



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

PhD THESIS

TEMPORAL EMOTION DYNAMICS IN
SOCIAL NETWORKS



DEBASHIS NASKAR

Departamento de Sistemas Informáticos y Computación

Supervisors : Eva Onaindía de la Rivaherrera
Miguel Rebollo Pedruelo

18th January 2022

Temporal Emotion Dynamics in Social Networks



DEBASHIS NASKAR

Departamento de Sistemas Informáticos y Computación

Universitat Politècnica de València

A thesis submitted for the degree of
Doctor of Philosophy in the subject of Computer Science
under the supervision of:
Dr. Eva Onaindía de la Rivaherrera
Dr. Miguel Rebollo Pedruelo

18th January 2022

PhD Thesis

Title: Temporal Emotion Dynamics in Social Networks

Author: Debashis Naskar

Advisors: Dr. Eva Onaindia de la Rivaherrera
Dr. Miguel Rebollo Pedruelo

Reviewers:

Dr. Biswanath Dutta
Dr. Ashish Anand
Dr. Juan Manuel Corchado Rodríguez

Day of the defense: 18th January 2022

A la mí mamá

Acknowledgements

I would like to express my sincere gratitude to my PhD supervisor, *Prof. Eva Onaindia* and my co-supervisor *Dr. Miguel Rebollo*. This thesis would not have been complete without their patient guidance and enthusiastic encouragement on my research work. Their continuous support and useful-positive criticism have influenced my attitude, nature and personality in many ways. I am genuinely grateful to them for putting me back on track during the hard times of my PhD, and it is my honour to be their student.

I would also like to express my deep gratitude to *Prof. Sukumar Nandi* and *Dr. Sanasam Ranbir Singh* for giving me the privilege to work with them. I have received helpful advice and assistance through their inspiring discussions, constructive suggestions and kind help on my work. This thesis would not have been possible without joining the research internship program at the *Indian Institute of Technology Guwahati* and having a nice research environment. I am forever grateful for their support and valuable times.

I am very grateful to the *NAMASTE Consortium* - coordinated by *Georg-August-Universität Göttingen*, has awarded me *Erasmus Mundus Action 2* scholarship for pursuing my PhD research at *Universitat Politècnica de València*. It was fantastic to have this opportunity with

lots of facilities and a chance to interact with eminent researchers from different countries.

I am forever thankful to *Dr. Sidahmed Mokaddem* for supporting at the beginning of my research and *Mr. Diego Aineto García* for helping me in writing code. I would like to thank every member of the DSIC lab for providing me with a healthy work environment and helping me out during my needs. I am also thankful to *Mr. Durgesh Kumar* and *Ms. Mala Das* for helping and supporting my research work in many ways. I had the privilege to collaborate with *Dr. Subhashis Das* whose interest and knowledge has enriched me a lot. I am forever thankful to *Dr. Subhrendu Chattopadhyay* for the countless coffee sessions and the round of discussions on various topics of interests. I express my sincere gratitude towards OSiNT Group members for their motivation, inspirational energy, and strong support during my research stay.

Last but not the least, I dedicate this thesis to my parents for all their everlasting and unconditional love. From the childhood they have raised me with all their love and care and gave me a lot of freedom to pursue my own interests. Many of the things I appreciate in research can be traced back to an idealized version of my parents' upbringing. Without them, I cannot achieve what I achieve today.

Abstract

Sentiment analysis in social networks has been widely analysed over the last decade. Despite the amount of research done in sentiment analysis in social networks, the distinct categories are not appropriately considered in many cases, and the study of dissemination patterns of emotions is limited. Therefore, understanding the significance of specific emotions will be more beneficial for various marketing activities, policy-making decisions and political campaigns.

The current PhD thesis focuses on designing a theoretical framework for analyzing the broad spectrum of sentiments and explain how emotions are propagated using concepts from temporal and multilayer networks. More precisely, our goal is to provide insights into emotion influence modelling that solves emotion estimation problems and its temporal dynamics nature on social conversation. To exhibit the efficacy of the proposed model, we have collected posts related to different events from Twitter and build a temporal network structure over the conversation.

Firstly, we perform sentiment analysis with the adaptation of a lexicon-based approach and the circumplex model of affect that enhances the effectiveness of the sentiment characterization. Subsequently, we investigate the social dynamics of emotion present in users' opinions by analyzing different social influential characteristics. Next, we design a temporal emotion-based stochastic model in order to investigate the engagement pattern and predict the significant emotions. Our ultimate contribution is the development of a sequential emotion-based influence model with the advancement of recurrent neural networks. It offers to predict emotions in a more comprehensive manner.

Finally, the document presents some conclusions and also outlines future research directions.

Resumen

El análisis de sentimientos en redes sociales se ha estudiado ampliamente durante la última década. A pesar de ello, las distintas categorías de sentimientos no se consideran adecuadamente en muchos casos, y el estudio de patrones de difusión de las emociones es limitado. Por lo tanto, comprender la importancia de emociones específicas será más beneficioso para diversas actividades de marketing, toma de decisiones empresariales y campañas políticas.

Esta tesis doctoral se centra en el diseño de un marco teórico para analizar el amplio espectro de sentimientos y explicar cómo se propagan las emociones utilizando conceptos de redes temporales y multicapa. Particularmente, nuestro objetivo es proporcionar información sobre el modelado de la influencia de las emociones y como esta afecta a los problemas de estimación de las emociones y a la naturaleza dinámica temporal en la conversación social. Para mostrar la eficacia del modelo propuesto, se han recopilado publicaciones relacionadas con diferentes eventos de Twitter y hemos construido una estructura de red temporal sobre la conversación.

En primer lugar, realizamos un análisis de sentimientos adoptando un enfoque basado en el léxico y en el modelo circunflejo de emociones de Russell que mejora la efectividad de la caracterización del sentimiento. A partir de este análisis investigamos la dinámica social de las emociones presente en las opiniones de los usuarios analizando diferentes características de influencia social. A continuación, diseñamos un modelo estocástico temporal basado en emociones para investigar el patrón de participación de los usuarios y predecir las emociones significativas. Nuestra contribución final es el desarrollo de un modelo de influencia secuencial basado en emociones mediante la utilización de redes neuronales recurrentes que permiten predecir emociones de una manera más completa.

Finalmente, el documento presenta algunas conclusiones y también describe las direcciones de investigación futuras.

Resum

L'anàlisi de sentiments en xarxes socials s'ha estudiat àmpliament durant l'última dècada. Malgrat això, les diferents categories de sentiments no es consideren adequadament en molts casos, i l'estudi de patrons de difusió de les emocions és limitat. Per tant, comprendre la importància d'emocions específiques serà més beneficiós per a diverses activitats de màrqueting, presa de decisions empresarials i campanyes polítiques.

Aquesta tesi doctoral se centra en el disseny d'un marc teòric per a analitzar l'ampli espectre de sentiments i explicar com es propaguen les emocions utilitzant conceptes de xarxes temporals i multicapa. Particularment, el nostre objectiu és proporcionar informació sobre el modelatge de la influència de les emocions i com aquesta afecta als problemes d'estimació de les emocions i a la naturalesa dinàmica temporal en la conversa social. Per a mostrar l'eficàcia del model proposat, s'han recopilat publicacions relacionades amb diferents esdeveniments de Twitter i hem construït una estructura de xarxa temporal sobre la conversa.

En primer lloc, realitzem una anàlisi de sentiments adoptant un enfocament basat en el lèxic i en el model circumflex d'emocions de Russell que millora l'efectivitat de la caracterització del sentiment. A partir d'aquesta anàlisi investiguem la dinàmica social de les emocions present en les opinions dels usuaris analitzant diferents característiques d'influència social. A continuació, dissenyem un model estocàstic temporal basat en emocions per a investigar el patró de participació dels usuaris i predir les emocions significatives. La nostra contribució final és el desenvolupament d'un model d'influència seqüencial basat en emocions mitjançant la utilització de xarxes neuronals recurrents que permeten predir emocions d'una manera més completa.

Finalment, el document presenta algunes conclusions i també descriu les direccions d'investigació futures.

Contents

List of Figures	xiii
List of Tables	xv
1 Introduction	1
1.1 Background	1
1.2 Research Motivation	5
1.3 Research Objectives and Contributions	6
1.4 Research Related Activities	9
1.4.1 Related Publications	10
1.4.2 Scientific Research Stays	11
1.5 Structure of Thesis	12
2 State of the Art in Sentiment Analysis and Opinion Dynamics	13
2.1 Sentiment Analysis on Social Networks	13
2.2 Social Networks Influence and Information Propagation	18
2.2.1 Social Networks Influence Analysis	19
2.2.2 Social Networks Propagation Phenomena	22
2.3 Technique and Modelling for Opinion/Emotion Dynamics	24
2.3.1 Technique for Opinion/Emotion Dynamics	25

CONTENTS

2.3.2	Models for Opinion/Emotion Dynamics	29
2.3.3	Other Related Models	30
2.4	Conclusions	30
3	Emotion Dynamics of Public Opinions	32
3.1	Introduction	32
3.2	Data Preparation and Emotional Modeling	36
3.2.1	Data Pre-processing and Emotion Extraction	38
3.2.2	Temporal Emotional State Chain	46
3.3	Emotion Transition on Twitter	50
3.4	Understanding Influence of Incoming Tweets	57
3.4.1	Can Majority Opinions Influence an Individual's Opinion?	57
3.4.2	Which Community Channel is More Influential?	62
3.4.3	Which Conversational Channel is More Influential?	63
3.5	Characteristic of Incoming Tweets towards State Transition	63
3.5.1	State Transition vs. Incoming Dominant Emotion	64
3.5.2	Which Community Channel is More Influential during Emotion State Transition?	67
3.5.3	Which Conversational Characteristics are More Influential during Emotion State Transition?	68
3.6	Conclusions	69
4	Sequential Influence Model for Emotion Dynamics using Hidden Markov Model	71
4.1	Introduction	71
4.2	Proposed Framework	74
4.2.1	Hidden Markov Model	74
4.2.2	Emotion-based Hidden Markov Model	79

4.3 Experiments and Discussions	80
4.3.1 Experimental Set-up	80
4.3.2 Experimental Results	84
4.4 Conclusions	96
5 Sequential Influence Model for Emotion Dynamics using Deep Learning	97
5.1 Introduction	97
5.2 Proposed Framework	100
5.2.1 Emotion-Based User Sequential Influence Model (E-USIM)	100
5.2.2 Emotion-Based Prediction	105
5.2.3 Parameter Estimation	105
5.3 Experiments and Discussions	106
5.3.1 Experimental Set-up	106
5.3.2 Experimental Results	107
5.4 Conclusions	117
6 Conclusions and Future Work	119
6.1 Summary of Contributions	121
6.1.1 Emotion Dynamics of Public Opinions	121
6.1.2 Sequential Influence Model for Emotion Dynamics using Hidden Markov Model	122
6.1.3 Sequential Influence Model for Emotion Dynamics using Deep Learning	123
6.2 Future Work	124
Appendices	126

CONTENTS

A Tweet Retrieval Methods	126
A.1 Data Collection 1	127
A.2 Data Collection 2	128
A.3 Data Collection 3	128
A.4 Data Collection 4	129
References	131

List of Figures

2.1	Sentiment Analysis Approaches	15
3.1	Russell’s Model of Affect (55)	39
3.2	A Semantic Structure of Affect (22, 55). The letters x and y represent semantic components: x = Pleasantness; y = Activation . . .	40
3.3	Global Overview of Sentiment Analysis in Tweets	45
3.4	Temporal Tweet Chain	48
3.5	Temporal Emotional State Chain	49
3.6	Macro average transition probabilities over all topics. Initial state probability is represented by dot arrow.	51
3.7	Transition probability of a user over topics in different categories (Part-1).	53
3.8	Transition probability of a user over topics in different categories (Part-2).	54
3.9	Frequency of the user changing emotion region across topics	56
3.10	Average probabilities of emotional agreement over all topics	64
3.11	Average probabilities of topological agreement over all topics	66
3.12	Average probabilities of conversational agreement over all topics . .	68

LIST OF FIGURES

4.1	Hidden Markov Model	75
4.2	Performances on emotional regions in different categories (Part-1).	89
4.3	Performances on emotional regions in different categories (Part-2).	90
5.1	Emotion-based User Sequential Influence Model	101
5.2	Performances on emotional regions in different categories (Part-1).	110
5.3	Performances on emotional regions in different categories (Part-2).	111
5.4	Effect on Different Data Partitions	117
A.1	Data Collection Method 1	127
A.2	Data Collection Method 2	127
A.3	Data Collection Method 3	128
A.4	Data Collection Method 4	129

List of Tables

3.1	Size of the all datasets	35
3.2	Notation of different attributes	43
3.3	Agreement and disagreement by all topics with average transition in Figure 3.6 P_{11} considers self transition with high positive emotion, i.e., $R1$, P_{ij} considers transition from self state to other states, $P(\rightarrow i)$ considers from others state to $R1$ and π_i considers initial state.	55
3.4	It shows the probability of an individual's emotion state in a post not matching with the emotion state of any of the incoming message	58
3.5	The probability of an individual's emotion state in a post matching with the emotion state of any of the incoming message	60
3.6	Continued	61
4.1	Log-likelihood $P(O \lambda)$ of three different initialization methods (E-HMM ₁ , E-HMM ₂ and E-HMM ₃) and other methods to be compared using DeGroot, Flocking, Voter and AsLM.	85
4.2	Overall performance of estimated models for each dataset.	87
4.3	Performance metrics for Policy Category	91
4.4	Performance metrics for Movie Category	92

LIST OF TABLES

4.5	Performance metrics for Sport Category	92
4.6	Performance metrics for Terror Attack Category	93
4.7	Performance metrics for Accident Category	93
4.8	Performance metrics for Politics Category	94
5.1	Overall performance of estimated models for each dataset.	108
5.2	Performance metrics for Policy Category	113
5.3	Performance metrics for Movie Category	113
5.4	Performance metrics for Sport Category	114
5.5	Performance metrics for Terror Attack Category	114
5.6	Performance metrics for Accident Category	115
5.7	Performance metrics for Politics Category	115

1

Introduction

"Education is the manifestation of the perfection already in man."

— Swami Vivekananda

1.1 Background

Social media have become an essential medium for people to acquire and broadcast information with explosive growth popularity. Social media occupies a large part of everyone's life, and human behaviour in social networks has been widely analyzed over the last decade. The progress from a read-only to a read-write web allowed internet users to interact, share their opinions through social networks, blogs and wikis. Especially a famous and sophisticated microblog service Twitter is profoundly used in various tasks. Users through the use of limited vocabulary, are able to express their opinion or views about a particular topic/event. This online expression is applicable for the healthcare organisation, businessman, stockbroker, product seller and so on. In 2020 statistics, the number of social network users

1. INTRODUCTION

is 3.6 billion worldwide¹. The way people communicate has been changed by the popularity of social media and how people run their business. At the same time, social influence in online communication has been increasing profoundly and can gradually change someone's opinion. Therefore, understanding social influence by analyzing someone's opinion in social media is a significant study and it's ranging from businessman to politicians. Automatic analysis of online opinions involves a deep understanding of natural language text by machines, from which we are still so far. Many analyses have been made that measure the impact of opinion on a topic/event in the social conversation. Nevertheless, this opinion is often expressed through the binary classification of sentiment (positive or negative), but multiple classes (happy, sad, angry, etc.) are not appropriately considered. Sentiment analysis includes this type of information, but usually just as a polarity measure to determine the general sense of the opinion. This has led to the emerging fields of opinion mining and sentiment analysis (64), including information retrieval, text mining and natural language processing (NLP) methods to extract sentiments and opinions from the vast mass of textual information on the internet.

Sentiment Analysis (SA) is the computational study of people's opinions, appraisals, evaluations and emotions toward entities. Existing approach of sentiment analysis has been classified by several way, for instance, affective valence of social media content as negative or positive (59), studies the subjectivity or objectivity of text (106), extracts the emotion and its arousal (111). Sentiment analysis became very attractive research field and also a number of techniques are increasing rapidly to identify the reaction of the users' on twitter dataset (3). The accessibility to Twitter's information brings up a whole new class of SA tasks related to the changing nature of sentiments over time. Particularly, the analysis of in-

¹<https://www.statista.com/markets/424/topic/540/social-media-user-generated-content/#statistic1>

fluence and sentiment propagation in Twitter became a very attractive research field. With the growing use of the social network (Twitter), we have evidenced the enormous growth of sentiment propagation (45, 157, 161, 177) research on Twitter. For instance, Giachanou et al. (66) propose two approaches for propagating sentiment signals to estimate tweets' reputation polarity. In addition, in the last few years, many researchers exploited the Twitter social network under the assumption that people influence one another in the same conversation and have witnessed a massive growth in research on measuring user influence (136), measuring opinion dynamics (29), tracking opinion dynamics (30), predicting opinion dynamics (191) etc. However, these existing studies are therefore targeted to information propagation of the individual user without capturing neighbours' long-term historical information which is another important criterion. Because in social platform, the user generally follows his/her neighbour during propagation, and if any user changes an opinion and sentiment on a specific topic, s(he) is most likely to get influenced by his/her neighbour. Secondly, a user's opinion is not static, and it could change over time. Thirdly, a user's future opinion predicts with the aggregated of long-term past historical opinions of her/his neighbour on the topic/event.

Despite the concreteness of the online expressions in written form, it remains unpredictable which kinds of emotions will arise/be expressed in individual messages of Twitter users. Exploring the mechanism of predicting a user's emotion by influencing his/her neighbours and tracking dynamics changes of their opinion is practically useful but technically challenging. The existing opinion influence models are therefore targeted to model the opinion dynamics of each user individually, and learns the interpersonal influence receives from her/his neighbours by capturing their long-term historical information. Therefore, uncovering users' dynamics opinion from their neighbours' long-term past history and represent the modified one is very important to better guesstimate the actual opinion on the

1. INTRODUCTION

topic/event. Identifying influential features on social networks can help us to understand factors causing people to change their opinions and its exploit users for updating their opinion. In general, Twitter predicts events by searching for an interrelationship between the emotion and the events over time. However, these works paid little attention to predicting users' emotion evolution in a social network. Observing emotions is a form of intelligence task which requires sequence labeling. These tasks are processed in layers where only a constrained amount of information can be held. Besides, due to a large number of social media data that have been collected to perform our task, therefore, it's necessary to create computational methods to process the data.

Social media generates various data types related to users' profiles, such as their dynamic opinion with the emotion expressed in the text messages and their social relationships every day. Mining opinions and sentiments from natural language is an extremely non-trivial process in social media. It concerns a deep comprehension of most syntactic and semantic rules specific to each language. A considerable amount of research has been done. There are many commercial companies that supply sentiment analysis services. Therefore, the development of such systems is important for commercial purposes and for government intelligence applications to track temporal positive or negative communications. All of these tools are still mainly keyword-based and, consequently, frequently fail to meet the key standards of human annotators. The principal aim of this thesis is to go behind such techniques by designing *a novel system to identify user's influential opinion and predict their dynamic emotion from the temporal written conversation.*

1.2 Research Motivation

Measuring users influence and emotion detection of public opinion in the social network are two became a very attractive study in the current research area. Most of the work on Twitter influence measures have been conducted by considering a limited number of influence features. On the other hand, users' opinions have been classified based upon the sentiment of a positive or negative posting. As a result, both areas have some limitations to understand the underlying mechanism of user influence and dynamics of emotion in social conversation. Therefore, the limitations mentioned above have been taken care of in this thesis.

In social media, a pair of users' posts are linked to each other, allowing us to study the propagation of stories and determine who is an expert or an influencer on given topics. It is observed from various studies that influential messages or users create a high impact on the social ecosystem. However, it is still unclear *"if social media or networking platform can be effectively used for opinion polarization or social contagion"*. There are many possibilities that how people are reacting to the potential influencer's tweet and effect to user's emotion which appear in the next tweet message. Specifically, how much amount of emotion influence effect to the user's opinion. We surmise that a public opinion on a social issue on Twitter carries a certain degree of emotion, and there is an emotion flow underneath the Twitter network. The last few years have evidenced a massive growth of propagation study on Twitter and came up with several proposed diffusion models. The original tweet propagated rapidly by retweeting (150), and this mechanism is immensely powerful for information diffusion on Twitter. In addition, the accessibility to the information provided by Twitter brings up a whole new class of SA tasks related to the changing nature of sentiments over time and examine the different role of sentiment in information propagation (149). For

1. INTRODUCTION

instance, some works that focus on tracking sentiments over time and performing temporal prediction in Twitter conclude the existence of a correlation between sentiment variation with cultural, social, economic and political events (18) or with major climate events (36). These works have some limitations to cope with the dynamic temporal emotion that users are continuously changing to being influenced by their neighbours. Because number of researchers worked on influence measure and emotion prediction by building several models. But combination of these two approaches is a new idea where we consider the degree of different emotional state to identify the user's actual emotional state. Therefore, this kind of problem gives both challenges and opportunities to the research in this thesis.

In this thesis, we concentrate our studies on: (1) building a model to extract emotion from the conversation on different events; (2) identifying influential features to understand factors causing people to change their opinions on social media; (3) uncovering emotional influence by monitoring temporal emotion dynamics; and (4) predicting individual's future emotions by estimating internal/external influence.

1.3 Research Objectives and Contributions

This section presents the precise objectives that guided the development of this PhD thesis, along with the associated tasks conducted throughout this research and the resulting contributions:

Objective 1: Emotion Dynamics of Public Opinions: This initial objective entails collecting a massive number of messages from Twitter. We primarily used hashtags (#) for retrieving tweets through Twitter *Search-API*. In this work, we proposed a framework to identify affective emotional feelings associated with

1.3 Research Objectives and Contributions

a message according to Russell’s affection model using the ANEW lexical dictionary. This model enables us to define a more accurate sentiment of the messages and discriminate between cases in which the general sentiment is just positive or negative (for example, when a catastrophe occurs).

The first step is to have a word representation for each message in order to facilitate machine manipulation as well as the elimination of noisy words. Processing each individual tweet message involves extracting bags of words, filtering all stop words and extracting stems by using the stemmer tool of the WordNet. Finally, the emotional role of the stem sets of tweets is obtained through the emotional base ANEW dictionary and words are annotated by their valence and arousal values. Words are then classified according to Russell’s emotion model. Several sentiment models have been proposed from social sciences. The most suitable one has to be chosen according to the scenario. The sentiment will be inferred from the participation that people has on social networks when talking about this topic/event. Finally, we build a ***Temporal Emotional State Chain*** (TESC) framework where every participating user’s data is defined by the sequence of *outgoing* and *incoming* tweets arranged in the order of posting time.

Our work focuses on emotion dynamics of a user while posting comments against an event/topic from three different social characteristics: (i) emotional (excitement, contentment, depression and distress); (ii) community (follower followee, membership relationship); and (iii) conversational (tweet, retweet, mention, reply). In this work, we investigate social dynamics of emotion present in users’ opinions and attempt to understand (i) changing characteristics of users’ emotions toward a social issue over time; (ii) influence of public emotions on individuals’ emotions; (iii) cause of changing opinion by social factors, and so on. Our hypothesis that public opinion on a social issue on Twitter carries a certain degree of emotion, and there is an emotion flow underneath this network.

1. INTRODUCTION

We used TESC framework as a backbone to accomplish this objective. From this emotion state chain, we analyse the transition probability of users emotional state change in subsequent tweets and the relationship between user’s emotional state and the nature of the topic. This analysis also focuses on the user’s participation in the network conversation and their frequency while transiting from one state to another. Next, we attempt to understand if majority opinion can influence an individual’s opinion or be biased by the opinions coming from the group/community that (s)he belongs. Finally, we investigate the distribution of the incoming tweets that an individual receives at the time of changing his/her emotional state while posting a tweet against a topic.

Objective 2: Sequential Influence Model for Emotion Dynamics using

Hidden Markov Model: The objective of the work implies proposing a mathematical model that given a set of tweets related to some events, determines how those sentiments will be distributed and determine the impact of the emotion reflected in the read/written messages of a person within a conversation. More specifically, we aim to systematically attempt to answer some research questions: (i) Given a set of tweets related to some events, how do we optimize the model parameters for learning? (ii) How can we infer how emotion will be distributed from the observed sentiment-labelled tweets? and (iii) How can we evaluate a user’s emotional state on different events or topics?

In order to perform an investigation on observing emotions unfolding in a consecutive sequence of tweets for a particular user and predict emotion evaluation at the user-level, we proposed E-HMMs with the adaptation of HMM. The backbone of this proposed model relies on the TESC framework, which has been introduced in Objective 1. Finally, Baum-Welch forward-backward algorithm applied in order to learn the HMMs and evaluation is performed using the MLE.

Objective 3: Sequential Influence Model for Emotion Dynamics using Deep Learning: The ultimate objective is to deeply analyze users' emotional dynamics derived within a conversation influenced by neighbours' opinion and predicting updated emotion on Twitter. The process of measuring a person's probable emotion and recognising how a person changes his/her emotion under the influence of neighbours' opinions have been studied in the previous two works. But deep analysis and a unified framework still required to increase the accuracy. In this work, we aim to study the components mentioned along with the interpersonal/personal emotion influence and uncover the users' significant emotion by capturing their historical information.

To address this objective, we proposed a novel framework called E-USIM, which captures temporal properties of emotion dynamics and the meaningful information included in the long-term sequence. In this model, we adopt RNN architecture and its variant Gated Recurrent Units GRUs which has an able to integrate the historical information with the new coming information for prediction. To train our model, we utilise the vanishing gradients problem using the BPTT approach and maximize the log-likelihood of the sequences of emotions. This fine-grained prediction strategy with the ability to forecast future emotion may benefit to uncover actual opinion about a topic/event in the social network.

1.4 Research Related Activities

This section lists the research activities performed during this PhD thesis development, namely the related scientific publications, research stays, and other publications.

1. INTRODUCTION

1.4.1 Related Publications

The following subsections list all the scientific publications related to this research. We classify articles according to the type of publication, such as the articles appearing in journals listed in the Science Citation Index (SCI). The papers published in the proceedings of relevant conferences included in the Computing Research and Education Association of Australasia (CORE) rankings.

Finally, the section [1.4.1.3](#) lists other relevant scientific articles without an impact factor or/not published in a ranked conference.

1.4.1.1 Publications in SCI Journals

- D. Naskar, S. Ranbir Singh, D. Kumar, S. Nandi, and E. Onaindia. **Emotion Dynamics of Public Opinions on Twitter**. *ACM Transactions on Information Systems (TOIS)*. Volume 38(2), pages 18:1-24, 2019.

1.4.1.2 Publications in CORE Conferences

- D. Naskar, E. Onaindia, M. Rebollo, and S. Ranbir Singh. **Predicting Emotion Dynamics Sequence on Twitter via Deep Learning Approach**. In *Proceedings of the 18th International Conference on Advances in Mobile Computing and Multimedia (MoMM)*. Chiang Mai, Thailand, Pages 20-24, 2020.
- D. Naskar, E. Onaindia, M. Rebollo, and S. Das. **Modelling Emotion Dynamics on Twitter via Hidden Markov Model**. In *Proceedings of the 21st International Conference on Information Integration and Web-based Applications Services (iiWAS)*. Munich, Germany, Pages 245-249, 2019.
- D. Naskar, S. Mokaddem, M. Rebollo, and E. Onaindia. **Sentiment Analysis in Social Networks through Topic Modeling**. In *Proceedings of*

1.4 Research Related Activities

the 10th International Conference on Language Resources and Evaluation (LREC). Portorož, Slovenia, Pages 46-53, 2016.

1.4.1.3 Other Publications

- D. Naskar, N. Hasan and A. Das. **Pattern of social media engagements by the learners of a library and information science MOOC course: an analytical study**. *Annals of Library and Information Studies (ALIS)*. Volume 68(1), pages 56-66, 2021.
- N. Hasan and D. Naskar. **ARPIT Online Course on Emerging Trends & Technologies in Library & Information Services (ETTLIS): A Case Study**. *DESIDOC Journal of Library & Information Technology (DJLIT)*. Volume 40(3), pages 160-168, 2020.
- D. Naskar and S. Das. **HNS Ontology using Faceted Approach**. *Knowledge Organization (KO)*. Volume 46(3), pages 187-198, 2019.
- D. Naskar, M. Rebollo, and E. Onaindia. **Análisis de Sentimientos en Twitter Mediante Modelos Ocultos de Markov**. In *Proceedings of the XXI Congreso Nacional de Física Estadística (FisEs)*. Sevilla, España, Pages 92, 2017.
- D. Naskar, and B. Dutta. **Ontology And Ontology Libraries: A Study From An Ontofier And An Ontologist Perspective**. In *Proceedings of the 19th International Symposium on Electronic Theses and Dissertations (ETD)*. Lille, France, Pages 1-12, 2016.

1.4.2 Scientific Research Stays

The following research stay was completed during the research period associated with this PhD thesis:

1. INTRODUCTION

- **01-10-2017 to 10-05-2019.** *Indian Institute of Technology Guwahati, Assam, India.* Research stay in Open Source Intelligence (OSiNT) Group of the Department of Computer Science and Engineering under the supervision of Professor Sukumar Nandy and Dr. Sanasam Ranbir Singh on Twitter sentiment analysis.

1.5 Structure of Thesis

The overall picture of the PhD thesis is as follows. Chapter [1](#) briefly introduces the background of the studies on emotion dynamics modelling. The research motivation, objectives and contributions and research-related activities are also explained. Chapter [2](#) explains the surveys of existing work on sentiment analysis, information propagation, sentiment propagation, opinion dynamics models and other influence models. After a brief summary of the literature, we point out the differences between the existing studies with opinion influence modelling explored in the thesis work. Chapter [3](#) presents data pre-processing and emotion extraction method. It also investigates the emotion transition process and attempt to understand emotion influences from three different social characteristics, i.e., emotional, community and conversational. Chapter [4](#) shows the emotion dynamics model through HMM and adopting temporal emotion properties of user interactions. Chapter [5](#) explores the emotion dynamics model through deep learning mechanisms and predicts modified opinion by adapting temporal emotion influence. Chapter [6](#) summarizes the proposed methods, conclusions, and contributions of the work. The potential future extensions of the current research are suggested at last.

2

State of the Art in Sentiment Analysis and Opinion Dynamics

“A sentimentalist is simply one who desires to have the luxury of an emotion without paying for it.”

— Oscar Wilde

2.1 Sentiment Analysis on Social Networks

Sentiment analysis (SA) is a type of text classification that deals with subjective statements (127). Sentiment analysis also known as *Opinion Mining* (OM) that quantify the people’s opinions through natural language processing (NLP) (90), computational linguistics and text analysis. Opinion mining and sentiment from natural language, is an extremely non-trivial process as it concerns a deep comprehension of most syntactical and semantic rules specific to each language. Pang

2. STATE OF THE ART IN SENTIMENT ANALYSIS AND OPINION DYNAMICS

and Lee (127) consider that SA or OM is the "computational study of opinions, feelings and subjectivity in text". It's a computational study of people's opinions, feelings, appraisals, attitudes, evaluations and emotions toward entities(107). It refers to classify the psychological state from text and determine someone's attitude on a particular topic or event. In general, the common task is to classifying the text through positive or negative orientation and in some cases neutral orientation. Main focus of sentiment analysis is to construct lexicon framework, extract selected feature, and to determined polarity level (116).

Sentiment analysis became very attractive research field and also a number of techniques are increasing rapidly to identify the reaction of the users' on twitter dataset (3). With the growing use of social networks in which there is a huge explosion of sentiments and opinion reviews among individuals, issues, events, topics. Existing approach of sentiment analysis has been classified by several way, for instance, affective valence of social media content as negative or positive (59), studies the subjectivity or objectivity of text (106), extracts the emotion and its arousal (111).

A considerable amount of research has been done (106). The automatic analysis of online opinions involves a deep understanding of natural language text by machines, from which we are still very far (81). This has led to the emerging fields of opinion mining and sentiment analysis, which includes information retrieval, text mining and natural language processing (NLP) methods to extract sentiments and opinions from the huge mass of textual information. Mohammad and Yang (114) proposed that sentiment analysis is applied to emails for gender identification through emotion analysis. Hockenbury and Hockenbury (78) define sentiment or emotion is a complex psychological state that involves three distinct components: a subjective experience, a physiological response, and a behavioural or expressive response. There are also many commercial companies that supply

2.1 Sentiment Analysis on Social Networks

sentiment analysis services. Thus, the development of such systems is important for commercial purposes and also for government intelligence applications to track positive or negative communications (1).

The general goal of textual emotion detection and classification is to detect and identify the type of emotion, for example, happiness, pleasure and arousal (109). Generally, the most common methods adopted by researchers are mainly statistics based models. Pang and Lee (127) were the pioneers to employ machine learning methods like Support Vector Machine (SVM) (153), Maximum Entropy (ME) (15) and Naive Bayesion (NB) (57) are some of the well used methods for text classification.

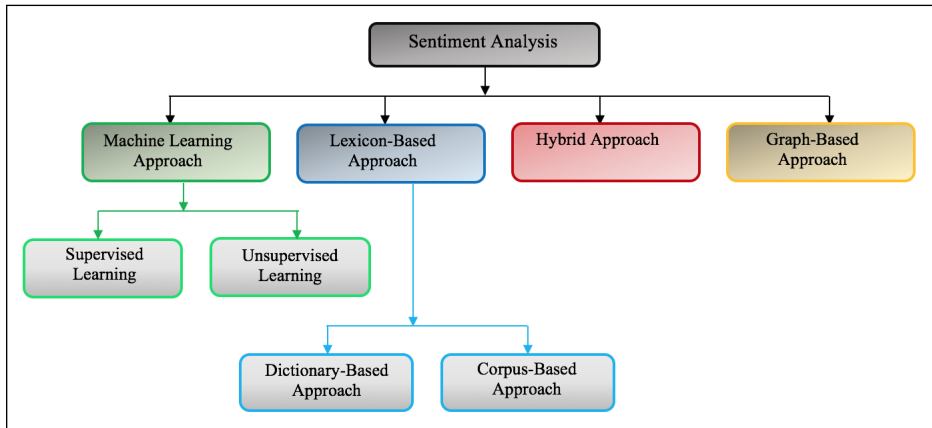


Figure 2.1: Sentiment Analysis Approaches

Different approaches are employed in order to extract sentiment and opinion mining, as shown in Figure 2.1. The existing approaches to social emotion mining either can be classified depending on the purpose (181) or the exploiting sentiments level. The following method can be identified by studying of sentiment analysis

2. STATE OF THE ART IN SENTIMENT ANALYSIS AND OPINION DYNAMICS

(64):

Machine learning based approach: The machine learning based approach is one of the most flexible and adaptable, well known approach for sentiment analysis. Mostly, this approach deal with sentiment analysis in twitter on the basis of various supervised learning. In order to classify the unseen training dataset, the classifier from the field of machine learning has presented a series of feature vectors of tweets. Some most prominent feature vector is classified by n-gram (sequence of N words), unigrams (single word phrases), bigrams (two consecutive phrases), trigrams (three consecutive phrases), polarity based on several algorithm. Many researches have been conducted their research by using most appropriate names of the classifiers such as Support Vector Machines(SVM)(68), Naïve Bayes (NB)(51), Multinomial Naïve Bayes (MNB)(164), Maximum Entropy (ME)(94) and Artificial Neural Network (95). The purpose of the machine learning approach is to retrieve accurate results by using different classification methods. Accuracy reported always depending upon the combination of various features selected. Mainly, this approach has follows five stages to complete the process of retrieving result, such as Collection of Data, Pre-processing, Training Dataset, Data Classification and Results.

Lexicon based approach: Lexicon based approach is simple and coherent, mainly extract sentiment from the text. This method annotated by polarity score of input text to obtain overall opinion. The input text tokenize by using tokenization method. This method depends on overall score in order to classify the text and this supervised classification approach imply on text or sentence for classification purpose (128). The method generally count positive and negative words from

2.1 Sentiment Analysis on Social Networks

the text (46). The set of words is obtained through WordNet¹ dictionary and SentiWordNet (54), MPQA (Multi-Perspective Question Answering) (174) lexicons. This method mainly follows four stages to complete the retrieving process of the result, such as Dataset input, pre-processing, polarity detection and data classification. Different classifier used different lexicon based algorithm in order to detect sentiment and classify the words. For instance, SentiStrength is one of the most conventional lexicon based algorithm to identify the sentiment of informal text(tweets) (159). SentiCircles (142) is one of the lexicon approach to automatically captures patterns of words of similar contextual semantics and sentiment in tweets.

