

# A Study of Sentiment and Trend Analysis Techniques for Social Media Content

**Asad Mehmood**

Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Islamabad, Pakistan  
Email: [asadmehmood16@hotmail.com](mailto:asadmehmood16@hotmail.com)

**Abdul S. Palli and M.N.A. Khan**

Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Islamabad, Pakistan  
Email: [abdulsattarpalli@gmail.com](mailto:abdulsattarpalli@gmail.com), [mna2010@gmail.com](mailto:mna2010@gmail.com)

**Abstract**—The social media networks have evolved rapidly and people frequently use these services to communicate with others and express themselves by sharing opinions, views, ideas etc. on different topics. The social media trend analysis is generally carried out by sifting the corresponding or interlinked events discussed on social media websites such as Twitter, Facebook etc. The fundamental objective behind such analyses is to determine the level of criticality with respect to criticism or appreciation described in the comments, tweets or blogs. The trend analysis techniques can also be systematically exploited for opinion making among the masses at large. The results of such analyses show how people think, assess, orate and opine about different issues. This paper primarily focuses on the trend detection and sentiment analysis techniques and their efficacy in the contextual information. We further discuss these techniques which are used to analyze the sentiments expressed within a particular sentence, paragraph or document etc. The analysis based on sentiments can pave way for automatic trend analysis, topic recognition and opinion mining etc. Furthermore, we can fairly estimate the degree of positivity and negativity of the opinions and sentiments based on the content obtained from a particular social media.

**Index Terms**—Trend Analysis, Sentiment Analysis, Social Media Analysis, Semantic Web, Opinion Mining.

## I. INTRODUCTION

Social media has evolved into a vibrant platform where people communicate freely with each other, share ideas and comment on various events and issues. Twitter is one of the evolving social media which, on average, hosts around 200 million tweets per day. Tweets generally correspond to infinite number of products, services, social issues, news, incidents and reviews etc. Further, people also comment and share views about tweets pertaining to various topics/issues. Twitter share these tweets in a unique way, which cannot be out rightly used to judge the essence of the tweets and topics being discussed on social media. Facebook supports five different types of post

status, video, link, image and music; and this information is very useful for trend analysis.

There is a verity of events discussed on social media (such as Twitter, Facebook) which can be grouped into two categories: planned events and unplanned events. Planned events include general elections in a country, music concert, educational or employment workshops, sports tournament etc. and unplanned events pertain to unanticipated, out of the blue and sudden incident such as earthquakes, hurricanes, bomb blasts, spot-fixing etc. Sakaki *et al.* [1] divide these events in two categories – *social events* such as large parties, sports events, exhibitions, accidents and political campaigns, and *natural events* such as storms, tornadoes and earthquakes.

Twitter is one of the evolving social media which, on average, hosts around 200 million tweets per day. It is an evolving social media platform that acts as communication source where the real-time information from users becomes available instantly. People express their views/comments about any event of interest and share the latest information about the particular event/incident. This can be quite useful for creating awareness and getting solution to the problem as well as ascertaining general public's trend on that issue. Tweets can correspond to innumerable products, services, social issues, news, incidents and reviews etc. Further, people also comment and share views about tweets pertaining to various topics/issues. Different organizations rely on such information to analyze and evaluate the customers' and consumers' views about their products or services. Twitter share tweets in a unique way, which cannot out rightly be used to judge the essence of the tweets and topics being discussed on social media. For example, TV channels use tweets as a major source of feedback about their programs and talk shows. Although a single tweet consists of maximum 140 characters, but this information becomes reasonably huge due to numerous tweets being made by thousands of users on a single issue in a shorter span of time. Eventually, analyzing such a large amount of data which is totally in textual form becomes a colossal task. According to Pohl *et al.* [2], social media platform can be used to manage crisis by (i) sharing useful information about a particular event or disaster before actually it happened so that people get awareness,

(ii) to control effects of the event if it had happened, and  
 (iii) get out from the disastrous situation.

Facebook is one the most popular social media network that started evolving in 2004. People use Facebook for several purposes — such as posting reviews about products/services on dedicated pages; vote on different kinds of polls launched by users etc. Facebook offers users three types of actions (“like”, “comments” and “share”) against a post to express themselves. The number of “likes” for a post shows how many people viewed and liked it as a positive gesture whereas some user comment to express likes and dislikes about the post. Some people use posts frequently to update their status to share their mood, thoughts, activities, criticism, likes, dislike about any related event or situation which is associated with them. This medium not only serves as a source of expressing for large number of users, but has also become a business for some people.

