

## **Chapter 2**

### **Literature Review**

Online social media has changed the way we communicate with the rest of the world. Today social media (SM) is a pervasive mode of online communication and connects all walks of life. Increased use of 'smartphones' has been accepted as an essential component of life that enables access to social media. The world's youth are often the most prepared to embrace and learn new technologies, and they are certainly at the forefront of all forms of social media. They have more technical know-how than many older generations. So, social media is definitely making a big impact on their daily lives for good and bad. We focused on the impact of social media in our daily life in some areas like e-governance, election, and disaster management. E-governance and disaster management in social media has been the vision of computer science research community for more than four decades. This chapter surveys the related work in this area focusing on e-governance, election and disaster management infrastructures and their requirements for social media usage. Our goal is to distill the key concepts and analyze their applicability, scope and challenges. A thorough understanding and a precise characterization of social media are essential to carry forward the lessons learned from the rich literature in disaster management and E-governance. Then we discuss the detailed literature survey of work done in the area covering social media attributes.

## 2.1 Use of Social Media in e-governance

In 1993, the US government launched its e-government initiative. The Internet has been seen as an essential medium to re-examine governance and transform the relationship between government and civil society [37]. E-government is generally seen as a new type of government concerned with information, and the implementation of the e-government system has become an important policy issue for governments around the world [38]. e-Governance, also known as "electronic Governance," is the use of Information and Communication Technologies (ICTs) to give individuals and associations easier access to essential services and information. The usage of the internet delivers the service faster and increases openness between the government and citizens. However, in developing nations such as India, where literacy levels are low and the majority of people live in poverty, it is difficult for the government to give its services to such residents over the internet [22]. India's e-Readiness Rank is also weak. ICTs, particularly the internet, are used in e-Government to improve the delivery of government services to residents, industries, and governmental organizations. It is not restricted to the public sector but also includes administration and management standard operating procedures in the private sector [39].

E-Governance is an effective technique for improving overall governance. It can not only increase accountability, openness, and efficiency of government operations, but it can also promote sustainable and equitable growth. E-Governance also provides a platform for direct delivery of governmental services to the most marginalised parts of society in the most remote areas, without the need for middlemen. The government uses e-government in practically every aspect of its operations. Government has extended its roots everywhere, from metropolitan regions to rural ones, and from politics to education. Whether it is the public or private sector, the ordinary man or the businessman, e-governance is heavily reliant on it [40]. The main goals of adopting microblogs in government organisations, according to Godwin et al., are to facilitate brief and rapid dialogues, engage the community,

provide outreach, announce vital information, report major events, upgrade material (through hashtags) for users, track information on microblog sites, and expand the reach of communication to wireless devices [41]. The use of Twitter by the government can be seen as a unique medium that enables interaction and collaboration with citizens. For many governments, Twitter has proved to be a new and essential place for electronic communication to reach citizens. Twitter fills a unique and important gap in electronic communication by allowing users to stay connected at all times. When it comes to communicating with the public, government agencies and public officials find Twitter to be an excellent, quick, timely, and beneficial medium for getting the message out and engaging citizens [42]. Table 2.1 and Table 2.2 shows the summarizing of e-governance literature surveys.

E-government systems are requisite to revive direct democracy, which is considered an ideal political system but is yet to be implemented because of time and space limitations associated with society's growth and increasing complexity [39]. While many government agencies have recognized social media as an essential information source and outlet, there is no comprehensive accounting of the needs and practices of social media. In times of natural disasters and/or emergencies, government agencies communicate with people through social media to provide alerts, warnings, understanding their needs, and mitigate distress [44–47]. Today's social technologies are generally transparent, especially regarding transparency and corruption. Although there are challenges and hurdles in implementation, it is possible to overcome these challenges through a combination of political will and technology [48]. It has been observed that the media can significantly influence governments' responses to the needs or demands of the people. Informed and situationally aware citizens get a better deal from the government, irrespective of the affluence of the states [49]. Information and news of geopolitical events in Twitter streams reflect the views and reactions of Twitter users to current events. Depending on the volume,