Hybrid/Combined approach: Hybrid and combined analysis defines a fusion of machine learning and lexicon based approach. This analysis advance level sentiment analysis, in contrast to corpus based and dictionary based methods which could collectively explore for more accuracy. Interesting hybrid approach (62) that uses ngram analysis for feature extraction and a dynamic artificial neural network (61) algorithm as alternative approach to classify each tweet into a sentiment class. Another interesting combined approach given by (89), where they represent Twitter opinion mining framework (TOM) by using three layers of classifier process, i.e., Enhanced Emoticon Classifier (EEC), Improved Polarity Classifier (IPC) and SentiWordNet Classifier (SWNC).

Graph based approach: Graph-based analysis is selective, non-parametric and decent approach on sentiment analysis in twitter. This method can be applied in both semi-supervised and cross domain tasks. Graph based learning exploits the ability of the annotated data to be represented through label propagation

¹<http://wordnet.princeton.edu/>

2. STATE OF THE ART IN SENTIMENT ANALYSIS AND OPINION DYNAMICS

method (64). Label propagation is a semi-supervised method which represented as a weighted graph where instances are nodes and edges reflect interrelation. From this method, many researchers can be assumed the influence of people through their conversation by using label of propagation. Different feature has been applied to fit the model of classification, propagation into the twitter social graph such as, hashtags, emoticons, Twitter follower graph, punctuation, followers/followees. Speriosu et al. (148) proposed one interest label propagation (LPROP) method for analyzing influence of the people through Twitter follower graph. Another interesting propagation method represented by Wang et al. (165) where they used different algorithm (Loopy Belief Propagation, Relaxation Labeling, Iterative Classification Algorithm) to perform hashtag-level sentiment classification which initialize the sentiment polarity distribution for every hashtag.

However, our proposed framework based on three approaches, where lexicon approach used for sentiment classification, machine learning approaches used for predicting user's emotion and graph based approach used for understand users dynamics influence pattern. We mainly focus on unsupervised leaning, because it's time-saving and easy to be adapted over different datasets (155).

2.2 Social Networks Influence and Information Propagation

This section briefly reviews the earlier literature that exploits different characteristics of opinion dynamics, measuring social influence and its propagation in social media. The study of social influence analysis and information propagation has become an attractive research area in sociology, physics, marketing and computer science. On Twitter, several studies have been conducted on changing opinion and

2.2 Social Networks Influence and Information Propagation

discussion on user influence. Many researchers have examined how message content affects individual retweeting decisions. They show users can influence brand content diffusion via retweets (8), the role of content influence on social media via retweets behaviour (189). Authors in the study (124) measures of public opinion derived from polls with sentiment measured from text analysis from the popular microblogging site Twitter.

2.2.1 Social Networks Influence Analysis

Social Behavioral Aspect: Emotional contagion influences individual and group-level communication behaviour in terms of information coordination and sharing (151). In the same direction, (56) conducted a study on the dynamics of emotional contagion using a random sample of Twitter users and measure the emotional valence of content the users are exposed to before posting their own tweets. A high level of cognitive involvement, such as anger, anxiety, awe, or amusement, might also trigger a high physiological arousal level. In contrast, low arousal or deactivation is characterized by relaxation, and high arousal or activation is denoted by activity (16). Social networking is a multidimensional concept where users share a different behavioural aspect (149) over a topic. An individual emotion on behavioural concept possible to utilize through social networks in viral marketing. In this part, we study different emotions transition of a user shows in the social network, and also the other type of emotions detection (37). (115) represent a model in which information can reach a node via the links of the social network or through the influence of external sources. The model used to infer the quantify the external influences over time and describe how they affect information adoption. Compared with other studies, we consider four regions based on 16 emotions of Russell’s model of affect. This emotions model combined with two main dimensions (i.e. valence and arousal) in a 2D circular space.

2. STATE OF THE ART IN SENTIMENT ANALYSIS AND OPINION DYNAMICS

The emotion transition leads to finding out how likely users express their emotions after receiving a tweet responding. A study conducted by authors in (93) where they show social aspects of the user's emotion by using Plutchik's wheel model and also examine that the conversational partners can influence each others' emotions and topics. (119) investigate collective public mood states derived from large-scale collections of daily Twitter posts over time. They analyze tweets using two mood tracking tools, namely OpinionFinder and 6 dimensions mood measures Google-Profile of Mood States (GPOMS). Similarly, (5) attempt to study the nine basic human emotions and their significance in various social network activities to determine the right strategies of marketing in e-business.

Social Influential Characteristics: Understanding influential factors is an important task to understand the user's dynamics opinion on a social network. A comprehensive comparison of various influential factors (indegree, retweets and mentions) on users' social dynamics is studied in (25). A similar study is conducted by (182) where they measure propagation patterns of tweet messages and social influence by following three metrics, i.e. follower, reply and retweet. (129) examined a set of distinct characteristics (Dynamic, Propagative, Composable, Measurable, Subjective, Asymmetric and Event-sensitive) to understand user's dynamics and identify influential users on Twitter. (99) study information diffusion pattern of topological features, namely singleton, reply, mention and retweet. They also study the temporal behaviour of trending topics. Unlike the above studies, (93) explores the effect of users' social and conversational characteristics on emotional dynamics. Specifically, they look into the social conversational features that lead to the transition of emotion states within a discussion chain. (144) propose a theoretical framework to systematically investigate the determinants of individual dissemination behaviour in a Twitter network. They found information

2.2 Social Networks Influence and Information Propagation

related to topical preference and homophily value are most influential on individual dissemination behaviour.

From the above discussion, we observe that while several studies have been conducted to study the community channel, not many of them study the relationship between the community channel and the tweets' emotional aspect. Few studies (25, 93) that have been conducted in this direction do not consider an extensive range of characteristics. Remarkably, this study evaluates the effect of a wide range of community and conversational features on users' emotion dynamics while posting their opinion. Our experiment is completely based on Twitter datasets, and all the possible sources are accommodated into this analysis. A popular study (25) defined that the majority of the people are influenced by three essential activities such as followers influence, retweets influence and mentions influence. Including these three sources of influence, our study also covers some more extra parameters of influences such as hashtag tweets, replies, member lists etc. Our finding shows that the member-list is one of the essential community channels, which shows that more influential to the user and retweets are more influential characteristics.

Social Influence Evaluation Measure: Identifying influential users is an important aspect of social media-related studies. Identifying influential users can aid in tasks like social or political campaigns or viral marketing etc. In this direction, (189) study the influence of content and users on the rebroadcasting pattern of a message. They observe that along with the content of a message, the rebroadcasting of a message by a user is also dependent on other users and the relevance of the message to the user. Authors conducted a similar study in (8) where they study users' influence in the diffusion of information in a Twitter network. (171) proposed Twitter Rank algorithm, an extension of the PageRank algorithm to find the influential users in a Twitter network for a given topic. (99) propose different

2. STATE OF THE ART IN SENTIMENT ANALYSIS AND OPINION DYNAMICS

measures for ranking influential users and report a comparison among them. (48) proposed a novel random walk model to measure the users' influence. For measuring a user's influence, they take into account the follower network of the user and the popularity of the tweets. A method for measuring user influence is also proposed in (187). This paper presents TrueTop, the first Sybil-resilient system to measure the influence of Twitter users. ProfileRank, a random walk based method inspired by PageRank, is proposed in (145) to find influential users and relevant content.

Rather than finding influential users, (141) propose a method to find trendsetters in information networks. Trendsetters are different from other influential users in that they need not necessarily be famous but can spread a new idea over a social network successfully. While all of the above studies have considered Twitter as the experimental framework, (105) investigates the influence mechanisms in Pinterest social media platform. Another influence study conducted by (120), where the author propose the computation of the Influence Spectrum algorithm for seeking a set of influential people on several social networks such as NetHEPT, NetPHY, Epinions, DBLP and Twitter. Similarly, (167) define two influence maximization queries to track influential users over Twitter and Reddit datasets. Other than the community structure of a network, (163) propose a linear-time shell decomposition method based on the layer structure to maximize the influence in large scale networks. Their approach can explain the different behaviours of real networks and predict the saturation dynamics in the networks.

2.2.2 Social Networks Propagation Phenomena

Information Propagation: Information propagation in Twitter is indispensable for making a strategy to explicitly spread it. In this platform, original tweets are rapidly propagated through the retweeting mechanism. Zarrella (186) focused

2.2 Social Networks Influence and Information Propagation

mainly on direct content analysis of the retweets and the original tweets themselves, such as the most likely words to be retweeted, types of URL shortening services used, and reading grade level of the retweets. Sadikov and Martinez (140) examined URL and tag propagation on Twitter where they focus on external and network influences. Jansen et al. (82) analyzed more than 150,000 tweets to investigate Twitter as a form of electronic information propagation platform for sharing consumer opinions concerning brands. Suh et al. (150) represented retweeting as a powerful key mechanism disseminating information in the Twitter social network. Ye and Wu (182) conducted the propagation patterns of Michael Jackson's death news through Twitter and also evaluate different social influences correlate with each other. Cha et al. (25) investigated the dynamics of user influence across topics and time based on in-depth comparison of three measures (indegree of users, retweets, and mentions) of Twitter users. Stieglitz and Dang-Xuan (149) presented the different roles of sentiments in information propagation in political tweets. Bakshy et al. (10) reconstructed a cascade in Twitter and examine the relative role of strong and weak ties in information propagation of users. Another interesting approach capturing by Kim and Yoo (91), where they analyzed role of sentiment information propagation of political communication in Twitter and also differentiate user reply vs. retweet with respect to sentiment variables. Naruse (117) represented an information flow model for Twitter by the combination of single layered network where they estimate an information propagation range of a tweet. Weng et al. (172) employed agent-based model to analyze the competition of information diffusion of different tweets from Twitter. Toole et al. (115) present a novel model with geolocation information by adopting Twitter data. Myers et al. (2012) present a model in which information can reach a node through the influence of external sources. Another interesting approach capturing by Jin et al. (84), they demonstrated that their approach is accurate at capturing diffusion

2. STATE OF THE ART IN SENTIMENT ANALYSIS AND OPINION DYNAMICS

of news and rumors. Tang et al. (158) proposes a dynamic micro-blog network model which simulated the information dissemination in the artificial microblog with different scenes. Wu and Shen (176) identified the characteristics of news propagation from Twitter and built a news popularity prediction model to predict the final number of retweets of a news tweet very quickly.

Sentiment Propagation: Nowadays, sentiment propagation in social networks is very hot topic. Very few works has been done on sentiment propagation. Zafarani et al. (184) analyze sentiment propagation study in Livejournal where they provide methodologies for measuring kind of propagation. Wu et al. (177) conducted the study to extract the sentiment seeds by integrating several common sentiment dictionaries for sentiment propagation. For the purpose of understanding customers' opinion and subjectivity, Jung et al. (161) proposed a fuzzy propagation model for opinion mining by sentiment analysis on online social networks. Deng and Wiebe (45) applied Loopy Belief Propagation to propagate sentiments among entities combined with opinion inferences. Another interesting work proposed by Tang et al. (157), where authors provide a propagating process to incorporate various types of emotional signals in microblogging data into a coherent model. In order to determine the reputation polarity of factual information, Giachanou et al. (67) proposed two approaches that implement sentiment propagation from sentiment-bearing texts to factual texts.

2.3 Technique and Modelling for Opinion/Emotion Dynamics

Opinion dynamics is a psychological and sociological process which firstly studied by some statics communities. Recently, many researchers devoted their work on

2.3 Technique and Modelling for Opinion/Emotion Dynamics

investigating dynamics user influence and influence modeling on human interaction in social network. Most of the approach predict a user's current opinion under analytics of past history or influenced by neighbors' opinions.

2.3.1 Technique for Opinion/Emotion Dynamics

Statistical Technique: Conversation on Twitter became more commercial, and massive study performs on sentiment analysis for information propagation in a social network. The coupled HMM takes the advantages of HMM on tracking the dynamics of every single component, and it has the ability to capture the interactions between various components. It was first proposed by Brand et al. (21) for coupling and training hidden Markov models (HMMs) to model interacting processes and demonstrate their superiority to conventional HMMs in a vision task classifying two-handed actions. Coupled HMMs provide an efficient way to resolve many of these vision (and speech) applications problems and offer superior training speeds, model likelihoods, and robustness to initial conditions. Ren and Xu (134) present a new approach named Primitivebased Coupled-HMM for human natural context-dependent action recognition. Similarly, Natarajan and Nevatia (118) have proposed SemiMarkov models and coupled HMMs as suitable tools for handling a sub-event and directly encoding interactions among multiple agents, and the results could be integrated into the outdoor visual surveillance system. Following this idea, a simplified coupled-HMM influence model was theoretically studied by Asavathiratham et al. (9), and was employed for understanding the behaviours of many interacting components in a complex network, such as a communication network, transportation systems, and power grids.

Despite the coupled HMMs in modelling the interacting components within a system, many studies successfully proved that HMMs has the ability to detect sentiment or opinion mining from the textual data. For instance, Ho and Cao (76)

2. STATE OF THE ART IN SENTIMENT ANALYSIS AND OPINION DYNAMICS

propose a method using the high-order Hidden Markov Model whose states are automatically generated to model the process that a mental state sequence causes an emotion. Vanzo et al. (162) modelled the polarity detection problem as a sequential classification task over streams of tweets by formulating the Support Vector Machine discriminative model as embodied by the SVM^{hmm} algorithm that has been employed to assign the sentiment polarity to entire sequences. A modified version of self-adaptive HMM has been proposed by Liu et al. (108) where Particle Swarm Optimization algorithms optimize the parameters and classify the emotion on Microblogs. Jose and Chooralil (86) use an approach that automatically classifies tweets' sentiment related to party and politicians by a new combination of classifiers: SentiWordNet classifier, naive bayes classifier, and hidden markov model classifier. Interestingly, a study conducted by Kim et al. (92) where they propose an approach for sentiment analysis in microblogs that learns patterns of syntactic and sentimental word transitions. For this, the authors build HMMs by using similar syntactic and sentimental information groups called SIGs. Rustamovt et al. (139) proposed three machine learning algorithms and their combinations for the classification of movie reviews: Fuzzy Control System (FCS), Adaptive Neuro-Fuzzy Inference System (ANFIS), and Hidden Markov Model (HMM). With the facilitate of spreading information under the environment of wireless communication in microblog platform, Wang et al. (166) proposed a framework to calculate sentiment for aspects of events using some effective technologies in processing natural language, such as wordvec, HMM, and TextRank. Chen et al. (27) developed a novel temporal influence model (TIM) based on a Markov chain to track and understand users' opinion behaviours about a specific topic via their dynamic interactions on Twitter. Due to the popularity of e-commerce business on various products, customer reviews grow rapidly. By following the same path, Jin et al.

2.3 Technique and Modelling for Opinion/Emotion Dynamics

(85) propose a novel machine learning framework using lexicalized HMMs to extract, learn and classify product-related entities from product reviews. Similarly, Soni and Sharaff (146) conducted a sentiment analysis study of customer’s online reviews about the product, and they used a Stochastic Hidden Markov model to train the data and reveal the comment for analyzing consumer opinions about the Products on test data. A recent study conducted by Zhao and Ohsawa (190) on online-shopping reviews of Amazon Japan’s tea category and result shows the adapted hidden Markov model has the highest fl score among the other baseline methods.

Apart from sentiment analysis, opinion mining and emotion prediction on various topic, HMM also applied to predict stock market price (121), stock selection (122), insulin chart prediction for diabetic patients (119), fault diagnosis (104) etc. Users have different habits of using social media, and it isn’t easy to divide their posting records into sequences with fixed time intervals. Thus, the specific properties of the emotion behaviours should be considered for modelling emotion dynamics. We present Hidden Markov Model (HMM) to understand the nature of changing emotions expressed in Twitter messages and investigate how the sentiments are propagated across the users that take part of the conversation.

Machine Learning Technique: The recurrent neural network (RNN) was first proposed by Elman (53) to solve modelling sequences with arbitrary lengths. When the length of the sequence increases, RNN suffers from the vanishing or exploding gradient problem. Many variants of Recurrent Neural Networks (RNNs) (70), including Long-Short Term Memory (LSTM) (77) and Gated Recurrent Units (GRUs) (34), are proposed to overcome this problem. LSTMs are explicitly designed to avoid the long-term dependency problem. The LSTM can remove or add information to the cell state, carefully regulated by structures called gates,

2. STATE OF THE ART IN SENTIMENT ANALYSIS AND OPINION DYNAMICS

and it replaces the hidden unit with a memory block, which contains three gates, including the input, output, and forget gates provide the write, read and reset operations for the cells. The hidden units allow the memory cells to access and forget historical information. However, the architecture of LSTM is complex and improves the cost of training. In this circumstance, GRU (34) is proposed to balance the effectiveness and efficiency in sequence modelling by maintaining two gates, i.e., reset and update gates. The GRU is the newer generation of Recurrent Neural networks and is pretty similar to an LSTM.

In recent years, the application of neural networks such as RNNs (70), LSTM, Convolutional Neural Networks (CNNs) (97) and GRUs have been used widely in different applications, such as language modeling (14), sentiment classification (49, 135) and tracking opinion (30). A study proposed by Colnerič and Demšar (38), where they used large emotion-labelled data sets and investigated the transferability of the final hidden state representations between different classifications of emotion. Ghosal et al. (63) present Dialogue Graph Convolutional Network (DialogueGCN) using current RNN-based methods to deeply capture the inter-relationship utterances/emotions in a dialogue conversation. A recent study conducted by Fu et al. (58), where authors proposed audio sample augmentation and an emotion-oriented encoder-decoder to improve emotion recognition performance and discussed an inter-modality, decision-level fusion method based on a graph attention network (GAT). Another interesting study conducted by Jiao et al. (83), where propose an Attention Gated Hierarchical Memory Network (AGHMN) with a bidirectional GRU (BiGRU) to recognize real-time emotion in the conversations. However, the opinion influence mechanism over the emotional state sequence formation has not been explored properly.

2.3 Technique and Modelling for Opinion/Emotion Dynamics

2.3.2 Models for Opinion/Emotion Dynamics

Most of the approach predict a user's current opinion under analytics of past history or influenced by neighbors' opinions. Many researchers proposed several statistical physics based models or linear models and data-driven models or non-linear models. The most well known linear models is DeGroot model (43) which assumes update opinion of a user based on the averaged value of his/her neighbors' opinion. Flocking model with similar to DeGroot model proposed by Hegselmann and Krause (74), where they follow averaging strategy of the neighbors' opinion as a update opinion. By considering opinion randomly, Holley and Liggett (79) proposed voter model which allows a user to update his/her opinion by randomly choosing neighbor's opinion at each timestamp. Later on, a modification voter model proposed by (96), where a user adopts the majority opinion in his/her neighborhood. Another updated voter model called Biased Voter Model, where Das et al. (40) mainly focus on nature of stable equilibria and rate of convergence of user's opinion. All these models developed on theoretical assumption by capturing only real life phenomena; e.g. consensus or polarization not fine-grained real data. To fill this gap, De et al. (41) proposed a linear (not stochastic) asynchronous linear model (AsLM) where authors assumes realistic by estimating edge influence strength. Similarly, De et al. (42) proposed a nonlinear model called SLANT which represents users' opinions over time and parameter estimation from historical fine grained event data. Another extension model called SLANT+ proposed by Kulkarni et al. (98), which captured nonlinear interaction and predict next message time along with opinion. Apart from these nonlinear models, some researchers (50, 168, 179) also proposed nonlinear models on different contexts.

2. STATE OF THE ART IN SENTIMENT ANALYSIS AND OPINION DYNAMICS

2.3.3 Other Related Models

Recently, many researchers devoted their work on investigating dynamics user influence and influence modeling on human interaction in social network. For instance, Cha et al. investigate the dynamics of user influence across topics and time (25). Though some research suggests influence models (11, 126) incorporated with dynamical parameters to uncover the different states of the user. Interestingly, diffusion model such as Independent Cascade (IC) and Linear Threshold (LT) models also became very popular in viral marketing where parameters are learned by maximizing the likelihood of observation (69, 143). Similarly, different heuristics based methods applied to provide truly scalable solutions to the influence maximization problems (32, 88). Myers et al. represent a external influential model by linking with infected neighbors (115). Other literature also include epidemiology model of influence (169). Many influence models considered link structure and the temporal order by adopting information in Twitter (100, 101). All these existing models uncover the state transition between different points in time and measuring user influence on social network.

So, basically these value-based and content-based sequential models I have used as a baseline in order compare with our proposed model. In our propose model, we have integrated the concept of both value-based and content-based properties in order to obtain better result.

2.4 Conclusions

Human behavior in social networks has been widely analyzed over the last decade, particularly in the field of politics, games, products and reviews. The progress from a read-only to a read-write web allowed internet users to interact, share their opinions through social networks, blogs and wikis. Existing approach of sentiment

analysis has been classified by several way, for instance, affective valence of social media content as negative or positive, studies the subjectivity or objectivity of text, extracts the emotion and its arousal (112). However, it is still not much discover the inherit mechanism of how precisely how sentiments are form through a network and how take measures to control it.

Some of the recent works focus on opinion dynamics model and sentiment prediction by capturing many valuable features like internal state of the user. For instance, opinion model with dual identity proposed by Chen et al. (29), where they consider user’s personal identities and social identities for a better understanding of opinion behaviors. Similarly, Chen et al. (30) developed content-based sequential influence model which based on two prediction strategies, i.e. sentiment based and opinion based prediction strategies.

In this thesis, we analyze the sentiments or emotions on tweet messages and infer behavior patterns by studying influencing process of the users. We also focus on the learning of emotional influence through users opinion dynamic interactions.

3

Emotion Dynamics of Public Opinions

“Human behavior flows from three main sources: desire, emotion, and knowledge.”

— Plato

3.1 Introduction

Recently, social media has been considered the fastest medium of information broadcasting and sharing. Considering the wide ranges of applications such as viral marketing, political campaigns, social advertisement, etc., studying the influencing characteristics of users or tweets has attracted several researchers. It is observed from various studies that influential messages or users create a high impact on a social ecosystem. Tweets coming from the member community have higher influencing capability to others than the other sources. It is also observed that retweet influences users more than hashtag, mention and reply. People often

tend to rely on these platforms for retrieving information about topics of their interest and often make their decisions/opinions based on the acquired information. Social network sites are used in various tasks such as political campaigns (71, 149), social advertisement (103), social aspects of emotions (93), expert finding (125), viral marketing etc. (17, 31, 47, 144, 152) for influencing people. Different studies give credits to the success of Arab spring (175), Brazilian protests (39), Nirbhaya justice (4) etc. to social networking platforms. Social movements are believed to be highly influenced by social media sites, particularly in their organization and communication. *Do social media posts about current events, news, and sociopolitical debates influence people's opinions?* This one is the core questions that many of the studies on social media data analysis attempt to understand. Several studies (99, 170) observe that social activities and interactions greatly effect people's day-to-day activities, lifestyle, reading habit etc. In regards to political and social issues, public policies, studies show different observations. Based on the finding study (87), what people say or post on social media highly influences one's support of public policies. The same is also found to be true for political leaders while supporting or opposing a public policy. Study (25) also noted that the influence pattern is different for different countries and leaderships. However, *has social media activities on a topic or news story ever changed one's opinion on a political issue?* Study (192) analyzed the propagation of trust and distrust on social networks, which can be considered the first paper in which sentiment propagation was studied. Interesting conclusions, like that positive and negative sentiments, follow a different propagation pattern (75), have been drawn from the various investigations on social networks' sentiments. Other works studied the correlation between emotions and information diffusion, finding that those messages emotionally charged were re-tweeted more often (149), or investigated if the topic

3. EMOTION DYNAMICS OF PUBLIC OPINIONS

and the opinion of the user's contacts affect the own user's opinion (156). Motivated by the above studies, this study investigates the changing characteristics of people's opinion against an event on Twitter and how mass discussions/interactions influence changing one's opinion against an event. We use a user's emotion reflected in the post as the matrix to indicate his/her opinion in support/oppose of a social event.

This study focuses on the emotion dynamics of a user while posting comments against an event/topic from three different social characteristics: i) emotional (excitement, contentment, depression and distress), ii) community (follower followee, membership relationship), and iii) conversational (tweet, retweet, mention, reply). Identifying influential features can help us understand factors causing people to change their opinions and help agencies like advertisers and marketers design more effective campaigns. This chapter systematically explores the user's changing characteristics of emotion over time and attempts to find answer to the following three questions.

- Do people change their opinion against an event/issue over time?
- Which type of opinions against which type of events are more prone to change?
- If people change their opinion, which of the social factors cause them to change their opinion?

To investigate the above questions, we collect posts related to twelve various Twitter events. Emotional states of the users reflected in the post are determined for each tweet using Russell's model of affect, which correctly classifies the emotions expressed in over 90% of text messages (73). For each user against a topic/event, *Temporal Emotional State Chain* (TESC) is prepared, as mentioned on Section

3.1 Introduction

Table 3.1: Size of the all datasets

Categories	Events/Topics	Starting Date	Ending Date	Total # of Tweets	Total # of Sent Tweets	# of Users	Avg # of Sent Tweets
Policy	#BlackMoneyDebate	14-11-2016 09:10:56	27-11-2016 16:51:03	616343	15936	1260	12.6
	#Brexit	18-07-2016 13:29:26	24-07-2016 14:58:39	686434	17053	2688	6.3
Movie	#AlienCovenant	16-05-2017 14:20:16	20-05-2017 12:21:33	74286	4957	1504	3.3
	#Baahubali2	08-05-2017 21:44:17	10-05-2017 11:05:39	53391	2750	679	4.0
Sport	#BadmintonRio2016	13-08-2016 19:11:38	23-08-2016 12:24:54	66679	7413	784	9.4
	#UCLFinal	04-06-2017 07:32:05	11-06-2017 15:21:09	100547	7481	1996	3.7
Terror Attack	#SyriaGasAttack	06-04-2017 04:35:33	07-04-2017 04:19:38	10823	1477	557	2.6
	#StockholmAttack	07-04-2017 15:24:49	12-04-2017 18:13:00	2092	375	128	2.9
Accident	#GrenfellTower	15-06-2017 10:59:25	17-06-2017 23:34:56	136821	6499	2297	12.8
	#UnitedAirlinesAssault	10-04-2017 16:17:14	13-04-2017 15:43:57	7176	7481	583	2.7
Politics	#MacronPresident	08-05-2017 06:38:01	10-05-2017 22:57:53	6171	643	235	2.7
	#Trumpregrets	13-03-2017 17:16:29	23-03-2017 12:11:11	4884	768	201	3.8

3.2.2. All the experimental analysis are conducted over the collection of TESC's across different users and different events. From various experimental setups, this chapter makes the following contributory observations.

- We show that 63% of the user change their opinions and if an individual shares positive emotion against a topic, (s)he is likely to stay in the same emotional state in his/her subsequent tweets.
- Tweets coming from a member community have higher influential ability to an individual than the other sources like followers.
- Retweeted tweets can also influence a user higher than the tweets received through the hashtag, reply and mention.

3. EMOTION DYNAMICS OF PUBLIC OPINIONS

3.2 Data Preparation and Emotional Modeling

This section describes the experimental dataset used in this study. For our purpose of analyzing real-time events, we chronologically retrieved tweets through Twitter using *Search-API*, as describe on Appendices [A](#). We have collected tweets related to various events which contain a specific hashtag. Our objective is to analyze the sentiments derived within a conversation on Twitter and investigate how the emotions are changed dynamically across the users that take part in the conversation. The datasets included in our experimentation cover a wide range of topics. We tried to address a different type of topics to check whether the user’s emotion acts differently. This study considers twelve events/topics belonging to six different categories (1) Policy; (2) Movie; (3) Sports; (4) Terror Attack; (5) Accident and (6) Politics. Some of the hashtags representing the events/topics are manually identified. These Hashtags are further used to collect the related post. The collected tweets mainly contain the following information (i) user information, i.e., the user who posted the tweet (ii) tweet text (iii) type of tweet, i.e., direct tweet, retweet, reply, quoted tweet (iii) time of posting the tweet. To study opinion dynamics, one should post at least two tweets. We therefore first identify users who have posted at least two tweets against a topic. The dataset consists of about 17.65 million tweets and 69.36K number of users. Out of the total number of tweets, 72.83K number of tweets sent by 12.91K users and the rest of the tweets (i.e. 10.36 million) have been received by the same 12.91K users. Since our study focuses on the emotional dynamics of these users only, therefore we require users to have sent at least two tweets within a two-time frame. From the raw datasets, we filtered out users who sent 0 or only one message and tweets that did not show any emotion (null values of valence and arousal). It means, out of total users, 12.91K users have posted at least two number of tweets. Table [3.1](#) shown the final

3.2 Data Preparation and Emotional Modeling

figures after filtering. The description of the datasets are given below:

- *#Blackmoneydebate* comprises the tweets of the user are related to demonetization which started in India from November 10 till December 30, 2016.
- *#Brexit* tweets are about the referendum about the United Kingdom's withdrawal from the European Union held on June 23 2016.
- *#AlienCovenant* is an American science fiction horror film directed by Ridley Scott. It was released in the United States on May 19, 2017.
- *#Baahubali2* is an Indian historical fiction film was theatrically released over 9000 screens worldwide on 28 April 2017.
- *#BadmintonRio2016* comprises the tweets of the user about the final championship of badminton at Rio2016.
- *#UCLfinal* is about 2017 UEFA Champions League Final football tournament between Italian side Juventus and Spanish side Real Madrid which played at Millennium Stadium in Cardiff, Wales on 3 June 2017.
- *#SyriaGasAttack* is about a gas attack in north-western Syria where more than 80 people were killed on April 4, 2017. Survivors and aid workers shared their stories of horror and shock after a suspected chemical attack in Syria.
- *#StockholmAttacks* tweets are related to attack which happened in Stockholm, the capital of Sweden on April 7, 2017. A hijacked truck was deliberately driven into crowds and killed four people, including many more injured.
- *#GrenfellTower* is a 220-foot high tower block of public housing flats in North Kensington, west London. Collected tweets are about the Grenfell

3. EMOTION DYNAMICS OF PUBLIC OPINIONS

Tower fire on 14 June 2017 which caused at least 80 deaths and over 70 injuries.

- *#UnitedAirlinesAssault* is about an Asian passenger Dr. David Dao who was violently dragged off by security officers from an overbooked United Airlines flight on April 10, 2017.
- *#MacronPresident* is a French politician who won the second round of the presidential election on 7 May, 2017.
- *#Trumpregrets* is a conversation about those American citizens who voted for Trump and now regret their decision.

3.2.1 Data Pre-processing and Emotion Extraction

This section presents the overview of our sentiment/emotion classification model. For each participating user, we extract and arrange the tweets posted by the user in the order of posting time. As mentioned above, we assign an emotional state to each of the selected tweets to enable us to investigate a user’s emotion dynamics while participating in social discussions. The details of the data preparation are discussed below.

3.2.1.1 Russell’s Circumplex Model of Affect

A circumplex model is perfect when looking at different emotional states, and it can be demonstrated along with their respective relationships. We use the well known 16-state Russell’s circumflex model of affect (55) to estimate the state of emotion present in a given tweet [1]. In the Russell’s circumflex model of affect, emotions are understood as a combination of varying degrees of two main dimensions, *valence* (pleasure dimension) and *arousal* (activation dimension), which are distributed in

¹This is the extension of the original eight states Russell’s model (137)

3.2 Data Preparation and Emotional Modeling

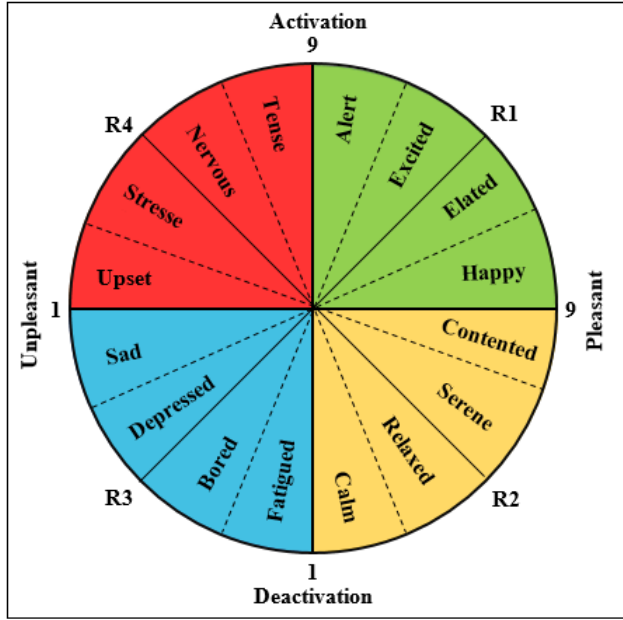


Figure 3.1: Russell's Model of Affect (55)

a 2D circular space (137). According to Russell's model, every affective experience is the consequence of a linear combination of valence and arousal dimensions (the so-called *core affect space*), which is then interpreted as representing a particular emotion. According to this model, every affective experience is defined by *valence* and *arousal* coordinate in the 2D circumplex shown in Figure 3.1. A numerical value for valence ranges from 1 (*unpleasant*) to 9 (*pleasant*) and arousal ranges from 1 (*deactivation*) to 9 (*activation*). The emotional state or sentiment label of a given entity (message or user) has been formed according to its valence (x -axis) and arousal (y -axis) values. Figure 3.2 shown on the right-hand side are the more pleasant states (+ve x); on the left-hand side the more unpleasant ones (-ve x) (22). The upper half shows the more activated states (+ve y), the lower half

3. EMOTION DYNAMICS OF PUBLIC OPINIONS

the more deactivated ones ($-ve\ y$). To reduce the number of emotional states, we identify four quadrants or regions defined in (137) and named them as *excitement* (R1), *contentment* (R2), *depression* (R3) and *distress* (R4). Excitement is a state with a high positive affect ($x + y$), while contentment is a state with a low negative affect ($x - y$). Similarly, distress is a state with high negative affect ($-x + y$) and depression is a state with low positive affect ($-x - y$) (55, 183). Particularly, it has been found evidence that many emotions lie on the perimeter of a circle and these are labeled with a particular name (*sad, unhappy, bored, stressed, etc.*) as can be observed in Figure 3.1.

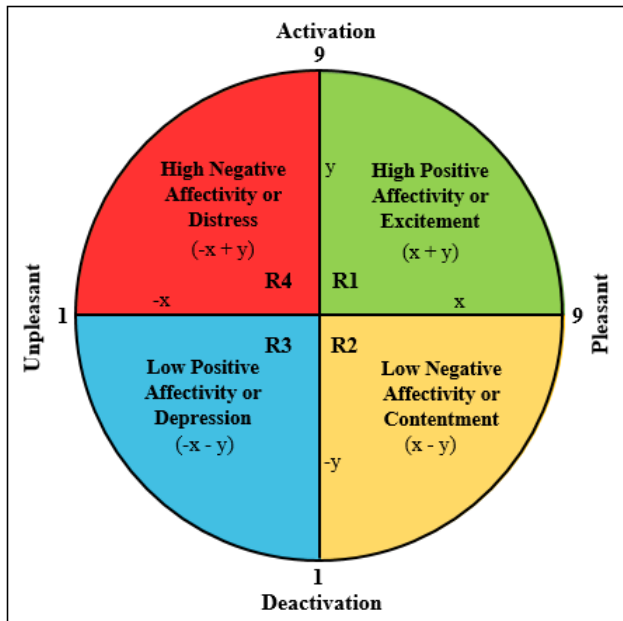


Figure 3.2: A Semantic Structure of Affect (22, 55). The letters x and y represent semantic components: x = Pleasantness; y = Activation

Assuming that affect can be modified by the degree of valence and arousal, it seems reasonable to assume that emotions have the potential to lie across all

3.2 Data Preparation and Emotional Modeling

positions in the two-dimensions rather than just on a perimeter (138). The core affect map in Figure 3.1 will be used to identify the emotional state or sentiment label(s) of a given entity (message, topic or user) according to its valence and arousal values. A particular entity will be associated with only one sentiment label when it falls onto the portion of the perimeter labelled with such a sentiment and it will be assigned two sentiment labels when it lies in a position outside the perimeter, particularly with the labels that result from the projection of the valence (X-axis) and arousal (Y-axis). For instance, an entity e with valence and arousal values $(v_e, a_e) = (7.26, 3.56)$ would fall within region R2, and the sentiment label associated is $S_e = \{serene\}$. In other cases, an entity will be assigned two sentiment labels. For instance, given the values $(v_e, a_e) = (6.08, 5.13)$ of an entity e , the point would graphically fall in a position placed between (*Happy*) and (*Excited*) and thus $S_e = \{happy, excited\}$.

