Modeling The Persona In Persuasive Discourse on Social Media
Using Context-aware and Knowledge-driven Learning

by

Ugur Kursuncu

(Under the Direction of I. Budak Arpinar)

Abstract

Social media has reshaped communication in the last decade, supporting interaction and community development among participants who would never otherwise meet. It provides opportunities to users for sharing information and expressing their opinions on specific topics. Recent studies show that social media is immensely instrumental in changing, and measuring public opinions on particular issues. These open platforms bring the freedom to users to disseminate information for changing the public opinion and the normative behaviors of users through implementing a persuasive discourse on certain topics. While some accounts choose to share promotional information about their products to influence the public opinion, some other malicious-intended accounts share misinformation or propaganda to persuade others. In this research, we use marijuana and radicalization related

communications as focal cases, employing a context-aware and knowledge-driven approach for modeling the persona in these persuasive discussions on social media.

INDEX WORDS: Knowledge-infused Learning Context-aware Learning, Persuasive Discourse Analysis, User Modeling

Modeling The Persona In Persuasive Discourse on Social Media Using Context-aware and Knowledge-driven Learning

by

Ugur Kursuncu

B.S., Ankara University, Turkey, 2006

M.S., Stevens Institute of Technology, 2011

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2018

Modeling The Persona In Persuasive Discourse on Social Media Using Context-aware and Knowledge-driven Learning

by

Ugur Kursuncu

Approved:

Major Professor: I. Budak Arpinar

Committee: John A. Miller
Krys J. Kochut
Dilshod Achilov

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2018

Modeling The Persona In Persuasive Discourse on Social Media Using Context-aware and Knowledge-driven Learning

Ugur Kursuncu

December 2018

Anneme ve Babama

For Mom and Dad

Acknowledgements

I have been grateful that I was surrounded by wonderful people during the years of my PhD. Although I can only mention names of a few of them here, I would like to express my gratitude to each of every loved ones for their immense support that made this accomplishment possible.

I would like to thank my advisor, Dr. I. Budak Arpinar, for his guidance and advisory throughout my PhD studies, that further strengthened the quality of my research. You put a great confidence in me and supported me in each and every step of this steep and curvy road from the beginning. I would also like to thank each of my committee members, Drs. John A. Miller, Krys J. Kochut, and Dilshod Achilov, for their precious time and feedback for my research, that greatly impacted my learning experience at UGA.

I would like to provide my deepest gratitude to my mentor, Prof. Amit P. Sheth, at Kno.e.sis Center, Wright State University, for an immense impact on my self-growth. You have not only provided academic mentorship, but also shed the light on the way to success. Your mentorship in the last one and half years of my PhD has been greatly reflected on this particular accomplishment. I would also like

to thank my colleagues at Kno.e.sis, that I have been honored to have the opportunity to work with. Especially, Manas Gaur, Dr. Krishnaprasad Thirunarayan, Dr. Valerie Shalin, Dr. Raminta Daniulaityte and Amanuel Alambo: it has been pleasure to have been working and having insightful and fruitful discussions with you, that resulted high-quality research with our hard work.

Athens and the University of Georgia have been a special place to me for its authentic community and culture. I have met great people and made friends that I will remember forever, and I thank each one of them for making me truly feel home. Specifically, Jittery Joe's(JJ) east side where I had done my daily work and met great humans, will remain a special place to remember.

And my family... I have no way to thank my mom, dad and two brothers enough for their endless patience during these bumpy and long years. You deserved to be proud long before, and I am happy to have made this happen. Therefore, this dissertation is dedicated to you.

Contents

[1]This chapter is published as:
Ugur Kursuncu, Manas Gaur, Usha Lokala, Krishnaprasad Thirunarayan, Amit Sheth and I. Budak Arpinar. Predictive Analysis on Twitter: Techniques and Applications. Book Chapter in *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining*, Springer, 2018.

[2]To appear:
Ugur Kursuncu, Manas Gaur, Usha Lokala, Anurag Illendula, Thirunarayan Krishnaprasad, Raminta Daniulaityte, Amit Sheth, and I. Budak Arpinar. 'What's ur type?' Contextualized Classification of User Types in Marijuana-related Communications using Compositional Multi-view Embedding. *IEEE/WIC/ACM International Conference on Web Intelligence (WI) (2018).*

List of Figures

List of Tables

Chapter 1

Introduction

Social media has reshaped communication in the last decade, supporting interaction and community development among participants who would never otherwise meet. It provides opportunities to users for sharing information and expressing their opinions on specific topics. Recent studies show that social media is immensely instrumental in changing [Shirky, 2011], and measuring public opinions [Kursuncu et al., 2019] on certain topics. These open platforms bring the freedom to users to disseminate information or misinformation for changing the public opinion and the normative behaviors of users through implementing a persuasive discourse on certain topics. Persuasive messages are usually used in traditional media productions by companies to manage the public opinion of their brands through ads in TV, newspapers and online ads. Similarly, they choose to share promotional information about their products to influence the public opinion on social media platforms as well. On the other hand, malicious-intended organiza-

tions (e.g., terrorist groups) also take advantage of social media by sharing their propaganda and misinformation to persuade users and eventually recruit them into their ideology. Hence, persuasive discourse online can be a powerful tool to change opinion of masses on a critical issue, thus understanding the underlying content and interactional dynamics of such communications is crucial.

1.1 What is Persuasive Communication?

According to Miller [1980], persuasive communication is defined as any message that is intended to shape, reinforce, or change the responses of another or others. Based on his definition, the intention of the persuader and the response of the recipient are two important factors in a persuasive process. Stiff and Mongeau [2016] points out the *intention* in Miller's definition as a limitation. All communication can be considered as persuasive, since many activities might inadvertently affect other's responses regardless of the intention. Stiff and Mongeau [2016] also focuses on the *response* in Miller's definition as it puts emphasis on the outcome of the persuasive process such as perception of the source, emotions, beliefs, behavioral intentions and behaviors.

Considering social media platforms, a similar persuasive process is also implemented to change recipients' beliefs, attitudes and behaviors. On social media, such process is *intentionally* conducted by the persuader and the recipient is exposed to such information shared in the communication. Recipients may change their opinions, beliefs, attitudes and behaviors upon the information being con-

sumed indicating a positive outcome of the persuasive process. For example, companies use social media to promote their products and brands by sharing promotional information while media outlets utilize these platforms to share informative content. These communications serve as medium to shape public opinion on particular topics. Moreover, social media platforms have been intensely used in forming and evolving social movements that have had real-world impacts, such as arab springs in middle eastern countries, [Howard et al., 2011; Tufekci, 2014; Arpinar et al., 2016], gezi protests in Turkey [Haciyakupoglu and Zhang, 2015; Tufekci, 2014], and Russian influence in the US 2016 elections [Allcott and Gentzkow, 2017]. Individuals involved in these events by taking action upon communications on social media.

Different forms of persuasion include propaganda [Gass and Seiter, 2015] that is usually implemented by organizations or groups, towards changing general belief and behavior of masses in their ideology. We often see such processes on social media conducted by extremist organizations such as Islamist radical [cite] and far-right groups [cite].

To gain better understanding the underlying dynamics of such persuasive processes on social media requires modeling users as persuader and recipient by identifying characteristics of their content, network interactions and personal metadata.

1.2 Modeling the Persona

The term *persona* was first coined by a Swiss psychiatrist Jung [2014][1] defining as a kind of *mask*, designed on the one hand to make a definite impression upon others, and on the other to conceal the true nature of the individual. According to Jung, the persona is consciously formed as personality or identity through socialization, acculturation and experience, within a community. Information a person is exposed to and the interactions with other individuals in a community are main factors that shape the persona of an individual. However, as Jung pointed out the two states of the persona above; it might be just a mask that a person wears to make an impression on others, or to hide the true nature of the self as an expression of the collective psyche. The process of the formation of the persona is called *individuation*. This process involves various factors related to the individual's self and its interactions with people in her/his surroundings, and these factors need to be identified and factored in for a better understanding. Jung also defined archetypes of people as major characters based on patterns of behaviors in a community, and exemplified the archetypal figures as great mother, father, child, devil, god, wise old man, wise old woman, the trickster, the hero.

User modeling has been studied in the Human Computer Interaction community. In software development, characteristics of the targeted user audience who will use the product is an essential factor for making critical development-related decisions. For this purpose, persona have been modeled in development of software as abstract user type representations, so that requirements can be set and

[1] `https://en.wikipedia.org/wiki/Carl_Jung`

customizations to the product can be made for these user types accordingly [Junior and Filgueiras, 2005]. This approach requires the software development process to consider user requirements prior to the technical requirements. Normally, in such software development process, characteristics of users is identified through traditional methods such as interviews, to assess the usability of the product [Garrett, 2010]. Traditional user modeling techniques have been developed towards identifying certain characteristics of the population that is targeted within the development and business plan, and it includes, user roles, user profile, user segments, marketing segments, personas [Nielsen, 1994]. Constantine and Lockwood [1999] defines user role as "a collection of attributes that characterize certain user populations and their intentional interactions with the system". User profile is defined as highlighting the user's individual personal characteristics that includes information related to age, gender, skills, education, experience, and cultural level [Brusilovsky, 1996; Shneiderman, 2010]. User segments and marketing segments are used for marketing purposes (e.g., buyer persona, seller persona), to identify characteristics of user groups that will allow the companies assess the needs of those users in various segments.

Virtual Persona was defined by Cooper and Reimann [2003] as realistic representations of characters through collecting realistic representative information that can include demographic and biographical characteristics of the personality under modeling. Representations of persona also include a picture to make it more realistic. According to [Cooper et al., 1999; Cooper and Reimann, 2003], some of the important factors to create representations of persona are personal

information, technical information, relationship information, opinion information, and traditional technique to collect such information is interviews.

Hence, characterization of the persona and creating representations depend upon gathering relevant personal, opinions, and interactional information. Social media data such as Twitter, provides variety of data and metadata that can be leveraged to obtain or extract such information.

Users on social media share their certain information in their personal profile and content interacting with other users in their networks. Such information can be used to generate representations of personas on online platforms. For example, users on Twitter: (i) explicitly or implicitly define their personal attributes in their profile (e.g., their job, age, gender, personal interest) in textual (e.g., user description) and pictorial format(e.g., profile picture, emoji), (ii) share information or their opinions on particular topics in their content using text, images and emoji, (iii) interact with other users in their network by mentioning their handles in their tweets or retweeting their tweets. Such experiences of users on Twitter that we can organize as three layers of data above, forms persona, and we can leverage related information to generate representations and eventually model it on social media. In the subsequent subsection, we explain Person, Content and Network formally defined by Purohit et al. [2011], corresponding the three categories of information to model persona, and we call them as *views*. Moreover, we also explain *perspectives* to represent content in different contexts.

1.3 Person, Content, Network (PCN)

Purohit et al. [2011] defined the framework Person, Content and Network to gain better understanding of dynamics of user activities on social media for measuring the user engagement. Findings of their experiments to measure user engagement, show the need of incorporation of all PCN features in such analysis.

As discussed earlier, generating representations and modeling persona requires information corresponding to features Purohit's PCN. Therefore, similarly, incorporation of PCN features is critical in modeling persona as each component provides valuable information to generate proper representations. In this research, we operationalize the people-content-network paradigm Purohit et al. [2011] through compositional multiview embeddings Kursuncu et al. [2018] to model different user types in marijuana-related communications on Twitter. Our approach uses several building blocks for an in-depth analysis of tweet content to extract relevant context in marijuana dataset. The PCN framework provides a systematic organization of features as it will provide required information for modeling persona.

On social media, communities are being formed around various topics of interest through network interactions Purohit et al. [2011]. The information being shared in tweets by a user in the marijuana community displays an intent based on the user type Purohit et al. [2015]. For instance, *personal users* share their experiences and opinions on marijuana, whereas *retail accounts* usually promote the use of marijuana and other related products that they sell, and *media accounts* disseminate information on marijuana-related events and festivals, and legalization processes. Accordingly, as these personas show different characteristics, it is

critical to bring to bear different views, such as person, content, and network, for reliable analysis and insights. We describe a systematic organization and analysis of these features in Section 3.5.3.

1.4 Multimodality

The freedom of users to share information in different forms (e.g., text, image, emoji, interactions), provided by the platforms such as Twitter, Reddit and Facebook, also creates a rich multi-modal nature of social media. Therefore, retrieval of meaningful information from such heterogeneous content is critical for making sense of big social data and eventually modeling characteristics of users.

Users on social media select a collection of words, terms, phrases, images and emoji associated with their sentiments and emotions as reflection of their opinions on certain topics. They pick their profile pictures or use a description statement as they see fit into the perception of their preference. They also interact with other users with similar interests and characteristics. For example, as marijuana has been one of popular topics on social media due to ongoing legalization debates across the nation and legislative processes in certain states, users with different opinions on the issue of legalization of marijuana will show different sentiment and emotions in their content, interactions as well as their personal profile information. These users interact with each other forming a community for this particular topic. Capturing such communications and modeling their users in their networks requires customized retrieval techniques for information and learning techniques

for modeling.

1.5 Challenges

Conventional learning mechanisms detect target content from such social media data, permitting for example the analysis of public opinion. However, a certain class of detection problems–persuasive social data–challenges the state of the art. Although learning rich representations using relevant information extracted from social media data is essential for modeling users, certain challenges pose as obstacles in the ultimate goal of maximized performance. In a learning scheme, these challenges are: (i) Appropriate incorporation of multimodal data in the views of person, content and network, (ii) Ambiguity in the meaning of significant concepts in the content, (iii) Sparsity of important lexical and semantic cues in the domain-specific corpus, (iv) Noisy nature of social media data, that threatens performance of learning process, (v) Imbalance in a training dataset that is randomly selected representing each persona as some predefined persona can be minority in a population.

We address these challenges using (i) Marijuana-related communications, (ii) Islamist Radicalization communications. The use of social media to spread Islamist extremism and radicalization is one example of persuasive social data, extended over time and at least initially, cloaked by ambiguous intentionality. For instance, the concept "jihad" commonly appears in mainstream Islam, as well as radical discourse, albeit with a different context-dependent meaning. Contempo-

rary bottom-up analysis is ineffective in the face of such ambiguity and target sparsity, further challenged by a process of persuasion that starts out benign and over time turns increasingly radical. We model this process as the interaction among connected agents with a mix of perspectives and influence on each other, each one of which exemplifies a degree of radicalization and depends crucially on the proper identification of relevant message features. We infuse domain knowledge of Islamist radical ideology in deep learning models to relate linguistic features spanning religion, ideology, and violence to classify discourse along an established 5-level radicalization scale. Combined with a network of agent models, a carefully constructed sequence of discourse content persuades the primed recipient to descend into radicalism. Using Islamist extremism and radicalization as the focal use case, our knowledge-driven and context-aware learning approach generalizes to persuasive social data problems in other domains such as politics and economics.

We address fundamental *data science* challenges that are common to a particular set of data-related grand social problems, such as (Islamic) religious extremism, white supremacy, and trolling activities of oppressive regimes such as Russia and China. Although we only present examples from the religious extremism domain, the challenges and proposed solutions are very similar among this particular subset of problems, for which we coin the term *persuasive social data*, as online social (media) data is used to persuade individuals into a particular religious, racial or political doctrine.

Neither knowledge-graph or (deep) learning based methods alone can provide sufficient accuracy to address persuasive social data challenges. Traditionally, the

Knowledge-Base (KB) deep learning (DL) communities have worked in isolation from each other while tackling various analytical problems in scientific and social heterogeneous data sets including text, structured and multimedia data. Although KB and DL based approaches alone have demonstrated recent significant success especially in commercial domains such as speech recognition and autonomous vehicles, they have had minimal success in understanding and deciphering online human interactions. The persuasive social data challenge exhibits the following common characteristics, which prevent separate KB and DL approaches from reaching the level of success achieved in other domains mentioned earlier.

First, persuasive social data involves unconstrained doctrinal concepts and relationships with contextual meanings from religion, history and politics. For example, the concept of "Jihad" can mean (i) self-spiritual struggle, (ii) defensive war to protect lives and property from aggression, or (iii) act of (unprovoked) violence, depending on its contextual use. Classification of the first and second uses of Jihad as extreme or radical would be a grave mistake.

Secondly, actors in persuasive social data challenges frequently disguise themselves as true representatives of a religion, doctrine or ideology (e.g., radicals posing as true (mainstream) believers in Islam). This means persuasive (propagandist) data will be very similar to data produced by common agents with no hidden persuasive agenda, except they will contain concepts and relations that are twisted in their meaning, or presented out of context (e.g., Jihad) or sometimes outright misinformation. This leads to sparse true signals within the training data sets. For example, it is very difficult to distinguish social media posts from Russian trolls

disguised as American citizens during 2016 US presidential election. Further, propagandist data commonly show non-stationary patterns that dynamically change over time. For example, adherents of Islamic extremism have shifted their attention from promoting the caliphate established by ISIS to encouraging violence in the West recently. For this reason, the limited number of labeled instances available for training can often fail to represent the true nature of concepts and relationships in these persuasive social data sets.

Thirdly, a process of persuasion usually starts out benign and turns increasingly intense and radical over time. We model this process as the interaction among connected agents with a mix of perspectives and influence on each other, each one of which exemplifies a degree of radicalization and depends crucially on the proper identification of relevant message features. In our work, we measure the degree of radicalization (varying from vague support for extremism to violent extremism) and also capture the process of radicalization to understand the recruitment process better. This requires a more in-depth classification in which an agent's radicalization stage and timeline are also identified.

Based on these observations, we believe standard KB and DL only methods break down on persuasive social data and lead to misleading conclusions. In particular, it is easy to deduce or learn spurious concepts and relationships that look deceptively good on a knowledge-base or training and test sets, yet do not provide adequate results when the data set contains contextual and dynamically changing concepts and relations. In our approach, we infuse domain knowledge of radical ideology in deep learning models to relate features spanning religion, ideology and

violence, addressing domain-specific lexical and semantic challenges, such as sparsity, ambiguity and noise to classify discourse along a radicalization scale informed by political science.

Chapter 2

Predictive Analysis on Twitter[1]

With the growing popularity of social media and networking platforms as an important communication and sharing media, they have significantly contributed to the decision making process in various domains. In the last decade, Twitter has become a significant source of user-generated data. The number of monthly active users was 330 million as of third quarter of 2017, and the number of daily active users was 157 million as of second quarter of 2017. Moreover, nearly 500 million tweets per day are shared on Twitter. Accordingly, significant technical advancements have been made to process and analyze social media data using techniques from different fields such as machine learning, natural language processing, statistics, and semantic web. This amalgamation and interplay of multiple techniques within a common framework have provided feature-rich analytical tools [Purohit

[1]This chapter is published as:
Ugur Kursuncu, Manas Gaur, Usha Lokala, Krishnaprasad Thirunarayan, Amit Sheth and I. Budak Arpinar. Predictive Analysis on Twitter: Techniques and Applications. Book Chapter in *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining*, Springer, 2018.

and Sheth, 2013a; Davis et al.], leading to valid, reliable and robust solutions.

Twitter provides multimodal data containing text, images, and videos, along with contextual and social metadata such as temporal and spatial information, and information about user connectivity and interactions. This rich user-generated data plays a significant role in gleaning aggregated signals from the content and making sense of public opinions and reactions to contemporary issues. Twitter data can be used for predictive analysis in many application areas, ranging from personal and social to public health and politics. Predictive analytics on Twitter data comprises a collection of techniques to extract information and patterns from data, and predict trends, future events, and actions based on the historical data.

Gaining insights and improving situational awareness on issues that matter to the public are challenging tasks, and social media can be harnessed for a better understanding of the pulse of the populace. Accordingly, state-of-the-art applications, such as Twitris [Sheth et al., 2018] and OSoMe [Davis et al.], have been developed to process and analyze big social media data in real time. Regarding availability and popularity, Twitter data is more common than data from web forums and Reddit[2]. It is a rich source of user behavior and opinions. Although analytical approaches have been developed to process Twitter data, a systematic framework to efficiently monitor and predict the outcome of events has not been presented. Such a framework should account for the granularity of the analysis over a variety of domains such as public health, social science, and politics, and it has been shown in Figure 2.1.

[2]`https://goo.gl/Jo1h9U`

We discuss a predictive analysis paradigm for Twitter data considering prediction as a process based on different levels of granularity. This paradigm contains two levels of analysis: *fine-grained* and *coarse-grained*. We conduct fine-grained analysis to make tweet-level predictions on domain independent aspects such as sentiment, topics, and emotions. On the other hand, we perform coarse-grained analysis to predict the outcome of a real-world event, by aggregating and combining fine-grained predictions. In the case of fine-grained prediction, a predictive model is built by analyzing social media data, and prediction is made through the application of the model to previously unseen data. Aggregation and combination of these predictions are made from individual tweets form signals that can be used for coarse-grained predictive analysis. In essence, low-level signals from tweets, such as sentiment, emotions, volume, topics of interest, location and timeframe, are used to make high-level predictions regarding real-world events and issues.

In this chapter, we describe use of Twitter data for predictive analysis, with applications to several different domains. In Section 2, we discuss both processing and analytic techniques for handling Twitter data and provide details of feature extraction as well as machine learning algorithms. In Section 3, we explain a predictive analysis paradigm for Twitter that comprises two levels: fine-grained and coarse-grained. We also provide use cases, based on real-world events, of how coarse-grained predictions can be made by deriving more profound insights about a situation from social media using signals extracted through fine-grained predictions. We also describe common domain-independent building blocks that can serve as the foundation for domain-specific predictive applications. In Section 4,

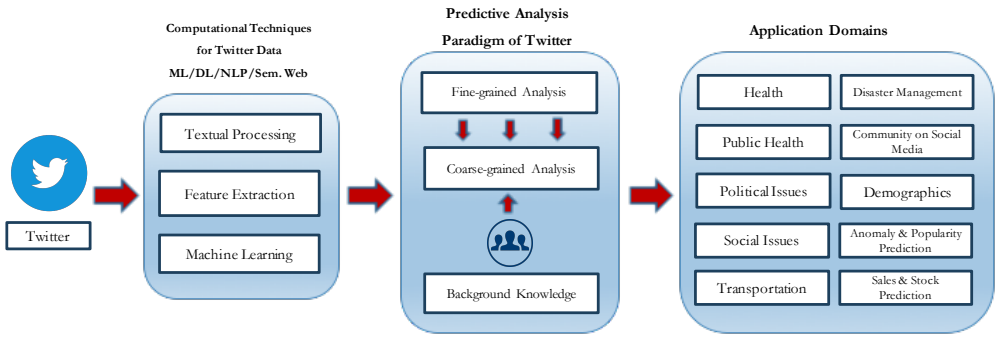


Figure 2.1: Overview of Predictive Analysis on Twitter Data.

we give further details on specific state-of-the-art applications of Twitter analytics that have been developed for different domains, such as public health, social and political issues. In Section 2.4, we conclude with a discussion of the impact of social media on the evolvement of real-world events and actions, challenges to overcome, for broader coverage and more reliable prediction. We also provide a comparative table relating techniques used with corresponding applications.

2.1 Language Understanding of Tweets

Novel processing and analysis techniques are required to understand and derive reliable insights to predict trends and future events from Twitter data due to their

unique nature – it contains slangs, unconventional abbreviations and grammatical errors as a matter of course. Moreover, due to the evolving nature of many events, may it be political, sports, or disaster-related, collecting relevant information as the event unfolds is crucial [Penuel and Statler, 2011; Malilay et al., 2014]. Overcoming the challenges posed by the volume, velocity, and variety of incoming social, big data is non-trivial [Wang et al., 2012b]. Sole keyword-based crawling suffers from low precision as well as low recall. For instance, obtaining tweets related to marijuana legislation [Lamy et al., 2017] using its street name "spice" pulls irrelevant content about "pumpkin spice latte" and "spice" in food. To improve recall without sacrificing precision, Sheth and Kapanipathi [2016b] provided a solution for adapting and enhancing filtering strategies that (a) obtains customized tweet streams containing topics of user interest [Kapanipathi et al., 2011] by constructing a hierarchical knowledge base by analyzing each user's tweets and profile information [Kapanipathi et al., 2014a], (b) selects and employs a domain-specific knowledge graph (e.g., using the Drug Abuse Ontology for opioid related analysis [Cameron et al., 2013]) for focus, and (c) reuses a broad knowledge graph such as DBPedia for coverage and generality. In Twitter data analysis, the processing phase includes natural language processing using techniques such as TF-IDF, word2vec, stemming, lemmatization, eliminating words with a rare occurrence, and tokenizing. On the other hand, some of the commonly used techniques, such as removal of stop-words, have proven ineffective. [Saif, 2017] has compared six different stop words identification methods over six different Twitter datasets using two well-known supervised machine learning methods and assessed the impact

of removing stop words by observing fluctuations in the level of data sparsity, the size of the classifier's feature space and the classifier performance. [Saif, 2017] concludes that in most cases that removing stop words from tweets has a negative impact on the classification performance.

2.1.1 Unique Nature of Tweets

Twitter's limit on the number of characters in a message encourages the use of unconventional abbreviations, misspellings, grammatical errors and slang terms. For instance, since a tweet was limited to 140 characters (until recent doubling to 280 character in December 2017), different sets of techniques and metadata have been considered to identify the best features to optimize the overall performance of the model being built. Due to the heterogeneous nature of the Twitter content, one can develop a variety of features [Wijeratne et al., 2017d] ranging from textual, linguistic, visual, semantic, network-oriented, to those based on the tweet and user metadata. Further, to handle tweet's textual data, the extracted features, techniques and tools [Sheth et al., 2018; Gimpel et al., 2011; Wagner et al., 2013] have been customized to exploit as well as being robust concerning misspellings, abbreviations, and slangs. Gimpel et al. [2011] addressed this problem in the context of part-of-speech (PoS) tagging, by developing a new tagset along with features specific to tweets, and reported 89% accuracy as opposed to Stanford tagger with 85% accuracy.

Tweets also include hashtags, URLs, emoticons, mentions, and emoji in their content. As these components contribute to the meaning of a tweet, it is imperative

that we incorporate them in the analysis, on a par with textual content.

Hashtags are meant to help in categorizing tweet's topics. They are frequently used to collect and filter data as well as for sentiment [Wang et al., 2011; Davidov et al., 2010; Kouloumpis et al., 2011], emotion [Wang et al., 2012b], and topical analysis [Romero et al., 2011; Morstatter et al., 2013]. Wang et al. [2011] used hashtags in their topical hashtag level sentiment analysis incorporating co-occurrence and literal meaning of hashtags as features in a graph-based model and reported better results compared to a sentiment analysis approach at the tweet level. In emotion analysis, Wang et al. [2012b] collected about 2.5 million tweets that contain emotion-related hashtags such as #excited, #happy, and #annoyed, and used them as the self-labeled training set for developing a high accuracy, supervised emotion classifier.

URL presence in a tweet is usually indicative the content being an index for a longer explanatory story pointed to by the URL. Researchers found URLs in a tweet to be discriminative in various studies such as sentiment analysis [Go et al., 2009; Agarwal et al., 2011], popularity prediction [Suh et al., 2010; Naveed et al., 2011], spam detection [Thomas et al., 2011]. They reported that the feature for URL presence in a tweet appeared as a top feature or has a substantial contribution to the accuracy of the model.

Emoticons (e.g., :), < 3) have been exploited by Liu et al. [Liu et al., 2012] in their Twitter sentiment analysis study, such as by interpreting ":)" as conveying positive sentiment and ":(" as conveying the negative sentiment. They used all tweets containing those emoticons as self-labeled training set and integrated them

with the manually labeled training set [Zhai et al., 2004]. They have achieved significant improvement over the model trained with only manually labeled data. Go et al. [2009], and other researchers [Boia and Faltings, 2013; Pak and Paroubek, 2010] conducted sentiment analysis on Twitter in 2009, and they found that they were able to achieve a better accuracy using models trained with emoticon data.

Emoji is a pictorial representation of facial expressions, places, food and many other objects, being used very often on social media to express opinions and emotions on contemporary issues of contentions and discussions. The use of emoji is similar to emoticon since they both provide a shorter means of expression of an idea and thought. The difference is that an emoji use a small image for the representation as opposed to emoticon that uses a sequence of characters. Kelly and Watts [2015] studied the use of emoji in different contexts by conducting interviews and found that the use of emoji goes beyond the context that the designer intended. Novak et al. [2015] created an emoji sentiment lexicon analyzing the sentiment properties of emojis, and they pointed that the emoji sentiment lexicon can be used along with the lexicon of sentiment-bearing words to train a sentiment classifier. On the other hand, Miller et al. [2016] found that the emoji provided by different platforms are not interpreted similarly. Wijeratne et al. [2017c] gathered possible meanings of 2,389 emojis in a dataset called EmojiNet, providing a set of words (e.g., smile), its POS tag (e.g., verb), and its definition, that is called its "sense." It associates 12,904 sense labels with 2,389 emojis, addressing the problem of platform-specific meanings by identifying 40 most confused emoji to a dataset.

2.1.2 Metadata for Tweet and User

There are mainly two types of metadata in a tweet object, namely, tweet metadata[3] and user metadata[4]. Tweet metadata contains temporal and spatial information along with user interactions and other information such as replies and language. On the other hand, user metadata contains information pertaining to the user that authored the tweet, such as screen-name and description. Some of the available metadata are described below.

Tweet Metadata

createdAt: This field contains the information on when the tweet was created, which is especially important when a time series analysis is being done [Varol et al., 2017b].

favoriteCount: The users on Twitter can like a tweet, and this is one way of interacting with the platform. The number of likes for a tweet has been used as a feature in various applications that includes trend detection [Varol et al., 2017b], identification of influence and popularity.

inReplyToScreenName: If this field of the tweet object is not null, it is a reply to another tweet, and this field will hold the username of the user that authored the other tweet. This information is valuable, especially to predict the engagement of the audience over an issue that tweets relate to, and to find influential users.

geoLocation: the Twitter platform has a feature that can attach the users' geolocation to the tweet, but this is up to the users to make it publically available.

[3]Tweet Object: `http://bit.ly/2QduwWd`
[4]User Object: `http://bit.ly/2JzEQVG`

Most of the users prefer not to share their geolocation.

retweet_count: Twitter allows users to repost a tweet by retweeting to their audience, and the original tweet holds this field to keep how many times this tweet has been retweeted. This information is useful to incorporate the prediction of popularity and trending topics.

User Metadata

description: This field holds the description of the account. As this metadata carries information on characteristics of the user, it is mostly used in user classification.

followers_count: This field holds the number of followers the user has, and as it is changeable information over time, the information located in a specific tweet may not be up to date.

friends_count: Twitter calls the accounts that a user follows as "friends," but it is also known as "followees." The numbers of followers and followees are used to determine the popularity of user and topics.

statuses_count: Twitter also calls tweets as "status," and in this case, status count refers to the number of tweets that a user has posted.

