Received November 28, 2018, accepted December 11, 2018, date of publication January 21, 2019,
date of current version March 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2892624

Emotion Classification and Crowd Source Sensing; A Lexicon Based Approach

RASHID KAMAL[1], MUNAM ALI SHAH[2], CARSTEN MAPLE[3], MOHSIN MASOOD[1], ABDUL WAHID[4], AND AMJAD MEHMOOD[5], (Senior Member, IEEE)

[1]Department of Computing and Technology, Abasyn University, Peshawar 25000, Pakistan
[2]Department of Computer Science, COMSATS University Islamabad, Islamabad 44000, Pakistan
[3]Cyber Security Centre at WMG, The University of Warwick, Coventry CV4 7AL, U.K.
[4]School of Electrical Engineering and Computer Science, National University of Science and Technology, Islamabad 44000, Pakistan
[5]Institute of Computing, Kohat University of Science & Technology, Kohat 26000, Pakistan

Corresponding author: Amjad Mehmood (dramjad.mehmood@ieee.org)

This work was supported by the Alan Turing Institute under EPSRC grant EP/N510129/1 and partially funded through the U.K. Hub for Cyber Security of the Internet of Things, PETRAS, under grant (EP/N02334X/1).

ABSTRACT In today's world, social media provides a valuable platform for conveying expressions, thoughts, point-of-views, and communication between people, from diverse walks of life. There are currently approximately 2.62 billion active users' social networks, and this is expected to exceed 3 billion users by 2021. Social networks used to share ideas and information, allowing interaction across communities, organizations, and so forth. Recent studies have found that the typical individual uses these platforms between 2 and 3 h a day. This creates a vast and rich source of data that can play a critical role in decision-making for companies, political campaigns, and administrative management and welfare. Twitter is one of the important players in the social network arena. Every scale of companies, celebrities, different types of organizations, and leaders use Twitter as an instrument for communicating and engaging with their followers. In this paper, we build upon the idea that Twitter data can be analyzed for crowd source sensing and decision-making. In this paper, a new framework is presented that uses Twitter data and performs crowd source sensing. For the proposed framework, real-time data are obtained and then analyzed for emotion classification using a lexicon-based approach. Previous work has found that weather, understandably, has an impact on mood, and we consider these effects on crowd mood. For the experiments, weather data are collected through an application-programming-interface in R and the impact of weather on human sentiments is analyzed. Visualizations of the data are presented and their usefulness for policy/ decision makers in different applications is discussed.

INDEX TERMS Big data, crowd-sourced sensing, lexicon-based approach, Twitter, social networks.

I. INTRODUCTION

In recent years, social media has influenced every field of life, and has become the main vehicle for individuals to express their views [1]–[4]. Microblogging sites, such as Twitter, provide an opportunity for sharing ideas and interacting with people, communities, groups and organizations [5]. The number of active users attained by microblogging sites already numbers in the billions, as shown in Figure 1, and the total number of distinct users of social networking platforms is predicted to exceed 3 billion by 2021, as shown in figure 2. In this paper, we will use the term microblogging to cover social media and social networks.

The associate editor coordinating the review of this manuscript and approving it for publication was Tariq Ahamed Ahanger.

Microblogging is considered as a popular source of socializing in today's culture. Microblogging provides a platform for the user to share his/her ideas, views, express feelings and communicate with people. In recent work, researchers have performed sentimental analysis, developed crowd-sensing techniques and other interesting analyses from Twitter data, for example [6] and [7].

Organizations are looking for ways to harness the power of big data (BD) to improve their decision making [8], [9]. Public opinion is a critical source of information to guide decision making. Opinion, and its connected concepts, such as sentiment and emotions have been identified as important derivations of the data and as such there is growing interest in sentimental analysis and opinion mining as topics of study.

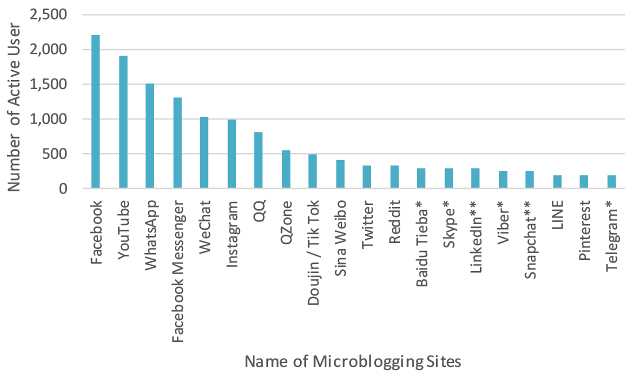

FIGURE 1. Number of users on microblogging sites.

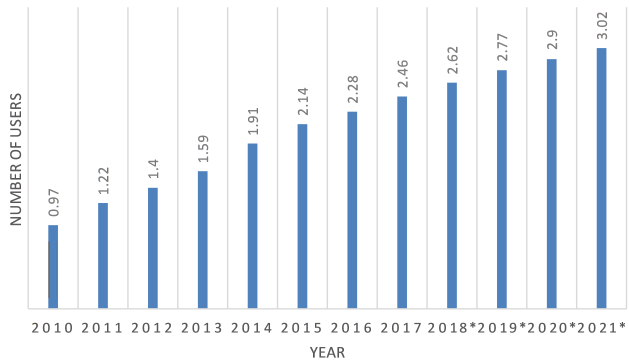

FIGURE 2. Number of social network users worldwide from 2010 to 2021 (in billions).

The importance and prevalence of microblogging has heightened in a variety of areas including business, food, entertainment, welfare and politics, and its reach extending, and connecting communities globally. The largest number of social media users are currently resident in China, India and United States respectively as shown in Figure 3.

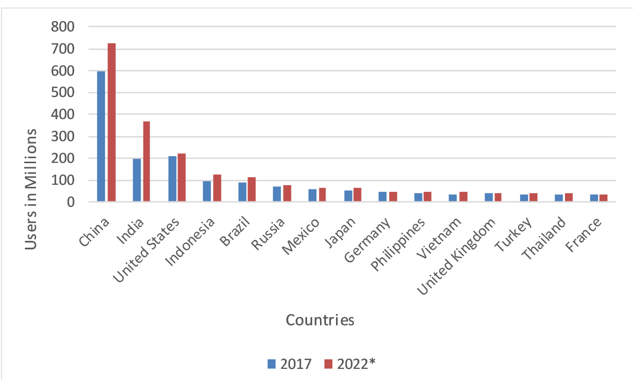

FIGURE 3. Users of different Microblogging sites in different countries 2017 and expected by 2022.

Social media is also becoming more frequently used as a platform for reporting live events [10], [11]. Politicians also use social media for their political events, issues, election campaigns etc. [12]. Crowd-source sensing techniques can be used for analyzing specific crowds, through microblogging, for various purposes [7], [13].

The senses and response messages of a crowd, with regard to specific events, issues, topics, and so on, can be analyzed by using techniques such as sentiment analysis, either in real time, or asynchronously. However, traditional sentiment analysis fails to give a complete perspective about the text beyond whether it is either positive or negative [14], [15]. To gain a deeper understanding of user's views, an emotion mining technique text-data is used.

Emotion mining is a field of getting emotions of the user from given data. In 1980 Robert Plutchik constructed a wheel-like diagram for emotions visualizing based on eight basic emotions. These eight emotion can be seen in Figure 4.