The fad of microblogging is becoming popular among the Internet users. Microblogging is broadcast medium and is an underlying form of blogs. Microblog also known as micropost, has smaller size and is quite different from the traditional blogs. The microblogs are usually a miniature form of the actual content like short sentences, images or video links [3, 4]. Millions of people across the world use this medium to express their views on different events and daily-life routine matter. Though this trend has emerged recently yet a number of approaches related to sentiment analysis of microblogs have been devised and explored.

Social media such as Facebook, Twitter etc. is one of the fast evolving phenomena for sentiment analysis to know how people think about a particular event. In the present day technology driven culture, we can get opinions from different polls and advertisements placed on blogs and social media sources. In general, human beings have natural instinct to share information or give feedback about the product or items they purchase in their daily lives. And this very trait of sharing information has now moved on to social media sites like Twitter, Facebook, LinkedIn and other microblogging sites. By this means, these sources are becoming very useful in identifying and analyzing diverse opinions on different topics and areas. Twitter is one of the important sources for getting opinion from microblogging data available in different languages. Such data can be obtained through the Twitter’s “Tweet Entities” using various applications.

Two types of analyses, trend analysis and sentimental analysis, can be highly beneficial to determine how people think and get emotional on certain social, religious or political issues. One reason for trend analysis can be to detect an emergent or suspicious behavior happening on the social media platform. For example, trend analysis can be used to see how certain groups of people are using it to launch their propaganda or forging facts about certain political or religious issues. Corporate sector can also use it to get feedback about their products.

Social media trend can be broadly categorized into four types namely positive, negative, neutral and uninterested. The level of trend can further be classified into low,

medium, high which will mainly be linked to the critical comments/tweets. Sentiments are evaluated and extracted from the social media content, which can either be in positive or negative attitude. The positive attitude of a person can be conceived as being happy or pleased with the content expressed by someone on certain issue. On the other hand, negative expressions can pertain to being unhappy or angry with the content posted by someone on certain issue.

Trend analysis in a traditional sense can be defined as the frequently mentioned topics throughout the stream of user activity [5]. Hence, for generating an effective trend out of the social media content, the need for an automated classifier becomes necessary to reduce the time for analyzing the large amount of data and improving efficiency of the analysis process.

Sentiment analysis basically tries to judge different aspects of natural language which help people to find valuable information from large amount of unstructured data [6]. It is an emerging concept in which different human emotions are determined from textual content. It enables us to extract opinions and sentimental feelings of the people. To know people’s opinion about a particular event and its future impact (commonly termed as social media trends), there is a need for an automated system that can analyze such a huge amount of data and produce desired results with certain level of accuracy so that such results can be made acceptable by the masses. On Internet, people use blog posts and forums for promoting products or services as well as discussing any topic and expressing their views. The sentiment analysis on this platform possesses very important information for security analysts to keep an eye on the activities of miscreants and terrorists etc. However, it becomes a serious challenge to perform such types of analysis on a big data. Numerous events take place regularly in our daily life, therefore, it is not possible to manually analyze every event and predict its future impact. It is really hard for the computing machines to automatically extract the meaning and tone of content as people express so many things in many different ways and styles etc. Sentiment analyses can prove very useful when we analyze search engine results, different blogs, social networks, web forums, different review of people on books, movies, sport and products etc [7]. This can help reduce the efforts required to go through large amount of documents to generate an opinion about the nature of the contents.

## II. LITERATURE REVIEW

Topic identification on social media helps understand what is being discussed and it also helps users to grab the broader picture without reading all the available information. Using network topology in trend deduction method can distinguish “viral topics” (topic which spreads via peer influence) from the shared information topics (spreads via news media). In the context of social media, Twitter is ranked as the second most famous social network [8]. The tweets made on it can be used to

predict the future impact of different events/issues by applying data mining techniques. The authors designed a tool that performs trend analysis for social celebrities to find most influential among them as well as trend analysis of national and international issues, and recent events. The system proposed by Lin *et al.* [8] has two layers: data processing layer (for data collection and applying data mining techniques for performing trend analysis) and information display layer (for representing or visualization the results). After collecting the required data which mostly pertains to message properties, it uses Term Frequency- Inverse Document Frequency (TF-IDF) and fixed keywords to analyze the tweets. The mined results are presented to the user in four sections: top news section, trending topics section, active users section and top sources section.