2.1.3 Network and Statistical Features

The users interact on the social networking platform Twitter with each other through follows, replies, retweets, likes, quotes, and mentions. Centrality metrics have been developed to compute and reveal users' position and their importance

based on their connections in their network. These centrality measures can help identify influential users. These metrics include in-degree, out-degree, closeness, betweenness, PageRank and eigenvector centrality. Closeness centrality is defined by Freeman [1978] as the sum of distances from all other nodes, where the distance from a node to another is defined as the length (in links) of the shortest path from one to the other. The smaller the closeness centrality value, the more central the node. Betweenness [Freeman, 1977] measures the connectivity of a node by computing the number of shortest paths which pass through the node. This aspect makes this node, a user in a Twitter social network, an essential part of the network as it controls the flow of information in the network. Therefore, removing this node would disconnect the network. EigenVector [Bonacich, 1987; Lawyer, 2015] metric measures the importance of a node based on the importance of its connections within the network. Therefore, the more critical connections a node gets, the more critical the node becomes. These metrics were used in a user classification application as features by Wagner et al. [2013] because of the intuition that similar users would have similar network connectivity characteristics.

Statistical features such as min, max, median, mean, average, standard deviation, skewness, kurtosis, and entropy can be computed for several data attributes [Varol et al., 2017b]. Machine learning determines a subset of these features that have the discriminative power necessary for particular applications and domains, especially for predicting user behaviors and user types [Pennacchiotti and Popescu, 2011a]. For instance, Varol et al. [2017b] extracted statistical features of a user, tweet, network. The statistical analysis was done over attributes

such as sender's follower count, originator's followee count, the time between two consecutive tweets, and the number of hashtags in a tweet. They conducted a time series analysis to predict if a trending meme is organic or promoted by a group. On the other hand, Pennacchiotti and Popescu [2011a] utilized statistical features to predict the type of users on social media based on their political leanings, ethnicity, and affinity for a particular business. As they classified users, they computed statistical characteristics of tweeting behavior of users such as average number of messages per day, average number of hashtags and URLs per tweet, average number and standard deviation of tweets per day.

2.1.4 Machine Learning and Word Embeddings

Machine learning algorithms play a crucial role in the predictive analysis for modeling relationships between features. It is well-known that there is no universal optimal algorithm for classification or regression task, and in fact requires us to tailor the algorithm to the structure of the data and the domain of discourse. Survey papers [Irfan et al., 2004; Nassirtoussi et al., 2014; Franch, 2013; Bravo-Marquez et al., 2012] and our comparative analysis (see Table 2.1) of related influential studies show what algorithms we found to perform well for various applications. As can be seen, this covers a wide variety – Random Forest, Naive Bayes, Support Vector Machine, Artificial Neural Networks, ARIMA and Logistic Regression.

Furthermore, deep learning (a.k.a advanced machine learning) enhanced the performance of learning applications. Deep learning is a strategy to minimize the human effort without compromising performance. It is because of the ability of

deep neural networks to learn complex representations from data at each layer, where it mimics learning in the brain by abstraction[5]. The presence of big data, GPU, and sufficiently large labeled/unlabeled datasets improve its efficacy. We discuss some of the applications that make use of deep learning for prediction task on social media in section 2.3.

Textual data processing benefits from the lexico-semantic representation of content. TF-IDF [Hong et al., 2011], Latent/Hierarchical Dirichlet Allocation (LDA/HDA) [Sokolova et al., 2016], Latent Semantic Analysis (LSA) [Dumais, 2008] and Latent Semantic Indexing have been utilized in prior studies for deriving textual feature representations. Mikolov et al. [2013b], they put forward a word embedding approach called Word2Vec that generates a numerical vector representation of a word that captures its contextual meaning incorporating its nearby words in a sentence. Training the word embedding model on a problem-specific corpus is essential for high-quality domain-specific applications, since the neighborhood set of words for an input term impacts its word embedding. For instance, pre-trained models of word2vec on news corpora generate poor word embeddings over a Twitter corpus. Wijeratne et al. [2016b] used word embeddings to further enhance the prediction of gang members on Twitter by training their model on a problem-specific corpus.

[5]How do Neural networks mimic the human brain?
`https://www.marshall.usc.edu/blog/how-do-neural-networks-mimic-human-brain`

2.1.5 Multi-modality on Twitter

Visual elements such as images and videos are often used on social media platforms. While users can attach images and videos to their tweets, they can also upload a profile image and a header image. Since the latter images are mostly related to the user's characteristics, personality, interest or a personal preference, these images are mostly used for classification of account type (e.g., media, celebrity, company), detection of user groups [Balasuriya et al., 2016; Wijeratne et al., 2016b] and identification of demographic characteristics (e.g., gender, age) [Sakaki et al., 2014]. Balasuriya et al. [2016] used the profile image of users in their feature set for finding street gang members on Twitter since gang members usually set their profile image in a particular way to intimidate other people and members of rival gangs. They retrieved a set of 20 words and phrases for each image through the Clarifai[6] web service to be used as features. As image processing is costly regarding time and computational resources required for training a model to retrieve information from images, it is usually preferred to use off-the-shelf web services that provide cheaper, yet effective alternative, for scalable social media analytics.

2.2 Prediction on Twitter Data

Gaining understanding about and predicting an event's outcome and its evolvement over time using social media, requires incorporation of analysis of data that may differ in granularity and variety. As tools [Gimpel et al., 2011; Bontcheva

[6] `https://www.clarifai.com`

et al., 2013] are developed and customized for Twitter, its dynamic environment requires human involvement in many aspects. For instance, verification of a classification process [Mitra and Gilbert, 2016] and annotation of a training dataset [Wijeratne et al., 2017c; De Choudhury et al., 2013; Lewenberg et al., 2015] are essential in the predictive analysis that can benefit from human expert guidance in creating ground truth dataset. Social media analysis in the context of complex and dynamic domains [Wang et al., 2012a; Ebrahimi et al., 2017; Vieweg et al., 2010] is challenging. Our approach to overcoming this challenge and dealing with a variety of domains is to customize domain independent building blocks to derive low-level/fine-grained signals from individual tweets. Then, we aggregate and combine these signals to predict high-level/coarse-grained domain-specific outcomes and actions with a human in the loop.

2.2.1 A Predictive Analysis Paradigm for Twitter

We consider predictive analysis on Twitter data as a two-phase approach: The first phase is fine-grained predictive analysis and the second phase is coarse-grained predictive analysis. An illustration of this paradigm is depicted in figure 2.2. The fine-grained analysis is a tweet-level prediction for individual signals, such as sentiment and emotions, about an event that is being monitored. This low-level prediction is made by building a predictive model that employs feature engineering and machine learning algorithms. Aggregating the tweet-level predictions for a specific time frame and location generates signals. For instance, a predictive model for sentiment predicts the sentiment of each tweet about an event in question as

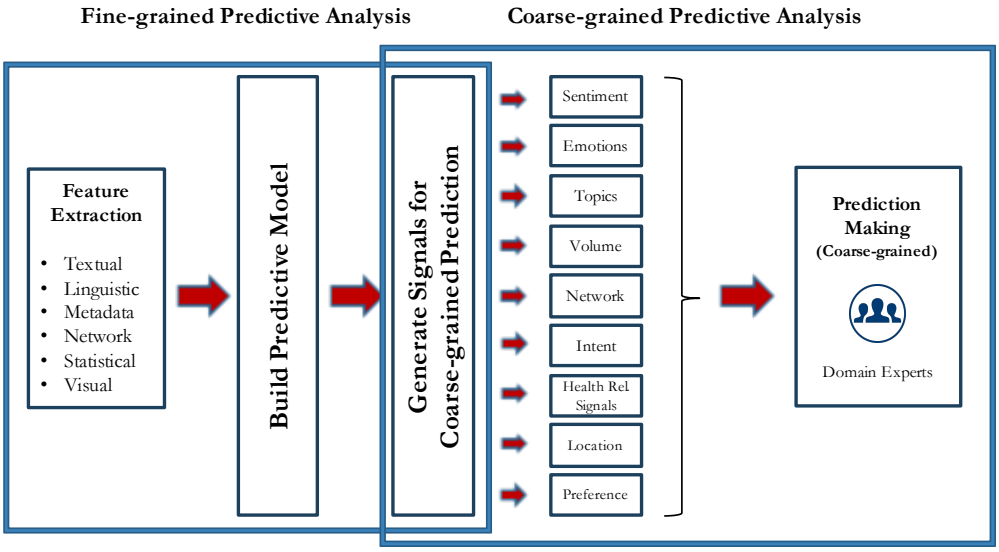


Figure 2.2: Two level Predictive Analysis Paradigm for Twitter.

negative (-1) neutral (0) or positive (+1), and we produce a signal between -1 and +1 for a particular location and time frame. A collection of such signals (e.g., emotions, topics) helps domain experts form insights while monitoring or predicting the outcome of an event, in their higher level analysis. Extraction of these signals is discussed further in subsequent section.

Coarse-grained analysis is a higher level prediction involving outcomes and trends of a real-world event, such as elections [Chen et al., 2012b], social movements [De Choudhury et al., 2016] and disaster coordination [Purohit et al., 2013, 2014b,a; Bhatt et al., 2014]. In this case, we gather the signals which we generated from the fine-grained predictions and make a judgment call for the outcome by

making sense of these signals in the context of the event and the related domain. Sentiment, emotions, volume, topics, and interactions between Twitter users can be considered as signals, while the importance and informativeness of each of these parameters may vary depending on the event and its domain. For instance, gauging the sentiment of a populace towards an electoral candidate would be very significant to predict the outcome of an election [Ebrahimi et al., 2017], but the same kind of information may not be as critical in the context of disaster management because, in the latter case, the sentiment may be largely negative. Further, for reliable decision making, the sentiment must be interpreted in a broader context. Predominantly positive sentiment towards democratic candidates in California is not as significant as that in Ohio. Similarly, the context provided by county demographics may be crucial in generalizing, predicting, and determining the outcome of an election. Moreover, temporal and spatial context plays an important role to understand the ongoing events better and obtain more profound insights. In US presidential elections, some states, called the "swing states" (as the electorate's choice has changed between Republican and Democratic candidates through the previous elections in these states), typically determine the eventual outcome of US elections. Therefore, narrowing down the analysis to the state level and gathering signals from these particular states would meaningfully contribute to the prediction of the outcome of the Presidential election and the future direction of the country.

In general, prediction analytics requires domain-specific labeled datasets created with the assistance of domain experts, and customization of feature space,

classification algorithm, and evaluation. Real world events have a dynamic nature in which critical happenings may change the course of discussions on social media. For example, breaking news about a candidate in an election may change the vibe in echo chambers of Twitter; thus, affecting the public opinion in one or another direction. For this reason, it is imperative to conduct the analysis accounting for essential milestone events happening during the process. Therefore, the analysis of such events would require an active learning paradigm that incorporates evolving domain knowledge in real-time.

2.2.2 Use Cases for Coarse-grained Prediction

Coarse-grained prediction requires taking into account many signals, and evaluating them concerning both present and historical context that varies with location and time frame. Importance of the signals in some domains and their related events may vary, and sole use of these signals would not be sufficient to make a reliable judgment call, although these signals are essential parameters in a real-world event context. For instance, an election usually whips up discussions on various sub-topics, such as unemployment, foreign policy; and necessitates proper cultivation of a diverse variety of signals following contextual knowledge of the domain [Ebrahimi et al., 2017]. We provide two use cases in this subsection to illustrate how a coarse-grained or high-level predictive analysis can be conducted.

US 2016 Presidential Election

During the 2016 US Presidential elections where "swing states" played a key role in determining the outcome, many polling agencies failed to predict it accurately[7][8]. On the other hand, researchers[9] conducted a real-time predictive analysis using a social media analytics platform [Sheth et al., 2018], making the prediction accurately before the official outcome was announced, by analyzing the state-level signals, such as from Florida and Ohio. Temporal aspect was also important in this use case to explain the evolution of the public opinion based on milestone events over the period of the election, as well as the election day because people tend to express, who they voted in the same day. They analyzed 60 million tweets by looking at the sentiment, emotions, volume, and topics narrowing down their analysis to state-level. On the election day, they focused on specific states such as Florida, which, before the election day, they predicted would be a pathway for Donald Trump to win the election[10]. In their analysis of Florida, volume and positive emotion (joy) for Trump was higher, whereas positive sentiment for Clinton was higher, eliciting report[11] such as "limited to professed votes from Florida until 1pm is not looking in her favor". Later in the day, the volume of tweets for Trump increased to 75% of all tweets based on the hashtag "#ivoted". Particularly in critical states of Florida, North Carolina, and Michigan, volume and positive emotions for Trump were significantly higher than for Clinton, although the sentiment was

[7] https://pewrsr.ch/2SGFxka
[8] https://goo.gl/mFtzvb
[9] https://goo.gl/AJVpKf
[10] https://goo.gl/sh7WNr
[11] https://goo.gl/iCqzk3

countering the overall signal. They made the call that the winner of Presidency and Congress as Donald Trump and the GOP respectively. While conducting this analysis [Ebrahimi et al., 2017], they noticed that the predictive model that they have built for sentiment signal was not successful due to the dynamic nature of the election with changing topics in conversations. A similar analysis was made for UK Brexit polls in 2012 by the same researchers, correctly predicting the outcome utilizing the volume and sentiment signals[12] [13] [14].

US Gun Reform Debate 2018

Researchers[15] monitored gun reform discussions on Twitter to predict the public support using the Twitris platform after the tragic shooting at a high school in Parkland, Florida, in February 2018. The public started demanding a gun control policy reform, and it has attracted the attention of legislative and executive branches of both state and federal governments. As polls measured the public opinion[16], researchers reported that the public support for a gun reform on social media was increasing over time since the Parkland shooting, confirming the overall outcome of these polls. They observed that reactions from public on social media in terms of the volume, sentiment, emotions and topics of interest, are strongly aligned with the milestone events related to this issue such as (i) POTUS' (President of the United States) meeting with families of the victims on February 21,

[12] https://goo.gl/i2Ztm6
[13] https://goo.gl/dFCGL9
[14] https://goo.gl/2EhSma
[15] http://blog.knoesis.org/2018/04/debate-on-social-media-for-gun-policy.html
[16] https://ti.me/2EYtD2B

(ii) CPAC(Conservative Political Action Conference) between February 22 and 24, (iii) POTUS' meeting with lawmakers on February 28 expressing strong support for a gun control policy change. These events significantly affected the public opinion on social media based on the aforementioned signals.

At the beginning of the gun reform discussions on social media, sentiment for pro-gun reform tweets was strong whereas the sentiment for anti-gun reform was relatively weak. However, the CPAC meeting changed the climate on social media, and it significantly boosted the momentum of anti-gun reform tweets, especially after the NRA (National Rifle Association) CEO Wayne LaPierre's speech in the morning of February 22[17]. Overall the volume of tweets for pro-gun reform was mostly higher than the anti-gun overhaul, except between February 22 and February 24[18], which covers the CPAC meeting where NRA CEO, VP Pence, and POTUS gave speeches. It surged the volume, positive sentiment and emotions in anti-gun reform posts radically, and those parameters for pro-gun reform posts dropped in the same manner. Effect of the meeting lasted a few days, and boycott calls for NRA and NRA's sponsors started to pick up in the meantime. After the meeting, sentiment for pro-gun reform tweets increased consistently, and the emotions expressed in pro-gun reform tweets became more intensified.

Emotions in anti-gun reform tweets were intense especially during and after the CPAC meeting, but later emotions in pro-gun reform tweets took over. Especially volume, positive sentiment, and emotions were overwhelmingly high right after the POTUS meeting with lawmakers on Wednesday, February 28, expressing his

[17] `https://goo.gl/kgbqWC`
[18] `https://goo.gl/LMFu3B`

support for a gun policy reform.

Furthermore, some of the most popular topics that users were discussing in their tweets included “midterm elections”, “parkland students”, “boycott the nra”, “stupid idea” and “individual freedoms”, where pro-gun reform arguments were expressed more frequently. The topic of “midterm elections” being one of the most popular topics on social media in gun reform discussions, also suggests that politicians from both Democrats and Republicans sensed the likely effect of this public opinion change on the midterm elections on November 2018. They have concluded in their predictive analysis that the public support for gun reform was significantly higher based on the signals they observed in the context of related events.

2.2.3 Extraction of Signals

We make predictions for the outcome of real-world events based on the insights we collect from big social data, and these insights are extracted as various signals such as sentiment, emotions, volume, and topics. The sentiment is a qualitative summary of opinions on a particular issue, and sentiment analysis techniques are utilized to extract such information computationally. The emotional analysis provides another stream of qualitative summary that is expressed by users about a particular event. The volume of tweets is an important signal about the engagement of the public in an event or an issue of consequence. Topical analysis is a process that extracts topics that contain particular themes in the domain of interest. We can produce and make use of more specific signals depending on the

domain such as preference, intent, and symptoms. The signals described below are commonly used parameters in higher level prediction tasks, and we describe related state-of-the-art applications and their technical details in the following.

Sentiment Analysis

Sentiment is one of the essential signals that can be used to measure the public opinion about an issue. As users on Twitter express their opinions freely, sentiment analysis of tweets attracted the attention of many researchers. Their approaches differ regarding the feature set, machine learning algorithm, and text processing techniques. Considering feature set, [Kouloumpis et al., 2011] used n-grams, POS-tags, emoticons, hashtags and subjectivity lexicon for sentiment analysis. For machine learning, Naive Bayes, SVM, and Conditional Random Fields (CRF) have been employed, and Naive Bayes has shown good performance [Pak and Paroubek, 2010]. Also, text processing techniques like stopwords removal, word-pattern identification, and punctuation removal have shown to improve sentiment analysis in [Davidov et al., 2010]. Nguyen et al. [2012] used time series analysis to be able to predict the public opinion so that the stakeholders on a stock market can react or pro-act against the public opinion by "posting balanced messages to revert the public opinion" based on the measurement that they performed using social media. Their objective was to use the sentiment change over time by identifying key features that contribute to this change. They measured the sentiment change regarding the fraction of positive tweets. They employed SVM, logistic regression and decision tree, and found that SVM and logistic regression provided similar

results outperforming the decision tree. They modeled the sentiment change overall twitter data and achieved around 73% F-score on sentiment prediction using time series analysis. [Stojanovski et al., 2016] employs a deep learning approach combining convolutional and gated recurrent neural network (CGRNN) for a diverse representation of tweets for sentiment analysis. Such a system was trained on GloVe word embedding created on a crawled dataset. The system was ranked among the top 10, evaluated using average F1 score, average recall, mean absolute error (MAE), Kullback-Leibler divergence (KLD), and EMD score [Esuli et al., 2015] for SemEval-2016 sub-tasks B, C, D, E. Exclusion of hand-crafted features and improved performance on SemEval 2016 shows the potency of the approach.

Emotion Analysis

Identification of emotions in tweets can provide valuable information about the public opinion on an issue. Wang et al. [2012b] predicted seven categorical emotions from the content of tweets using 131 emotion hashtags and utilizing the features such as n-grams, emotion lexicon words, part-of-speech tags, and n-gram positions. They used two machine learning algorithms: LIBLINEAR and Multinomial Naive Bayes. In a similar study, Lewenberg et al. [2015] examined the relationship between the emotions that users express and their perceived areas of interest, based on a sample of users. They used Ekman's emotion categories and crowdsourced the task of examining the users and their tweet's content to determine the emotions as well as their interest areas. They created a tweet-emotion dataset consisting of over 50,000 labeled tweet-emotion pairs, then trained a lo-

gistic regression model to classify the emotions in tweets according to emotion categories, using textual, linguistic and tweet metadata features. The model predicted a user's emotion score for each emotion category, and they determined the user's interest in areas such as sports, movies, technical computing, politics, news, economics, science, arts, health, and religion.

Topical Analysis

Topical analysis is one of the essential strategies under the umbrella of information extraction techniques that capture semantically relevant topics from the social media content [Griffiths et al., 2007]. Extraction of topics in the context of social media analysis helps understand the subtopics associated with an event or issue and what aspects of the issue have attracted the most attention from the public. As discussed in use cases for elections and gun reform debate, it is imperative to have the topics extracted from tweets for a better understanding of the underlying dynamics of relevant discussions. Chen et al. [2012a] associated topics of interest with their relative sentiment to monitor the change in sentiments on the extracted topics. Furthermore, utilizing the extracted topics as features for a supervised model improved the performance of the classification task in [Hong and Davison, 2010]. In [Zhao et al., 2011], researchers assessed quality of topics using coherence analysis, context-sensitive topical PageRank based ranking and probabilistic scoring function. This approach was used in a crime prediction application [Wang et al., 2012c].

Engagement Analysis

The volume is the size of the dataset that has been collected and indicates the user engagement on an event being monitored. In general, the larger the dataset, the better is the accuracy and consistency of a predictive model because it minimizes the possibility of bias. Engagement analysis enables human experts to improve their confidence in the learned representations/patterns for an accurate high-level prediction. However, while maintaining the sufficient size of the dataset to make reliable predictions from representative data is critical, data collection strategies need to be chosen strategically since relying solely on keyword-based crawling can bring in noise and irrelevant [Bhattacharya et al., 2017] data from a different context into the dataset. Therefore, a suitable filtering mechanism is essential for better quality data with high recall as well as precision. A semantic filtering mechanism [Sheth and Kapanipathi, 2016b; Phillips et al., 2017] as in the Twitris platform, can be implemented that selects and employs a domain-specific knowledge graph (e.g., using the Drug Abuse Ontology for related opioid analysis [Cameron et al., 2013]) for precision, and reuses a broad knowledge graph such as DBPedia for coverage and generality (see section 2). Thus, a significant and relevant dataset can be collected with high recall and precision that will allow one to obtain insights on the user engagement.

2.3 Applications

Twitter data has enabled researchers and analysts to deal with diverse domains ranging from healthcare, finance, and economy to socio-political issues and crisis management. Approaches to retrieve as much information as possible requires the inclusion of domain-specific features as well as the use of domain knowledge in the analysis. In this section, we provide a list of domains where predictive analysis applications on Twitter were implemented, along with the technical details. A comprehensive table is also included at the end to give a comparative overview of application domains, the features and machine learning algorithms being used, and their performance. The included applications were selected because they were state-of-the-art in their respective domains or had been influential. The applications that we describe in this section combine a variety of signals that can be the basis for coarse-grained predictive analysis. Since some of the applications in this section make use of the Twitris platform; therefore, we first provide background information about the platform. Purohit and Sheth [2013a] introduced the Twitris platform for citizen sensing that performs analysis of tweets, complemented by shared information from contextually relevant Web of Data and background knowledge. They describe it as a scalable and interactive platform which continuously collects, integrates, and analyzes tweets to give more profound insights. They demonstrate the capabilities of the platform with an analysis in various dimensions including spatio-temporal-thematic, people-content network, and sentiment-emotion-subjectivity, with examples from business intelligence including brand tracking, advertising campaigns, social/political unrests, and disaster

events.

2.3.1 Healthcare

Twitter data can be employed to shed light on many healthcare and disease-related aspects of contemporary interest, ranging from Alzheimer and dementia progression [Robillard et al., 2013] to eating disorders [Prieto et al., 2014] and mental health problems [Yazdavar et al., 2017; Coppersmith et al., 2015]. We focus on applications to glean depression in individuals or at a community level using self-reports about these conditions, their consequences, and patient experiences on Twitter.

Depression is a condition that a sizable population in all walks of life experiences in their daily life. Social media platforms including Twitter has been used to voluntarily express the mood changes and feelings as they arise. From these tweets, it is possible to predict whether a user is depressed or not, what symptoms they show as well as the reasons for their depressive mood. Some examples indicative of depression as expressed in tweets[19] include: "I live such a pathetic life.", "Cross the line if you feel insecure about every aspect of your life." ,"That's how depression hits. You wake up one morning afraid that you're going to live.", and "Secretly having a mental breakdown because nothing is going right and all motivation is lost.". These tweets epitomize the expression of emotional tumult that may underlie subsequent conscious actions in the physical world.

An interesting study by Yazdavar et al. [2017] explored the detection of clinical

[19]These tweets were modified before we share them in this chapter.

depression from tweets by mimicking the PHQ-9 questionnaire which clinicians administer to detect depression in patients. This study is different from traditional clinical studies that use questionnaires and self-reported surveys. They crawled 23M tweets over 45K twitter users to uncover nine significant depressive symptoms; (1) Lack of Interest, (2) Feeling Down, (3) Sleep Disorder, (4) Lack of Energy, (5) Eating Disorder, (6) Low Self-esteem, (7) Concentration Problems, (8) Hyper/Lower Activity, and (9) Suicidal Thoughts. A probabilistic topic model with a semi-supervised approach is developed to assess clinical depression symptoms. This hybrid approach is semi-supervised in that it exploits a lexicon of depression symptoms as background information (top-down) and combines it with generative model gleaned from the social media data (bottom-up) to achieve a precision of 72% on unstructured text.

De Choudhury et al. [2013] predicted the depression in an individual by exploiting their tweets. For ground truth dataset, they used, crowdsourcing to collect and label data. They utilized tweet metadata, network, statistical, textual and linguistic features, and time series analysis over a year of data to train an SVM model, obtaining an accuracy of 0.72.

The extraction of the location of people who experience depression using textual and network features can further assist in locating depression help centers. Do et al. [2017] utilizes a multiview[20] and deep learning based model, to predict the user location. The multi-entry neural network architecture (MENET) developed for location prediction uses words, the semantics of the paragraph (using doc2vec

[20] `http://www.wcci2016.org/document/tutorials/ijcnn8.pdf`

[Lau and Baldwin, 2016]), network features and topology (using node2vec [Grover and Leskovec, 2016]) and time-stamps to deduce user's location. They achieved an accuracy over 60% for GeoText[21], UTGeo11[22] and 55% for TwitterWorld [Bo et al., 2012]. Furthermore, MENET achieves an accuracy of 76% in region classification and 64.4% in state classification using GeoText dataset.

2.3.2 Public Health

Social media platforms including web forums, Reddit and Twitter, has become a venue where people seek advice and provide feedback for problems concerning public health. These conversations can be leveraged to predict trends in health-related issues that may threaten the well-being of the society. Moreover, caregivers have also seen these sources to be a game changer in its potential for actionable insights because of the information circulation. Particularly, cannabis legalization issue in the U.S. has been a trending topic[23] in the country as well as social media. Prior research on Twitter data analysis in this domain proved that it is an essential tool for epidemiological prediction of emerging trends.

Existing studies have involved identifying syntactic and statistical features for public health informatics, such as PREDOSE (PRescription Drug abuse Online Surveillance and Epidemiology) which is a semantic web platform that uses the web of data, background domain knowledge and manually created drug abuse ontology for extraction of contextual information from unstructured social media

[21] https://www.cs.cmu.edu/~ark/GeoText/README.txt
[22] http://www.cs.utexas.edu/~roller/research/kd/corpus/README.txt
[23] https://pewrsr.ch/2SAJuqV

content. PREDOSE performs lexical, pattern recognition (e.g., slang term identification), trend analysis, triple extraction (subject-predicate-object) and content analysis. It is helpful in detecting substance abuse involving marijuana and related products. Not only can it analyze generic marijuana but also its concentrates like butane hash oil, dabs, and earwax that are used in the form of vaporizers or inhalers. In a similar analysis of Twitter data, the marijuana concentrate use and its trends were identified in states where cannabis was legalized as well as not legalized. In 2014, utilizing the eDrugTrends[24] Twitris platform, researchers collected a total of 125,255 tweets for a two-month period, and 22% of these tweets have state-level location information [Daniulaityte et al., 2015]. They found that the percentage of dabs-related tweets was highest in states that allowed recreational or medicinal cannabis use and lowest in states that have not passed medical cannabis laws, where the differences were statistically significant. A similar study in 2015 [Lamy et al., 2016] reported adverse effects of Cannabis edibles and estimated the relationship between edibles-related tweeting activity and local cannabis legislation. Another study [Daniulaityte et al., 2015] was to automatically classify drug-related tweets by user type and the source of communication as to what type of user has authored the tweet, where the user types are defined as user, retailer and media. They employed supervised machine learning techniques incorporating the sentiment of tweets (e.g., positive, negative, neutral).

[24] `http://wiki.knoesis.org/index.php/EDrugTrends`

2.3.3 Political Issues

Political discussions on Twitter, which capture dynamic evolvement of public opinion, can directly impact the outcome of any political process. Arab Spring demonstrations [Howard et al., 2011; Tufekci, 2014; Arpinar et al., 2016] in the middle eastern countries, Gezi protests [Haciyakupoglu and Zhang, 2015; Tufekci, 2014] in Turkey, as well as US Presidential elections in 2016 involving influence peddling on several social media platforms [Allcott and Gentzkow, 2017] provide impactful illustrative examples. Researchers have explored user classification and profiling in the context of such political events on Twitter to predict the issue trends and eventual outcome.

Researchers [Pennacchiotti and Popescu, 2011a; Hoang et al., 2013; Makazhanov and Rafiei, 2014] used Twitter data to predict political opinions of users based on linguistic characteristics (e.g., Tf-IDF) of their tweet content. While classification of users based on their political stance on Twitter has been well studied, Cohen and Ruths [2013] have claimed that much of the studies and their datasets to date have covered very narrow portion of the Twittersphere, and their approaches were not transferable to other datasets. Pennacchiotti and Popescu [2011a] focused on the user profiling task on Twitter, and used user-centric features such as profile, linguistic, behavioral, social and statistical information, to detect their political leanings, ethnicity, and affinity for a particular business.

Moreover, prediction of dynamic groups of users has been employed [Chen et al., 2012b] to monitor the polarity during a political event by analyzing tweeting behavior and content through clustering. Usage of hashtags and URL, retweeting

behaviors and semantic associations between different events were key to clustering. 56% of the Twitter users participated in 2012 US Republican Primaries by posting at least one tweet, while 8% of the users tweeted more than 10 tweets. 35% of all users mostly retweet, separating them from the remaining. In terms of dynamic user groups, they formed the following bilateral groups: silent majority and vocal minority, high and low engaged users, right and left-leaning users, where users were from different political beliefs and ages. They analyzed these dynamic groups of users to predict the election outcomes of Super Tuesday primaries in 10 states. They also reported that the characterization of users by tweet properties (frequency of engagement, tweet mode, and type of content) and political preference provided insights to make reliable predictions. 8 weeks of data comprising 6,008,062 tweets from 933,343 users about 4 Republican candidates: Newt Gingrich, Ron Paul, Mitt Romney and Rick Santorum, was analyzed to assess the accuracy of predicting the winner. Prediction of user location using a knowledge base such as LinkedGeoData[25] in tweets also contributed to the election prediction. Furthermore, an error of 0.1 between the prediction and actual votes attest to the efficacy of the approach. Such a low error rate in prediction is attributed to original tweets (not retweets) from users who are highly engaged and right leaned.

2.3.4 Social Issues

Social issues and related events have been a part of discussions on Twitter, which gives opportunities to the researchers to address problems concerning individuals

[25] `http://linkedgeodata.org/About`

as well as the society at large. Solutions to such problems can be provided by measuring public opinion and identification of cues for detrimental behavior on Twitter by employing predictive analysis. We explain three problems and their respective solutions in this subsection.

Harassment

Harassment[26] is defined as an act of bullying an individual through aggressive offensive word exchanges leading to emotional distress, withdrawal from social media and then life. According to a survey from Pew Research Center[27], 73% of the adult internet users have observed, and 40% have experienced harassment, where 66% percent of them are attributed to social media platforms. Also, according to a report from Cyberbullying[28] research center, 25% of teenagers claimed to be humiliated online. While it is imperative to solve this problem, frequency and severe repercussions of online harassment exhibit social and technological challenges.

