Moment-to-Moment Mood Change Modelling in Mobile Mental Health Network

ALANOUD ALHARBI
N0705969

A thesis submitted in partial fulfilment of the requirements of Nottingham Trent University for the degree of Doctor of Philosophy

January, 2022

Abstract

Human interests and behaviour change over time and often affected by multiple factors. In particular, human emotions, mood and its constituent processes change and interact over time. Therefore, modelling human behaviour should take into account the changes over time for customization and adaptation of systems to the users' specific needs. Understanding and assessing the temporal dynamics of mood are critical for modelling human behaviour for both individuals and group of people who share similar habits, life style and personal circumstances. Thus, in order to construct a personalized recommendation for a given user, it is first necessary to have some knowledge about previous user interests and behaviours. However, the challenge of obtaining large-scale data on human emotions has left the most fundamental questions on emotions less explored: How do emotions vary across individuals, evolve over time, and are connected to social ties? We address these questions using a large-scale dataset of users that contains both their users' interactions with momentary emotions and topical labels. Using this dataset, we identify patterns of human emotions on different levels, starting from the network level, group-level (cluster) and moving towards the user level. At the user-level, we identify how human emotions are distributed and vary over time. In particular, we model changes in mood using multi-level multimodal features including users' sentimental status, engagement and linguistic queries. We also utilise language models to model and understand patterns of mood change. We model the changes of users' mental states based on replies and responses to posts over time and predict future states. We find that the future mental states can be predicted

with reasonable accuracy given users' historical posts, current participation features. Our findings form a step forward towards better understand the interplay between user behaviour and mood change exhibited while interacting on mental health network and providing some interpretable summaries that can be used in the future by health experts and individuals and work on possible medical interventions together with clinical experts.

Acknowledgements

I would like to express my sincere gratitude to Qassim University for sponsoring me and Nottingham Trent University for providing me with this opportunity to complete the PhD.

I would not have been able to go through the past years without the immense support of my husband and mother who have stepped up and helped me when I needed them the most; this PhD would not have been possible without them. Of course, a big thanks to my kids, family and friends for your ongoing prayers as that has sustained me this far.

Most of all, I would like to thank Professor Eiman Kanjo and Professor Ahmad Lotfi, my supervisors for their understanding, patience and encouragement and for pushing me to do my best throughout the years.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

Behavior has been defined as a range of actions and mannerisms made by individuals, and it plays an essential role in people's lives and health. For example, smoking and a poor diet can contribute to many diseases. Human behavior is a general term that includes various aspects of people's lives, including eating, movement, and technology-use behaviors. In addition, the advent of the computing revolution and the internet engaged multiple types of behaviors, including internet-use browsing behaviors and social networks.

To suggest behavioral change interventions that can improve well-being, we need to understand human behavior. Creating and modifying a conceptual understanding of user behavior is called "User Modeling" (UM). For a given user, modeling infers unobservable information from observable behaviors, including the user's actions and utterances. In addition, user modeling inferences a user's preferences, goals, and future activities and locations from understanding that user's behavior. This understanding can help to develop practical actions or behavioral change interventions. For example, Rothengatter [189] modeled road-use behavior to establish suitable measures to influence driver behavior. Furthermore, modeling user behavior in networked games has led to providing and charging for the quality of network services [95].

Monitoring and assessing mental health and emotion through smart devices and social media is an active and promising research area [217]. A lot of work in this area aims to distinguish between users of different types (e.g., users with a condition vs. controls). Consequently, it increases the ability of computer devices to improve the quality of people's lives by understanding them and providing intuitive interventions. In recent years, with the advances in technologies and smartphones with various computational and communicational capabilities, specifically in social networking platforms, the opportunities to model people and their behavior and emotion have been raised. As smartphone ownership and social media users have increased exponentially, massive data about users become available. The improved user data enhance the ability to understand, model people and their emotions, and then contribute to their lives by introducing the help and suggesting behavioral change interventions to improve their mental well-being. The improved user data exhibits a range of attractive properties, facilitating the delivery of state-of-the-art models and methods to monitor and model users. Another line of work aims to use data derived from users (e.g., mobile phone sensor data) to predict a target value such as mood or stress scores based on a psychometric test or discrete labels.

People with mental health issues turn to online forums or mobile social media networks not only for advice but also to find people who will listen and provide emotional support. Previous research has indicated that the dynamic of emotional status can be studied through the change in patterns of an individual's posts overtime [217]. Social media's growing popularity has fundamentally introduced social media as a tool for (mental) health and behavior monitoring. As these networks are considered as a critical user information source, this provides opportunities to model users and emotions on a large scale. Furthermore, people interact and describe their emotional experiences in text through the network; therefore, the text is the primary source of data from social networks. It is rich in both emotion and subjectivity. Given that background, this work will focus on modeling users and emotion based on text, which is a part of affective computing [224].