Huberman *et al.* [9] used tweets to predict revenue generated for a newly released movie. The proposed approach entails selecting and analyzing tweets, before and after the release of a movie, that match the specific keywords taken from title of the movie. Tumasjan *et al.* [10] suggest that Twitter can be used as a platform for political discussion, and tweets can be used to detect election results. In this regard, the party with highest rate of tweets in its favor has fair chance of winning the elections. Wegrzyn-Wolska *et al.* [11] designed a sentimental analysis system based on tweets for French Presidential Election held in 2012. The aim of study was to correlate what is being discussed at Twitter-sphere. By using REST API, the authors collected specific tweets matching the user supplied keywords. The system detects trend by calculating both the frequency of a searched keyword in the dataset and its sentiments on the basis of positive, negative or neutral comments made against a post.

Asur *et al.* [12] argue that the topic being discussed quite often on twitter during certain timeframe becomes the trend. Re-tweets on the same topic from multiple users such as news from various media sources can also be the reason of setting trend of the topic. If multiple topics being discussed on Twitter among the different user groups are divided according to the region, then it results into multiple trends in tandem. The frequently discussed topic becomes the principal trend in the trends list. Likewise, Asur *et al.* [13] compared SinaWeibo (Chinese social media) with Twitter and found differences in trends. On SinaWeibo, users share jokes, images, videos, and re-tweet most of the time. But on twitter people mostly amplify the news which they obtain from other media sources.

Sakaki *et al.* [1] propose earthquake detection and reporting system which sends email alerts to the registered users when it gets any tweet originated from Japan only about an earthquake. In the proposed system, all the Twitter users are considered as a sensor because they send sensory information. After every second, the system searches tweets which match the given keywords and applies semantic analysis to get accurate results. To know the location or area where that event has occurred, the system uses event detection algorithm. The proposed

algorithm uses search API to get the time and location of a tweet and the same is automatically attached with the tweet when it is posted via iPhone or phone that has GPS system. The other alternate it uses for finding event location is to get the registered location of the user through Kalman filter or particle filtering algorithms.

Achrekar *et al.* [14] designed a system to predict the ratio of flu disease in USA for a specific time period using tweets. For data collection, the authors developed a crawler using the Twitter real time search API to retrieve tweets which match the keywords and to collect patient details such as name, age, location etc. from their profiles in order to identify the affected area and the number of expected patients in that area.

In order to evaluate large volume of Twitter data, Hao *et al.* [15] introduced three techniques which were mainly based on visual sentiment analysis. The suggested technique include topic-based analysis that involves natural language processing to determine nature of the topic of discussion by extracting different opinion-related attributes to measure the degree of sentiments. Then a stream analysis is performed on large number of tweets to extract information of interest based on positive/negative attitude and the influencing characteristics that it possesses within the larger density of tweets. Pixel cell-based sentiment calendars and high density geomaps are used to visualize and depict large number of tweets in a single view.

Since Twitter data is one of the important sources of microblogging platform, so it has been used for sentiment analysis and Pak *et al.* [16] used this corpus for sentiment analysis and opinion mining. The corpus containing emotions like happy smiley “:)” or sad smiley “:(” are readily evaluated as positive or negative sentiments respectively. After corpus collection, it is analyzed to check how data has been distributed into subjective corpus (containing positive or negative set) and objective corpus (containing neutral set). The authors calculated the presence of n-gram for extracting binary feature and keyword frequency was used to obtain rest of the general information. The analytics reported by the authors showed that objective sets contained more common and proper nouns which in turn have often used personal pronouns. Similarly, the objective sets bloggers addressed themselves as third person while the subjective sets bloggers described themselves as first person or second person. Lima *et al.* [17] propose a classifier for Twitter messages that comprises three modules: Support Counting, Database Selection and Classification modules. Support counting module counts percentage of the tweets that contain at least one word or emotion in the tweet. Database Selection module divides data into two sets: training set and testing set. Classification module classifies data using Naïve Bayes algorithm.

Cvijikj and Michahelles [18] figured out discovering trends on Facebook using features of shared posts. They categorized topic of interest into three groups: descriptive events, popular events and daily routines. For this purpose, Graph API search feature is used to find the posts against the specific keyword after every 10 minutes.