Prior work [Xu et al., 2012] has modeled harassment on social media to identify the harassing content which was a binary classification approach. However, in their predictive analysis, the context, network of users and dynamically evolving communities shed more light on the activity than pure content-based analysis. For instance, sarcastic communication between two friends on social media may not be conceived as harassment while the aggressive conversation between two strangers can be considered as an example of bullying. For reliably identifying and predict-

[26] http://bit.ly/2AG1gSC
[27] http://www.pewinternet.org/2014/10/22/online-harassment/
[28] http://cyberbullying.us/facts

ing harassment on Twitter, it is essential to detect language-oriented features (e.g., negation, offensive words), emotions, and intent. [Chen et al., 2012c] employs machine learning algorithms along with word embedding, and DBpedia knowledge graph to capture the context of the tweets and user profiles for harassment prediction.

Edupuganti [2017] focused on reliable detection of harassment on Twitter by better understanding the context in which a pair of users is exchanging messages, thereby improving precision. Specifically, it uses a comprehensive set of features involving content, profiles of users exchanging messages, and the sequence of messages, we call conversation. By analyzing the conversation between users and features such as change of behavior during their conversation, length of conversation and frequency of curse words, the harassment prediction can be significantly improved over merely using content features and user profile information. Experimental results demonstrate that the comprehensive set of features used in our supervised machine learning classifier achieves F-score of 88.2 and Receiver Operating Characteristic (ROC) of 94.3. Kandakatla [2016] presents a system that identifies offensive videos on YouTube by characterizing features that can be used to predict offensive videos efficiently and reliably. It exploits using content and metadata available for each YouTube video such as comments, title, description, and the number of views to develop Naïve Bayes and Support Vector Machine classifiers.The training dataset of 300 videos and test dataset of 86 videos were collected, and the classifier obtained an F-Score of 0.86.

Gang Communities & Their Members and Gun Violence

Gang communities and their members have been using Twitter to subdue their rivals, and identification of such users on Twitter facilitates the law enforcement agencies to anticipate the crime before it can happen. Balasuriya et al. [2016] investigated conversations for finding street gang members on Twitter. A review of the profiles of gang members segregates them from rest of the Twitter population by checking hashtags, YouTube links, and emojis in their content [Wijeratne et al., 2015]. In [Balasuriya et al., 2016], nearly 400 gang member profiles were manually identified using seed terms, including gang affiliated rappers, their retweeters, followers as well as followees. They used textual features of the tweet, YouTube video descriptions and comments, emojis and profile pictures to power various machine learning algorithms including Naive Bayes, Logistic Regression, Random Forest and Support Vector Machines, to train the model. Random Forest performed well for Gang and Non-Gang classification. It is interesting to notice that gang members usually make use of their profile images in a specific way to intimidate other people and members of rival gangs.

As gun control policies in big cities, such as Chicago, have changed over the years, the volume of the taunting and threatening conversations on social media has also relatively increased [Blevins et al., 2016]. Such conversations can be leveraged to assist law enforcement officers by providing insights on situational awareness as well as predicting a conflict between gang groups for a possible gun violence incident. Blevins et al. [2016] used a Twitter dataset that was manually labeled by a team of researchers with expertise in cyber-bullying, urban-based

youth violence and qualitative studies. Their strategy was to collect all tweets, mentions, replies, and retweets from a specific user profile between 29 March and 17 April 2014. Three experts developed the key types of content and used the work by Bushman and Huesmann [2006] to identify and categorize types of aggression. To overcome the challenge of recognizing special slang terms and local jargons in tweets as mentioned in Blevins et al. [2016], where they developed a part-of-speech (POS) tagger for the gang jargon and mapped the vocabulary they use to Standard English using machine translation alignment. They developed emotion classifier that uses the extracted POS tags, and Dictionary of Affect in Language (DAL) quantitative scores (Whissell, 2009) as key features. Ternary classification is applied to the whole dataset (TCF) and binary classification on the aggression-loss subset (BCS). Then they use a cascading classifier (CC), which uses two SVM models. Initially, one SVM model is used to filter the tweets into aggression/loss tweets, and all other tweets fall into the other category. After this filtration, only aggression/loss tweets is passed to second SVM model which is again a binary classifier for loss or aggression. So this Aggression Supervised classifier is able to categorize loss with 62.3% F-score and aggression with 63.6% F-score which beats the baseline model (Unigrams) by 13.7 points (aggression) and 5.8 points (loss) [Blevins et al., 2016].

2.3.5 Transportation

Congestion due to traffic is one of the prevalent problems in the United States (U.S.). Even after having structured rules that govern the flow of the traffic in

the U.S., congestion due to non-recurring activities still affects the schedules of people. However, the stationing of police officers to smooth the traffic is a probable solution, although it would not be long-term. Having an estimation of the flow of traffic in the advent of an event can help people to re-route their path to the destination. Leveraging social media and machine learning to estimate traffic is one such long-term solution that can be drafted for active traffic monitoring. Social media is flooded with posts from people about an event. Such posts can provide the location of the event or the tweeter, and it can be used along with other textual features to estimate the traffic flow. In [Ni et al., 2014], textual features, tweet and user-metadata such as text, hashtags, URLs, number of users and retweeted tweets were used by combining with live event data to predict traffic dynamics. They utilized autoregressive model, neural network, support vector regression, and K-nearest neighbor for traffic prediction. The evaluation was performed using mean absolute percentage error (MAPE), and root means square error (RMSE), with support vector regression (SVR) performing better over other regression models. SVR reduced the error in traffic prediction by 24% in terms of RMSE.

2.3.6 Location Estimation

Social media serves a vital role in times when people struggle to survive a disastrous event such as hurricane or earthquake, to provide solutions for assisting the public in recovery efforts. These solutions include identification of the demand and its location, and mapping the identified demands with suitable suppliers analyzing Twitter data.

In particular, location extraction plays a significant role in identifying the area that is impacted by a disaster as well as providing assistance [Krishnamurthy et al., 2014]. Mahmud et al. [2012] developed an approach to predict the location of users at the city level on Twitter combining several classifiers. They removed stop words, performed part-of-speech tagging, extracted hashtags, and extracted a feature called local term, a term used by local people to refer to the city. For detecting the local terms, several classification algorithms and found Naïve Bayes, SVM and Decision trees (J48) as the best performing algorithms. Al-Olimat et al. [2017a] developed a tool called LNEx (Location Name Extraction), that extracts the location from the tweet content by utilizing the OpenStreetMap [Haklay and Weber, 2008], GeoNames [Ahlers, 2013] and DBpedia [Lehmann et al., 2012] for disambiguation. The information retrieval process from the tweet is two-fold, which are toponym extraction and geoparsing. Toponym is a process to extract city and street names, points of interest, from unstructured text, tweets in particular for this study. Location names are usually abbreviated on Twitter; hence, a text normalization procedure is used for expansion of such brevity. For instance, tweets may contain "Rd" as an abbreviation, and it is normalized to "road". Furthermore, ambiguous location problems are resolved by employing the geoparsing procedure using the OpenStreetMap API[29]. LNEx improved the average F-Score by 98-145%, outperforming all the state of art taggers.

[29] `https://wiki.openstreetmap.org/wiki/API_v0.6`

2.3.7 Community on Social Media

People with distinct feelings, expression, solutions, and intelligence, share their opinions on Twitter. Such a diversified content can be related to elections, football game or a domain that is influenced by public views. With the abundance of textual data, one can envision the power of collective intelligence that can be harnessed for a wise recommendation, judgment and strategy building. Also, it is a known fact that a judgment call made by a crowd is superior to an individual's decision [Lee and Lee, 2017]. Formation of a diverse group can improve the decision-making process through what is known as Wisdom of Crowd (WoC). WoC is meant to minimize regional biases that may cloud objectivity associated with individual's judgment and bring together different perspectives and knowledge that can enhance coverage and comprehensiveness of the analysis. For example, WoC can be used to design a portfolio of stocks that maximize the profit in the stock market trading. However, no existing work illustrates the notion of WoC statistically and analytically. A methodological way for measuring the diversity of the crowd is crucial to the rise of human social engagement on social media. According to a recent survey from Pew Research Center, 76% of the American population is active on social media. It attributes success to a significant amount of online data and can aid in creating WoC of the social system. In [Bhatt et al., 2017a], fantasy premier league (FPL) is considered to exercise the better judgment of the diverse crowd. In their work, they predict the best performing team captain in the premier league, an element dictating the success of a team, based on the scores retrieved from the fantasy football and content of Twitter users. They utilized Word2vec

similarity measure to quantify the diversity of two groups of users during captain selection in FPL. Furthermore, They defined and validated their statistical objective scoring criteria to measure the quality of crowd judgment.

2.3.8 Demographics

In many applications, demographic information is a key to analysis that depends on different segments of the population concerning age groups, ethnicity, and gender. For example, age is critical for understanding drug abuse, while gender is critical to understand vulnerability to depression. Twitter in its current state does not require users to provide any demographic information.

Age Estimation using social media

Researchers developed a machine learning system coupled with the DBpedia knowledge graph utilizing the user follower-followee networks to predict the most probable age of a Twitter user, in [Smith and Gaur, 2018]. They gathered pre-identified famous people from DBpedia, based on their occupations and areas of interest, which also included their birth dates. Then they extracted a sample of 23,120 users who are in one/two hops of follower-followee network of famous people. Some of the user profiles were spam/bot and hence they were removed. Then they selected 16K users among the followers of the top 50 famous figures as their training set and 8K as their testing set. They achieved 84% accuracy in predicting the age of these users. They selected Support Vector Regression (SVR) with K-Fold Cross Validation [Refaeilzadeh et al., 2009] as their best performing model after evaluat-

ing using Linear Regression [Nguyen et al., 2011], Least Absolute Shrinkage and Selection Operator (LASSO) [Chen et al., 2010], and ElasticNet [Culotta et al., 2016].

Zhang et al. [2016b] studied the problem of age prediction on Twitter, using SVM and least square optimization algorithm in building the model. They utilized various features such as linguistic, textual, and network, to improve their model, achieving an F1 score of 0.81. They discovered that the characteristics of users in the same age groups have similar content and interactions between each other. On the other hand, Nguyen et al. [2013] investigated the relationship between the language used in tweets of a user and his/her age. They annotated the dataset that was collected following a guideline formed based on the tweet content of users in different age groups such as explicit or implicit age or life stage mentions. They found that the language use of people in same age groups is similar regarding the word and phrase selections as well as the topics that they are talking about. For instance, the following two sets of words, "school, son, daughter, wish, enjoy, thanks, take care" and "haha, xd, internship, school" have been used by users in two different age groups. In their analysis, they used linear and logistic regression models with unigram feature only, achieving an F1 score of 0.76.

Gender Estimation using Twitter

Estimation of the gender of a twitter user is beneficial to the analysis of Twitter data for health-related, drug abuse, and harassment activities. Existing approaches utilized statistical features [Bamman et al., 2012] and seldom involved

background knowledge along with social information. In [Li and Dickinson, 2017], a dataset from Sina Weibo, which is a counterpart of the micro-blogging platform Twitter, in China, was used to assess their methodology for gender prediction. [Li et al., 2014] exploits online behavioral and textual features and choice of vocabulary for each user. Online behavioral features include the number of fans, attention, messages, comments, forwards and a ratio of original/forwarded messages. Textual features include hashtags, URLs, emoticons, and sentence length. They also made use of username and pictures in content. Lexical features were extracted from the content using TF-IDF. They used four algorithms for predicting gender: Decision Tree, Naive Bayes, Logistic Regression and Support Vector Machines (SVMs), and found that SVM outperformed other classifiers by attaining accuracy of 94.3%.

2.3.9 Anomaly & Popularity Prediction

Twitter has become a playground for spammers. While public conversations on Twitter are diverse and challenging to analyze and summarize, spammers and bots further complicate the reliability of the outcome. Bots are automated software that is programmed to post a predefined content. They are being used mostly to propagate or promote bias and skew votes in politics, views on social issues, or provide impetus to promotional campaigns. On the other hand, prediction of the popularity of trending topics or issues requires robust analysis that takes into account anomalous accounts.

Thomas et al. [2011] collected 1.8 billion tweets sent by 32.9 million users and manually identified 1.1 million suspended accounts as spammer accounts along with 80 million anomalous tweets. They used user behavior regarding interactions with other users, public Twitter handler service usage and textual features of tweet content such as shortened URLs created using free web hosting services. Volkova and Bell [2017] also studied this problem by applying a deep learning technique, Recurrent Neural Networks (RNN), using tweet metadata and network features. They compared their approach with state-of-the-art machine learning methods such as log-linear models. Their RNN model outperformed all the machine learning models built using various combinations of features with 0.95 F1 score. Sentiment has also been used in spam detection works [Dickerson et al., 2014; Varol et al., 2017a] as a feature to detect bots on Twitter. [Varol et al., 2017a] also studied the detection of online bots on Twitter, and utilized Random Forests, AdaBoost, Logistic Regression and Decision Tree algorithms. They found Random Forest classifier achieved the best performance with 0.95 AUC score. They made use of sentiment features that they extracted from the text beside tweet and user metadata, textual, linguistic and network features.

Castillo et al. [2011] have investigated the tweet credibility issue in the news disseminated on the platform. They crowdsourced the task of evaluating the credibility of each tweet to determine if it has newsworthy topics, labeling each tweet using automated credibility analysis. Labels given by crowdsourcing process were used in the training phase. They used SVM, decision trees, decision rules and Bayesian networks, and best results were given by J48 decision tree, achieving

an 86% F1 score. Ross and Thirunarayan [2016] created a robust and general feature set for learning to rank tweets based on credibility and newsworthiness. In previous works by Gupta et al. [2014]; Gupta and Kumaraguru [2012]; Gupta et al. [2013], they have demonstrated that when the training and testing data are from two distinct time periods, the ranker performs poorly. Ross et al. [Ross and Thirunarayan, 2016] improved upon this by creating a feature set that does not overfit a particular year or a set of topics, which is critical for robust analysis of social media over time and across different domains.

Varol et al. [2017b] conducted a time series analysis to predict if a trending meme is organic or promoted by a group. They aimed to predict meme's that have potential to trend before it becomes trending; therefore, the task of predicting trends is naturally forced to utilize a sparse dataset. For this reason, they had to reliably extract textual, linguistic, tweet and user metadata, network and statistical features, from a small dataset. They used three learning algorithms namely, K-Nearest Neighbor (KNN) with Dynamic Time Warping (KNN-DTW), Symbolic Aggregate approXimation with Vector Space Model (SAX-VSM) and KNN. KNN is a machine learning algorithm for classification and DTW for multi-dimensional time series. They found KNN-DTW and KNN showed the best performance in prediction. They used AUC as evaluation metric to measure accuracy because it is not biased by the imbalance in classes (e.g., 75 promoted trends versus 852 organic ones). Weng et al. [2014] studied the prediction of the popularity of meme on Twitter. They relied mostly on network features besides tweet and user metadata, using random forest and linear regression. They extracted 13 features such as some

early adopters, average shortest network path length between users, the diameter between users, and the number of infected communities. They built their model using random forest and tested against five different baselines that used linear regression along with different combinations of the 13 features. Their model achieved 0.85 F1 score, outperforming the baselines. Kobayashi and Lambiotte [2016] predicted the popularity of a tweet in terms of the number of retweets in a time window in the future. They used time series analysis using a method called time-dependent Hawkes process (TiDeH) calculating infectious rate and using tweet and user metadata such as temporal information from a tweet and number of followers of a user. They evaluated their system against other existing methods that incorporated linear regression and Poisson process and reported that it outperformed other approaches achieving around 5% mean error rate. Tsur and Rappoport [2015] also studied the popularity of hashtags on Twitter, through linguistic features of the tweet text, specifically hashtags. They obtained promising results using a modified version of Gradient Boosted Trees called Gradient Boosted rank. They compared their approach with SVM and Least-effort algorithms, obtaining 0.11 mean error rate. Ruan et al. [2012] predicted the volume of tweets, analyzing the user behavior on individual as well as collective level. Besides tweeting activity and content analysis of users, they utilized the underlying follower-followee network, user network structure, neighboring friends' influence and user past activity as features. They used linear regression model with multiple features that include network structure, user interaction, content characteristics and past activity, and found that combining features yields the best performance.

Table 2.1: Comparative Analysis of Applications and their Evaluation. Acronyms for Algorithms and Features are described in Table 2.3

Ref. & Evaluation	**Application**	**Algorithms**	**Features**
[Pennacchiotti and Popescu, 2011a] F1=0.88 [Zhang et al., 2016b] F1=0.81 [Pattisapu et al., 2017] F1=0.59 [Gilani et al., 2017] F1=0.83 [Balasuriya et al., 2016] F1=0.77 [Alowibdi et al., 2014] Acc=0.82 [Hoang et al., 2013] Acc=0.92 [Makazhanov and Rafiei, 2014] F1=∼0.75 [Nguyen et al., 2013] F1=0.76 [Lewenberg et al., 2015] AUC=∼0.7 [Wagner et al., 2013] AUC=0.8 [Smith and Gaur, 2018] Acc=0.84	User Profiling User Classification	SVM LinR CNN RF NB LogR, LASSO	UsM, TwM Ling, Nw, Stat, Txt Vis
[De Choudhury et al., 2013] Acc=0.72 [Mahmud et al., 2016] F1=0.62	User Attitude, Personality, Mood Prediction	SVM, NB, RF	TwM, Txt, Nw, Ling, Stat
[Georgiev et al., 2014] AUC=0.8	Sales & Stock Price Prediction	NB, RF, SVM	TwM
[De Choudhury et al., 2016] Med error=0.32 [Yang et al., 2017]] F1=0.58 [Korolov et al., 2016] Acc=0.85 [Kallus, 2014] AUC=0.91	Social/Political events, Elections, Collective action	PosR NBR, SVM, LogR, CNN, RF	TwM, UsM Txt, Ling, Nw
[Varol et al., 2017b] AUC=0.95 [Weng et al., 2014] F1=∼0.85 [Kobayashi and Lambiotte, 2016] Mean err=∼0.179 [Tsur and Rappoport, 2015] Mean err=∼0.11	Popularity prediction	KNN-DTW, SAX-VSM,RF, TiDeH, KNN, LinR, GrB,SVM,LogR	Txt Nw and Stat, TwM and UsM Ling
[Thomas et al., 2011] Acc=0.89 [Varol et al., 2017b] AUC=0.95 [Dickerson et al., 2014] AUC=0.73 [Volkova and Bell, 2017] F1= 0.95 [Echeverria and Zhou, 2017] F1=0.99 [Varol et al., 2017a] AUC=0.95 [Castillo et al., 2011] F1=0.86	Spam bot detection Troll detection Credibility prediction	KNN-DTW, KNN,LinR,DT NB, GrB RF, AdB, BNet, RNN, SVM, LogR SAX-VSM	Text, Nw, Ling,TwM,UsM Stat, Vis

Table 2.2: (Continued from Table 2.1) Comparative Analysis of Applications and their Evaluation. Acronyms for Algorithms and Features are described in Table 2.3

Ref. & Evaluation	**Application**	**Algorithms**	**Features**
[Davidov et al., 2010] Pre=∼0.80 [Wang et al., 2011] F1=0.77 [Kouloumpis et al., 2011] F1=0.67 [Agarwal et al., 2011] F1=∼0.60 [Pak and Paroubek, 2010] F0.5=0.62 [Go et al., 2009] Acc=0.82 [Wang et al., 2012b] Acc=0.61 [Gao and Sebastiani, 2015] Mean error=0.071 [Hassan et al., 2013] Acc=∼0.71 [Kothari et al., 2013] F0.5= 0.76 [Nguyen et al., 2012] F1=∼0.73 [Liu et al., 2012] F1=0.79	Sentiment analysis Emotion Detection	NB,SVM,CRF, LibLin, AdB, RT, REPTree, BNet, LogR RBF-NN	Ling, Txt, TwM and UsM
[Mahmud et al., 2012] Acc=∼0.83 [Georgiou et al., 2015] Mean error=0.39 [Aiswal et al., 2013] Acc=0.88 [Rout et al., 2013] Acc=0.79	Location Estimation Traffic Estimation	NB, SVM, DT, LinR, MxEnt	Ling,TwM,UsM Txt
[Rath et al., 2017] F1= 0.70 [Bizid et al., 2015] Pre=0.86	Finding Key Users, Community Detection	RNN, SVM	Nw, UsM, Txt
[Ferrara et al., 2013] LFK-NMI=0.13 [Yamamoto and Satoh, 2015] F1=0.63	Topic Extraction, Meme Extraction	HiCl, KM, LDA, SVM	Txt,Ling,TwM, UsM, Nw
[Beykikhoshk et al., 2014] Acc=0.84 [Yin et al., 2016] AUC=0.83 [Daniulaityte et al., 2016] F= 0.8736 [Lamy et al., 2016] K Alpha=0.84	Public Health Health-care	NB, SVM, RF LDA, ssToT	Txt,Ling,TwM Sentiment
[Al-Olimat et al., 2017a] F1=0.81 [Hu et al., 2015] R2= 0.67	Disaster Management	LogR	Txt,Ling,UsM, TwM, Nw

2.3.10 Sales & Stock Price Prediction

As social media, particularly Twitter, users share their satisfaction or frustration with products on the platform, these user reviews can be exploited by companies to generate actionable insights to meet customer expectations and eventually provide better quality products and services. Industrial applications of predictive analysis of social media have been gradually adopted, to gain the understanding of the market. Some of the use-cases that have adopted social media data for decision making are for:

1. Improvement of Customer Service: Delta Airline exploited social media to identify the reasons for customer frustration. For instance, lost luggage or poor service.

2. New Products Research and Development: JD Power quality assessment has determined that car company modify car seats based user sentiments on the social sphere [Kessler et al., 2010].

3. Key Influencers: A cosmetics company L'Oreal uses social media follower-followee network to find Influencers for promotions[30].

4. Recommendations through deep learning: YouTube utilizes the deep neural network to enhance their recommendation system using implicit feedback by analyzing users' comments and videos of interest [Covington et al., 2016].

Georgiev et al. [2014] investigated the question of how the Olympic Games impacted the sales of businesses in London. They used Twitter posts along with the

[30]http://bit.ly/2zkRfZ3

check-ins through Foursquare platform to extract mostly location-specific features from tweet text and tweet metadata, such as the distance of businesses to stadiums and sponsor businesses, transitions to entertainment places and social areas. They evaluated their work using AUC, for Naïve Bayes, Random Forest, and SVM algorithms and reported that SVM performed best with 0.8 of AUC.

Korpusik et al. [2016] employs feed-forward network (FF) for predicting the likelihood of a customer to buy a product. They restricted their dataset to tweets about mobile phones and cameras, expensive products that people often buy after doing some research online. Before predicting the likelihood of purchasing a product, they predicted whether a tweet represents the respective user's purchasing behavior. Then they predict whether the user will purchase the product after 60 days time window of tweeting. They compared the performance of their approach with Long Short Term Memory (LSTM), Recurrent Neural Networks (RNNs) (with varying dropout rates) based implementation and observed that their approach with FF surpasses others by small margins. FF learning cycle involved RMSprop [Tieleman and Hinton], sigmoid activations and negative log-likelihood function with batch training.

2.4 Conclusion

Twitter has positioned itself as an essential part of the social media environment becoming an emergent communication medium. This development has opened up new opportunities for researchers to gauge the pulse of the populace reliably

and use that to study public opinion, form policies, understand the impact of events, and find newer ways to address certain problems. Social media data has already enabled researchers to predict the trends and outcomes of several critical real-world events, and its reliability and coverage can further be improved by incorporating background knowledge [Tufekci, 2014; Morstatter et al., 2013]. Specifically, monitoring the engagement and public opinion about ongoing events from temporal and spatial perspectives can foretell their evolution as well as the outcome. Moreover, this information can complement traditional surveys or polls that are conducted by non-government agencies to improve our confidence in the prediction, as traditional methods alone can be misleading or sluggish in reacting to rapidly changing events. In order to account for the complex decision making that requires consideration of a number of factors that can impact a situation or an event, incorporation of as many signals as possible in comprehending the big picture is necessary. We have explored a predictive analysis paradigm that comprises two levels of prediction, using coarse-grained analysis built upon fine-grained analysis. Such analysis have been conducted with creditable success for events such as elections, gun violence, drug misuse or illicit drug use [Sheth et al., 2018].

In this chapter, we have discussed processes, algorithms, and applications concerning predictive analysis in different domains. We illustrated fine-grained analysis by customizing domain-independent approaches to extract signals such as sentiment, emotions, and topics through the application of machine learning models, and coarse-grained analysis by aggregating and cultivating the signals to make

predictions. We have also provided details of related prominent studies in ten different domains such as healthcare, public health, political and social issues, disaster management, sales and stock prediction, and demographics. The following table summarizes related work describing various applications and methods used.

Table 2.3: The Acronyms used in the comparative table.

Acronym	**Algorithm Description**	**Acronym**	**Feature Description**
LinR	Linear Regression	UsM	User metadata
RF	Random Forest	TwM	Tweet metadata
NB	Naïve Bayes	Ling	linguistic
LogR	Logistic Regression	Nw	Network
PosR	Poisson Regression	Stat	Statistical
NBR	Negative Binomial Reg.	Txt	Textual
GB	Gradient Boosting	Vis	Visual
AdB	AdaBoost		
DT	Decision Trees		
BNet	Bayes Net		
LibLin	LIBLINEAR		
HiCl	Hierarchical Clustering		
KM	K-Means		
RT	Random Tree		

Chapter 3

Use Case 1: User Modeling on Marijuana Communications[1]

3.1 Introduction

"It's 4/20, and that means everyone is talking about marijuana[2]*,"* highlights the state of marijuana-related communication on Twitter, especially around the time marijuana legalization polls were conducted in the USA. As more evidence is gathered through research studies on the safety and benefits of the medical and recre-

[1]To appear:
Ugur Kursuncu, Manas Gaur, Usha Lokala, Anurag Illendula, Thirunarayan Krishnaprasad, Raminta Daniulaityte, Amit Sheth, and I. Budak Arpinar. 'What's ur type?' Contextualized Classification of User Types in Marijuana-related Communications using Compositional Multi-view Embedding. *IEEE/WIC/ACM International Conference on Web Intelligence (WI) (2018).*

[2]https://goo.gl/JGSs3X

ational uses of cannabis, there is a rise in public demand for broader legalization of marijuana and its variants. Accordingly, it is useful to study the engagement of users on social media to understand public opinion and its influence on policies better.

Characterization of marijuana concentrate users on social media can enable researchers to describe the patterns of use, reasons for symptoms, risk factors, and side effects using spatio-temporal analysis. Specifically, classification of user types into retail, informed agency and personal accounts, using marijuana communications on social media can aid in selectively analyzing their content-network dynamics. Focus of analysis can include assessing homophily in these communities, differences concerning their marijuana conversations, the information flow, and interactions between user types. This can eventually help better situate their characteristics and understand the implications. For instance, in the case of predicting the outcome of a state legalization process [Kursuncu et al., 2019], understanding public opinions of the residents towards marijuana related topics are critical as these opinions translate to votes. We associate personal user type (P) with an account handled by an individual user expressing their opinions, retail user type (R) with an account managed by a business entity to promote and market marijuana-related products, and informed agency user type (I) with an account handled by a group or organization to disseminate marijuana related information. Throughout the paper, we use informed agency and media interchangeably.

In this study, we make three key contributions: (i) Model the multiview aspect of the Twitter data and features through people, content, and network dimensions.

(ii) Exploit multimodal and diverse data on Twitter such as text, image, emoji and network interactions for effective user classification task. (iii) Extensively leverage context with the help of multimodality and multiple views of the social media data, and derive comprehensive representation from the three orthogonal views (People, Content, Network) through Compositional Multiview Embedding (CME).

The multimodality stems from the inclusion of text, image (profile pictures), emoji and network interactions between accounts pertaining to different user types [Benton et al., 2016]. Hence, for a reliable classification, we create compositions of vector embeddings for these views of the Twitter data, called *Compositional Multiview Embedding* (CME) as it can represent the context coherently [Mitchell and Lapata, 2010]. In our approach, we create two CMEs: (i) one using tweet text, emoji and network interactions of users, and (ii) another using user description and emoji. To assess which combinations of features can be utilized in generating the CMEs, we performed correlation analysis, as explained in Section 3.5.2. For instance, we found that descriptions and network interactions of users are highly correlated, suggesting that their combination can affect the performance of the classifier over the validation and test data. Therefore, we did not create the embedding using these two views. We evaluated the classifiers based on the individual F-scores of user type classes. We also generated word embedding vectors for profile pictures of users, which significantly improved the performance of classification of the informed agency user type. Details of our approach and results are discussed in Sections 3.5 and 3.6 respectively.

The remainder of the paper is organized as follows: Section 3.3 provides pre-

liminaries about the concepts and technologies that are used. Section 3.4 provides an exploratory analysis that includes statistics on our dataset. Section 3.5 explains features and our experimental setting, and Section 3.6 discusses the results of our analysis. Section 3.7 concludes the paper with a summary and future research directions.

3.2 Related Work

In this section, we describe prior studies that are broadly related to user classification, under three prominent sub-headings: (i) Embedding based Approaches to User Classification, (ii) Diverse Features for User Classification, and (iii) User-level Approaches.

3.2.1 Embedding based Approaches to User Classification

The profile of a user on Twitter consists of user description, tweets and profile picture. Researchers [Zhang et al., 2016a] utilized user tweets to learn an embedding model using Long Short-Term Memory (LSTM) and Recurrent Neural Network (RNN) to classify users based on their gender and age information achieving an accuracy of 91% and 82% respectively. In contrast, [Rizos et al.] employed interactional features to generate embeddings for a semi-supervised approach. Specifically, they utilize a small number of seed users with labels (e.g., news agency, person, genres) and interactions via "mentions" in their tweets. [Liao et al., 2017] proposed an approach to learn the interactional features of users by optimizing

the structural and attribute level properties of their networks that characterizes homophily in their communication. In another study [Benton et al., 2016], researchers utilized person-level multiview embedding to predict engagement, friend selection and demographic information of users. In contrast, our study gleans person, content and network-level features, creating a composition of multiview embeddings through *vector addition* operation that characterizes users in the context of marijuana-related communications on social media.

3.2.2 Diverse Features for User Classification

Prior work related to user classification on social media has involved different sets of features: (i) Person-level features included profile [Pennacchiotti and Popescu, 2011b], user behavior, first and last names [Bergsma et al., 2013], and demographics; (ii) Content level features included linguistic, domain-specific and generic LDA topics; and (iii) Network level features comprised follower-followee connections [Pennacchiotti and Popescu, 2011b]. These features were utilized to glean political affinity, ethnicity, and favorability towards a particular profession, to generate machine-readable user-profiles for improving the user classification [Pennacchiotti and Popescu, 2011b], and to cluster users based on their conversations and predict demographics [Bergsma et al., 2013]. Combination of these features with network interactions results in a better-contextualized representation of the dataset [Campbell et al., 2013], which in turn improves user classification.