People living with mental health conditions have a smaller social network than the general population. They hence face a higher risk of loneliness and isolation, even though the benefit of social support in mental health and psychological wellbeing is well-documented [13]. For instance, studies have found that people with higher informal social support are more likely to recover from psychiatric symptoms [34]. Furthermore, social support may have enhanced mental wellbeing by promoting a sense of meaning and purpose in life. In addition, social connections could foster a sense of responsibility and concern for others, which leads individuals to care for others and their own health issues.

Beyond controlled experiments, researchers have examined social interactions that happened naturally "in the wild" on social media platforms to infer some understanding of health behavior expressed online. A growing number of people are now describing their healthcare experiences on platforms such as Facebook and Twitter. The increasing availability of patients' health accounts online in a public forum presents an unprecedented opportunity to gather new forms of health data and advance the patient-centered care agenda. Social media provides online communication in real-time at low cost, which allows researchers to monitor and track disease outbreaks [202], detect the public's moods and responses to certain health issues, and through these identify target communities to disseminate health information. We believe that social media can be used as a tool for mental health monitoring, where population data can be captured, serving as a record and sensor for events in peoples' lives [53].

More specifically, linguistic features, interaction patterns, and sentiment analyses of social media posts can offer a novel methodology for measuring and understanding mood changes when engaging with online peer-to-peer mental health support. For example, Pruksachatkun et al. [181] proposed a model that can predict whether a conversation thread in an online mental health forum is associated with 'cognitive change' where people reframe their negative belief into a positive one. Various work has also been done using social media data to predict people's suicidal ideation, depression level, and sleep issues. These studies examine four categories of features: i) user profile, e.g., age, gender, length of membership, number of posts and followers; ii) post characteristics, e.g., linguistic characteristics of texts, analysis of images; iii) patterns of engagement, e.g., the time gap between post, passive use (browsing) vs. active use (posting); iv) relational activity, e.g., social network analysis of the interaction between users. Studies have found that features relating to how social media are used (i.e., patterns of engagement and relational activity, and to a specific extent post characteristics) are often more indicative of people's wellbeing. For instance, it was found that passive forms of use contribute to a depressive mood. It is claimed that passively browsing others' photos of vacations might have triggered envy and loneliness. While social relationships are an essential source of emotional support, they can also be stressful. A systematic review of social media use in adolescents highlighted potential harms such as social isolation, depression, and cyber-bullying. Possible adverse effects on people using social media for healthcare have also been raised, including loss of privacy and being targeted for advertising.

Therefore, it is important that we understand how mental health social media is used to maximize the benefits and mitigate the negative impacts. Concretely within the context of this study, we will investigate the use of social media data to characterize negative and positive changes in people living with mental health problems, along the four features dimensions: user profile, post characteristics, patterns of engagement, and relational activity.

The current study aims to address the following research questions:

- Do mental health social networks enable us to model people's mental wellbeing and the change in the mood during the time(i.e., do people feel more positive or negative over time?)
- If so, what factors/features contribute in modelling users and what factors/features are driving this difference/change, and how can this inform better social media design for mental healthcare?. In addition, do all users interact similarly, and do they get impacted by the same triggers?

Advances in deep learning present new opportunities for the inference of wellbeing by alleviating the need for manual feature extraction. When combined with physical manipulation tools, it opens the door for new natural interactions and responsive interventions. This work will take advantage of the latest developments in deep learning, natural language processing, and user modeling to understand and model user behavior when interacting on Mental Health Social Network and its impact on their emotion and wellbeing.

1.2 Aims and Objectives

The overall aim of the present research is firstly to utilize existing theory and techniques in user modeling, Natural Language Processing and deep learning to understand and model user behavior on Mental health social networks. In addition, we also want to investigate the patterns of mood changes triggered while interacting with their peers to enable a variety of interventions and recommendations. This thesis then explores a real and live large dataset from a mobile social network platform using a range of state-of-the-art multi-level features extraction, feature fusion, and deep learning classifiers. To achieve the project aim, the identified objectives include:

- To extract various multi-model features at multi-levels based on users' interests, language use, app interaction, and communication with others
- To investigate efficient affective classification models, including the transfer learning with pre-trained language models.
- Explore the potential of Deep Learning in studying the temporal dynamics of users' mood over time.
- Explore the capabilities to build robust longitudinal NLP models to capture temporal dynamics and changes in language use and other online behavior over time.
- To explore the ability of modeling users of social networks using unsupervised approaches.