Trend deduction is a two pronged strategy which involves topic identification and cluster detection. Only the 'status' attribute used with Facebook posts was used to identify trend. The methodology uses TF-IDF for assigning weights to the terms on the basis of two quanta: frequency of occurrence of a term within a single document and the number of documents in the corpus which contain the given term. Finally, the trend analysis was done using LingPipe API. Li *et al.* [19] applied clustering-based sentiment analysis approach on dataset which was based on the review of a movie. The dataset was divided into two primary clusters corresponding to the positive and negative user comments and remarks about that movie. For experimental purpose, the technique of TF-IDF (Term Frequency – Inverse Document Frequency) was applied. The stability of results was increased when the voting mechanism was used, and finally a symbolic technique was applied to enhance the clustering results. Symbolic technique, as described by Li *et al.* [19], is the technique in which each term (or word) is assigned some value based on the negative and positive connotation as well as the intensity/criticality of that term (or word). Then, an aggregation function is used to obtain cumulative word scores to draw conclusions about the overall negativity or positivity of the comments.

Goorha *et al.* [20] used entity extraction system for extracting more relevant and most occurred phrases from tweets, blogs and newspaper articles to find out people's opinion about a product or company. The proposed system used cluster streaming to identify related terms and used TF-IDF for assigning and calculating weights of terms which were useful for making cluster. Pohl *et al.* [2] proposed a system to detect sub-events related to a crisis situation based on YouTube and Flickr data such as picture or video. Video or picture data item has two parts: the coordinates and the terms. For extraction of the coordinates, it uses two-phase clustering approach which is based on longitude and latitude; and for the terms extraction, it uses textual metadata fields of a particular data item and calculates TF/IDF for clustering. To present results visually, it uses OpenStreetMap which is capable to display cluster data and its location.

A proximity-based sentiment analysis is proposed by Hasan *et al.* [7] that uses features based on word proximities within a sentence. The authors used three proximity-based features which are called proximity distribution, mutual information between proximity types, and proximity patterns. The dataset was divided into number of segments in which each segment contained over 100 words. The distance between positive and negative pair of word was calculated. Three proximity-based features were used. In *Proximity Distributions*, different numbers of bin were considered which returned the distribution of pair-wise distances from the proximity models. In *Mutual Information between Proximity Types*, the relationship between the proximity types was used to determine the polarity of the document. Then, the theoretic quantities of entropy for each sequence, was used to get mutual information between pairs of

proximity types. In *Proximity Patterns*, it described polarity of words which were used in the document.

Corley *et al.* [21] proposed a technique for finding Flu cases discussed over blogs and finding the relationship among the outcome of blog posts and data reported by Center for Disease Control and Prevention. Python script with combination of pyMPI was used for data extraction. The pyMPI is software that integrates the Message Passing Interface into the Python interpreter". Suzumura *et al.* [22] developed a system for processing large amount of data on-the-fly by using web services including Twitter Search Service and Twitter Streaming API, and then displayed the tweets on Google Maps using Google Maps API and Ajax components.

Karamibekr *et al.* [23] proposed an approach to classify the sentiments using verbs from the sentence. The approach uses the verbs as core opinion terms in social domains on the pretext that verb is considered as the main opinion term. The authors used lexical knowledge to extract the sentiment terms present in the sentence. With the help of bootstrapping process, authors collect the list of opinion verbs using an English dictionary and its synonyms in the WordNet [24]. A *subject* in the sentence describes action of the doer or what the predicate does, whereas the *object* in sentence is what or whom the verb is acting. Cai *et al.* [25] proposed an approach to create an effective sentiment based taxonomy that employs statistical based approach for the sentiment analysis. The sentiments expressed by the words, are measured on the scale of positive or negative. Then, the list of positive and negative words is created by using two external NLP resources. In order to score the relative sentiment between the posts which have the positive and negative words, the author characterizes the degree of positive/negative sentiment that each word conveys. Koncz *et al.* [26] proposed a feature selection approach that used frequencies of terms in a particular document. The values are normalized in terms of total number of documents in category.

Mizumoto *et al.* [27] proposed the system in which the author created polarity dictionary to determine the sentimental polarities of stock market. While constructing the dictionary, the author has used the semi-supervised learning approach from which small polarity dictionary is made. Using the co-occurrence frequency with words in polarity dictionary, those new words are added to the dictionary whose polarities are unknown. For estimating the polarity of the text, the author has used sentiment analysis method. The polarity of article is determined according to the frequency of words in the polarity dictionary; hence, the articles are determined as positive, negative or neutral. In [28], the authors introduced text sentiment classification for the contextual information. For this purpose, the flow of the sentiments and keywords in the paragraph were taken out from the contextual information. Finally, by computing the contextual information degree (linearly combined weighted sum of contextual information), the overall sentiment of sentence was classified.