3.2.3 User-level Approaches

For particular problems such as identification of user interests and event detection, user-level understanding of the content as well as the network dynamics is pivotal. In [De Choudhury et al., 2012], they classified users into three classes, namely, organization, media personnel, and an ordinary person, to identify variation in characteristics across multiple events. Engagement of users on a particular subject on social media is considered as an important signal, and has been used for user classification in [Tinati et al., 2012]. The authors developed a model to categorize users as Idea Starter, Commentator, Curator, Amplifier, and Viewer. In the election domain, political homophily on social media forms a feature for user classification, and [Colleoni et al., 2014] illustrates its significance for resolving reciprocated and non-reciprocated ties in the network of users. Homophily creates social echo chambers polarizing the users, which can be used to discriminate ordinary users (or information seekers) from information providers (e.g., journalist). Topical analysis of the user-generated content is another informative approach about user characteristics. In [Fang et al., 2015], topic-centric Naive Bayes classifier was developed to identify the topics to categorize unknown users based on closeness of their topics to those of the users in the training dataset. Similar to the use of marijuana concentrates, in recent years, there has been a surge in the use of e-cigarettes among smokers, and Twitter has emerged as a cost-effective platform for sharing and promoting information. Researchers [Kim et al., 2017] developed an approach to classify users as individuals, informed agencies, marketers, spammers, and vapor enthusiasts, employing tweet and user metadata, and tweeting

behavior.

3.3 Preliminaries

We discuss the people-content-network paradigm [Purohit et al., 2011] and compositional word embeddings. Our approach uses several building blocks for an in-depth analysis of tweet content to extract relevant context in marijuana dataset. Specifically, we discuss the people-content-network paradigm [Purohit et al., 2011] and compositional word embeddings for expressiveness, EmojiNet for interpreting emoji, Clarifai for processing profile pictures, and SMOTE for oversampling.

3.3.1 People-Content-Network

On social media, communities are being formed around various topics of interest through network interactions [Purohit et al., 2011]. The information being shared in tweets by a user in the marijuana community displays an intent based on the user type [Purohit et al., 2015]. For instance, *personal users* share their experiences and opinions on marijuana, whereas *retail accounts* usually promote the use of marijuana and other related products that they sell, and *media accounts* disseminate information on marijuana-related events and festivals, and legalization processes. Accordingly, as these user types show different characteristics, it is critical to bring to bear different perspectives, such as person, content, and network, for reliable analysis and insights. We describe a systematic organization and analysis of these features in Section 3.5.3.

3.3.2 EmojiNet

Emoji are pictorial representations of facial expressions, places, foods and other objects. These are often used by marijuana community on social media to express opinions and emotions about marijuana-related topics. Emoji contribute to the interpretation of the content created by users and better recognition of characteristics of the user types. To achieve this goal, we make use of EmojiNet [Wijeratne et al., 2017a], which gathers meanings of 2,389 emoji. Specifically, EmojiNet provides a set of words (e.g., smile), with the corresponding POS tags (e.g., verb), and their sense definitions. It maps 12,904 sense definitions to 2,389 emoji, to capture platform-specific interpretations.

3.3.3 Clarifai Web API

We use information gleaned from profile images of the users in our training dataset as profile pictures, which can provide additional insight about a user. For instance, retail accounts usually use marijuana related pictures in their profiles. We also use Clarifai[3], a web service for image processing, to recognize objects in an image and generate a textual representation of the image using a subset of 20 tags. We also used the feedback endpoint[4] of Clarifai to send end-user responses on the tags generated by the API to improve the quality of the tags. We eventually generated word embeddings utilizing these tags, for each user.

[3] `https://clarifai.com/developer/guide/`

[4] `https://clarifai.com/developer/guide/feedback#feedback`

3.3.4 Word Embedding Model

A word embedding model created using word2vec can learn a rich low dimensional representation of words in a tweet corpus. Initially, the word embedding procedure was developed to generate distributional representations over corpora such as Wikinews, News articles, and Google News corpus that represent the current state-of-the-art. [Mikolov et al., 2013a] also shows that vector arithmetic over the word vectors can be used to generate analogies. For instance, word embedding of "Queen" can be obtained by summing the word embeddings of "Man" and "Woman" and subtracting from it the word embedding of "King."

The model takes the corpus as input, identifies unique words (with vocabulary V of size v) and generates k-dimensional word vectors in v-dimensional word-space. k-dimensional word vector is also termed as k-dimensional word embedding or k-dimensional word representation. Similarities (e.g., cosine similarity) between the word vectors of two words from V reflect the semantic relationships between them. This improves upon bag-of-words approach or n-gram models that is unable to capture deep semantic relationships satisfactorily.

In recent studies [Lilleberg et al., 2015; Wang et al., 2016], the researchers have shown that word embedding models perform well over short texts. In another study [Godin et al., 2015], the authors have created a "named entity recognition shared task" for data from microblogging platforms using distributed word representations. These recent and prior successes in modeling words as computable vectors have encouraged us to utilize a pre-trained word2vec model trained over a generic Twitter corpus [Godin et al., 2015] or train a new word-embedding model

over our domain-specific Twitter corpus. Depending on the type of the corpus (characterized using sentence level statistics and word frequency counts), we can use one of two neural network architectures for learning word2vec embeddings: (i) Continuous Bag-Of-Words model [Mikolov et al., 2013a] (CBOW) (ii) Skip-gram model [Mikolov et al., 2013a]. In our study we have used skip-gram architecture. First, is the continuous bag-of-words model (CBOW) which learns the embeddings by predicting the target word, given the contextual words (neighboring words within a predefined window size). Second, is the Skip-gram model which learns the embeddings of the contextual words, given the target words. The latter has been proven to work well over short text using negative sampling [Wijeratne et al., 2017b].

3.3.5 Compositional Word Embedding

We utilize compositional word embedding (CWE) [Mitchell and Lapata, 2010] to combine feature-level embedding vectors and to generate a comprehensive representation of a data point (e.g., user, tweet, user descriptions). Specifically, we employ weighted vector addition, a linear composition function detailed in [Mitchell and Lapata, 2010]. Formally, we define Z, the weighted composition of word embeddings of U and V as follows: $\mathbf{Z} = W_0 \cdot U + W_1 \cdot V$, where $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{m \times 300}$($\mathbf{m}$ represents the number of users) are two embeddings composed using $\mathbf{W_0}, \mathbf{W_1} \in \mathbb{R}^{m \times m}$, respectively. Note that in such a composition, the dimension of input and output representation is unaltered. As detailed in Section 3.5.2, it is essential to consider the correlation between different view embeddings before composing them. For

instance, for **Z**, the weight matrices will be optimized; however, if the embeddings U and V are uncorrelated, hence independent, the optimization function will fail to converge. Hence, we perform a linear composition, vector addition, to generate the representation of **Z**. Since the classification is insensitive to the position of emoji and words in the content, we consider such composition appropriate. Formally, **Z**= U+V is a vector addition of U and V.

As discussed later, CWE provides a reasonable semantic basis for combining text and emojis in a contextually meaningful way. Our domain-specific word embeddings are used to capture the semantics of words directly, and of emojis via its textual description from emojinet, when processing tweet content. A different User description-specific word embeddings are used for both of them when processing them in the user description context. This approach ensures that words and emojis are treated similarly in each context, and the distinction between the two contexts is also captured. Note that the emojis are not directly mapped to their word embeddings because the scarcity of data would have made such embeddings unreliable. A more thorough investigation of the relative semantic adequacy of the various word embedding models – domain independent vs domain specific, direct emoji embeddings vs indirect emojis embeddings via its definition in emojinet – is out of scope of this work. [Scheepers et al., 2018; Senel et al., 2018] provide relevant intuitions for our work.

3.3.6 Managing Imbalance

Social media platforms are an epitome of real-life communication channels for information dissemination. So machine learning algorithms need to deal with imbalanced data (or uneven data distribution) that naturally arise due to atypical behaviors of different groups of users on social networks [Krawczyk, 2016]. For instance, considering the users interested in rock music, the youth population on Twitter will form a majority class while the older adults will constitute the minority class. Hence, we require a sampling approach that can handle such data imbalances. In the context of our corpus, the training dataset was imbalanced with respect to retail user type. For this reason, we used the popular sampling approach, called Synthetic Minority Over-sampling TEchnique (SMOTE). It is a procedure in which the minority class is oversampled by creating additional synthetic examples from existing data utilizing feature-level properties and operations (e.g., feature correlation, covariance). Oversampling of the minority class is performed by considering the k-nearest neighbors of minority class samples identified by the line segments on which they lie [Chawla et al., 2002].

3.4 Exploratory Analysis

Figure 3.1 captures the word cloud synthesized using the tweet content of users pertaining to Informed Agency user type that can be used to glean related topics.

We have conducted an analysis of our dataset by extracting statistical, textual

Figure 3.1: Word Cloud of tweets from Informed Agency.

and topical information. We have three classes of user types, namely, Personal (P), Informed Agency (I), and Retail Accounts (R). Our corpus contains tweets crawled in June, July, and August of 2017, covering all states in the U.S. During this time frame, the volume of communications related to marijuana was high due to ongoing events (e.g., Cannabis Cup, The 420 Games)[5]. Data collection involved semantic filtering [Sheth and Kapanipathi, 2016a] utilizing the DAO[6] on the eDrugTrends[7]/Twitris platform[8]. The corpus comprised of a total of 4,106,566

[5]https://goo.gl/Xzsd5V
[6]https://goo.gl/9cXQcT
[7]http://wiki.knoesis.org/index.php/EDrugTrends
[8]http://twitris.knoesis.org

tweets from 1,066,615 unique users. Out of nearly 4.1M tweets, nearly 1.9M tweets were identified as unique based on the content.

Table 3.1: Descriptive Information on the Training Set.

Features ('#' is "number of")	**P**	**R**	**I**	**Total**
#Tweets (T)	9836	1928	338	12102
#Profile Pictures (PP)	4394	476	111	4981
#Users use Emoji (E)	1085	37	17	1139
#Users with Descriptions (D)	3884	461	108	4453
#Retweets	955	24	964	1943
#Mentions	94	6	307	407

We randomly selected 6000 users from our pool of 1M unique users for training. The domain experts from CITAR[9] manually annotated only 4982 users into the following three types: Personal Accounts, Informed Agency, and Retail Accounts, since the others were ambiguous. After the annotation process, the distribution per user type was as follows: 4395 personal, 476 informed agency, and 111 retail accounts. Effectively, the distribution of user types in the training set is highly skewed. The reason for sparsity among retailers (i.e., retail business twitter accounts) is that marijuana is a schedule I[10] drug according to the federal law, and thus its promotion on social media platforms is restricted due to its federal status as an illegal drug. Similarly, media accounts are significantly smaller compared to personal accounts, but still significantly higher than the retail accounts. Such data imbalance can bias the classifier towards the majority class, which is a challenge.

Upon our initial exploratory analysis of the corpus, we observed that the content in tweets and description of users are adequate to identify the characteristics of different user types. The average number of words in descriptions and tweets

[9]https://medicine.wright.edu/citar
[10]https://goo.gl/UQhR4D

are 9.6 and 12.8, while the average number of emoji in descriptions and tweets are 0.46 and 0.26, respectively. 88% of the users have their descriptions complete, and these user descriptions carry information containing emoji and text that can be utilized for classification.

Further, interactions among users can play an essential role in disseminating information, and influence other connected users in the network. The median number of followers and friends for users are 367 and 376 respectively, and the average number of tweets per user is 3.85. Our corpus includes 2,837,734 interactions (mentions, retweets) between users, 83% of which are retweets, and the rest are mentions. This suggests that there is much communication among users that can contribute to the classification of user types.

3.5 Methodology

Our approach to the user classification problem leverages the multiview and multimodal aspects of the Twitter data by creating compositions of embeddings for different views using data in different modality. As depicted in the overall architecture in figure 3.2, this section provides details of critical steps in our approach.

3.5.1 Preprocessing

At this stage, we trained two Word Embedding(WE) models – one for *Content* view and another for *People* view – using our domain-specific Twitter corpus. (i) The Content WE model is based on 1.8 M unique pre-processed tweets, and (ii) the

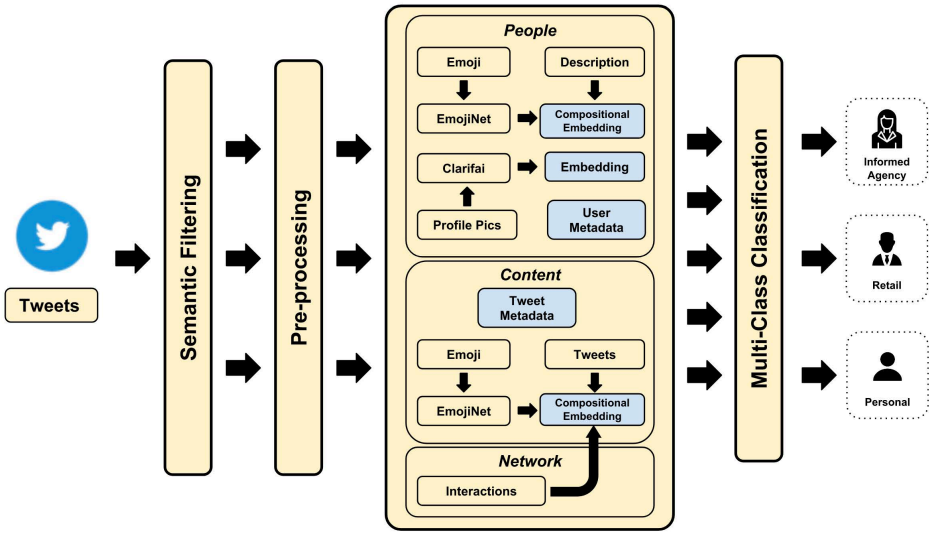


Figure 3.2: Overall Architecture. The workflow shows composition of embeddings across People-Content-Network views for User Classification

People WE model is based on pre-processed user descriptions of 1 M unique users. We built two separate WE models because we observed that user descriptions were more complete and contained less jargon and slang terms as compared to tweets.

To obtain discriminative features for user classification, we removed stop words, punctuations, and alphanumeric characters from tweets and user descriptions. We also extracted URLs, mentions of screen names, retweeted user screen names, contact information (e.g., phone number, email), and emoji. Then, we lemmatized the tweets and user descriptions in the corpus. Moreover, we employed EmojiNet [Wijeratne et al., 2017a] to retrieve senses and keywords from emoji, and Clarifai[11] to process profile pictures. The overall goal is to enable gleaning of semantically relevant information about users from their tweets for reliable determination of

[11] `https://www.clarifai.com/demo`

user types.

3.5.2 Correlation Analysis

We perform correlation analysis between embeddings of features from different views to assess which compositional operation is appropriate. The similarity between embedding vectors derived from the textual representation of features constrains the operations that can be used to combine them since the resulting vector needs to be representative of the components. For example, when two embedding vectors are highly uncorrelated, dimensionality reduction does not generate representative vector space. However, uncorrelated embeddings can be composed with vector addition, to make resulting vector space more representative.

For instance, researchers [Goikoetxea et al., 2016] made use of operations such as addition and concatenation, to combine word embedding vectors of the input text. These word embeddings were generated from text corpora and knowledge bases for more contextually rich representation of the input text. Similarly, [Faruqui et al., 2014] retrofits word vectors, using the WordNet embeddings to enrich word embeddings of the input text.

The creation of embedding vectors is performed through probabilistic calculations [Bamler and Mandt, 2017], and the embedding of each view (Section 3.5.4) may or may not correlate with that of the other views.

We conducted correlation analysis between different pairs of view embedding vectors. Table 3.2 shows Spearman correlation and their corresponding p values for these pairs. We use Spearman as our correlation metric to measure the similarity

Table 3.2: Spearman (ρ) Correlation Analysis for View Pairs

View Pairs	ρ	**p-value**
User Description & Emoji	0.002	< 0.01
Tweets & Emoji	0.02	< 0.01
Tweets & Network	0.04	< 0.01
User Description & Network	0.0001	> 0.01

between view embeddings at each data point since our embeddings do not follow the Gaussian distribution. In this analysis, our alternative hypothesis (H_1) is that the two embedding vectors are uncorrelated, and similarly the null hypothesis (H_0) is that they are correlated. Having the p-value less than 0.01 suggests the rejection of H_0. Hence, based on Spearman, we see from the Table 3.2 that, for the first three pairs, the null hypothesis of correlation H_0 can be rejected, while for the pair – User description and Network, we are unable to reject the null hypothesis of correlation (H_0). In fact, the data indicates that people interact closely based on their similar user characteristics rather than the shared tweet content in marijuana-related communications.

3.5.3 Feature Engineering

In our analysis, we have organized our features under three main categories: Person, Content, and Network, since we consider these as the main views of the Twitter communication that contribute to the context.

Figure 3.3: Example profile pictures(2 each) of P(left), R(center) and I(right)

People

This set of features are user-level that contributes to differentiating the user types from each other on social media. Specifically, we consider user descriptions, name, screen name, contact information and profile pictures as discriminative based on the exploratory analysis (Section 3.4).

- User Descriptions: This field holds the description of the account that was defined by the user. As this metadata carries information on characteristics of the user, we exploit the elements of this feature such as text, emoji, and contact information by employing text processing techniques.

- Name: This field holds the name of each user where users can enter their full personal, business, or organization name, or have an arbitrary entry. We use this information to discriminate person users utilizing a lexicon[12] of commonly used person's first and last names. In fact, we found that 68% of the person users can be identified using names listed in the lexicon.

[12] `https://goo.gl/8MY5Cz`

- Contact Information: We extract this information from the description of users as it includes a phone number, email and web addresses. Usually, retail accounts provide this information in their profile for their customers to reach out to them, making this feature a discriminative factor in classification.

- Profile Pictures: This visual form of Twitter data can reflect feelings, emotions, intentions, and other characteristics of a user. We consider this feature as discriminative as there is a noticeable difference in profile pictures of personal, retail, and informed agency accounts (see figure 3.3 for examples).

Content

To glean discriminatory features from tweet content, we first separated text, emoji and URLs, and then processed them separately. The number and frequency of URLs, and the number and senses of emoji in tweets of users were contributing factors in discriminating user types.

- Tweet text: We first extracted tweet text, by filtering other elements such as mentions, URLs, and emoji, and concatenate tweets of each user. Then we created word embedding vectors out of this textual data.

- URLs: Retail and Media accounts usually use URLs in their tweets to direct clients to their web page, more often than personal accounts. The number and frequency of URLs in a tweet can help to discriminate among user types.

- URLs: Users usually provide URLs in their character-limited tweets to refer to a more detailed version of their stories. For instance, retail and media

accounts use URLs in their tweets to direct clients to their web page, more often than personal accounts. The number and frequency of URLs in a tweet can help to discriminate among user types.

- Emoji: The use of emoji provides a concise and precise expression of opinions, reactions, sentiments, and emotions concerning a topic of discussion. It is a discriminative feature in our study capturing the number and senses of emoji used by different user types. The most commonly used emoji in Tweets across all the three user types are: 😂, 😭, 🏾, 🔥, and 🤦‍♀️. The most commonly used emoji in User Descriptions are:
👦🏽, 🐶, 🙏, 🌬️, and 🌄.

Network

As users on Twitter primarily interact using replies, mentions, and retweets, we utilize these interactions as our features to identify communication patterns for each user type. We consider replies as mentions. We generate network embeddings by creating the adjacency matrix based on these interactions between users. This procedure is further explained in Section 3.5.4.

- Mentions: It is a derived feature where the author mentions the screen-name of another user and is considered as direct interaction.

- Retweets: It is a derived feature where the retweeting user forwards another user's tweet and is considered a direct interaction between these two users.

3.5.4 Compositional Multiview Embedding (CME)

The Twitter data contains multiple dimensions that we call views, such as People, Content, and Network. These views can be leveraged to contextualize a comprehensive and multi-level analysis of the Twitter social network. In our study, we employed the Content and People WE models for generating embeddings for Content view (e.g., Tweets) and People view (e.g., User Descriptions and Profile Pictures), respectively.

The tweet content and user descriptions involve emoji, which we regard as critical for interpreting the meaning. For this reason, we extracted the textual representation of emoji from EmojiNet [Wijeratne et al., 2016a], and generated emoji embeddings using the embeddings of the words in the description. We also generated word embeddings for profile pictures of users utilizing Clarifai. We generated comprehensive CMEs by combining the embeddings of different views of the Twitter data, as formulated below.

For Person and Content views (T), word embedding vector (WV) in each data point (WV_{T_i}, i represents an index of a data point in a view) is calculated by averaging the word-vectors of each word that is present in the view. For instance, we preprocess the tweets of a user and generate word vectors of each word in 300 dimensions. Then we sum these vectors and divide by the number of words to generate the embedding vector for tweets of the user. However, while we perform the average operation to generate separate embedding vectors for Person and Content views, we do not perform average for the Network view. For generation of network embeddings, we utilized interactional features (mentions and retweets)

and performed t-SVD to generate dense embeddings, where each embedding has 300 dimensions. The procedure is detailed in Section 3.5.4.

We formally define the calculation of WV_{T_i} as $\mathbf{WV_{T_i}} = \frac{\sum_{w \in T_i \cap V} \vec{\mathbf{v}}_w}{|T_i \cap V|}$, where $\vec{\mathbf{v}}_w$ is the embedding of word $\mathbf{w}$ and V is the vocabulary of the Content WE model trained over the marijuana-related tweet corpus.

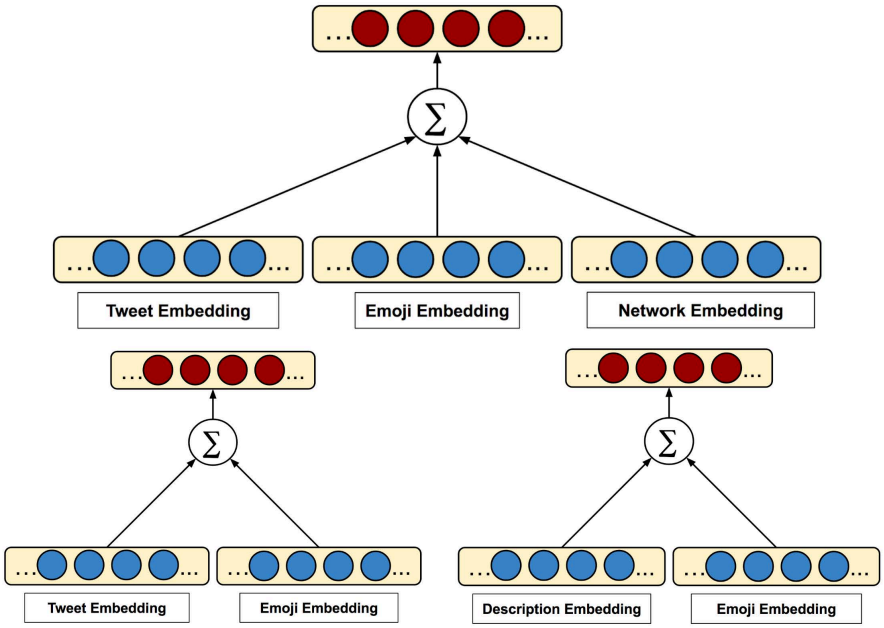


Figure 3.4: Creation of CMEs for Tweet, Description, Emoji and Network

Tweet-Emoji(T+E) & User Description-Emoji(D+E)

We explained the procedure for generating WEs for Tweets, User Descriptions and Emoji earlier in this section, given the component word embeddings. We now explain in more detail how we obtained the Content WE and Person WE that were used to generate CMEs for Tweet & Emoji, and a separate CMEs for User Description & Emoji.

In building the Content and Person WE models, we used Skip-gram model with negative sampling. The rate of negative sampling was set to 10 and the window size was set to 5. Such a set up is desirable for datasets of average-size [Mikolov et al., 2013a]. The Content WE model was trained on a pre-processed corpus of 1.8M unique tweets generated from 1M unique users creating a vocabulary (V) of 16,531 words. The People WE model was trained on 946,975 pre-processed user descriptions obtained from 1M unique users, generating a vocabulary (V) of 16,903 words. Recall that, apart from linguistic differences between user descriptions and tweets, another reason to build two WE models is multiview aspect of our dataset that also includes profile pictures and emoji in a profile that reflects different contextual meanings as compared to the tweets of a user. In order to create an embedding of a profile picture, we used Clarifai to generate text caption and then apply the Person WE model on the text caption.

Network Embedding (N)

The user types that we characterize have different volumes of network activities. For instance, while average retweet and mention rates (derived from Table 3.1) per user are 0.9 and 0.09 respectively on personal accounts, they are 11.08 and 3.53 on informed agency accounts. Clearly, the network activity can be used to distinguish and recognize these user types. Thus, combining the network activity information with tweet content and user information can contribute to a reliable classification.

For representing the network activities of users, we created the weighted adjacency matrix of interactions, which was sparse. For generating a low dimensional dense representation, we utilize truncated Singular Value Decomposition (t-SVD) such as in GloVe [Pennington et al., 2014] and [Tsitsulin et al., 2018].

Formally, we define the adjacency matrix as $\mathbf{A} \in \mathbb{R}^{m \times n}$, where **m** and **n** denote source users and target users respectively (capturing direction of communication).

$$A_{u_i,u_j} = InteractionCount_{u_i,u_j} \tag{3.1}$$

where, for a pair of users u_i, u_j, A_{u_i,u_j} represents a cell in the matrix $\mathbf{A}$ of dimension $|m| \times |n|$ representing interaction counts, which includes both retweets and mentions, for the corresponding users.

As the adjacency matrix $\mathbf{A}$ is sparse and non-stochastic ($\sum_{j=1}^{n} A_{i,j} \neq 1$), and we need to create a dense and stochastic representation of the network activities, we normalize the values in a row such that they will all sum up to 1. In our training set, only 1149 users have interactions with other 1701 users. As the source and target users are mostly different in $\mathbf{A}$ (1149×1701), we convert $\mathbf{A}$ to square cosine matrix, denoted by $\mathbf{A^{cosine}} \in \mathbb{R}^{m \times m}$ obtaining a matrix 1149×1149, since we want to measure the similarity between users in our training set. Transformation of $\mathbf{A}$ to $\mathbf{A^{cosine}}$ is formulated as follows: $A^{cosine} = \frac{A \cdot A^T}{||A||||A^T||}$. Each cell value in $\mathbf{A^{cosine}}$ lies between 0 and 1 and is symmetric.

However, the similarity between users represented in $\mathbf{A^{cosine}}$ are not representative of their degree(d) as present in $\mathbf{A}$. A degree of a user is defined by the number of outgoing edges in terms of interactions with other users and not

self. In order to normalize $\mathbf{A}^{\mathbf{cosine}}$ by the degree (d) of each user, we generate a normalized adjacency matrix, denoted by $\mathbf{A}^{\mathbf{norm}} \in \mathbb{R}^{M \times M}$ and is formulated as, $\mathbf{A}^{\mathbf{norm}} = D^{-1/2} A^{cosine} D^{-1/2}$, where $\mathbf{D}$ is a diagonal matrix of degree of each user in $\mathbf{M}$ and is represented as: $D^{-1/2} = diag(d_1^{-1/2}, ..., d_M^{-1/2})$

The factorization of the matrix $\mathbf{A}^{\mathbf{cosine}}$ using SVD yields three square matrices: U, Σ, $V^T \in \mathbb{R}^{M \times M}$, where $\Sigma = \{\sigma_1, \sigma_2, ..., \sigma_M\}$ is a set of $\mathbf{M}$ singular values or eigenvalues of the matrix $\mathbf{A}^{\mathbf{cosine}}$. Not all singular values are non-zeros, removing the zero values or near zero values reduce the dimension of the matrix. Since, the required network embedding needs to be of the same dimension as word embedding (i.e. k=300), we truncate Σ by M-300. U is reduced by removing M-300 columns and V^T is reduced by removing M-300 rows. After the truncation, the reduced matrix generated has dimension $M \times 300$. We denote reduced matrix as $\mathbf{A}^{\mathbf{reduced}} \in \mathbb{R}^{M \times 300}$ and its value is determined by: $\mathbf{A}^{\mathbf{reduced}} = U_{M \times 300} \cdot (\Sigma^{-1}_{300 \times 300})^T$. The 300 dimensional embedding present in $\mathbf{A}^{\mathbf{reduced}}$ is considered as network embedding of the users and is used in user type classification.

As our adjacency matrix $\mathbf{A}^{\mathbf{cosine}}$ is 1149×1149, we reduce its dimension down to 300. Therefore, we apply t-SVD over the matrix $\mathbf{A}^{\mathbf{cosine}}$ resulting three square matrices: U, Σ, $U^T \in \mathbb{R}^{m \times m}$, where $\Sigma = \{\sigma_1, \sigma_2, ..., \sigma_m\}$ is a set of $\mathbf{m}$ singular values. After we apply the dimensionality reduction, the reduced matrix becomes of dimension $m \times 300$. We denote the reduced matrix as $\mathbf{A}^{\mathbf{reduced}} \in \mathbb{R}^{m \times 300}$ and its value is determined by: $\mathbf{A}^{\mathbf{reduced}} = U_{m \times 300} \cdot (\Sigma^{-1}_{300 \times 300})^T$.

The 300 dimensional embeddings in $\mathbf{A}^{\mathbf{reduced}}$ is considered as the network embedding of users, and is used to create a CME in our user type classification.

Network-Tweet-Emoji(N+T+E)

We use the WEs for Tweets and User Descriptions, and network embeddings (NE) of users to generate a comprehensive Network-Tweet-Emoji CME. Embeddings for Network, Tweets and Emoji all have 300 dimensions because the former was explicitly represented using 300 dimension reduced space, while the latter two very created using the standard word embedding approach.

3.5.5 Experimental Setting

To the best of our knowledge, the problem of user type classification in marijuana-related communications on Twitter has not been investigated so far. Our experiments using clustering algorithms such as [Wang et al., 2018b], MeanShift, K-means, and DBSCAN, for a baseline, significantly under-performed, partly because of their instability. For this reason, we created an *empirical baseline* that utilizes word embeddings of the textual content of tweets and descriptions.

We conducted two sets of experiments depending on whether CME with network level features were included or not. The first set of experiments do not include the CME with network level features, and we incrementally add the Person and Content level features. We used 10-fold stratified cross-validation with same proportions of all types in all folds, utilizing all data points in our training set that is comprised of 4982 users. The second set of experiments included CMEs which contain Network level features, where we take the best performing classification

setting from the first set of experiments as a baseline for comparison. At this stage, we had to reduce the size of the training set down to 1149 users where the sizes of P, I, and R classes were 1045, 87 and 17 respectively.

Since our training set was highly imbalanced, we applied the oversampling algorithm SMOTE to avoid biasing towards the majority class at the expense of the minority classes.

In our experiments, to illustrate the improvement that the domain specific WE models provide, we also utilized a generic word2Vec model, called Tweet2Vec [Godin et al., 2015], for a comparison, which is explained in detail in Section 3.6.

3.6 Results

Table 3.3 and Table 3.4 present the results of the two sets of experiments. The first set of experiments involve only user profile and tweet content level features, whereas the second set of experiments involve the addition of network features. To illustrate the improvement obtained by the addition of network level features into the classification, we take the best performing approach of the first set of experiments as the baseline for the second set of experiments.

We systematically and gradually include person-content-network features to observe their individual contributions to the outcome of the classification.

The baseline approach that we empirically chose achieved an overall F-score of 88% using the word embeddings of tweets content and user descriptions. The

Table 3.3: Results on Classification of User Types with 4982 Users.

