base HMM improved to 54.3% through proposed approach	Han et al. (2014)
12	Multimodal system	Both audio and visual information have been used. Results were integrated through fusion. Emotions were classified into 4 categories	Accuracy of 70.9% through acoustic source, 85% through facial source, 89.1% through bimodal system	Busso et al. (2004)

emotions using SVM, the problem is converted into a binary classification problem (Table 16).

4.3.2 HMM

The use of HMM for speech emotion recognition has become predominant for the last several years. HMM is a Markov process with states hidden from the observer. Each state has a probability distribution over the possible output tokens. The internal behavior of HMM refers to the state sequence through which the model passes. Features such as pitch contours are fed into a HMM machine to train the model, which are then used to recognize the rest of the audio samples.

4.3.3 GMM

Gaussian mixture (Vlassis and Likas 2002) model is another method used widely for speech emotion recognition. One of the most popular methods is Maximum Likelihood (ML) (Akaike 1998) for parametric estimation of GMM. Using ML, we can estimate model parameters using the training data. During testing, GMM is used to get probability of new utterances. Emotion with the highest probability is assigned to that utterance.

4.3.4 Hybrid systems

There are some hybrid systems (Mao et al. 2007) which use multiple sources of information such as speech and linguistic to classify emotions. Once a model is trained for each information source, they are fused by means of Multilayer perceptron (MLP) or any other technique to produce better results.

4.3.5 Ensemble systems

Some research efforts utilized ensemble systems to tackle this challenge. In an ensemble system (Hayden 1998), multiple classifiers are used to produce results, which are then combined by various methodologies depending on the ensemble system being used, such as majority voting and stacking. The base-level classifiers can be different classifiers or variations of the same classifier. The idea behind ensemble systems is that with multiple classifiers, even if some classifiers make a mistake, there is still a chance to get correct results. Ensemble systems are built with the hope that they will produce better results than any single classifier present in the system.

5 Conclusions

People express their emotions more frequently and explicitly now-a-days using social networking platforms. They post about their daily life activities, events, achievements, losses, failures, perspectives about social, national or international issues and lots of other things related to their day-to-day impressions. In spite of the use of different audio-visual technologies, people still use text in most of their interactions within their social networks and for their posting on social media forums. For long time, various organizations related to business, psychology, healthcare, politics, security and other fields have been trying to extract emotion of people from their social network interactions. Due to the necessity of detecting the correct emotion from a piece of text in various socio-economical areas, researchers have developed automatic emotion detection systems using different approaches. Researchers are still working on this topic to optimize existing systems and to increase their accuracy. Since it is nearly impossible to identify all variations of text representing all possible human emotions, the current research described in the literature has a wide range of possibilities for improvement.

Humans themselves intentionally or unintentionally perform emotion recognition every time they interact with others. Hints may be picked up to classify emotions based on facial expression, speech, and even text. There has been a lot of research done to categorize emotions by considering these information sources. In this paper, we focused on reviewing research performed on categorizing emotions based on text and speech.

For speech-based emotion recognition, we looked into two main types of databases: simulated and natural. Majority of the databases are simulated to provide a wide range of emotions to work with, but regardless of how well simulated they are, they are not same as natural databases. More research should focus on natural databases to show how well the system will perform in real scenario. There are many features in speech such as pitch, volume, energy, formant. Many of the research paper performed feature ranking to use the important features for categorizing emotions. It might be worth considering finding the most important features for categorizing each emotion separately, rather than finding most important features for the classification of all emotions.

For text-based emotion detection, we have covered some recent works on this topic and some possible areas where improvements are possible and needed. Existing works of different researchers have used novel approaches or hybrids of existing approaches to address text-based emotion detection. They have used only one word or multiple words from the given sentences, they have applied lexicons, implemented machine learning methods, created new datasets for various domains, described emotions with diverse models and tried some unique ideas. Many different methodologies, such as, HMM, SVM, Hybrid systems, Ensemble systems have been used for categorizing emotions. Hybrid systems seem to perform better than systems using just one source of information. Ensemble systems also seem to perform better than individual classifier. We have found that there is a lot of overlap in the methodologies and features used in the research on this topic.