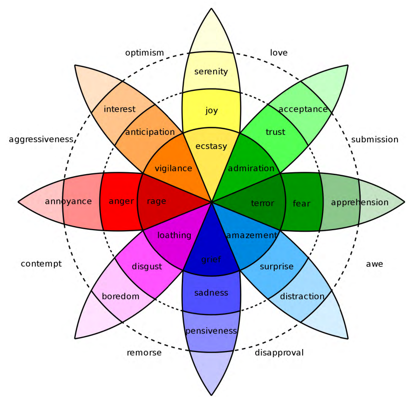

FIGURE 4. In 1980 Robert Plutchik constructed a wheel-like diagram of emotions visualizing eight basic emotions, plus eight derivative emotions each composed of two basic one.

The exponential growth of the field aligns with the growth of social media commentary on the web in the form of reviews, forum discussions, blogs, and microblogs. However, this data source requires processing. Without exception, all the forums of microblogging (such as Twitter) contain so-called noisy data. Spelling, grammar and punctuation mistakes are examples of errors that can cause data in microblogging sites to become noisy.

In text, noisy data presents the biggest hurdle for data analysis. Users of microblogging sites, belongs to various communities throughout the world and express their thoughts with more formal or more casual approaches, depending on a variety of factors. This lack of conformance across individuals and communities creates noisy-data which cannot be easily analyzed using naïve language processing algorithms [16]. Natural Language Processing (NLP) is the branch of Artificial Intelligence (AI) [17] which deals with the understanding and processing of human language and is the foundation for many techniques.

A further challenge in data analysis lies in incomplete sources for the data. For example, using Twitter as microblogging application, users have an option to disable his/her geolocation. Such scenarios generate a challenge for crowd-sourced sensing rending many techniques ineffective.

TABLE 1. An overview of some crowd sensing applications.

Reference	Technique(s)	Data set(s)
Social sensing of urban land use based on analysis of Twitter users'mobility patterns [23]	Filtration (based on geo-location)	Twitter data
Differences between Android and iOS Users of the TrackYourTinnitus Mobile Crowd Sensing mHealth Platform [24]	Diary study	TrackYourTinnitus (app data)
Using on-the-move mining for mobile crowd sensing [26]	Mining on mobile data	Facebook and Twitter data
Crowd-sourced sensing and collaboration using twitter [7]	Diary study	Twitter data

This paper presents a block diagram and pseudo-code for the proposed framework, ''Crowd-Sourced Sensing using Twitter (CSST)''. The suggested CSST framework not only presents crowd-sourced data using a word model technique but also gives emotion classification for the Twitter data. This framework also produces statistical results based on human emotions. Furthermore, the work is implemented in R and simulate results are discussed in later sections.

To enhance the CSST framework, weather data is included. There is significant evidence in the literature that asserts that weather has impact on human emotions. The contributions of the paper can be summarized as:

- A word model technique is implemented for experimentation.
- Classification of emotions in Twitter data using a lexicon-based approach.
- Weather data included in the model to measure impact of weather on crowd sentiment.

The remainder of the paper is organized as follows: Section II presents a literature review of the state of the art in the area. Section III provides the problem statement ahead of Section IV covering a detailed description and design of the proposed framework. Section V illustrates and discusses the results of the proposed framework, before the paper concludes with section VI.

II. LITERATURE REVIEW

A. CROWD-SOURCED SENSING

Mobile crowd-sourced sensing is considered as a new and emerging sensing paradigm. It provides a mechanism for the public to participate in data collection, fusion and generation through their mobile devices [18]. For effective mobile crowd-sourced sensing, a large number of sources are required, and techniques are used which build upon ideas from traditional physical sensors [19]. Modern techniques also utilize images from smart phone users as a source of information. Crowd sensing mechanisms also benefit from the participation of many mobile devices, efficiently collecting data for mass applications such as Careem [20], Uber [21], Food Panda [22] and so forth. Soliman *et al.* [23] used Twitter data to gain information concerning land use in Chicago city. They collected Tweets over North America using the Twitter API from 2013 to 2016. By using geolocation as a parameter, they filtered out all the Tweets except those from Chicago.

In recent work of Pyrss *et al.* [24], Shu *et al.* [25] compared Android and iOS users by using a platform known as TrackYourTinnitus (TYT). TYT is the mobile crowd sensing platform which tracks the individual tinnitus of its registered users; the information is obtained through the use of questionnaires. They used the TYT data set to conduct tests and concluded that there is no difference in sex but in the age between Android and iOS users.

B. REVIEW OF EMOTION THEORIES

Most researchers in computer science have used the terms feelings, mood, emotions and affect, interchangeably. However, in reality those terms do not yield the corresponding meaning. In affective neuroscience the terms are more rigourously defined, for example Fox [27] clarifies the meanings:

- **Feeling:** The personal illustration of emotions, private to the specific experiencing them. Like emotion, it has short span duration.
- **Mood:** A wordy sentimental state that is related to the emotion, usually less concentrated but with long duration.
- **Emotion:** Separate and reliable retorts to inward or outward actions which have a specific consequence of the organism. Emotion has a short term period.
- **Affect:** A surrounding span, to define the themes of emotion, moods, and feeling composed. It frequently has extended term duration.

Traditionally, two emotion theories are dominant in the field of psychology: discrete emotion theory and dimensional emotion theory. Discrete emotion theory states that various emotions flow from distinct neural systems, while the dimensional model says that a shared and unified neurophysiology system is accountable for all effective states.

Although people originate from, and resider in, different countries having different cultures they still have basic emotions regardless of their race, geographical location and languages. Famous psychologists, of both discrete and dimensional schools, have proposed sets of emotions as shown in Table 2.

TABLE 2. Emotion models suggested by different theorists.

Theorist	Year	Emotions	Type
Lovheim	2011	fear, disgust, joy, anger, distress, shame, interest, surprise	dimensional
Shaver	1987	love, joy, anger, sadness, fear, surprise	discrete
Plutchik	1986	anticipation, trust, joy, disgust, sadness, anger, fear, surprise	dimensional
Ekman	1972	joy, fear, sadness, disgust, surprise, anger	discrete

C. EMOTION MINING FROM TEXT

Regarding text data, emotions can be analyzed from either the author's perspective or the reader's view. From the reader's stance, there has, to date, been less research conducted. However, in many applications the reader's view can help to provide significant information about the audience's emotions that can change the overall perspective of the text data [28]–[32].

There are examples where text data can provide sentimental-based data as an output, relying on different emotions. There are two common approaches for the mining of text emotions, one of which is termed lexical-oriented methods. These methods consider a word or collection of words, to make inferences about feelings. Such approaches are often described as keyword-based methods. The other common approach to emotion mining are learning-based techniques, that utilize machine-learning algorithms. These algorithms use training data to build models for the prediction of emotions conveyed in data.

Pang and Lee [33] conducted comprehensive research on sentiment analysis and noteworthy improvements have since been made, see [34], [35]. While these works are concerned with sentiment analysis, the fundamental techniques can be applied to the area of emotion mining from texts.