**Table 2.1 Use of SM in e-Governance**

Study type	Techniques	Ref	Dataset	Assumption	Observations
Empirical analysis of the government's Twitter activity	Content analysis, semantic network analysis, ego network analysis	[37]	MFAFF tweets (217 RT and 330 mentions)	Use of SM by Triple Helix Sectors	MFAFF use-fulness of Twitter in realizing digital co-operation government.
Socio technical theory	Scientometrics	[38]	145 articles from 7 core e-government journals	e-Government research that deals with developing countries issues/topics	e-Government studies pertaining to developing countries issues/ topics have rapidly increased
Domains of e-governance	Systemic Approaches	[22]	Case studies	Government processes (e-administration); connecting citizens (e-citizens and e-services); and building external interactions (esociety)	1. improving the internal workings of the public sector, 2. relationship between government and citizens, 3. the relationship between public agencies and other institutions
Case study	Classification	[39]	-	Challenges for e-Governance in India	1. Environmental and Social Challenges, 2. Economical Challenges, 3. Technical challenges.
Case study	-	[40]	E-Governance projects which is currently being used in India	Challenges for the implementation of e-government in India	Challenges are like low literacy, lack of awareness, low broadband penetration, lack of system integration within a department, and all other reasons.
Preliminary study	-	[43]	government projects	challenges faced in implementing e-governance in rural areas of India	The e-government projects successful implementation of e-governance in rural India

content embedded in Twitter streams can help track or predict public behavior if it can be extracted efficiently [50]. Recent analyzes of Twitter streams have shown that a large proportion of tweets are generally flattering, of interest only to relevant parties, contain user-to-user conversations, links to interesting pieces of news content, or spam and self-

promotion [51]. The social network systems enable massive scale distributed collaboration, information distribution, and collective intelligence in government areas at all levels, from local to national. Governments are facing extraordinary transparency claims and openness, further encouraged by electronic grassroots mobilizations using social technology [52]. In another work, Twitter account of 93 Saudi Ministries was examined using web-based analytical devices. In another work, Twitter accounts of ninety-three Saudi ministries were examined using web-based analytical tools. It was an exploratory network analysis, using quantitative data covering aspects of attempting to answer questions about the number of tweets (original, retweet, reply), number of followers/followers, and hashtag usage [53].

Facebook is another top-rated social media platform in India. India has over 270 million Facebook users, making it the top country among Facebook users, followed by the United States (190 million), Indonesia (190 million), and Brazil (120 million) [54]. The study explores the reality of social media usage (especially Facebook) in local government and the factors affecting its use. Studies found that people, economic freedom, government level, the organization operating social media. The use of SM by most of the governments has been very cautious and measured as the government faces many difficulties while using new technology - the massive cost of service. Lack of trust and/or resistance to change from public employees, inadequate IT infrastructure, coupled with lack of citizen participation with the lack of expertise, issues of the cultural and digital divide, etc. [55]. E-Government is the online operation of a government or the delivery of government services to citizens at their doorstep. E-Government, often known as e-commerce, refers to the online availability of government services. The technology and methodologies utilised in the E-Government initiative provide a road map for delivering services efficiently at the point of need.

**Table 2.2 Use of SM in e-Governance cont...**

Study type	Techniques	Ref	Dataset	Assumption	Observations
Theoretical Overview	Diffusion of Innovations, Social Presence, Social Influence, and Collective Intelligence	[42]	Twitter	Twitter's use in governments	Illustrates these developments utilizing four theories: Diffusion of Innovations, Social Influence, Social Presence, and Collective Intelligence theories.
Case study	Theoretical analysis	[49]	Panel data over the period 1958-1992 of the 16 major Indian states	enhancing government responsiveness based on asymmetric information between citizens and government	States that are more responsive tend to also be those with high levels of newspaper circulation, electoral turnout, and literacy
Case study	UTAUT2 and hypothesis testing	[55]	Facebook data	Influence of social networks represented by Facebook on e-government in Jordan	social influence, effort expectancy, personal innovativeness, and enjoyment significantly predicted the continuous use of e-government pages on Facebook

## 2.2 Social Media during Elections

The stakes are high during political elections, and proponents may try to gain support for their cause by actively manipulating social media. A viewer's conclusion of a candidate's popularity can have an impact by altering the followers. Recently, the number of followers for a presidential contender in the United States increased by more than 110 thousand in a single day, according to reports, and analysis revealed that the majority of these followers are unlikely to be real people. Two out of every three people in the United States use social networking sites like Facebook [56], Twitter, and YouTube, while search engines are utilised on a daily basis [57]. Monitoring what people share or hunt for on social media and the Internet has given us a better understanding of what people really care about and pay attention to at any moment in time. It also enables segments of the global population

in becoming more informed, arranging, and reacting quickly. However, social networking sites and search queries can be easily manipulated, which the press and the general public have failed to acknowledge [58]. Online working of a government or providing its services online to its citizens at their door step is known as e-Governance. E-Governance is E-Commerce technology means online availability of government services. The technology and the methods used in E-Governance project provide a roadmap for efficient delivery of services at the door step.