Iqbal *et al.* [30, 31] proposed performance metrics for

software design and software project management. Process improvement methodologies are elaborated in [32, 43] and Khan et al. [33] carried out quality assurance assessment. Amir et al. [34] discussed agile software development processes. Rehman et al. [35] and Khan et al. [44] analyzed issues pertaining to requirement engineering processes. Umar and Khan [36, 37] analyzed non-functional requirements for software maintainability. Khan et al. [38, 39] proposed a machine learning approaches for post-event timeline reconstruction. Khan [40] suggests that Bayesian techniques are more promising than other conventional machine learning techniques for timeline reconstruction. Rafique and Khan [41] explored various methods, practices and tools being used for static and live digital forensics. In [42], Bashir and Khan discuss triaging methodologies being used for live digital forensic analysis.

### III. CRITICAL ANALYSIS

Budak [5] Used Independent Trend Formation Model (ITFM) and nearest neighbor model to identify structural trends and compared them with traditional trends by analyzing tweets. However, it does not defined structural trends in generalized form so one cannot identify gap among the discussed trend types. Lin et al. [8] used natural language processing, semantic analysis, and TF-IDF to Analyze tweets of celebrities for finding trend on Twitter. Same approach can also be used to analyze tweets of common people or public to figure out public trend on national and international issues and events. Cvijikj [18] used linguistic analysis, TF-IDF, clustering by distribution, and clustering by co-occurrence to identify the topic of discussion by analyzing the 'status' posts over Facebook. Nevertheless, selecting only the 'status' posts does not provide such a dataset to analyze trends. One should collect other types also such as posted image or audio/video and users comments on that post.

Asur et al. [9] used linguistic analysis and sentiment analysis on tweets which matches the given keywords to analyze and predicts the revenue of a movie. The author selects tweets by matching the keywords that are present in only the title of a movie. User may include director, producer, actor/actress, or character titles in their tweets, but there is a possibility that that those tweet will be left over.

Tumasjan et al. [10] used linguistic analysis and sentiment analysis techniques to predict the election results by using twitter as a platform for political discussion as well as a data source for finding people's opinion about a particular political party. Dataset contains only those tweets which contain names of political parties and names of some well-known politicians. However, the tweets which miss the parties name can also be very useful for this analysis, since people may use polling symbols instead of a party name or people may use slogans of the parties in their tweets. Furthermore, the selected tweets only come from the same group of users instead of variety of other social media content. Asur et al.

[12] analyzed twitter data to know the reasons which are important for different stages of trend by keywords matching. There should be some mechanism for matching misspelled words, and sentimental analysis of tweets for better results. Yu et al. [13] observed trends over Chinese social media and compared with Twitter trends. Research is more focused towards analyzing new tweets and retweets over SinaWeibo (Chinese social media). Trends can also be found through location as well. Sakaki et al. [1] used Kalman filter (a particle filtering algorithms) for semantic analysis of users tweets that belongs to Japan and performed semantic analysis to generate email alerts about real time event such as earthquake. There is a very less chance of getting GPS data (location of tweet) from tweet because every user could not have i-Phone to tweet. User may be able to tweet from any location other than the registered location.

Achrekar et al. [14] Used auto-regression with exogenous inputs (ARX) model to design a system to predict the ratio of flu disease in USA using tweets of a specific time period. Pohl et al. [20] used streaming clustering algorithm and TF-IDF to design a system which tells the public opinion about any product or company by extracting more relevant and most occurred phrases from tweets, blogs and newspaper's articles, and displays result in a user interface by plotting the clusters. It decides popularity of an entity (product or company) based on the number of tweets which discuss that particular entity. There is no any mechanism used for semantic analysis to decide the positivity and negativity of tweets about the entity. There is a great chance that tweets might be about the closing down of a company or product. Pohl et al. [2] proposed a system to detect sub-events in crisis situation over YouTube and Flickr data such as picture or video using two-phase clustering and TF-IDF approach. However, the approach lacks mechanism for filtering the duplication of data. Suzumura et al. [22] designed the system architecture for processing the large amount of data on the fly rather than store and process using probabilistic models (such as Temporal Model, Spatial Model). The authors developed a system for predicting real time events based on their designed architecture. But, such techniques used for finding location are unable to find the specific location of the posted messages. Corley et al. [21] proposed a technique for finding Flu cases discussed over blogs and to find relationship among the outcome of blog posts and data reported by Center for Disease Control and Prevention. It filters only English language words from blog and compares them with the given keywords. No procedure was defined for matching any misspelled or informal short words.