Using this model enables us to define a more accurate sentiment of the messages and discriminate between cases in which the general sentiment is just positive or negative (for example, when a catastrophe occurs).

3.2.1.2 Finding Emotional State of a Tweet

To determine the emotional state of a tweet using Russell’s circumflex model, we first need to estimate the valence and arousal score of the tweet. The aim of sentiment extraction is to compile sentiment words. One of the most efficient approaches for this purpose is the dictionary-based approach. Dictionary-based approaches use dictionaries of emotional words which are associated with a sentiment score. To estimate the valence and arousal score of a tweet, we use the ANEW dictionary of affect (20). The new version of the Affective Norms for English Words (ANEW) dictionary (123) is being developed to provide the mean and standard deviation of normative emotional ratings (valence v and arousal a)

3. EMOTION DYNAMICS OF PUBLIC OPINIONS

for 2477 unique words in English. The word list of ANEW initiated from a set of obscene words as well as a few positive words. Later on, different slang words were included, such as WTF, LOL and ROFL. The entries of this dictionary match by applying Porter word stemming and WordNet lemmatization. It also contains another less strongly-related dimension was called dominance. However, for the purpose of our experiment, we only concentrate on two primary dimensions.

The dictionary-based approach’s performance can be evaluated according to two aspects: 1) the number of emotional words covered by the dictionary and 2) the nature of the sentiment score provided by the dictionary. ANEW computes sentiment score with the word’s valence and arousal values, which range from 1 to 9. ANEW allows us to calculate a more accurate sentiment value that fits better our aim of having a bi-dimensional representation of sentiments and measuring the intensity of expressed emotions.

3.2.1.3 Calculating Sentiment Scores

The average sentiment score of a message is calculated with the stem words’ valence and arousal that appear in the ANEW dictionary. Following, we show an example of a tweet message that comprises three emotional words in the ANEW dictionary. **Hazirah Affah @AzieFifa (Fri Aug 19 19:11:27 CEST 2016): ”Good job guys!!!! We Malaysian re so proud!!!! #MalaysiaBoleh #badmintonRio2016”**.

- **Good**, $v = [\mu : 7.47, \sigma : 1.45], a = [\mu : 5.43, \sigma : 2.85]$
- **job**, $v = [\mu : 5.83, \sigma : 2.15], a = [\mu : 5.20, \sigma : 2.23]$
- **proud**, $v = [\mu : 8.03, \sigma : 1.56], a = [\mu : 5.56, \sigma : 3.01]$

The aim of this phase is to associate each entity e with a tuple (v_e, a_e) . Specifically, we calculate the sentiment score of a tweet message and then the user’s

3.2 Data Preparation and Emotional Modeling

score based on his/her tweet messages. The average sentiment score of an entity e is calculated with the stem words of e that appear in the ANEW dictionary, i.e., the emotional words of S_e , the sentiment label associated with e . In order to combine the mean values of the valence and arousal of the words within S_e , we have to assume that individual mean values reported for each stem form a normal distribution. Supposedly, if a stem has a high σ of valence (equivalently for arousal), then the valence ratings of the word are distributed over a wider range of values; and lower values of σ imply that ratings are closer to μ . Thus, we used a probability weight based on each word’s probability density function in ANEW to estimate that the stem’s valence (arousal) falls exactly at the mean.

$$X = \frac{\sum_{i=1}^N \frac{\mu_i}{\sigma_i}}{\sum_{i=1}^N \frac{1}{\sigma_i}} \quad (X, \mu, \sigma) \quad (3.1)$$

Table 3.2: Notation of different attributes

X	mean value of valence (Y , mean value of arousal)
N	total number of emotional words within the message
μ	word’s mean value of valence (equivalently for arousal)
σ	word’s standard deviation of valence (equivalently for arousal)

Formula (3.1) calculates the sentiment score of a message by estimating the overall mean value of all emotional words within the message (see Table 3.2). Particularly, X is the mean value of valence (similarly, Y mean value of arousal), N is the total number of emotional words within the message, μ is the word’s mean value of valence (equivalently for arousal) and σ is the word’s standard deviation of valence (equivalently for arousal).

The sentiment score of a user is calculated as the average emotional value of all the tweets sent by the user. For example, if we amalgamate the three words

3. EMOTION DYNAMICS OF PUBLIC OPINIONS

Good, job and **proud** of the above message d , the result of the weighted average formula (3.1) for the valence and arousal is $X_d = 7.30$ and $Y_d = 5.41$, respectively.

3.2.1.4 Overview of Sentiment Analysis Model

Several approaches for extracting emotion from tweet messages exist in the literature. In (113), authors create a large lexicon from tweets with sentiment-word hashtags added by tweeters corresponding to the six basic Ekman emotions (52). Using emotion-related hashtags to identify the topic of the message has been the predominant choice to create emotion-labelled datasets from tweet messages (35, 130, 131). Here, we aim to extract meaningful emotions in a *single-hashtag* tweet collection against different topic or event.

The overview of the sentiment extraction model is shown in Figure 3.3. It includes tweet processing mechanisms, calculating sentiment score and sentiment extraction tool for classification of sentiments. We consider two types of entities: tweets or messages and users, all associated with an opinion orientation expressed either with an individual sentiment or a set of sentiments represented through valence and arousal values.

For our purpose of analyzing real-time events, we chronologically retrieved as many tweets through the Twitter *Search API*. We used *hashtags* for collecting tweets that several users post based on a particular domain (conversation), and we built a corpus of D tweets. The first step is to have a word representation for each message to facilitate machine manipulation and eliminate noisy words. Processing each individual tweet message involves extracting bags of words, filtering all stop words and extracting stems by using the stemmer tool of the WordNet dictionary (Figure 3.3, top). As a result, we obtain the vocabulary for our particular dataset composed of V stem words. After obtaining the stem words, we have performed

3.2 Data Preparation and Emotional Modeling

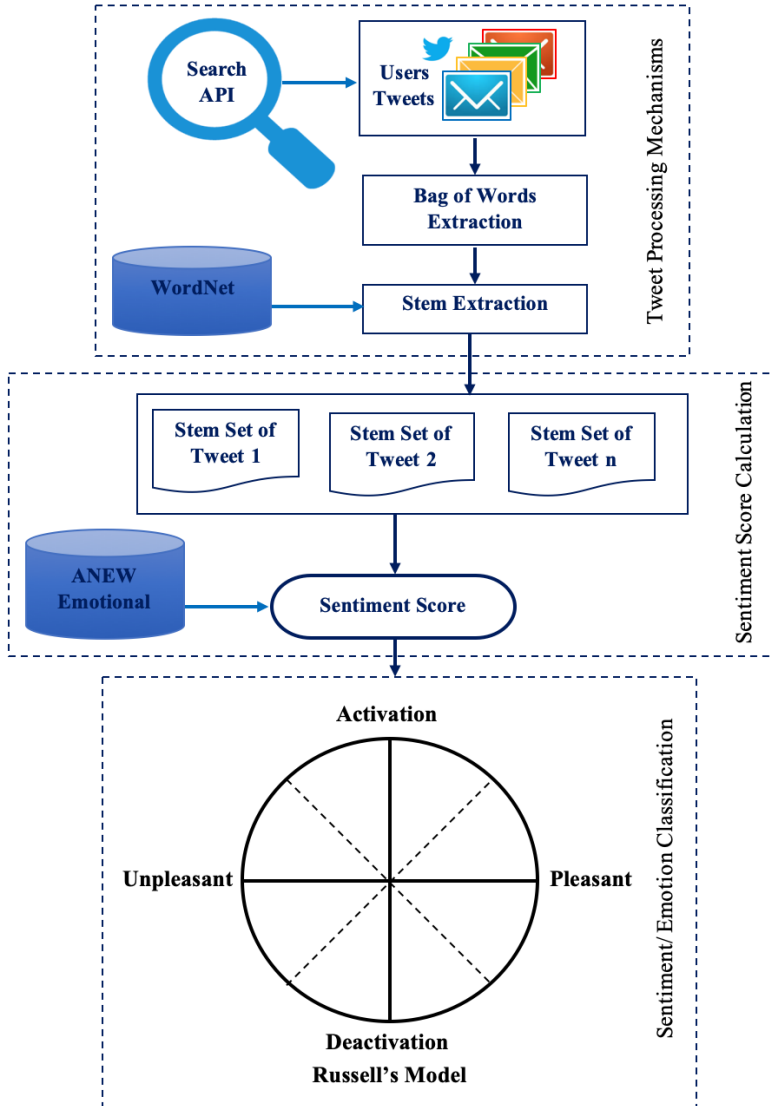


Figure 3.3: Global Overview of Sentiment Analysis in Tweets

3. EMOTION DYNAMICS OF PUBLIC OPINIONS

a sentiment score calculation. The emotional role of the stem sets of tweets is obtained through the emotional base ANEW dictionary (Figure 3.3 middle), and words are annotated by their valence and arousal values (see section 3.2.1.2). Finally, words are then classified according to Russell’s emotion model (Figure 3.3 bottom).

Formally, let’s consider a given tweet $d \in D$, an entity that is tokenized and filtered by eliminating stop words to a bag of words w_d . We transform w_d to a stem set, s_d , through the WordNet stemmer tool, which allows us to get the appropriate stem for each token. The sentiment words associated with s_d are spotted based on the ANEW dictionary, where emotional words are annotated by their valence and arousal values (examples of such emotional words are agreement, love, sad, quite, disagree, etc.). Finally, the user’s overall opinion orientation or emotional tweet status is determined by combining the content of each emotional word identified in the tweets sent by the user and classified according to Russell’s model.

3.2.2 Temporal Emotional State Chain

This section describes formation of a *Temporal Emotional State Chain* (TESC). Our presumption is that a person’s emotion depends on his/her personal opinion and the past history s/he received from neighbors by following different emotions, popularity and characteristics of messages. The Temporal Emotional State Chain is the sequence of *outgoing* and *incoming* tweets within two time frame. Every participating user’s data is defined by the sequence of *outgoing* and *incoming* tweets arranged in the order of posting time. Outgoing tweets are those tweets posted by the target user about the topic under consideration. Whereas the incoming tweets are those tweets posted by other users about target topic, and are received by the user through one of the following : hashtag, mention, reply, retweet, following-list,

3.2 Data Preparation and Emotional Modeling

member-list, other-list. We use the following terminologies to describe the chain formation.

- *User (u)*: User u is the current user for which we want to analyse the temporal emotional state chain.
- *Incoming mention (InM)*: If a tweet is posted by another user with $@u$, then this tweet is an *incoming mention (InM)* tweet to user u .
- *Incoming retweet ($InRT$)*: A quoted retweet with $@u$ is referred to as *incoming retweet ($InRT$)* for user u .
- *Incoming reply (InR)*: An *incoming reply (InR)* to an user u is direct reply to u 's post.
- *Incoming hashtag (InH)*: A tweet bearing the same hashtag as that of u 's tweet is the *incoming hashtag (InH)* for u .
- *Incoming member-list ($InML$)*: Any user u of a group post a tweet and correspondingly another user post another tweet from the same group, we refer to this tweet as coming from *incoming member-list ($InML$)*.
- *Incoming following-list ($InFL$)*: The user u who is following someone and list of tweets bearing the same hashtag seen by followee (u who is being followed), we refer to these tweets as coming from *incoming following-list ($InFL$)*.
- *Incoming other-list ($InOL$)*: Any user u who is mentioned by someone but not associated with followee or member list, we refer to as a *incoming other-list ($InOL$)*.

In TESC, we wanted to register the message that causes the reaction of a user and, at the same time, whether the user reacts to this message or not. Consequently, we needed to know the message received by the user and the possible

3. EMOTION DYNAMICS OF PUBLIC OPINIONS

message sent by the user. More specifically, we consider the time at which tweets were posted. Given a direct graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where each vertex $u \in \mathcal{V}$ represents a user and each edge $(u, v) \in \mathcal{E}$ represents a user u following to another user v . The each user $u \in \mathcal{V}$ post a tweet message m with sentiment region R at time t . Taken into account \mathcal{G} , we considered a collection of all external set for each user $u \in \mathcal{V}$, which denoted as a $\mathcal{Z}_u = \{v | (u, v) \in \mathcal{E}\}$. The total number of users is \mathcal{N} and size of the neighbor set \mathcal{Z}_u is $n(u)$.

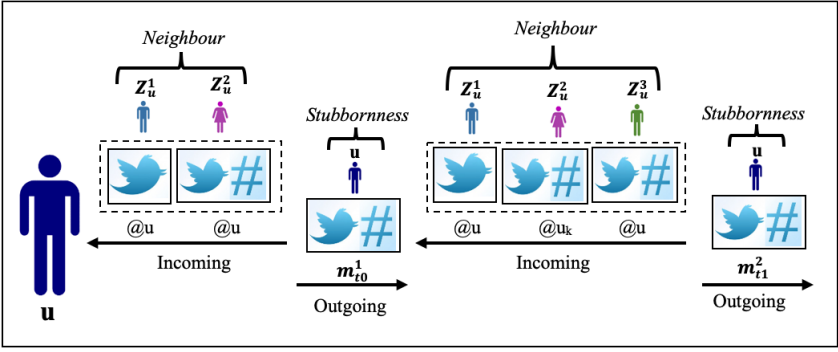


Figure 3.4: Temporal Tweet Chain

Given a user u , we estimate the incoming tweets posted by v that the user u has received between two consecutive tweets (i.e. m^1 and m^2). The time of the two consecutive tweets sent by a user u is referred to as t_0 and t_1 . Additionally, we also include the incoming tweets posted by v before posting u 's first tweet at time t_0 . Given a user u and a topic $\#h$, as we construct a typical temporal tweet chain (as shown in Figure 3.4) or tuple chain of u 's posted a message at each timestamp $t_u(i)$, where \downarrow denotes (incoming) tweets and \uparrow denotes the (outgoing) tweets.

$$\langle u, \#h \rangle \rightarrow \langle \dots, \downarrow m_{t_0-1}^{in} \rangle, \uparrow m_{t_0}^{out}, \langle \downarrow m_{t_0+1}^{in}, \dots \rangle, \uparrow m_{t_1}^{out}, \langle \downarrow m_{t_1+1}^{in}, \dots \rangle, \uparrow m_{t_2}^{out}, \dots$$

3.2 Data Preparation and Emotional Modeling

where $C^{in} \in \{inH, inRT, inR, inM, inFL, inML, inOL\}$. When the user u posts his first tweet at t_0 on topic $\#h$, public discussion on the topic $\#h$ might have already taken place. It is denoted by the tuple $\downarrow m_{t_0-i}^{C^{in}}, i = 1, 2, \dots$ and $C^{in} \in \{inH, inRT, inR, inM, inFL, inML, inOL\}$. Similarly, incoming tweets between the user's tweet $\uparrow m_{t_k}$ and $\uparrow m_{t_{k+1}}$, is denoted by the tuple $\langle \downarrow m_{t_{k+0}}^c, \downarrow m_{t_{k+1}}^c, \dots \rangle$. An example is also shown in Figure 3.4

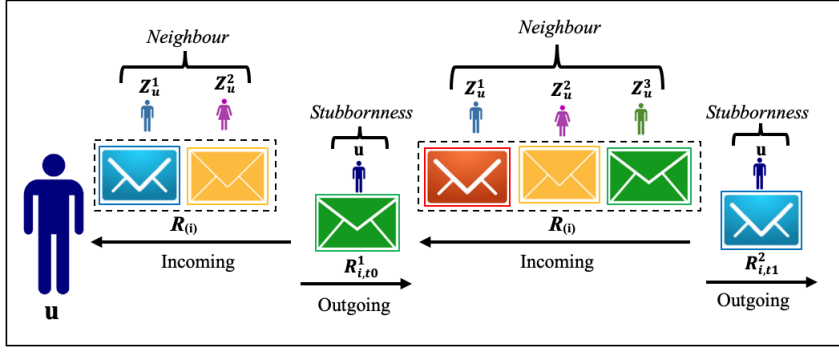


Figure 3.5: Temporal Emotional State Chain

The emotional state of a tweet in a *temporal tweet chain* is determined using Russell's circumplex model of affect as described in Section 3.2.1.4. If $\mathcal{R}_i, i \in 1, 2, 3, 4$ denotes one of the four emotional states for a given tweet, the above *temporal tweet chain* can be transformed into the following *temporal emotional state chain*.

$$\langle u, \#h \rangle \rightarrow \langle \dots, \downarrow \mathcal{R}_{i,t_0-1}^{C^{in}} \rangle, \uparrow \mathcal{R}_{i,t_0}^{C^{out}}, \langle \downarrow \mathcal{R}_{i,t_0+1}^{C^{in}}, \dots \rangle \uparrow \mathcal{R}_{i,t_1}^{C^{out}}, \langle \downarrow \mathcal{R}_{i,t_1+1}^{C^{in}}, \dots \rangle, \uparrow \mathcal{R}_{i,t_2}^{C^{out}}, \dots$$

All the incoming set of emotion tweets from neighbor are denoted by the tuple

3. EMOTION DYNAMICS OF PUBLIC OPINIONS

$\langle \downarrow \mathcal{R}_{t_{k+0}}^{C^{in}}, \downarrow \mathcal{R}_{t_{k+1}}^{C^{in}}, \dots \rangle$ and fall between the user's personal emotion tweets $\uparrow \mathcal{R}_{t_k}^{C^{out}}$ and $\uparrow \mathcal{R}_{t_{k+1}}^{C^{out}}$.

Finally, our assumptions of user's emotion is based on past information s(he) receives and his/her personal emotions. Therefore, our objective is to predict u 's future emotion state $\mathcal{S}_u(i)$ at the next timestamp t_u based on personal emotion sequence $\uparrow \mathcal{P}R_{u,i}^{out}(t) = \langle \uparrow \mathcal{R}_{i,t_0}^{C^{out}} \dots, \uparrow \mathcal{R}_{i,t_k}^{C^{out}} \rangle$ sent by the user u at t and all the neighboring emotion of the messages $\downarrow \mathcal{N}R_{u,i}^{in}(t) = \langle \downarrow \mathcal{R}_{i,t_{0-1}}^{C^{in}} \dots, \downarrow \mathcal{R}_{i,t_k}^{C^{in}} \rangle$ received from $t - 1$ to t .

3.3 Emotion Transition on Twitter

This section analyzes the characteristics of a user's emotional state transition in his/her subsequent tweets against a topic. We focus on analyzing the following characteristics: (i) transition probability of users emotional state change in subsequent tweets, (ii) likely initial state of user's emotion while posting a tweet against a topic, (iii) relationship between user's emotional state and nature of the topic, and (iv) participation of the user into the network conversation and their frequency while transiting from one state to another.

To perform the user's state transition, a probabilistic sequence model, i.e. the Markov model, is adopted. The simplest Markov model is the Markov chain (7). According to the Markov model, the next state is solely chosen based on the current state. The transition probabilities control how the hidden state at time t is chosen given the hidden state at time $t - 1$. The set of transition probabilities for transitions from any given state must sum to 1. In this study, users' emotion states are likewise labelled with Russell's regions $S = R1, R2, R3$ and $R4$. The process starts in one of these states and moves successively from one state to another. Each move is called a step. If the chain is currently in state s_i , then it moves to

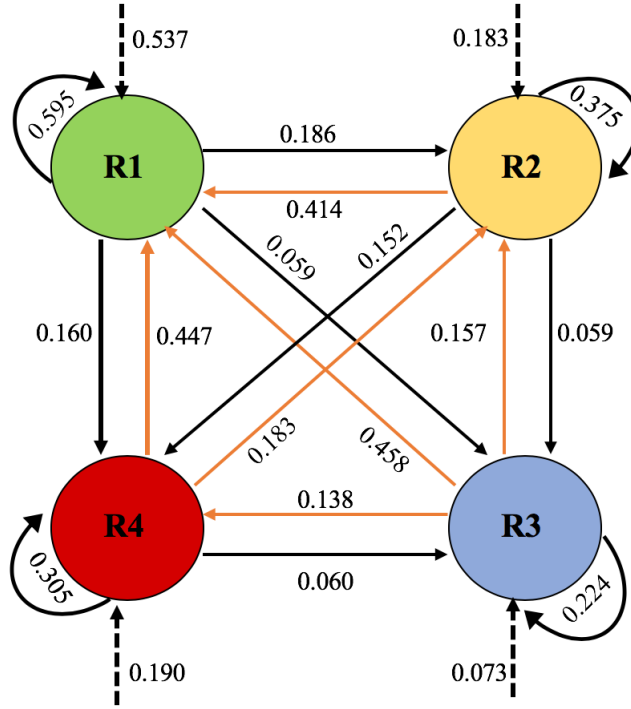


Figure 3.6: Macro average transition probabilities over all topics. Initial state probability is represented by dot arrow.

state s_j at the next step with a probability denoted by p_{ij} , and this probability does not depend upon which states the chain was in before the current state. The probabilities p_{ij} are called transition probabilities. The process can remain in the state it is in, and this occurs with probability p_{ii} . Let $\langle R1 \rangle$ is a current emotion state of the tweets sent by the user at t , and then next move can be towards $\langle R2 \rangle$, $\langle R3 \rangle$ or $\langle R4 \rangle$ or this transition can remain in the $\langle R1 \rangle$. However, the emotional state of the user can start with any of these four states.

Figure [3.6](#) shows the average transition probability from one emotional state to

3. EMOTION DYNAMICS OF PUBLIC OPINIONS

another over the topics in different categories. First, Figure 3.6 shows the macro average transition probabilities over all topics. It is clearly evident that if a user is in a state with high positive emotion $R1$, the probability of staying in the same state in the subsequent tweets from the same user is higher than that of the highly negative emotion state $R4$ (with probability 0.60 for staying in $R1$ and 0.31 for staying in $R4$). Further, it can also be seen that if a user makes a transition from one state to another, the user is more likely to move toward the state with highly positive emotion as compared to other states (on average probability 0.43, 0.17, 0.5 and 0.15 towards $R1$, $R2$, $R3$ and $R4$, respectively). Interestingly, $R1$ has got the highest initial transition probability. It means when a random user posts his/her opinion on a random topic, s(he) is likely to start with a highly positive emotional state. However, for the topics like *Terror Attack*, the observations deviate from the above-average pattern.

The above observation may be biased by the nature of the topics/events that we consider in the experimental dataset. Users' emotional states may depend on the nature of the topic under consideration. To understand the topic dependent characteristics, we further investigate topic-wise (Figure 3.7 and 3.8) transition probability as follows.

- For the majority of the topics except in *Terror Attack* category, like in average case, a user in a highly positive emotional state ($R1$) is likely to continue in the same state with higher probability (with probability more than 0.5) as compared to that of the highly negative emotional state ($R4$).
- The topics in *Terror Attack* category show slightly different characteristics where probabilities of a user staying in the highly positive emotion state ($R1$) and highly negative emotion state ($R4$) are comparable. The probability of staying in $R4$ is even slightly higher than that of $R1$.

3.3 Emotion Transition on Twitter

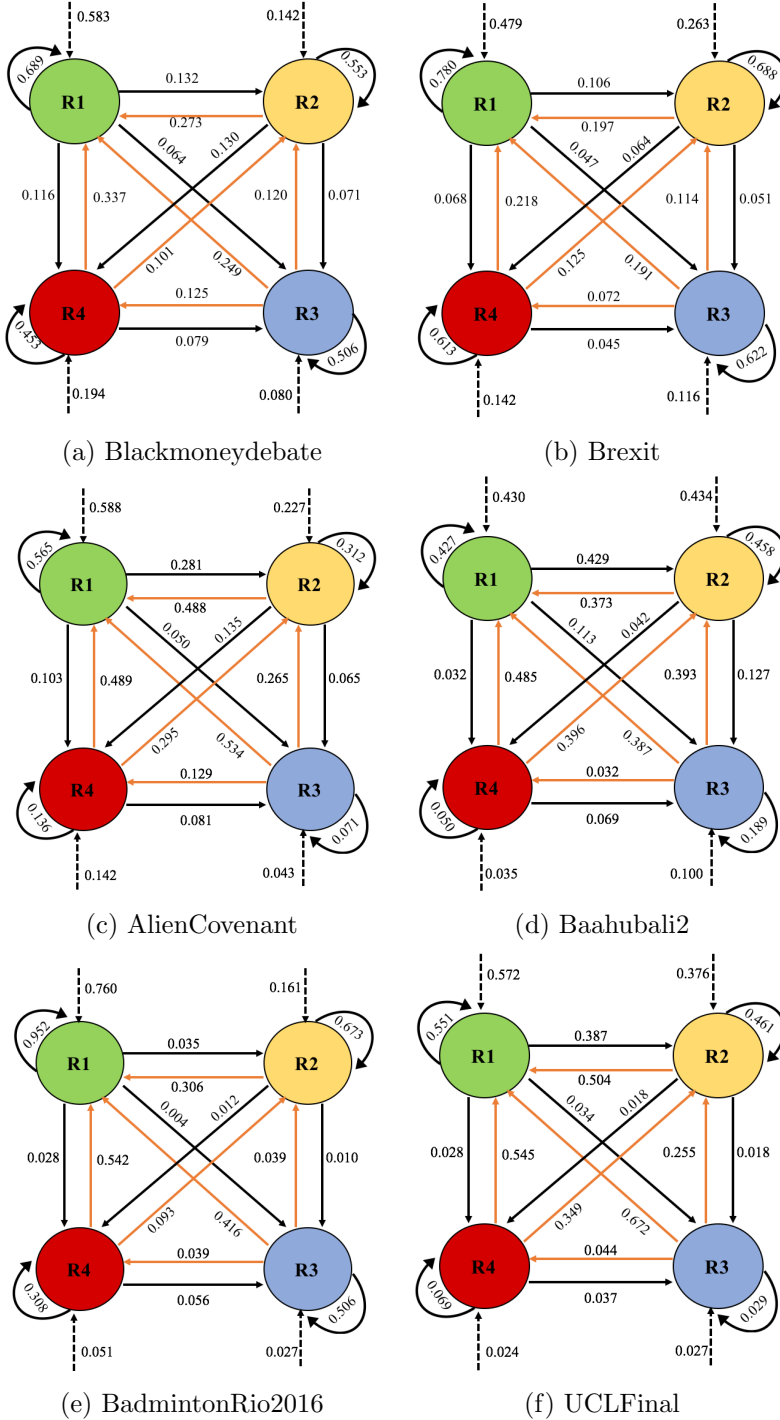


Figure 3.7: Transition probability of a user over topics in different categories (Part-1).

3. EMOTION DYNAMICS OF PUBLIC OPINIONS

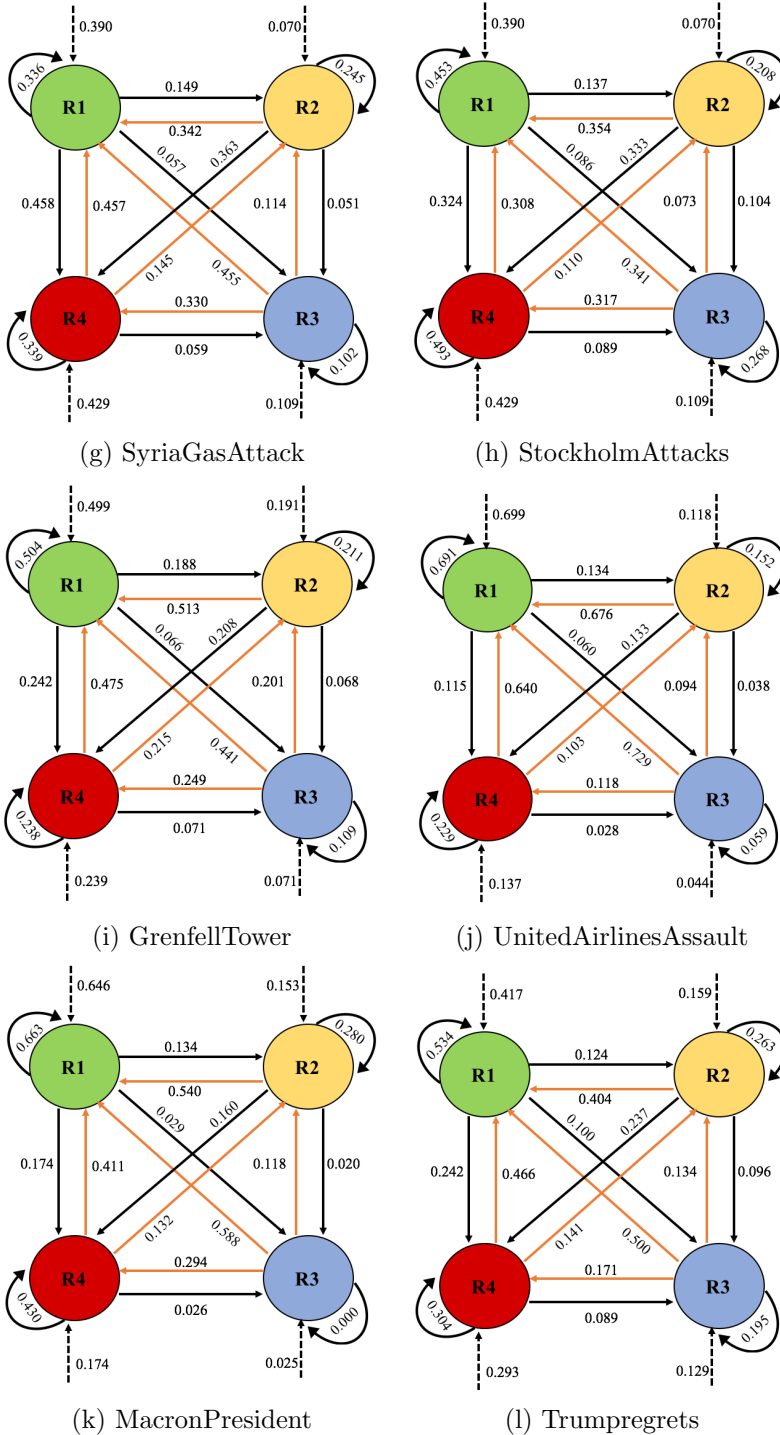


Figure 3.8: Transition probability of a user over topics in different categories (Part-2)

3.3 Emotion Transition on Twitter

Table 3.3: Agreement and disagreement by all topics with average transition in Figure 3.6. P_{11} considers self transition with high positive emotion, i.e., $R1$, P_{ij} considers transition from self state to other states, $P(\rightarrow i)$ considers from others state to $R1$ and π_i considers initial state.

Dataset	Category	P_{11}	P_{ij}	$P(\rightarrow i)$	π_i
#Blackmoneydebate	Policy	✓	×	✓	✓
#Brexit	Policy	✓	×	✓	✓
#AlienCovenant	Movie	✓	✓	✓	✓
#Baahubali2	Movie	×	×	✓	×
#BadmintonRio2016	Sport	✓	×	✓	✓
#UCLFinal	Sport	✓	✓	✓	✓
#SyriaGasAttack	TA	×	×	✓	×
#StockholmAttack	TA	×	×	✓	×
#GrenfellTower	Accident	✓	✓	✓	✓
#UnitedAirlinesAssault	Accident	✓	✓	✓	✓
#MacronPresident	Politics	✓	×	✓	✓
#Trumpregrets	Politics	✓	✓	✓	✓

- Another interesting observation for the topics related to a terror attack is that when users make state transitions, the probabilities of moving towards both $R1$ and $R4$ are also comparable.
- For the topics related to *Policy*, in the majority of the cases, users continue to stay in the earlier state (with a probability higher than 0.5, users continue to take self transition.)
- In the majority of the cases, users in states $R2$ and $R3$ are more prone to transition to other states than that of $R1$ and $R4$. Further, users in $R3$ are more prone to change state than the users in other states.
- Unlike other topics, for the topics (*#StockholmAttack*, *#SyriaGasAttack*) *Terror Attack* category, the initial probability is quite high for region $R4$.

3. EMOTION DYNAMICS OF PUBLIC OPINIONS

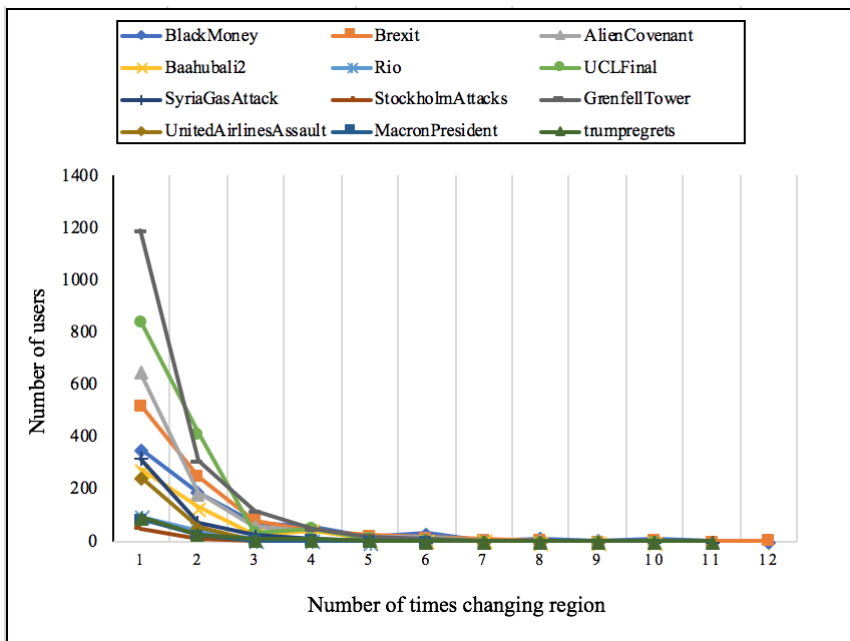


Figure 3.9: Frequency of the user changing emotion region across topics

Table 3.3 summarizes agreement and disagreement of all the topics with that of the observation on average. It clearly shows that except for topics related to terror attacks, the majority of the cases agree with the average observations over all topics. Further, Figure 3.9 shows the number of times the users have changed their emotional state. We see that, for all the topics, the majority of the users change their emotion state against a topic only once. Comparatively, a very small number of users change their emotional state more than once for the same topic. Further, it is observed that 63% of users change his/her emotion state at least once against a topic. Users in emotion state $R3$ has the highest likelihood of changing state with a probability of 32%.

3.4 Understanding Influence of Incoming Tweets

On Twitter, users receive messages from other users through various channels like mention, reply, member-list, follower-list, etc. From the studies (25), it is noted that the user’s opinion is often influenced by the incoming messages (s)he receives. This section attempts to understand the influential characteristics of different channels over the user’s opinion by estimating the probability of the incoming emotion state coming through a channel agreeing with the user emotion state present in his post. For example, given that the emotion state $R1$ of the user and the majority of the tweets received by the users, what is the likelihood that the user is carrying the majority on an emotion state $R1$. In this section, we examine three types of social features; (i) incoming tweets with emotion state (i.e., $R1$, $R2$, $R3$ and $R4$), (ii) incoming tweets only from community (i.e., $InFL$, $InML$ and $InOL$), and (iii) incoming tweets only through conversation (InM , $InRT$, InR and InH). From the first features, we attempt to understand, if majority opinion can influence an individual’s opinion. The second features try to understand if an individual’s opinion can be biased by the opinions coming from the group/community that (s)he belongs to. Lastly, we investigate if responses from the general public on previous posts of an individual influence the individual. We systematically explore the above questions by exploiting the dataset we prepare in section 3.1.