The term “Google bomb” refers to a type of Web spam that is widely used today and can be applied to all major search engines. Web spammers make associations between anchor words or phrases and linked Web pages by leveraging the explanatory power of anchor text (the phrase directly associated with a hyperlink). These groups compel a search tool to allow higher relevancy to often unconnected findings, causing them to appear in the “top 10” search engine results. The association of the phrase "miserable failure" with President George W. Bush’s Web page initially, and later with those of Michael Moore, Hillary Clinton, and Jimmy Carter, was a well-known Google bomb [59]. It was impossible to launch any successful Google bombs on politicians during the 2008 and 2010 elections, and it is wished that this trend will continue. We saw attempts by out-of-state political groups to influence voters just before the elections during the 2010 Massachusetts Special Election (MASEN) to fill the seat vacated by Senator Ted Kennedy’s death. Many political decisions, particularly those related to governance or elections, have been informed and influenced by it. According to studies, political opinion on social media has a predictive power that outperforms traditional, expensive polls [60].

Social media sentiments can influence others’ views on the current government [61] or who ultimately wins an election [62]. Sentiment Analysis [63], a widely used technique for automatically detecting these sentiments, has revealed potential voters’ political preferences more accurately than public opinion polls or surveys [64]. Over the years, sentiment

analysis has been used on the components of popular social media platforms like Twitter and Facebook with the goal of understanding and predicting election outcomes in a number of countries. The Dutch Senate election [65], the Pakistan general election [66], the Indonesia presidential election [67], and the UK general election [68] are examples of elections that used sentiment analysis. During and after elections, sentiment analysis has also investigated public attitudes toward successful and unsuccessful candidates. According to the findings, people are more likely to distance themselves from a failed candidate and embrace a successful counterpart, while others will remain loyal to the unsuccessful candidate [69]. Tumasjan et al. [62] conducted three research studies in the context of the 2009 German federal election:

- They gathered tweets that either mentioned the six political parties or famous politicians from those parties to see if Twitter facilitates political debate.
- They looked at whether tweets reflect political sentiments in the real world.
- They looked at whether the volume of tweets tends to reflect party popularity in the real world and can predict election outcomes.

Their findings support the widely held belief that social media can be used to discuss political issues and that social messages closely reflect offline sentiments.

Razzaq et al. used public sentiment expressed toward political parties on social media to analyse and predict the Pakistan general election. They used supervised machine learning techniques to sort tweets into three categories: positive, negative, and neutral. They compared the average accuracies of Nave Bayes and SVM, as well as other machine learning algorithms. With an average accuracy of 70% for binary classification and around 55% for multiclass classification, Nave Bayes was the most accurate [66]. Hate speech and offensive posts on social media are increasing day by day. People post messages or tweets, often targeting people with hate and lewd words. Such messages often hurt people, causing

enormous psychological distress and trauma to the users at times. Instead of bringing people together, it causes a digital divide and social alienation. Such practices should be minimized if they cannot be stopped entirely for reasons like maintaining the courtesy and decorum of any platform so that everyone feels at home to participate. However, the absence of any moderator to flag a post objectionable often makes the job difficult. Efforts are, therefore, on to automatically detect the use of various forms of abusive languages in social networks, micro-blogs, and blogs so that prevention can also be thought. Since manual filtering takes much time and can cause human annotators symptoms such as post-traumatic stress disorder, several research efforts have been made to automate this process [70]. The task is designed as a supervised classification problem, where the system is trained for the presence of any kind of offensive or objectionable material. Hate speech in communication is considered harmful (at a personal or social level) based on defined 'protected characteristics' such as race, disability, sexuality, etc., whereas offensive speech is simply any communication that harms someone. A review of the available research on hate speech and offensive content detection shows in Table 2.3.

**Table 2.3 Summary of the existing work on Hate Speech and Offensive content identification**

Model Name	Single task	Multi-task	Pretrained Embedding	Pretrained contextual model	Emotion Features
MTL-CN (proposed)		✓	✓		✓
MTL [71]		✓	✓	✓	✓
fBERT [72]		✓	✓		✓
Transformed Word Embedding Model (TWEM) [73]	✓		✓		
CNN,LSTM,BLSTM [74]	✓		✓		
LSTM [75]	✓		✓		
LSTM and BERT [76]	✓		✓		
stacked CNN and GRU [77]		✓	✓		
XLM-R [78]	✓		✓		
BERT [79]	✓		✓		