Feature Set	Precision			Recall			F-score			Avg.F
	P	I	R	P	I	R	P	I	R	
E(T),E(D)	0.91	0.86	0.79	0.99	0.27	0.67	0.95	0.42	0.73	0.88
T2V(T),T2V(D)	0.89	0.87	0.87	0.99	0.10	0.66	0.94	0.18	0.75	0.86
E(T+E),E(D+E)	0.89	0.96	0.88	0.99	0.10	0.60	0.94	0.18	0.71	0.85
E(T+E),E(D+E), TMD,UMD	0.89	0.95	0.84	0.99	0.09	0.63	0.94	0.17	0.72	0.85
E(T+E),E(D+E), TMD,UMD,PP	0.97	0.99	0.88	0.99	0.77	0.92	0.98	0.87	0.90	0.97

Table 3.4: Results for Classification of User Types with 1149 Users

Feature Set	Precision			Recall			F-score			Avg.F
	P	I	R	P	I	R	P	I	R	
E(T+E),E(D+E), TMD,UMD,PP	0.96	0.98	0.93	0.99	0.57	0.82	0.97	0.72	0.87	0.95
E(N),E(T+E),E(D+E) TMD,UMD,PP	0.95	0.95	0.95	1.0	0.52	0.80	0.97	0.67	0.87	0.95
E(N+T+E),E(D+E) TMD,UMD,PP	0.96	0.98	0.97	1.0	0.58	0.86	0.98	0.73	0.91	0.96

E:Embedding, T:Tweet, D:Description, N:Network, T2V:Tweet2Vec, TMD:Tweet Metadata, UMD:User Metadata, PP: Profile Pictures

F-scores for individual classes of P, I, and R were 95%, 42%, and 73%, respectively. We generated these embedding vectors using the domain-specific word embedding models. Table 3.3 shows that the classifier built with the embeddings of tweets and descriptions generated through the Tweet2Vec model obtained an average F-score of 86%, and underperformed for P and I classes. Therefore, we continued experiments using Content and People WE models.

The inclusion of profile pictures as a feature in the experiments showed a significant improvement in the overall F-score to 97%, where F-scores for P, I, and R were 98%, 87%, and 90%, respectively. The inclusion of textual data, emoji and profile pictures in our approach by combining them through CMEs for classifica-

tion of user types, has impacted the outcome significantly.

Furthermore, recall that, in the second set of experiments, we have extended our study with the addition of network interactions between users. We have used the best performing classifier from the first set of experiments (Table 3.3) as a baseline for the second set of experiments, to compare our approach that incorporates the network embeddings.

In our second set of experiments, we have first added the network embedding as a separate feature along with the features from the second baseline approach, and it did not affect the performance. Then we created CME from the embeddings of tweets, emoji, and network, and it boosted the performance of each class, P, I, and R in terms of their F-scores, by 1%, 6%, and 4%, respectively. It also improved the overall F-score by 1%. The improvement that we achieved by applying CMEs is significant since the F-score for the second baseline was already significantly high, and our approach has improved upon that performance.

3.7 CONCLUSION

Our overarching goal was to utilize people, content, and network related features in marijuana-related communications on Twitter to classify the user types into three prominent categories: Personal, Informed Agency, and Retail accounts. Such a classification provides support for analyzing the dynamics of issues related to marijuana and its variants from location and temporal perspectives ultimately. Furthermore, dominant and trending topics can be identified separately for each

user type for more precise and reliable subjective analysis of related events and their impacts.

In this chapter, we introduced an approach to classify user types utilizing Compositional Multiview Embedding (CME). For the purpose of integrating multiple views and multimodal data, we learned a domain-specific embedding for tweet text, a separate embedding for user profile descriptions to adequately capture a different context, and a mapping of profile images to tags to obtain their embeddings and a mapping of emojis to Emojinet textual description to obtain two separate embeddings for them – one for content-view and another for person-view. We also incorporated interactional features by creating network embeddings. Overall, our comprehensive approach achieved 7% improvement over the empirical baseline, when we used the CMEs without network embedding and 8% improvement when we used the CMEs with network embedding. The latter also resulted in an F-score of 0.96.

Chapter 4

Use Case 2: User Modeling on Radical Communications

4.1 Introduction

Radicalization is a social and psychological process through which individuals experience incremental commitment or adaptation of extremist views and ideologies [Horgan, 2009]. Radical networks have effectively and strategically utilized social media [Vidino and Hughes, 2015] to disseminate highly persuasive ideological content and recruit new members. The new generation of homegrown terrorist is ethnically diverse and technologically savvy and actively use social media [Hafez and Mullins, 2015]. Individuals change their belief systems, adopt a radical view,

and advocate for action, including the willingness to use violence to achieve a radical societal change [Helfstein, 2012; Porter and Kebbell, 2011; Vidino, 2011]. Often, demanding action against the enemies of Islam (e.g., the West) completes the radicalization process of mostly young and ordinary individuals, turning them to lone wolves.

We address fundamental *data science* challenges that are common to a particular set of data-related grand social problems, such as (Islamic) religious extremism, white supremacy, and trolling activities of oppressive regimes such as Russia and China. Although we only present examples from the religious extremism domain, the challenges and proposed solutions are very similar among this particular subset of problems, for which we coin the term *persuasive social data*, as online social (media) data is used to persuade individuals into a particular religious, racial or political doctrine.

Persuasive social data involves unconstrained doctrinal concepts and relationships with contextual meanings from religion, history and politics. For example, the concept of "Jihad" can mean (i) self-spiritual struggle, (ii) defensive war to protect lives and property from aggression, or (iii) act of (unprovoked) violence, depending on its contextual use. Classification of the first and second uses of Jihad as extreme or radical would be a grave mistake.

Actors in persuasive social data challenges frequently disguise themselves as true representatives of a religion, doctrine or ideology (e.g., radicals posing as true (mainstream) believers in Islam). This means persuasive (propagandist) data will be very similar to data produced by common agents with no hidden persuasive

agenda, except they will contain concepts and relations that are twisted in their meaning, or presented out of context (e.g., Jihad) or sometimes outright misinformation. This leads to sparse true signals within the training data sets. For example, it is very difficult to distinguish social media posts from Russian trolls disguised as American citizens during 2016 US presidential election. Further, propagandist data commonly show non-stationary patterns that dynamically change over time. For example, adherents of Islamic extremism have shifted their attention from promoting the caliphate established by ISIS to encouraging violence in the West recently. For this reason, the limited number of labeled instances available for training can often fail to represent the true nature of concepts and relationships in these persuasive social data sets.

A process of persuasion usually starts out benign and turns increasingly intense and radical over time. We model this process as the interaction among connected agents with a mix of perspectives and influence on each other, each one of which exemplifies a degree of radicalization and depends crucially on the proper identification of relevant message features. In our work, we measure the degree of radicalization (varying from vague support for extremism to violent extremism) and also capture the process of radicalization to understand the recruitment process better. This requires a more in-depth classification in which an agent's radicalization stage and timeline are also identified.

Based on these observations, we believe standard KB and DL only methods break down on persuasive social data and lead to misleading conclusions. In particular, it is easy to deduce or learn spurious concepts and relationships that look

deceptively good on a knowledge-base or training and test sets, yet do not provide adequate results when the data set contains contextual and dynamically changing concepts and relations. In our approach, we incorporate domain knowledge of radical ideology in deep learning models to relate features spanning religion, ideology and violence, to address domain-specific lexical and semantic challenges, such as sparsity, ambiguity and noise to classify discourse along a radicalization scale informed by political science.

While carefully selected verses from the Qur'an or Hadith (prophetic narrations) inspire individuals to be included in their echo chambers, deviant interpretations and commentaries of such Islamic knowledge painted in a radical ideology are used to recruit their followers (see Table 1). At the macro-level, three dimensions inform the conceptualization of Online Radicalization Index(ORI): (a) religion, (b) ideology, and (c) degree of support for violence. Religiosity is the superset of overall Muslims attitudes which range from "mainstream" through more "extreme" interpretations of Islamic scriptures. Attitude toward political ideology (i.e., Islamism) is an essential dimension of radicalization. Conceptualization and measurement of variation in political, ideological attitudes toward Islamism are drawn from Achilov and Sen [2017]'s novel (concept building) study of Political Islamism. Finally, support for violence is the third critical dimension that marks vital benchmarks on radicalization with the potential of carrying out violent terrorist acts [Helfstein, 2012; Hafez and Mullins, 2015].

Further complicating identifying the relationship between lexical feature and radicalization class is that the meaning of an individual term depends upon its

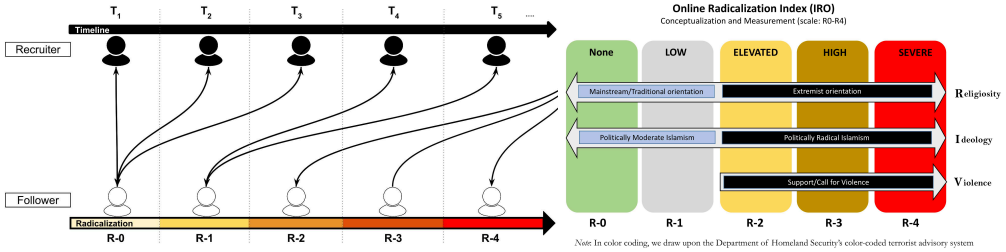
Table 4.1: Example tweets from verified radical social media users. They are annotated for religious (**R**), ideological (**I**) and violent (**V**) terminology. **Jihad** appears in multiple perspectives.

Radical Content Examples	**R**	**I**	**V**
"Here is the fragrance of **Paradise**,Here is the field of **Jihad**. Here is the **land of #Islam**,Here is the **land of the Caliphate**"		✓	
"Reportedly, a number of **apostates** were **killed** in the process. Just because they like it I guess.. #Spring**Jihad** #CountrysideCleanup"			✓
"and **Jihad** means to **sacrifice** YOURSELF in **war** to save your **country** (or **religion**)"		✓	✓
"I asked about the paths to **Paradise** It was said that there is no path shorter than **Jihad**"	✓	✓	
"**God** honored us w/ **Jihad Khilafah** in this era of **Fitnah**"		✓	
"By the **Lord** of **Muhammad** (blessings and **peace be upon him**) The **nation** of **Jihad** and **martyrdom** can never be **defeated**"		✓	
"Anyone who prefers to raise secularism over **Islam** is a **kafir**, whether he's from **Saudi, Sudan, Somalia, Mexico, Burma, Hawaii**, or elsewhere."		✓	

surrounding in the content. Consider the term "jihad". For example (see examples in Table 4.1), when it co-occurs with "kill" and "attack", the term "jihad" connotes *violence*. In the presence of "Allah" and "Islam", the meaning of the term "jihad" is its original, *religious* concept of self-struggle. "Jihad" can also co-occur with "imam_anwar_al_awlaki" who is considered [Bowman-Grieve and Conway, 2012] as a prominent *ideologue* of radical Islamist groups. Highly diagnostic terms such as "martyrdom" can *sparsely appear* in a corpus, so that mere frequency does not convey importance. Moreover, considering that keyword-based social media data can bring *noise*, traditional learning mechanisms will likely overlook such significant indicator terms in the content. Not surprisingly, efforts from social media platforms to detect radicalism remain inadequate, limited in scope, opaque, and mostly ineffective, [Hussain and Saltman, 2014].

4.1.1 The Radicalization Classification Problem

In this research, we lay the groundwork for a comprehensive classification of radicalization where we can characterize users, and model recruiter and follower personas in each stage of radicalization. It will also allow us to understand the underlying dynamics of the radicalization process and progression of radicalized users over time, from religious, ideological and violence perspectives. We adopt a conceptual model of Online Radicalization Index (ORI), drawing on Achilov and Sen [2017] and Klein et al. [2006] three-level concept building framework: (a) concept, (b) concept intentions, (c) data indicator levels. Concept intentions are informed by current radicalization research in social and behavioral sciences. At



(a) (b)

Figure 4.1: (a) The radicalization process that *Recruiters* of radical groups pursuing over their *Followers* over periods (e.g., T1, T2) of time through *Radicalization stages* (R-0 to R-4) where R-0 being not radical and R-4 most radical based on our Online Radicalization index (see Table 4.2). The same individuals participate throughout the persuasion process, as the follower proceeds through stages of radicalization. (b) Working Conceptual Model for Islamist Radicalization on Social Media (adopted from Achilov and Sen [2017])

the macro-level, three dimensions inform the conceptualization of ORI: (i) religion, (ii) ideology, and (iii) degree of support for violence. Table 4.2 defines each category/scale of ORI. These define the classes against which we will classify the social media content of a potentially radicalized follower.

4.2 Related Work

This section focuses on the related computer science research. Neural networks provide a standard approach to a set of machine learning problems. This powerful class of algorithms can classify feature patterns (e.g., relationships) within the data and correct itself through backpropagation Halevy et al. [2009]. As stated in Halevy et al. [2009], the deeper the network, denser the representation and

Table 4.2: Five-Level Conceptualization of ORI

ORI	**Concept Intention (core defining features)**	**Data Indicator**
R-O: None	Mainstream religious views and orientations Explicit opposition to religious extremist views, politically-radical Islamist ideology and the use of violence in the name of Islam	Islam; Allah; jihad (as internal struggle); (inclusive role of) Shariah; halal; democracy_islam; salah; fatwa; hajj;
R-1: Low	Attitudinal support for politically-moderate Islamism Vague or passive support for exclusive rule for the Shari'a law, known radical networks, known radical clerics and/or ideologues	hadith; Caliphate (Khilafah) justified; Sharia_better (than secular law); hypocrisy_west; fitnah_government;
R-2: Elevated	Emergent or implicit support for the exclusive rule for the Shari'a law Symbolic support for religious radical networks, radical clerics and/or ideologues, migration to Darul-Islam.	Shariah_best; qisas; revenge (justified); jihad (against West); justify Said Qutb; Daesh (ISIS), sahwah; fitnah_west;
R-3: High	Explicit support for the Shari'a law as the only legitimate form of government, religious radical networks (e.g., ISIS, Al-Qaida, Al-Shabab, etc.), clerics, migration to Darul-Islam	Kafir; infidel; hijrah to Darul-Islam; (supporting) fatwa_Al-Awlaki; mushrikeen; imamah_daesh/obey;
R-4: Severe	Operational support (call for action) for installing the exclusive rule of Sharia as the only legitimate form of government, labelling critics as "Kafirs" (infidels); call for action to join the fight and the use of violence.	apostate; sahwat; taghut; kill; kafir; kuffar; takfiri; murtadd; tawaghitt; al_baghdadi; mushrikeen; jihad (against West); (support for) fatwa_Al-Awlaki, fatwa_Al-Zawahiri, fatwa_Al-Baghdadi, osama_ibn, bin_Laden, Muqtada fatwa_as-Sadr; mujahedeen; darul-harb; (must) migrate_darul_Islam; martyrdim_for_khilafah; puppet_west;

the better the learning. A large number of parameters and the layered nature of neural networks make them modifiable based on specific problem characteristics. However, they are dependent upon an adequately captured feature space and large amounts of data. It makes them vulnerable to the sudden appearance of relevant sparse and/or ambiguous features in noisy big social media data, for example, an emerging concept. Novel techniques are required to compensate for this vulnerability, to add knowledge in principled fashion, based on a measurable discrepancy between the knowledge captured in the neural network and external resources. In this section, we describe existing research related to this problem.

4.2.1 Radicalization on Social Media

Harnessing the social media data, in particular Twitter, for Islamist Extremism, has become an important problem in parallel with the rise of threat posed by terrorist organizations such as ISIS and Al-Qaeda. In solving this problem, detection of posts and user accounts that are disseminating these posts, and countering this with extremist narrative have been the main problems that researchers have tackled; however, a thorough and substantial approach to solving those problems is lacking. One challenge that we have is the diverse nature of social media as there are many platforms that offers different capabilities (i.e. micro-blogging, macro-blogging etc.), which makes the form of data different as well. For example, posts on Twitter had had a limitation of 140 character until November 2017, and later it has been increased to 280. This limitation makes the language style being used on Twitter very informal with the use of abbreviations and slang terms;

thus, results in noisy data. Researchers have studied and discussed the problem of online religious extremism in various social media platforms. Although there have been qualitative studies analyzing the content in this problem, there have not been many studies [Mahmood, 2012] that computationally develop a quantitative approach. The works with computational approach used machine learning [Omer, 2015; Scanlon and Gerber, 2014, 2015] and social network analysis techniques [Agarwal et al., 2015] to be able to detect communities of extremist users. Agarwal et al. developed a crawler for mining hate and extremist content shared on Youtube [Agarwal and Sureka, 2014], Tumblr [Agarwal and Sureka, 2016] and Twitter [Agarwal and Sureka, 2015; Sureka and Agarwal, 2014] using their content as well as metadata, and utilized machine learning algorithms to classify posts. They were able to obtain classification accuracy of 0.74 for Youtube, 0.8 for Tumblr and 0.83 for Twitter in identifying hate promoting extremist content. For Twitter, they have conducted only classification of tweets using two classification algorithms, namely SVM (Support Vector Machine) and KNN (k-nearest neighbor), and achieved 0.6 and 0.83 accuracy respectively [Agarwal and Sureka, 2015]. Wadhwa and Bhatia [2013] employed machine learning and social network analysis in classification of tweets and monitoring the detected extremist networks. Anwar and Abulaish [2015] ranked user accounts through a customized version of PageRank algorithm [Brin and Page, 1998], within a forum which is used by Islamist extremists to promote their ideology, influence and eventually recruit individuals. Another study [Cano Basave et al., 2013] investigated the identification of violent content in tweets using a weakly supervised learning model, and this model per-

formed better on classifying violence tweets than the baseline models that were used. Kaati et al. [2015] utilized AdaBoost algorithm to classify extremist content on Twitter.

In a more recent study, Ferrara et al. [2016] developed a framework to predict extremist users, adoption of extremist content and interaction reciprocity between extremists and regular users. They employed Random Forest and Logistic Regression algorithms in building predictive models. They extracted 52 features that includes user and tweet metadata, network, statistical and temporal features. They predict the extremist users in a binary classification, and adoption prediction is performed based on the behavior of regular users in retweeting the tweets posted by extremist users while prediction of interactions with extremists is based on reply tweets. For three prediction tasks, Random Forest outperformed logistic regression. For prediction of extremist users, it performed with 0.87 AUC, for prediction of adoption 0.77 AUC and for prediction of interactions 0.69 AUC. This study has been the most structured and effective study in this domain.

In order to quantify radicalization signals through drifts of users based on their perspective towards favoring Pro vs. Anti-extremist stances, Rowe and Saif [2016] performed a study over 154K users on Twitter. Out of this 154K, 727 users showed Pro-ISIS ideation, especially around an event when ISIS was in news. Such user-level study accounted for homophily among Pro-ISIS users. Saif et al. [2017] proposed a graph-based semantic approach for detection of radicalization on Twitter, which highlights the importance of knowledge bases (e.g. DBpedia) over prior lexical, sentiment, topic or network-based approaches to classifying users as

Pro/Anti-ISIS. They performed the study on 1132 (566 pro-ISIS / 566 anti-ISIS) users totaling a sum of 1.9M (0.6M pro-ISIS / 1.3M anti-ISIS) tweets achieving an F-measure of 0.923 showing significant improvement with semantics over network-based, topic, sentiment, and lexical. However such approaches are vulnerable to closing of Twitter accounts as seen in these studies. First study had 727 Pro-ISIS users while the subsequent study was conducted on 566 users as 161 Twitter accounts were shut down in a period of 1 year.

4.2.2 Neural Language Models (NLMs)

NLMs are a subcategory of neural networks capable of learning sequential dependencies in a sentence, and preserve such information while learning a representation. In particularly, LSTMs (Long Short-Term Memory) networks Hochreiter and Schmidhuber [1997] have emerged from the failure of RNNs (Recurrent Neural Networks) in remembering long-term information. Concerned with the loss of contextual information while learning, Cho et al. [2014] proposed a context-feed forward LSTM architecture in which context learned by the previous layer is merged with forgetting and modulation gates of the next layer. However, if the erroneous contextual information is learned in previous layers, it is difficult to correct. Noisy data and sparsity of terms in the content magnify this problem, resulting in a saddle point of sequential local minima in learning. The inclusion of structured knowledge (e.g., Knowledge Graphs) can be the remedy to such problems in deep learning, as it has improved information retrieval from social media data (e.g., semantic filtering [Sheth and Kapanipathi, 2016b; Phillips et al., 2017])

improved information retrieval.

The ESP algorithm generates groups of features based on their statistical distributions. Enforced Subpopulations [Schmidhuber et al., 2005] together with genetic algorithms form a neuroevolution approach [Gomez and Miikkulainen, 1998], involving the selection of neurons from different tasks to create a subpopulation and performing mutation and crossover on each subpopulation in each generation of learning. Thus, Enforced SubPopulations act as a population generator through hybridizing neurons. Genetic Algorithms generate suitable parameters for smooth and faster convergence. Such a neuroevolution approach has proven to perform well in various applications: gene phenotyping [Schmidhuber et al., 2007], evolving opponents in computer games [Schrum and Miikkulainen, 2008] and the domain of reinforcement learning [Hoekstra, 2011]. Further, Schmidhuber et al. [2005] showed improvement in sequential LSTMs using ESP for generating representations in context-free language learning and using SVM as a predictor of outcome. Fan et al. [2003] proposes a Rule-based Enforced Subpopulations (ESP) scheme for utilizing prior knowledge in an evolving NN to add diversification. The ESP scheme has shown improvement in reinforcement learning, and thus together they make NN generalizable [Stanley and Miikkulainen, 2002].

4.2.3 Neural Attention Models (NAM)

NAM Rush et al. [2015] highlights particular features that are important for pattern recognition/classification based on a hierarchical architecture of the content. The manipulation of attentional focus is effective in solving real-world problems

involving massive amounts of data [Halevy et al., 2009; Sun et al., 2017]. On the other hand, some applications demonstrate the limitation of attentional manipulation in a set of problems such as sentiment (mis)classification [Maurya, 2018] and suicide risk [Corbitt-Hall et al., 2016], where feature presence is inherently ambiguous, just as in the radicalization problem. For example, in the suicide risk prediction task, references to the suicide-related terminology appear in the social media posts of both victims as well as supportive listeners, and the existing NAMs fail to capture semantic relations between terms to help differentiate the suicidal user from a supportive user. To overcome such limitations in a sentiment classification task, Vo et al. [2017], augments sentiment scores in the feature set for enhancing the learned representation and modifies the loss function to respond to values of the sentiment score during learning. However, Sheth et al. [2017]; Kho et al. [2019]; Perera et al. [2016] have pointed out the importance of using domain-specific knowledge especially in cases where the problem is complex. In an empirical study, Bian et al. showed the effectiveness of combining richer semantics from domain knowledge with morphological and syntactic knowledge in the text, by modeling knowledge assistance as an auxiliary task that regularizes learning of the main objective in a deep neural network [Bian et al., 2014].

4.2.4 Knowledge-Guided Neural Networks

Yi et al. [2018] introduced a knowledge-based recurrent attention neural network (KB-RANN) that modifies the attentional mechanism by incorporating domain knowledge to make the model generalize better. However, their domain-knowledge

is statistically derivable from the input data itself and is analogous to merely learning an interpolation function of the existing data. Dugas et al. [2009] proposed a modification in the neural network by adopting Lipschitz functions for its activation function. Zhiting Hu et al. suggested a combination of deep neural networks with logic rules by employing Hinton's Knowledge Distillation procedure [Hinton et al., 2015] for transferring the structured knowledge to the weights of the NN [Hu et al., 2016]. These studies for incorporating knowledge in a deep learning process have not involved declarative knowledge structures in the form of knowledge graphs such as DBpedia. However, Casteleiro et al. [2018] recently showed how the Cardiovascular Disease Ontology (CDO) provided context and reduced ambiguity, improving performance on a synonym detection task. Researchers [Shen et al., 2018] employed embeddings of entities in a knowledge graph, derived through Bi-LSTMs, to enhance the efficacy of neural attention models. Sarker et al. [2017] presents a conceptual framework for explaining artificial neural networks' classification behavior using background knowledge on the semantic web. Makni and Hendler explains a deep learning approach to learn RDFS rules from both synthetic and real-world semantic web data. They also claim their approach improves noise-tolerance capabilities of RDFS reasoning.

4.2.5 Contextual Modeling

Generating embeddings of the content provide a rich numerical vector representation that captures context. Recent embedding algorithms have emerged to create such representations such as Word2Vec by Goldberg and Levy [2014], GLoVe by

Pennington et al. [2014] and FastText by Athiwaratkun et al. [2018]. Besides these state-of-the-art contextual modeling algorithms, for creating high-quality embedding models of a specific domain, it is essential to capture the domain information based on domain-specific corpora.

Modeling the problem of radicalization, which is a pragmatic, context-sensitive phenomenon, vastly depends on carefully designed features to extract meaningful information. We will utilize our existing expertise in designing features based on characteristics of the problem and a ground truth dataset (see Section 4.3). Moreover, differentiating the users and their content requires different levels of granularity in the organization of features. For instance, the information being shared in tweets by a user in radicalization networks displays an intent that depends on the user's type (e.g., recruiter, follower), as recruiters intentionally disseminate information to impress followers and eventually recruit. Hence, as these user types show different characteristics, for reliable analysis, it is critical to bring to bear different perspectives, such as person, content, and network (P-C-N) [Purohit and Sheth, 2013b; Kursuncu et al., 2018] at this higher level, in addition to the content issues already noted at religion, ideology and violence at the lower level.

Our wide expertise on modeling certain problems on social media (User modeling on Twitter [Kapanipathi et al., 2014b; Kursuncu et al., 2018], Modeling Depression [Yazdavar and Hussein, 2017], Disaster preparedness [Al-Olimat et al., 2017b], Intent Mining [Jadhav, 2016], blasphemy on Twitter [Wang et al., 2014a], subjective information on social media [Chen, 2016]) will assure the use of different perspective representations of the content.

4.2.6 Knowledge-Guided Machine Learning

As knowledge involvement in learning processes is one of the critical aims of the proposed research, we leverage our expertise in incorporating knowledge in machine learning applications on social media data. Manually curated medical KGs (UMLS [McInnes et al., 2009], ICD-10 [Brouch, 2000] and DataMed [Ohno-Machado et al., 2017]) provide rich knowledge resource to assist classification of social media posts in healthcare domain. We leveraged these KGs to classify Reddit posts into 21 Mental Disorders (defined in the DSM-5) to facilitate web-based intervention for clinicians. Typical approaches to such classifications employ word embeddings, such as Word2Vec, resulting in sub-optimal performance when used in domain-specific tasks. We have infused knowledge into the embeddings of Reddit posts using Zero Shot learning [Palatucci et al., 2009], and obtained a significant reduction in the false alarm rate, from 30% (without knowledge) to 2.5% (with knowledge) [Gaur et al., 2018]. We also infused knowledge from medical KGs into the content representations by *modulating* (e.g., re-weighting) their embeddings similar to NAMs [Gaur et al., 2018].

4.3 Methodology

Our positive ground truth dataset includes 538 verified radical users and their 47,376 tweets (see Table 4.3) spanning seven years between Oct 2010 and Aug 2017. To test the representations we will create in our experiments, we complement this dataset with 539 non-radical users for our experiments for classification, leveraging

Table 4.3: Statistics of our ground truth dataset

Tweet Metadata	**Value**
Number of Tweets	47376
Number of Unique tweets	25302
Number of Users	538
Time frame	Oct 2010 - Aug 2017
Most Prevalent Concepts and Topics	”kill”, ”muslim”, “attack”, “aleppo”, “allah”, “fight”, “force”, “islam”, “support”, “usa”, “the_islamic_state”, “imam_anwar_al_awlaki”, “join_islamic_state”
Features	Number of follower, Number of following, Gender, Psycholinguistic information derived from LIWC, AFINN, and LabMT (illustrated in [Gaur et al., 2018]) (People)
	Religion, Ideology, Violence, Knowledge (Content)
	Interactions: Retweets, Mentions (Network)

the religious dataset by Chen et al. [2014]. In our positive samples, the prevalent concepts and topics in positive instances usually refer to persons (e.g., ideologue, historic person), locations (region, city) and verbs (fight, join). These terms might appear with different terms in their surrounding in different contexts. Moreover these concepts may be housed in a knowledge graph with their related and similar concepts.

Our approach in modeling users in radical networks on social media incorporates contextual perspectives and domain-specific knowledge. In this section we explain our methodology in detail. The overall architecture in Figure 4.2, guides the description of our methodology depicting critical steps in our approach.

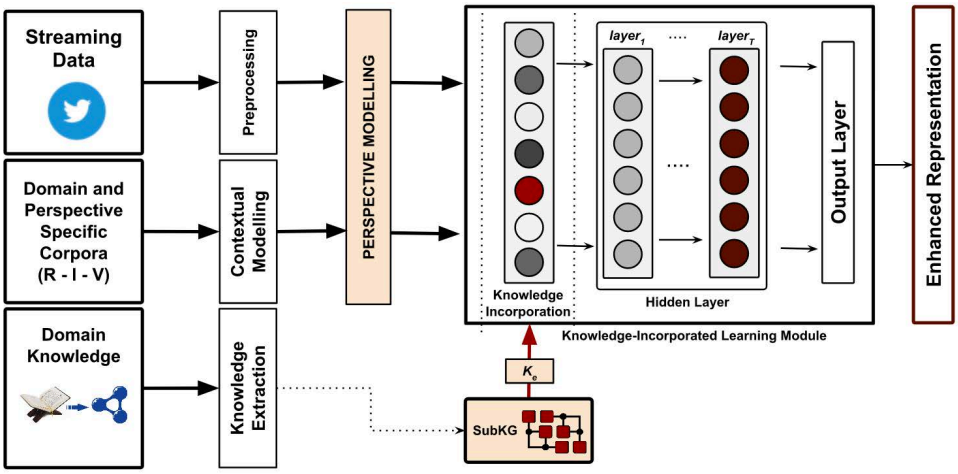


Figure 4.2: Overall Architecture.

As this dataset comprises only radical users, we create 538 Non-Radical users from an annotated muslim religious dataset [Chen et al., 2014] that contains 6040 Muslim users, using Hierarchical Dirichlet Processing (HDP) clustering [Teh et al., 2005]. HDP was employed with an intuition that topics across users are similar. Hierarchical relationship between users can be deduced probabilistically using topical similarity. We obtained 20 topics with 30 sub-topics, and randomly picked 538 from the 600 user clusters.

4.3.1 Perspective Modeling

Terms in the content for each perspective have different weights for importance; however, some diagnostic terms are sparsely distributed or their meaning can be ambiguous in the general domain context. We create three contextual perspectives: Religion, Ideology and Violence. Their representations are generated

through word embedding models created based on perspective-specific corpora as it will reflect the accurate semantic meaning of terms in the content. We use: (i) authentic Islamic resources (e.g., Qur'an, Hadith and their commentary) for the religion perspective, (ii) books and transcribed lectures of radical ideologues that are available online for the ideology perspective, (iii) hate speech and violence corpora for the violence perspective.

Considering the example of term "jihad", it will have different representations in each perspective-specific embedding model since it is being used with different frequency (sparsely occurs in religious context compared to ideology and violence) and surrounded by different terms, which differs its semantic meaning in the contexts of religion, ideology and violence. Traditional approaches for representing the content will fail to capture the nuances in semantic meanings of such essential vocabulary, thus resulting an under-performing model. Generating the three contextual perspective representations of a social media post will emphasize the weights of such essential lexical cues.