3.4.1 Can Majority Opinions Influence an Individual’s Opinion?

To answer this, we examine the distribution of the emotion states over all the incoming tweets that an individual receives before posting his next tweet and checks if the emotion state in his post agrees with any of the incoming tweets’ emotions. To systematically investigate the influential characteristics of the incoming

3. EMOTION DYNAMICS OF PUBLIC OPINIONS

Table 3.4: It shows the probability of an individual’s emotion state in a post not matching with the emotion state of any of the incoming message

Emotion State	BlackMoneyDebate	Brexit	AlienCovenant	Baahubali2	BadmintonRio2016	UCLFinal	SyriaGasAttack	StockholmAttacks	GrenfellTower	UnitedAirlinesAssault	MacronPresident	Trumpregrets	Avg. Across Topics
R1	0.464	0.510	0.422	0.600	0.194	0.397	0.664	0.466	0.538	0.237	0.351	0.594	0.477
R2	0.830	0.758	0.668	0.565	0.813	0.639	0.824	0.883	0.790	0.707	0.845	0.725	0.760
R3	0.903	0.888	0.898	0.819	0.974	0.976	0.970	0.755	0.917	0.918	0.925	0.858	0.896
R4	0.793	0.833	0.890	0.949	0.937	0.977	0.483	0.570	0.740	0.804	0.651	0.760	0.798
Avg. Across Emotion State	0.672	0.669	0.553	0.623	0.246	0.521	0.639	0.566	0.667	0.332	0.489	0.683	0.645

message, we intend to understand the following sub-questions?

When an individual posts an opinion, how likely does his emotion state agree with that of any of the incoming messages? To answer this question, whenever a user posts a tweet, we check how many cases his emotion state matches the emotion state of the incoming tweets he receives. Table 3.4 shows the probability of an individual’s emotional state when he posts a tweet not matching the incoming message’s emotion state across different topics. When an individual posts a tweet, if the emotional state of his tweet matches the emotion state of any of the incoming tweet, we refer to it as **Matching**, otherwise **UnMatching**. The last row of the Table 3.4 shows the percentage of **UnMatching** posts for each topic. It clearly shows that except for two out of twelve topics, the percentage of **UnMatching** posts is higher than that of the **Matching** posts. Across all topics, 64.5% of the cases, user posts do not agree with any of the incoming tweets he

3.4 Understanding Influence of Incoming Tweets

receives. It indicates that the majority of the user’s opinion does not depend on the incoming opinion.

In rows ($R1$, $R2$, $R3$, and $R4$), it further shows the probability of **UnMatching**, if user posts a tweet with emotion state $R_i, i = 1, 2, 3, 4$ i.e., $Pr(UnMatching|R_i) = \frac{Pr(R_i|UnMatching)Pr(UnMatching)}{Pr(R_i)}$. It shows that user has higher chances of **Matching** when he posts a message with positive emotion (i.e., across all topics, the average of $Pr(UnMatching|R_1)$ is smaller than that of $R2, R3$, and $R4$). Similarly, chances of **UnMatching** is higher when he posts with low positive or high negative emotion state ($R3$ or $R4$). These observations are true for majority of the topics.

If individual’s emotion state matches with incoming emotion, does (s)he agree with majority? To answer this question, we further estimate the probability of user emotion state matching with the majority emotion state of the tweets that user receives. Table 3.5(a) shows the probability of matching with the majority emotion state. In the table, M_i denotes i^{th} majority emotion states i.e., 1^{st} , 2^{nd} , 3^{rd} , and 4^{th} majority. An entry in R_i row and M_j column in Table 3.5(a) is the probability that user posts a tweet with emotion R_i and the emotion state of M_j is also R_i i.e., $Pr(M_j = R_i|e(\uparrow m) = R_i)$ where $e(\uparrow m)$ is the emotion state of the user’s outgoing tweet m .

The row with *total average* in Table 3.5 and 3.6 column (a) shows the probability of agreeing with $M_i, i = 1, 2, 3, 4$ if the user’s outgoing post has an agreement with an incoming tweet over all topics across different emotional state. It shows that probability of agreeing with M_1 is higher than that of M_2, M_3 and M_4 . It means that if the user’s emotional state in the tweet that he posted has an agreement with the emotional state of some of the incoming tweet that he received, it is likely to agree with the dominant emotional state among all the tweets he receives.

3. EMOTION DYNAMICS OF PUBLIC OPINIONS

Table 3.5: The probability of an individual’s emotion state in a post matching with the emotion state of any of the incoming message

Dataset (#h)	Emotion State	(a) Agreement with Majority					(b) Community Channels			(c) Conversational Channels			
		Pr(R _i)	M1	M2	M3	M4	ML	FL	OL	RT	H	R	M
(1) BlackMoneyDebate	R1	0.445	0.883	0.095	0.022	0.001	0.555	0.441	0.004	0.877	0.079	0.022	0.022
	R2	0.217	0.300	0.423	0.075	0.040	0.481	0.515	0.004	0.901	0.068	0.014	0.017
	R3	0.161	0.154	0.290	0.381	0.018	0.481	0.518	0.002	0.796	0.027	0.005	0.173
	R4	0.177	0.058	0.347	0.417	0.044	0.541	0.457	0.002	0.868	0.090	0.018	0.024
	Avg.		0.690	0.169	0.089	0.051	0.542	0.454	0.004	0.874	0.076	0.020	0.030
(2) Brexit	R1	0.465	0.964	0.033	0.003	0.000	0.845	0.153	0.003	0.707	0.177	0.056	0.060
	R2	0.267	0.122	0.816	0.057	0.006	0.802	0.195	0.003	0.728	0.169	0.039	0.064
	R3	0.116	0.073	0.052	0.249	0.626	0.814	0.181	0.005	0.711	0.157	0.064	0.067
	R4	0.153	0.042	0.124	0.686	0.148	0.870	0.127	0.003	0.663	0.203	0.057	0.077
	Avg.		0.694	0.193	0.076	0.037	0.837	0.160	0.003	0.708	0.177	0.053	0.063
(3) AlienCovenant	R1	0.593	0.973	0.027	0.000	0.000	0.935	0.035	0.030	0.767	0.166	0.015	0.052
	R2	0.269	0.171	0.816	0.012	0.001	0.823	0.052	0.125	0.814	0.128	0.010	0.047
	R3	0.043	0.027	0.036	0.149	0.788	0.939	0.039	0.021	0.252	0.055	0.140	0.553
	R4	0.095	0.038	0.165	0.772	0.026	0.895	0.025	0.080	0.799	0.149	0.005	0.048
	Avg.		0.781	0.188	0.022	0.008	0.911	0.038	0.050	0.767	0.156	0.016	0.061
(4) Baahubali2	R1	0.447	0.769	0.229	0.002	0.000	0.953	0.038	0.008	0.889	0.073	0.012	0.027
	R2	0.400	0.501	0.495	0.003	0.000	0.925	0.045	0.030	0.851	0.115	0.006	0.027
	R3	0.127	0.128	0.108	0.753	0.011	0.826	0.047	0.127	0.852	0.129	0.002	0.017
	R4	0.027	0.019	0.038	0.192	0.750	0.817	0.087	0.096	0.750	0.163	0.048	0.038
	Avg.		0.604	0.344	0.049	0.003	0.932	0.042	0.026	0.869	0.096	0.009	0.027
(5) BadmintonRio2016	R1	0.918	0.996	0.004	0.000	0.000	0.595	0.391	0.014	0.947	0.025	0.002	0.026
	R2	0.075	0.190	0.805	0.006	0.000	0.622	0.359	0.020	0.907	0.063	0.001	0.028
	R3	0.003	0.000	0.500	0.500	0.000	0.500	0.000	0.500	0.250	0.250	0.500	0.000
	R4	0.004	0.250	0.750	0.000	0.000	0.417	0.333	0.250	0.750	0.000	0.083	0.167
	Avg.		0.981	0.019	0.000	0.000	0.595	0.391	0.014	0.946	0.026	0.002	0.026
(6) UCLFinal	R1	0.563	0.972	0.028	0.000	0.000	0.975	0.022	0.004	0.943	0.040	0.003	0.014
	R2	0.383	0.118	0.882	0.000	0.000	0.982	0.016	0.002	0.974	0.021	0.000	0.004
	R3	0.031	0.000	0.148	0.796	0.056	0.926	0.056	0.019	0.963	0.000	0.037	0.000
	R4	0.023	0.053	0.079	0.711	0.158	0.816	0.132	0.053	0.947	0.053	0.000	0.000
	Avg.		0.723	0.275	0.002	0.000	0.977	0.020	0.003	0.952	0.034	0.002	0.011
(7) SyriaGasAttack	R1	0.432	0.864	0.129	0.007	0.000	0.873	0.057	0.070	0.872	0.088	0.016	0.023
	R2	0.149	0.263	0.165	0.557	0.015	0.805	0.063	0.132	0.910	0.057	0.012	0.021
	R3	0.055	0.286	0.095	0.095	0.524	0.619	0.095	0.286	0.952	0.000	0.048	0.000
	R4	0.365	0.458	0.534	0.007	0.001	0.862	0.103	0.034	0.890	0.076	0.018	0.016
	Avg.		0.606	0.343	0.047	0.004	0.861	0.082	0.057	0.884	0.079	0.017	0.019
(8) StockholmAttacks	R1	0.491	0.890	0.079	0.031	0.000	0.567	0.009	0.424	0.842	0.113	0.023	0.022
	R2	0.073	0.542	0.083	0.125	0.250	0.708	0.042	0.250	0.625	0.292	0.042	0.042
	R3	0.120	0.844	0.000	0.065	0.091	0.083	0.455	0.117	0.909	0.039	0.026	0.026
	R4	0.310	0.438	0.556	0.006	0.000	0.651	0.009	0.340	0.683	0.228	0.052	0.037
	Avg.		0.741	0.220	0.027	0.011	0.588	0.017	0.395	0.793	0.147	0.033	0.027
(9) GrenfellTower	R1	0.486	0.991	0.009	0.001	0.000	0.935	0.046	0.018	0.873	0.087	0.016	0.024
	R2	0.186	0.018	0.282	0.690	0.010	0.947	0.046	0.007	0.786	0.146	0.019	0.048
	R3	0.091	0.012	0.022	0.127	0.839	0.904	0.076	0.020	0.857	0.076	0.025	0.042
	R4	0.237	0.025	0.910	0.064	0.001	0.946	0.047	0.006	0.862	0.095	0.013	0.030
	Avg.		0.676	0.208	0.096	0.021	0.938	0.047	0.015	0.860	0.095	0.016	0.029

3.4 Understanding Influence of Incoming Tweets

Table 3.6: Continued

(10) UnitedAirlinesAssault	R1	0.721	0.999	0.001	0.000	0.000	0.672	0.014	0.314	0.723	0.160	0.038	0.079
	R2	0.138	0.819	0.118	0.063	0.000	0.173	0.007	0.819	0.889	0.048	0.044	0.018
	R3	0.029	0.000	0.125	0.125	0.750	1.00	0.000	0.000	0.375	0.563	0.000	0.063
	R4	0.112	0.388	0.374	0.238	0.000	0.687	0.034	0.279	0.707	0.177	0.041	0.075
	Avg.		0.983	0.010	0.005	0.003	0.641	0.014	0.345	0.732	0.155	0.038	0.075
(11) MacronPresident	R1	0.655	0.985	0.015	0.000	0.000	0.876	0.022	0.102	0.795	0.155	0.016	0.034
	R2	0.122	0.161	0.452	0.387	0.000	0.790	0.032	0.177	0.782	0.177	0.032	0.008
	R3	0.039	0.474	0.000	0.368	0.158	0.474	0.053	0.474	0.842	0.105	0.053	0.000
	R4	0.184	0.327	0.604	0.066	0.002	0.780	0.033	0.187	0.687	0.130	0.156	0.026
	Avg.		0.869	0.105	0.025	0.001	0.858	0.024	0.118	0.782	0.152	0.035	0.032
(12) Trumpregrets	R1	0.485	0.932	0.064	0.003	0.000	0.753	0.081	0.166	0.842	0.228	0.052	0.037
	R2	0.171	0.279	0.358	0.257	0.106	0.690	0.097	0.212	0.625	0.292	0.042	0.042
	R3	0.094	0.109	0.266	0.281	0.344	0.734	0.078	0.188	0.909	0.039	0.026	0.026
	R4	0.251	0.298	0.540	0.131	0.031	0.796	0.097	0.107	0.683	0.228	0.052	0.037
	Avg.		0.681	0.207	0.077	0.036	0.751	0.087	0.163	0.793	0.147	0.033	0.027
Total Average	R1	(0.483)	0.941	0.051	0.008	0.000	0.766	0.222	0.012	0.814	0.118	0.030	0.039
	R2	(0.242)	0.193	0.657	0.105	0.046	0.772	0.205	0.023	0.809	0.125	0.023	0.043
	R3	(0.117)	0.106	0.154	0.323	0.417	0.680	0.305	0.014	0.747	0.084	0.034	0.135
	R4	(0.158)	0.061	0.429	0.397	0.113	0.764	0.226	0.010	0.802	0.128	0.029	0.042
	Avg.		0.711	0.188	0.070	0.032	0.764	0.222	0.013	0.810	0.118	0.029	0.043

Further, the $Pr(R_i)$ column in Table 3.5 and 3.6(a) shows the probability of a user posting a comment/tweet with the emotion state R_i . It shows that in majority of the cases user posts tweets with positive emotion R_1 irrespective of the topics. Overall, 48.3% percent of the tweets are in R_1 . Interestingly, even for the events like (7) *SyriaGasAttack* or (8) *StockholmAttacks*, majority of the tweets are with positive emotion i.e., 43.2% for (7) *SyriaGasAttack* and 49.1% for *StockholmAttacks*.

For different emotion states, it is observed that when a user posts a tweet with $R1$, it mostly agrees with the dominant emotion in the incoming tweets, i.e., M_1 . However, this is not the case when a user posts a tweet with other emotion states (other than $R1$). For example, when users post a tweet with $R3$ emotion state, it agrees mostly with either M_3 or M_4 . It indicates that a significant number of users do not get influenced by what he receives.

3. EMOTION DYNAMICS OF PUBLIC OPINIONS

Remarks Coming back to our earlier question, i.e., *can majority opinions influence an individual's opinion?* From the above observations in Table 3.5 and 3.6(a), i.e., the emotional state of a significant number of posts do not agree with the dominant emotion, which indicates that the influential ability of social propaganda on the social network is questionable.

3.4.2 Which Community Channel is More Influential?

In the above section, we have considered all the incoming messages irrespective of the channels through which a user receives the messages. However, as indicated in the studies (25), different channels may have a different influential pattern. In this section, we investigate the influential characteristics of different channels over individual opinion. Table 3.5 and 3.6(b) shows the probability of agreement with the dominant emotion state in each of the community channel namely *member-list (ML)*, *following-list (FL)* and *other-list (OL)*. An entry in R_i row and ML column in Table 3.5 and 3.6(b) is the probability that user posts a tweet with emotion R_i and majority of the tweets coming from ML have emotion state R_i , i.e., percentage of agreement with the dominant state in ML . Similarly, entries at FL and OL represent the percentage of agreement with their respective dominant state.

It is evident from Table 3.5 and 3.6(b), ML has the highest probability of agreement as compared to FL and OL for all the emotion state $R_i, i = 1, 2, 3, 4$. It means that message coming from ML has a higher potential for influencing users than the tweets coming from FL and OL . From the rows with *total average* in Table 3.5 and 3.6(b), it is observed that ML has 76% of agreements, whereas FL has only 22%. An interesting observation is that, though ML dominates FL for almost all the topics, the topics like *BlackMoneyDebate*, *Brexit* and *Badminton-Rio2016* have comparable distribution between ML and FL . This is due to the

3.5 Characteristic of Incoming Tweets towards State Transition

fact that popular tweets propagate multiple times from the source by retweeting throughout the network. A study (169) indicates that retweeting is a powerful mechanism in the social network where a group of users re-post the same tweet.

3.4.3 Which Conversational Channel is More Influential?

In the previous section, it is observed that tweets coming from ML has a higher potential of influencing a user's opinion. Further, tweets might be coming through various conversational channels like hashtags (H), retweet (RT), mention (M) and reply (R). In this section, we further investigate the influential characteristics of these conversational channels. Table 3.5 and 3.6(c) shows the probability of agreement with the dominant emotional state in each of the above conversational channel. An entry in R_i row and conversational channel in Table 3.5 and 3.6(c) is the probability that a user posts a tweet with emotion R_i and the majority of the tweets coming from the channel also have emotion state R_i .

In almost all the cases over all the topics, RT dominants H , R and M . It means that tweets coming from RT have a higher potential for influencing user's opinion than the tweets coming from others (H , R , M). Similar observations have also been reported in the study (150). Among H , R and M , H has higher potential. From the row with *total average* in Table 3.5 and 3.6(c), it is observed that RT contributes in 81% of the agreements, H contributes in 12%, and rest from M and R .

3.5 Characteristic of Incoming Tweets towards State Transition

In section 3.3 we observe that 63% of users change their opinions towards a topic at least once. This section investigates the distribution of the incoming tweets

3. EMOTION DYNAMICS OF PUBLIC OPINIONS

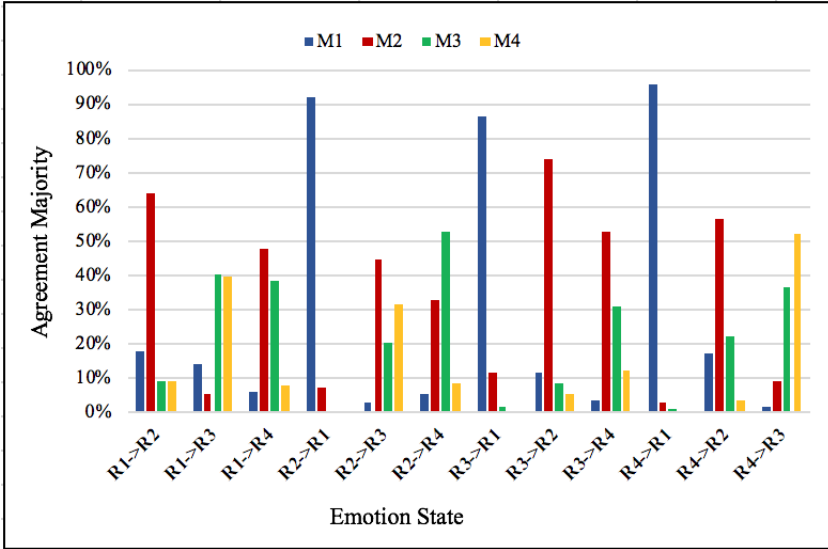


Figure 3.10: Average probabilities of emotional agreement over all topics

that an individual receives at the time of changing his/her emotional state while posting a tweet against a topic. It will help us to understand possible causal influence from different channels through which an individual receives incoming tweets. Like in section [3.4](#), this section also considers the same three types of cases; *whole incoming tweets*, *community channels* and *conversational channels*.

3.5.1 State Transition vs. Incoming Dominant Emotion

Why does an individual change his opinion from his/her earlier emotion state against a topic? Has it been influenced by the emotional state of the majority opinion? To answer these questions, we study the distribution of the emotional state of all the incoming tweets that an individual receives before posting his/her next tweet with a different emotion state (emotion state different from his/her previous tweet). Figure [3.10](#) shows the distribution of the emotional state of the

3.5 Characteristic of Incoming Tweets towards State Transition

incoming tweets across different dominant emotion states for each possible state transition.

In Figure 3.10, the emotion transitions of a user is represented as $R_i \rightarrow R_j$ (where $i, j = 1, 2, 3, 4$). For each $R_i \rightarrow R_j$ transition, we have shown its agreement with the distribution of the incoming emotion state. A bar chart with M_k (where $k = 1^{st}, 2^{nd}, 3^{rd},$ and 4^{th}) against a state transition $R_i \rightarrow R_j$ represents the number of the cases (in percentage) in which the emotion state changes from R_i to R_j , and emotion state of the k^{th} majority incoming tweet is R_j .

A user's agreement with 1^{st} majority i.e., M_1 at the time of state transition (change in emotion state of the user i.e., $R_i \rightarrow R_j, i \neq j$) potentially reflects possible influenced of changing state from the dominant incoming emotion. Similarly, a disagreement with the dominant emotion state at the time of changing the state of a user may also indicate a user's ability to make his/her own opinion (not biased by the dominant incoming emotion). Figure 3.10 shows two interesting observations. Whenever users change their emotions state to extreme positive from any other state, it always agrees with the majority (M_1), i.e., $R_i \rightarrow R_1, i \neq 1$. However, for any other changes where $R_i \rightarrow R_j, j \neq 1$, users agreement with M_1 is very low. It potentially means that sharing positive emotion is more general, and sharing negative emotion is more personal. It is interesting to see that transition to extreme negative has agreement mostly with 2^{nd} and 3^{rd} dominant emotions, not with the majority. Figure 3.10 shows that when users change their emotion state from $R4 \rightarrow R1$ and $R1 \rightarrow R4$, the percentage of agreements with dominant incoming emotion are 96% and 6% respectively. We can also see that the percentage of agreement with dominant incoming emotion while the transition from low positive $R3$ to high positive $R1$ is higher than that of $R1$ to $R3$ (i.e. 87% and 14%).

3. EMOTION DYNAMICS OF PUBLIC OPINIONS

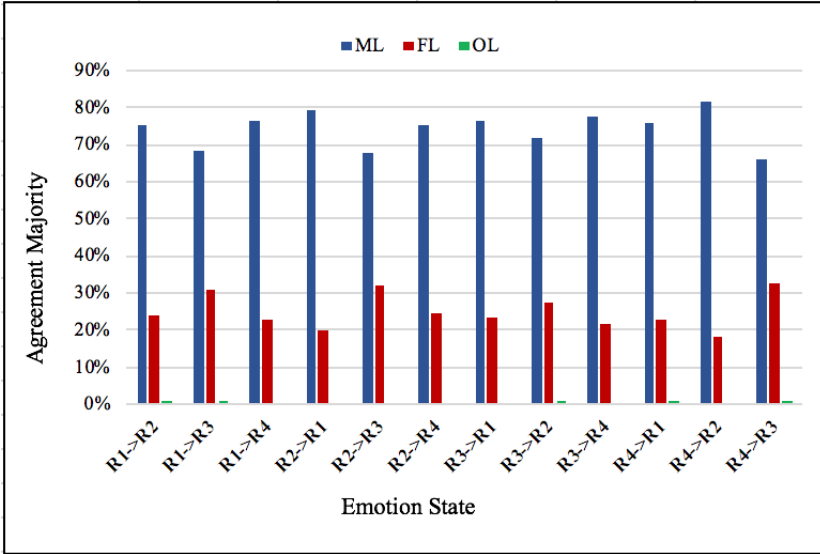


Figure 3.11: Average probabilities of topological agreement over all topics

From the above observations, it is evident that when a user changes his/her emotional state to non-positive states (i.e., $R_j, j \neq 1$), the user tends to share personal opinion. *How large is this proportion?* In our dataset, 40% of the state transitions do not agree with M_1 , and 41% of these transitions belong to a non-positive state transition. Further, *how likely a user who had an agreement with the majority tends to change his/her state?* It is observed from the dataset that in 2% of the cases a user tends to change his/her state from agreement to disagreement with the M_1 i.e., $R_i \rightarrow R_j, R_i = M_1, R_j \neq M_1$. It shows a significant proportion of the users are not influenced by incoming emotions.

3.5 Characteristic of Incoming Tweets towards State Transition

3.5.2 Which Community Channel is More Influential during Emotion State Transition?

In the previous section, we have investigated over entire incoming tweets irrespective of the channels through which they receive the incoming tweets. In this section, we study the emotion distribution of the incoming tweets with respect to community channels, i.e., member list, following, and others. Like in section 3.5.1, we estimate the probability of user emotion state matching with a different channel through (s)he receives during emotion transition. We calculate the total average probability of agreeing with the different channels over all topics in different categories. In Figure 3.11, the emotion transition refers to $R_i \rightarrow R_j$, where $i, j = 1, 2, 3, 4$ and for each $R_i \rightarrow R_j$ transitions, we have shown its agreement with majority for each community channel i.e., *ML* for member list, *FL* for following and *OL* for others. Each bar corresponding to a channel for a given transition $R_i \rightarrow R_j$ represents the percentage of its emotional agreement with the dominant emotional state in each channel, i.e., and the dominant emotion state is R_j .

Figure 3.11 shows that if an individual changes their emotional state from one region to another, the average majority agreement of the member-list channel is always higher than other channels (i.e. 74%). It is true for all the $R_1 \rightarrow R_j, i \neq j$ pairs. From this observation, it is evident that *ML* has maximum contribution in causing the state change. Interestingly emotion agreement with tweets coming from other (*OL*) channels is negligible. It shows that users are mostly influenced by tweets coming from either member list and following. Between member-list and following-list, member-list significantly dominates the following-list.

3. EMOTION DYNAMICS OF PUBLIC OPINIONS

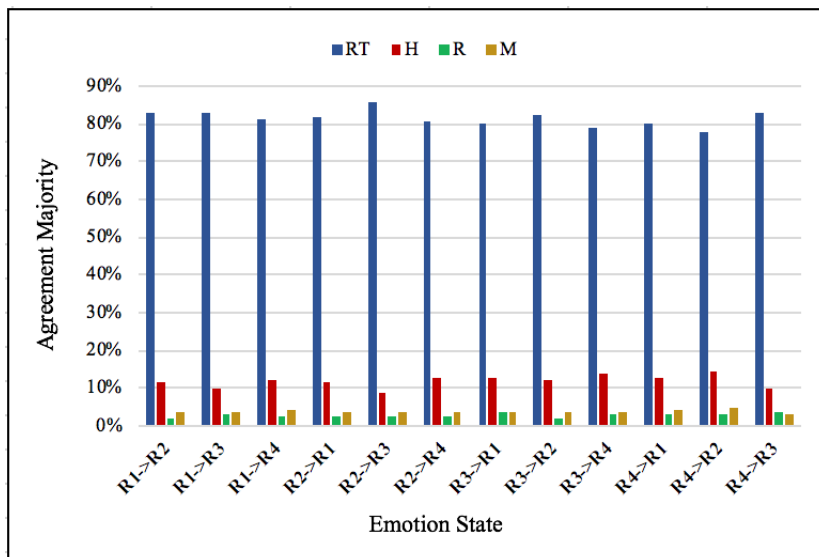


Figure 3.12: Average probabilities of conversational agreement over all topics

3.5.3 Which Conversational Characteristics are More Influential during Emotion State Transition?

Individuals also receive tweets through conversational characteristics like *Re-tweet*, *Hashtag*, *Reply* and *Mention*. Which one of these is more influential in causing state transition? To acknowledge the above question, we investigate of distribution of dominant emotion states over these channels in Figure 3.12. We calculate the total average probability of agreeing with different conversational characteristics over all topics in different categories. Like above, each bar in Figure 3.12 shows the percentage of the instances of agreement with the majority in each characteristic against different possible state transition $R_i \rightarrow R_j, i \neq j$. It is evident from the figure that if an individual changes their emotional state from one region to another, then the retweet contributes the most with about 81% on average over

all transitions. Further, Hashtags contribute about 10%. The contribution from the reply and mention are negligible.

3.6 Conclusions

In this study, we analyze the emotion dynamics of users while posting public opinion through Twitter. This study has considered a dataset consisting of 17.65 million tweets with about 69.36k users over 12 various topics. From this dataset, the emotion dynamics of 12.91k users who have posted at least two tweets have been studied. Analysis has been investigated from three different perspectives; the user's emotion transition, the influence of the emotion of public opinion on an individual's opinion, social parameters causing a change in an individual's opinion. First, we observe that 63% of the user change their opinion against a topic. People who share positive emotion against topics are likely to stay in the same emotional state in his/her subsequent tweets. Users in the highly negative state have the highest probability of changing state. If users change their emotional state, the probability of changing towards the highly positive state is higher than the probability of changing towards the highly negative state. It is observed that tweets coming from the member community have higher influential capability than the tweets coming from the follower community and other sources. Further, it is also observed that retweeted tweets can influence users higher than hashtag, reply and mention.

In order to investigate systematically, we perform three different experiments over several categorical social topics/events on Twitter. First, we perform emotion state transition over the entire conversation. Focusing on user emotion transition helps others find someone's indisputable interest on the topic in the future. From the transition, we infer that people have shown positive opinion into their conver-

3. EMOTION DYNAMICS OF PUBLIC OPINIONS

sation except *terrorism attack* topics. We also infer that people at the initial stage show high positive emotion while continuing their conversation, most people going towards high positive direction. Second, we perform an analysis to understand the influential characteristics of different channels over the user's opinion by estimating the incoming emotion state coming through a channel. From this analysis, we infer that the emotional state of a significant number of posts does not influence the dominant emotion on social networks. Finally, we investigate the influential characteristics towards state transition by approaching three types of cases (whole incoming tweets, community channels and conversational characteristics). From this investigation, we infer that most people agree with positive emotion when users change their emotional state from negative to positive. As a result, it is significant that the majority people of state transition potentially reflects possible influenced by changing state from the dominant incoming emotion.

4

Sequential Influence Model for Emotion Dynamics using Hidden Markov Model

"Public sentiment is everything. With public sentiment, nothing can fail. Without it, nothing can succeed."

— Abraham Lincoln

4.1 Introduction

Twitter Sentiment Analysis (TSA) is a currently hot research topic that embraces a large variety of tasks such as sentiment classification or opinion mining (65, 127), opinion dynamics (30), emotion measure (33) etc. Exploring this mechanism of users' emotion dynamics towards social events and further predicting their future emotions have attracted significant attention to the researchers. The last few years have evidenced a massive growth of sentiment propagation on Twitter. This

4. SEQUENTIAL INFLUENCE MODEL FOR EMOTION DYNAMICS USING HIDDEN MARKOV MODEL

propagation of sentiment into networks has spread on many different areas from professional to everyday life like viral marketing (144), political campaigns (149) and social advertising (103) and so on. Several studies (99, 170) observe that social activities and interactions greatly effect people's day-to-day activities, lifestyle, reading habit etc. Moreover, the accessibility to the information provided by social media brings up a whole new class of sentiment analysis tasks related to the changing nature of sentiments over time. People's opinion towards public events or products may change over time, rather than staying in the same state. Even tracking users' dynamics opinion (30) helps a company monitor critical feedback of the product and further adjust its marketing plans. A government can also utilize users' feedback about the new policy and forecast upcoming development (188). Therefore, understanding the essential mechanism of emotion dynamics and further predicting an individual's future emotion is of great importance.

In general, tracking emotion over time in Twitter has been used to predict events by finding a correlation between the events' sentiments and events. Many researchers mainly concentrated on social aspects of emotion (93), opinion dynamics (98) and predicting user opinion (133) from observed tweets. Social dynamics not even restricted only to the Twitter conversation, an important study performed by (44), where they proposed a new technique that combines automatic topic extraction and sentiment analysis of Brexit debate discussion Facebook public pages. Some interesting studies have been conducted on modelling and predicting opinion dynamics behaviours over time in social network (30, 42). One of the unexplored components of human communication found online in written form is an emotional expression. Despite the concreteness of the online expressions in written form, it remains unpredictable which kinds of emotions will be expressed in individual messages of Twitter users. However, a user's opinion prediction on a specific topic is a more difficult task and demand in different application scenarios. For example, an

individual political opinion may help to predict the proportion of potential Bised voter during the election time. Motivated by these observations on the effectiveness of using the social network in a different context, this chapter proposed a mathematical model to predict user's emotion dynamics on various social issues on Twitter. We adopted HMM to understand how user's spread their sentiment from one message to another. The HMM is defined as a system visualized as an expected Markov process with hidden states in which the system can directly produce an output that must return to its current state before moving on to the next state. As part of the essentials of HMMs, we will be concerned with producing an HMM models with the highest probable outcomes of the evident and observable transitions of state sentiments from the user's messages. The HMM application will understand the most likely sequence occurring of sentiments of a particular user that will transition from one state of emotions to the next based on a history of tweets with emotions contained in sent and received messages. We especially interest in predicting the user's emotion expressed in their written tweet and attempts to understand the changing of user's emotion over time. We aim to systematically attempts to answer the following research questions:

- Given a set of tweets related to some events, how do we optimize the model parameters for learning?
- Can we infer how emotion will be distributed from the observed sentiment-labelled tweets?
- How can we evaluate a user's emotional state on different events or topics?

To investigate the above questions, we collected our dataset from Twitter using *Search-API*¹ related to various events, as mentioned in Section 3.1. Emotional

¹<http://twitter4j.org>

4. SEQUENTIAL INFLUENCE MODEL FOR EMOTION DYNAMICS USING HIDDEN MARKOV MODEL

states of the users reflected in the post are determined for each tweet using Russell’s model of affect, which correctly classifies the emotions expressed in over 90% of text messages (73). For each user against a topic/event, *Temporal Emotional State Chain* (TESC) is prepared. Next, we adopted the Hidden Markov Model (132) over the collection of TESC’s across different users and proposed different initialization methods. We called them the *Emotion-based Hidden Markov Models* (E-HMMs). The E-HMMs unfold the emotional states of the users’ list from the training observation sequence. We evaluate the model by using a testing set of users to identify the evolution of the sentiments. This proposed E-HMMs analyzes the impact of varying initialization across different users against various events. Finally, Baum-Welch forward-backward algorithm (178) applied in order to learn the HMMs and evaluation is performed using the maximum-likelihood estimation (MLE). The performance of our proposed method is compared with different linear models such as DeGroot model (43), Flocking model (74), Bised Voter model (40) and Asynchronous Linear Model (41). The result shows that the adapted HMM has the highest accuracy among others.

4.2 Proposed Framework

This section describes one of the most paramount stochastic model and our proposed approach based on this HMM architecture.

4.2.1 Hidden Markov Model

HMM is an extension of Markov chains. HMM is a doubly stochastic process with an underlying stochastic process hidden but can only be observed through another set of stochastic processes that produce the sequence of observed symbols (132). This model is more effective when we can’t directly observe the state (hidden

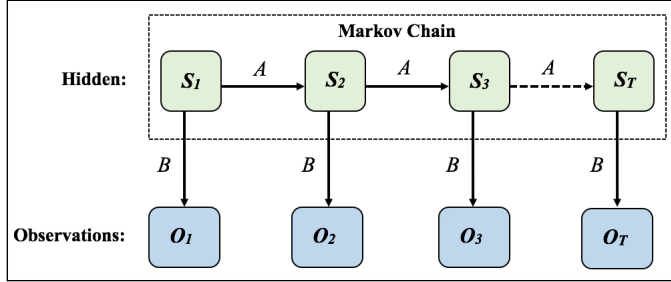


Figure 4.1: Hidden Markov Model

state). In HMM, the state is not directly visible; rather, the output result is visible, depending on the finite state stochastic sequence, as shown in Figure 4.1. To characterize an HMM completely, we define the following notation for a distinct observation HMM:

Let,

T = length of observation sequence

N = number of states in the model

M = number of observation symbols

$S = (s_1, s_2, \dots, s_T)$ = hidden state sequence

$O = (o_1, o_2, \dots, o_T)$ = observation sequence

$\{\vartheta_i, i = 1, 2, \dots, N\}$ = Possible values of each state

$\{\varphi_k, k = 1, 2, \dots, M\}$ = Possible symbols per state

$A = [\text{state transition probability distribution}]_{N \times N}$

$$a_{ij} = P(s_{t+1} = \vartheta_j | s_t = \vartheta_i), \text{ where } i, j = 1, 2, \dots, N$$

$B = [\text{observation symbol probability distribution}]_{N \times M}$

$$b_j(k) = P(o_t = \varphi_k | s_t = \vartheta_j), \text{ where } i=1, 2, \dots, N, k=1, 2, \dots, M$$

$\pi = [\text{initial state probability distribution}]_{N \times 1}$

4. SEQUENTIAL INFLUENCE MODEL FOR EMOTION DYNAMICS USING HIDDEN MARKOV MODEL

$$\pi_i = P(s_1 = \vartheta_i), \text{ where } i=1,2,\dots,N$$

So, HMM may be denoted as 5-tuple $\lambda = \langle N, M, A, B, \pi \rangle$. Note that components N, M, A, B , and π are often called parameters of HMM in which A, B , and π are essential parameters. HMM with N states $S = (s_1, s_2, \dots, s_N)$ can be characterized by a set of parameters (A, B, π) is called the model of HMM $\lambda = (A, B, \pi)$. Stochastic sequences are called observation sequences, i.e. $O = (o_1, o_2, \dots, o_T)$ whose element o_t is an observation at time point t . The matrices λ, A , and B is row stochastic, which means that each element is a probability distribution and the elements of each row sum to 1. An HMM has three fundamental problems:

1. Evaluation problem: Given the model $\lambda = (A, B, \pi)$ and observation sequence $O = (o_1, o_2, \dots, o_T)$, how we compute $P(O|\lambda)$, the likelihood probabilities of observation sequence.