But despite the so many existing works described in the literature, the accuracy of emotion detection from text when the emotion is implicit is not good. Detecting sarcasms correctly still needs huge contributions. Classifying emotions according to their intensities is a comparatively new topic. Detecting multiple emotions from the same text, realizing the cause behind an emotion, detecting the personality of a person, relating emotion with different social and individual parameters, detecting the mood of an individual or public in general, predicting the action of a person based on his/her emotion—these can be some possible issues to focus on in future. We hope that through this review paper, the reader will have a better understanding of the research done on this topic, databases used, features extracted, methodologies used, and the results reported by various researchers.

References

Abak FS, Evrim V (2016) HONET-ICT. IEEE, pp 154–158

Agrawal A, An A (2012) Unsupervised emotion detection from text using semantic and syntactic relations. In: Proceedings of the IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology, pp 346–353

Akaike H (1998) Information theory and an extension of the maximum likelihood principle. In: Selected papers of Hirotugu Akaike. Springer, New York, pp 199–213

Anagnostopoulos C-N, Iliou T, Giannoukos I (2015) Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011. Artif Intell Rev 43(2):155–177

An Y, Sun S, Wang S (2017) Naive Bayes classifiers for music emotion classification based on lyrics. In: IEEE/ACIS international conference on computer and information science (ICIS), pp 635–638

Ayadi ME, Kamel MS, Karray F (2011) Survey on speech emotion recognition: features, classification schemes, and databases. Pattern Recognit 44:572–587

Bell C (1824) Essays on the anatomy and philosophy of expression. John Murray, London

Bellegarda J (2013) Large-scale personal assistant technology deployment: the Siri experience. ISCA INTERSPEECH, pp 2029–2033

Binali H, Wu C, Potdar V (2010) Computational approaches for emotion detection in text. In: IEEE international conference on digital ecosystems and technologies (DEST), pp 172–177

Borod JC (2000) The neuropsychology of emotion. Oxford University Press, Oxford

Broad CD (1954) Emotion and sentiment. J Aesthet Art Crit 13(2):203–214

Busso C, Deng Z, Yildirim S, Bulut M, Lee C, Kazemzadeh A, Lee S, Neumann U, Narayanan S (2004) Analysis of emotion recognition using facial expression, speech and multimodal information. ACM Multimodal Interfaces 6:205–211

Calvo RA, Kim SM (2013) Emotions in text: dimensional and categorical models. Comput Intell 29(3):527–543

Calvo R, Kim S (2013) Emotions in text: dimensional and categorical models. Comput Intell 29(3):527–543

Canales L, Martinez-Barco P (2014) Emotion detection from text: a survey. In: Processing in the 5th Information Systems Research Working Days (JISIC 2014), pp 37–43

Chen LS, Tao H, Huang TS, Miyasato T, Nakatsu R (1998) Emotion recognition from audiovisual information. In: IEEE second workshop on multimedia, signal processing, pp 83–88

Chopade CR (2015) Text based emotion recognition: a survey. Int J Sci Res (IJSR) 4(6):409–414

Cohen I, Garg A, Huang T (2000) Emotion recognition from facial expression using multilevel HMM. In: Neural information processing system

Cowie R, Douglas-Cowie E, Tsapatsoulis N, Votsis G, Kollias S, Fellenz W, Taylor JG (2001) Emotion recognition in human–computer interaction. IEEE Signal Process 18(1):32–80

Daga D, Hudait A, Tripathy HK, Das MN (2016) Automatic emotion detection model from facial expression. In: International conference on advanced communication control and computing technologies, pp 77–85

Darwin C (1998) The expression of the emotions in man and animals. Oxford University Press, Oxford

Desmet B, Hoste V (2013) Emotion detection in suicide notes. Expert Syst Appl 40(16):6351–6358

Devijver PA, Kittler J (1982) Pattern recognition: a statistical approach. Prentice-Hall, London