Figure 4.3 details the perspective modeling workflow. These contextual perspective models will represent the content in their respective context. Thus, in our example tweets, the terms "jihad" and "paradise" will be represented differently in these three representations based on their lexical and semantic relations in the content. It will assign higher or lower weight to the term depending on its similarity and relatedness with particular concepts such as "war", "martyrdom", "Caliphate" and "struggle".

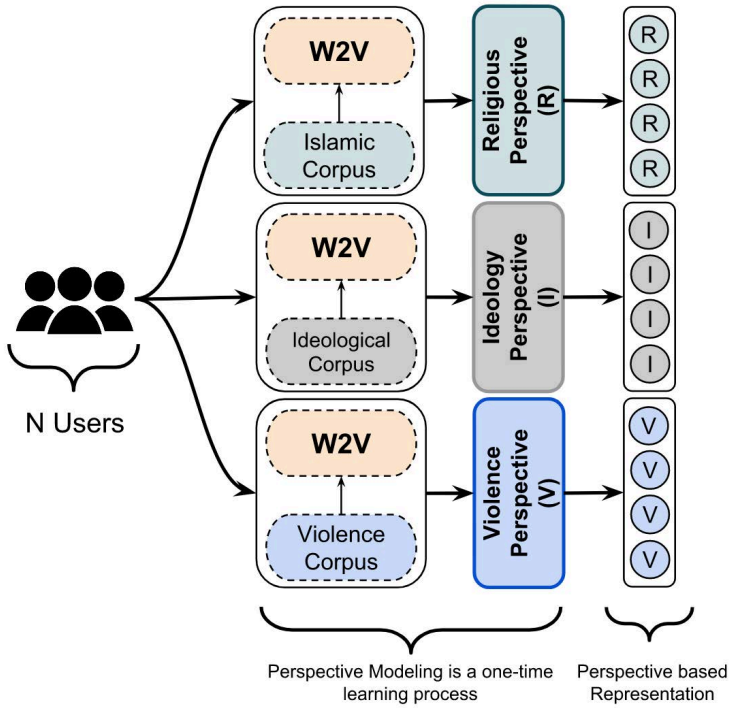


Figure 4.3: Perspective Modeling.

4.3.2 K_e: Knowledge Embedding Creation

We extract related knowledge (e.g., concepts, relations) as a sub-knowledge graph (SubKG) from the KG for the content of each user. We generate representation of knowledge in the SubKG as embedding vectors, and create an embedding of each concept using the perspective models (R, I, V). We combine the embedding vectors of concepts in the SubKG through average operation to obtain representation of each perspective. We leverage the existing structural information of the graph.

This procedure is formally defined:

$$K_e = \frac{1}{n}\sum_{i=0}^{n} C_i^R \oplus \frac{1}{n}\sum_{i=0}^{n} C_i^I \oplus \frac{1}{n}\sum_{i=0}^{n} C_i^V \tag{4.1}$$

where K_e is the representation of the relevant knowledge that is formed through concatenation ($\oplus$) of the three perspectives. C_i is representation of relevant concepts generated through perspective models (e.g., religion, ideology and violence), and n is the number of concepts in the SubKG.

4.3.3 LSTMs for Natural Language Models

Long Short Term Memory network (LSTM) is a special type of RNN. It is specifically designed to learn long-term dependencies by Hochreiter and Schmidhuber [1997]. It is widely used for natural language applications, and its another variant Bidirectional LSTMs (Bi-LSTMs) have also become popular for same goals [1]. We particularly explain LSTM because it provided the best performance in our experiments.

As depicted in figure 4.4, it has C_{t-1}, x_t and h_{t-1} as inputs, and outputs C_t and h_t. The cell state C, horizontal line running through the top of the diagram plays an important role. Gates in LSTMs provides the ability to forget or keep information in the cell state based on the statistical significance of the information piece. It contains sigmoid layer, tanh layer and a pointwise multiplication operation. A sigmoid layer (f_t) outputs numbers between zero and one, deciding how much of each information piece should be forgotten. A tanh layer(i_t) forms a vector with

[1] https://nlp.stanford.edu/manning/talks/Simons-Institute-Manning-2017.pdf

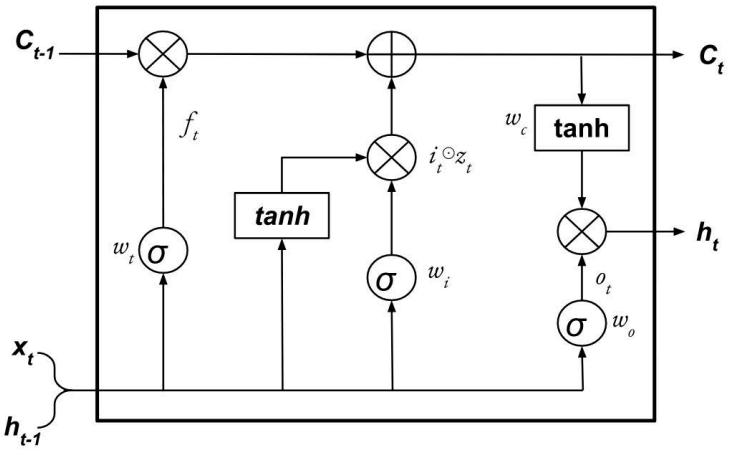


Figure 4.4: Inner Mechanism of an LSTM Cell.

another sigmoid gate (z_t) to update the state.

4.3.4 Experimental Setting

In our learning scheme, we train our model the best possible representations of users through their domain-specific characteristics. In our experimental setting, we utilize uni-bi-tri-perspective representations of a user through concatenation. to assess the effect of each perspective as well as their combination in the learning process. Then, we incorporate knowledge using a domain-specific knowledge graph (e.g., Qur'an ontology by Hakkoum and Raghay [2015]) in the learning process. As the creation of knowledge embedding (K_e) is explained in Section 5.1.1, we concatenate K_e with the perspective representations to see the impact of the relevant knowledge in our learning.

We employ deep neural networks in our experiments since we observed overfit-

ting in traditional machine learning algorithms (SVM, Random Forest and Naive Bayes) based on our perspective-based modeling scheme. For our experiments, we generated representations of the three perspectives for our dataset of 1077 accounts (538 radicals and 539 non-radicals). The Violence perspective model (V) comprises of a vocabulary of 13,255 words, Ideology (I) has a vocabulary of 21,836 words, and Religion (R) has a vocabulary of 186,075 words. Using these perspective-based embedding models, we generate three representations of the user to evaluate the following models: Feed Forward Neural Network (FFNN), Long Short-Term Memory (LSTM), and Bi-Directional Long Short-Term Memory (Bi-LSTM). Selection of these models is influenced by their success in natural language modeling tasks [Zhou et al., 2015; Wen et al., 2018; Chen et al., 2018; Schmidhuber, 2015].

4.4 Results

We have chosen the state-of-the-art baseline model defined in Fernandez et al. [2018] for comparing our approach. Fernandez et al. [2018] utilized a frequency-based weighing scheme with Naive Bayes model for dichotomous classification. Further, we detail our iterative evaluation pipeline involves combinatorial concatenation of different perspectives influencing the persuasive discourse on radicalization.

From the table 4.4, the baseline model holds a significant F-score of 0.84 using a feature size of 23K dimensions through bag of n-grams (unigrams, bi-grams, tri-

grams, and four-grams) because of the independence assumption of Naive Bayes.

In Table 4.4, for violence perspective (V), we noted an improvement of 11% and 3% in precision over the baseline by Bi-LSTM and LSTM respectively, while FFNN did not improve over the baseline. Moreover, deep neural networks under-perform compared to the baseline concerning Recall.

Table 4.4: Evaluation of the Learning Process for User Representations. R: Religious Representation, V:Violence Representation, I:Ideology Representation, K_e: Knowledge Representation (Embedding). $\oplus$ indicates concatenation.

Representation Set	**Algorithm**	**Precision**	**Recall**	**F-score**
Baseline	Naive Bayes	0.88	0.82	0.84
I	FFNN	1	0.51	0.67
V	FFNN	0.88	0.72	0.75
R	FFNN	0.8	0.82	0.81
I	LSTM	0.7	0.66	0.68
V	LSTM	0.914	0.68	0.78
R	LSTM	0.95	0.93	0.94
I	Bi-LSTM	0.75	0.66	0.7
V	Bi-LSTM	0.99	0.78	0.87
R	Bi-LSTM	0.95	0.94	0.94
V $\oplus$ I	FFNN	1	0.53	0.7
R $\oplus$ I	FFNN	1	0.51	0.67
R $\oplus$ V	FFNN	0.86	0.67	0.75
V $\oplus$ I	LSTM	0.87	0.7	0.78
R $\oplus$ I	LSTM	0.83	0.75	0.79
R $\oplus$ V	LSTM	0.875	0.75	0.81
V $\oplus$ I	Bi-LSTM	0.87	0.82	0.84
R $\oplus$ I	Bi-LSTM	0.91	0.85	0.88
R $\oplus$ V	Bi-LSTM	0.9	0.88	0.89
R $\oplus$ I $\oplus$ V	FFNN	0.88	0.79	0.8
R $\oplus$ I $\oplus$ V	LSTM	0.94	0.9	0.92
R $\oplus$ I $\oplus$ V	Bi-LSTM	0.96	0.94	0.95
R $\oplus$ I $\oplus$ V $\oplus$ K_e	Bi-LSTM	0.96	0.96	0.96

For Ideology perspective (I), only FFNN outperforms the baseline regarding

precision while LSTM and Bi-LSTM improve 7% each upon the baseline in terms of precision for Religious perspective (R). Further, 12% and 13% improvement in the recall were shown by these models in comparison to the baseline. However, FFNN finished close to the baseline. Although we see a significant improvement on the F-score 0.94 for R with LSTM and Bi-LSTM, we investigate further improvement employing multiple perspectives and knowledge.

Employing perspectives V and I, for FFNN, we don't see significant thrust in recall with high precision. On the other hand, for I and R, we observed Bi-LSTM outperforming the baseline with an increase of 3% and 3.5% in precision and recall respectively. Bi-LSTM showed similar behavior for V and R with a rise of 2% and 7% in precision and recall respectively compared to baseline. This Bi-perspective integration experiment showed the essence of perspective-based analysis for identifying radicalism in social media. Further, such an integration semantically combines the representation of the words rather than considering them as individual pieces of information as in [Fernandez et al., 2018].

We further extend our study with Tri-perspective integration involving V, I, and R. We observed a compensating improvement of 6% and 9% for LSTM, and 8% and 13% for Bi-LSTM in precision and recall outperforming the baseline. Our perspective-focused integration procedure minimizes fp and fn improving the classification results.

As Bi-LSTM with the tri-perspective representation performed best, we incorporate knowledge representation in the form of embedding vectors, derived from Qur'an ontology and our contextual perspective models. Complementing the tra-

ditional bottom approaches that rely on sole corpora, with a top-down approach such as declarative domain-specific knowledge, even raises our bar for the performance of the model by 8% and 15% for precision and recall.

4.5 Conclusion

The goal of this study was to create representations of users through strategically identified domain-specific perspectives and relevant knowledge representation for more robust classification. Such rich representations of users will enable to classify users further into multiple classes of radicalization, based on a scale determined by domain experts. Furthermore, it can allow researchers to understand the radicalization process over time through radicalization stages concerning religion, ideology and violence.

In this research, we provide an approach to generate representations of users in radical and non-radical networks on social media, that will improve classification upon the state-of-the-art. We identified three contextual perspectives of the problem of radicalization on social media and learned three domain-specific embedding model for content of users. We also incorporated knowledge embeddings created through a domain-specific knowledge graph and the three perspective models that we created. Overall, our comprehensive approach achieved 8%, 14% and 12% improvement in precision, recall and F-score respectively over the baseline, when we used R,I,V representations with the knowledge representation.

Chapter 5

Conclusion

Social media has become a major communication pathway in recent years providing instant access to masses. Conventional learning mechanisms detect target content from such social media data, permitting for example the analysis of public opinion. However, a certain class of detection problems–persuasive social data–challenges the state of the art. We have conducted two case studies in the marijuana and radicalization related communications to test effectiveness of our approach that involves domain specific information. We leverage the rich nature of social media data to extract and design feature and incorporate domain knowledge in the learning scheme.

Moreover, the research presented in this document can be extended that will further enhance modeling techniques and strategies using generative and declarative modeling techniques in one scheme. In the following section, we will discuss remaining challenges exemplifying from the use case of radicalization and future

directions that we foresee.

5.1 Future Directions

The use of social media to spread Islamist extremism and radicalization is one example of persuasive social data, extended over time and at least initially, cloaked by ambiguous intentionality. For instance, the concept "jihad" commonly appears in mainstream Islam, as well as radical discourse, albeit with a different context-dependent meaning. Contemporary bottom-up analysis is ineffective in the face of such ambiguity and target sparsity, further challenged by a process of persuasion that starts out benign and over time turns increasingly radical. We model this process as the interaction among connected agents with a mix of perspectives, each one of which exemplifies a degree of radicalization and depends crucially on the proper identification of relevant message features. We infuse domain knowledge of Islamist radical ideology in deep learning models to relate linguistic features spanning religion, ideology, and violence to classify discourse along an established 5-level radicalization scale. Combined with a network of agent models, a carefully constructed sequence of discourse content persuades the primed recipient to descend into radicalism. Using Islamist extremism and radicalization as the focal use case, our knowledge-driven and context-aware learning approach generalizes to persuasive social data problems in other domains such as politics and economics.

Religious radicalization represents a class of rarely studied problems that do not yield to contemporary bottom-up machine learning methods. The initial discourse

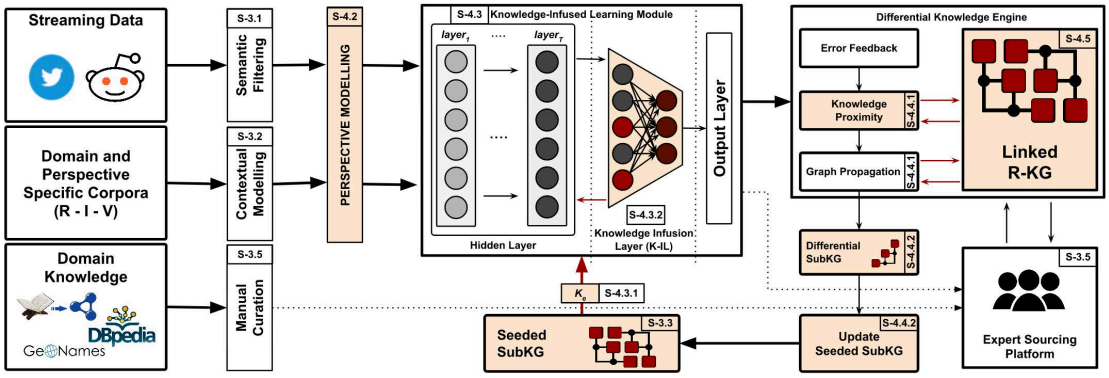


Figure 5.1: Overall Architecture

corpus is large, but the signals are sparse, creating an extreme class imbalance. The signals are ambiguous, properly members of at least two subsuming domains (i.e., radical vs. main-stream) that provide message context, threatening precision. The process of persuasion between a recruiter and a target also offers unique content characteristics.

Our methods will disambiguate important concepts defined in the Radicalization KG (R-KG) with their different semantic meanings through its structural relations. Knowledge incorporation will redefine the emphasis of sparse but essential and irrelevant but frequently occurring terms and concepts, boosting recall without reducing precision. Our novel, transformative learning approach will marry top-down and bottom-up approaches in one framework, providing explanatory insight into the model, robustness to noise and reducing dependency on frequency in the learning process.

Our innovations seek to operationalize more abstract models of behavior from

religion, political and psychology to render them computationally accessible. A knowledge-based approach structures search within the feature space for deep learning. If successful, this will transform how we study social media discourse with respect to content, person and networks. We advance content analysis on social media by using levels as classes to reflect the persuasion process. Perspective modeling coupled with knowledge infusion in a neural networks allows us to address a challenging class of problems with inherent sparsity and ambiguity in which the relations are implicit in the data.

5.1.1 Knowledge-Infused Learning Module

Each layer in a neural network architecture produces a latent representation of the input vector. As neural network consists of an input layer, hidden layers and output layer, external information has been incorporated before the input layer and after the output layer. Infusion after the input, within the hidden layer or before the output layer have not been investigated. we infuse knowledge within the neural network while the latent representation is transmitted between layers including hidden layers. The infusion of knowledge during the representation learning phase raises the following central research questions, (i) *Knowledge-Aware Loss Function (K-LF):* How do we decide whether to infuse knowledge or not at a particular stage in learning between layers, and how to measure the incorporation of knowledge? (ii) *Knowledge Modulation Function (K-MF):* How to merge latent representations with knowledge representations, and How to propagate the knowledge through the learned representation?

Configurations of neural networks can be designed in various ways depending on the problem. As our aim is to infuse knowledge within the neural network, such operation can take place (i) before the output layer (e.g., SoftMax), (ii) between hidden layers (e.g., reinforcing the gates of an NLM layer, modulating the hidden states of NLM layers, Knowledge-driven NLM dropout and recurrent dropout between layers). To illustrate (i), we describe our initial approach to neural language models that fuses knowledge before the output layer.

In the subsequent subsections, we explain: (a) Creation of Knowledge representations (e.g., Knowledge embeddings, K_e), (b) Knowledge Infusion Layer is responsible for the two proposed functions. In these subsections, we provide an initial approach that, we believe, will shed the light towards a reliable and robust solutions with more research and rigorous experimentations.

K_e: Knowledge Embedding Creation

We generate representation of knowledge in the Seeded SubKG as embedding vectors. We create an embedding of each concept and their relations in the Seeded SubKG using the perspective models (R, I, V), and merge these embeddings through the proximity of their concepts and relations in the graph. Unlike traditional approaches that compute the representation of each concept in the KGs by simply taking average of embedding vectors of concepts, we leverage the existing structural information of the graph. This procedure is formally defined:

$$K_e = \sum_{ij} [C_i, C_j] \bigotimes D_{ij} \tag{5.1}$$

where K_e is the representation of the concepts enriched by the relationships in the Seeded-KG, (C_i, C_j) is the relevant pair of concepts in the Seeded-KG, $\mathbf{D_{ij}}$ is the distance measure (e.g., Least Common Subsumer [Baader et al., 2007]) between the two concepts C_i and C_j. We will further examine novel methods building upon our initial approach above as well as existing tools that include TRANS-E [Bordes et al., 2013], TRANS-H [Wang et al., 2014b], and HOLE [Nickel et al., 2016] for the creation of embeddings from KGs.

Knowledge Infusion Layer

In a many-to-one NLM [Shivakumar et al., 2018] network with **T** hidden layers, the $\mathbf{T^{th}}$ layer contains the learned representation before the output layer. The output layer (e.g., SoftMax) of the NLM model will estimate the error to be back-propagated. As discussed above, knowledge infusion can take place between hidden layers or just before the output layer. We will explore techniques for both scenarios. In this subsection, we explain the Knowledge Infusion Layer (K-IL) which takes place just before the output layer.

Algorithm 1 takes the **PSP**, the type of neural language model, number of epochs, iterations and the seeded knowledge graph embedding $\mathbf{K_e}$ as input, and returns a knowledge fused representation of the hidden state $\mathbf{M_T}$. In line 4, the fusion of the knowledge happens after each epoch without obstructing the learning of the vanilla NLM model and is explained by line 5-10. Within the knowledge fusion process (line 7-9), we optimize the loss function in equation (1) with convergence condition defined as the reduction in the difference between the $\mathbf{D_{KL}}$ of

$\mathbf{h_T}$ and $\mathbf{h_{T-1}}$ in the presence of K_e. Considering the vanilla structure of a NLM [Greff et al., 2017], $\mathbf{M_T}$ is utilized by the fully connected layer for classification.

To illustrate our initial approach in Figure 5.2b, we use LSTMs as NLMs in our neural network. K-IL functions an additional layer before the output layer of our proposed neural network architecture. This layer takes the latent vector ($\mathbf{h_{T-1}}$) of the penultimate layer, the latent vector of the last hidden layer ($\mathbf{h_T}$) and the knowledge embedding (K_e), as input.

Algorithm 1 Routine for Infusion of Knowledge in NLMs

procedure KNOWLEDGEINFUSION
  $Data : PSP, NLM_{type}, \#Epochs, \#Iterations, K_e$
  $Output : \overrightarrow{M_T}$
  for ne=1 to #Epochs **do**
    $\overrightarrow{h_T}$, $\overrightarrow{h_{T-1}} \leftarrow$ TrainingNLM(PSP, NLM_{Type},$\#Iterations$)
    while $(\mathbf{D}_{KL}(\overrightarrow{h_{T-1}}||\overrightarrow{K_e}) - \mathbf{D}_{KL}(\overrightarrow{h_T}||\overrightarrow{K_e}) > \varepsilon)$ **do**
      $h_T \leftarrow \sigma(W_{hk} * (\overrightarrow{h_T} \oplus \overrightarrow{K_e}) + b_{hk})$
      $W^{hk} \leftarrow W^{hk}$ - $\eta_k \nabla(K - LF)$
    $\overrightarrow{M_T} \leftarrow \overrightarrow{h_T} \odot W^{hk}$
  return: $\overrightarrow{M_T}$

In this layer, we define two particular functions that will be critical for merging the latent vectors from the hidden layers and the knowledge embedding vector from the R-KG. Note that dimensions of these vectors are same because they are created from same embedding models (e.g., R, I, V perspective word embedding models)(see Sections 4.3.1 and Section 5.1.1), which makes the merge operation of those vectors possible and valid.

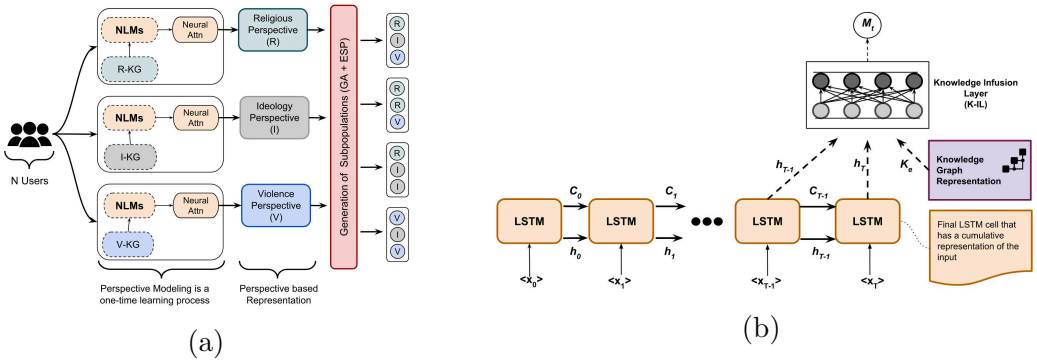


Figure 5.2: (a) Perspective Modeling Diagram (b) Inner Mechanism of the Knowledge Infusion Layer

K-LF: Knowledge-Aware Loss Function: In DL hidden layers of neural networks may de-emphasize important patterns due to the sparsity of certain features during the learning process, which causes information loss. In some cases, such patterns may not even appear in the data. However, such relations or patterns may be defined in KGs with the relevant knowledge. We call this information gap between the learned representation of the data and knowledge representation as *differential knowledge*. Information loss in a learning process is relative to the distribution that suffered the loss. Hence, we plan to develop a measure to be used to determine the *differential knowledge* and guide the degree of knowledge infusion in learning. As our initial approach to this measure, we developed a two-state regularized loss function by utilizing Kullback Leibler (KL) divergence. Our choice of KL divergence measure is largely influenced by the Markov assumptions made in language modeling and have been highlighted in [Longworth, 2010]. The K-LF measure estimates the divergence between the hidden representations ($\mathbf{h_{T-1}}, \mathbf{h_T}$)

and knowledge representation (K_e), to determine the differential knowledge to be infused.

Formally we define it as: $\arg\min(\vec{h_{T-1}}, \vec{h_T}, \vec{K_e}) \equiv K - LF$, where $\mathbf{h_{T-1}}$ is an input for convergence constraint.

$$\mathbf{K - LF} = min\ \mathbf{D}_{KL}(\vec{h_T}||\vec{K_e});\quad s.t.\ \ \mathbf{D}_{KL}(\vec{h_T}||\vec{K_e}) < \mathbf{D}_{KL}(\vec{h_{T-1}}||\vec{K_e}) \tag{5.2}$$

We minimize the *relative entropy* for information loss to maximize the information gain from the knowledge representation (e.g., K_e). We will compute differential knowledge ($\nabla\mathbf{K - LF}$) through such optimization approach; thus, the computed differential knowledge will also determine the degree of knowledge to be infused in the K-IL. $\nabla\mathbf{K - LF}$ will be computed in the form of embedding vectors, and the dimensions from K_e will be preserved.

K-MF: Knowledge Modulation Function: We need to merge the differential knowledge representation with the learned representation. However, such operation cannot be done arbitrarily., We explain an initial approach for the K-MF to modulate the learned weight matrix of the neural network with the hidden vector through an appropriate operation (e.g., Hadamard pointwise multiplication). This operation at the $\mathbf{T^{th}}$ layer can be formulated as:
Equation for $W^{hk} = W^{hk} - \eta_k * \nabla\mathbf{K - LF}$, where W^{hk} is the learned weight matrix infusing knowledge, η_k is learning momentum Sutskever et al. [2013], $\nabla\mathbf{K - LF}$ is differential knowledge. The weight matrix (W^{hk}) is computed through the learning epochs utilizing the differential knowledge embedding ($\nabla\mathbf{K - LF}$). Then we

merge W^{hk} with the hidden vector $\mathbf{h_T}$ through the K-MF. Considering that we use Hadamard pointwise multiplication as our initial approach, we formally define the output $\mathbf{M_T}$ of K-MF as: This operation at the $\mathbf{T^{th}}$ layer can be formulated as:

$$\vec{M_T} = \vec{h_T} \odot W^{hk} \tag{5.3}$$

where $\mathbf{M_T}$ is Knowledge-Modulated representation, $\mathbf{h_T}$ is the hidden vector and W^{hk} is the learned weight matrix infusing knowledge. Further investigations of techniques for K-MF, will be one of the main research topics in the agenda of this proposed research.

5.1.2 Differential Knowledge Engine

In deep neural networks, each epoch generates an error that is back-propagated until the model reaches a saddle point in the local minima, and the error is reduced in each epoch. The error indicates the difference between probabilities of actual and predicted labels, and such difference can be used to enrich the Seeded SubKG in our proposed knowledge-infused deep learning framework.

In this section, we discuss the sub-knowledge graph operations that are based on the difference between the learned representation of our knowledge-infused model ($\mathbf{M_T}$), and the representation of the relevant sub-knowledge graph from the R-KG, which we call as differential sub-knowledge graph. We define *Knowledge Proximation function* to generate the *Differential Sub-knowledge Graph*, and *Update Seeded SubKG* to insert the differential sub-knowledge graph into the Seeded

SubKG.

Knowledge Proximity

Upon the arrival of the learned representation from the knowledge-infused learning model, we query the R-KG for retrieving related information to the respective data point. In this particular step, it is important to find the optimal proximity between the concept and its related concepts. For example, from the "martyrdom" concept, we may traverse the surrounding concepts with different number of hops (empirically decided). We plan to investigate finding the optimal number of hops towards each direction from the concept in question. As we find optimal proximity of a particular concept in the KG, we propagate R-KG based on the proximation starting from the concept in question.

Differential SubKG

Once we obtain the SubKG from the graph propagation, we create differential SubKG that will reflect the difference in knowledge from the Seeded SubKG. For this procedure, we plan to carry out research formulating the problem using variational autoencoders to extract such SubKG as we call *differential subKG*($\mathbf{D_{kg}}$) and, we believe it will provide missing information in the Seeded-KG.

Update function

The differential subKG generated as a result of minimizing knowledge proximation is considered as input factual graph to the update procedure. As a result, the

procedure dynamically evolves the Seeded subKG with missing information from differential subKG. We plan to utilize *Lyapunov stability theorem* [Liu et al., 2014] and *Zero Shot learning* to update the Seeded-KG using D_{kg}. D_{kg} and Seeded-KG represent two knowledge structures requiring a process of transfer the knowledge from one structure to another [Hamaguchi et al., 2017]. We define it as the process of generating semantic mapping weights that encodes and decodes the two semantic spaces. We plan to utilize the Lyapunov stability constraint and Sylvester optimization approach: Given two semantic spaces belonging to a domain D (in this case radicalization), we tend to attain an equilibrium position defined as:

$$||S_{kg} - W * D_{kg}||_F = \alpha * ||W * S_{kg} - D_{kg}||_F \tag{5.4}$$

$||.||_F$ represents Frobenius norm and α is a proportionality constant belong to $\mathbb{R}$. Equation 5.4 reflects lyapunov stability theorem and to achieve such a stable state we define our optimization function as follows:

$$L = min(||S_{kg} - WD_{kg}||_F - \alpha * ||WS_{kg} - D_{kg}||_F), \alpha > 0, W \in \mathbb{R}X\mathbb{R} \tag{5.5}$$

Equation 5.5 is solvable using Sylvester optimization and its derivation is defined in a recent study [Gaur et al., 2018].

5.1.3 R-KG: Radicalization Knowledge Graph

The Radicalization Knowledge Graph plays a key role in our framework as it will be extensively used by several functions. Radical text includes content that appears

as leaves in two pathways of a subsuming graph: the extremist parents and the legitimate Islamic religious practice. We will develop the R-KG by capturing and manually curating the frequently used terms and concepts in radical content, such as (but not limited to) *jihad, kafir/kufar [infidel], Al-Baghdadi, caliphate, murtad, haram, Sharia* and their relationships. Manual curation will be conducted through our expert-sourcing platform by our domain experts (see support letters). The R-KG will differentiate semantic and contextual nuances of concepts in potential radical content. Domain experts in our team will maintain and monitor the evolution of the R-KG as new concepts are added with their relations. R-KG will also be linked to domain-specific and general knowledge graphs such as DBPedia and the Islamic knowledge sources including the Qur'an and the books of Hadith (Prophetic Narrations) [Harrag et al., 2011]. An Islamic KB that links these Islamic resources in electronic format at the macro and micro levels will be utilized [Basharat et al., 2016]. Linking the R-KG with these knowledge graphs will provide access to more related knowledge. We will leverage our expert-sourcing tool for knowledge that is extracted from the classified social media data. The extracted information as candidate concepts and relations in the Radicalization KG (R-KG), will be evaluated for its relevance to Islamist radicalization by three domain experts. Then the information will be passed on to the R-KG to be attached if approved by majority of the experts; if not it will be discarded.

5.1.4 Evaluation Plan

We have identified three key issues that challenge conventional machine learning algorithms. Hence our evaluation focuses on how well our approach handles these three issues. In addition, we evaluate our contributing models.

Sparsity Evaluation

Defined as the model's ability to function precisely in the presence or absence of sparsity in the content. This is an essential problem concerning our proposal as domains like Radicalization do not generate a large amount of positive instances. For evaluating the model's sensitivity to sparsity, we plan to utilize the following two metrics described in prior research: (1) AUC [Krishnan et al., 2017], (2) Gini Ratio [Guest and Love, 2017], (3) Kolgomorov-Smirnov (K-S Test) (or Chi-square) test [Gómez et al., 2008], (4) Information Divergence measures [Karacan et al., 2015]. The K-S Test is proposed with an assumption that ground truth annotated data represent one distribution and other is generated by our approach.

