Chapter 7

Conclusions and Future Directions

We present conclusions and future work at the end of each of the five chapters (Chapters 2 to 5) describing five broad applications on open-source social media intelligence in government domain. In this chapter, we summarize our work, present general conclusions, specific conclusions already covered at the end of each chapter and provide future directions.

7.1 Abstract and General Conclusions

In this dissertation, we demonstrate how open-source social media intelligence (OSSMInt) can be applied in government domain. While there has been a lot of work done in the area of social media intelligence by private companies and private sector for developing consumer-based applications, our dissertation makes a unique contribution in the relatively unexplored area of OSSMInt in government and security informatics domain. We categorize our proposed applications into three broad classes: identification, prediction and response applications. We propose and implement five broad applications and usage scenarios spanning these three categories: (1) citizen complaints and grievance redressal [response] (2) religious conflict detection [identification] (3) hate and extremism promoting content, users and community detection [identification] (4) secret message and communication detection [identification] (5) early civil protest forecasting [prediction].

We formulate these applications as text analytics based problems and propose information retrieval (IR), machine learning (ML) and natural language processing (NLP) based solutions to develop these applications. However, due to the free-form nature of social media data, building such applications is a technically challenging problem. The presence of noisy content such as spelling errors, incorrect grammar, unstructured text, use of slang and abbreviations makes it challenging to develop applications without ignoring the false positives. Further, due to the presence of disguised emotions and ambiguity in the content, it’s hard to label the data by human annotation as well. Our work demonstrates that government can gather a lot of intelligence from publicly available user-generated data on social media platforms. We show how government can automatically analyze and monitor social media data in real time for improving their decision making. The chapters in the dissertation are organized in the form of applications and case-studies and show how useful intelligence can be gathered without invading into anybody’s privacy as open-source social media data is public. Hence one of the major and broad conclusions of our work is the following:

There are several untapped opportunities as well as technical challenges in exploiting open-source social media data by the government for intelligence gathering.
We present general technical challenges [Chapter 1 - Introduction] which covers all the five broad applications described in this dissertation and also includes the unique technical challenges within the context of each application [respective chapters]. Another significant contribution of our thesis is a demonstration of how the open-source social media data can be systematically collected and processed by applying different IR, ML and NLP based techniques. Crawling, filtering, cleaning, annotating and pre-processing (unlike stopwords removal and text tokenization) social media data is non-trivial. We present solutions and approaches to enhance this that varies for different platforms. The text enhancement and enrichment of user-generated data includes language detection and translation in multilingual scripts, removal of non-informative content, expanding joint keywords and slang without affecting the problem specific terms and handling incorrect spelling and grammar.

Advanced multi-step pre-processing of social media data is mandatory before it can be used for analysis or statistical model building. There are no one-size-fits-all pre-processing tools, and each step in the pre-processing pipeline requires problem and domain specific solutions.

Despite having a lot of similarity in terms of the usability and reachability of the websites, several online social media (OSM) platforms facilitate unique features to their users. One of such features is the amount of text that can be uploaded to the website at once (micropost, short text or long text). NLP based techniques are the integrated part of building government applications by mining social media data. The performance of NLP-based approaches varies based on the amount of input text. The variation in length of each data input impedes the accuracy of linguistic features.

In addition to the enrichment of raw text, in order to make content identification approaches generalized, it is mandatory to identify platform independent features that are generalized for all sizes and structures (formal, informal, unstructured) of the input text.

OSSMInt in Security Informatics and government domain covers the applications that require experts to annotate the results and evaluate the performance of proposed methodologies. However, due to the high velocity and massive size of available social media data, creating ground truth for each record is overwhelmingly impractical. Further, the lack of annotated data impacts the performance of feature extraction and supervised learning based classification techniques. Based on our proposed approaches we conclude that the unsupervised, semi-supervised and active learning based approaches are able to learn the features on the way. The use of such classification techniques efficiently categorizes the content reducing the efforts taken into annotation of each record. Furthermore, due to the excess of irrelevant content on social media, the data is highly imbalanced, and a very small percentage (0.000001%) of records belong to the topic and relevant class (hate promoting, complaint reports, religion and race-targeted posts, discussion on protest and unrest events). Therefore, it is impractical to identify the class of each record and extract discriminatory features for them.

To identify the relevant content by applying OSSMInt in government domain, unsupervised and semi-supervised learning based techniques are reasonably effective. While, training the classifier on the features of targeted class (one-class classification, identifying relevant records and classifying other inputs as unknown) are practically feasible and reduce the noise in the classification.