In [36], a study has been conducted using 40 undergraduate male and female students, to analyze emotions text data, according to the state of mind. The findings showed that students used 5 times more positive words and accentuation marks, regardless of the gender differences. These outcomes are aligned with Social Information Processing (SIP) theory, introduced by Walther [37], Ullah *et al.* [38], and Mehmood *et al.* [39]. The theory states that to send relational based information in computer systems, people use verbal clues to convey their communication rather than non-verbal clues. Jain and Kulkarni [40] provide a brief survey on emotion mining, introducing models from Information Retrieval (IR). They propose an architecture based upon attaining a bag of words and then applying a vector space method.

He and Xia [41] proposed a Joint Binary Neural Network for multi-labeled data. They argue that their model can perform better than classic neural network models. Sahay *et al.* [42] propose a Recurrent Neural Network (RNN) model for application in the video domain and gathers emotions on the data. They generate audio and text from video in addition to emotion analysis, and the performance results in an F score of 0.61.

D. WEATHER IMPACT ON MOOD

According to the literature, see [47] and [48] for example, weather has a decisive impact on the mood of individuals. The study of Ettema *et al.* [47] examines weather-based travel and weather effects on an individual's mood. The author analyzes 562-time sample data from 363 random people in three different Swedish cities. During experiments, people are asked to use smart phones in their homes and after arriving at destinations. The data is analyzed with weather data and the experimental results reveal that there is a significant impact of weather on each person's mood.

Barrington-Leigh and Behzadnejad [49] in his study, analyzed Canadian people's mood according to the weather in different parts of the country. The results obtained show that the sample of Canadian people have a noteworthy change in mood according to weather condition experienced in their area of residence.

Extraordinary research has been done by Baylis regarding the relationship between weather conditions and the spirit of human emotions. For their work, they analyze three and half billion social media posts from millions of people on Facebook and Twitter between 2009 and 2016. The results indicate that warm temperatures, hot temperatures, humidity, and moisture cloud cover are connected with emotional illness [48].

E. WEATHER IMPACT ON BUSINESS

Weather can been a major factor on business trends and weather data can be used to inform business direction for the future. Tian *et al.* [50] show how the weather, including amount of sunlight, temperature, and air quality, is a parameter that can influence the search for different types of customers in their purchases. Using public weather data and supermarket panel data for five general retail products, they conducted their research on 3 hypothesis: H1: consumers buy more when there is less sunlight; H2: more hot and cold temperatures leads to more consumers buying products; and H3: Consumers propensity to purchase a product depends on air quality. Their final analysis shows that weather conditions have a large impact on the mood of consumers.

In a recent study by Dong and Tremblay [51], the authors conclude that weather has the largest impact on the global trading index. They estimate the profitability of global index trading across 49 countries. They claimed that weather can impact 20.5 % annual profit as opposed to the mean of 3.75% during 1993-2012. Their findings included that weather has

TABLE 3. An overview of emotion mining.

Reference(s)	Technique(s)	Data set(s)
Joint Binary Neural Network for Multi-label Learning with Applications to Emotion Classification [41]	Joint binary neural network	Ren-CECps
Multi modal Relational Tensor Network for Sentiment and Emotion Classification [42]	Recurrent neural network and tensor fusion network	CMU-MOSEI
Lexical and Learning-based Emotion Mining from Text [43]	Support vector machine and lexical based	CBET
Intentional Learning to Efficiently Build up Automatically Annotated Emotion Corpora [44]	Word vector and Support vector machine	NRC
Lexicon based Feature Extraction for Emotion Text Classification [45]	Lexicon based	Twitter, SemEval 2007, ISEAR and blogs
Ensemble learning on visual and textual data for social image emotion classification [46]	Ensemble learning	Web data

TABLE 4. An overview of the impact of weather on mood.

Reference(s)	Methodology(s)	Impact on Mood(s)
Season and weather effects on travel-related mood and travel satisfaction [47]	Analyzed 562 time-samples which offered by 363 people	Significant impact on mood
The impact of daily weather conditions on life satisfaction: Evidence from cross-sectional and panel data [49]	Survey-based	Minor impact on mood
Weather impacts expressed sentiment [48]	Social media posts (2009 to 2016)	Significant impact on mood

a large impact on mood, and investors can enhance profits by utilizing weather data.

III. PROBLEM STATEMENT

The existing crowd-sourced sensing techniques based upon social network data have numerous deficiencies including;

- Not all Tweets provide geographical coordinates of users or posts, causing difficulty in detecting user geolocation [52];
- Existing text classification techniques such as [53] and [54] take into account informal words and communication formats on social media, which is noise in the data and can lead towards low accuracy and imprecision; and
- While there is sufficient evidence in the literature that weather influences the mood of the individual (see [47], [49], [55]), limited work has been undertaken analyzing crowd-sourced data from social media, and as such this is an area that requires consideration.

The work in this paper presents a framework utilizing Twitter data to address these problems. Experiments has been conducted using the proposed framework along with simulated results in section V.

IV. METHODOLOGY:

The CSST framework proposed in this paper is illustrated in Figure 5. Further details of the framework are described as follows:

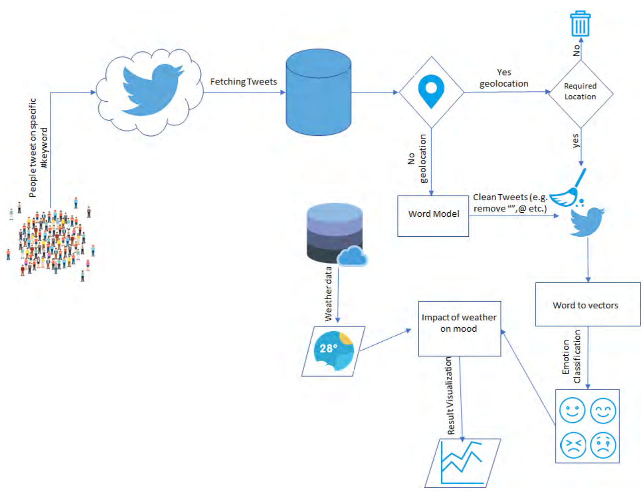

FIGURE 5. Proposed framework.

A. REAL-TIME TWEETS

To fetch Tweets in real time TwitterR has been used in R which provides access to the Twitter API [56]. The application provides a variety of functionality, and allows parameters such as how many Tweets required, the time interval for the Tweets being posted and geolocation of Tweets to be prescribed. Geolocation can be very helpful in crowd-sourced sensing, but as discussed in previous sections not all users enable the geolocations function in Twitter.

B. CROWD-SOURCED SENSING AND THE WORD MODEL

As described earlier crowd-sourced sensing from Twitter data is non-trivial for a number of reasons, including the fact that not all users share their geolocation information. To solve this problem a word model technique is included in the proposed framework. The pseudo-code of the word model technique is given in algorithm 1. According to algorithm 1 if a user has disabled the geolocation function, this algorithm helps to identify the user's location (location can be city, name of venue and country) from the Tweet.