Ambiguity Evaluation

Defined as the model's ability to distinctively characterize a user on the radicalization scale. As there are multiple outcomes (or called as actions) based on the scale, we employ counterfactual assessment measures for evaluating the model. Based on the prior literature, initially we plan on utilizing; (1) Inverse Probability (or Propensity) Weighting [Braun et al., 2016], (2) Cumulative Reward or Regret [Guo, 2017], (3) Hamming Loss [Saxena, 2018] or Jaccard Score [Issa et al., 2018],

and (4) Mean Absolute Error (MAE). Although MAE is generic, we believe it to be convincing in evaluating the model's tendency to discriminate between different outcome labels and appropriately classify the user.

Noise Sensitivity Evaluation

Defined as the consistency in the outcome of the model with or without the presence of noise in the data. F-measure is a good metric to start, with but it does not account for noise sensitivity. One can abstractly explain with reference to Precision and Recall. However, we plan to utilize R-squared and adjusted R-squared metric using the continuously valued metrics that our model returns. Using this metric, we evaluate two models: (1) In the presence of noisy data, and (2) In the absence of noisy data. In the event of non-normal distributed values, may employ a Rank correlation coefficient.

Evaluation of Perspective Models

Considering an analogy of subpopulations from GA-EsP to diverse crowds, we propose to evaluate the subpopulations using Monte-Carlo methods illustrated in [Bhatt et al., 2017b]. In such an evaluation we consider a null-scenario baseline of uniform populations (e.g. R-R-R, V-V-V, I-I-I) for comparison. Further, we plan to create a gold standard of perspectives in radicalization domain for supervised evaluation using an F1-measure.

Evaluation of Subgraph

In our planned study, we define an architecture of dynamically evolve the seeded knowledge graph ($SeededK_e$, a subgraph). But, this assumes a quality subgraph that is evolved by supervising the learning from NLMs. Recent work highlights the use of error detection, completeness and Information-theoretic (Normalized Mutual Information [Wang et al., 2018a], Jaccard Similarity, Jenson-Shannon Divergence) approaches for KG quality evaluation [Paulheim, 2017]. We plan to align our preliminary groundwork along two evaluation measures: Count of Temporal Conflicts: Measured as a factor of increase in misclassification after missing knowledge is added to $SeedK_e$ from reference knowledge graph. Minimum Information Discrepancy: We measure the noticeable information discrepancy by employing Shannon entropy and similarity measures elucidated in [Chowdhury et al., 2017].

Bibliography

Dilshod Achilov and Sedat Sen. Got political islam? are politically moderate muslims really different from radicals? *International Political Science Review*, 38(5):608–624, 2017.

Apoorv Agarwal, Boyi Xie, and Ilia Vovsha. Sentiment analysis of twitter data. In *Proceedings of the Workshop on Language in Social Media (LSM 2011)*, number June, pages 30–38, 2011.

Swati Agarwal and Ashish Sureka. A focused crawler for mining hate and extremism promoting videos on youtube. In *Proceedings of the 25th ACM conference on Hypertext and social media*, pages 294–296. ACM, 2014.

Swati Agarwal and Ashish Sureka. Using knn and svm based one-class classifier for detecting online radicalization on twitter. In *International Conference on Distributed Computing and Internet Technology*, pages 431–442. Springer, 2015.

Swati Agarwal and Ashish Sureka. Spider and the flies: Focused crawling on tumblr to detect hate promoting communities. *arXiv preprint arXiv:1603.09164*, 2016.

Swati Agarwal, Ashish Sureka, and Vikram Goyal. Open source social media

analytics for intelligence and security informatics applications. In *International Conference on Big Data Analytics*, pages 21–37. Springer, 2015.

Dirk Ahlers. Assessment of the accuracy of geonames gazetteer data. In *GIR*, 2013.

Anuj J Aiswal, Wei Peng, and Tong Sun. Predicting Time-sensitive User Locations from Social Media. In *ASONAM*, 2013.

Hussein S Al-Olimat, Krishnaprasad Thirunarayan, Valerie Shalin, and Amit Sheth. Location Name Extraction from Targeted Text Streams using Gazeeer-based Statistical Language Models. *arxiv preprint*, 11(17), 2017a.

Hussein S Al-Olimat, Krishnaprasad Thirunarayan, Valerie Shalin, and Amit Sheth. Location name extraction from targeted text streams using gazetteer-based statistical language models. *arXiv preprint arXiv:1708.03105*, 2017b.

Hunt Allcott and Matthew Gentzkow. Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives—Volume*, 31(2—Spring):211–236, 2017. doi: 10.1257/jep.31.2.211.

Jalal S Alowibdi, Ugo A Buy, Philip S Yu, and Leon Stenneth. Detecting Deception in Online Social Networks. In *ASONAM*, 2014.

Tarique Anwar and Muhammad Abulaish. Ranking radically influential web forum users. *IEEE Transactions on Information Forensics and Security*, 10(6):1289–1298, 2015.

I.B. Arpinar, U. Kursuncu, and D. Achilov. Social media analytics to identify and counter islamist extremism: Systematic detection, evaluation, and challenging of extremist narratives online. In *Proceedings - 2016 International Conference on Collaboration Technologies and Systems, CTS 2016*, 2016. ISBN 9781509022991. doi: 10.1109/CTS.2016.113.

Ben Athiwaratkun, Andrew Gordon Wilson, and Anima Anandkumar. Probabilistic fasttext for multi-sense word embeddings. *arXiv preprint arXiv:1806.02901*, 2018.

Franz Baader, Baris Sertkaya, and Anni-Yasmin Turhan. Computing the least common subsumer wrt a background terminology. *Journal of Applied Logic*, 5 (3):392–420, 2007.

Lakshika Balasuriya, Sanjaya Wijeratne, Derek Doran, and Amit Sheth. Finding Street Gang Members on Twitter. In *ASONAM*, 2016.

Bamler and Mandt. Dynamic word embeddings. In *ICML*, 2017.

David Bamman, Jacob Eisenstein, and Tyler Schnoebelen. GENDER IN TWITTER: STYLES, STANCES, AND SOCIAL NETWORKS. *CoRR*, 2012.

Amna Basharat, Bushra Abro, Ismailcem Budak Arpinar, and Khaled Rasheed. Semantic hadith: Leveraging linked data opportunities for islamic knowledge. In *LDOW@ WWW*, 2016.

A Benton, R Arora, and M Dredze. Learning multiview embeddings of twitter users. In *ACL*, 2016.

S Bergsma, M Dredze, B Van Durme, T Wilson, and D Yarowsky. Broadly improving user classification via communication-based name and location clustering on twitter. In *NAACL-HLT*, 2013.

Adham Beykikhoshk, Ognjen Arandjelovi, Dinh Phung, and Svetha Venkatesh. Data-Mining Twitter and the Autism Spectrum Disorder: A Pilot Study. In *ASONAM*, 2014.

Shreyansh Bhatt, Hemant Purohit, and Andrew Hampton. Assisting Coordination during Crisis: A Domain Ontology based Approach to Infer Resource Needs from Tweets. In *Web Science*, 2014.

Shreyansh Bhatt, Brandon Minnery, Srikanth Nadella, Beth Bullemer, Valerie Shalin, and Amit Sheth. Enhancing crowd wisdom using measures of diversity computed from social media data. In *Proceedings of the International Conference on Web Intelligence*, 2017a. doi: 10.1145/3106426.3106491.

Shreyansh Bhatt, Brandon Minnery, Srikanth Nadella, Beth Bullemer, Valerie Shalin, and Amit Sheth. Enhancing crowd wisdom using measures of diversity computed from social media data. In *Proceedings of the International Conference on Web Intelligence*, pages 907–913. ACM, 2017b.

N. Bhattacharya, I.B. Arpinar, and U. Kursuncu. Real Time Evaluation of Quality of Search Terms during Query Expansion for Streaming Text Data Using Velocity and Relevance. In *Proceedings - IEEE 11th International Conference on Semantic Computing, ICSC 2017*, 2017. ISBN 9781509048960. doi: 10.1109/ICSC.2017.105.

Jiang Bian, Bin Gao, and Tie-Yan Liu. Knowledge-powered deep learning for word embedding. In *Joint European conference on machine learning and knowledge discovery in databases*, pages 132–148. Springer, 2014.

Imen Bizid, Nibal Nayef, Patrice Boursier, Sami Faiz, and Jacques Morcos. Prominent Users Detection during Specific Events by Learning On-and Off-topic Features of User Activities. In *ASONAM*, 2015.

Terra Blevins, Robert Kwiatkowski, Jamie Macbeth, Kathleen Mckeown, Desmond Patton, and Owen Rambow. Automatically Processing Tweets from Gang-Involved Youth: Towards Detecting Loss and Aggression. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2196–2206, 2016.

Han Bo, Paul Cook, T Imoth, and Bal Dw. Geolocation Prediction in Social Media Data by Finding Location Indicative Words. In *Proceedings of COLING 2012*, pages 1045–1062, 2012.

Marina Boia and Boi Faltings. A :) Is Worth a Thousand Words: How People Attach Sentiment to Emoticons and Words in Tweets. In *SocialCom*, 2013. doi: 10.1109/SocialCom.2013.54.

Phillip Bonacich. Power and centrality : A family of measures. *American Journal of Sociology*, 92(5):1170–1182, 1987.

Kalina Bontcheva, Leon Derczynski, Adam Funk, Mark a Greenwood, Diana Maynard, and Niraj Aswani. TwitIE : An Open-Source Information Extraction

Pipeline for Microblog Text. *Proceedings of Recent Advances in Natural Language Processing*, (September):83–90, 2013. ISSN 13138502.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems*, pages 2787–2795, 2013.

Lorraine Bowman-Grieve and Maura Conway. Exploring the form and function of dissident irish republican online discourses. *Media, War & Conflict*, 5(1):71–85, 2012.

Danielle Braun, Corwin Zigler, Francesca Dominici, and Malka Gorfine. Using validation data to adjust the inverse probability weighting estimator for misclassified treatment. *Using Validation Data to Adjust the Inverse Probability Weighting Estimator for Misclassified Treatment*, 2016.

Felipe Bravo-Marquez, Daniel Gayo-Avello, Marcelo Mendoza, and Barbara Poblete. Opinion Dynamics of Elections in Twitter. In *Eighth Latin American Web Congress*, 2012. doi: 10.1109/LA-WEB.2012.11.

Sergey Brin and Lawrence Page. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks and ISDN Systems*, 1998.

Kathy L Brouch. Where in the world is icd-10? *Where in the World Is ICD-10?/AHIMA, American Health Information Management Association*, 2000.

Peter Brusilovsky. Methods and techniques of adaptive hypermedia. *User modeling and user-adapted interaction*, 6(2-3):87–129, 1996.

BJ Bushman and LR Huesmann. Short-term and long-term effects of violent media on aggression in children and adults. *Arch Pediatr Adolesc Med*, 2006. doi: doi:10.1001/archpedi.160.4.348.

Delroy Cameron, Gary A Smith, Raminta Daniulaityte, Amit P Sheth, Drashti Dave, Lu Chen, Gaurish Anand, Robert Carlson, Kera Z Watkins, and Russel Falck. PREDOSE: A semantic web platform for drug abuse epidemiology using social media. *Journal of Biomedical Informatics*, 46:985–997, 2013. doi: 10.1016/j.jbi.2013.07.007.

WM Campbell, E Baseman, and K Greenfield. Content+ context networks for user classification in twitter. In *NIPS*, 2013.

Amparo Elizabeth Cano Basave, Yulan He, Kang Liu, and Jun Zhao. A weakly supervised bayesian model for violence detection in social media. 2013.

Mercedes Arguello Casteleiro, George Demetriou, Warren Read, Maria Jesus Fernandez Prieto, Nava Maroto, Diego Maseda Fernandez, Goran Nenadic, Julie Klein, John Keane, and Robert Stevens. Deep learning meets ontologies: experiments to anchor the cardiovascular disease ontology in the biomedical literature. *Journal of biomedical semantics*, 9(1):13, 2018.

Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. Information Credibility on Twitter. In *Proceedings of the 20th international conference on World wide web*, pages 675–684. ACM, 2011.

NV Chawla, KW Bowyer, LO Hall, and WP Kegelmeyer. Smote: synthetic minority over-sampling technique. *JAIR*, 2002.

Cuixian Chen, Yaw Chang, Karl Ricanek, and Yishi Wang. Face age estimation using model selection. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, pages 93–99, 2010. doi: 10.1109/CVPRW.2010.5543820.

Lu Chen. Mining and analyzing subjective experiences in user generated content. 2016.

Lu Chen, Chen@knoesis Org, Wenbo Wang, Wenbo@knoesis Org, Meenakshi Nagarajan, Shaojun Wang, Amit P Sheth, and Amit@knoesis Org. Extracting Diverse Sentiment Expressions with Target-Dependent Polarity from Twitter. In *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*, 2012a.

Lu Chen, Wenbo Wang, and Amit P Sheth. Are Twitter Users Equal in Predicting Elections? A Study of User Groups in Predicting 2012 U.S. Republican Presidential Primaries. In *Social Informatics*, 2012b.

Lu Chen, Ingmar Weber, and Adam Okulicz-Kozaryn. Us religious landscape on twitter. In *International Conference on Social Informatics*, pages 544–560. Springer, 2014.

Ying Chen, Sencun Zhu, Yilu Zhou, and Heng Xu. Detecting Offensive Language in Social Media to Protect Adolescent Online Safety. In *Privacy, Security, Risk*

and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom), 2012c.

Zhourong Chen, Xiaopeng Li, and Nevin L Zhang. Learning parsimonious deep feed-forward networks. 2018.

Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

FA Rezaur Rahman Chowdhury, Chao Ma, Md Rakibul Islam, Mohammad Hossein Namaki, Mohammad Omar Faruk, and Janardhan Rao Doppa. Select-and-evaluate: A learning framework for large-scale knowledge graph search. In *Asian Conference on Machine Learning*, pages 129–144, 2017.

Raviv Cohen and Derek Ruths. Classifying Political Orientation on Twitter: It's Not Easy! In *ICWSM*, 2013.

E Colleoni, A Rozza, and A Arvidsson. Echo chamber or public sphere? predicting political orientation and measuring political homophily in twitter using big data. *Journal of Communication*, 2014.

Larry L Constantine and Lucy AD Lockwood. *Software for use: a practical guide to the models and methods of usage-centered design*. Pearson Education, 1999.

Alan Cooper and Robert Reimann. About face 2.0. *The Essentials of Interaction Design*, 2003.

Alan Cooper et al. The inmates are running the asylum:[why high-tech products drive us crazy and how to restore the sanity](vol. 261). *Sams Indianapolis*, 1999.

Glen Coppersmith, Mark Dredze, Craig Harman, and Kristy Hollingshead Ihmc. From ADHD to SAD: Analyzing the Language of Mental Health on Twitter through Self-Reported Diagnoses. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 1–10, 2015.

Darcy J Corbitt-Hall, Jami M Gauthier, Margaret T Davis, and Tracy K Witte. College students' responses to suicidal content on social networking sites: an examination using a simulated facebook newsfeed. *Suicide and Life-Threatening Behavior*, 46(5):609–624, 2016.

Paul Covington, Jay Adams, and Emre Sargin. Deep Neural Networks for YouTube Recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*, 2016. doi: 10.1145/2959100.2959190.

Aron Culotta, Nirmal Kumar Ravi, and Jennifer Cutler. Predicting Twitter User Demographics using Distant Supervision from Website Traffic Data. *Journal of Artificial Intelligence Research*, 55:389–408, 2016.

Raminta Daniulaityte, Ramzi W Nahhas, Sanjaya Wijeratne, Robert G Carlson, Francois R Lamy, Silvia S Martins, Edward W Boyer, G Alan Smith, and Amit Sheth. "Time for dabs": Analyzing Twitter data on marijuana concentrates across the U.S. HHS Public Access. *Drug Alcohol Depend*, 155:307–311, 2015. doi: 10.1016/j.drugalcdep.2015.07.1199.

Raminta Daniulaityte, Lu Chen, Francois R Lamy, Robert G Carlson, Krishnaprasad Thirunarayan, and Amit Sheth. "When 'Bad' is 'Good'": Identifying Personal Comm unication and Sentiment in Drug-Related Tweets. *JMIR PUBLIC HEAL TH AND SURVEILLANCE*, 2016.

Dmitry Davidov, Oren Tsur, and Ari Rappoport. Enhanced Sentiment Learning Using Twitter Hashtags and Smileys. In *Proceedings of the 23rd international conference on computational linguistics. ACM*, pages 241–249, 2010.

Clayton A Davis, Giovanni Luca Ciampaglia, Luca Maria Aiello, Keychul Chung, Michael D Conover, Emilio Ferrara, Alessandro Flammini, Geoffrey C Fox, Xiaoming Gao, Bruno Gonçalves, Przemyslaw A Grabowicz, Kibeom Hong, Pik-Mai Hui, Scott Mccaulay, Karissa Mckelvey, Mark R Meiss, Snehal Patil, Chathuri Peli Kankanamalage, Valentin Pentchev, Judy Qiu, Jacob Ratkiewicz, Alex Rudnick, Benjamin Serrette, Prashant Shiralkar, Onur Varol, Lilian Weng, Tak-Lon Wu, Andrew J Younge, and Filippo Menczer. OSoMe: the IUNI observatory on social media. *PeerJ Computer Science*. doi: 10.7717/peerj-cs.87.

M De Choudhury, N Diakopoulos, and M Naaman. Unfolding the event landscape on twitter: classification and exploration of user categories. In *ACM CSCW*, 2012.

Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. Predicting Depression via Social Media. In *ICWSM*, 2013.

Munmun De Choudhury, Shagun Jhaver, Benjamin Sugar, and Ingmar Weber.

Social Media Participation in an Activist Movement for Racial Equality. In *ICSWM*, number Icwsm, pages 92–101, 2016. ISBN 9781577357582.

John P Dickerson, Vadim Kagan, and V S Subrahmanian. Using Sentiment to Detect Bots on Twitter: Are Humans more Opinionated than Bots? In *ASONAM*, 2014.

Tien Huu Do, Duc Minh Nguyen, Evaggelia Tsiligianni, Bruno Cornelis, and Nikos Deligiannis. Multiview Deep Learning for Predicting Twitter Users' Location. *arXiv preprint*, 2017.

Charles Dugas, Yoshua Bengio, François Bélisle, Claude Nadeau, and René Garcia. Incorporating functional knowledge in neural networks. *Journal of Machine Learning Research*, 10(Jun):1239–1262, 2009.

Susan T. Dumais. Latent semantic analysis. *Annual Review of Information Science and Technology*, 3(11):4356, 2008. ISSN 1941-6016. doi: 10.4249/scholarpedia.4356.

Monireh Ebrahimi, Amir Hossein Yazdavar, and Amit Sheth. On the Challenges of Sentiment Analysis for Dynamic Events. *IEEE Intelligent Systems*, 2017.

Juan Echeverria and Shi Zhou. Discovery, Retrieval, and Analysis of the 'Star Wars' Botnet in Twitter. In *ASONAM*, 2017.

Venkatesh Edupuganti. *Harassment Detection on Twitter using Conversations*. PhD thesis, 2017.

Andrea Esuli, Fabrizio Sebastiani, Consiglio Nazionale, and Delle Ricerche. Optimizing Text Quantifiers for Multivariate Loss Functions. *ACM Transactions on Knowledge Discovery from Data ACM Trans. Knowl. Discov. Data. VV*, 26, 2015. doi: 10.1145/0000000.0000000.

James Fan, Raymond Lau, and Risto Miikkulainen. Utilizing domain knowledge in neuroevolution. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pages 170–177, 2003.

A Fang, I Ounis, P Habel, C Macdonald, and N Limsopatham. Topic-centric classification of twitter user's political orientation. In *ACM SIGIR*, 2015.

Manaal Faruqui, Jesse Dodge, Sujay K Jauhar, Chris Dyer, Eduard Hovy, and Noah A Smith. Retrofitting word vectors to semantic lexicons. *arXiv preprint arXiv:1411.4166*, 2014.

Miriam Fernandez, Moizzah Asif, and Harith Alani. Understanding the roots of radicalisation on twitter. 2018.

Emilio Ferrara, Mohsen Jafariasbagh, Onur Varol, Vahed Qazvinian, Filippo Menczer, and Alessandro Flammini. Clustering Memes in Social Media. In *ASONAM*, 2013.

Emilio Ferrara, Wen-Qiang Wang, Onur Varol, Alessandro Flammini, and Aram Galstyan. Predicting online extremism, content adopters, and interaction reciprocity. In *International conference on social informatics*, pages 22–39. Springer, 2016.

Fabio Franch. (Wisdom of the Crowds) : 2010 UK Election Prediction with Social Media. *Journal of Information Technology & Politics*, 10(1):57–71, jan 2013. doi: 10.1080/19331681.2012.705080.

L Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40 (1):35–41, 1977.

Linton C Freeman. Centrality in Social Networks Conceptual Clarification. *Social Networks*, 179:215–239, 1978.

Wei Gao and Fabrizio Sebastiani. Tweet Sentiment: From Classification to Quantification. In *ASONAM*, 2015.

Jesse James Garrett. *Elements of user experience, the: user-centered design for the web and beyond*. Pearson Education, 2010.

Robert H Gass and John S Seiter. *Persuasion: Social influence and compliance gaining*. Routledge, 2015.

Manas Gaur, Ugur Kursuncu, Amanuel Alambo, Amit Sheth, Raminta Daniulaityte, Krishnaprasad Thirunarayan, and Jyotishman Pathak. " let me tell you about your mental health!" contextualized classification of reddit posts to dsm-5 for web-based intervention. 2018.

Petko Georgiev, Anastasios Noulas, and Cecilia Mascolo. Where Businesses Thrive: Predicting the Impact of the Olympic Games on Local Retailers through Location-based Services Data. In *ICWSM*, pages 151–160, 2014. ISBN 9781577356578.

Theodore Georgiou, Amr El Abbadi, Xifeng Yan, and Jemin George. Mining Complaints for Traffic-Jam Estimation: A Social Sensor Application. In *ASONAM*, 2015. doi: 10.1145/2808797.2809404.

Zafar Gilani, Ekaterina Kochmar, and Jon Crowcroft. Classification of Twitter Accounts into Automated Agents and Human Users. In *ASONAM*, 2017. doi: 10.1145/3110025.3110091.

Kevin Gimpel, Nathan Schneider, Brendan O 'connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A Smith. Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments. *Proceedings of ACL*, 2011.

Alec Go, Richa Bhayani, and Lei Huang. Twitter Sentiment Classification using Distant Supervision. Technical report, 2009.

F Godin, B Vandersmissen, W De Neve, and R Van de Walle. Multimedia lab @ acl wnut ner shared task: Named entity recognition for twitter microposts using distributed word representations. In *Proceedings of the Workshop on Noisy User-generated Text*, 2015.

J Goikoetxea, E Agirre, and A Soroa. Single or multiple? combining word representations independently learned from text and wordnet. In *AAAI*, 2016.

Yoav Goldberg and Omer Levy. word2vec explained: deriving mikolov et al.'s negative-sampling word-embedding method. *arXiv preprint arXiv:1402.3722*, 2014.

Faustino Gomez and Risto Miikkulainen. 2-d pole balancing with recurrent evolutionary networks. In *ICANN 98*, pages 425–430. Springer, 1998.

Vicenç Gómez, Andreas Kaltenbrunner, and Vicente López. Statistical analysis of the social network and discussion threads in slashdot. In *Proceedings of the 17th international conference on World Wide Web*, pages 645–654. ACM, 2008.

Klaus Greff, Rupesh K Srivastava, Jan Koutník, Bas R Steunebrink, and Jürgen Schmidhuber. Lstm: A search space odyssey. *IEEE transactions on neural networks and learning systems*, 28(10):2222–2232, 2017.

Thomas L Griffiths, Mark Steyvers, Joshua B Tenenbaum, and Tom Griffiths. Topics in semantic representation Topics in semantic representation. *Psychological review*, 2007.

Aditya Grover and Jure Leskovec. node2vec: Scalable Feature Learning for Networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 2016. doi: 10.1145/2939672.2939754.

Olivia Guest and Bradley C Love. What the success of brain imaging implies about the neural code. *Elife*, 6:e21397, 2017.

Xiaoxiao Guo. Deep learning and reward design for reinforcement learning. 2017.

Aditi Gupta and Ponnurangam Kumaraguru. Credibility Ranking of Tweets during High Impact Events. In *PSOSM*, 2012.

Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, and Anupam Joshi. Faking Sandy: Characterizing and Identifying Fake Images on Twitter during Hurricane Sandy. In *WWW*, 2013.

Aditi Gupta, Ponnurangam Kumaraguru, Carlos Castillo, and Patrick Meier. TweetCred: A Real-timeWeb-based System for Assessing Credibility of Content on Twitter. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8851 (November), 2014. ISSN 16113349. doi: 10.1007/978-3-319-13734-6.

Gulizar Haciyakupoglu and Weiyu Zhang. Social Media and Trust during the Gezi Protests in Turkey. *Journal of Computer-Mediated Communication*, 20(4): 450–466, 2015. ISSN 10836101. doi: 10.1111/jcc4.12121.

Mohammed Hafez and Creighton Mullins. The radicalization puzzle: a theoretical synthesis of empirical approaches to homegrown extremism. *Studies in Conflict & Terrorism*, 38(11):958–975, 2015.

Aimad Hakkoum and Said Raghay. Ontological approach for semantic modeling and querying the qur'an. In *Proceedings of the International Conference on Islamic Applications in Computer Science And Technology*, 2015.

Mordechai Haklay and Patrick Weber. Openstreetmap: User-generated street maps. *IEEE Pervasive Computing*, 2008.

Alon Halevy, Peter Norvig, and Fernando Pereira. The unreasonable effectiveness of data. *IEEE Intelligent Systems*, 24(2):8–12, 2009.

Takuo Hamaguchi, Hidekazu Oiwa, Masashi Shimbo, and Yuji Matsumoto. Knowledge transfer for out-of-knowledge-base entities: a graph neural network approach. *arXiv preprint arXiv:1706.05674*, 2017.

Fouzi Harrag, Eyas El-Qawasmeh, and Abdul Malik Salman Al-Salman. Extracting named entities from prophetic narration texts (hadith). In *International Conference on Software Engineering and Computer Systems*, pages 289–297. Springer, 2011.

Ammar Hassan, Ahmed Abbasi, and Daniel Zeng. Twitter Sentiment Analysis: A Bootstrap Ensemble Framework. In *SocialCom*, 2013. doi: 10.1109/SocialCom.2013.56.

Scott Helfstein. Edges of radicalization: Ideas, individuals and networks in violent extremism. Technical report, MILITARY ACADEMY WEST POINT NY COMBATING TERRORISM CENTER, 2012.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

Tuan-Anh Hoang, William W Cohen, Ee-Peng Lim, Doug Pierce, and David P Redlawsk. Politics, Sharing and Emotion in Microblogs. In *ASONAM*, 2013.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

Vincent Hoekstra. An overview of neuroevolution techniques. 2011.

Liangjie Hong and Brian D Davison. Empirical Study of Topic Modeling in Twitter. In *1st Workshop on Social Media Analytics (SOMA '10)*, 2010.

Liangjie Hong, Ovidiu Dan, and Brian Davison. Predicting Popular Messages in Twitter. In *WWW*, 2011.

John Horgan. Disengaging from terrorism. *The faces of terrorism: Multidisciplinary perspectives*, pages 257–276, 2009.

Philip N Howard, Muzammil Hussain, and Will Mari. Opening Closed Regimes What Was the Role of Social Media During the Arab Spring? 2011.

Yuheng Hu, Shelly Farnham, and Kartik Talamadupula. Predicting User Engagement on Twitter with Real-World Events. In *ICWSM*, 2015.

Zhiting Hu, Xuezhe Ma, Zhengzhong Liu, Eduard Hovy, and Eric Xing. Harnessing deep neural networks with logic rules. *arXiv preprint arXiv:1603.06318*, 2016.

Ghaffar Hussain and Erin Marie Saltman. *Jihad trending: A comprehensive analysis of online extremism and how to counter it*. Quilliam, 2014.

Rizwana Irfan, Christine K. King, Daniel Grages, Sam Ewen, Samee U. Khan, Sajjad A. Madani, Joanna Kolodziej, Lizhe Wang, Dan Chen, Ammar Rayes, Nikolaos Tziritas, Cheng-Zhong Xu, Albert Y. Zomaya, Ahmed Saeed Alzahrani, and Hongxiang L. A Survey on Text Mining in Social Networks. *The Knowledge Engineering Review*, 000:1–24, 2004. doi: 10.1017/S000000000000000.

Fuad Issa, Marco Damonte, Shay B Cohen, Xiaohui Yan, and Yi Chang. Abstract meaning representation for paraphrase detection. In *Proceedings of the 2018*

Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), volume 1, pages 442–452, 2018.

Ashutosh Jadhav. Knowledge driven search intent mining. 2016.

Carl Gustav Jung. *Two essays on analytical psychology*. Routledge, 2014.

Plinio Thomaz Aquino Junior and Lucia Vilela Leite Filgueiras. User modeling with personas. In *Proceedings of the 2005 Latin American conference on Human-computer interaction*, pages 277–282. ACM, 2005.

Lisa Kaati, Enghin Omer, Nico Prucha, and Amendra Shrestha. Detecting multipliers of jihadism on twitter. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pages 954–960. IEEE, 2015.

Nathan Kallus. Predicting Crowd Behavior with Big Public Data. *Proceedings of the 23rd International Conference on World Wide Web*, pages 625–630, 2014. doi: 10.1145/2567948.2579233.

Rajeshwari Kandakatla. *Identifying Offensive Videos on YouTube*. PhD thesis, 2016.

Pavan Kapanipathi, Fabrizio Orlandi, Amit Sheth, and Alexandre Passant. Personalized Filtering of the Twitter Stream. In *SPIM Workshop at ISWC 2011*, 2011.

Pavan Kapanipathi, Prateek Jain, Chitra Venkataramani, and Amit Sheth. User Interests Identification on Twitter Using a Hierarchical Knowledge Base. In *European Semantic Web Conference*, 2014a.

Pavan Kapanipathi, Prateek Jain, Chitra Venkataramani, and Amit Sheth. User interests identification on twitter using a hierarchical knowledge base. In *European Semantic Web Conference*, pages 99–113. Springer, 2014b.

Levent Karacan, Aykut Erdem, and Erkut Erdem. Image matting with kl-divergence based sparse sampling. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 424–432, 2015.

Ryan Kelly and Leon Watts. Characterising the Inventive Appropriation of Emoji as Relationally Meaningful in Mediated Close Personal Relationships. *Experiences of Technology Appropriation: Unanticipated Users, Usage, Circumstances, and Design*, 2015.

Jason S Kessler, Miriam Eckert, Lyndsie Clark, and Nicolas JD Nicolov Power. The ICWSM 2010 JDPA Sentiment Corpus for the Automotive Domain. In *4th International AAAI Conference on Weblogs and Social Media Data Workshop Challenge (ICWSM-DWC)*, 2010.

Soon Jye Kho, Swati Padhee, Goonmeet Bajaj, Krishnaprasad Thirunarayan, and Amit Sheth. Domain-specific use cases for knowledge-enabled social media analysis. In *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining*, pages 233–246. Springer, 2019.