In [15], the authors primarily focused on visual opinion analysis and sentiment analysis. The technique visualized large number of Tweets to a single view that depicts the overall sentiments. This approach does not help identify opinion association. The major focus area of research in [16] was sentiment analysis in which the proposed technique automatically collected corpus to train a sentiment classifier. Syntactic structures are useful in

describing emotions in the Twitter content. The approach needs to develop multilingual version of such technique or to include an auto-translator with the proposed technique. Li and Liu [19] focused on sentiment analysis and clustering. The cluster based technique was applied which had more accurate results as compared to techniques that involve human intervention. The dataset used for evaluation was small which would otherwise have produced more accurate results. Sentiment Analysis was major focus in the research and technique proposed by Lima and deCastro [17]. The proposed technique automatically trained the classifier based on Naïve Bayes algorithm to categorize datasets. However, the combination of different proposed techniques could have produced accurate results while using the neural classification.

The techniques proposed in [28] are based on sentiment analysis and security informatics which estimate sentiments that were expressed in content. It does not require labeling and collecting the document. It is a semi-supervised sentiment estimation technique which lacks multi-lingual sentiment analysis as well as the criteria/procedure that indicates when to smooth the polarity estimates. The major area of focus of the techniques proposed by Hasan and Adjeroh [7] is based on sentiment analysis and text mining. The technique used proximity-based sentiment analysis. The approach depends on the polarity of dictionary that was created from the corpus. Again, the major area of focus in [26] was Sentiment Analysis. The approach used the feature selection in comparison with Information Gain (IG) feature selection. The approach showed slightly poor results than the Information Gain (IG).

Mizumoto et al. [27] used semi-supervised learning for sentiment analysis. The technique created the polarity dictionary to estimate the polarity of stock market contextual information. This technique does not deal with the negation and adversative conjunction. The major research focus in [23] was on sentiment analysis and opinion structure. The technique used verb oriented sentiment classification approach for social domains. The technique did not use main verbs of the sentences to classify the sentiments related to sentence. Cai et al. [25] carried out sentiment analysis through Topic Detection. The technique for sentiment analysis used sentiment classification and sentiment detection scheme. The technique does not use part-of-speech and word syntactic relationships doing sentiment analysis and topic detection.

#### IV. CONCLUSION AND FUTURE WORK

In this study, we looked into different techniques used for trend analysis on social media such as Facebook and Twitter. We also studied that how data which is available on social media can be used in different ways to analyze and predict future trends. We observed that there is not any flexible system that has data dictionary with more appropriate keywords for predicting trends over Facebook. As most of the work is done using Twitter as a source, we have plan to design and develop a system to

predict future trends over Facebook by collecting all type of public posts and related user comments and applying sentiment analysis. We also intend to propose a model that is capable to analyze different events being discussed on social media along with detecting trends to make future prediction about the outcome of such events and issues. Since, most of the content available on Facebook and Twitter are unstructured text, therefore, we have planned to develop the system which can automatically analyze sentiments from the available content and verify the opinions that are expressed in the contextual information.