Quite a few works have been done in the field of irony detection from social media posts. Twitter and Facebook are two favorite sources of data generation for this task, as they are easily accessible and provide a fast and convenient way to search for (potentially) irony posts using hashtags like #irony, #scasm etc. Regardless of the methodologies and algorithms, irony detection often involves binary classification, where irony is defined as instances that convey the meaning of the opposite [80, 81]. Used [82] a multilayer perceptron with many features (lexical, syntactic, semantic and polarity features) and attained an accuracy of 70.15% for binary classification. It is to be noted that irony detection research often relies on automatically annotated (i.e., based on irony-related hashtags) corpora, as described in [83]. Described [81] new approaches to the irony detection that can be broadly classified as either rule-based or (supervised) machine learning-based. While rule-based approaches rely mostly on lexical information and do not require any training, machine learning relies on training data. It uses a variety of information sources (or features), such as bag of words, syntax patterns, sensitive information, or meaningfulness. Described [84] the use of Bi-LSTM at the word and character level to identify tweets as ironic and non-ironic and achieved an accuracy of 0.7321 for Sub-task A. Used [85] maximum entropy classifier to achieve an accuracy of 0.788 for binary classification of tweets.

Gonzalez et al. (2018) [86] used two layers of convolutional neural network and a LSTM layer. Their work on English and Spanish tweets using multilayer perceptron for the binary classification of tweets and achieved an accuracy of 0.5727 [87]. There have also been some works on application of ensembling-based techniques [88] at SemEval-2018 used a soft ensemble-based voting classifier with Support Vector Machine and Logistic Regression. They ranked third for binary classification of tweets with a  $F_1$  score of 0.65 and [89] used LSTM for the binary classification of tweets and achieved an accuracy of 0.6173. Used [90] the majority voting ensemble and coverage based ensemble and

achieved  $F_1$ -scores of 0.65 and 0.63 respectively. A work at SemEval-2018 for Task 3 by YNU-HPCC [91] proposed an ensemble of various LSTM based models that achieved an accuracy of 0.66 for their ensemble-based system to detect irony and sarcasm in micro-blogs. A multi-layer approach and a hierarchical Bayesian model averaging for the classification of tweets into irony and non-irony. They achieved an  $F_1$ -score of 0.5253 for the Bayesian model averaging and 0.5261 for hierarchical Bayesian model averaging respectively [92]. Ahmed et. al. [93] used an unweighted graph for each class, and the comparison of the classes resulted in a vector that used as a feature for the model training. Their  $F_1$ -score was 0.5955 for the task of binary classification of irony and non-irony. A Siamese model for classifying irony and non-irony and achieved a  $F_1$ -score of 0.7234 [94]. Zhang et al. used transfer learning for the detection of irony and built a model that was previously trained on sentiment resources like sentiment word lexicon and sentiment tweets corpora. They trained two different models, one for each above. Another model is trained that inherited learning from the above two models to detect irony posts and achieved an  $F_1$  score of 67.55% [95].

## 2.3 Use of Social Media in Disaster Management

When disasters (natural or man-made) strike, the use of SM increases manifold. People try to communicate with their family and friends in disaster areas and inquire about the safety of their lives. They share or seek information about the shelter, food, transportation, medical and other needs. While many phone networks are unable to deal with a sudden surge in calls (thousands of people trying to call at once) or texts immediately following a catastrophic event, online social media such as Facebook and Twitter can serve people. These media have the ability to handle large traffic, can stay online, and can serve as a medium of communication. It can also remove the barrier of one-way communication

between television and radio. In fact, in the case of natural calamities, when the power is cut or gets disconnected, these channels of communication (radio/television) can also get disrupted. Nevertheless, through smartphones, tablets, and other devices, people can access social media for real-time information exchange [96]. They can convey their well-being and location with friends and family.

In an emergency, everyone is a potential source of important information that collects and distributes content [28]. Online social media, thus, can play an all-important role during crises and emergencies, and it is, in fact, on the rise [29].

Increased use of 'smart phones' has been recognized as an essential factor that enables access to social media. In recent years, Twitter has been used to spread the number of casualties and damage, donation efforts, and alerts, including multimedia information like video and photos [97][98]. Recent incidents of natural disasters have seen the network of humanitarian organizations and volunteers setting up web-based manual crisis mapping sites like the Haiti 2010 Earthquake, the Russian 2010 Wildfire, 2012 Hurricane Sandy of New York, and Oklahoma for the 2013 Tornado [99].