A Kim, T Miano, R Chew, M Eggers, and J Nonnemaker. Classification of twitter users who tweet about e-cigarettes. *JMIR*, 2017.

James P Klein, Gary Goertz, and Paul F Diehl. The new rivalry dataset: Procedures and patterns. *Journal of Peace Research*, 43(3):331–348, 2006.

Ryota Kobayashi and Renaud Lambiotte. TiDeH: Time-Dependent Hawkes Process for Predicting Retweet Dynamics. Number ICWSM, pages 191–200, 2016. ISBN 9781577357582.

Rostyslav Korolov, Di Lu, Jingjing Wang, Guangyu Zhou, Claire Bonial, Clare Voss, Lance Kaplan, William Wallace, Jiawei Han, and Heng Ji. On Predicting Social Unrest Using Social Media. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2016.

Mandy Korpusik, Shigeyuki Sakaki, Francine Chen, and Yan-Ying Chen. Recurrent neural networks for customer purchase prediction on twitter. In *CBRecSys@ RecSys*, pages 47–50, 2016.

Alok Kothari, Walid Magdy, Kareem Darwish, Ahmed Mourad, and Ahmed Taei. Detecting Comments on News Articles in Microblogs. In *ICWSM*, 2013.

E Kouloumpis, T Wilson, and J Moore. Twitter sentiment analysis: The good the bad and the omg! *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM 11)*, pages 538–541, 2011. ISSN 03781097.

B Krawczyk. Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, 2016.

Revathy Krishnamurthy, Pavan Kapanipathi, Amit P Sheth, Krishnaprasad Thirunarayan, and Amit Sheth. Location Prediction of Twitter Users using Wikipedia. 2014.

Adit Krishnan, Ashish Sharma, and Hari Sundaram. Improving latent user models in online social media. *arXiv preprint arXiv:1711.11124*, 2017.

Ugur Kursuncu, Manas Gaur, Usha Lokala, Anurag Illendula, Krishnaprasad Thirunarayan, Raminta Daniulaityte, Amit Sheth, and I Budak Arpinar. " what's ur type?" contextualized classification of user types in marijuana-related communications using compositional multiview embedding. In *IEEE/WIC/ACM International Conference on Web Intelligence(WI'18)*, 2018.

Ugur Kursuncu, Manas Gaur, Usha Lokala, Krishnaprasad Thirunarayan, Amit Sheth, and I Budak Arpinar. Predictive analysis on twitter: Techniques and applications. In *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining*, pages 67–104. Springer, 2019.

Francois R Lamy, Raminta Daniulaityte, Amit Sheth, Ramzi W Nahhas, Silvia S Martins, Edward W Boyer, and Robert G Carlson Francois R Lamy. "Those edibles hit hard": Exploration of Twitter data on cannabis edibles in the U.S HHS Public Access. *Drug Alcohol Depend*, 1(164):64–70, 2016. doi: 10.1016/j.drugalcdep.2016.04.029.

Francois R Lamy, Raminta Daniulaityte, Ramzi W Nahhas, Monica J Barratt, Alan G Smith, Amit Sheth, Silvia S Martins, Edward W Boyer, and Robert G

Carlson. Increases in synthetic cannabinoids-related harms: Results from a longitudinal web-based content analysis. *International Journal of Drug Policy*, 2017. doi: 10.1016/j.drugpo.2017.05.007.

Jey Han Lau and Timothy Baldwin. An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation. *arXiv preprint*, 2016.

Glenn Lawyer. Understanding the influence of all nodes in a network. *Nature Scientific Reports*, 2015. doi: 10.1038/srep08665.

Michael D Lee and Megan N Lee. The relationship between crowd majority and accuracy for binary decisions. *Judgment and Decision Making*, 12(4):328–343, 2017.

Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, Sören Auer, and Christian Bizer. Dbpedia – a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web*, 1:1–5, 2012.

Yoad Lewenberg, Yoram Bachrach, and Svitlana Volkova. Using emotions to predict user interest areas in online social networks. In *Data Science and Advanced Analytics (DSAA)*, 2015. ISBN 9781467382731. doi: 10.1109/DSAA.2015.7344887.

Li Li, Maosong Sun, and Zhiyuan Liu. Discriminating Gender on Chinese Microblog: A Study of Online Behaviour, Writing Style and Preferred Vocabulary. In *10th International Conference on Natural Computation (ICNC)*, 2014.

Wen Li and Markus Dickinson. Gender Prediction for Chinese Social Media Data. In *Proceedings of Recent Advances in Natural Language Processing*, pages 438–445, 2017. doi: 10.26615/978-954-452-049-6_058.

L Liao, X He, H Zhang, and T Chua. Attributed social network embedding. *arXiv preprint arXiv:1705.04969*, 2017.

J Lilleberg, Y Zhu, and Y Zhang. Support vector machines and word2vec for text classification with semantic features. In *IEEE ICCI* CC*, 2015.

Kun-Lin Liu, Wu-Jun Li, and Minyi Guo. Emoticon Smoothed Language Models for Twitter Sentiment Analysis. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.

Mingxia Liu, Daoqiang Zhang, and Songcan Chen. Attribute relation learning for zero-shot classification. *Neurocomputing*, 139:34–46, 2014.

Chris Longworth. *Kernel methods for text-independent speaker verification*. PhD thesis, University of Cambridge, 2010.

Shah Mahmood. Online social networks: The overt and covert communication channels for terrorists and beyond. In *Homeland Security (HST), 2012 IEEE Conference on Technologies for*, pages 574–579. IEEE, 2012.

Jalal Mahmud, Jeffrey Nichols, and Clemens Drews. Where Is This Tweet From? Inferring Home Locations of Twitter Users. In *ICWSM*, 2012.

Jalal Mahmud, Geli Fei, Anbang Xu, Aditya Pal, and Michelle Zhou. Predicting Attitude and Actions of Twitter Users. In *Proceedings of the 21st International*

Conference on Intelligent User Interfaces - IUI '16, pages 1–6. ACM Press, 2016. ISBN 9781450341370. doi: 10.1145/2856767.2856800.

Aibek Makazhanov and Davood Rafiei. Predicting Political Preference of Twitter Users. *Social Network Analysis and Mining*, 2014.

Bassem Makni and James Hendler. Deep learning for noise-tolerant rdfs reasoning.

Josephine Malilay, Michael Heumann, Dennis Perrotta, Amy F Wolkin, Amy H Schnall, Michelle N Podgornik, Miguel A Cruz, Jennifer A Horney, David Zane, Rachel Roisman, Joel R Greenspan, Doug Thoroughman, Henry A Anderson, Eden V Wells, and Erin F Simms. The Role of Applied Epidemiology Methods in the Disaster Management Cycle. *American Journal of Public Health*, 104(10): 2092–2102, 2014.

Ashutosh K Maurya. Learning low dimensional word based linear classifiers using data shared adaptive bootstrap aggregated lasso with application to imdb data. *arXiv preprint arXiv:1807.10623*, 2018.

Bridget T McInnes, Ted Pedersen, and Serguei VS Pakhomov. Umls-interface and umls-similarity: open source software for measuring paths and semantic similarity. In *AMIA Annual Symposium Proceedings*, volume 2009, page 431. American Medical Informatics Association, 2009.

T Mikolov, I Sutskever, K Chen, GS Corrado, and J Dean. Distributed representations of words and phrases and their compositionality. In *NIPS*, 2013a.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed Representations of Words and Phrases and their Compositionality. *Advances in neural information processing systems*, 2013b.

GR Miller. On being persuaded: Some basic distinctions. 1980.

Hannah Miller, Jacob Thebault-Spieker, Shuo Chang, Isaac Johnson, Loren Terveen, and Brent Hecht. "Blissfully happy" or "ready to fight": Varying Interpretations of Emoji. *International AAAI Conference on Web and Social Media*, (ICWSM):259–268, 2016. ISSN 2152-2715. doi: 10.1089/cyber.2011.0179.

J Mitchell and M Lapata. Composition in distributional models of semantics. *Cognitive science*, 2010.

Tanushree Mitra and Eric Gilbert. CREDBANK: A Large-Scale Social Media Corpus with Associated Credibility Annotations. In *ICWSM*, 2016.

Fred Morstatter, Jürgen Pfeffer, Huan Liu, and Kathleen M. Carley. Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose. In *ICWSM*, pages 400–408, 2013. ISBN 9783319055787. doi: 10.1007/978-3-319-05579-4_10.

Arman Khadjeh Nassirtoussi, Saeed Aghabozorgi, Teh Ying Wah, David Chek, and Ling Ngo. Text mining for market prediction: A systematic review. *EXPERT SYSTEMS WITH APPLICATIONS*, 41:7653–7670, 2014. doi: 10.1016/j.eswa.2014.06.009.

Nasir Naveed, Thomas Gottron, Jerome Kunegis, and Arifah Che Alhadi. Bad News Travel Fast : A Content-based Analysis of Interestingness on Twitter. In *Proceedings of the 3rd International Web Science Conference. ACM*, 2011.

Dong Nguyen, Noah A Smith, and Carolyn P Rosé. Author Age Prediction from Text using Linear Regression. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities. Association for Computational Linguistics*, 2011.

Dong Nguyen, Rilana Gravel, Dolf Trieschnigg, and Theo Meder. "How Old Do You Think I Am?": A Study of Language and Age in Twitter. In *ICWSM*, 2013.

Le T. Nguyen, Pang Wu, William Chan, Wei Peng, and Ying Zhang. Predicting collective sentiment dynamics from time-series social media. In *Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM)*, pages 6:1–6:8, 2012. ISBN 9781450315432. doi: 10.1145/2346676.2346682.

Ming Ni, Qing He, and Jing Gao. Using Social Media to Predict Traffic Flow under Special Event Conditions. In *The 93rd Annual Meeting of Transportation Research Board*, 2014.

Maximilian Nickel, Lorenzo Rosasco, Tomaso A Poggio, et al. Holographic embeddings of knowledge graphs. In *AAAI*, volume 2, pages 3–2, 2016.

Jakob Nielsen. *Usability engineering*. Elsevier, 1994.

Petra Kralj Novak, Jasmina Smailović, Borut Sluban, and Igor Mozetič. Sentiment of Emojis. *PLOS One*, 2015. doi: 10.1371/journal.pone.0144296.

Lucila Ohno-Machado, Susanna-Assunta Sansone, George Alter, Ian Fore, Jeffrey Grethe, Hua Xu, Alejandra Gonzalez-Beltran, Philippe Rocca-Serra, Anupama E Gururaj, Elizabeth Bell, et al. Finding useful data across multiple biomedical data repositories using datamed. *Nature genetics*, 49(6):816, 2017.

Enghin Omer. Using machine learning to identify jihadist messages on twitter, 2015.

Alexander Pak and Patrick Paroubek. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *LREc*, 10, 2010.

Mark Palatucci, Dean Pomerleau, Geoffrey E Hinton, and Tom M Mitchell. Zero-shot learning with semantic output codes. In *Advances in neural information processing systems*, pages 1410–1418, 2009.

Nikhil Pattisapu, Manish Gupta, Ponnurangam Kumaraguru, and Vasudeva Varma. Medical Persona Classification in Social Media. In *ASONAM*, 2017.

Heiko Paulheim. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic web*, 8(3):489–508, 2017.

Marco Pennacchiotti and Ana-Maria Popescu. Democrats, Republicans and Starbucks Afficionados: User Classification in Twitter. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2011a.

Marco Pennacchiotti and Ana-Maria Popescu. Democrats, republicans and starbucks afficionados: user classification in twitter. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2011b.

Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014.

K. Bradley. Penuel and Matthew. Statler. *Encyclopedia of disaster relief*. Sage Publications, 2011. ISBN 9781412971010.

Sujan Perera, Pablo N Mendes, Adarsh Alex, Amit P Sheth, and Krishnaprasad Thirunarayan. Implicit entity linking in tweets. In *International Semantic Web Conference*, pages 118–132. Springer, 2016.

Lawrence Phillips, Chase Dowling, Kyle Shaffer, Nathan Hodas, and Svitlana Volkova. Using Social Media To Predict the Future: A Systematic Literature Review. In *arXiv preprint*, 2017.

Louise E Porter and Mark R Kebbell. Radicalization in australia: Examining australia's convicted terrorists. *Psychiatry, Psychology and Law*, 18(2):212–231, 2011.

Víctor M Prieto, Sé Rgio Matos, Manuel Lvarez, Fidel Cacheda, José Luís Oliveira, and Juan A Añ. Twitter: A Good Place to Detect Health Conditions. *PLoS ONE*, 9(1), 2014. doi: 10.1371/.

H Purohit, Y Ruan, A Joshi, S Parthasarathy, and A Sheth. Understanding user-community engagement by multi-faceted features: A case study on twitter. In *WWW Workshop SoME*, 2011.

H Purohit, G Dong, V Shalin, TK Prasad, and A Sheth. Intent classification of short-text on social media. In *Smart City/SocialCom/SustainCom (SmartCity), 2015 IEEE International Conference on*. IEEE, 2015.

Hemant Purohit and Amit Sheth. Twitris v3: From Citizen Sensing to Analysis, Coordination and Action. In *ICWSM*, 2013a.

Hemant Purohit and Amit P Sheth. Twitris v3: From citizen sensing to analysis, coordination and action. In *ICWSM*, 2013b.

Hemant Purohit, Andrew Hampton, Valerie L Shalin, Amit P Sheth, John Flach, and Shreyansh Bhatt. What kind of #conversation is Twitter? Mining #psycholinguistic cues for emergency coordination. *Computers in Human Behavior*, 29:2438–2447, 2013. doi: 10.1016/j.chb.2013.05.007.

Hemant Purohit, Shreyansh Bhatt, Andrew Hampton, Valerie L Shalin, and Amit P Sheth. With Whom to Coordinate, Why and How in Ad- Hoc Social Media Communications during Crisis Response. In *Proceedings of the 11 th International ISCRAM Conference*, pages 787–791, 2014a.

Hemant Purohit, Andrew Hampton, Shreyansh Bhatt, Valerie L Shalin, Amit P Sheth, and John M Flach. Identifying Seekers and Suppliers in Social Media

Communities to Support Crisis Coordination. In *Computer Supported Cooperative Work (CSCW)*, 2014b. doi: 10.1007/s10606-014-9209-y.

Bhavtosh Rath, Wei Gao, Jing Ma, and Jaideep Srivastava. From Retweet to Believability: Utilizing Trust to Identify Rumor Spreaders on Twitter. In *ASONAM*, 2017. doi: 10.1145/3110025.3110087.

P. Refaeilzadeh, L. Tang, and H. Liu. Cross-Validation. *Encyclopedia of Database Systems. Springer*, 2009.

G Rizos, S Papadopoulos, and Y Kompatsiaris. Learning to classify users in online interaction networks.

Julie M Robillard, Thomas W Johnson, Craig Hennessey, B Lynn Beattie, and Judy Illes. Aging 2.0: Health Information about Dementia on Twitter. *Plos One*, 20(87), 2013. doi: 10.1371/journal.pone.0069861.

Daniel M. Romero, Brendan Meeder, and Jon Kleinberg. Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter. In *Proceedings of the 20th international conference on World wide web.*, 2011.

Jacob Ross and Krishnaprasad Thirunarayan. Features for Ranking Tweets Based on Credibility and Newsworthiness. In *International Conference on Collaboration Technologies and Systems*, 2016. doi: 10.1109/CTS.2016.21.

Dominic Rout, Daniel Preoiuc-Pietro, Kalina Bontcheva, and Trevor Cohn. Where's @wally? A Classification Approach to Geolocating Users Based on

their Social Ties. In *24th ACM Conference on Hypertext and Social Media*, Paris, France, 2013.

Matthew Rowe and Hassan Saif. Mining pro-isis radicalisation signals from social media users. In *Proceedings of the tenth international AAAI conference on web and social media (ICWSM 2016)*, pages 329–338, 2016.

Yiye Ruan, Hemant Purohit, David Fuhry, Srinivasan Parthasarathy, Amit P Sheth, and Amit Sheth. Prediction of Topic Volume on Twitter. In *4th International ACM Conference on Web Science*, pages 397–402, 2012. ISBN 978-1-4503-1228-8.

Alexander M Rush, Sumit Chopra, and Jason Weston. A neural attention model for abstractive sentence summarization. *arXiv preprint arXiv:1509.00685*, 2015.

Hassan Saif. *Semantic Sentiment Analysis in Social Streams*. 2017.

Hassan Saif, Thomas Dickinson, Leon Kastler, Miriam Fernandez, and Harith Alani. A semantic graph-based approach for radicalisation detection on social media. In *European semantic web conference*, pages 571–587. Springer, 2017.

Shigeyuki Sakaki, Yasuhide Miura, Xiaojun Ma, Keigo Hattori, and Tomoko Ohkuma. Twitter User Gender Inference Using Combined Analysis of Text and Image Processing. In *Proceedings of the 25th International Conference on Computational Linguistics*, pages 54–61, 2014.

Md Kamruzzaman Sarker, Ning Xie, Derek Doran, Michael Raymer, and Pascal

Hitzler. Explaining trained neural networks with semantic web technologies: First steps. *arXiv preprint arXiv:1710.04324*, 2017.

Ankita Saxena. *A Semantically Enhanced Approach to Identify Depression-Indicative Symptoms Using Twitter Data*. PhD thesis, Wright State University, 2018.

Jacob R Scanlon and Matthew S Gerber. Automatic detection of cyber-recruitment by violent extremists. *Security Informatics*, 3(1):5, 2014.

Jacob R Scanlon and Matthew S Gerber. Forecasting violent extremist cyber recruitment. *IEEE Transactions on Information Forensics and Security*, 10(11): 2461–2470, 2015.

Thijs Scheepers, Evangelos Kanoulas, and Efstratios Gavves. Improving word embedding compositionality using lexicographic definitions. In *WWW*, 2018.

Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.

Jürgen Schmidhuber, Matteo Gagliolo, Daan Wierstra, and Faustino Gomez. Evolino for recurrent support vector machines. *arXiv preprint cs/0512062*, 2005.

Jürgen Schmidhuber, Daan Wierstra, Matteo Gagliolo, and Faustino Gomez. Training recurrent networks by evolino. *Neural computation*, 19(3):757–779, 2007.

Jacob Schrum and Risto Miikkulainen. Constructing complex npc behavior via multi-objective neuroevolution. *AIIDE*, 8:108–113, 2008.

Lutfi Kerem Senel, Ihsan Utlu, Veysel Yucesoy, Aykut Koc, and Tolga Cukur. Semantic structure and interpretability of word embeddings. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2018.

Ying Shen, Yang Deng, Min Yang, Yaliang Li, Nan Du, Wei Fan, and Kai Lei. Knowledge-aware attentive neural network for ranking question answer pairs. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 901–904. ACM, 2018.

A Sheth and P Kapanipathi. Semantic filtering for social data. *IEEE Internet Computing*, 2016a.

Amit Sheth and Pavan Kapanipathi. Semantic Filtering for Social Data. *IEEE Internet Computing*, 2016b.

Amit Sheth, Sujan Perera, Sanjaya Wijeratne, and Krishnaprasad Thirunarayan. Knowledge will propel machine understanding of content: Extrapolating from current examples. *arXiv preprint arXiv:1707.05308*, 2017.

Amit Sheth, Hemant Purohit, Gary Alan Smith, Jeremy Brunn, Ashutosh Jadhav, Pavan Kapanipathi, Chen Lu, and Wenbo Wang. Twitris: A System for Collective Social Intelligence. *Encyclopedia of Social Network Analysis and Mining*, 2018. doi: 10.1007/978-1-4614-6170-8_345.

Clay Shirky. The political power of social media: Technology, the public sphere, and political change. *Foreign affairs*, pages 28–41, 2011.

Prashanth Gurunath Shivakumar, Haoqi Li, Kevin Knight, and Panayiotis Georgiou. Learning from past mistakes: Improving automatic speech recognition output via noisy-clean phrase context modeling. *arXiv preprint arXiv:1802.02607*, 2018.

Ben Shneiderman. *Designing the user interface: strategies for effective human-computer interaction*. Pearson Education India, 2010.

Alan Smith and Manas Gaur. What's my age?: Predicting Twitter User's Age using Influential Friend Network and DBpedia. *arxiv preprint*, 2018.

Marina Sokolova, Kanyi Huang, Stan Matwin, Joshua Ramisch, Vera Sazonova, Renee Black, Chris Orwa, Sidney Ochieng, and Nanjira Sambuli. Topic Modelling and Event Identification from Twitter Textual Data. *ArXiv preprint*, 2016.

Kenneth O Stanley and Risto Miikkulainen. Efficient reinforcement learning through evolving neural network topologies. In *Proceedings of the 4th Annual Conference on Genetic and Evolutionary Computation*, pages 569–577. Morgan Kaufmann Publishers Inc., 2002.

James B Stiff and Paul A Mongeau. *Persuasive communication*. Guilford Publications, 2016.

Dario Stojanovski, Gjorgji Strezoski, Gjorgji Madjarov, and Ivica Dimitrovski. Finki at SemEval-2016 Task 4: Deep Learning Architecture for Twitter Sentiment Analysis. In *Proceedings of SemEval*, pages 149–154, 2016.

Bongwon Suh, Lichan Hong, Peter Pirolli, and Ed H Chi. Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network. In *IEEE international conference on Social Computing Social computing (SocialCom)*, 2010.

Chen Sun, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta. Revisiting unreasonable effectiveness of data in deep learning era. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pages 843–852. IEEE, 2017.

Ashish Sureka and Swati Agarwal. Learning to classify hate and extremism promoting tweets. In *Intelligence and Security Informatics Conference (JISIC), 2014 IEEE Joint*, pages 320–320. IEEE, 2014.

Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. On the importance of initialization and momentum in deep learning. In *International conference on machine learning*, pages 1139–1147, 2013.

Yee W Teh, Michael I Jordan, Matthew J Beal, and David M Blei. Sharing clusters among related groups: Hierarchical dirichlet processes. In *Advances in neural information processing systems*, pages 1385–1392, 2005.

Kurt Thomas, Chris Grier, and Vern Paxson. Suspended Accounts in Retrospect: An Analysis of Twitter Spam. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement*, 2011.

T Tieleman and G Hinton. Divide the gradient by a running average of its recent magnitude. coursera: Neural networks for machine learning. Technical report.

R Tinati, L Carr, W Hall, and J Bentwood. Identifying communicator roles in twitter. In *WWW*, 2012.

A Tsitsulin, D Mottin, P Karras, and E Müller. Verse: Versatile graph embeddings from similarity measures. In *WWW*, 2018.

O Tsur and A Rappoport. Don't Let Me Be #Misunderstood: Linguistically Motivated Algorithm for Predicting the Popularity of Textual Memes. In *ICWSM, Ninth International AAAI Conference on Web and Social Media*, pages 426–435, 2015. ISBN 9781577357339.

Zeynep Tufekci. Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. In *ICWSM*, 2014.

Onur Varol, Emilio Ferrara, Clayton A Davis, Filippo Menczer, and Alessandro Flammini. Online Human-Bot Interactions: Detection, Estimation, and Characterization. In *ICWSM*, 2017a.

Onur Varol, Emilio Ferrara, Filippo Menczer, and Alessandro Flammini. Early detection of promoted campaigns on social media. *EPJ Data Science*, 6(1), 2017b. ISSN 21931127. doi: 10.1140/epjds/s13688-017-0111-y.

Lorenzo Vidino. Radicalization, linkage, and diversity: Current trends in terrorism in europe. Technical report, RAND NATIONAL DEFENSE RESEARCH INST SANTA MONICA CA, 2011.

Lorenzo Vidino and Seamus Hughes. *ISIS in America: From retweets to Raqqa*. Program on Extremism, The George Washington University, 2015.

Sarah Vieweg, Amanda L Hughes, Kate Starbird, and Leysia Palen. Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness. In *CHI - Crisis Informatics*, 2010.

Khuong Vo, Dang Pham, Mao Nguyen, Trung Mai, and Tho Quan. Combination of domain knowledge and deep learning for sentiment analysis. In *International Workshop on Multi-disciplinary Trends in Artificial Intelligence*, pages 162–173. Springer, 2017.

Svitlana Volkova and Eric Bell. Identifying Effective Signals to Predict Deleted and Suspended Accounts on Twitter across Languages. In *ICWSM, Association for the Advancement of Artificial Intelligence*, number Icwsm, pages 290–298, 2017. ISBN 9781577357889.

Pooja Wadhwa and MPS Bhatia. Tracking on-line radicalization using investigative data mining. In *Communications (NCC), 2013 National Conference on*, pages 1–5. IEEE, 2013.

Claudia Wagner, Sitaram Asur, and Joshua Hailpern. Religious Politicians and Creative Photographers: Automatic User Categorization in Twitter. In *SocialCom*, 2013. doi: 10.1109/SocialCom.2013.49.

Hao Wang, Dogan Can, Abe Kazemzadeh, François Bar, and Shrikanth Narayanan. A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 115–120, 2012a.

P Wang, B Xu, J Xu, G Tian, CL Liu, and H Hao. Semantic expansion using word embedding clustering & convolutional neural network for improving short text classification. *Neurocomputing*, 2016.

Ruijie Wang, Yuchen Yan, Jialu Wang, Yuting Jia, Ye Zhang, Weinan Zhang, and Xinbing Wang. Acekg: A large-scale knowledge graph for academic data mining. *arXiv preprint arXiv:1807.08484*, 2018a.

Wenbo Wang, Lu Chen, Krishnaprasad Thirunarayan, and Amit P Sheth. Harnessing Twitter 'Big Data' for Automatic Emotion Identification. In *IEEE International Confernece on Social Computing (SocialCom)*, 2012b.

Wenbo Wang, Lu Chen, Krishnaprasad Thirunarayan, and Amit P Sheth. Cursing in english on twitter. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 415–425. ACM, 2014a.

Xiaofeng Wang, Matthew S Gerber, and Donald E Brown. Automatic Crime Prediction using Events Extracted from Twitter Posts. In *International conference on social computing, behavioral-cultural modeling, and prediction. Springer*, 2012c.

Xiaolong Wang, Furu Wei, Xiaohua Liu, Ming Zhou, and Ming Zhang. Topic Sentiment Analysis in Twitter: A Graph-based Hashtag Sentiment Classification Approach. In *Proceedings of the 20th ACM international conference on Information and knowledge management. ACM*, 2011.

Yiqi Wang, Zhan Shi, Xifeng Guo, Xinwang Liu, En Zhu, and Jianping Yin. Deep embedding for determining the number of clusters. In *AAAI*, 2018b.

Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph embedding by translating on hyperplanes. In *AAAI*, volume 14, pages 1112–1119, 2014b.

Yang Wen, An Xu, Wei Liu, and Leiting Chen. A wide residual network for sentiment classification. In *Proceedings of the 2018 2nd International Conference on Deep Learning Technologies*, pages 7–11. ACM, 2018.

Lilian Weng, Filippo Menczer, and Yong-Yeol Ahn. Predicting Successful Memes using Network and Community Structure. In *IC*, pages 535–544, 2014. ISBN 9781577356578.

S Wijeratne, L Balasuriya, A Sheth, and D Doran. Emojinet: Building a machine readable sense inventory for emoji. In *SocInfo*, 2016a.

S Wijeratne, L Balasuriya, A Sheth, and D Doran. Emojinet: An open service and api for emoji sense discovery. *arXiv preprint arXiv:1707.04652*, 2017a.

S Wijeratne, L Balasuriya, A Sheth, and D Doran. A semantics-based measure of emoji similarity. *arXiv preprint arXiv:1707.04653*, 2017b.

Sanjaya Wijeratne, Derek Doran, Amit Sheth, and Jack L Dustin. Analyzing the Social Media Footprint of Street Gangs. In *Intelligence and Security Informatics (ISI)*, 2015.

Sanjaya Wijeratne, Lakshika Balasuriya, Derek Doran, Amit Sheth, and Amit@knoesis Org. Word Embeddings to Enhance Twitter Gang Member Profile Identification. In *IJCAI Workshop on Semantic Machine Learning*, 2016b.

Sanjaya Wijeratne, Lakshika Balasuriya, Amit Sheth, and Derek Doran. EmojiNet: An Open Service and API for Emoji Sense Discovery. In *ICWSM*, 2017c.

Sanjaya Wijeratne, Amit Sheth, Shreyansh Bhatt, Lakshika Balasuriya, Hussein S. Al-Olimat, Manas Gaur, Amir Hossein Yazdavar, and Krishnaprasad Thirunarayan. Feature Engineering for Twitter-based Applications. In *Feature Engineering for Machine Learning and Data Analytics*, page 35. 2017d.

Jun-Ming Xu, Kwang-Sung Jun, Xiaojin Zhu, and Amy Bellmore. Learning from Bullying Traces in Social Media. In *2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 656–666, 2012.

Shuhei Yamamoto and Tetsuji Satoh. Hierarchical Estimation Framework of Multi-Label Classifying: A Case of Tweets Classifying into Real Life Aspects. In *ICWSM*, 2015.

Xiao Yang, Richard Mccreadie, Craig Macdonald, and Iadh Ounis. Transfer Learning for Multi-language Twitter Election Classification. In *ASONAM*, 2017. doi: 10.1145/3110025.3110059.

Amir Hossein Yazdavar and S Hussein. Al-olimat, monireh ebrahimi, goonmeet bajaj, tanvi banerjee, krishnaprasad thirunarayan, jyotishman pathak, and amit

sheth. semi-supervised approach to monitoring clinical depressive symptoms in social media. in 2017 ieee. In *ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Sydney, Australia*, 2017.

Amir Hossein Yazdavar, Hussein S Al-Olimat, Monireh Ebrahimi, Goonmeet Bajaj, Tanvi Banerjee, Krishnaprasad Thirunarayan, Jyotishman Pathak, and Amit Sheth. Semi-Supervised Approach to Monitoring Clinical Depressive Symptoms in Social Media. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2017.

Kai Yi, Zhiqiang Jian, Shitao Chen, Yu Chen, and Nanning Zheng. Knowledge-based recurrent attentive neural network for traffic sign detection. *arXiv preprint arXiv:1803.05263*, 2018.

Zhijun Yin, You Chen, Daniel Fabbri, Jimeng Sun, and Bradley Malin. #PrayForDad: Learning the Semantics Behind Why Social Media Users Disclose Health Information. In *ICWSM*, 2016.

Chengxiang Zhai, John Lafferty, J Lafferty, and C Zhai. A Study of Smoothing Methods for Language Models Applied to Information Retrieval. *ACM Transactions on Information Systems*, 22(2):179–214, 2004.

D Zhang, S Li, H Wang, and G Zhou. User classification with multiple textual perspectives. In *COLING*, 2016a.

J Zhang, X Hu, Y Zhang, and H Liu. Your age is no secret: Inferring microbloggers'

ages via content and interaction analysis. *Proceedings of the 10th International Conference on Web and Social Media, ICWSM 2016*, (Icwsm):476–485, 2016b.

Wayne Xin Zhao, Jing Jiang, Jing He, Yang Song, Palakorn Achananuparp, Ee-Peng Lim, and Xiaoming Li. Topical Keyphrase Extraction from Twitter. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pages 379–388, 2011.

Chunting Zhou, Chonglin Sun, Zhiyuan Liu, and Francis Lau. A c-lstm neural network for text classification. *arXiv preprint arXiv:1511.08630*, 2015.