2. Decoding problem: Given the model $\lambda = (A, B, \pi)$ and observation sequence $O = (o_1, o_2, \dots, o_T)$, find the optimal state sequence $S = (s_1, s_2, \dots, s_T)$. In other words, we attempt to uncover the hidden part of the model.

3. Learning problem: Given an observation sequence $O = (o_1, o_2, \dots, o_T)$ and the dimensions N and M , adjust the model parameters $\lambda = (A, B, \pi)$ to maximize $P(O|\lambda)$.

Considering the above mentioned fundamental problems, we adopted the third problem, i.e., learning problem to adjust parameters by maximizing the probability of observation sequence O . Formally, given an observation sequence O and the set of possible states in the HMM λ , learn the HMM parameters A and B , as follows:

$$(A, B, \pi) = \underset{(A, B, \pi)}{\operatorname{argmax}} P(O|\lambda) \tag{4.1}$$

4.2 Proposed Framework

The Expectation-Maximization (EM) algorithm is applied successfully into solving HMM learning problem, which is equivalently well-known Baum-Welch algorithm (13). The EM algorithm is interpreted in HMM learning problem by efficiently re-estimating the model $\hat{\lambda} = (\hat{a}_{ij}, \hat{b}_j(k), \hat{\pi}_i)$. To solve EM algorithm for HMM learning problem, we must follow two steps, i.e., E-step and M-step.

E-Step : Given the current parameter $\lambda = (a_{ij}, b_j(k), \pi_i)$, the joint probabilities $\xi_t(i, j)$ and $\gamma_t(j)$ with observation sequence O have been calculated.

$$\xi_t(i, j) = \alpha_{t-1}(i)a_{ij}b_j(O_{t+1})\beta_t(j) \text{ where } t \geq 2 \quad (4.2)$$

$$\gamma_t(j) = P(O, s_t = \vartheta_j | \lambda) = \alpha_t(j)\beta_t(j) \quad (4.3)$$

Where forward variable α_t and backward variable β_t are calculated by

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^n \alpha_t(i)a_{ij} \right] b_j(o_{t+1}) \quad (4.4)$$

$$\beta_t(i) = \sum_{j=1}^n a_{ij}b_j(o_{t+1})\beta_{t+1}(j) \quad (4.5)$$

M-Step : Given the joint probabilities $\xi_t(i, j)$ and $\gamma_t(j)$ determined at E-step, the parameters $\hat{\lambda} = (\hat{a}_{ij}, \hat{b}_j(k), \hat{\pi}_i)$ have been re-estimated.

$$\hat{a}_{ij} = \frac{\sum_{t=2}^{T-1} \xi_t(i, j)}{\sum_{t=2}^{T-1} \sum_{k=1}^N \xi_t(i, k)} \quad (4.6)$$

4. SEQUENTIAL INFLUENCE MODEL FOR EMOTION DYNAMICS USING HIDDEN MARKOV MODEL

$$\hat{b}_j(k) = \frac{\sum_{t=1}^T \sum_{o_t=\varphi_k} \gamma_t(i)}{\sum_{t=1}^T \gamma_t(i)} \quad (4.7)$$

$$\hat{\pi}_i = \gamma_1(i) \quad (4.8)$$

The solution to Learning problem can be summarized as follows.

1. Initialize, $\lambda = (A, B, \pi)$.
2. Compute $\alpha_t(i)$, $\beta_t(i)$, $\xi_t(i, j)$ and $\gamma_t(j)$.
3. Re-estimate the model $\hat{\lambda} = (\hat{a}_{ij}, \hat{b}_j(k), \hat{\pi}_i)$.
4. If $P(O|\lambda)$ increases, go to (2).

Unfortunately, given observation data, there is no way to find the global maximum of $P(O|\lambda)$. However, it might be desirable to stop if $P(O|\lambda)$ does not increase by at least some predetermined threshold and/or to set a maximum number of iterations. Baum Welch algorithm is basically a hill-climbing technique in the parameter space of the A , B matrices, so it converges at least to a local maximize of the likelihood function.

The advantage of choosing HMM has a strong statistical foundation with efficient learning algorithms where learning can take place directly from raw sequence data. But the transition of the hidden states is only influenced by the previous hidden states. However, in reality, the transition of the user's hidden emotion states from their past long-term history is more relevant to better guesstimate the actual emotion on the topic or event.

4.2.2 Emotion-based Hidden Markov Model

To create the HMM, the probabilities of the transitions among the hidden states and the emissions of the observable symbol must be calculated. If we would assume that the sentiment state of a user at time t is perfectly reflected by the sentiment of the tweet sent at t , then states would be directly observable. However, it seems reasonable to assume that the sentimental state of a user is also influenced by the tweet messages (s)he has received (incoming tweets). Although we have no means to know whether or not the user has read all the received messages mentioned by the u inside the body of text, we will assume so. Our method establishes that the sentimental state of a user who sends a message i at time t is a combination of \mathcal{R}_i and the sentiments of the received messages from t_{0-1} . This chapter mainly uses Russell’s model core affect and the psychological construction of emotion (138) for labelling the tweets.

Given a collection of tweets, we extract the sequence of tweets sent by each user $u_i \in U$ in the collection. Thus, for each user u we have a sequence of observations $O_u = \langle \uparrow m_{t_1} \dots, \uparrow m_{t_k} \rangle$ that represent the sentiments of the k tweets sent by user u at a time t . More specifically, after labeling tweets with their corresponding Russell’s region we will have $O_u = \{\mathcal{R}_i\}_{i=1}^k$, where each message i is sent at a time t that we prepare in section 3.2.2. Sentimental states of users are likewise labeled with Russell’s regions $\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3, \mathcal{R}_4$. Therefore, the HMM is defined with four hidden states $S_i = \langle s_i^1, s_i^2, s_i^3, s_i^4 \rangle$ and corresponding observations $O_k = \langle o_k^1, o_k^2, o_k^3, o_k^4 \rangle$. Let $\langle \uparrow \mathcal{R}_{i,t_1}^{out} \dots, \uparrow \mathcal{R}_{i,t_k}^{out} \rangle$ be the sentiment of the tweet sent by the user at t and let $\langle \downarrow \mathcal{R}_{i,t_{0-1}}^{in} \dots, \downarrow \mathcal{R}_{i,t_k}^{in} \rangle$ be the average sentiment of the messages received from $t - 1$ to t . The initial sentimental state of a user is calculated as a weighted average w of the sentiments of the send (w_s) and the received (w_r) messages. The initialization of the sentimental state corresponding to the

4. SEQUENTIAL INFLUENCE MODEL FOR EMOTION DYNAMICS USING HIDDEN MARKOV MODEL

emission is computed as the Russell’s region within which the pair valence-arousal $\langle w_r \times \downarrow \mathcal{R}_{i,t_{0-1}} + w_s \times \uparrow \mathcal{R}_{i,t_1}, \dots, w_r \times \downarrow \mathcal{R}_{i,t_k} + w_s \times \uparrow \mathcal{R}_{i,t_k} \rangle$ falls. We are aimed to testing whether a higher weight to the sentiment of the sent tweet reports a more accurate HMM. Because we surmise that emotionally-oriented discussions may affect a given user’s general emotional state by their neighbour. Therefore, we design different E-HMMs: E-HMM₁ with $w_r = 0.6$ and $w_s = 0.4$; E-HMM₂ with $w_r = 0.5$ and $w_s = 0.5$; and E-HMM₃ with $w_r = 0.3$ and $w_s = 0.7$.

Learning : To learn E-HMM and re-estimate the parameters of this model, we applied the most commonly used algorithm for HMMs, a form of Expectation-Maximization called the Baum-Welch algorithm (13). E-HMM is trained with the 80% of the sequences O_i and S_i of the users, using the Baum-Welch algorithm until it converges. Once the E-HMM is fitted, it is tested with the remaining 20% of the samples. The best model estimator is computed as $\hat{\lambda} = \operatorname{argmax}_{\lambda} \mathcal{L}(\lambda)$ where $\mathcal{L}(\lambda)$ is the log-likelihood of λ given by $\mathcal{L}(\lambda) = \sum_{j=1}^m P(O_j|\lambda)$ for the observation sequences O_j of the m users. The best model is the one with the highest log-likelihood value.

4.3 Experiments and Discussions

4.3.1 Experimental Set-up

We have conducted our experiments on various datasets related to policy, movie, sport, terror attack, accident and policies. The data collection method and the statistics of dataset have been introduced in Appendices A. This experiment considers twelve events/topics belonging to six different categories as shown in Table 3.1

4.3.1.1 Compared Methods

To testify our E-HMM’s effectiveness, we compare it with existing value-based models for sentiment prediction, i.e., Degroot, Flocking, Biased Voter model and AsLM.

Degroot model: This model is a network interaction model which allow us to understand information transmission, opinion formation into the network. The core idea of the DeGroot model suggest a classical method of opinion propagation. The form of updated a user’s opinion is initiated by averaging his/her own opinion including the mean opinions of his/her neighbors (43). Formally, given a undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where each link $(u, v) \in \mathcal{E}$ has weight $w_{uv} = w_{vu}$. $\mathcal{N}(u)$ denote the set of users who have an edge to the user u . At each timestamp $t_u(i)$, let $\mathcal{Y}_u(i)$ is updated opinion of the users which calculated by average emotions scores of his/her neighbour, which is denoted as:

$$\mathcal{Y}_u(i) = w_{uu}\mathcal{Y}_u(i-1) + \sum_{v=1}^{\mathcal{N}(u)} w_{uv}\mathcal{Z}_u(t_v) \quad (4.9)$$

where $\mathcal{Y}_u(i-1)$ represents the opinion score of u ’s previous tweet and $\mathcal{Z}_u(t_v)$ represents the opinion score of u ’s neighbors. We assume $w_{uu} \geq 0, w_{uv} \geq 0$ and $\sum_{v=1}^{\mathcal{N}(u)} w_{uv}\mathcal{Z}_u(t_v) = 1$ as its elements represent the factors of the weighted averaging process the DeGroot model. This model always reaches a consensus assuming the graph induced by links with $w_{uv} \geq 0$ is strongly connected. Their range is from 0 to 1.

Flocking model: This bounded confidence model is similar with the DeGroot model where updates the opinion by averaging neighbors’ similar opinions with her/himself. This model formalizes opinion of the user by assigning more weights

4. SEQUENTIAL INFLUENCE MODEL FOR EMOTION DYNAMICS USING HIDDEN MARKOV MODEL

to opinions close to user's own opinion lie on a real time (74). Let the user u updates his/her opinion by captured the set of averaging the confident neighbors' \mathcal{Z}_u^v opinion in $\mathcal{X}_u(\bar{\mathcal{Z}})$. For fixed ε , the confidence region of a user u is captured by the set

$$\mathcal{Y}_u(\bar{\mathcal{Z}}) = \{|\bar{\mathcal{Y}}_u - \bar{\mathcal{Z}}_u^v| \leq \varepsilon\} \quad (4.10)$$

where $\bar{\mathcal{Y}}_u$ and $\bar{\mathcal{Z}}_u^v$ represent the average sentiment score of user u and her/his neighbor \mathcal{Z}_u^v accordingly. This phenomenon is a special case of flocking behavior (or influence systems (26)) and focused on characterizing the convergence of bounded confidence dynamics to either absolute consensus or some clustering (polarization).

Voter and Biased Voter model: In Voter model, an individual adopt its opinion by randomly choosing opinion from one of his/her neighbor (79). The voter model can be used to model the interactions and diffusion of conflicting opinions (e.g., opinions on political candidates) in the network. A study showed (102) that people's opinion might change due to the interaction with their friends, and it has been adopted as the basis for influence diffusion. This model follow if the user's opinion lie within the small range of neighboring opinion and it cannot recognize actual dynamics opinion.

To overcome this problem, BiasedVoterModel (40) proposed which is combination of DeGroot and the voter model. In this model, an individual update his/her opinion by recovering three distinct types of opinion, i.e. stubbornness (user do not change opinion), compromise (user's initial and their neighbours' opinion) and biased conformity (user adopt neighbor's opinion without considering own opinion).

AsLM (Asynchronous Linear Model): This is a linear opinion propagation model. This model is another variant of DeGroot model, but algorithm of this model only estimate the strengths of influential edges. In this model, the opinion of each user is known, but the update opinion have no boundary and it could be different for different users (41).

4.3.1.2 Evaluation Metrics

After training the HMM model, the procedure can be described as for the predicting the observation sequence. For predicting O_j , we use $P(O_j|\lambda_{best})$ (O_j is j^{th} the observation sequence) to find those sequence which have highest $P(O_j|\lambda_{best})$ value. The emotion prediction performance for all methods is evaluated after optimizing from the best model $P(O_j|\lambda_{best})$ to the sequence of O_j observation of the users. We evaluate with evaluation metrics called prediction accuracy method which is determined if the accuracy of correctly predicted test instances. The tweet label evaluation metrics describe as follow.

$$Accuracy = \frac{\text{Number of correctly predicted tweets' emotion}}{\text{Number of test tweets' emotion}}$$

For a more detailed analysis, we further evaluate the performance of predicting different emotion regions against all influence models. We further evaluated the performance of different emotions The $F1$ -score which considers both precision and recall, is used as the measurement on each emotional region, i.e., *excitement* (R1), *contentment* (R2), *depression* (R3) and *distress* (R4).

On the other hand, we have used two different measure of errors to evaluate the performance of our proposed method on a different topic.

Mean Square Error (MSE): MSE (12) is the most simple and common metric for regression evaluation and it measures the average squared error of predictions.

4. SEQUENTIAL INFLUENCE MODEL FOR EMOTION DYNAMICS USING HIDDEN MARKOV MODEL

It can measure the quality of an estimator. The better value signifies when the value is close to zero, as smaller values imply smaller magnitudes of error. MSE at the time t is defined as follows,

$$MSE(\text{MeanSquareError}) = \frac{\sum_{t=1}^{\mathcal{N}} (\mathcal{R}_u(i) - \mathcal{S}_u(i))^2}{\mathcal{N}}$$

Mean Absolute Percentage Error (MAPE): MAPE (1154) is statistical measures which calculate the average of the absolute percentage errors of forecasts. Error is defined as the actual value minus the forecast value and the smaller value imply the better forecast. MAPE at the time t is defined as follows,

$$MAPE(\text{MeanAbsolutePercentageError}) = \frac{\sum_{t=1}^{\mathcal{N}} \left| \frac{\mathcal{R}_u(i) - \mathcal{S}_u(i)}{\mathcal{R}_u(i)} \right|}{\mathcal{N}} \times 100$$

where for both measures, $\mathcal{R}_u(i)$ is the actual value, $\mathcal{S}_u(i)$ is the predicted value and \mathcal{N} is the number of users on each emotional sequence.

4.3.2 Experimental Results

4.3.2.1 Overall Performance of Estimated Models

Our application of the HMM will be to understand the nature of changing emotions expressed in Twitter messages and analyze the transition probabilities between emotional states of the users. In this section, we represent the log-likelihood values for the final estimated models of the testing samples based on its overall performance.

Table 4.1 shows the log-likelihood values for final estimated models built with the training samples as well as the log-likelihood of the testing samples. The model obtained after the Baum-Welch process is labelled as a final model and in both

4.3 Experiments and Discussions

Table 4.1: Log-likelihood $P(O|\lambda)$ of three different initialization methods (E-HMM₁, E-HMM₂ and E-HMM₃) and other methods to be compared using DeGroot, Flocking, Voter and AsLM.

Type	Dataset (#h)	DeGroot	Flocking	B/Voter	AsLM	E-HMM ₁	E-HMM ₂	E-HMM ₃
Train	#BlackMoneyDebate	-13,601.283	-13,601.285	-13,612.688	-13,964.005	-13,285.286	-13,285.286	-13,285.283
	#Brexit	-9,758.087	-9,850.980	-9,923.680	-9,801.694	-9,758.076	-9,758.075	-9,758.074
	#AlienCovenant	-4,316.688	-4,316.699	-4,316.692	-4,316.681	-4,316.680	-4,316.670	-4,316.625
	#Baahubali2	-2,392.612	-2,392.617	-2,392.617	-2,392.623	-2,392.600	-2,392.616	-2,392.599
	#BadmintonRio2016	-1,471.439	-1,521.273	-1,471.439	-1,471.439	-1,442.761	-1,442.761	-1,442.760
	#UCLFinal	-5,152.839	-5,152.843	-5,152.849	-5,152.810	-5,152.843	-5,152.838	-5,151.643
	#SyriaGasAttack	-1,365.803	-1,365.822	-1,371.185	-1,371.201	-1,365.796	-1,365.791	-1,365.796
	#StockholmAttack	-370.323	-375.656	-372.765	-370.318	-364.972	-362.509	-370.313
	#GrenfellTower	-6,214.840	-6,214.843	-6,214.842	-6,214.839	-6,214.838	-6,214.838	-6,214.835
	#UnitedAirlinesAssault	-1,163.379	-1,163.382	-1,164.767	-1,164.765	-1,163.377	-1,163.377	-1,163.376
#MacronPresident	-485.453	-511.506	-485.459	-485.453	-485.440	-485.450	-485.452	
#Trumpregrets	-697.895	-707.757	-703.006	-697.912	-697.884	-697.882	-697.888	
Test	#BlackMoneyDebate	-1,836.186	-1,842.393	-1,842.393	-1,849.774	-1,816.937	-1,816.941	-1,816.936
	#Brexit	-2,972.115	-2,996.456	-3,012.564	-2,993.838	-2,972.113	-2,972.112	-2,972.111
	#AlienCovenant	-1,126.992	-1,127.022	-1,126.986	-1,126.984	-1,126.983	-1,126.983	-1,126.982
	#Baahubali2	-649.039	-649.063	-649.103	-649.042	-649.038	-649.036	-649.032
	#BadmintonRio2016	-431.935	-439.351	-431.935	-431.935	-411.760	-411.760	-411.759
	#UCLFinal	-1,450.196	-1,450.198	-1,450.639	-1,450.196	-1,450.182	-1,450.178	-1,450.173
	#SyriaGasAttack	-350.074	-350.099	-350.684	-350.687	-350.044	-350.012	-350.046
	#StockholmAttack	-76.243	-76.733	-75.701	-75.459	-75.451	-73.746	-75.457
	#GrenfellTower	-1,555.284	-1,555.293	-1,555.287	-1,555.284	-1,555.283	-1,555.284	-1,555.282
	#UnitedAirlinesAssault	-349.171	-349.181	-349.192	-349.185	-348.493	-349.164	-348.488
#MacronPresident	-118.831	-119.829	-118.836	-118.831	-118.820	-118.829	-118.830	
#Trumpregrets	-213.538	-213.541	-213.541	-213.539	-210.848	-210.519	-213.412	

cases (training and testing) can be seen that obtains a higher value for the log-likelihood. The models are then applied to the testing sets and the obtained results are coherent with the training dataset. It can be seen that E-HMM₃ yields to the best model for *BlackMoneyDebate*, *Brexit*, *AlienCovenant*, *Baahubali2*, *adminton-Rio2016*, *UCLFinal*, *GrenfellTower* and *UnitedAirlinesAssault* dataset. The initialization using a weight of 0.3 for the sum of all the read messages and an 0.7 for the written tweet consider to the best model. Whereas, E-HMM₂ is the best model for *SyriaGasAttack*, *StockholmAttack* and *Trumpregrets* dataset where

4. SEQUENTIAL INFLUENCE MODEL FOR EMOTION DYNAMICS USING HIDDEN MARKOV MODEL

received and sent messages reflected the same weight(0.5). However, E-HMM₁ shows as a best model for one dataset, i.e., *MacronPresident* where a weight of 0.6 reflected for received messages and 0.4 weight used for sent messages. We observed the best model only is to identify based on different datasets from different categories. It means that the sentiments do not reflect the same impact for the received messages and the message written by the user to predict the sentimental state of a person within a conversation. In three cases, we found that sentiment of the read and write messages have the same impact within the conversation. Most of the datasets are associated with high impact on written messages than reading messages. It seems reasonable that if an individual changes his/her sentiment into their conversation, stubbornness plays a major role than neighbour's opinion except in political related category. We observe that the value of $P(O_j|\lambda)$ varies for each generative model. It confirms that our proposed initialization E-HMM₁₋₃ methods result in better models compare to other baseline methods. However, some of them act differently against topics and it means sentiment of the received messages shown different impact as the sentiment reflected in the written message by the user. It can be observed that the final model obtained by E-HMM₁ hardly represents an improvement against other initial models. Therefore, we can conclude that E-HMM₂ and E-HMM₃ are better estimator of the training set for each dataset and the values obtained by E-HMM₃ support our hypothesis. The overall performance-wise E-HMM₃ performs better on each topic and converses very much fast to achieve a local maxima/global maxima in comparison to DeGroot, Flocking, Biased Voter and AsLM model.

It can be observed from experimental data that DeGroot and AsLM model performs good in terms of consistency, but achieve a local maximum with a minimum convergence point is difficult. The DeGroot and AsLM model is good enough for opinion formation and consistently obtains a good result compared to other

4.3 Experiments and Discussions

Table 4.2: Overall performance of estimated models for each dataset.

Dataset/Methods	DeGroot	Flocking	B/Voter	AsLM	E-HMM ₁	E-HMM ₂	E-HMM ₃
BlackMoneyDebate	0.5917	0.5017	0.5445	0.6225	0.6249	0.6514	0.6687
Brexit	0.5547	0.5006	0.5029	0.5828	0.6274	0.6500	0.6668
AlienCovenant	0.6257	0.5536	0.5762	0.6552	0.6787	0.7032	0.7203
Baahubali2	0.6414	0.5783	0.5908	0.6702	0.7043	0.7278	0.7448
BadmintonRio2016	0.4763	0.4698	0.4720	0.4905	0.5720	0.5715	0.5819
UCLFinal	0.5039	0.4905	0.4901	0.5182	0.5853	0.5888	0.6030
SyriaGasAttack	0.6236	0.5978	0.5948	0.6140	0.6378	0.6517	0.6415
StockholmAttacks	0.6338	0.5795	0.6154	0.6362	0.6659	0.6848	0.6747
GrenfellTower	0.6913	0.6728	0.6857	0.6923	0.6994	0.6999	0.7090
UnitedAirlinesAssault	0.6890	0.6946	0.6888	0.6813	0.7059	0.7022	0.7147
MacronPresident	0.5840	0.5728	0.5807	0.5898	0.6517	0.6424	0.6417
Trumpregrets	0.6890	0.6796	0.6713	0.6671	0.7034	0.7245	0.6997

baselines, but it’s not well fitted for emotion dynamics. Whereas, Bias voter is substantially lower than AsLM in all cases in most cases. Similarly, the performance of the Flocking is model is much poor among all and wouldn’t be sufficient for the dynamics model. As a result, our proposed methods consider the effect of emotion dynamics for propagation and accurately capture the emotion sequence on reading/writing message, which none of these existing baselines could do. Therefore, it shows a much higher fitted model in terms of performance.

Table 4.2 shows the overall performance of sentiment prediction on each dataset. The performance of E-HMM₃ is relatively impressive as compared to other E-HMMs in terms of all evaluation metrics on each event. Compared with other methods which only learns the social relation factors from opinion behaviours for each user individually, whereas E-HMMs consider different weight factors through learning from the opinion behaviours of all users. The good performance of E-HMM demonstrates its better ability to capture hidden states of the user and predictable pattern of emotional expression in twitter’s message history.

4. SEQUENTIAL INFLUENCE MODEL FOR EMOTION DYNAMICS USING HIDDEN MARKOV MODEL

4.3.2.2 Performances on Emotional Regions

The best performances of all evaluation metrics are obtained by different models, as shown in Figure 4.2 and 4.3. The E-HMMs consistently achieves better performances in sentiment prediction compare to other influence models over all the categories. However, we have found several observation on each topic.

In Figure 4.2 (Part-1), the E-HMMs perform well on the prediction of high positive region (i.e. excitement (R1)) compared to other three regions. It means that the weakness of learning other regions formation process when the communication is insufficient. In the comparison between the three E-HMMs, we can observe that the performance of E-HMM₃ is much better than E-HMM₁ and E-HMM₂. It means a higher weight to the sentiment of an outgoing tweet is better than an incoming one. We also find that the performance of AsLM shows quite better results than other baseline models. It means a person will accept the mainstream view of her/his neighbours, but future emotion may differ from others. On the other hand, most of the influence models simply follow the aggregation method, and it also has a lack of ability to uncover the temporal properties of user interactions. On the contrary, our models prove that the temporal sequence is important to understand the significant behaviour of the user.

In Figure 4.3 (Part-2), we find the performance of E-HHMs achieve better but act differently against different topics. For instance, the E-HMMs perform well on the prediction of low positive/high negative emotion regions (i.e., Depression (R3) and Distress (R4)) on the topic *#SyriaGasAttack* and *#StockholmAttack*. It seems reasonable that the tweet communication on the positive emotional aspect is insufficient. In details, we observed that the E-HMM₂ is much better than E-HMM₁ and E-HMM₃. It means people show equal weight to the sentiment of the sent and received tweets. Interestingly, in the case of *#GrenfellTower* and

4.3 Experiments and Discussions

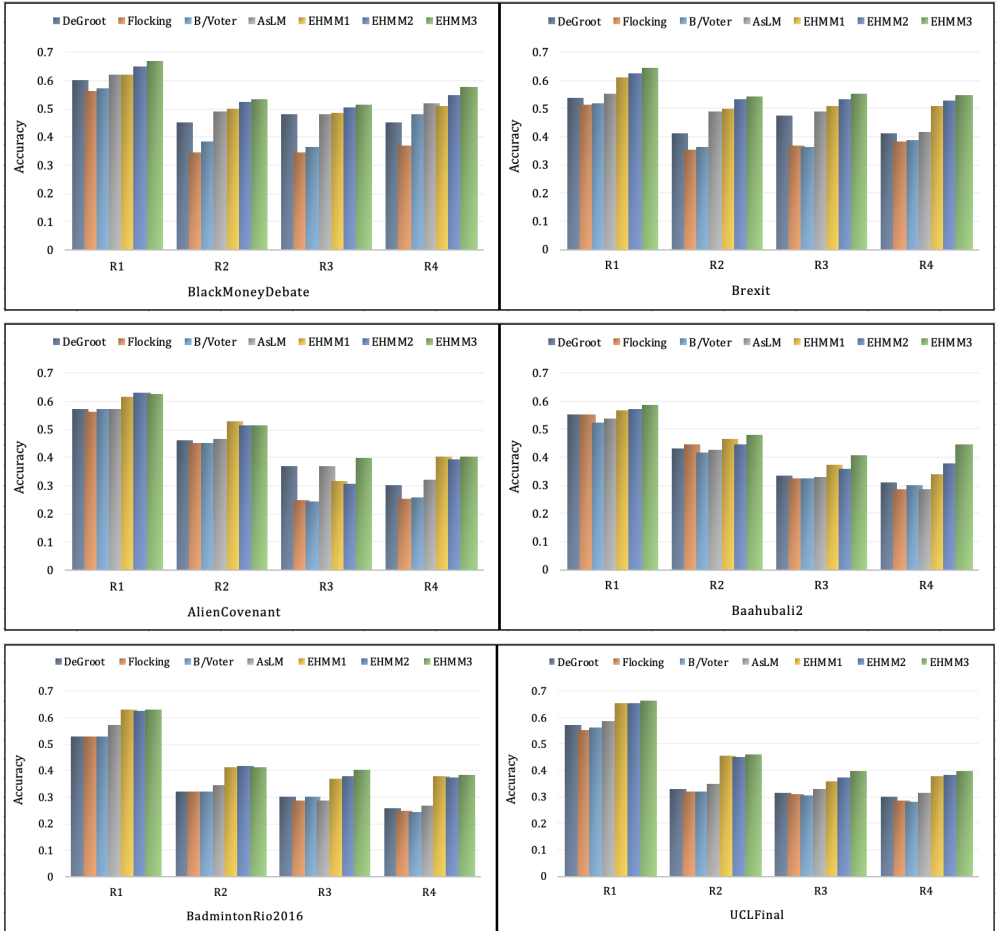


Figure 4.2: Performances on emotional regions in different categories (Part-1).

4. SEQUENTIAL INFLUENCE MODEL FOR EMOTION DYNAMICS USING HIDDEN MARKOV MODEL

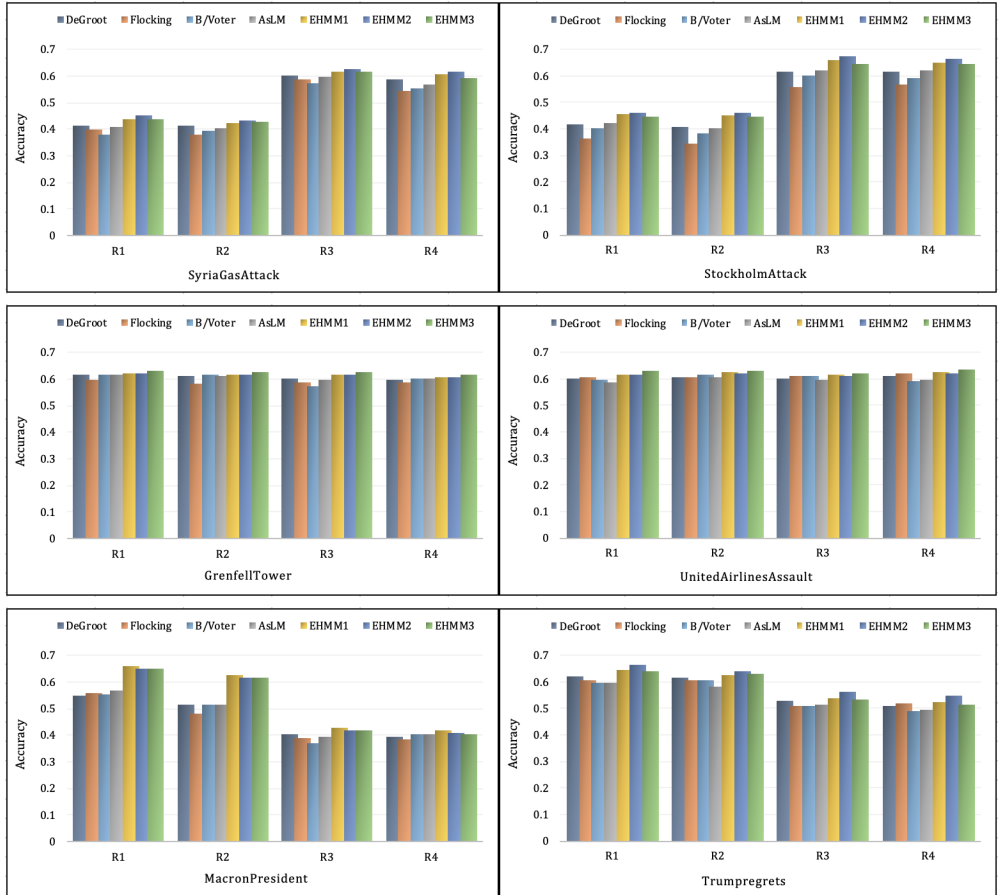


Figure 4.3: Performances on emotional regions in different categories (Part-2).

4.3 Experiments and Discussions

#UnitedAirlinesAssault topics, the prediction result for each emotional region comes up with quite an equal measure, and the performance E-HMMs show nearly similar accuracy compared to each baseline model. The result reveals that all kind of emotions is equally distributed inside the conversations. It signifies that the people mostly gave their priorities to personal opinion instead of accepting neighbours' opinions. Thus, we can assume that the topics may not contain interpersonal impact inside the conversation. The overall results indicate that the group of people may dominate when the influential conversation is insufficient. Therefore, it motivates us to utilize the interpersonal influence to benefit the emotion prediction in the insufficient communication situation.

Table 4.3: Performance metrics for Policy Category

Methods	#BlackMoneyDebate		#Brexit	
	MSE	MAPE(%)	MSE	MAPE(%)
DeGroot	2.2062	9.066%	3.9980	8.235%
Flocking	2.5452	10.577%	3.9989	8.239%
B/Voter	2.5464	10.577%	3.9989	8.239%
AsLM	2.6692	10.634%	3.9985	8.237%
E-HMM₁	2.2062	9.062%	3.7549	7.720%
E-HMM₂	2.5450	9.066%	2.6769	7.686%
E-HMM₃	1.3025	8.289%	2.6338	7.282%

4.3.2.3 Evaluation Metrics

The emotion prediction performance for all methods is evaluated after optimizing from the best model $P(O_j|\lambda_{best})$ to the sequence of O_j observation of the users. To evaluate the performance of our proposed method on a different topic, we have represented a comparative analysis of the prediction error (MSE and MAPE) of

4. SEQUENTIAL INFLUENCE MODEL FOR EMOTION DYNAMICS USING HIDDEN MARKOV MODEL

Table 4.4: Performance metrics for Movie Category

Methods	#AlienCovenant		#Baahubali2	
	MSE	MAPE(%)	MSE	MAPE(%)
DeGroot	0.1689	7.378%	0.1991	7.903%
Flocking	0.1709	7.415%	0.1995	7.903%
B/Voter	0.1691	7.397%	0.2001	7.905%
AsLM	0.1690	7.383%	0.2010	7.908%
E-HMM ₁	0.1688	7.377%	0.1993	7.907%
E-HMM ₂	0.1688	7.374%	0.1994	7.909%
E-HMM ₃	0.1684	7.364%	0.1990	7.902%

Table 4.5: Performance metrics for Sport Category

Methods	#BadmintonRio2016		#UCLFinal	
	MSE	MAPE(%)	MSE	MAPE(%)
DeGroot	2.6234	21.332%	0.0386	3.546%
Flocking	2.6234	21.332%	0.0386	3.546%
B/Voter	2.6234	21.332%	0.0385	3.546%
AsLM	0.5424	19.868%	0.0385	3.553%
E-HMM ₁	0.4069	19.868%	0.0385	3.546%
E-HMM ₂	0.4069	19.868%	0.0385	3.545%
E-HMM ₃	0.4069	19.868%	0.0383	3.534%

four state-of-the-art models along with three additional variations of our proposed models.

Table 4.3 reported that the performance of the E-HMM₃ model is better than any other models and the error for *#BlackMoneyDebate* and *#Brexit* shows the minimum errors. The variant of E-HMM₃ shown a significant performance for policy category w.r.t. E-HMM₁ and E-HMM₂, which confirms our hypothesis. However, the error rate for other models shown relatively balanced performance in this category.

4.3 Experiments and Discussions

Table 4.6: Performance metrics for Terror Attack Category

Methods	#SyriaGasAttack		#StockholmAttack	
	MSE	MAPE(%)	MSE	MAPE(%)
DeGroot	0.1059	7.128%	0.9763	28.273%
Flocking	0.1065	7.165%	1.0645	25.771%
B/Voter	0.1065	7.642%	0.8258	28.275%
AsLM	0.1077	7.649%	0.8304	28.969%
E-HMM ₁	0.1058	7.124%	0.8253	28.210%
E-HMM ₂	0.1054	7.089%	0.7143	26.438%
E-HMM ₃	0.1063	7.126%	0.8257	28.269%

Table 4.7: Performance metrics for Accident Category

Methods	#GrenfellTower		#UnitedAirlinesAssault	
	MSE	MAPE(%)	MSE	MAPE(%)
DeGroot	0.0874	5.705%	0.3949	10.459%
Flocking	0.0874	5.705%	0.3945	10.456%
B/Voter	0.0875	5.710%	0.3950	10.465%
AsLM	0.0874	5.706%	0.3954	10.464%
E-HMM ₁	0.0874	5.705%	0.3924	10.175%
E-HMM ₂	0.0874	5.705%	0.3941	10.455%
E-HMM ₃	0.0873	5.702%	0.3857	10.083%

Similarly, in the movie category, the error rate of E-HMM₃ shown a minimum value compared to other models, as shown in Table 4.4. In the case of E-HMM₁ and E-HMM₂ models, the error rates are pretty much close to each other. Thus, it's true that the performance of our initial models is substantially better than all the baselines.