Information ceaselessly flows in SM and it can be used to benefit emergency managers [100–102]. With the spread of network technologies, researchers have started focusing on the elements of network components, as well as reducing the impact of network elements on the effects and accelerating recovery from extreme events [103–106]. Social media, as an emerging and evolving communication technology, offers the possibility of better emergency communication for its capacity, dependence and interactivity [107]. Although a few surveys on the topic of disaster and/or crisis management exist in the literature, they primarily focus on some specific domain of administration or management like public health [108], use of social media by the public during disasters [109–111], or case studies related to a particular event ([112, 113]), etc. These papers cover very

few technicalities involved in using social media data as far as information processing is concerned. Some of these surveys are also a little old, dating back to 2007 [114]. Among the recent surveys, Kankanamge et al. (2019) [115] discussed data collection through volunteer crowdsourcing from the disaster domain along with from other domains like marketing and communications. Another survey by Chaves et al. (2019) [116] looked at how an urban emergency management infrastructure can be designed and built through crowdsourcing involving people in a participatory model. Both the surveys look at the problem from the point of view of policy administration and management with little or no discussion on computational techniques or methodologies involved therein.

In the social media gets a surge during emergency situation. Getting useful information out of this huge data in real-time is critical, at the same time, an increasingly difficult problem. How this sudden influx of information can be mined to get what one actually requires, what are the effective and efficient techniques to achieve this – are explored in this section. We cover the papers that primarily focus on different information accumulation techniques from microblogging sites.

### **2.3.1 Extraction**

In text mining, information extraction is an important task. The general goal of information extraction is to find structured information within uncontrolled or semi-structured text. Two basic functions of information extraction have been named entity recognition and relation extraction [117]. We classify papers focusing on the techniques into a number of categories. Table 2.4 and Table 2.5 summarize the techniques.

#### **Information Extraction techniques**

Vieweg et al (2010) [102] discussed the manual identification of relevant information during emergency situations: specifically, through identifying geo-locations, location

**Table 2.4 Extraction Techniques**

Study type	Techniques	Ref	Dataset	Assumption	Observations
Improving situational awareness	Information Extraction techniques	[102]	4,592,466 tweets for Red River Floods (RR) and 1,986,091 tweets for the Oklahoma Grassfires (OK)	extracting useful, relevant information during emergencies	The events identify the characteristics of the information generated during the emergency and lead to the development of a working framework to inform the design and implementation of software systems that employ information extraction strategies.
Extracting Information Nuggets from Disaster	Naïve Bayesian classifiers	[118]	206,764 tweets	A system to automatically extract information nuggets from microblogging messages during disaster times(7 Extractor)	The precision of all the extractors are above 0.7 except for the Damaged Object Extractor(0.47)
Preliminary study	Spectral clustering	[119]	England riots = 1,165,628, hurricane Irene = 90,237 and earthquake in Virginia = 277,604 tweets	identify and characterize communities from a set of users who post messages on Twitter during crisis events	81% accuracy
Improving situational awareness	Information Extraction techniques	[102]	4,592,466 tweets for Red River Floods (RR) and 1,986,091 tweets for the Oklahoma Grassfires (OK)	extracting useful, relevant information during emergencies	The events identify the characteristics of the information generated during the emergency and lead to the development of a working framework to inform the design and implementation of software systems that employ information extraction strategies.
Extracting Information Nuggets from Disaster	Naïve Bayesian classifiers	[118]	206,764 tweets	A system to automatically extract information nuggets from microblogging messages during disaster times(7 Extractor)	The precision of all the extractors are above 0.7 except for the Damaged Object Extractor(0.47)
Preliminary study	Spectral clustering	[119]	England riots = 1,165,628, hurricane Irene = 90,237 and earthquake in Virginia = 277,604 tweets	identify and characterize communities from a set of users who post messages on Twitter during crisis events	81% accuracy

**Table 2.5 Extraction Techniques continue..**

Study type	Techniques	Ref	Dataset	Assumption	Observations
Analysis of public behavior	Visual Analytics Approach	[120]	Real-time tweets, collect about 2.2 million geo-tagged Tweets with in the United States per day	How people prepare and respond to disasters, plays an important role in crisis management, disaster response, and evacuation planning	Demonstrated an integrated visualization that allows spatial and temporal aspects with in a single view
Case studies	Counter-disaster systems	[121]	events of the disaster in order to show the state of the country after March 11	comprehensive survey of Counter-disaster systems	Future disasters: i) authorities are advised to disclose statistical information as quickly as possible, ii) coordination among developers must be provided, iii) interconnection of databases is essential for efficiency
Real-Time Crisis Mapping	Geospatial data-extraction tools & Open-StreetMap to access street-level	[122]	877,527 tweets in 5 days	Compare street-level Twitterbased crisis maps to a verified ground truth based on post-event expert assessment	90% precision
Compares the capabilities of two popular systems- Project Epic's Tweak the Tweet (TtT) and Ushahidi	Geospatial Technology	[123]	Twitter Data or SMS Text	Earthquake data	Both provide rudimentary situational awareness and visualization features

referencing and situational updates. Through a comprehensive study of two emergency events they came up with a number of features that can be used to identify useful tweets in any information extraction (IE) techniques.