Dini L, Bittar A (May 2016) Emotion analysis on twitter: the hidden challenge. In: Language Resources and Evaluation Conference (LREC)

Ekman P (1992) An argument for basic emotions. Cogn Emot 6(3–4):169–200

Ghazi D, Inkpen D, Szpakowicz S (2014) Prior and contextual emotion of words in sentential context. Comput Speech Lang 28(1):76–92

Grover S, Verma A (2016) Design for emotion detection of punjabi text using hybrid approach. In: International Conference on Inventive Computation Technologies (ICICT), vol 2, pp 1–6

Gupta U, Chatterjee A, Srikanth R, Agrawal P (2017) A sentiment-and-semantics-based approach for emotion detection in textual conversations. Neu-IR: Workshop on Neural Information Retrieval, SIGIR 2017, ACM, arXiv preprint arXiv:1707.06996

Hajar M (2016) Using youtube comments for text-based emotion recognition. Procedia Comput Sci 83:292–299

Han K, Yu D, Tashev I (2014) Speech emotion recognition using deep neural network and extreme learning machine. In: Interspeech

Hasan M, Rundensteiner E, Agu E (2014) Emotex: detecting emotions in twitter messages. In: 2014 ASE BIGDATA/SOCIALCOM/CYBERSECURITY Conference

Hasan M, Rundensteiner E, Kong X, Agu E (2017) Using social sensing to discover trends in public emotion. In: IEEE 11th International Conference on Semantic Computing (ICSC), pp 172–179

Hayden M (1998) The ensemble system. Cornell University

Hu H, Xu M, Wu W (2007) GMM supervector based SVM with spectral features for speech emotion recognition. In: ICASSP '07, pp 4-413–4-416

Jain VK, Kumar S, Fernandes SL (2017) Extraction of emotions from multilingual text using intelligent text processing and computational linguistics. J Comput Sci 21:316–326

Joshi A, Tripathi V, Soni R, Bhattacharyya P, Carman MJ (2016) Emogram: an open-source time sequence-based emotion tracker and its innovative applications. In: Knowledge extraction from text, AAAI Workshop