We also find that E-HMM₁, E-HMM₂, and E-HMM₃ models outperform in terms of all evaluation metrics on #BadmintonRio2016 in the sports category as shown in Table 4.5. However, the performances of the other models comparatively

4. SEQUENTIAL INFLUENCE MODEL FOR EMOTION DYNAMICS USING HIDDEN MARKOV MODEL

Table 4.8: Performance metrics for Politics Category

Methods	#MacronPresident		#Trumpregrets	
	MSE	MAPE(%)	MSE	MAPE(%)
DeGroot	0.7565	25.857%	1.4637	11.746%
Flocking	0.7564	25.822%	1.4642	11.749%
B/Voter	0.7564	25.819%	1.4646	12.720%
AsLM	0.7559	25.810%	1.4646	12.451%
E-HMM₁	0.7267	23.456%	1.3047	11.746%
E-HMM₂	0.7552	25.799%	1.0219	10.882%
E-HMM₃	0.7559	25.810%	1.4166	11.746%

shown outperform on *#UCLFinal* event in the same category and it is hard to say which model performs the best on all the evaluation metrics. The existing opinion influence models are developed according to their assumption on the final state of a dynamic system. They cannot be generalized to model of emotion influence on different topics. However, our proposed model achieves a balanced performance on all evaluation metrics.

Another interesting observation is that E-HMM₂ model outperforms compared to other baseline and proposed models on the Terror attack category as shown in Table 4.6. It means people shown similar sentiments impact on the read and written messages within the conversation. We also find that the percentage error for *#StockholmAttack* is much higher against all other categories.

Table 4.7 represents the mean error rate for the accident category is almost similar for all models except for a bit difference E-HMM₃ model. When error rate is very close to each other, it is difficult to choose which model perform better among all. It demonstrates that emotional influence actually exists during a user’s decision making process.

Despite the categories mentioned above, only politics perform differently in

4.3 Experiments and Discussions

terms of performance metrics. Table 4.8 shown that the E-HMM₁ and E-HMM₂ models are outperformed of all evaluation metrics on *#MacronPresident* and *#Trumpregrets* events. It demonstrates that emotional influence plays a different role in different political scenarios. We compare our proposed methods with baselines in terms of prediction.

- The prediction performance of DeGroot is consistently good with compare other baselines. But during the training phase, this model was iterate multiple times to converge and predict the emotion.
- Flocking is comparatively better than the Biased voter model and this model updates the emotion of a user by calculating the average value of his/her neighbours. If any user does not have any neighbours, then this model makes difficult to predict the right emotion of that user. Therefore, the performance of predicting polarity comparatively poor.
- The performance of Biased Voter model not much impressive than others. Since this model update user’s emotion randomly, thus it cannot judge the actual emotional sequence for prediction. Thus, the performance of this model is unsatisfactory.
- In the case of AsLM model, the performance-wise is similar to DeGroot model. But the error rate on overall evaluation metrics is high in some cases.

Finally, we can conclude that E-HMM₃ outperforms the other baselines and proposed methods in term of predicting emotion dynamics evaluation of the users. Apart from this, we can also apply E-HMM₂ method when the average sequence of user’s post is much higher over the conversation.

4. SEQUENTIAL INFLUENCE MODEL FOR EMOTION DYNAMICS USING HIDDEN MARKOV MODEL

4.4 Conclusions

In this chapter, we have presented the Hidden Markov Model to express the nature of changing emotions from single-hashtag Twitter conversations. One of the most common Baum-Welch algorithms is applied for examining basic techniques for parameter estimation in HMMs. We propose a Temporal Emotional State chain framework to incorporate with the sentiment of the user's sequential tweets for sentiment prediction in emotion dynamics. HMM with multiplicative factor to force the hidden sentiment to generate the same emission with a higher probability. Our main result contrasts with the MLE approach to identify the best model by comparing it with the different initialization methods. From our experiment, we examine several interesting observations. First, we found a proportion linear combination of reading and written tweets are more feasible than other baseline methods. The initialization method highlights that the sentiments of the read and write messages have a different effect as the sentiment reflected in the message sent by the user within a conversation. It also demonstrates that an individual's emotional influence does not always depend on the neighbour. Many times the decision can be made on its own. As per our assumption, a higher weight to the sentiment of the sent tweet reports a more accurate than the written tweet. Finally, we also discovered that performances of emotion prediction play a different role for different events with our proposed framework.

5

Sequential Influence Model for Emotion Dynamics using Deep Learning

“No study is likely to be fruitful of results if carried on without a system. The majority of those who pursue knowledge for its own sake pursue it after an aimless and desultory fashion.”

— Bankim Chandra Chatterjee

5.1 Introduction

In the age of social network, a famous and sophisticated microblog service Twitter is profoundly used in various tasks. At the same time, social influence in online communication can gradually change someone’s opinion (6). Especially, a famous and sophisticated microblog service on Twitter, the users can interact with each other and share their opinions by expressing a few words to a par-

5. SEQUENTIAL INFLUENCE MODEL FOR EMOTION DYNAMICS USING DEEP LEARNING

ticular topic or event. Social network sites are profoundly used in various tasks such as political campaigns (71, 149), social advertisement (103), social aspects of emotions (93), propaganda news (160), expert finding (125), viral marketing, etc. (17, 31, 47, 144, 152) for influencing people. This online expression is applicable for the businessman, stockbroker and product seller for predicting customers' future opinion. With the help of these, they can appropriately modify future marketing strategies. This online expression is applicable for the businessman, stockbroker, and product seller to predict customers' future opinions and further modify the marketing strategies. Exploring the mechanism of predicting user's opinion by influencing his/her neighbours (28) and tracking dynamics changes of their opinion is practically useful (30) but technically challenging. For instance, users may not explicitly express their opinions in the text instead of showing their opinion through social interactions such as retweet on Twitter or clicking on the "Like" button on Facebook (110). This opinion dynamic in social networks has spread on many different areas from professional to everyday life such as product, movie, politics and serials (98).

The last few years have evidenced a massive growth of research on opinion propagation (24), emotion and information diffusion (149), measuring user influence (136), dynamical influence prediction (126) and opinion dynamics measure (2, 29, 40) in Twitter. In social platform, users generally follow their neighbours, and if any user changes an opinion on a specific topic or product, s(he) most likely influenced by his/her neighbours (25). Seeking opinion about necessary information, users usually influenced by their neighbours (40). But if we deeply concentrate on the practical situation, the number of users also influenced by the particular influential activity of a person (25), such as retweet, mention and reply. Therefore, uncovering user's dynamics opinion from their neighbours and represent the modified one is very important. Based on this study, researchers have proposed

several influence models by addressing neighbor's opinion (23, 40, 43). Some researchers also proposed a model that captures long-term historical information, including capturing current opinion status (28, 30, 98).

The existing study mainly focus on emotion evolution (180, 188), opinion dynamics circumstance and characterizing sentiment polarities (41, 74) such as positive, negative and neutral sentiments. However, sentiment polarities status has some limitation to capture someone's particular emotion while re-posting and replying to tweet. In reality, automatically identifying emotion parameters and estimating emotion evolution dynamics is still challenging through their conversation. With the increasing study on measuring emotion through social media, the researcher also explore to identify depression (33), stock market situation (19), political opinion (124) by incorporating prediction measure of basic emotion from Twitter posts over time. Although predicting users' emotion from their long-term past history is relevant better to guesstimate the actual opinion on the topic or event. Thus, it is important to explore deep analysis to capture the user's hidden emotion state from their conversation sequence.

In this chapter, we are especially interested in modelling emotion influence and predicting the evolution of emotion dynamics on Twitter. This study aims to identify users' emotions expressed in their written tweets and attempt to understand the change in their emotions over time. More specifically, given an incoming sequence of sentiment-labelled observation tweets before/after posting by a user, can we determine the user's dynamics updated emotion? Correspondingly, we will predict how emotion influence distributed across various social issues or events. The ultimate purpose is to predict a person's emotion and recognize how a person changes his/her emotion under the influence of neighbours' opinions. To compare with other studies, the existing works mostly concern about influence models based

5. SEQUENTIAL INFLUENCE MODEL FOR EMOTION DYNAMICS USING DEEP LEARNING

on neighbours opinion and they only focus on discrete categories of sentiment polarities. However, it is important to know which significant emotion play major role to influence the users. Therefore, we focus on emotion influential role of the neighbors and predict how emotions are dynamically transmitting from one state to another state across different topics.

To address the gap, we proposed a prediction model called *Emotion-based User Sequential Influence Model* (E-USIM) to capture four important influence factors for the prediction of user-level opinion dynamics, i.e., (i) prior neighboring emotion influence, (ii) prior personal emotion influence, (iii) current neighboring emotion influence, and (iv) historical emotion influence. In this model, we adopt Recurrent Neural Network (RNN) architecture and its variant Gated Recurrent Units (GRUs) (34) which has an able to integrate the historical information with the new coming information for prediction. Our proposed model contains a set of primary emotions which have 90% accuracy to classify the emotions (73). To exhibit the efficacy of the proposed model, we perform the experiments on the collected Twitter datasets on different topics.

5.2 Proposed Framework

In this section, we describe a sequential influence model by integrating it with the temporal emotion state. This model has the ability to solve the dynamics emotion prediction problem by considering the long history of information.

5.2.1 Emotion-Based User Sequential Influence Model (E-USIM)

This chapter proposes a novel sequential model of influence based on training data to solve the emotion dynamics prediction problem. Predicting an individual's

5.2 Proposed Framework

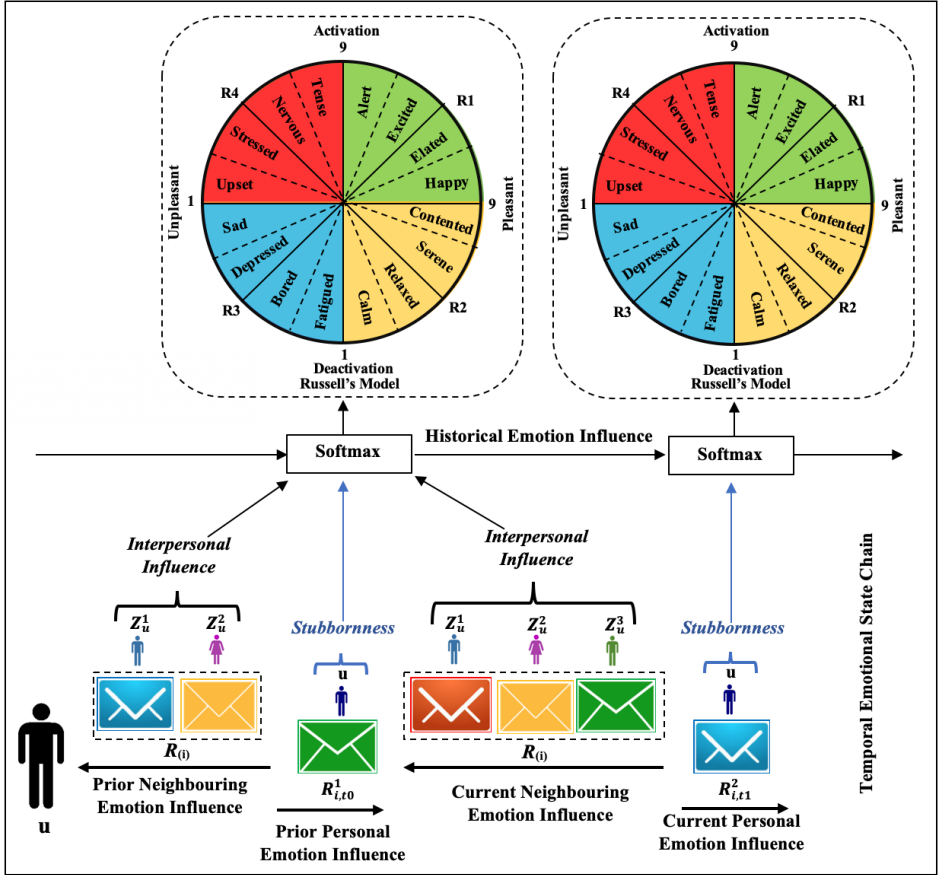


Figure 5.1: Emotion-based User Sequential Influence Model

future emotion is depending on different factors over the different time-stamp. It is also challenging to build an influence model by considering the long history of information. Therefore, we proposed a *Temporal Emotional State Chain* (TESC)(mentioned in section [3.2.2](#)) as a backbone of the E-USIM model and adopted an effective sequential model is the Recurrent Neural Network (RNN) ([147](#)), which has an ability to predict new upcoming information by amalgamat-

5. SEQUENTIAL INFLUENCE MODEL FOR EMOTION DYNAMICS USING DEEP LEARNING

ing the long history of opinion. Finally, by taking advantage of these model and recurrent neural network, we proposed a novel sequential influence model which can solve the emotion prediction problem by capturing four important influence factors, i.e., (1) prior neighbouring emotion influence, (2) prior personal emotion influence, (3) current neighbouring emotion influence, and (4) historical emotion influence, as shown in Figure [5.1](#)

Prior Neighboring Emotion Influence: The user u may updates his/her emotion based on previous opinion s(he) received. When the user u posts his/her first tweet at t_0 on topic $\#h$, s(he) might have received a tweet message from the neighbor $Z_u(i)$ on the similar topic at time t_{0-1} . We use $\alpha_{u,v}$ to denote as a degree influence which consider as a user’s incoming emotional influence $\downarrow \mathcal{R}_{i,t_{0-1}}^{C_{in}}$ by his/her neighbor. The \tanh function helps to limit the strength of the interpersonal influence between -1 and $+1$. The prior neighboring emotion vector $\mathcal{PN}_{u,i}(t)$ is obtained before a user u sending his/her first message and the influence based on average of all prior neighbor’s emotion in the set $\downarrow \mathcal{NR}_{u,i}^{in}(t)$ at time t_{0-1} as:

$$\mathcal{PN}_{u,i}(t) = \sum_{v=1}^{n(u)} \tanh(\alpha_{u,v}) \sum \downarrow \mathcal{R}_{i,t_{0-1}}^{C_{in}} \quad (5.1)$$

Prior Personal Emotion Influence: The user u posts a personal message at each timestamp $t_u(i)$ and modify his/her emotion on the particular event. We denote $\alpha_{u,0}$ as a degree of u ’s stubbornness on his/her personal post and \tanh functions helps to limit the strength of stubbornness between -1 and $+1$. The prior personal emotion vector $\mathcal{PP}_{u,i}(t)$ of u is obtained by his/her personal outgoing post $\uparrow \mathcal{PR}_{u,i}^{out}(t)$ at time t_0 . Thus, it denoted as:

$$\mathcal{PP}_{u,i}(t) = \tanh(\alpha_{u,0}) \sum \uparrow \mathcal{R}_{i,t_0}^{C_{out}} \quad (5.2)$$

Current Neighboring Emotion Influence: The user u updates his/her emotion based on the **incoming** tweet that s(he) received between two **outgoing** tweets. It means a user u may received a tweets at time t_{0+1} from the neighbor $\mathcal{Z}_u(i)$ after post his/her first tweet at time t_0 on topic $\#h$. The emotional information received from neighbor at $t_u(i)$ is a forms of the neighbor set $\downarrow NR_{u,i}^{in}(t)$. The current neighboring emotion vector $\mathcal{CN}_{u,i}(t)$ is obtained based on average of all neighbor's emotion in the set $\downarrow NR_{u,i}^{in}(t)$ between two time frame, i.e., t_0 and t_1 which devised as follow:

$$\mathcal{CN}_{u,i}(t) = \sum_{v=1}^{n(u)} \tanh(\alpha_{u,v}) \sum \downarrow \mathcal{R}_{i,t_{0+1}}^{C_{in}} \quad (5.3)$$

Historical Emotion Influence: The purpose of this section is to appropriately embed the entire history of emotion influence. The user u updates his/her emotion by considering all the past history. The historical emotion influence $\mathcal{HJ}_{u,i}(t)$ represent as a function f at time t_u by including personal emotion sequence $\uparrow \mathcal{PR}_{u,i}^{out}(t-1)$ sent and all the neighboring emotion sequence $\downarrow NR_{u,i}^{in}(t-1)$ of the messages received by the user u . Thus, it is denoted as follows:

$$\mathcal{HJ}_{u,i}(t) = f(\downarrow NR_{u,i}^{in}(t-1), \uparrow \mathcal{PR}_{u,i}^{out}(t-1)) \quad (5.4)$$

5.2.1.1 Sequential Influence Model of Emotion Dynamics

Recurrent Neural Network : For the purpose of updating user's emotion, we proposed a sequential influence model of emotion dynamics based on *Recurrent Neural Network* (RNN). The RNN is combined with prior neighboring $\mathcal{PN}_{u,i}(t)$, prior personal $\mathcal{PP}_{u,i}(t)$, current neighboring $\mathcal{CN}_{u,i}(t)$ and historical emotion influence $\mathcal{HJ}_{u,i}(t)$. To proposed the sequential influence model, we already taken place personal and neighboring emotion influence, i.e., derived by $\mathcal{X}_{u,i}(t) \in$

5. SEQUENTIAL INFLUENCE MODEL FOR EMOTION DYNAMICS USING DEEP LEARNING

$\langle \mathcal{PN}_{u,i}(t), \mathcal{PP}_{u,i}(t), \mathcal{CN}_{u,i}(t) \rangle$. To update u 's next emotion at $t_u(i+1)$, the historical emotion influence $\mathcal{HJ}_{u,i}(t-1)$ could be replaced by the internal emotion state $\mathcal{C}_{u,i}(t-1)$. Thus, it is derived as follow:

$$\mathcal{C}_{u,i}(t) = \tanh(\mathcal{W}_a [\mathcal{C}_{u,i}(t-1), \mathcal{X}_{u,i}(t)] + b_a) \quad (5.5)$$

where, $\mathcal{W}_a \in \mathfrak{R}^{d_e}$, and $b_a \in \mathfrak{R}^d$, and d_e is the dimension of emotion state region.

Gated Recurrent Unit: In our sequential influence model, we acknowledge the GRU to prove more appropriate than the simple RNN model. This gated recurrent unit is more effective and compact to drop irrelevant information and update the relevant on with affordable computation cost. This unit is formally expressed as follows:

$$[r_{u,i}(t) = \sigma(\mathcal{W}_r [\mathcal{C}_{u,i}(t-1), \mathcal{X}_{u,i}(t)] + b_r) \quad (5.6)$$

$$[g_{u,i}(t) = \sigma(\mathcal{W}_g [\mathcal{C}_{u,i}(t-1), \mathcal{X}_{u,i}(t)] + b_g) \quad (5.7)$$

$$\tilde{\mathcal{C}}_{u,i}(t) = \tanh(\mathcal{W}_c [[r_{u,i}(t) \cdot \mathcal{C}_{u,i}(t-1), \mathcal{X}_{u,i}(t)] + b_c) \quad (5.8)$$

$$\mathcal{C}_{u,i}(t) = [g_{u,i}(t) \cdot \tilde{\mathcal{C}}_{u,i}(t) + (1 - [g_{u,i}(t)]) \cdot \mathcal{C}_{u,i}(t-1) \quad (5.9)$$

where, $\mathcal{W}_r, \mathcal{W}_g, \mathcal{W}_c \in \mathfrak{R}^{d_e}$, and $b_r, b_g, b_c \in \mathfrak{R}^d$. The reset gate $[r_{u,i}(t)$ determines how much past information forget, i.e., not relevant for the future. The update gate $[g_{u,i}(t)$ helps the model to determine how much of the past information needs to process for the future. These are two important gates which decide what information should be passed to the output and its' control long-term and short-term memory accordingly. The current memory content gate $\tilde{\mathcal{C}}_{u,i}(t)$ which will use the reset gate to store the relevant information from the past. The final gate

used to calculate the internal emotion state $\mathcal{C}_{u,i}(t)$ which holds information for the current unit and passes it down to the network. In order to do that the update gate is needed. It determines what to collect from the current memory content $\tilde{\mathcal{C}}_{u,i}(t)$ and what from the previous steps $\mathcal{C}_{u,i}(t-1)$, as mention on equation [5.9](#).

5.2.2 Emotion-Based Prediction

For the purpose of predicting u 's future emotion state $\mathcal{S}_u(i)$, we formulate the problem as a predicting probability distribution by given internal emotion state $\mathcal{C}_{u,i}(t)$ and the output layer for the E-USIM is a softmax function over all categories of Russell's model of affect [\(55\)](#). The graphical representation of E-USIM is shown in Figure [5.1](#).

$$\mathbb{P}(\mathcal{S}_u(i)|\mathcal{C}_{u,i}(t)) = \text{Softmax}(\mathcal{C}_{u,i}(t) + b) \quad (5.10)$$

Where, softmax function, i.e., $\text{Softmax}(\Theta^{(i)}) = \exp(\Theta^{(i)}) / \sum_{j=1}^{\mathcal{R}_k} \exp \Theta^{(i)}$ and $b \in \mathbb{R}^{\mathcal{R}}$. \mathcal{R}_k denotes the sixteen emotion regions.

5.2.3 Parameter Estimation

To train our model, we utilise the gradients by using the Back Propagation Through Time (BPTT) approach [\(173\)](#) and maximize the log-likelihood of the sequences of emotions. To implement this algorithm, we use Tensor-Flow^{[1](#)} open-source software library with a few lines of code.

$$\mathcal{L} = \sum_{u=1}^N \sum_{i=1}^{\mathcal{R}_u} \log \mathbb{P}(\mathcal{S}_u(i) | \downarrow \mathcal{N}R_{u,i}^{in}(t), \uparrow \mathcal{P}R_{u,i}^{out}(t)) \quad (5.11)$$

¹<https://www.tensorflow.org/>

5.3 Experiments and Discussions

5.3.1 Experimental Set-up

We tried to address different types of topics/events for our experiments to check whether the user’s emotion acts differently. The data processing methods have been mentioned in Chapter 3, and the dataset’s details, as shown in Table 3.1.

5.3.1.1 Compared Methods

To testify the effectiveness of our E-USIM and explore how different components affect emotion dynamics influence, we provide the results of all the compared methods used in previous Chapter 4.

Value-based models: includes Degroot, Flocking, Biased Voter model and AsLM. The details of the models as mentioned in Section 4.3.1.1.

Value-based sequential model: is the temporal emotion-based sequential models i.e., E-HMM₁₋₃ which mentioned in Section 4.2.2. It only considers the temporal properties such as the sentimental state which is computed as four region of Russell’s model. The initial sentimental state of a user is calculated as a weighted average of the sentiments of the send and the received messages.

Content-based sequential model: is a sequential model for opinion dynamics based on content-based method. CSIM_S model which integrates the personal prior opinion influence and neighboring opinion influence are directly utilized for the sentiment prediction (30). The fine-grained opinion words strictly constrain the learning of CSIM_S model, and it uses the sentiment-based prediction strategy in the output layer.

5.3.1.2 Evaluation Metrics

Given the sequence of u 's emotion, we split the each topic of collected dataset into training set and testing data based on the posting time. The training dataset is made by emotional sequence from the previous steps $\mathcal{C}_{u,i}(t-1)$ for each user $u_i \in U$. The first 80% of the sequences of emotion state of the user are used as the training data and the remaining 20% are used as test data. The emotion prediction performance for all methods are evaluated after optimizing from $\mathbb{P}(\mathcal{S}_u(i)|\mathcal{C}_{u,i}(t))$ to the sequence of emotion state $\mathcal{S}_u(i)$ of the users.

Similar to the previous chapter, the performance is still demonstrated on the model's ability of emotion prediction and evaluated in terms of the prediction accuracy and the $F1$ -score over four emotion regions.

On the other hand, we have used two different measure of errors which determine if the predicted emotional sequence is accurate compare it to the actual sequence, as mentioned on Section [4.3.1.2](#)

5.3.2 Experimental Results

5.3.2.1 Overall Performance of Estimated Models

In Section [5.2.1.1](#), we propose to employ two popular recurrent units to capture the historical emotion influence. One is the standard Recurrent Neural Network (presented in the Equation [5.5](#)) which aggregates the entire past histories. Another one is the Gated Recurrent Unit which controls the integration of the long-term and short-term memory (presented in Equation [5.6](#), [5.9](#)). In our proposed models, E-USIM_R refers to standard RNN, and E-USIM_G refers to GRU. We explore the overall performance of these two standard units by comparing four baseline methods along with E-HMMs and CSIM_S models.

5. SEQUENTIAL INFLUENCE MODEL FOR EMOTION DYNAMICS USING DEEP LEARNING

Table 5.1: Overall performance of estimated models for each dataset.

Dataset/Methods	DeGroot	Flocking	B/Voter	AsLM	E-HMM ₁	E-HMM ₂	E-HMM ₃	CSIM.S	E-USIM _R	E-USIM _G
BlackMoneyDebate	0.5917	0.5017	0.5445	0.6225	0.6249	0.6514	0.6687	0.6841	0.6741	0.7334
Brexit	0.5547	0.5006	0.5029	0.5828	0.6274	0.6500	0.6668	0.6884	0.6839	0.7659
AlienCovenant	0.6257	0.5536	0.5762	0.6552	0.6787	0.7032	0.7203	0.6204	0.6186	0.6706
Baahubali2	0.6414	0.5783	0.5908	0.6702	0.7043	0.7278	0.7448	0.6407	0.6466	0.7005
BadmintonRio2016	0.4763	0.4698	0.4720	0.4905	0.5720	0.5715	0.5819	0.6892	0.6741	0.7252
UCLFinal	0.5039	0.4905	0.4901	0.5182	0.5853	0.5888	0.6030	0.6297	0.6323	0.6531
SyriaGasAttack	0.6236	0.5978	0.5948	0.6140	0.6378	0.6517	0.6415	0.6607	0.6924	0.6797
StockholmAttack	0.6338	0.5795	0.6154	0.6362	0.6659	0.6848	0.6747	0.6960	0.7354	0.7220
GrenfellTower	0.6913	0.6728	0.6857	0.6923	0.6994	0.6999	0.7090	0.7370	0.7211	0.7684
UnitedAirlinesAssault	0.6890	0.6946	0.6888	0.6813	0.7059	0.7022	0.7147	0.7393	0.7209	0.7471
MacronPresident	0.5840	0.5728	0.5807	0.5898	0.6517	0.6424	0.6417	0.6641	0.6706	0.6852
Trumpregrets	0.6890	0.6796	0.6713	0.6671	0.7034	0.7245	0.6997	0.7359	0.7464	0.7512

Table 5.1 presented the performance of the models on emotion prediction. As we can see, the overall performance of E-USIM_G outperforms compared to other methods in terms of all evaluation metrics. The experimental results signify that the emotional-based personal/interpersonal influences and the aggregation of historical information in E-USIMs can better describe the future emotion of the user. In this section, we also explore some detailed findings from the experiments.

Our first observation is the sequential models E-USIM_R and E-USIM_G performed better as compared to our value-based sequential models E-HMM₁₋₃ against all events. It seems reasonable that the standard RNN framework provides a better solution to track emotion dynamics and capture historical information much more efficiently than other value-based models. We also observe that standard E-USIM_R model performs well on *SyriaGasAttack* and *StockholmAttack* topics. It implies that due to insufficient conversation into the network, the E-USIM_R captures all the past histories into consideration. In the case of E-USIM_G, we find the emotion-based prediction strategy significantly outperforms on all the events other than *SyriaGasAttack* and *StockholmAttack*. It indicates that the model only captures the relevant information due to the sufficient number of instances and

forgets the irrelevant ones from the communication network. This model also has the ability to capture the correlation between the historical emotion behaviour and the future emotion of the user.

Finally, we also observe that the CSIM_S performs well with compare to standard E-USIM_R model. It is because the CSIM_S model is based on GRUs, and it has the ability to capture sentiment polarity from content-based information. However, the weakness of this model is unable to verify specific emotion instead of only summarized sentiment category labels which may harm the performance of the learned model. From this point of view, our proposed model significantly outperforms on the prediction strategy and consistently achieves the best results on all topics.

5.3.2.2 Performances on Emotional Regions

The best performances of all evaluation metrics for each emotion region are obtained by different baseline models, as shown in Figure 5.2 and 5.3. The overall performance of the E-USIM_G model is significantly impressive compared to other value-based models, CSIM_S and E-HMM₁₋₃ methods. However, we have found several details observation on each topic in term of different emotional regions.

In Figure 5.2 (Part-1), the E-USIM_G performs well on the prediction of all emotion regions compared to other models. However, due to insufficient conversation on a negative emotion, the E-USIM_G performs differently than the E-USIM_R model. For instance, the E-USIM_R model performs well in the case of the *Baahubali2* and *UCLfinal* events in respect of highly negative emotions. It indicates very few people are involved in the negative emotional conversation on these topics. On the other hand, We find that CSIM_S with sentiment-based prediction strategy significantly outperforms the model E-HMM₁₋₃ with value-based sequential on all topics. The results verify that the weight of the sentiment labels

5. SEQUENTIAL INFLUENCE MODEL FOR EMOTION DYNAMICS USING DEEP LEARNING

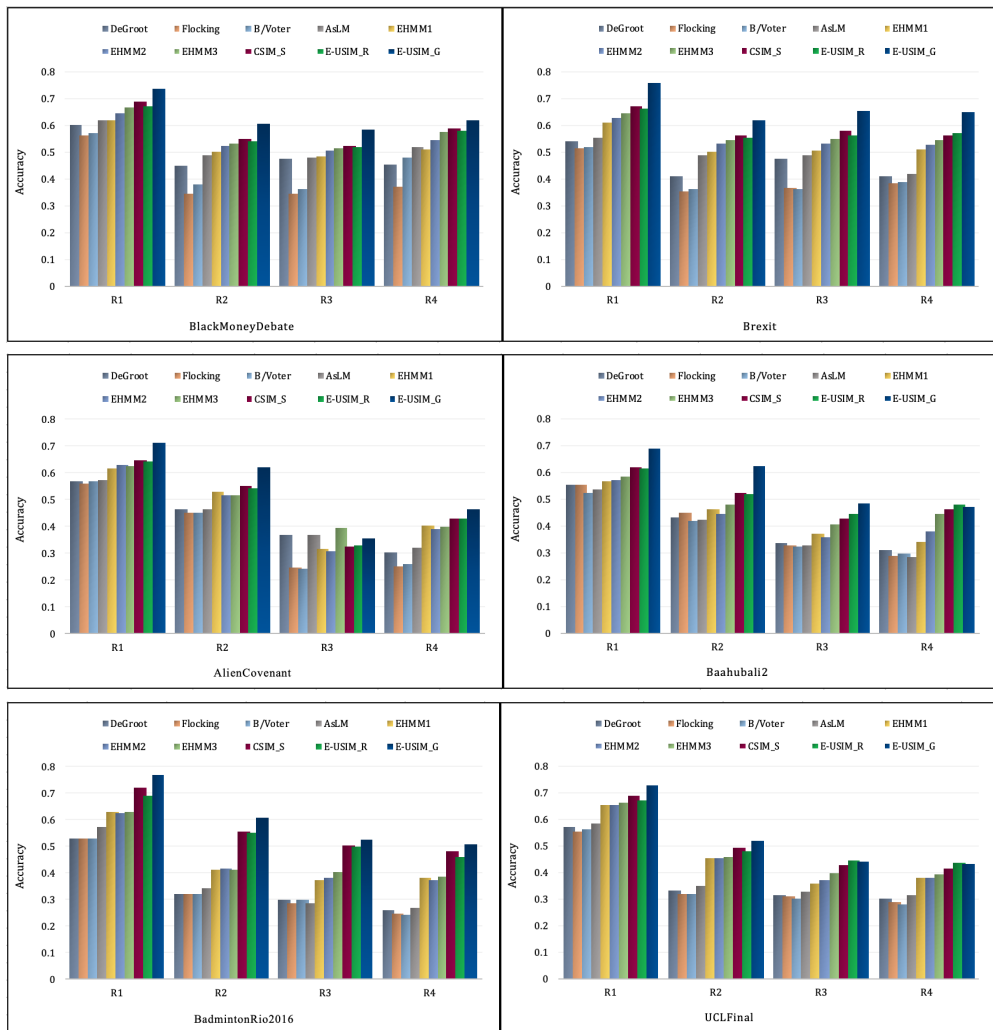


Figure 5.2: Performances on emotional regions in different categories (Part-1).

may harm the performance of the learned model. Therefore, the consistency in the overall performance of E-USIM motivates us to extend the emotional-based sequential influence model to the other scenarios on social media in our future

5.3 Experiments and Discussions

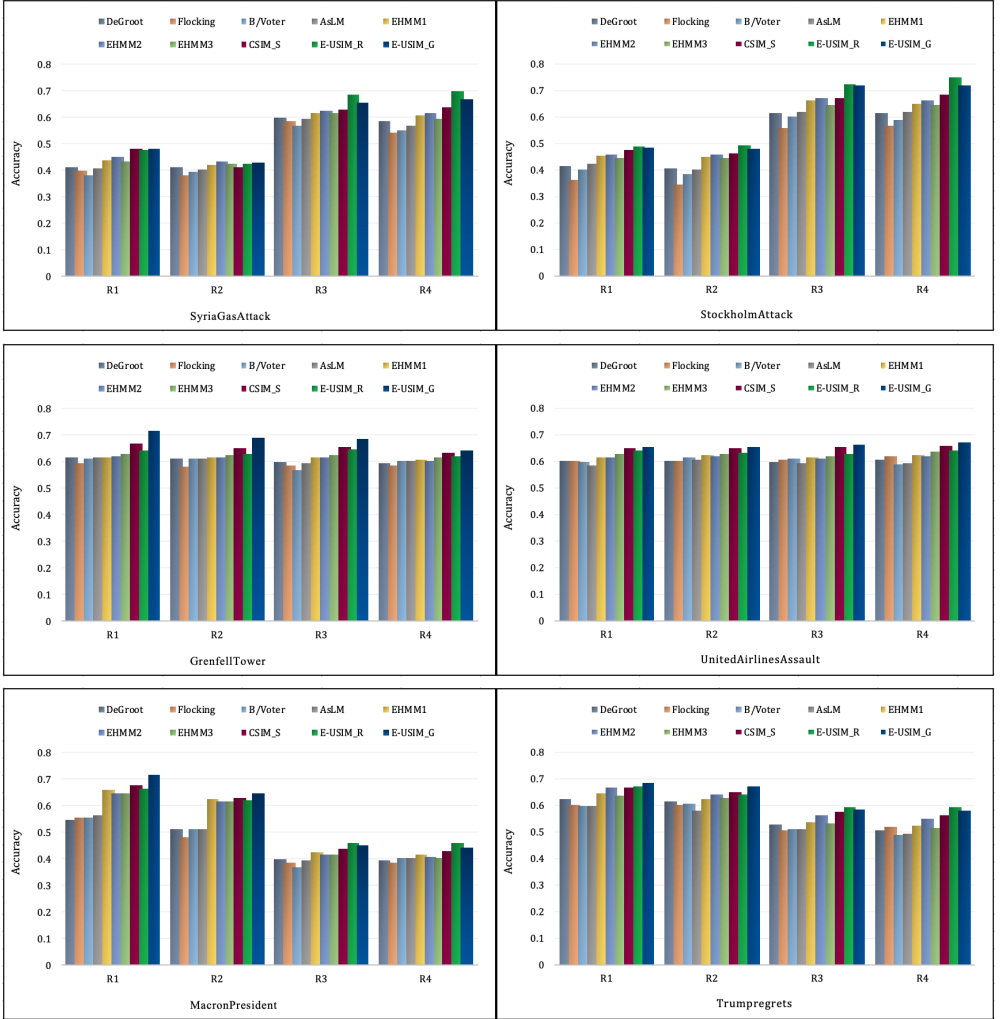


Figure 5.3: Performances on emotional regions in different categories (Part-2).

work.

In Figure 5.3 (Part-2), the overall performance of E-USIM_R and E-USIM_G is quite significant as compare to other models. Due to different strength in

5. SEQUENTIAL INFLUENCE MODEL FOR EMOTION DYNAMICS USING DEEP LEARNING

emotional-based sequential conversation of the network, the performance of our proposed models act accordingly. For instance, E-USIM_R perform better than E-USIM_G on the prediction of low positive/high negative emotion regions (i.e., Depression(R3) and Distress (R4)) on the topic *SyriaGasAttack* and *StockholmAttack*. On the other hand, E-USIM_R perform well on the prediction of low positive/high negative emotion regions on the topics *MacronPresident* and *Trumpregrets* but the accuracy measure comparatively lower than other high positive emotion regions. Based on the dataset in Table 3.1 the result high accuracy of E-USIM_R reveals that the users mainly involve in negative conversation. Whereas, low accuracy reveals that when the conversation is relatively low-active and may not have enough influential information for E-USIM_G to predict emotion regions accurately, and as a result, it harms the accuracy of emotion prediction. However, the emotional-based sequential model E-USIM consistently achieves the best results for all events/topics compared to other models. Due insufficient historical information over the network, it's difficult to learn the actual influence for emotion prediction. Therefore, prediction performance may even harm the results for the topic with small dataset.