### **Naïve Bayesian classifiers**

Imran et al (2013) [118] described a method of machine learning based extraction of information nuggets from Twitter data on Joplin tornado 2011 data. For a filtering task, they first manually created training data through crowdsourcing and labeled the raw tweets into five categories: personal, informative (direct), informative (indirect), informative (direct or indirect) and other. They found 60% of the training data belonging to informative. Next they trained a Naive Bayes multi-class classifier to automatically classify informative tweets into further sub-categories based on a set of binary, scalar and text features. They also extracted information nuggets of different categories: caution & advice, casualty & damage, donation & offer etc with reasonable accuracy. The experiments established application of machine learning based filtering and extraction of structured information nuggets in crisis informatics with acceptable performance.

### **Spectral clustering**

Gupta et al (2012) [119] identified, extracted and characterized prominent components of topics and opinions that people share during crisis events on Twitter. The aim of the study was to identify different user community based on the topic of discussion and mark them by top central users based on Twitter data during England Riots, Hurricane Irene and Earthquake in Virginia, occurring in 2011. To detect communities, first user-user similarity calculation was done using content-similarity (common words, hashtags, URLs), link similarity (based on # retweets, mentions, replies between two users) and meta-data

similarity (common location). Spectral clustering was then applied to find clusters. All the users were thus divided into clusters. Degree centrality was used to find top users in each cluster. The experiments suggested that the top users represent the topics and opinions of an entire community with an average 81% accuracy.

### **Visual Analytics Approach**

Chae et al (2014) [120] developed a tool for visual analysis of spatiotemporal distribution of tweets to identify public behavior patterns during natural disasters. The authors collected geo-temporal information of the Twitter users and plotted them on a map using different colors based on posting time. On a single frame, the tool offered visualization of both location and time. The main features of this approach are spatial analysis and decision support, temporal pattern analysis and spatiotemporal visualization. All methods are tightly integrated based on a user-centered design in order to enhance the ability to analyze huge social media data. The system provides effective analysis for exploring and examining the spatial distribution of Twitter users and supporting spatial decision making using a large volume of geo-located tweets and multiple types of supplementary information during specific time periods. Using user movement data of two emergency events: hurricane Sandy of 2012 in New York city and moore tornado of 2013 they showed how the visualization tool can help understand public movement from normal time to time of crisis. Also it can help in understanding the effect of public warnings, design evacuation plan and disaster management.

### **Geospatial Technology**

McClendon et al (2013) [123] compared capabilities of Project Epic's Tweak the Tweet (TtT) and Ushahidi, both of which collect and visualize crowd-sourced information. TtT

system takes the tweets from Twitter streaming API with content is mapped to locations using Google map. Ushahidi can be configured to work with Twitter API and provides APIs for data exchange in XML and JSON. The raw data can be saved in Google spreadsheet and integrated with Googlemap. Both the software are comparable in their capability of creating crisis maps and provide rudimentary situational awareness and visualization features.

### **Counter-disaster systems**

Middleton et al (2014) [122] came up with a social media crisis mapping system in real-time for natural disasters. They collected disaster-prone areas or locations from the gazetteer, road map and voluntary geographical information (VGI) sources and matched them with real-time tweets. GoogleGeoCoding API is used for necessary geo-coding. With case studies on hurricane Sandy and Oklahoma tornado, they demonstrated more than 90% precision in geo-parsing from real-time tweets.

### **Machine Learning Methods**

Utani et al (2011) [121] did a comprehensive survey on a plethora of systems developed during the Japan earthquake 2011 triggered by real world needs of victims and survivors and shared by Twitter, Facebook and Mixi (a Japanese networking service). The authors identified 162 counter-disaster systems use Google Person Finder (a database service with details of missing persons and identified persons), Traffic/Road information service based on car-navigation data and Google map, Electric Power information tool to forecast power outages, radioactivity monitoring systems.