OSSMInt based applications for government and law enforcement agencies include the identification of many such contents on social media that includes sensitive topics like religion, race and anti-national posts. While it is important for the security analysts to monitor each and every post that might lead to the violence
and unrest; it is also important to identify the intent and objective of the author before removing the content or account from social media. Due to the presence of ambiguity, references, and sarcasm, the use of shallow keyword-spotting techniques is discouraging and requires deeper analysis.

Identification of authors’ personality and psychological behaviors including social tendencies, emotions, social references, and writing style is mandatory for capturing the dependencies between content and intent of a post.

Online radicalization, planning of civil protests, conflicts of religious beliefs, online terrorism activities are not specific to only a one or a few countries or specific regions but are the topics of concern for the security analysts at a global level. Social media websites allow their users to post textual content in a variety of languages and scripts. Several groups and communities of like-minded people and individuals take leverage of such features and post ill-disposed content in the regional language of targeted audiences. Further, the users expressing their religious opinions and sentiments in their community on social media use their regional language for communications and discussions.

In order to remove the bias from data collection and geographical independent analysis, it is mandatory to analyze the multilingual texts and scripts (by translating them into a base language) on social media and not just analyzing the content posted in the defined base language.

In addition to the users publishing sensitive and harmful content on social media, there are several users who do not actively post any malicious content on their channel but silently contribute to the community. Online social media platforms allow their users to see the different activity feeds of their followings and subscriptions varying for different platforms. For example, YouTube and Facebook allow users to see the ‘like’, ‘favorite’ and ‘comment’ activities unless kept as private. Therefore, users can disseminate information and spread the already published content among their followers without actively posting them on their channel.

It is required to monitor and analyze the activity feeds of the users and not just the uploaded content on their channels for identifying the right communities and reduce the number of false negatives (silent contributors identified as naive bloggers).

7.2 Specific Conclusions

In addition to the general take aways from this thesis, we also discuss some of the specific conclusions of the work presented in each chapter of this dissertation.

Mining Twitter to Identify Citizen Centric Complaints and Grievances

1. In the first application of OSSMInt for government discussed in Chapter 2, we perform a case study on 4 different Twitter accounts of Indian public service agencies for identifying complaints and grievances of citizens. Based on our results, we conclude that computational linguistic and natural language processing-based techniques are an efficient method for identifying discriminatory features from complaints reports.
2. We also conclude that a prior identification of tweets that are certainly not the complaint reports (appreciation, information sharing, and promotional tweets) improves the accuracy of complaint tweet classification.

3. In Chapter 2, we also conclude that for the same features and training data, the accuracy of baseline SVM classification method is improved by ensembling different kernels functions of SVM. In the first case-study conducted in Chapter 2, we were able to boost the accuracy of our classifier up to 20%.

4. In addition to the identification of response-seeking based complaint reports, in Chapter 2, we also conduct a case-study on mining Twitter for identifying complaints on bad road conditions. Based on the performance of proposed approach, we conclude that identification of pinpoint location (or landmark), region (or city), and key issue reported in the complaints are the three primary features for identifying the road related complaints.

5. Based on our experiments conducted on the Twitter dataset, we also conclude that there are several complaints which are incomplete and do not contain all three primary features. However, based on the type of missing information (such as missing region or city information) these tweets can be enriched and used for further analysis.

6. We conduct a characterization study on the tweets identified as complaint reports and conclude that a maximum number of complaints are reported about the risky, dangerous and accident prone roads while most of them are due to the poor condition of amenities.

Mining Public Opinions on Tumblr for Identifying Religious Conflicts

1. In Chapter 3, we conduct a case-study on open source data of Tumblr website for identifying religious conflicts within society. Based on our experimental results, we conclude that social media is a rich source of information for identifying religious beliefs of people and mining these religious beliefs and sentiments on OSM websites fills the gaps of offline surveys for religious conflicts identification.

2. Based on our survey conducted among people from different groups (graduate students, Tumblr bloggers and users, and random individuals from society), we conclude that the presence of positive and negative sentiments in religion based posts do not reveal the religious beliefs of the author. But it requires deeper analysis of the context in terms of the emotions present in a post. For example, disgust, insult, disappointment, defensive, ashamed and disbelief.

3. In order to identify the religious conflicts on social media and fill the gaps of offline studies, it requires a significant number of posts from different groups of users. However, it is overwhelming impractical to annotate each post individually. In the work presented in Chapter 3, based on our experimental results we conclude that unsupervised and semi-supervised learning based approaches are the efficient techniques to process and large-scale data.