Algorithm 1 Word Model

Require: R - a raw Tweet obtained from Twitter data
Require: L - a Location (where available)
Require: K - the required city
Require: P - the required place
Ensure: TB - filtered Tweets from the desired location

```
for all i to n do
  if R[i] == L then
    TB ← R
  end if
end for
for all i to n do
  if R[i] == K then
    TB ← R
    if R[i] == P then
      TB ← R
    else
      Discard
    end if
  end if
end for
```

Algorithm 2 takes raw Tweets as an input which are fetched from Twitter. If the geolocation of a Tweet is available, the Tweet goes directly into the cleaning step. In all other cases, the proposed algorithm obtains the location information from within the Twitter text. It is not unusual for a user to mention the location name, such as a city name, a famous monument etc. within the Twitter text. Consider the case in which it is required to obtain crowd-source sensing of a baseball match and the analyst wishes to determine the emotions of the fans of each team. In this case, the **Stadium Name, Team Names** and **City Name** are all keywords for a specific location.

C. PREPROCESSING OF DATA

Since Tweets are written by the general public there is a possibility, indeed a reasonable probability, that a large number of casual words, abbreviations and short forms, special characters and spelling mistakes, are present in their Tweets. These add to the noise in the input data, which will be used for the classification in both lexicon-based and learning-based approaches. For reasons previously mentioned, it is important to clean Twitter data before classification. The steps of our preprocessing method are given below:

Algorithm 2 Lexicon-Based Emotion Classification

Require: D - a word from the lexicon
Require: I - a pointer function
Require: S - a sample set
Require: W_j - a number of times the word W occurs in the text
Ensure: T - emotion classification in Tweet

```
for all i to n do
  if D[i] == S[i] then
    I ← 1
  else
    I ← 0
  end if
end for
for all i to N do
  I * W_j
end for
for all i to N do
  ∑_{s∈S} T * I
end for
```

- All special characters are removed.
- Http links are removed.
- Stopping words, useless punctuation marks, numbers etc. are removed.
- All capital letters are changed into small letters.

D. EMOTION CLASSIFICATION: LEXICON-BASED

In this section, a practice is adopted that automatically infers emotions from the Tweets as given in algorithm 2. This problem can be formulated mathematically as: $E = e1, e2, \ldots, ek$, where k is number of labels. A sample set, $S = s1, s2, \ldots, sn$, is used to predict emotions and attach test samples. The substantial work in the framework is emotion classification from the text, using a lexicon-based method. A simple combination back process is used to obtain emotional words within the text. In other methods, one or more external resources can be utilized. Several methods use the structure of the lexicon, which collects data about emotions or at a minimum level expresses words or phrases. Using these lexicon-based approaches, the content of a message, and the feelings and emotions in these words can be identified. Classification can then be made on the basis of the information derived.

For this purpose, a single letter from the words V of all single words (words) in V alone This word expresses Wj emotions. Weight the $T(ei|wj)$ label is counted as the number of times that Wj has occurred in the text that is:

$$T(e_i|w_j) = \sum_{s \in S} T(e_i|S) * I_s(W_j) \tag{1}$$

where T $(e_i|s)$ emotions present in Tweet given sample s and I_s is a pointer function which is just to 1 if x $\in$ s and 0 otherwise.

E. REAL TIME WEATHER DATA

To capture real-time weather data an open weather map API has been used [57] in R. This function fetches real-time weather data for the desired location, and this is then used to determine the impact of weather on the mood in that location.

V. RESULTS

A. MOVIE REVIEWS USING TWEET DATA

To test the framework we have presented, experiments were conducted that involved collecting Twitter data and applying the framework. Different experiments were performed, and these are explained in this results and analysis section. As a first experiment, a scenario is considered in which viewers'responses to the Hollywood movie ''A Star Is Born'' through analysis of their Tweets. The film is a 2018 American musical romantic drama film produced and directed by Bradley Cooper, and is a remake of a 1937 film of the same name, it stars Cooper, Lady Gaga and follows a hard-drinking musician (Cooper), who falls in love with a young singer (Gaga). It marks the fourth remake of the original 1937 film, after the 1954 musical, the 1976 rock musical and the 2013 Bollywood romance film.

For the given trial, 1000 Tweets are gathered using an implementation of the framework. After analyzing the Twitter data, results are obtained about the movie, and these are presented in Figure 6. Figure 6, shows the results of the analysis and demonstrates that the emotions of joy, trust and anticipation are most prevalent. In addition, the graph also shows some other emotions that are present such as surprise, sadness, disgust, fear and anger. These emotions give an insight into audience perception of the movie. For data validation, the records on rotten-tomatoes [58] and IMDB [59], [60] shows that the movie was a hit and have ratings of 8.3/ 10. This shows that the Tweets collected and analyzed using the framework concur with other data sources, and offers an indication of reliability for application in future consideration of other films.

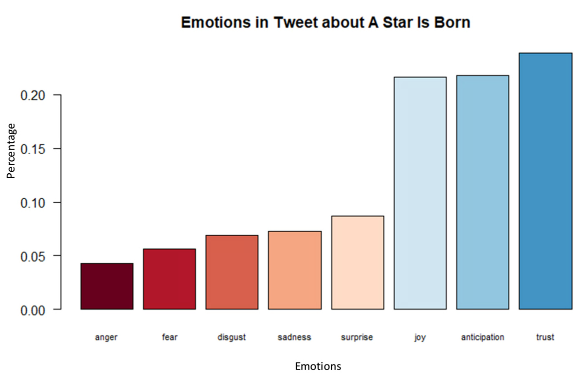

FIGURE 6. Emotions in tweets about a star is born.

While examining the Twitter data, the Figure 7 is obtained, which shows the frequency of words used in the Tweets. From Figure 7, it can be easily seen that the most common word/phrase used in the Tweets is ''Lady Gaga''. Other dominant words in the Tweets are Love, Good, great, Shallow, Oscar, Finally and Best. This shows, that people mostly discuss about Lady Gaga along with positive words, and the overall film response is captured in Figure 7.

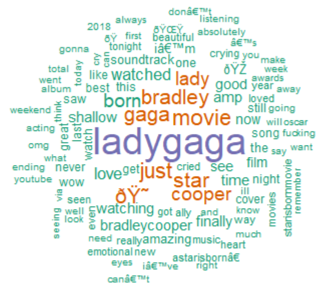

FIGURE 7. Word cloud a star is born.

Figure 8 shows the results of analysis of the 1000 Tweets about the movie *A Star Is Born*. The upper spikes show positive emotions while lower spikes in negative emotions in a tweet. Since the emotions were positive about the movie, it is expected that the data presented in Figure 8 has a large number of positive spikes.

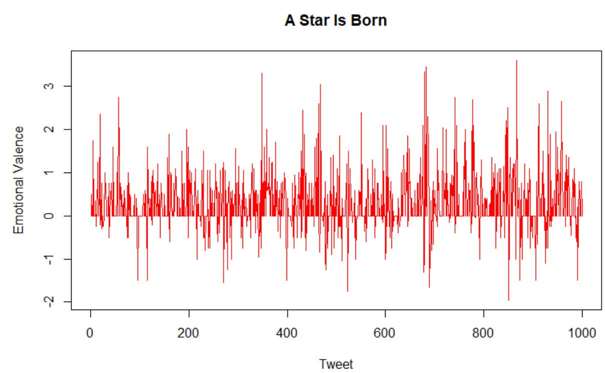

FIGURE 8. Emotions valence over all tweets for a star is born.