Kahou SE, Bouthillier X, Lamblin P, Gulcehre C, Michalski V, Konda K, Jean S, Froumenty P, Dauphin Y, Boulanger-Lewandowski N, Ferrari RC, Mirza M, Warde-Farley D, Courville A, Vincent P, Memisevic R, Pal C, Bengio Y (2016) Emonets: multimodal deep learning approaches for emotion recognition in video. J Multimodal User Interfaces 10(2):99–111
Kanger N, Bathla G (2017) Recognizing emotion in text using neural network and fuzzy logic. Indian J Sci Technol 10(12). https://doi.org/10.17485/ijst/2017/v10i12/100526
Kang X, Ren F, Wu Y (2017) Exploring latent semantic information for textual emotion recognition in blog articles. IEEE/CAA J Autom Sin 5(1):204–216
Kao EC-C, Liu C-C, Yang T-H, Hsieh C-T, Soo V-W (2009) Towards text-based emotion detection a survey and possible improvements. In: IEEE international conference on information management and engineering (ICIME), pp 70–74
Kessous L, Castellano G, Caridakis G (2010) Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis. J Multimodal User Interface 3(1–2):33–48
Kim J, Wagner J, Vogt T, Andre E, Jung F, Rehm M (2009) Emotional sensitivity in human–computer interaction. Inf Technol Methods Appl Inform Inf Technol 51(6):325–328
Kjeldskov J, Graham C (2003) A review of Mobile HCI Research Methods. Springer Human–Computer Interaction with Mobile Devices and Services, Mobile HCI, pp 317–335
Kononenko I (1994) Estimating attributes: analysis and extensions of RELIEF. In: European conference on machine learning. Springer, Berlin, pp 171–182
Kudiri KM, Said AM, Nayan MY (2016) Human emotion detection through speech and facial expressions. In: IEEE international conference on computer and information sciences (ICCOINS), pp 351–356
Kwon O, Chan K, Hao J, Lee T (2003) Emotion recognition by speech signals. In: EUROSPEECH Speech Communication and Technology, p 8
Lalitha S, Madhavan A, Bhushan B, Saketh S (2014) Speech emotion recognition. In: ICAECC, pp 1–4
Lalitha S, Patnaik S, Arvind TH, Madhusudhan V, Tripathi S (2015) Emotion recognition through speech signal for human–computer interaction. In: IEEE fifth international symposium on electronic system design, vol 5, pp 217–218
LeDoux JE (1984) Cognition and emotion. Handbook of cognitive neuroscience. Springer, Berlin, pp 357–368
Lee CM, Narayanan S, Pieraccini R (2001) Recognition of negative emotions from the speech signal. In: IEEE workshop on automatic speech recognition and understanding. ASRU'01, pp 240–243
Li W, Xu H (2014) Text-based emotion classification using emotion cause extraction. Expert Syst Appl 41(4):1742–1749
Lin Y, Wei G (2005) Speech emotion recognition based on HMM and SVM. In: International conference on machine learning and cybernetics, vol 8, pp 4898–4901
Li X, Pang J, Mo B, Rao Y (2016) Hybrid neural networks for social emotion detection over short text. In: International Joint Conference on Neural Networks (IJCNN), pp 537–544
Lovheim H (2012) A new three-dimensional model for emotions and monoamine neurotransmitters. Med Hypotheses 78(2):341–348
Maedche A, Morana S, Schacht S, Werth D, Krumeich J (2016) Advanced user assistance systems. Bus Inf Syst Eng 58(5):367–370
Manser AR (1963) Sketch for a theory of the emotions. Anal Philos 4(1):27–28
Mao X, Zhang B, Luo Y (2007) Speech emotion recognition based on a hybrid of HMM/ANN. In: Proceedings of the 7th conference on 7th WSEAS international conference on applied informatics and communications, vol 7, pp 367–370
Mohammad SM, Bravo-Marquez F (2017) Emotion intensities in tweets. In: Proceedings of the sixth Joint Conference on Lexical and Computational Semantics (*Sem)
Mohammad SM, Kiritchenko S (2015) Using hashtags to capture fine emotion categories from tweets. Comput Intell 31(2):301–326
Munezero MD, Montero CS, Sutinen E, Pajunen J (2014) Are they different? Affect, feeling, emotion, sentiment, and opinion detection in text. IEEE Trans Affect Comput 5(2):101–111
Nahin NH, Alam JM, Mahmud H, Hasan K (2014) Identifying emotion by keystroke dynamics and text pattern analysis. Behav Inf Technol 33(9):987–996
Nahin N, Alam J, Mahmud H, Hasan K (2015) Identifying emotion by keystroke dynamics and text pattern analysis. Behav Inf Technol 33(9):301–326
Neiberg D, Elenius K, Laskowski K (2006) Emotion recognition in spontaneous speech using GMMs. In: Proceedings of Interspeech, pp 809–812
Oatley K, Johnson-Laird PN (1987) Towards a cognitive theory of emotions. Cogn Emot 1(1):29–50
Ortony A, Clore GL, Collins A (1988) The cognitive structure of emotions. Cambridge University Press, Cambridge
Pane J, Myers B, Miller L (2002) Using HCI techniques to design a more usable programming, system. IEEE Human Centric Computing Language and Environment, HCC, pp 198–206
Pan Y, Shen P, Shen L (2012) Speech emotion recognition using support vector machine. Int J Smart Home 6(2):101–108
Perikos I, Hatzilygeroudis I (2016) Recognizing emotions in text using ensemble of classifiers. Eng Appl Artif Intell 51:191–201
Perikos I, Hatzilygeroudis I (2016) Recognizing emotions in text using ensemble of classifiers. Eng Appl Artif Intell 51:191–201
Petrushin VA (2000) Emotion recognition in speech signal: experimental study, development, and application. In: Proceedings of ICSLP, pp 222–225
Plutchik R (1980) Emotion: a psychoevolutionary synthesis. Harper and Row, New York
Poria S, Cambria E, Howard N, Huang GB, Hussain A (2016) Fusing audio, visual and textual clues for sentiment analysis from multimodal content. Neurocomputing 174:50–59
Poria S, Chaturvedi I, Cambria E, Hussain A (2016) Convolutional MKL based multimodal emotion recognition and sentiment analysis. In: IEEE 16th international conference on data mining (ICDM), pp 439–448
Preotiuc-Pietro et al (2016) The valence and arousal facebook posts. In: Proceedings of the 7th workshop on computational approaches to subjectivity, sentiment and social media analysis, San Diego, California, pp 9–15
Rosenblum M, Yacoob Y, Davis L (1994) Human emotion recognition from motion using a radial basis function network architecture. In: IEEE motion of non-rigid and articulated objects, pp 43–49
Russell JA (1980) A circumplex model of affect. J Pers Soc Psychol 39(6):1161–1178
Schmitt JJ, Hartje W, Willmes K (1997) Hemispheric asymmetry in the recognition of emotional attitude conveyed by facial expression, prosody and propositional speech. Cortex 33(1):65–81
Schuller B, Rigoll G, Lang M (2004) Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture. In: ICASSP, vol 1, pp I-577–I-580
Seehapoch T, Wongthanavasu S (2013) Speech emotion recognition using support vector machines. IEEE Knowl Smart Technol 5:86–91
Semwal N, Kumar A, Narayanan S (2017) Automatic speech emotion detection system using multi-domain acoustic feature selection and classification models. In: IEEE international conference on identity, security and behavior analysis (ISBA), pp 1–6