#### REFERENCES

- [1] Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: real-time event detection by social sensors. In Proceedings of the 19th international conference on World Wide Web (pp. 851-860). ACM.
- [2] Pohl, D., Bouchachia, A., & Hellwagner, H. (2012). Automatic Identification of Crisis-Related Sub-Events using Clustering. In Machine Learning and Applications (ICMLA), 2012 11th International Conference on (Vol. 2, pp. 333-338). IEEE.
- [3] Kaplan, A. M., & Haenlein, M. (2011). The early bird catches the news: Nine things you should know about micro-blogging. *Business Horizons*, 54(2), 105-113.
- [4] Lohmann, S., Burch, M., Schmauder, H., & Weiskopf, D. (2012, May). Visual analysis of microblog content using time-varying co-occurrence highlighting in tag clouds. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (pp. 753-756). ACM.
- [5] Budak, C., Agrawal, D., & El Abbadi, A. (2011). Structural trend analysis for online social networks. *Proceedings of the VLDB Endowment*, 4(10), 646-656.
- [6] Bloom, K., Garg, N., & Argamon, S. (2007). Extracting appraisal expressions. *HLT-NAACL 2007*, 308-315.
- [7] Hasan, S. S., & Adjeroh, D. A. (2011). Proximity-based sentiment analysis. In *Applications of Digital Information and Web Technologies (ICADIWT), 2011 Fourth International Conference on the* (pp. 106-111). IEEE.
- [8] Lin, Y. C., Yang, P. C., Hsieh, W. T., & Seng-cho, T. C. Technology Trend Analysis Tool using Twitter as a Source.
- [9] Asur, S., & Huberman, B. A. (2010). Predicting the future with social media. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on* (Vol. 1, pp. 492-499). IEEE.
- [10] Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proceedings of the fourth international AAAI conference on weblogs and social media* (pp. 178-185).
- [11] Wegrzyn-Wolska, K., & Bougueroua, L. (2012). Tweets mining for French Presidential Election. In *Computational Aspects of Social Networks (CASoN), 2012 Fourth International Conference on* (pp. 138-143). IEEE.
- [12] Asur, S., Huberman, B. A., Szabo, G., & Wang, C. (2011). Trends in social media: Persistence and decay. In *5th International AAAI Conference on Weblogs and Social Media*.
- [13] Yu, L., Asur, S., & Huberman, B. A. (2011). What trends in chinese social media. arXiv preprint arXiv:1107.3522.
- [14] Achrekar, H., Gandhe, A., Lazarus, R., Yu, S. H., & Liu, B. (2011). Predicting flu trends using twitter data. In *Computer Communications Workshops (INFOCOM*