Figure 9 shows the results of the analysis of the 1000 Tweets, considered over time. From Figure 9 it can be seen that Tweets have positive emotions at 200, and then become negative until 450 Tweets. Nevertheless, the Tweets again got emotionally positive from 500, and reaches to its peak at 780, as shown in the Figure 9. then again Tweets are

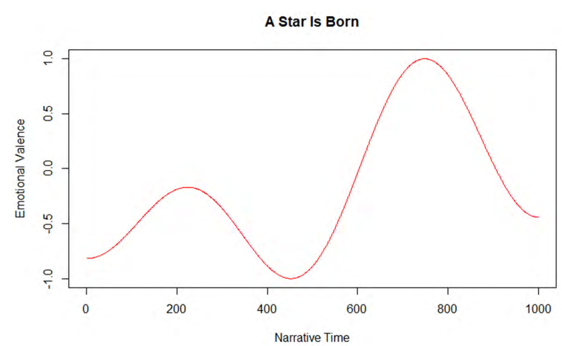

FIGURE 9. Overall emotions pattern in tweets for a star is born.

identified as very positive and emotion valence peaks. In the above figure most Tweets were very positive which illustrates the positive emotions about the movie.

The above figures show that more than 85% of emotions presented in the Tweets about *A Star Is Born* are positive. Emotion Valence is used in psychology while discussing emotions. This means the goodness (positive valence) or badness (negative valence) of an event, object, or situation. This term also characterizes and categorizes specific emotions. The findings and results of the proposed research are further validated with Rotten Tomatoes and IMDB data.

B. PERSON POPULARITY TESTING WITH TWITTER DATA

The proposed framework is further tested using Twitter data in an experiment that examines the popularity of a famous person. Mr. Sen. **Sherrod Brown**, a democrat senator from Ohio, is selected as a case study for these experiments. One thousand Tweets about Senator Brown were gather for analysis and to create a popularity graph. The Tweets were taken from Ohio state only and were processed using the word model as explain in algorithm 1. Figure 10 shows that people in the state of Ohio have emotions that are mainly concerning **Trust** in their senator, as well as having anticipation and joyful feelings. On the other hand, very small number of Tweets were found in the data regarding senator that express feelings of anger, sadness, fear, surprise and disgust. The data implies that Sen. Sherrod Brown has a highly positive popularity graph in the Ohio Community. Sen. Brown was recently re-elected in the 2018 midterm election garnering a total 2,286,730 votes out of 4,298,562 votes [61], which aligns with the data presented in Figure 10.

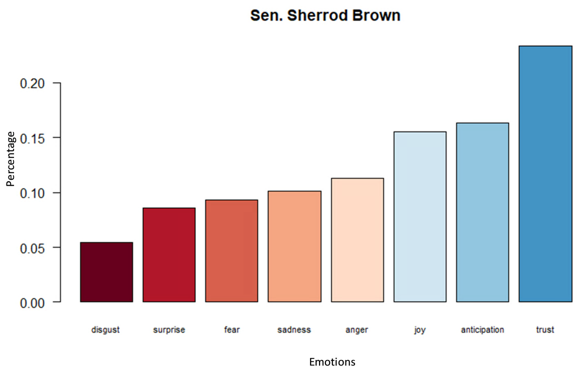

FIGURE 10. Overall emotions in tweets about sen. sherron brown.

Figure 11 illustrates the word cloud formed from anaylsis of 1000 Tweets (posted in Ohio) about Ohio Sen. Sherrod Brown. The word cloud shows the words which are most frequent in the Tweets. The most used word/phrases are presented in the center and can be seen to be *Chuck Todd* and *meet the press*. The popularity of these terms is due to an interview given by Sen. Sherrod Brown given to NBC TV channel. The other dominant words are thanks, love, 2020, good and Ohio. Mostly of the words are positive, which results in the positive emotions. However, there are small number of negative words too, and it is these that results in the detection of the emotions of anger and sadness seen in Figure 11.

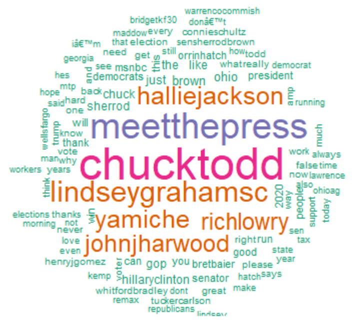

FIGURE 11. Word cloud formed by considering ohio-based tweets about sen. sherrod brown.

In Figure 12 each Tweet emotion valance can be seen for the 1000 Tweets about Sen. Sherrod Brown. Once again, the upper spikes show positive emotions, while lower spikes represent negative emotions in a Tweet.

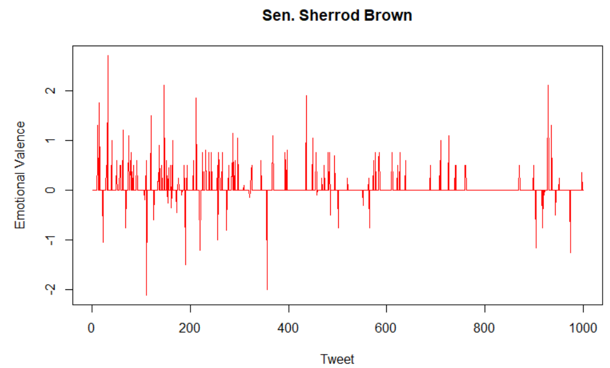

FIGURE 12. Emotion valence over all tweets about sen. sherrod brown.

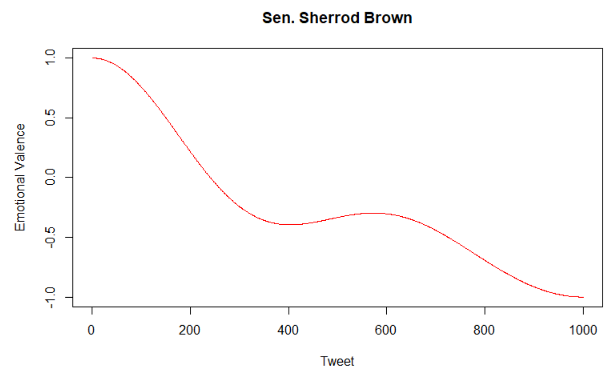

FIGURE 13. Overall emotion pattern in tweets about sen. Sherrod Brown.

At start there are very positive emotions, nonetheless there are some Tweets that have negative emotions such as anger, sadness and disgust; the same emotions can be seen in Figure 13.

C. CAPTURING EMOTIONS DATA ACCORDING TO WEATHER

An experiment is implemented to investigate human emotions with the support of Twitter data. This data that has been gathered has been analyzed and then classified into keywords representing **bad weather** and **good weather**. The weather impact on human emotions is then observed, from the data, and is organized in the form of results as shown in Figure 14 and Figure 15. In good weather, people express through their Tweets, emotions of anticipation, joy, trust and surprise, and Figure 14 clearly depicts that these emotions are prevalent. Conversely with bad weather, Tweets demonstrate emotions of anger, sadness and fear as illustrated in Figure 15, which shows the negative emotions.

For further scrutiny of the Twitter data, a word cloud is used to explore the impact of Good and Bad Weather on public sentiment. Again, the word cloud displays words which are most frequently found in the Tweets. The most frequent words used in the Tweets posted during good weather are beautiful, good, free, perfect and sunny, as shown in Figure 16. This aligns with the overall positive emotions present in the Tweets. On the other hand, during bad weather, the words that are most commonly used in the Tweets are shame, bad and cancel, as given the world cloud Figure 17, demonstrating the negative public thoughts.