5.3.2.3 Evaluation Metrics

The emotion prediction performance for all methods are evaluated after optimizing from $\mathbb{P}(\mathcal{S}_u(i)|\mathcal{C}_{u,i}(t))$ to the sequence of emotion state $\mathcal{S}_u(i)$ of the users. To evaluate the performance of our proposed method on different topics, we have used two different measures of error, as mentioned on previous Chapter in Section 4.3.1.2

Table 5.2-5.7 reports a comparative analysis of the prediction error (MSE and MAPE) of several baseline models along with our proposed method. We observe that for all the datasets, the overall performance of our proposal is substantially

5.3 Experiments and Discussions

Table 5.2: Performance metrics for Policy Category

Methods	#BlackMoneyDebate		#Brexit	
	MSE	MAPE(%)	MSE	MAPE(%)
DeGroot	2.2062	9.066%	3.9980	8.235%
Flocking	2.5452	10.577%	3.9989	8.239%
B/Voter	2.5464	10.577%	3.9989	8.239%
AsLM	2.6692	10.634%	3.9985	8.237%
E-HMM ₁	2.2062	9.062%	3.7549	7.720%
E-HMM ₂	2.5450	9.066%	2.6769	7.686%
E-HMM ₃	1.3025	8.289%	2.6338	7.282%
CSIM_S	1.2031	8.046%	2.2976	7.068%
E-USIM _R	1.2545	8.166%	2.4796	7.186%
E-USIM _G	0.9305	7.829%	1.3835	6.822%

Table 5.3: Performance metrics for Movie Category

Methods	#AlienCovenant		#Baahubali2	
	MSE	MAPE(%)	MSE	MAPE(%)
DeGroot	0.1689	7.378%	0.1991	7.903%
Flocking	0.1709	7.415%	0.1995	7.903%
B/Voter	0.1691	7.397%	0.2001	7.905%
AsLM	0.1690	7.383%	0.2010	7.908%
E-HMM ₁	0.1688	7.377%	0.1993	7.907%
E-HMM ₂	0.1688	7.374%	0.1994	7.909%
E-HMM ₃	0.1684	7.364%	0.1990	7.902%
CSIM_S	0.1565	7.236%	0.1769	7.866%
E-USIM _R	1.1596	7.316%	0.1795	7.896%
E-USIM _G	1.3250	7.1312%	0.1754	7.842%

better than all the baselines. Among the baselines, we find that E-USIM_G performs best. However, the variants of E-USIM_R show a significant performance on *Syria-GasAttack*, *StockholmAttack*, *UnitedAirlinesAssault* and *MacronPresident* which

5. SEQUENTIAL INFLUENCE MODEL FOR EMOTION DYNAMICS USING DEEP LEARNING

Table 5.4: Performance metrics for Sport Category

Methods	#BadmintonRio2016		#UCLFinal	
	MSE	MAPE(%)	MSE	MAPE(%)
DeGroot	2.6234	21.332%	0.0386	3.546%
Flocking	2.6234	21.332%	0.0386	3.546%
B/Voter	2.6234	21.332%	0.0385	3.546%
AsLM	0.5424	19.868%	0.0385	3.553%
E-HMM ₁	0.4069	19.868%	0.0385	3.546%
E-HMM ₂	0.4069	19.868%	0.0385	3.545%
E-HMM ₃	0.4069	19.868%	0.0383	3.534%
CSIM_S	0.3201	19.066%	0.0367	3.486%
E-USIM _R	0.3212	19.096%	0.0368	3.489%
E-USIM _G	0.3195	19.059%	0.0367	3.486%

Table 5.5: Performance metrics for Terror Attack Category

Methods	#SyriaGasAttack		#StockholmAttack	
	MSE	MAPE(%)	MSE	MAPE(%)
DeGroot	0.1059	7.128%	0.9763	28.273%
Flocking	0.1065	7.165%	1.0645	25.771%
B/Voter	0.1065	7.642%	0.8258	28.275%
AsLM	0.1077	7.649%	0.8304	28.969%
E-HMM ₁	0.1058	7.124%	0.8253	28.210%
E-HMM ₂	0.1054	7.089%	0.7143	26.438%
E-HMM ₃	0.1063	7.126%	0.8257	28.269%
CSIM_S	0.1036	7.042%	0.6989	25.398%
E-USIM _R	0.1032	7.033%	0.6938	25.338%
E-USIM _G	0.1034	7.037%	0.6969	25.382%

conforms to the utility of our proposal. Most importantly, E-USIMs properly captures all the past history of emotion influence for each user. As a result, it provides better result in terms of forecasting emotion even at a distant future. Despite the

5.3 Experiments and Discussions

Table 5.6: Performance metrics for Accident Category

Methods	#GrenfellTower		#UnitedAirlinesAssault	
	MSE	MAPE(%)	MSE	MAPE(%)
DeGroot	0.0874	5.705%	0.3949	10.459%
Flocking	0.0874	5.705%	0.3945	10.456%
B/Voter	0.0875	5.710%	0.3950	10.465%
AsLM	0.0874	5.706%	0.3954	10.464%
E-HMM ₁	0.0874	5.705%	0.3924	10.175%
E-HMM ₂	0.0874	5.705%	0.3941	10.455%
E-HMM ₃	0.0873	5.702%	0.3857	10.083%
CSIM.S	0.0868	5.692%	0.3745	9.683%
E-USIM _R	0.0872	5.698%	0.3739	9.635%
E-USIM _G	0.0868	5.686%	0.3743	9.679%

Table 5.7: Performance metrics for Politics Category

Methods	#MacronPresident		#Trumpregrets	
	MSE	MAPE(%)	MSE	MAPE(%)
DeGroot	0.7565	25.857%	1.4637	11.746%
Flocking	0.7564	25.822%	1.4642	11.749%
B/Voter	0.7564	25.819%	1.4646	12.720%
AsLM	0.7559	25.810%	1.4646	12.451%
E-HMM ₁	0.7267	23.456%	1.3047	11.746%
E-HMM ₂	0.7552	25.799%	1.0219	10.882%
E-HMM ₃	0.7559	25.810%	1.4166	11.746%
CSIM.S	0.7137	23.323%	0.9219	10.382%
E-USIM _R	0.7124	23.312%	0.9616	10.654%
E-USIM _G	0.7134	23.320%	0.9198	10.156%

above observation, we also report separately some important findings from the experiments.

We observe that CSIM.S performs better than other baselines including E-

5. SEQUENTIAL INFLUENCE MODEL FOR EMOTION DYNAMICS USING DEEP LEARNING

USIM_R on policy, movie and sport category. Since CSIM_S considers the message dynamics into consideration and describes how an individual is influenced by their neighbours to form the future sentiment. However, the weakness of the model is the limited variety of sentiment category labels which harms the performance of the learned model. Besides, the performance of the CSIM_S model also affects when the sequence of data is insufficient compared to E-USIM_G. For instance, E-USIM_G performs well on the *SyriaGasAttack*, *StockholmAttack* and *Trumpregrets* topics.

Interesting, we observe that the E-USIM_R gives proper leverage to accumulate the entire past histories, and as a consequence, it can properly anticipate the emotion dynamics from the short conversation. This model adopts a principle approach to exploit the continuous dynamics for forecasting user's emotion. Whereas, the E-USIM_G captures the sequential personal/interpersonal influences and most relevant past histories for each individual user. On the other hand, this model considers the effect of emotion propagation on dynamics conversation. Hence, it is able to accurately capture the significant emotion from the entire communication process, which none of the existing baselines could do. As a result, it consistently achieves the best results and more robust model than E-USIM_R.

5.3.2.4 Effect on Data Partitions

Our study obtained all the experimental results based on the 8:2 ratio of training and test data. To see the effectiveness of the proposed model, we experimented with two more data partitions where the ratio of training and test data is 7:3 and 9:1. The accuracy of prediction performance on the twelve events/topics are listed in Figure 5.4. We can see that predictive accuracy for all topics does not show significant differences except for *BlackMoneyDebate*, *BadmintonRio2016* and *GrenfellTower*. It stipulates that when the volume of training data not sufficient,

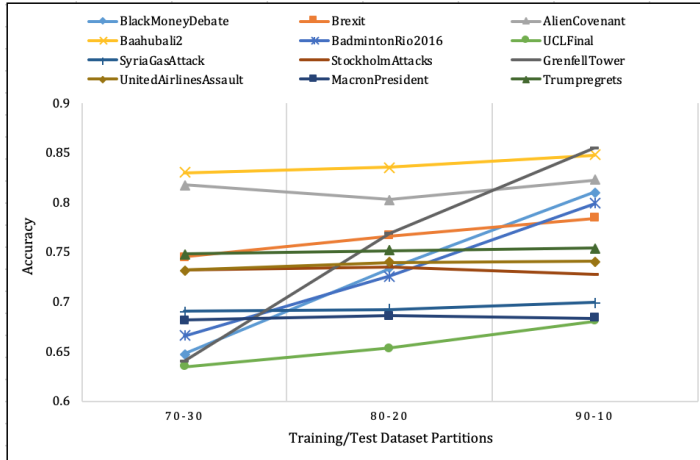


Figure 5.4: Effect on Different Data Partitions

it may not reflect much on prediction accuracy. The model could better predict when the volume of training data increases and the average number of sent tweets increases, as reflected on the topics mentioned above. The details of the data statistics mention on the Table [3.1](#)

5.4 Conclusions

In this chapter, we propose a Emotion-based User Sequential Influence Model (E-USIM) to incorporate with the historical information for predicting emotion dynamics from single-hashtag Twitter conversations. To devise such a model, we adopt the standard recurrent neural network architecture which integrate the historical opinion information with the new coming information for prediction. To train our model, we compute Back Propagation Through Time (BPPT) approach and maximize the log-likelihood of the sequences of emotions. To identify specific emotion of the user instead of measure only polarity, we employed a psychological

5. SEQUENTIAL INFLUENCE MODEL FOR EMOTION DYNAMICS USING DEEP LEARNING

model i.e. Russell's model of affect which presented his categorization in a wheel of emotions. Based on the current work, we examine users' expressions by considering different emotional influence factors. As a result, our approach outperforms the existing approaches in terms of all evaluation matrices.

6

Conclusions and Future Work

“Wisdom is not a product of schooling but of the lifelong attempt to acquire it.”

— Albert Einstein

In the past decade, the growth of social media became very attractive research field and it has revolutionized the analysis of influence and information diffusion among people. Specially, a famous and sophisticated microblog service in Twitter, the users are permitted to interact with each others and share their feelings/opinion by expressing in a few words to a particular topic or product. Sentiment extraction and opinion mining became a more challenging task in the social media platforms by Natural Language Processing (NLP), text analytics and computational linguistics. The research on sentiment analysis mainly focuses on extract sentiment and opinion mining. Sentiment analyzer or researchers are proposing and sharing different approaches to determine sentiment and opinion from Microblogs, Twitter, websites, forums, etc. However, it is also essential to understand people’s opinion

6. CONCLUSIONS AND FUTURE WORK

on various events. Because event-wise people, most people change their opinion and reflect on the message containing their emotional state. Begin with a concrete propagation, i.e., Dave posts a tweets in Twitter about pandemic situation which he is sad or depress about it. Some of his followers on Twitter reply to his tweet while others retweet it. This kind of diffusion process is profoundly used in various applications on social network sites such as political campaigns, social advertisement, the spread of rumors/news, expert finding, viral marketing, behavioral targeting etc., on an excellent understanding of propagation mechanisms by people influential activities and accurate predictions of dynamic information. This dynamic information in social networks has spread on many different areas from professional to everyday life, such as product, movie, politics, disaster, attack and pandemic. For example, governments can make a decision and maintain stability by tracking the temporal propagation of public emotions on the disaster. Understanding the mechanism of dynamics emotion being influenced by their neighbours' opinion and predicting the emotion dynamics process is a significant study to explore. With the adaptation of different algorithms and knowledge of the understanding psychological model, we have performed our research on the emotional influence process and predict the emotion dynamics on social media.

Our objective is to analyze the sentiments derived within a conversation on Twitter and investigate how the emotions are changed dynamically across the users that take part in the conversation. For our purpose of analyzing real-time events, we have collected our dataset from Twitter against different categories. The purpose of analyzing various events is to identify the clear vision of people's emotion and opinion on different circumstances in social media. We tried to address different types of topics to check whether the user's emotion acts differently.

In this thesis, we carefully and comprehensively study the problem of emotional influence and predicting emotion dynamics by analyzing network information and

proposing computational modelling on various events. More specifically, we solve some important sub-problems, i.e., understanding diffusion pattern by analyzing user's emotion present in interactions, unfolding how likely user's emotion evolved inside the temporal conversation, uncovering personal/interpersonal influence by proposing sequential emotion influence model and predicting dynamic propagation of emotional information by capturing user's historical information. In order to solve these problems, we have used a very powerful machine learning framework and also compared with different state-of-the-art studies in this thesis.

6.1 Summary of Contributions

The following sections summarize the contributions of each chapter.

6.1.1 Emotion Dynamics of Public Opinions

- For our purpose of analyzing real-time events, we chronologically started collecting our datasets over the year by addressing twelve events/topics belonging to six different categories (1) Policy; (2) Movie; (3) Sports; (4) Terror Attack; (5) Accident and (6) Politics.
- The TESC sequential network has been proposed in order to integrate every participating user's outgoing and incoming tweets information within a two-time frame.
- The purpose of analyzing various events is to identify the clear vision of people's emotion and opinion on different circumstances in social media.
- This work focuses on the emotional dynamics of a user while posting comments against an event/topic from three different social characteristics: (i) emotional, (ii) community, and (iii) conversational.

6. CONCLUSIONS AND FUTURE WORK

- We investigate social dynamics of emotion present in user's opinion and attempt to understand: (i) changing characteristics of user's emotion toward a social issue over time, (ii) influence of public emotions on individual's emotion, and (iii) causing of changing opinion by social factors etc.
- The first investigation shows that people at the initial stage display high positive emotion while continuing their conversation, most people going towards high positive direction.
- Another observation shows that the emotional state of a significant number of posts does not influence the dominant emotion on social networks.
- Finally, the experimental results show that most people agree with positive emotion when users change their emotional state from negative to positive against a topic.

6.1.2 Sequential Influence Model for Emotion Dynamics using Hidden Markov Model

- We have performed an investigation on observing emotions unfolding in a consecutive sequence of tweets for a particular user based on his/her temporal sequential information.
- We have developed a model which has the ability to understand the nature of changing emotions expressed in Twitter messages and analyze the transition probabilities between the emotional states of the users.
- Our model outperformed with baseline methods and considered the effect of emotion dynamics for propagation, and accurately captured the emotion sequence on reading/writing message.

- The initialization method highlights that the sentiments of the read and write messages have a different effect as the sentiment reflected in the message sent by the user within a conversation.
- The experimental results demonstrate that an individual’s emotional influence does not always depend on the neighbour. A higher weight to the sentiment of the sent tweet reports more accurate than the written tweet.
- We also discovered that performances of emotion prediction play a different role for different events with our proposed framework.

6.1.3 Sequential Influence Model for Emotion Dynamics using Deep Learning

- We propose a novel emotion-based sequential influence model with the combination of personal and interpersonal influential information.
- To prepare this model, we have adopted neural network architecture to integrate the historical information with the new coming information for prediction. It has a great ability to integrate historical information and capture the semantic information embedded in the message.
- Compared to our previous model, the E-USIM model can capture all the sixteen emotions from Russell’s model of affect instead of only measure polarity.
- The performance of the model shows the accurate result on each topic, and its novel prediction strategy provides a significant result compared to other baseline methods.

6. CONCLUSIONS AND FUTURE WORK

6.2 Future Work

Though we have studied considerable efforts to address sentiment propagation problems on various real-life events in the social network, there are still open issues to be explored in the future in a different direction. Finally, we present some potential trends of our current work.

There are vast amounts of social media data being collected today to observe and analyze the interests and opinions of millions of people about politicians, products, disaster, pandemic, etc. On the other hand, due to the exponential growth of social media interaction, various influencers or actors are weaponizing information to run false information campaigns with targeted manipulation of public opinion on specific topics and threatening social/political development (72, 185). For example, false information can hurt the image of a candidate at political events, and false information can cause panic and general chaos during disaster or pandemic situations. Exploring the mechanism of identifying false information and tracking user emotion on information propagation is technically challenging. The user dynamics emotion on information propagation in social networks has spread in different areas ranging from professional use to personal use in everyday life like viral marketing, political campaigns, propaganda news and social advertising, etc. Therefore, *detecting false information types by analysing different features and examining public opinions or emotions by understanding its propagation patterns* is a significant study to explore. In social media, many researchers have executed various studies on false information detection and opinion dynamics separately by adopting different features, including emotional features. In Reference (60), authors have been conducted for detecting types of false information by incorporating sentiment features. But the combination of detecting false information and tracking the emotion dynamics on social media is a kind of new direction of our

study.

By performing this research, we will improve the automatic extraction of false information and the types of emotion associated with it in the global community. This research aims to go behind such techniques by designing a novel system to link the cognitive and affective breach between word-level natural language data and concept-level opinions. There are so many commercial companies, welfare society that supplies false information detection tools and sentiment analysis services. The conceptual foundation of such new technology and unique scientific contributions that envision satisfying society's demands for transparency. Therefore, such development is essential for the news industry and government welfare society to track false information and the emotional status over social communications.

Appendix A

Tweet Retrieval Methods

Due to the increasing popularity of Twitter Sentiment Analysis, researchers have used the Twitter API services to crawl tweet messages and create their own datasets, particularly in the field of politics, games, products and reviews. All the tweets were collected in the english language. This section describes the experimental dataset used in our study. For our purpose of analyzing real-time events, we chronologically retrieved tweets through Twitter *Search-API*^[1] and created our own datasets. The collected tweets are stored in Neo4j (80), a robust and high-performance native graph database for storage, query performance from any scale. Neo4j^[2] is an open-source database implemented in Java and Scala. We have implemented different retrieval methods to extracts massive data from the Twitter conversation related *#hashtag* events. In the following subsections, we present the composition of each data collection.

¹<http://twitter4j.org>

²<https://neo4j.com/>

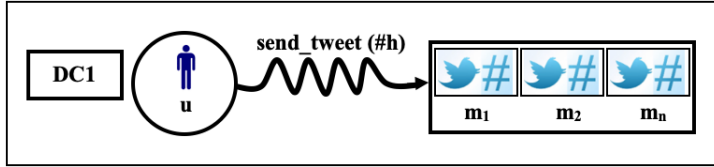


Figure A.1: Data Collection Method 1

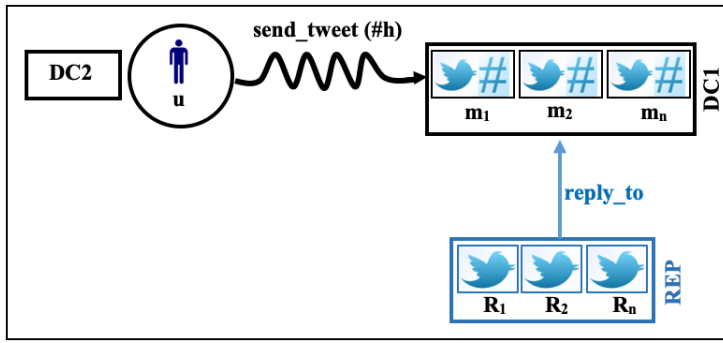


Figure A.2: Data Collection Method 2

A.1 Data Collection 1

The first data retrieval method is called *DC1*, which compose a list of the messages that explicitly mention the hashtag ($\#h$) within the text of the tweet message (see Figure [A.1](#)). *DC1* is a somewhat restrictive method because some of the messages related to the conversation that does not explicitly contain $\#h$ within the text's body. For example, if someone replies to a user's $\#h$ tweet and the reply do not include $\#h$, this message will not be retrieved as part of the *DC1* collection. Therefore, we have adopted another method to collect this missing information through our second data collection method.

A. TWEET RETRIEVAL METHODS

A.2 Data Collection 2

To perform the second data retrieval method $DC2$, we have collected the list of replies for a given message of $DC1$ by querying through the same Twitter Search API. We have retrieved the list of reply tweets REP for each message in $DC1$ and identified as R_1, R_2, \dots, R_n for all the replies received by the n messages (see Figure [A.2](#)). It seems reasonable to consider all the replies as a part of a user’s sentimental state, even though REP messages do not contain #h within the body of the text. Thus, $DC2$ is a combined set of messages retrieved in $DC1$ plus the list of all replies REP , i.e., $DC2 = DC1 \cup REP$.

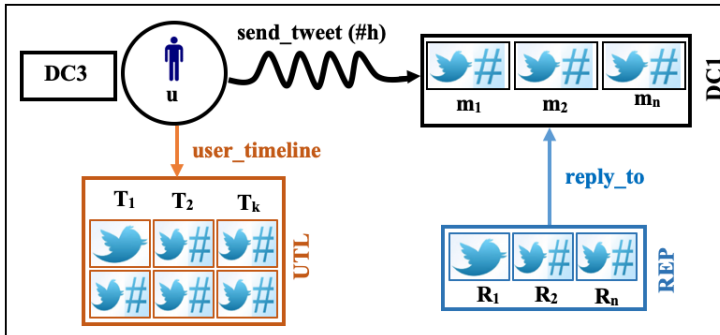


Figure A.3: Data Collection Method 3

A.3 Data Collection 3

In case of third data retrieval method $DC3$, we have retrieved data from the user’s timeline by querying each user $u \in U$ between two-time points associated with two posted tweets. The list of messages UTL received from the users’ timeline are identified as T_1, \dots, T_k , as shown in Figure [A.3](#). We assume that the messages in UTL may influence the sentimental state of the user, even though the messages

in the sets T_1, \dots, T_k do not contain #h within the body of the text. Thus, $DC3$ is a combination of the messages retrieved in $DC1$ and the list of replies REP to each message in $DC1$ plus the messages contained in the timeline REP of the users, i.e., $DC3 = DC1 \cup REP \cup UTL$.

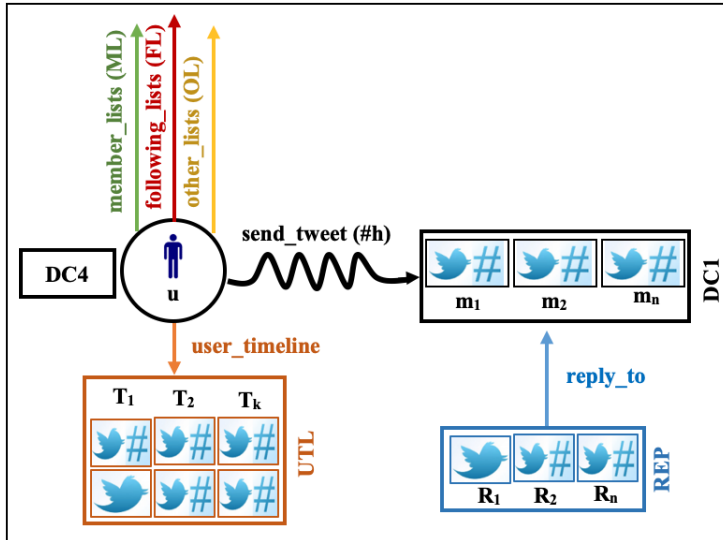


Figure A.4: Data Collection Method 4

A.4 Data Collection 4

Despite the methods mentioned above, we also have collected more data on each user in the fourth data collection method $DC4$. In this method, the member-lists ML , following-lists FL , and other-lists OL of data has been retrieved for each user between the two-time frame (see Figure [A.4](#)). We hypothesise that the tweets from the same group or tweets bearing by followee may affect the user’s sentiment at some point. Therefore, it’s also essential to consider this data

A. TWEET RETRIEVAL METHODS

for our study. Thus, $DC4$ is a combination of the messages retrieved in $DC1$ and the list of replies REP plus the messages contained in the timeline REP of the users including member-lists ML , following-lists FL , and other-lists OL , i.e., $DC4 = DC1 \cup REP \cup UTL \cup ML \cup FL \cup OL$.

References

- [1] AHMED ABBASI AND HSINCHUN CHEN. **Analysis of affect intensities in extremist group forums.** In *Terrorism informatics*, pages 285–307. Springer, 2008. [15](#)
- [2] DARON ACEMOGLU AND ASUMAN OZDAGLAR. **Opinion dynamics and learning in social networks.** *Dynamic Games and Applications*, **1**(1):3–49, 2011. [98](#)
- [3] APOORV AGARWAL, BOYI XIE, ILIA VOVSHA, OWEN RAMBOW, AND REBECCA PASSONNEAU. **Sentiment analysis of twitter data.** In *Proceedings of the workshop on languages in social media*, pages 30–38. Association for Computational Linguistics, 2011. [2](#), [14](#)
- [4] SAIFUDDIN AHMED, KOKIL JAIDKA, AND JAEHO CHO. **Tweeting India’s Nirbhaya protest: a study of emotional dynamics in an online social movement.** *Social Movement Studies*, **16**(4):447–465, 2017. [33](#)
- [5] ANTHONIRAJ AMALANATHAN AND S MARGRET ANOUNCIA. **Influence of human emotion expressed through social networks in viral marketing.** *International Journal of Business Information Systems*, **26**(1):15–32, 2017. [20](#)

REFERENCES

- [6] ARIS ANAGNOSTOPOULOS, RAVI KUMAR, AND MOHAMMAD MAHDIAN. **Influence and correlation in social networks**. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 7–15. ACM, 2008. [97](#)
- [7] CHRISTOPHE ANDRIEU, NANDO DE FREITAS, ARNAUD DOUCET, AND MICHAEL I JORDAN. **An introduction to MCMC for machine learning**. *Machine learning*, **50**(1-2):5–43, 2003. [50](#)
- [8] THEO ARAUJO, PETER NEIJENS, AND RENS VLIEGENTHART. **Getting the word out on Twitter: The role of influentials, information brokers and strong ties in building word-of-mouth for brands**. *International Journal of Advertising*, **36**(3):496–513, 2017. [19](#) [21](#)
- [9] CHALEE ASAVATHIRATHAM, SANDIP ROY, BERNARD LESIEUTRE, AND GEORGE VERGHESE. **The influence model**. *IEEE Control Systems Magazine*, **21**(6):52–64, 2001. [25](#)
- [10] EYTAN BAKSHY, ITAMAR ROSENN, CAMERON MARLOW, AND LADA ADAMIC. **The role of social networks in information diffusion**. In *Proceedings of the 21st international conference on World Wide Web*, pages 519–528. ACM, 2012. [23](#)
- [11] SUMIT BASU, TANZEEM CHOUDHURY, BRIAN CLARKSON, ALEX PENTLAND, ET AL. **Learning human interactions with the influence model**. NIPS, 2001. [30](#)
- [12] GLENN J BATTAGLIA AND JAMES M MAYNARD. **Mean Square Error: A useful tool for statistical process management**. *AMP Journal of Technology*, **34**:1256–1260, 1996. [83](#)

-
- [13] LEONARD E BAUM. **An equality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes.** *Inequalities*, **3**:1–8, 1972. [77](#), [80](#)
- [14] YOSHUA BENGIO, RÉJEAN DUCHARME, PASCAL VINCENT, AND CHRISTIAN JAUVIN. **A neural probabilistic language model.** *Journal of machine learning research*, **3**(Feb):1137–1155, 2003. [28](#)
- [15] ADAM BERGER, STEPHEN A DELLA PIETRA, AND VINCENT J DELLA PIETRA. **A maximum entropy approach to natural language processing.** *Computational linguistics*, **22**(1):39–71, 1996. [15](#)
- [16] JONAH BERGER. **Arousal increases social transmission of information.** *Psychological science*, **22**(7):891–893, 2011. [19](#)
- [17] BIN BI, YUANYUAN TIAN, YANNIS SISMANIS, ANDREY BALMIN, AND JUNGHO CHO. **Scalable topic-specific influence analysis on microblogs.** In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 513–522. ACM, 2014. [33](#), [98](#)
- [18] JOHAN BOLLEN, HUINA MAO, AND ALBERTO PEPE. **Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena.** In *Fifth International AAAI Conference on Weblogs and Social Media*, 2011. [6](#)
- [19] JOHAN BOLLEN, HUINA MAO, AND XIAOJUN ZENG. **Twitter mood predicts the stock market.** *Journal of computational science*, **2**(1):1–8, 2011. [20](#), [99](#)

REFERENCES

- [20] MM BRADLEY AND PJ LANG. **Affective norms for English Words (ANEW): Affective ratings of words and instruction manual.** *University of Florida, Gainesville. FLTechnical Report C-2*, 2010. [41](#)
- [21] MATTHEW BRAND, NURIA OLIVER, AND ALEX PENTLAND. **Coupled hidden markov models for complex action recognition.** In *Proceedings of IEEE computer society conference on computer vision and pattern recognition*, pages 994–999. IEEE, 1997. [25](#)
- [22] JAMES M CARROLL, MICHELLE SM YIK, JAMES A RUSSELL, AND LISA FELDMAN BARRETT. **On the psychometric principles of affect.** *Review of general psychology*, **3**:14–22, 1999. [xiii](#), [39](#), [40](#)
- [23] CLAUDIO CASTELLANO, SANTO FORTUNATO, AND VITTORIO LORETO. **Statistical physics of social dynamics.** *Reviews of modern physics*, **81**(2):591, 2009. [99](#)
- [24] DUMITRU-CLEMENTIN CERCEL AND STEFAN TRAUSAN-MATU. **Opinion Propagation in Online Social Networks: A Survey.** In *Proceedings of the 4th International Conference on Web Intelligence, Mining and Semantics (WIMS14)*, page 11. ACM, 2014. [98](#)
- [25] MEEYOUNG CHA, HAMED HADDADI, FABRICIO BENEVENUTO, AND P KRISHNA GUMMADI. **Measuring user influence in twitter: The million follower fallacy.** *Icwsn*, **10**(10-17):30, 2010. [20](#), [21](#), [23](#), [30](#), [33](#), [57](#), [62](#), [98](#)
- [26] B. CHAZELLE. **Natural algorithms and influence systems.** *Commun. ACM*, **55**:101–110, 2012. [82](#)

-
- [27] CHENGYAO CHEN, WENJIE LI, DEHONG GAO, AND YUOXIAN HOU. **Exploring interpersonal influence by tracking user dynamic interactions.** *IEEE Intelligent Systems*, **32**(3):28–35, 2017. [26](#)
- [28] CHENGYAO CHEN, ZHITAO WANG, YU LEI, AND WENJIE LI. **Content-based influence modeling for opinion behavior prediction.** In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2207–2216, 2016. [98](#), [99](#)
- [29] CHENGYAO CHEN, ZHITAO WANG, AND WENJIE LI. **Modeling opinion influence with user dual identity.** In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 2019–2022, 2017. [3](#), [31](#), [98](#)
- [30] CHENGYAO CHEN, ZHITAO WANG, AND WENJIE LI. **Tracking Dynamics of Opinion Behaviors with a Content-based Sequential Opinion Influence Model.** *IEEE Transactions on Affective Computing*, 2018. [3](#), [28](#), [31](#), [71](#), [72](#), [98](#), [99](#), [106](#)
- [31] WEI CHEN, CHI WANG, AND YAJUN WANG. **Scalable influence maximization for prevalent viral marketing in large-scale social networks.** In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1029–1038. ACM, 2010. [33](#), [98](#)
- [32] WEI CHEN, YAJUN WANG, AND SIYU YANG. **Efficient influence maximization in social networks.** In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 199–208. ACM, 2009. [30](#)

REFERENCES

- [33] XUETONG CHEN, MARTIN D SYKORA, THOMAS W JACKSON, AND SUZANNE ELAYAN. **What about Mood Swings: Identifying Depression on Twitter with Temporal Measures of Emotions.** In *Companion of the The Web Conference 2018 on The Web Conference 2018*, pages 1653–1660. International World Wide Web Conferences Steering Committee, 2018. [71](#), [99](#)
- [34] 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. [27](#), [28](#), [100](#)
- [35] MUNMUN DE CHOUDHURY, MICHAEL GAMON, AND SCOTT COUNTS. **Happy, Nervous or Surprised? Classification of Human Affective States in Social Media.** In *AAAI International Conference on Weblogs and Social Media, ICWSM, 2012*, 2012. [44](#)
- [36] EMILY M. CODY, ANDREW J. REAGAN, LEWIS MITCHELL, PETER SHERIDAN DODDS, AND CHRISTOPHER M. DANFORTH. **Climate Change Sentiment on Twitter: An Unsolicited Public Opinion Poll.** *PloS one*, **10**(8):e0136092, 2015. [6](#)
- [37] NIKO COLNERIĆ AND JANEZ DEMŠAR. **Emotion Recognition on Twitter: Comparative Study and Training a Unison Model.** *IEEE Transactions on Affective Computing*, 2018. [19](#)
- [38] NIKO COLNERIĆ AND JANEZ DEMŠAR. **Emotion recognition on twitter: Comparative study and training a unison model.** *IEEE transactions on affective computing*, **11**(3):433–446, 2018. [28](#)

-
- [39] JEAN MR COSTA, RAHMTIN ROTABI, ELIZABETH L MURNANE, AND TANZEEM CHOUDHURY. **It is not only about Grievances-Emotional Dynamics in Social Media during the Brazilian Protests.** In *Proceedings of the Ninth International AAAI Conference on Web and Social Media*, pages 594–597, 2015. [33](#)
- [40] ABHIMANYU DAS, SREENIVAS GOLLAPUDI, AND KAMESH MUNAGALA. **Modeling opinion dynamics in social networks.** In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 403–412. ACM, 2014. [29](#), [74](#), [82](#), [98](#), [99](#)
- [41] ABIR DE, SOURANGSHU BHATTACHARYA, PARANTAPA BHATTACHARYA, NILOY GANGULY, AND SOUMEN CHAKRABARTI. **Learning a linear influence model from transient opinion dynamics.** In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 401–410. ACM, 2014. [29](#), [74](#), [83](#), [99](#)
- [42] ABIR DE, ISABEL VALERA, NILOY GANGULY, SOURANGSHU BHATTACHARYA, AND MANUEL GOMEZ RODRIGUEZ. **Learning and forecasting opinion dynamics in social networks.** In *Advances in Neural Information Processing Systems*, pages 397–405, 2016. [29](#), [72](#)
- [43] MORRIS H DEGROOT. **Reaching a consensus.** *Journal of the American Statistical Association*, **69**(345):118–121, 1974. [29](#), [74](#), [81](#), [99](#)
- [44] MICHELA DEL VICARIO, FABIANA ZOLLO, GUIDO CALDARELLI, ANTONIO SCALA, AND WALTER QUATTROCIOCCHI. **Mapping social dynamics on Facebook: The Brexit debate.** *Social Networks*, **50**:6–16, 2017. [72](#)
- [45] LINGJIA DENG AND JANYCE WIEBE. **Sentiment Propagation via Implicature Constraints.** In *EACL*, pages 377–385, 2014. [3](#), [24](#)

REFERENCES

- [46] XIAOWEN DING, BING LIU, AND PHILIP S YU. **A holistic lexicon-based approach to opinion mining.** In *Proceedings of the 2008 international conference on web search and data mining*, pages 231–240. ACM, 2008. [17](#)
- [47] ZHAOYUN DING, YAN JIA, BIN ZHOU, JIANFENG ZHANG, YI HAN, AND CHUNFENG YU. **An influence strength measurement via time-aware probabilistic generative model for microblogs.** In *Asia-Pacific Web Conference*, pages 372–383. Springer, 2013. [33](#), [98](#)
- [48] ZHAOYUN DING, HUI WANG, LIANG GUO, FENGCAI QIAO, JIANPING CAO, AND DAYONG SHEN. **Finding Influential Users and Popular Contents on Twitter.** In *International Conference on Web Information Systems Engineering*, pages 267–275. Springer, 2015. [22](#)
- [49] JIACHEN DU, LIN GUI, YULAN HE, RUIFENG XU, AND XUAN WANG. **Convolution-based neural attention with applications to sentiment classification.** *IEEE Access*, 7:27983–27992, 2019. [28](#)
- [50] NAN DU, HANJUN DAI, RAKSHIT TRIVEDI, UTKARSH UPADHYAY, MANUEL GOMEZ-RODRIGUEZ, AND LE SONG. **Recurrent marked temporal point processes: Embedding event history to vector.** In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1555–1564. ACM, 2016. [29](#)
- [51] KATHLEEN T DURANT AND MICHAEL D SMITH. **Mining sentiment classification from political web logs.** In *Proceedings of Workshop on Web Mining and Web Usage Analysis of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (WebKDD-2006)*, Philadelphia, PA, 2006. [16](#)