Identifying Extremist Content, Users, and Communities on Social Media

1. In the first case-study (Section 4.4) presented in Chapter 4, we conduct experiments on Twitter data for identifying hate promoting content posted in form of tweets. We propose to use an n-gram based model and investigate the efficiency of various features like presence of war-related terms, religious mention, negative emotions and use of swear & offensive words. Our results reveal that due to the sparse and highly imbalanced nature of dataset, SVM outperforms K-NN with a margin of 23% in F-score.

2. To investigate the efficiency of selected features, we apply leave-p-out strategy and measure the performance of classifiers. Our results reveal that while, the presence of war-related terms, negative emotions and swear words are discriminatory features for identifying extremist content, classification based on the presence of certain hate promoting hashtags impedes the performance of classifiers.
7.2 Specific Conclusions

3. In the second case study (Section 4.5) presented in Chapter 4, we propose to use computational linguistic based approach for identifying the intent of the author in radicalized and racist posts. Based on our results, we conclude that shallow NLP and keyword spotting techniques are inefficient for determining the intent of the radicalized posts. Whereas, the use of author’s psychological behavior such as personality traits, social tendencies, writing cues and semantic role of terms used in the post are strong indicators for identifying the radicalized or racist intent-based posts.

4. We conduct our experiments on open source data of Tumblr website, and our results reveal that prior identification of topics (such as religion and race) in the posts and filtering non-topic posts improves the accuracy of intent classification.

5. In addition to the identification of radicalized, racist posts, in Chapter 4, we also present our work on identifying hate promoting users and uncover their hidden and virtual communities on social media. We conduct our case studies on YouTube (Section 4.6) and Tumblr (Section 4.7) websites. We propose to use link analysis and topical crawling based approaches for navigating through the website and identify the channels and bloggers promoting extremist content on the website. Our results reveal that while conducting our experiments on YouTube, Shark Search based graph traversal algorithm outperform Best First search with a margin of 5% (accuracy 74%). Whereas, Random Walk based algorithm on Tumblr gives an accuracy of 77%.

6. We perform social network analysis on the extremist users on YouTube and Tumblr website for identifying the central users playing major role in the community. Our results reveal that in cluster-based networks of YouTube users, shark search navigation is able to extract large number strongly connected users and communities in comparison to best first search navigation. While based on the social network analysis performed on the Tumblr bloggers, we conclude that in comparison to the like feature, re-blogging is a strong indicator and a discriminatory feature to capture strongly connected communities on the website.

Detecting Word Obfuscation in an Adversarial and Secret Message Communication

1. In the work presented in Chapter 5, we propose a conceptual similarity based approach to detect term obfuscation in adversarial communication. To evaluate the performance of proposed methodology, we conduct our experiments on three different datasets: Enron email corpus (EMC), Brown news corpus (BNC) and the examples used in previous literature. Our experimental results reveal an accuracy of 72.72%, 77.4% and 62.0% for EMC, BNC and examples dataset respectively.

2. Our results reveal that computing the conceptual and semantic similarity between the terms is an effective method for identifying out-of-context words in the sentences. Empirical evaluation and validation show that commonsense knowledgebase is an efficient lexical resource for identifying the obfuscated terms in formal as well as informally structured sentences.

3. Based on our results, we conclude that the proposed approach is also able to detect term obfuscation in long sentences containing more than 5 – 6 concepts. Furthermore, we conclude that the proposed approach is generalizable as we conduct experiments on nearly 2000 sentences belonging two three different datasets and diverse domains.

Mining Twitter Data for an Early Forecast of Civil Protest and Unrest

1. In the work presented in Chapter 6, we propose a trend analysis and time sliding window based approach for an early prediction of civil unrest and protest related events. We conduct our experiments on immigration-related tweets downloaded from publicly available data on Web Observatory. Based on
7.3 Future Directions

Our experimental results conclude that identification of crowdbuzz & commentary and mobilization & planning tweets posted prior to the event date are lead indicators for predicting a protest event.

2. We propose a frequency-based model and compute the correlation between topic, spatial location and temporal features for identifying the significant pairs of entities that co-occur together over a time period (for example, from a week before the event). Our experimental results reveal that identification of such significant pairs in lead indicator tweets and analyzing their trend in sliding window frame improves the performance and confidence score of the predictive model.

7.3 Future Directions

The work presented in this dissertation is interdisciplinary as it at the intersection of fields such as computer science (particularly machine learning, information retrieval, natural language processing), security, social media, e-governance, e-participation, counter-terrorism and law enforcement. We propose the following future directions of our work which draws from multiple domains.