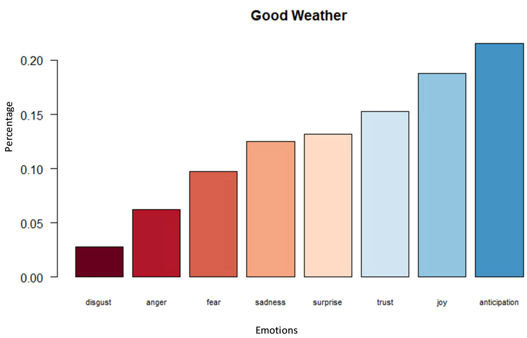

FIGURE 14. Emotions about good weather.

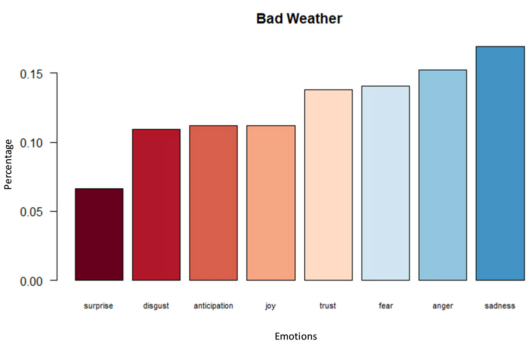

FIGURE 15. Emotions about bad weather.

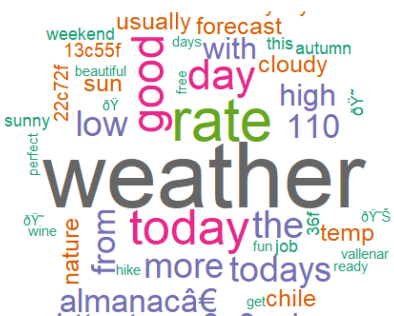

FIGURE 16. Word cloud of tweets posted during good weather.

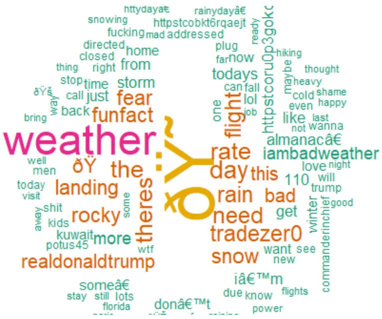

FIGURE 17. Word Cloud of tweets posted during bad weather.

VI. CONCLUSION

In this paper, a new framework is proposed that can be used for crowd-sourced sensing and to classify emotions in Tweets posted by the crowd. The sentiment of the Tweets that people posts varies among users, and in different environments and as a reaction to varying events and issues. In the experimental work we have conducted, Twitter data is captured and then analyzed using the proposed framework. Three different scenarios are considered for the experiments including Hollywood movie-based Tweets, public opinion about a political personality and weather impact on people emotions. For the aforementioned scenarios, Twitter data was collected and analyzed. The results obtained from the movie reviews revealed that the movie under consideration, *A Star is Born*, has Tweets represent a positive audience response. These outcomes are further corroborated with ratings obtained from the Rotten Tomatoes and IMDB websites. Secondly, the case of public opinion regarding a political personality implied that senator **Sherrod Brown** is a popular personality in his home state of Ohio, USA. His recent success in the midterm elections in his state reflect this popularity among his people. The third experiment was executed to investigate the impact the weather has on public feelings. The results obtained demonstrate that in **good weather**, people have positive emotions, while in **bad weather** they display negative emotions in their Tweets, as can be seen in the related word cloud. Using the presented framework provides a method to understand the emotions being conveyed in micro = blogging and this analysis can help to provide quick and easy insights into public opinions in a number of situations, and for a number of purposes.

REFERENCES

[1] M. N. Hajli, "A study of the impact of social media on consumers," *Int. J. Market Res.*, vol. 56, no. 3, pp. 387–404, 2014.

[2] A. Mehmood, J. L. Mauri, M. Noman, and H. Song, "Improvement of the wireless sensor network lifetime using leach with vice-cluster head," *Ad Hoc Sensor Wireless Netw.*, vol. 28, nos. 1–2, pp. 1–17, 2015.

[3] G. S. O'Keeffe and K. Clarke-Pearson, "The impact of social media on children, adolescents, and families," *Pediatrics*, vol. 127, no. 4, pp. 800–804, 2011.