- WKSHPs), 2011 IEEE Conference on (pp. 702-707). IEEE.
- [15] Hao, M., Rohrdantz, C., Janetzko, H., Dayal, U., Keim, D. A., Haug, L., & Hsu, M. C. (2011, October). Visual sentiment analysis on twitter data streams. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on* (pp. 277-278). IEEE.
- [16] Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of LREC* (Vol. 2010).
- [17] Lima, A. C., & de Castro, L. N. (2012, November). Automatic sentiment analysis of Twitter messages. In *Computational Aspects of Social Networks (CASoN), 2012 Fourth International Conference on* (pp. 52-57). IEEE.
- [18] Cvijikj, I. P., & Michahelles, F. (2011). Monitoring trends on facebook. In Dependable, Autonomic and Secure Computing (DASC), 2011 IEEE Ninth International Conference on (pp. 895-902). IEEE.
- [19] Li, G., & Liu, F. (2010, November). A clustering-based approach on sentiment analysis. In *Intelligent Systems and Knowledge Engineering (ISKE), 2010 International Conference on* (pp. 331-337). IEEE.
- [20] Pohl, D., Bouchachia, A., & Hellwagner, H. (2012). Automatic Identification of Crisis-Related Sub-Events using Clustering. In *Machine Learning and Applications (ICMLA), 2012 11th International Conference on* (Vol. 2, pp. 333-338). IEEE.
- [21] Corley, C. D., Mikler, A. R., Singh, K. P., & Cook, D. J. (2009). Monitoring influenza trends through mining social media. In *International Conference on Bioinformatics & Computational Biology* (pp. 340-346).
- [22] Suzumura, T., & Oiki, T. (2011). StreamWeb: Real-Time Web Monitoring with Stream Computing. In *Web Services (ICWS), 2011 IEEE International Conference on* (pp. 620-627). IEEE.
- [23] Karamibekr, M., & Ghorbani, A. A. (2012, December). Verb Oriented Sentiment Classification. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2012 IEEE/WIC/ACM International Conferences on* (Vol. 1, pp. 327-331). IEEE.
- [24] C. Fellbaum. Wordnet: An electronic lexical database.
- [25] Cai, K., Spangler, S., Chen, Y., & Zhang, L. (2008, December). Leveraging sentiment analysis for topic detection. In *Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT'08. IEEE/WIC/ACM International Conference on* (Vol. 1, pp. 265-271). IEEE.
- [26] Koncz, P., & Paralic, J. (2011, June). An approach to feature selection for sentiment analysis. In *Intelligent Engineering Systems (INES), 2011 15th IEEE International Conference on* (pp. 357-362). IEEE.
- [27] Mizumoto, K., Yanagimoto, H., & Yoshioka, M. (2012, May). Sentiment Analysis of Stock Market News with Semi-supervised Learning. In *Computer and Information Science (ICIS), 2012 IEEE/ACIS 11th International Conference on* (pp. 325-328). IEEE.
- [28] Colbaugh, R., & Glass, K. (2011, September). Agile Sentiment Analysis of Social Media Content for Security Informatics Applications. In *Intelligence and Security Informatics Conference (EISIC), 2011 European* (pp. 327-331). IEEE.
- [29] Toutanova, K., Klein, D., Manning, C. D., & Singer, Y. (2003, May). Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1* (pp. 173-180). Association for Computational Linguistics.
- [30] Iqbal S., Khalid M., Khan, M N A. A Distinctive Suite of Performance Metrics for Software Design. *International Journal of Software Engineering & Its Applications*, 7(5), (2013).
- [31] Iqbal S., Khan M.N.A., Yet another Set of Requirement Metrics for Software Projects. *International Journal of Software Engineering & Its Applications*, 6(1), (2012).
- [32] Faizan M., Ulhaq S., Khan M N A., Defect Prevention and Process Improvement Methodology for Outsourced Software Projects. *Middle-East Journal of Scientific Research*, 19(5), 674-682, (2014).
- [33] Khan K., Khan A., Aamir M., Khan M N A., Quality Assurance Assessment in Global Software Development. *World Applied Sciences Journal*, 24(11), (2013).
- [34] Amir M., Khan K., Khan A., Khan M N A., An Appraisal of Agile Software Development Process. *International Journal of Advanced Science & Technology*, 58, (2013).
- [35] Rehman T U., Khan M N A., Riaz N., Analysis of Requirement Engineering Processes, Tools/Techniques and Methodologies. *International Journal of Information Technology & Computer Science*, 5(3), (2013).
- [36] Umar M., Khan, M N A., A Framework to Separate Non-Functional Requirements for System Maintainability. *Kuwait Journal of Science & Engineering*, 39(1 B), 211-231, (2012).
- [37] Umar M., Khan, M. N. A., Analyzing Non-Functional Requirements (NFRs) for software development. In *IEEE 2nd International Conference on Software Engineering and Service Science (ICSESS), 2011 pp. 675-678*, (2011).
- [38] Khan, M. N. A., Chatwin, C. R., & Young, R. C. (2007). A framework for post-event timeline reconstruction using neural networks. *digital investigation*, 4(3), 146-157.
- [39] Khan, M. N. A., Chatwin, C. R., & Young, R. C. (2007). Extracting Evidence from Filesystem Activity using Bayesian Networks. *International journal of Forensic computer science*, 1, 50-63.
- [40] Khan, M. N. A. (2012). Performance analysis of Bayesian networks and neural networks in classification of file system activities. *Computers & Security*, 31(4), 391-401.
- [41] Rafique, M., & Khan, M. N. A. (2013). Exploring Static and Live Digital Forensics: Methods, Practices and Tools. *International Journal of Scientific & Engineering Research* 4(10): 1048-1056.
- [42] Bashir, M. S., & Khan, M. N. A. (2013). Triage in Live Digital Forensic Analysis. *International journal of Forensic Computer Science* 1, 35-44.
- [43] Faizan M., Khan M N A., Ulhaq S., Contemporary Trends in Defect Prevention: A Survey Report. *International Journal of Modern Education & Computer Science*, 4(3), (2012).
- [44] Khan, MNA., Khalid M., ulHaq S., Review of Requirements Management Issues in Software Development. *International Journal of Modern Education & Computer Science*, 5(1), (2013).

### Authors' Profiles

**Asad Mehmood** has completed his MS in Computing from Shaheed Zulfikar Ali Bhutto Institute of Science and Technology (SZABIST), Islamabad, Pakistan. He has over 7 years of industry experience at his credit. His research interests include Business Intelligence, Big Data Analytics and Sentimental Analysis.

**Abdul Sattar Palli** has completed his MS in Computing from Shaheed Zulfikar Ali Bhutto Institute of Science and Technology (SZABIST), Islamabad, Pakistan. His research interests include Data Mining and Software Engineering.

**M. N. A. Khan** obtained D.Phil. degree in Computer System Engineering. His research interests are in the fields of software engineering, data mining, cyber administration, digital forensic analysis and machine learning techniques.