7.3.1 Integration with Other Domains

1. **Information Visualization** - The end-users of the proposed applications and usage scenarios in this dissertation are the government agencies and security analysts. It is important for them to not only be able to build such applications that can detect relevant content on social media, but it is also important to visually analyze the current state-of-the-art in the form of dynamic and interactive visualizations [196]. We have not explored sophisticated visualization techniques and advanced topics in information visualization in our work. We believe that visual analytics and interactive visualization can further improve the value and usability of the applications presented in our work [197].

2. **Deep Learning** - We have applied several machine learning based techniques for statistical model building in our work. However, we have not used deep learning based methods which are recent and advanced developments in machine learning research and algorithms [198]. Deep learning has shown encouraging results in areas such as speech and image recognition, and we believe that the application of deep learning for solving problems defined in this dissertation can be interesting [199].

3. **Sensemaking** - The user-generated data on social media is written in free-form text and hence does not have a specific structure, vocabulary or defined terminology while reporting for complaint tweets. Further, due to the excessive use of sarcasm in the reports, the standard NLP techniques become shallow and are not able to capture the issues reported in the complaint reports. Our future work includes the application of sensemaking in identifying hidden and ambiguous complaints in micropost. Sensemaking is an area that comprises of developing theories, applying psychological research, developing interaction between human sensors and information technology and making sense of observed data [200] [201].

7.3.2 Novel Research Applications of OSSMInt for Government

1. **Intent vs. Impact** - In addition to the identification of posts and channels created with the intention of promoting hatred and a shared propaganda, we propose a novel use-case of OSSMInt in government domain. The aim of proposed application is to identify the posts and users who cause an extreme and grave impact on their viewers irrespective of the intent of the post. The proposed application not only detect the intent and content-based posts but it also aims to analyze the amount of attention received
on the post. The impact of such posts can be measured in terms of identifying the user engagement, discussion, and activities performed on and regarding the post.

2. Detection of Online Recruitment in Radical Groups- The aim of online radicalized communities on social media is to not only spread hatred and promote their ideologies, but it also includes recruiting young people in their groups. The aim the proposed application is to perform a real-time and dynamic analysis of the extremist communities on social media and capture the phenomena of variation in new participating nodes in the community and existing nodes leaving the community. We propose to use the advanced and interactive visualizations for analyzing the periodic changes in the community and users moving back and forth within multiple communities.

3. Corruption Barometer- Social Media platforms are gradually being used by government agencies to combat corruption by empowering citizens to report cases of bribery [202] [203]. Popular social media websites like Facebook and Twitter provides an effective and quick two-way communication between citizens and government wherein citizens can engage with the government. Our future work includes the automatic identification and extraction of useful information from bribery reports posted by citizens that can help the anti-corruption and law enforcement agencies in understating the extent of corruption in various departments, location and the volume of bribes being paid.

4. Fake News Detection- Prior literature shows that while there has been a lot of work done in the area of rumor, spam and phishing content detection, the spread of fake news has recently gained the attention of the researchers [204]. However, due to the lack of facts and ground truth, automatic identification of fake news is a technically challenging problem [205]. Our future work includes investigating the applications of contrast in contradictions, cross-platform data mining and parallel corpora for identifying the fake news on social media.

5. NLP Systems for Security Informatics- Existing NER tools such as Stanford CORE NLP Parser [91] and Indico [101] only extract the specific entities like people, location, money, date, time, and cardinal numbers. However, such NER tools are not able to capture the named entities when applied in security informatics domain where named entities have semantic meanings. For example, in a terrorist attack news, standard NER tools can identify the entities like a person, location, and organization. While the semantics of these entities such as 'defense organization', 'group who caused the attack', 'place where the attack happened' are not captured. Similarly, standard NERs can identify the cardinal number while the information like the number of people injured or died need to extracted. Since not all reports are structured and written by different teams; a dictionary based approach cannot be used to identify all relevant entities and requires a customized named entity recognizer for the government and security informatics based applications.

7.3.3 Modality and Dimensions

1. Multimedia Content- In the work presented in this dissertation, we have investigated the application of only contextual and linguistic metadata for building OSSMInt based applications. Social media websites allow users to post multimedia content (images with text, gifs, videos) to express their opinions and disseminate information. We believe that mining such content and extracting features from them can be used to improve the accuracy of classification. Mining such content includes the advanced techniques such as image processing [206] and optical character recognition [207] based methods.

2. Cross-Platform- to develop the proposed applications in this thesis, we have explored different and various social media platform for each case-study. However, we believe that despite the different structure and user behavior on the websites, analyzing data on multiple platforms can improve the performance of base approaches [208]. The future work includes the identification of similar or extended radicalized communities on multiple social media platforms.