[4] I. E. Erdoğmuş and M. Çiçek, "The impact of social media marketing on brand loyalty," *Procedia-Social Behav. Sci.*, vol. 58, pp. 1353–1360, Oct. 2012.
[5] J. Fotis, D. Buhalis, and N. Rossides, "Social media impact on holiday travel planning: The case of the Russian and the FSU markets," *Int. J. Online Marketing*, vol. 1, no. 4, pp. 1–19, 2011.
[6] *Monthly Active Twitter Users*, Statista, Hamburg, Germany, 2018.
[7] M. Demirbas, M. A. Bayir, C. G. Akcora, Y. S. Yilmaz, and H. Ferhatosmanoglu, "Crowd-sourced sensing and collaboration using twitter," in *Proc. IEEE Int. Symp. World Wireless, Mobile Multimedia Netw. (WoWMoM)*, Jun. 2010, pp. 1–9.
[8] M. Janssen, H. van der Voort, and A. Wahyudi, "Factors influencing big data decision-making quality," *J. Bus. Res.*, vol. 70, pp. 338–345, Jan. 2017.
[9] A. Mehmood, J. Lloret, and S. Sendra, "A secure and low-energy zone-based wireless sensor networks routing protocol for pollution monitoring," *Wireless Commun. Mobile Comput.*, vol. 16, no. 17, pp. 2869–2883, 2016.
[10] S. Petrović, M. Osborne, and V. Lavrenko, "Streaming first story detection with application to twitter," in *Proc. Hum. Lang. Technol., Annu. Conf. North Amer. Chapter Assoc. Comput. Linguistics*, Los Angeles, CA, USA: Association for Computational Linguistics, 2010, pp. 181–189.
[11] A. Mehmood, A. Khanan, M. M. Umar, S. Abdullah, K. A. Z. Ariffin, and H. Song, "Secure knowledge and cluster-based intrusion detection mechanism for smart wireless sensor networks," *IEEE Access*, vol. 6, pp. 5688–5694, 2018.
[12] D. Halpern, S. Valenzuela, and J. E. Katz, "We face, I tweet: How different social media influence political participation through collective and internal efficacy," *J. Comput.-Mediated Commun.*, vol. 22, no. 6, pp. 320–336, 2017.
[13] M. M. Umar, A. Mehmood, H. Song, and K.-K. R. Choo, "I-Marks: An iris code embedding system for ownership identification of multimedia content," *Comput. Elect. Eng.*, vol. 63, pp. 209–219, Oct. 2017.
[14] A. Yadollahi, A. G. Shahraki, and O. R. Zaiane, "Current state of text sentiment analysis from opinion to emotion mining," *ACM Comput. Surv.*, vol. 50, no. 2, p. 25, 2017.
[15] A. Mehmood, M. Nouman, M. M. Umar, and H. Song, "ESBL: An energy-efficient scheme by balancing load in group based WSNs," *KSII Trans. Internet Inf. Syst.*, vol. 10, no. 10, pp. 4883–4901, 2016.
[16] W. Li and H. Xu, "Text-based emotion classification using emotion cause extraction," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1742–1749, 2014.
[17] G. G. Chowdhury, "Natural language processing," *Annu. Rev. Inf. Sci. Technol.*, vol. 37, no. 1, pp. 51–89, 2005. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/aris.1440370103
[18] B. Guo, Z. Yu, X. Zhou, and D. Zhang, "From participatory sensing to mobile crowd sensing," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PERCOM WORKSHOPS)*, Mar. 2014, pp. 593–598.
[19] H. Ma, D. Zhao, and P. Yuan, "Opportunities in mobile crowd sensing," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 29–35, Aug. 2014.
[20] *Careem*. Accessed: Nov. 23, 2018. [Online]. Available: https://www.careem.com/
[21] *Uber*. Accessed: Nov. 23, 2018. [Online]. Available: https://www.uber.com/
[22] R. Wenzeld. *Food Panda*. Accessed: Nov. 15, 2018. [Online]. Available: https://www.foodora.com/
[23] A. Soliman, K. Soltani, J. Yin, A. Padmanabhan, and S. Wang, "Social sensing of urban land use based on analysis of Twitter users' mobility patterns," *PloS one*, vol. 12, no. 7, 2017, Art. no. e0181657.
[24] R. Pryss, M. Reichert, W. Schlee, M. Spiliopoulou, B. Langguth, and T. Probst, "Differences between android and iOS users of the TrackYourTinnitus mobile crowdsensing mHealth platform," in *Proc. IEEE 31st Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jun. 2018, pp. 411–416.
[25] L. Shu, Y. Chen, Z. Huo, N. Bergmann, and L. Wang, "When mobile crowd sensing meets traditional industry," *IEEE Access*, vol. 5, pp. 15300–15307, 2017.
[26] W. Sherchan, P. P. Jayaraman, S. Krishnaswamy, A. Zaslavsky, S. Loke, and A. Sinha, "Using on-the-move mining for mobile crowdsensing," in *Proc. IEEE 13th Int. Conf. Mobile Data Manage. (MDM)*, Jul. 2012, pp. 115–124.
[27] E. Fox, *Emotion Science: Cognitive and Neuroscientific Approaches to Understanding Human Emotions*. Basingstoke, U.K.: Palgrave Macmillan, 2008.
[28] Y. Rao, Q. Li, L. Wenyin, Q. Wu, and X. Quan, "Affective topic model for social emotion detection," *Neural Netw.*, vol. 58, pp. 29–37, 2014.
[29] Z. Zhang, A. Mehmood, L. Shu, Z. Huo, Y. Zhang, and M. Mukherjee, "A survey on fault diagnosis in wireless sensor networks," *IEEE Access*, vol. 6, pp. 11349–11364, 2018.
[30] A. Mehmood, M. Mukherjee, S. H. Ahmed, H. Song, and K. M. Malik, "NBC-MAIDS: Naïve Bayesian classification technique in multi-agent system-enriched IDS for securing IoT against DDoS attacks," *J. Supercomput.*, vol. 7, no. 10, pp. 5156–5170, 2018.
[31] G. Mishne and MDe Rijke, "Capturing global mood levels using blog posts," in *AAAI Spring Symp., Comput. Approaches Analyzing Weblogs*, vol. 6, 2006, pp. 145–152.
[32] J. Lei, Y. Rao, Q. Li, X. Quan, and L. Wenyin, "Towards building a social emotion detection system for online news," *Future Gener. Comput. Syst.*, vol. 37, pp. 438–448, Jul. 2014.
[33] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, nos. 1–2, pp. 1–135, 2008.
[34] B. Liu, "Sentiment analysis and opinion mining," *Synth. Lectures Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.
[35] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," *Pattern Recognit.*, vol. 44, no. 3, pp. 572–587, 2011.
[36] K. Gao, H. Xu, and J. Wang, "Emotion cause detection for chinese micro-blogs based on ECOCC model," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*. Chi Minh, Vietnam: Springer, 2015, pp. 3–14.
[37] J. B. Walther, "Interpersonal effects in computer-mediated interaction: A relational perspective," *Commun. Res.*, vol. 19, no. 1, pp. 52–90, 1992.
[38] I. Ullah, M. A. Shah, A. Wahid, A. Mehmood, and H. Song, "ESOT: A new privacy model for preserving location privacy in Internet of Things," *Telecommun. Syst.*, vol. 67, no. 4, pp. 553–575, 2018.
[39] A. Mehmood, M. M. Umar, and H. Song, "ICMDS: Secure inter-cluster multiple-key distribution scheme for wireless sensor networks," *Ad Hoc Netw.*, vol. 55, pp. 97–106, Feb. 2016.
[40] M. C. Jain and V. Y. Kulkarni, "TexEmo: Conveying emotion from text-the study," *Int. J. Comput. Appl.*, vol. 86, no. 4, pp. 1–8, 2014.
[41] H. He and R. Xia. (2018). "Joint binary neural network for multi-label learning with applications to emotion classification." [Online]. Available: https://arxiv.org/abs/1802.00891
[42] S. Sahay, S. H. Kumar, R. Xia, J. Huang, and L. Nachman. (2018). "Multi-modal relational tensor network for sentiment and emotion classification." [Online]. Available: https://arxiv.org/abs/1806.02923
[43] A. G. Shahraki and O. R. Zaiane, "Lexical and learning-based emotion mining from text," in *Proc. Int. Conf. Comput. Linguistics Intell. Text Process.*, 2017, pp. 1–14.
[44] L. Canales, C. Strapparava, E. Boldrini, and P. Martínez-Barco, "Intensional learning to efficiently build up automatically annotated emotion corpora," *IEEE Trans. Affect. Comput.*, to be published. doi: 10.1109/TAFFC.2017.2764470.
[45] A. Bandhakavi, N. Wiratunga, D. Padmanabhan, and S. Massie, "Lexicon based feature extraction for emotion text classification," *Pattern Recognit. Lett.*, vol. 93, pp. 133–142, Jul. 2017.
[46] S. Corchs, E. Fersini, and F. Gasparini, "Ensemble learning on visual and textual data for social image emotion classification," in *International Journal of Machine Learning and Cybernetics*. Springer, 2017, pp. 1–14.
[47] D. Ettema, M. Friman, L. E. Olsson, and T. Gärling, "Season and weather effects on travel-related mood and travel satisfaction," *Frontiers Psychol.*, vol. 8, p. 140, Jan. 2017.
[48] P. Baylis *et al.*, "Weather impacts expressed sentiment," *PLoS one*, vol. 13, no. 4, 2018, Art. no. e0195750.
[49] C. Barrington-Leigh and F. Behzadnejad, "The impact of daily weather conditions on life satisfaction: Evidence from cross-sectional and panel data," *J. Economic Psychol.*, vol. 59, pp. 145–163, Apr. 2017.
[50] J. Tian, Y. Zhang, and C. Zhang, "Predicting consumer variety-seeking through weather data analytics," *Electron. Commerce Res. Appl.*, vol. 28, pp. 194–207, Mar./Apr. 2018.
[51] M. Dong and A. Tremblay, *Are Weather-Based Trading Strategies Profitable?*, SSRN, 2018. [Online]. Available: https://ssrn.com/abstract=3111467
[52] E. Williams, J. Gray, and B. Dixon, "Improving geolocation of social media posts," *Pervasive Mobile Comput.*, vol. 36, pp. 68–79, Apr. 2017.
[53] S. Chawla and M. Mehrotra, "A conceptual model for understanding the impact of textual emotion mining.," *Int. J. Adv. Res. Comput. Sci.*, vol. 9, no. 1, pp. 860–863, 2018.
[54] F. Krebs, B. Lubascher, T. Moers, P. Schaap, and G. Spanakis, "Social emotion mining techniques for Facebook posts reaction prediction," in *Proc. 10th Int. Conf. Agents Artif. Intell. (ICAART)*, Funchal, Portugal: SciTePress, vol. 2, 2018, pp. 211–220.
[55] M. M. Umar, A. Mehmood, and H. Song, "SeCRoP: Secure cluster head centered multi-hop routing protocol for mobile ad hoc networks," *Secur. Commun. Netw.*, vol. 9, no. 16, pp. 3378–3387, 2016.
[56] streamR. *streamR Package in R*. Accessed: Oct. 27, 2018. [Online]. Available: https://cran.r-project.org/web/packages/streamR/index.html