**Table 2.6 Event Detection Techniques**

Study type	Techniques	Ref	Dataset	Assumption	Observation
Earthquakes detection	Short and Long-term-average algorithm	[130]	real-time earthquake & tweets containing the words “earthquake”, “gempa”, “temblor”, “terremoto”, or “sismo”	categories of alert dissemination, situational awareness, and event detection	75% occur within two minutes of the origin time
satellite observations and sensor measurements	GNOME model	[131]	NOAA/NOS medium resolution coastline data designed for 1 : 70 000 scales.	GNOME trajectory forecasts	oil trajectories, indicating the need for realistic quantitative mapping of the thickness of inferred oil spill images
Event Detection	Gradient boosting and Random forest	[132]	32,400 tweets	tweets in English and Hindi languages	81% accuracy
Event detection and tracking	Transaction-based Rule Change Mining (Association rule mining)	[133]	Twitter dataset from: FA Cup final(444,291 tweets), US Elections 2012 (3,837,291 tweets) and Super Tuesday 2012 (474,109 tweets)	Dataset divided into different tweets/time slots (FA: 2000 tweets/time slot of 1min; Election: 20,000 tweets/ slot of 10min; SuperT-day: 10,000 tweets/slot of 1hr)	FA cup(Prec - 96.6%, F-score 76.9%), US elections(Prec - 53.8%, F-score 60%),Super Tuesday ( Prec - 62.5%, F-score 58.79%)

### 2.3.2 Event Detection

An event is an arbitrary classification of the space-time zone. An event can be a actively participating agent, idle factor, products, and a location in space/time [124]. Events such as earthquakes, typhoons, traffic jams, and flood are visible through tweets. A target event is extracted from tweets with the help of hashtags and keywords. For example, the rapid and massive influx of tweets coming soon after the earthquake prompted many blog postings and articles, which use Twitter as a tool for earthquake detection [125–129]. Some of the reported event detection techniques are summarized in Table 2.6.

### **Short and Long-term-average algorithm**

Earle et al (2012) [130] ran an experiment to check whether Twitter can detect an earthquake and to what extent. They checked rapid increase in the number of tweets containing the word "earthquake" or similar in other languages. Rapid increase was detected using short-term-average over long-term average (STA/LTA) algorithm popularly used in seismic study. Although the technique missed vast majority of the earthquake events (48 detected out of 5175 occurred), it detected the major ones within tens of seconds of occurrence.

### **GNOME model through event detection**

Aulov and Halem (2012) [131] used social media (SM) as a human sensor network in oil spill prediction. SM was included in the geophysical model, with satellite observations and sensor measurements. Along with the use of a special software GNOME that is used to predict oil spill, the authors collected tweets, Flickr images, with their geo-tags to compare and verify the oil-spill trajectory of GNOME. The longitude, latitude, timestamp of SM data were taken as ground truths and GNOME forecasts were compared with them.

### **Transaction-Based Rule Change Mining**

Olowe et al (2016) [133] studied event detection and tracking for three events in Twitter: from two different domains; Sports (English FA Cup 2012) and Politics (US Presidential Elections 2012 and Super Tuesday 2012). They applied Transaction-Based Rule Change Mining (TRCM), motivated by Association Rule mining on the tweets. Since a football game is of 90 min to 120 min event, whereas election is of much longer duration, they applied different datatype to different time-windows and update rates for their study (Football 2000 tweets per 1 min slot; President election 20000 tweets/ 10min slots; 10000

tweets per 1 hr update rate). TRCM was able to correctly detect 54 out of 59 events for FA Cup data and for election detection accuracy was 45%.

### 2.3.3 Summarization

With a summary, we can make effective decisions in a short time. Although some tools are already available, with increasing volume of online information, it is becoming difficult to generate meaningful and timely summaries [134]. Some of the reported summarization techniques are summarized in Table 2.7.

#### Content Analysis

Yates et al (2011) [138] did a case study on Haiti Earthquake (2010) response that was a large effort co-ordinated by three major agencies involving Govt of Haiti, UN and a number of countries. The authors detailed how a social media tool SharePoint was used to share, reuse, different information in the real time and cut down time in decision making. The paper discussed in detail how knowledge from different agencies were captured and shared without redundancy/duplicacy, discussed and then decision was communicated across them with the use of SharePoint and Wiki. The authors also pointed out potential loopholes in using such systems and suggested how to plug them.

Vis (2013) [136] studied the use of Twitter as a tool for reporting breaking news during UK riots 2011. The author mainly focused on the Twitter activity of two journalists of mainstream news media and analysed their ways of data collection as well as sharing of mainstream news content with links, images and videos. Both the journalists use Twitter for real-time reporting. While one journalist restricted himself in expressing personal opinion (5.2%), the other was more expressive (22.4%).