Sen A, Sinha M, Mannarswamy S, Roy S (2017) Multi-task representation learning for enhanced emotion categorization in short text. In: Pacific-Asia conference on knowledge discovery and data mining. Springer, Cham, pp 324–336

Shaver P, Schwartz J, Kirson D, O'connor C (1987) Emotion knowledge: further exploration of a prototype approach. J Pers Soc Psychol 52(6):1061–1086

Shivhare SN, Khethawat S (2012) Emotion detection from text. arXiv:1205.4944 [cs.HC]

Silva LC, Ng PC (2000) Bimodal emotion recognition. In: Fourth IEEE international conference on automatic face and gesture recognition, pp 332–335

Soleymani M, Asghari-Esfeden S, Fu Y, Pantic M (2016) Analysis of eeg signals and facial expressions for continuous emotion detection. IEEE Trans Affect Comput 7(1):17–28

Sreeja PS, Mahalakshmi GS (2017) Emotion models: a review. Int J Control Theory Appl (IJCTA) 10(8):651–657

Stearns S (1976) On selecting features for pattern classifiers. In: 3-d international conference on pattern recognition, Coronado, CA, pp 71–75

Strapparava C, Mihalcea R (2008) Learning to identify emotions in text. In: Proceedings of ACM symposium on applied computing, pp 1556–1560

Summa A, Resch B, GIS Geoinformatics-Z, Strube M (2016) Microblog emotion classification by computing similarity in text, time, and space. In: Proceedings of the workshop on computational modeling of people's opinions, personality, and emotions in social media, pp 153–162

Tiwari SP, Raju MV, Phonsa G, Deepu DK (2016) A novel approach for detecting emotion in text. Indian J Sci Technol 9(29):1–5

Tripathi V, Joshi A, Bhattacharyya P Emotion analysis from text: a survey. http://www.cfilt.iitb.ac.in/resources/surveys/emotion-analysis-survey-2016-vaibhav.pdf

Vlassis N, Likas A (2002) A greedy EM algorithm for Gaussian mixture learning. Neural Process Lett 15(1):77–87

Yadollahi A, Shahraki AG, Zaiane OR (2017) Current state of text sentiment analysis from opinion to emotion mining. ACM Comput Surv (CSUR) 50(2):25:1–25:33

Yat S, Lee M, Chen Y, Huang C-R (2010) A text-driven rule-based system for emotion cause detection. In: Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text. Association for Computational Linguistics, pp 45–53