[57] E. E. LTD. *OpenWeatherMap API* Accessed: Oct. 27, 2018, [Online]. Available: https://openweathermap.org/api
[58] R. Tomatoes. *A Star is Born*. Accessed: Nov. 15, 2018, [Online]. Available: https://www.rottentomatoes.com/m/a_star_is_born_2018/
[59] IMDB. "*A Star is Born*. Accessed: Nov. 15, 2018. [Online]. Available: https://www.imdb.com/title/tt1517451/
[60] A. Mehmood, Z. Lv, J. Lloret, and M. M. Umar, "ELDC: An artificial neural network based energy-efficient and robust routing scheme for pollution monitoring in WSNs," *IEEE Trans. Emerg. Topics Comput.*, to be published. doi: 10.1109/TETC.2017.2671847.
[61] NYTimes. *Ohio Election Results*. Accessed: Nov. 23, 2018. [Online]. Available: https://www.nytimes.com/interactive/2018/11/06/us/elections/results

RASHID KAMAL received the B.S. degree in computer science from Iqra University, Islamabad Campus, in 2016. He is currently pursuing the M.S. degree with COMSATS University. He is currently a Research Associate with Abasyn University, Peshawar Campus. His area of specialization is artificial intelligence (AI). His research interests include big data, algorithms, AI, and natural language processing. He has done several projects in AI, natural language processing, and big data.

MUNAM ALI SHAH received the B.Sc. and M.Sc. degrees in computer science from the University of Peshawar, Pakistan, in 2001 and 2003, respectively, the M.S. degree in security technologies and applications from the University of Surrey, U.K., in 2010, and the Ph.D. degree from the University of Bedfordshire, U.K., in 2013.

Since 2004, he has been an Assistant Professor with the Department of Computer Science, COMSATS Institute of Information Technology, Islamabad, Pakistan, where he also leads the Performance Evaluation and Enhancements of Computing Systems Research Group. He is supervising three Ph.D. students and six master's students. Nine master's students already finished their dissertation under his supervision. He is a HEC Approved Supervisor. He has authored more than 110 research articles published in national and international conferences and journals. His research interests include MAC protocol design, the Internet of Things, and security issues in wireless communication systems.

Dr. Shah received the Best Paper Award from the International Conference on Automation and Computing, in 2012.

CARSTEN MAPLE is currently a Professor of cyber systems engineering with the WMG's Cyber Security Centre. He is the Director of Research of cyber security with organizations in key sectors such as manufacturing, healthcare, financial services, and the broader public sector to address the challenges presented by today's global cyber environment. He has an international research reputation and extensive experience of institutional strategy development and interacting with external agencies. He has published over 200 peer-reviewed papers. He has co-authored the U.K. Security Breach Investigations Report 2010, supported by the Serious Organized Crime Agency and the Police Central e-Crime Unit. He has co-authored *Cyberstalking* (U.K.), a report supported by the Crown Prosecution Service and Network for Surviving Stalking. His research has attracted millions of pounds in funding and has been widely reported through the media.

He has given evidence to government committees on issues of anonymity, child safety online. He is with various departments such as the Association of Chief Police Officers, the College of Policing, Interpol, the Equality and Human Rights Commission, and the Department of Business Innovation and Skills. He has advised executive and non-executive directors of public sector organizations and multibillion pound private organizations.

MOHSIN MASOOD received the M.Sc. degree in mobile communication from the University of East London, U.K., and the Ph.D. degree in artificial intelligence (AI)-based metaheuristic techniques for the optimization in MPLS/GMPLS networks from the University of Strathclyde, Glasgow, U.K. He is currently an Assistant Professor with Abasyn University, Pakistan. He is involved in industrial- and academic-based projects. His research interests include AI, machine learning, artificial neural networks, metaheuristic optimization techniques, computer, optical, wireless and wired networks (IP, MPLS/GMPLS networks) optimization by investigating/proposing and working on genetic, evolutionary, bio-inspired, swarm, modified/adaptive, and hybrid models-based optimization techniques. He received research awards including Erasmus funding from the Technical University of Ostrava, Czech Republic.

ABDUL WAHID received the Ph.D. degree from Kyungpook National University, South Korea. He is currently serving as an Associate Professor at School of Electrical Engineering and Computer Science, National University of Science and Technology, Islamabad, Pakistan. His research interests include but are not limited to vehicular ad-hoc networks, wireless sensor networks, underwater wireless sensor networks, cyber physical systems, software-defined networking, and information-centric networking.

AMJAD MEHMOOD (SM'16) received the Ph.D. degree in wireless networks from the Kohat University of Science and Technology, Kohat, in 2014, and the Ph.D. degree from the University of Virginia, USA. He is currently pursuing the Ph.D. degree with the Guangdong Provincial Key Laboratory on Petrochemical Equipment Fault Diagnosis, Guangdong University of Petrochemical Technology, Maoming, China. In 2003, he joined the Kohat University of Science and Technology, where he is currently a Professor and the Coordinator of M.S./Ph.D. Program. He is the KUST University Branch Coordinator. He is also the Director of the Reliable Optimal Smart Scalable and Efficient Lab, KUST. He has visited many countries for presenting his research work like Australia, Hong Kong, China, South Korea, Malaysia, and Thailand. He has published more than 60 academic articles in peer-reviewed international journals and conferences around the world. His research areas include cyber physical systems, the Internet of Things, connected vehicles, wireless communications and networking, optical communications and networking, smart grid communications and networking, security issues in wireless networks, big data, cloud computing, and fault diagnosis in WSNs. He has supervised many BCS, MCS, M.S., and Ph.D. students in the above-mentioned research areas. He is also been remained the part of reviewing and organizing different workshops, seminar, and training sessions on different technologies. He has served as a TPC Member, as a Reviewer, and as the Demo Chair for numerous international conferences, including CCNC, SCPA, WICOM, INFOCOM, and SCAN. He is a Reviewer or an Associate Editor for many peer-reviewed international journals. As far as his social experience is concerned, he is a Professional Member of ACM.