**Table 2.7 Summarization Techniques**

Study type	Techniques	Ref	Dataset	Assumption	Observation
Case study	Content Analysis	[135]	94,101 microblog posts and 41,817 re-posts	five major categories	1540 (35%) opinion-related messages, 1151 (26%) situated update messages, 811 (18%) general Yushu Earthquake related messages, 729 (17%) emotion-related messages, and 177 (4%) action-related messages
Twitter as a reporting tool	Content Analysis	[136]	Two datasets: 1. The Guardian/Twitter database of 2.6 million tweets, uploaded by 700,000 individual users 2. 731 tweets by two journalists	To study evolving nature of use of Twitter by instinctive journalists	Both journalists took to Twitter predominantly for eye-witness reporting (30%) with huge use of hastags (59% and 78%), also expressed personal opinion (5% and 22% respectively).
Australian Government to detect, assess, summarise, and report messages of interest for CC published by Twitter	Incremental clustering algorithm	[137]	Real-time data of Twitter	Cluster summaries with font size and colour used to indicate the relative number of tweets	Evaluation of this algorithm and approach is the subject of ongoing work.

### **Incremental clustering algorithm**

In a study of microblog use by Qu et al (2011) [135], the authors looked at three different questions: 1. what people post during disaster? 2. how the posting pattern changes over time during different stages of a disaster? 3. How information spreads in microblog-sphere? For Question 1, they iteratively developed a classification and coding scheme based on manual checking of few sample tweets randomly. For trend analysis, they looked at distribution of different types of post categories (done in categorization) over time. To understand information spread, they saw the frequency distribution of re-posts of different message categories followed by regression analysis to identify factors of reposts.

### **2.3.4 Classification**

Classification is the process of analyzing a particular input and assigning it to (one or more) category. Some of the reported classification techniques are summarized in Table 2.8

#### **Naive Bayes classifiers**

Yin et al. (2015) [139] discussed the relevant approaches for burst detection, tweet filtering and classification, online clustering, and geotagging. Burst detection focused on monitoring a feed of Twitter messages and raising an alert for immediate attention when an unexpected event was detected. When Twitter was monitored for a specific incident, then every other published tweet was considered to be largely irrelevant. These features were used for architecture classification: word unigram, word bigram, word length, number of hashtags, user number, whether the tweet has been retweeted, and other users have responded to a tweet.

**Table 2.8 Classification Techniques**

Study type	Techniques	Ref.	Datasets	Assumptions	Observations
Burst detection, tweet filtering and classification, online clustering, and geotagging	Support Vector Machines (SVM) and Naive Bayes classifiers	[139]	annotated dataset of 450 tweets	classification features: word unigrams, word bigrams, word length, the number of hashtags, the number of user mentions	Naive Bayes and SVM achieve classification accuracy of 86.2% and 87.5%
Classify and Review High-Value Messages	Support Vector Machines	[137]	near-real-time information from Twitter during February 2011 earthquake.	Classify and review high-value messages during an incident (e.g. messages describing infrastructure damage or cries for help)	Watch officers with a more targeted view of information during an incident.

### **Support Vector Machines**

Cameron et al (2012) [137] described a work with the Australian Government to identify, evaluate, summarise, and report tweets of interest for crisis coordination. The developed platform and devices, collectively termed as the Emergency Situation Awareness - Automated Web Text Mining (ESA-AWTM) system. Crisis coordinators required the tools and services that the social media needs to meet the following requirements:

- Find unexpected or unusual events, possibly beyond official communication;
- Provide messages and summaries about an incident to keep awareness of the content collected without reading individual messages;
- Classify and review high-value messages during an event (for example, message describing the infrastructure or cry for help), classification was done based on Support Vector Machines;
- Identify, track, and manage problems that arise in an event;
- Perform forensic analysis of events by analyzing social media content in advance during and after an incident.

### **Crowd-sourcing**

Gao et al (2011) [140] described the advantages and disadvantages of the crowd-sourcing application used for disaster relief coordination and challenges rumours, false information, baseless speculations, inaccurate geo-tags etc. Security is another problem which should be considered especially in the design of a crowdsourcing system in the Disaster Relief Management Plan. When making data publicly available, such systems should protect the privacy of NGOs and ensure the safety of their workers. Crowd-sourcing integrated with the map of crisis has been a powerful tool in the form of humanitarian aid and disaster. Future

crowding applications should provide capabilities to better manage untested messages and improve streaming data.