REFERENCES

- [52] PAUL EKMAN. **An argument for basic emotions.** *Cognition and Emotion*, **6**(3-4):169–200, 1992. [44](#)
- [53] JEFFREY L ELMAN. **Finding structure in time.** *Cognitive science*, **14**(2):179–211, 1990. [27](#)
- [54] ANDREA ESULI AND FABRIZIO SEBASTIANI. **Sentiwordnet: A publicly available lexical resource for opinion mining.** In *Proceedings of LREC*, **6**, pages 417–422. Citeseer, 2006. [17](#)
- [55] LISA FELDMAN BARRETT AND JAMES A RUSSELL. **Independence and bipolarity in the structure of current affect.** *Journal of personality and social psychology*, **74**(4):967, 1998. [xiii](#), [38](#), [39](#), [40](#), [105](#)
- [56] EMILIO FERRARA AND ZEYAO YANG. **Measuring emotional contagion in social media.** *PloS one*, **10**(11):e0142390, 2015. [19](#)
- [57] NIR FRIEDMAN, DAN GEIGER, AND MOISES GOLDSZMIDT. **Bayesian network classifiers.** *Machine learning*, **29**(2):131–163, 1997. [15](#)
- [58] CHANGZENG FU, CHAORAN LIU, CARLOS TOSHINORI ISHI, AND HIROSHI ISHIGURO. **Multi-modality emotion recognition model with GAT-based multi-head inter-modality attention.** *Sensors*, **20**(17):4894, 2020. [28](#)
- [59] RUI GASPAR, CLÁUDIA PEDRO, PANOS PANAGIOTOPOULOS, AND BEATE SEIBT. **Beyond positive or negative: Qualitative sentiment analysis of social media reactions to unexpected stressful events.** *Computers in Human Behavior*, **56**:179–191, 2016. [2](#), [14](#)

REFERENCES

- [60] BILAL GHANEM, P. ROSSO, AND F. M. R. PARDO. **An Emotional Analysis of False Information in Social Media and News Articles.** *ACM Transactions on Internet Technology (TOIT)*, **20**:1 – 18, 2020. [124](#)
- [61] M GHIASSI AND H SAIDANE. **A dynamic architecture for artificial neural networks.** *Neurocomputing*, **63**:397–413, 2005. [17](#)
- [62] MANOOCHEHR GHIASSI, JAMES SKINNER, AND DAVID ZIMBRA. **Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network.** *Expert Systems with applications*, **40**(16):6266–6282, 2013. [17](#)
- [63] DEEPANWAY GHOSAL, NAVONIL MAJUMDER, SOUJANYA PORIA, NIYATI CHHAYA, AND ALEXANDER GELBUKH. **DialogueGCN: A Graph Convolutional Neural Network for Emotion Recognition in Conversation.** In *EMNLP/IJCNLP*, 2019. [28](#)
- [64] ANASTASIA GIACHANOU AND FABIO CRESTANI. **Like It or Not: A Survey of Twitter Sentiment Analysis Methods.** *ACM Computing Surveys (CSUR)*, **49**(2):28, 2016. [2](#), [16](#), [18](#)
- [65] ANASTASIA GIACHANOU AND FABIO CRESTANI. **Like It or Not: A Survey of Twitter Sentiment Analysis Methods.** *ACM Comput. Surv.*, **49**(2):28:1–28:41, 2016. [71](#)
- [66] ANASTASIA GIACHANOU, JULIO GONZALO, AND FABIO CRESTANI. **Propagating sentiment signals for estimating reputation polarity.** *Information Processing & Management*, **56**(6):102079, 2019. [3](#)

-
- [67] ANASTASIA GIACHANOU, JULIO GONZALO, IDA MELE, AND FABIO CRESTANI. **Sentiment propagation for predicting reputation polarity**. In *European Conference on Information Retrieval*, pages 226–238. Springer, 2017. [24](#)
- [68] ALEC GO, RICHA BHAYANI, AND LEI HUANG. **Twitter sentiment classification using distant supervision**. *CS224N Project Report, Stanford*, 1:12, 2009. [16](#)
- [69] AMIT GOYAL, FRANCESCO BONCHI, AND LAKS VS LAKSHMANAN. **Learning influence probabilities in social networks**. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 241–250. ACM, 2010. [30](#)
- [70] ALEX GRAVES, ABDEL-RAHMAN MOHAMED, AND GEOFFREY HINTON. **Speech recognition with deep recurrent neural networks**. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 6645–6649. IEEE, 2013. [27](#), [28](#)
- [71] QI GU, JR EUGENE SANTOS, AND EUNICE E SANTOS. **Modeling opinion dynamics in a social network**. In *Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume 02*, pages 9–16. IEEE Computer Society, 2013. [33](#), [98](#)
- [72] BIN GUO, Y. DING, L. YAO, YUNJI LIANG, AND Z. YU. **The Future of Misinformation Detection: New Perspectives and Trends**. *ArXiv*, abs/1909.03654, 2019. [124](#)
- [73] MARYAM HASAN, EMMANUEL AGU, AND ELKE RUNDENSTEINER. **Using hashtags as labels for supervised learning of emotions in twitter**

REFERENCES

- messages. In *ACM SIGKDD Workshop on Health Informatics, New York, USA*, 2014. [34](#), [74](#), [100](#)
- [74] RAINER HEGSELMANN, ULRICH KRAUSE, ET AL. **Opinion dynamics and bounded confidence models, analysis, and simulation.** *Journal of artificial societies and social simulation*, **5**(3), 2002. [29](#), [74](#), [82](#), [99](#)
- [75] ROBERT HILLMANN AND MATTHIAS TRIER. **Dissemination patterns and associated network effects of sentiments in social networks.** In *IEEE International Conference on Advances in Social Networks Analysis and Mining, ASONAM, 2012*, pages 511–516, 2012. [33](#)
- [76] DUNG T HO AND TRU H CAO. **A high-order hidden Markov model for emotion detection from textual data.** In *Pacific rim knowledge acquisition workshop*, pages 94–105. Springer, 2012. [25](#)
- [77] SEPP HOCHREITER AND JÜRGEN SCHMIDHUBER. **Long short-term memory.** *Neural computation*, **9**(8):1735–1780, 1997. [27](#)
- [78] DON H HOCKENBURY AND SANDRA E HOCKENBURY. *Discovering psychology*. Macmillan, 2010. [14](#)
- [79] RICHARD A HOLLEY AND THOMAS M LIGGETT. **Ergodic theorems for weakly interacting infinite systems and the voter model.** *The annals of probability*, pages 643–663, 1975. [29](#), [82](#)
- [80] MICHAEL HUNGER. *Neo4j 2.0: Eine Graphdatenbank für alle*. entwickler. Press, 2014. [126](#)
- [81] NITIN INDURKHYA AND FRED J DAMERAU. *Handbook of natural language processing*. CRC Press, 2000. [14](#)

-
- [82] BERNARD J JANSEN, MIMI ZHANG, KATE SOBEL, AND ABDUR CHOWDURY. **Twitter power: Tweets as electronic word of mouth.** *Journal of the Association for Information Science and Technology*, **60**(11):2169–2188, 2009. [23](#)
- [83] WENXIANG JIAO, MICHAEL LYU, AND IRWIN KING. **Real-time emotion recognition via attention gated hierarchical memory network.** In *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, pages 8002–8009, 2020. [28](#)
- [84] FANG JIN, EDWARD R. DOUGHERTY, PARANG SARAF, YANG CAO, AND NAREN RAMAKRISHNAN. **Epidemiological modeling of news and rumors on Twitter.** In *Proceedings of the 7th Workshop on Social Network Mining and Analysis, SNAKDD 2013, Chicago, IL, USA, August 11, 2013*, pages 8:1–8:9, 2013. [23](#)
- [85] WEI JIN, HUNG HAY HO, AND ROHINI K SRIHARI. **A novel lexicalized HMM-based learning framework for web opinion mining.** In *Proceedings of the 26th annual international conference on machine learning*, **10**. Citeseer, 2009. [27](#)
- [86] RINCY JOSE AND VARGHESE S CHOORALIL. **Prediction of election result by enhanced sentiment analysis on twitter data using classifier ensemble Approach.** In *2016 international conference on data mining and advanced computing (SAPIENCE)*, pages 64–67. IEEE, 2016. [26](#)
- [87] MARIA KARAKIZA. **The Impact of Social Media in the Public Sector.** *Procedia - Social and Behavioral Sciences*, **175**, 02 2015. [33](#)
- [88] DAVID KEMPE, JON KLEINBERG, AND ÉVA TARDOS. **Maximizing the spread of influence through a social network.** In *Proceedings of the*

REFERENCES

- ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM, 2003. [30](#)
- [89] FARHAN HASSAN KHAN, SABA BASHIR, AND USMAN QAMAR. **TOM: Twitter opinion mining framework using hybrid classification scheme**. *Decision Support Systems*, **57**:245–257, 2014. [17](#)
- [90] MUHAMMAD TAIMOOR KHAN, MEHR DURRANI, ARMUGHAN ALI, IRUM INAYAT, SHEHZAD KHALID, AND KAMRAN HABIB KHAN. **Sentiment analysis and the complex natural language**. *Complex Adaptive Systems Modeling*, **4**(1):2, 2016. [13](#)
- [91] JIHIE KIM AND JAEBOONG YOO. **Role of sentiment in message propagation: Reply vs. retweet behavior in political communication**. In *Social Informatics (SocialInformatics), 2012 International Conference on*, pages 131–136. IEEE, 2012. [23](#)
- [92] NOO-RI KIM, KYOUNGMIN KIM, AND JEE-HYONG LEE. **Sentiment analysis in microblogs using HMMs with syntactic and sentimental information**. *International Journal of Fuzzy Logic and Intelligent Systems*, **17**(4):329–336, 2017. [26](#)
- [93] SUIN KIM, JINYEONG BAK, AND ALICE HAEYUN OH. **Do You Feel What I Feel? Social Aspects of Emotions in Twitter Conversations**. In *ICWSM*, 2012. [20](#), [21](#), [33](#), [72](#), [98](#)
- [94] SVETLANA KIRITCHENKO, XIAODAN ZHU, AND SAIF M MOHAMMAD. **Sentiment analysis of short informal texts**. *Journal of Artificial Intelligence Research*, **50**:723–762, 2014. [16](#)

-
- [95] SOTIRIS B KOTSIANTIS, I ZAHARAKIS, AND P PINTELAS. **Supervised machine learning: A review of classification techniques**, 2007. [16](#)
- [96] PAUL L KRAPIVSKY AND SIDNEY REDNER. **Dynamics of majority rule in two-state interacting spin systems**. *Physical Review Letters*, **90**(23):238701, 2003. [29](#)
- [97] ALEX KRIZHEVSKY, ILYA SUTSKEVER, AND GEOFFREY E HINTON. **Imagenet classification with deep convolutional neural networks**. In *Advances in neural information processing systems*, pages 1097–1105, 2012. [28](#)
- [98] BHUSHAN KULKARNI, SUMIT AGARWAL, ABIR DE, SOURANGSHU BHATTACHARYA, AND NILOY GANGULY. **SLANT+: A nonlinear model for opinion dynamics in social networks**. In *Data Mining (ICDM), 2017 IEEE International Conference on*, pages 931–936. IEEE, 2017. [29](#), [72](#), [98](#), [99](#)
- [99] HAEWOON KWAK, CHANGHYUN LEE, HOSUNG PARK, AND SUE MOON. **What is Twitter, a social network or a news media?** In *Proceedings of the 19th international conference on World wide web*, pages 591–600. ACM, 2010. [20](#), [21](#), [33](#), [72](#)
- [100] CHANGHYUN LEE, HAEWOON KWAK, HOSUNG PARK, AND SUE MOON. **Finding influentials based on the temporal order of information adoption in twitter**. In *Proceedings of the 19th international conference on World wide web*, pages 1137–1138. ACM, 2010. [30](#)
- [101] KRISTINA LERMAN AND RUMI GHOSH. **Information contagion: An empirical study of the spread of news on Digg and Twitter social networks**. *Icwsm*, **10**:90–97, 2010. [30](#)

REFERENCES

- [102] Y. LI, WEI CHEN, Y. WANG, AND ZHI-LI ZHANG. **Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships.** *ArXiv*, abs/1111.4729, 2013. [82](#)
- [103] YUNG-MING LI AND YA-LIN SHIU. **A diffusion mechanism for social advertising over microblogs.** *Decision Support Systems*, **54**(1):9–22, 2012. [33](#), [72](#), [98](#)
- [104] ZHINONG LI, ZHAOTONG WU, YONGYONG HE, AND CHU FULEI. **Hidden Markov model-based fault diagnostics method in speed-up and speed-down process for rotating machinery.** *Mechanical Systems and Signal Processing*, **19**(2):329–339, 2005. [27](#)
- [105] PANAGIOTIS LIAKOS, KATIA PAPAKONSTANTINOPOULOU, MICHAEL SIOUTIS, KONSTANTINOS TSAKALOZOS, AND ALEX DELIS. **Pinpointing Influence in Pinterest.** In *SocInf@IJCAI*, pages 26–37, 2016. [22](#)
- [106] BING LIU. **Sentiment analysis and subjectivity.** In *Handbook of Natural Language Processing, Second Edition*, pages 627–666. Chapman and Hall/CRC, 2010. [2](#), [14](#)
- [107] BING LIU. **Sentiment analysis and opinion mining.** *Synthesis lectures on human language technologies*, **5**(1):1–167, 2012. [14](#)
- [108] L. LIU, DASHI LUO, MING LIU, JUN ZHONG, YE WEI, AND LETIAN SUN. **A Self-Adaptive Hidden Markov Model for Emotion Classification in Chinese Microblogs.** *Mathematical Problems in Engineering*, **2015**:1–8, 2015. [26](#)

-
- [109] RADA MIHALCEA AND HUGO LIU. **A Corpus-based Approach to Finding Happiness.** In *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs, CAAW, 2006*, pages 139–144, 2006. [15](#)
- [110] YASUHIDE MIURA, MOTOKI TANIGUCHI, TOMOKI TANIGUCHI, AND TOMOKO OHKUMA. **Unifying text, metadata, and user network representations with a neural network for geolocation prediction.** In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1260–1272, 2017. [98](#)
- [111] SAIF M MOHAMMAD. **Sentiment analysis: Detecting valence, emotions, and other affectual states from text.** *Emotion Measurement*, pages 201–238, 2015. [2](#) [14](#)
- [112] SAIF M MOHAMMAD. **Sentiment analysis: Detecting valence, emotions, and other affectual states from text.** In *Emotion measurement*, pages 201–237. Elsevier, 2016. [31](#)
- [113] SAIF M. MOHAMMAD AND SVETLANA KIRITCHENKO. **Using Hashtags to Capture Fine Emotion Categories from Tweets.** *Computational Intelligence*, pages 301–326, 2015. [44](#)
- [114] SAIF M MOHAMMAD AND TONY WENDA YANG. **Tracking sentiment in mail: How genders differ on emotional axes.** In *Proceedings of the 2nd workshop on computational approaches to subjectivity and sentiment analysis*, pages 70–79. Association for Computational Linguistics, 2011. [14](#)
- [115] SETH A MYERS, CHENGUANG ZHU, AND JURE LESKOVEC. **Information diffusion and external influence in networks.** In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 33–41. ACM, 2012. [19](#) [23](#) [30](#)

REFERENCES

- [116] ZHU NANLI, ZOU PING, LI WEIGUO, AND CHENG MENG. **Sentiment analysis: A literature review**. In *Management of Technology (ISMOT), 2012 International Symposium on*, pages 572–576. IEEE, 2012. [14](#)
- [117] KEITARO NARUSE. **Modeling and analysis of information propagation in multilayered communities in twitter**. In *Soft Computing and Intelligent Systems (SCIS) and 13th International Symposium on Advanced Intelligent Systems (ISIS), 2012 Joint 6th International Conference on*, pages 1839–1843. IEEE, 2012. [23](#)
- [118] PRADEEP NATARAJAN AND RAMAKANT NEVATIA. **Coupled hidden semi markov models for activity recognition**. In *2007 IEEE Workshop on Motion and Video Computing (WMVC'07)*, pages 10–10. IEEE, 2007. [25](#)
- [119] RAVINDRA NATH AND RENU JAIN. **Insulin Chart Prediction for Diabetic Patients Using Hidden Markov Model (HMM) and Simulated Annealing Method**. In *Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), December 28-30, 2012*, pages 3–11. Springer, 2014. [27](#)
- [120] HUNG T NGUYEN, PREETAM GHOSH, MICHAEL L MAYO, AND THANG N DINH. **Social influence spectrum at scale: Near-optimal solutions for multiple budgets at once**. *ACM Transactions on Information Systems (TOIS)*, **36**(2):14, 2017. [22](#)
- [121] NGUYET NGUYEN. **An analysis and implementation of the hidden Markov model to technology stock prediction**. *Risks*, **5**(4):62, 2017. [27](#)
- [122] NGUYET NGUYEN AND DUNG NGUYEN. **Hidden markov model for stock selection**. *Risks*, **3**(4):455–473, 2015. [27](#)

-
- [123] FINN ÅRUP NIELSEN. **A new ANEW: Evaluation of a word list for sentiment analysis in microblogs.** In *ESWC Workshop on 'Making Sense of Microposts': Big things come in small packages, MSM, 2011*, pages 93–98, 2011. [41](#)
- [124] BRENDAN O'CONNOR, RAMNATH BALASUBRAMANYAN, BRYAN R ROUTLEDGE, AND NOAH A SMITH. **From tweets to polls: Linking text sentiment to public opinion time series.** *ICWSM*, 11(122-129):1–2, 2010. [19](#), [99](#)
- [125] ADITYA PAL AND SCOTT COUNTS. **Identifying topical authorities in microblogs.** In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 45–54. ACM, 2011. [33](#), [98](#)
- [126] WEI PAN, WEN DONG, MANUEL CEBRIAN, TAEMIE KIM, JAMES H FOWLER, AND ALEX SANDY PENTLAND. **Modeling dynamical influence in human interaction: Using data to make better inferences about influence within social systems.** *IEEE Signal Processing Magazine*, 29(2):77–86, 2012. [30](#), [98](#)
- [127] BO PANG AND LILLIAN LEE. **Opinion mining and sentiment analysis.** *Foundations and trends in information retrieval*, 2(1-2):1–135, 2008. [13](#), [14](#), [15](#), [71](#)
- [128] BO PANG, LILLIAN LEE, AND SHIVAKUMAR VAITHYANATHAN. **Thumbs up?: sentiment classification using machine learning techniques.** In *ACL conference on Empirical methods in natural language processing, EMNLP, 2002*, pages 79–86, 2002. [16](#)

REFERENCES

- [129] SANCHENG PENG, GUOJUN WANG, AND DONGQING XIE. **Social influence analysis in social networking big data: opportunities and challenges.** *IEEE Network*, **31**(1):11–17, 2017. [20](#)
- [130] MATTHEW PURVER AND STUART BATTERSBY. **Experimenting with Distant Supervision for Emotion Classification.** In *ACL Conference of the European Chapter of the Association for Computational Linguistics, EACL, 2012*, pages 482–491, 2012. [44](#)
- [131] ASHEQUL QADIR AND ELLEN RILOFF. **Learning Emotion Indicators from Tweets: Hashtags, Hashtag Patterns, and Phrases.** In *Conference on Empirical Methods in Natural Language Processing, EMNLP, 2014*, pages 1203–1209, 2014. [44](#)
- [132] LAWRENCE R RABINER. **A tutorial on hidden Markov models and selected applications in speech recognition.** *Proceedings of the IEEE*, **77**(2):257–286, 1989. [74](#)
- [133] FUJI REN AND YE WU. **Predicting user-topic opinions in twitter with social and topical context.** *IEEE Transactions on Affective Computing*, **4**(4):412–424, 2013. [72](#)
- [134] HAIBING REN AND GUANGYOU XU. **Human action recognition with primitive-based coupled-HMM.** In *Object recognition supported by user interaction for service robots*, **2**, pages 494–498. IEEE, 2002. [25](#)
- [135] YAFENG REN, YUE ZHANG, MEISHAN ZHANG, AND DONGHONG JI. **Context-sensitive twitter sentiment classification using neural network.** In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016. [28](#)

REFERENCES

- [136] FABIÁN RIQUELME AND PABLO GONZÁLEZ-CANTERGIANI. **Measuring user influence on Twitter: A survey.** *Information Processing & Management*, **52**(5):949–975, 2016. [3](#), [98](#)
- [137] JAMES A RUSSELL. **A circumplex model of affect.** *Journal of personality and social psychology*, **39**(6):1161–1178, 1980. [38](#), [39](#), [40](#)
- [138] JAMES A RUSSELL. **Core affect and the psychological construction of emotion.** *Psychological review*, **110**(1):145–172, 2003. [41](#), [79](#)
- [139] SAMIR RUSTAMOV, ELSHAN MUSTAFAYEV, AND MARK A CLEMENTS. **Sentiment analysis using Neuro-Fuzzy and Hidden Markov models of text.** In *2013 Proceedings of IEEE Southeastcon*, pages 1–6. IEEE, 2013. [26](#)
- [140] ELДАР САДИКОВ AND MARIA MONTSERRAT MEDINA MARTINEZ. **Information propagation on Twitter.** *CS322 Project Report*, 2009. [23](#)
- [141] DIEGO SAEZ-TRUMPER, GIOVANNI COMARELA, VIRGÍLIO ALMEIDA, RICARDO BAEZA-YATES, AND FABRÍCIO BENEVENUTO. **Finding trendsetters in information networks.** In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1014–1022. ACM, 2012. [22](#)
- [142] HASSAN SAIF, YULAN HE, MIRIAM FERNANDEZ, AND HARITH ALANI. **Semantic patterns for sentiment analysis of Twitter.** In *International Semantic Web Conference*, pages 324–340. Springer, 2014. [17](#)
- [143] KAZUMI SAITO, MASAHIRO KIMURA, KOUZOU OHARA, AND HIROSHI MOTODA. **Selecting information diffusion models over social networks**

REFERENCES

- for behavioral analysis.** In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 180–195. Springer, 2010. [30](#)
- [144] JUAN SHI, PING HU, KIN KEUNG LAI, AND GANG CHEN. **Determinants of users? information dissemination behavior on social networking sites: an elaboration likelihood model perspective.** *Internet Research*, 28(2):393–418, 2018. [20](#), [33](#), [72](#), [98](#)
- [145] ARLEI SILVA, SARA GUIMARÃES, WAGNER MEIRA JR, AND MOHAMMED ZAKI. **ProfileRank: finding relevant content and influential users based on information diffusion.** In *Proceedings of the 7th Workshop on Social Network Mining and Analysis*, page 2. ACM, 2013. [22](#)
- [146] SWATI SONI AND AAKANKSHA SHARAFF. **Sentiment analysis of customer reviews based on hidden markov model.** In *Proceedings of the 2015 International Conference on Advanced Research in Computer Science Engineering & Technology (ICARCSET 2015)*, pages 1–5, 2015. [27](#)
- [147] ALESSANDRO SORDONI, YOSHUA BENGIO, HOSSEIN VAHABI, CHRISTINA LIOMA, JAKOB GRUE SIMONSEN, AND JIAN-YUN NIE. **A hierarchical recurrent encoder-decoder for generative context-aware query suggestion.** In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pages 553–562. ACM, 2015. [101](#)
- [148] MICHAEL SPERIOSU, NIKITA SUDAN, SID UPADHYAY, AND JASON BALDRIDGE. **Twitter polarity classification with label propagation over lexical links and the follower graph.** In *ACL workshop on Unsupervised Learning in NLP@EMNLP*, pages 53–63. Association for Computational Linguistics, 2011. [18](#)

- [149] STEFAN STIEGLITZ AND LINH DANG-XUAN. **Emotions and information diffusion in social media?sentiment of microblogs and sharing behavior.** *Journal of management information systems*, **29**(4):217–248, 2013. [5](#), [19](#), [23](#), [33](#), [72](#), [98](#)
- [150] BONGWON SUH, LICHAN HONG, PETER PIROLLO, AND ED H. CHI. **Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network.** In *(IEEE) International Conference on Privacy, Security, Risk and Trust, PASSAT 2010, Minneapolis, Minnesota, USA, 2010*, pages 177–184, 2010. [5](#), [23](#), [63](#)
- [151] BONGWON SUH, LICHAN HONG, PETER PIROLLO, AND ED H CHI. **Want to be retweeted? large scale analytics on factors impacting retweet in twitter network.** In *2010 IEEE Second International Conference on Social Computing*, pages 177–184. IEEE, 2010. [19](#)
- [152] BEIMING SUN AND VINCENT TY NG. **Identifying influential users by their postings in social networks.** In *Ubiquitous social media analysis*, pages 128–151. Springer, 2013. [33](#), [98](#)
- [153] JOHAN AK SUYKENS AND JOOS VANDEWALLE. **Least squares support vector machine classifiers.** *Neural processing letters*, **9**(3):293–300, 1999. [15](#)
- [154] P. M. SWAMIDASS, editor. [MAPE \(mean absolute percentage error\)MEAN ABSOLUTE PERCENTAGE ERROR \(MAPE\)](#), pages 462–462. Springer US, Boston, MA, 2000. [84](#)
- [155] MAITE TABOADA, JULIAN BROOKE, MILAN TOFILOSKI, KIMBERLY VOLL, AND MANFRED STEDE. **Lexicon-based methods for sentiment analysis.** *Computational linguistics*, **37**(2):267–307, 2011. [18](#)

REFERENCES

- [156] JIE TANG AND ACM FONG. **Sentiment diffusion in large scale social networks.** In *IEEE International Conference on Consumer Electronics, ICCE, 2013*, pages 244–245, 2013. [34](#)
- [157] JILIANG TANG, CHIKASHI NOBATA, ANLEI DONG, YI CHANG, AND HUAN LIU. **Propagation-based sentiment analysis for microblogging data.** In *Proceedings of the 2015 SIAM International Conference on Data Mining*, pages 577–585. SIAM, 2015. [3](#), [24](#)
- [158] MINGSHENG TANG, XINJUN MAO, SHUQIANG YANG, AND HUIPING ZHOU. **A dynamic microblog network and information dissemination in ?@? mode.** *Mathematical Problems in Engineering*, **2014**, 2014. [24](#)
- [159] MIKE THELWALL, KEVAN BUCKLEY, GEORGIOS PALTOGLOU, DI CAI, AND ARVID KAPPAS. **Sentiment strength detection in short informal text.** *Journal of the American Society for Information Science and Technology*, **61**(12):2544–2558, 2010. [17](#)
- [160] JITHENDER J TIMOTHY. **How does propaganda influence the opinion dynamics of a population?** *arXiv preprint arXiv:1703.10138*, 2017. [98](#)
- [161] DUC NGUYEN TRUNG, JASON J JUNG, ATTILA KISS, ET AL. **Towards modeling fuzzy propagation for sentiment analysis in online social networks: A case study on TweetScope.** In *2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom)*, pages 331–338. IEEE, 2013. [3](#), [24](#)
- [162] ANDREA VANZO, DANILLO CROCE, AND ROBERTO BASILI. **A context-based model for sentiment analysis in twitter.** In *Proceedings of coling 2014, the 25th international conference on computational linguistics: Technical papers*, pages 2345–2354, 2014. [26](#)

-
- [163] ALI VARDASBI, HESHAAM FAILI, AND MASOUD ASADPOUR. **SWIM: Stepped weighted shell decomposition influence maximization for large-scale networks.** *ACM Transactions on Information Systems (TOIS)*, **36**(1):6, 2017. [22](#)
- [164] SIDA WANG AND CHRISTOPHER D MANNING. **Baselines and bigrams: Simple, good sentiment and topic classification.** In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*, pages 90–94. Association for Computational Linguistics, 2012. [16](#)
- [165] 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*, pages 1031–1040. ACM, 2011. [18](#)
- [166] XINZHI WANG, HUI ZHANG, SHENGCHENG YUAN, JIAYUE WANG, AND YANG ZHOU. **Sentiment processing of social media information from both wireless and wired network.** *EURASIP Journal on Wireless Communications and Networking*, **2016**(1):1–13, 2016. [26](#)
- [167] YANHAO WANG, YUCHEN LI, JU FAN, AND KIAN-LEE TAN. **Location-aware influence maximization over dynamic social streams.** *ACM Transactions on Information Systems (TOIS)*, **36**(4):43, 2018. [22](#)
- [168] YONGQING WANG, SHENGHUA LIU, HUawei SHEN, JINHUA GAO, AND XUEQI CHENG. **Marked temporal dynamics modeling based on recurrent neural network.** In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 786–798. Springer, 2017. [29](#)

REFERENCES

- [169] DUNCAN J WATTS AND PETER SHERIDAN DODDS. **Influentials, networks, and public opinion formation.** *Journal of consumer research*, **34**(4):441–458, 2007. [30](#), [63](#)
- [170] BRIAN E WEEKS, ALBERTO ARDÈVOL-ABREU, AND HOMERO GIL DE ZÚÑIGA. **Online influence? Social media use, opinion leadership, and political persuasion.** *International Journal of Public Opinion Research*, **29**(2):214–239, 2017. [33](#), [72](#)
- [171] JIANSHU WENG, EE-PENG LIM, JING JIANG, AND QI HE. **Twitterrank: finding topic-sensitive influential twitterers.** In *Proceedings of the third ACM international conference on Web search and data mining*, pages 261–270. ACM, 2010. [21](#)
- [172] LILLIAN WENG, ALESSANDRO FLAMMINI, ALESSANDRO VESPIGNANI, AND FILLIPO MENCZER. **Competition among memes in a world with limited attention.** *Scientific reports*, **2**:335, 2012. [23](#)
- [173] PAUL J WERBOS. **Backpropagation through time: what it does and how to do it.** *Proceedings of the IEEE*, **78**(10):1550–1560, 1990. [105](#)
- [174] THERESA WILSON, JANYCE WIEBE, AND PAUL HOFFMANN. **Recognizing contextual polarity in phrase-level sentiment analysis.** In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pages 347–354. Association for Computational Linguistics, 2005. [17](#)
- [175] GADI WOLFSFELD, ELAD SEGEV, AND TAMIR SHEAFER. **Social media and the Arab Spring: Politics comes first.** *The International Journal of Press/Politics*, **18**(2):115–137, 2013. [33](#)

-
- [176] BO WU AND HAIYING SHEN. **Analyzing and predicting news popularity on Twitter.** *International Journal of Information Management*, **35**(6):702–711, 2015. [24](#)
- [177] HUI-HSIN WU, ANGELA CHARNG-RURNG TSAI, RICHARD TZONG-HAN TSAI, AND JANE YUNG-JEN HSU. **Sentiment value propagation for an integral sentiment dictionary based on commonsense knowledge.** In *2011 International Conference on Technologies and Applications of Artificial Intelligence*, pages 75–81. IEEE, 2011. [3](#), [24](#)
- [178] Y WU, A GANAPATHIRAJU, AND J PICONE. **Baum-welch re-estimation of hidden markov model.** *Department of Electrical and Computer Engineering, Mississippi State University*, 1999. [74](#)
- [179] SHUAI XIAO, JUNCHI YAN, XIAOKANG YANG, HONGYUAN ZHA, AND STEPHEN M CHU. **Modeling the Intensity Function of Point Process Via Recurrent Neural Networks.** In *AAAI*, **17**, pages 1597–1603, 2017. [29](#)
- [180] XIAOBING XIONG, GANG ZHOU, YONGZHONG HUANG, HAIYONG CHEN, AND KE XU. **Dynamic evolution of collective emotions in social networks: a case study of Sina weibo.** *Science China Information Sciences*, **56**(7):1–18, 2013. [99](#)
- [181] MOHAMED YASSINE AND HAZEM HAJJ. **A framework for emotion mining from text in online social networks.** In *IEEE International Conference on Data Mining Workshops, (ICDMW), 2010*, pages 1136–1142, 2010. [15](#)

REFERENCES

- [182] SHAOZHI YE AND SHYHTSUN FELIX WU. **Measuring message propagation and social influence on twitter. com.** *SocInfo*, **10**:216–231, 2010. [20](#), [23](#)
- [183] MICHELLE SM YIK, JAMES A RUSSELL, AND LISA FELDMAN BARRETT. **Structure of self-reported current affect: Integration and beyond.** *Journal of personality and social psychology*, **77**(3):600, 1999. [40](#)
- [184] REZA ZAFARANI, WILLIAM COLE, AND HUAN LIU. **Sentiment propagation in social networks: a case study in livejournal.** *Advances in Social Computing*, pages 413–420, 2010. [24](#)
- [185] SAVVAS ZANNETTOU, M. SIRIVIANOS, J. BLACKBURN, AND NICOLAS KOURTELLIS. **The Web of False Information.** *Journal of Data and Information Quality (JDIQ)*, **11**:1 – 37, 2019. [124](#)
- [186] DAN ZARRELLA. **The science of retweets.** *Retrieved December*, **15**:2009, 2009. [22](#)
- [187] JINXUE ZHANG, RUI ZHANG, JINGCHAO SUN, YANCHAO ZHANG, AND CHI ZHANG. **Truetop: A sybil-resilient system for user influence measurement on twitter.** *IEEE/ACM Transactions on Networking*, **24**(5):2834–2846, 2016. [22](#)
- [188] LUMIN ZHANG, YAN JIA, XIANG ZHU, BIN ZHOU, AND YI HAN. **User-level sentiment evolution analysis in microblog.** *China Communications*, **11**(12):152–163, 2014. [72](#), [99](#)
- [189] YUCHI ZHANG, WENDY W MOE, AND DAVID A SCHWEIDEL. **Modeling the role of message content and influencers in social media rebroad-**

- casting.** *International Journal of Research in Marketing*, **34**(1):100–119, 2017. [19](#), [21](#)
- [190] XIAOYI ZHAO AND YUKIO OHSAWA. **Sentiment analysis on the on-line reviews based on hidden Markov model.** *Journal of Advances in Information Technology*, **9**(2), 2018. [27](#)
- [191] LIXING ZHU, YULAN HE, AND DEYU ZHOU. **Neural opinion dynamics model for the prediction of user-level stance dynamics.** *Information Processing & Management*, **57**(2):102031, 2020. [3](#)
- [192] CAI-NICOLAS ZIEGLER AND GEORG LAUSEN. **Propagation models for trust and distrust in social networks.** *Information Systems Frontiers*, **7**(4-5):337–358, 2005. [33](#)

Sentiment analysis in social networks has been widely analysed over the last decade. Despite the amount of research done in sentiment analysis in social networks, the distinct categories are not appropriately considered in many cases, and the study of dissemination patterns of emotions is limited. Therefore, understanding the significance of specific emotions will be more beneficial for various marketing activities, policy-making decisions and political campaigns.

The current PhD thesis focuses on designing a theoretical framework for analyzing the broad spectrum of sentiments and explain how emotions are propagated using concepts from temporal and multilayer networks. More precisely, our goal is to provide insights into emotion influence modelling that solves emotion estimation problems and its temporal dynamics nature on social conversation. To exhibit the efficacy of the proposed model, we have collected posts related to different events from Twitter and build a temporal network structure over the conversation.

Firstly, we perform sentiment analysis with the adaptation of a lexicon-based approach and the circumplex model of affect that enhances the effectiveness of the sentiment characterization. Subsequently, we investigate the social dynamics of emotion present in users' opinions by analyzing different social influential characteristics. Next, we design a temporal emotion-based stochastic model in order to investigate the engagement pattern and predict the significant emotions. Our ultimate contribution is the development of a sequential emotion-based influence model with the advancement of recurrent neural networks. It offers to predict emotions in a more comprehensive manner.

Finally, the document presents some conclusions and also outlines future research directions.