1.3 Key Contributions of the thesis

My Ph.D. project will model user emotions using interacting with Mental health social networks, and It will also seek to use these data to explore personalise user modeling. The main contributions of this thesis are as follows:

- We have processed and manipulated a large dataset acquired from an active Mobile Mental Health Network to characterize and model users' emotions.

- We have proposed deep learning architecture to classify emotion based on the text of mental health dataset social network dataset. A range of Machine learning and deep learning models, including SVM, LSTM, GRU, and others, have been trained using a huge labeled dataset. Moreover, the application of the state-of-the-art pre-trained Language models to explore the impact of these networks have on users' emotions.
- We have developed multilevel feature extraction and multi-model data fusion from unstructured Mobile Mental Health Network to improve emotion analysis from unstructured social network text feed. This will include the merging of the following features: linguistic, static (e.g., demographic), meta, and asynchronous information (e.g., network metrics) that can represent a user or a feature in a longitudinal fashion and study their effectiveness in capturing mental well-being in an intra-user and in an inter-user validation setting.
- We have investigated novel longitudinal NLP models for capturing changes in language use and other online behavior over time as proxies for assessing mental well-being. Moreover, our model predicts the direction of change to positive or negative emotion, and the influence of peers on the user mood has been explored.
- We have provided interpretable summaries of these findings that can be used by health experts and individuals and work on possible medical interventions together with clinical experts.
- We have built a user model to identify normal states and anomalies for individuals by considering changes in their emotional, network interaction style, and language use over time. This will be achieved by proposing a novel technique employing BERT's Users embeddings to cluster users.

1.4 Thesis Outline

This thesis consists of eight chapters that are summarised as follows:

Chapter 2 provides a literature review exploring the concept of user emotion, modeling user emotion, and user modeling in addition to methods to monitor emotions and model users using the text approaches.

Chapter 3 introduces the Mental health social network dataset, explores the network's details, and proposes a detailed analysis.

Chapter 4 explores the classification of emotion based on text data using a range of deep learning classifiers. Furthermore, a state-of-the-art pre-trained Language model is devised to model social network users' emotions as it improves performance by understanding the language.

Chapter 5 of this thesis investigates the fusing of Multi-feature with Language models to improve the classification of emotion based on unstructured social media text. This approach helps to reduce the traditionally challenging of social networks dataset to train the language models.

Chapter 6 explores the modeling temporal Dynamics of users and predicting change in mental wellbeing based on their activities in the network.

Chapter 7 investigates the potential of modeling mental health networks by exploring the use of unsupervised learning in addition to language models.

Chapter 8 concludes the work with a summary of each contribution and formulates future researchs.

Chapter 2

Literature Review

The following chapter will review the related work regarding the modelling users of social network starting with modelling emotion from social network, followed by modelling the change of emotion, then reviewing advances in modelling users of social network.

2.1 User emotions

It is most likely that there is no specific definition of emotion. Kleinginna [116] suggests a definition with a rather broad consensus: 'Emotions describe a complex set of interactions between subjective and objective variables that are mediated by neural and hormonal systems, which can (a) give rise to affective experiences of emotional valence (pleasure-displeasure) and emotional arousal (high-low activation/calming-arousing); (b) generate cognitive processes such as emotionally relevant perceptual affect, appraisals, labeling processes; (c) activate widespread psychological and physiological changes to the arousing conditions; and (d) motivate behavior that is often but not always expressive,) goal-directed and adaptive' [116]. Accordingly, Panksepp [168] suggested the following definition: 'Emotions are the psych-neural processes that are influential in controlling the vigor and patterning of actions in the dynamic flow of intense behavioral interchanges.' This definition seems more comprehensive to important aspects of emotional systems, such as how emotions create subjectively experienced feelings and how they control personality dimensions.

Emotion considered one of the most complicated and attractive human behavior. It has an infrastructure that includes neural systems dedicated to emotion processes, at least in part, motivates cognition and action and recruits response systems [103]. Therefore, it has attracted many studies from different aspects such as the emotion mechanisms, expression of emotions, and emotions recognition [103]. Identifying what emotion is and how emotion is generated is mainly a research area in psychology and neural science. Thus, this thesis will not review the details on this area of emotion. However, emerging of Affective Computing, which is the field of creation and interaction with machine systems that sense, recognize, respond to, and influence emotions [176], made several researchers from different computer science areas interested more and more in emotions, especially in Natural Language Processing (NLP), and Human-Computer Interaction.

There are several related terms to emotion, such as moods, feelings, and effects. The terms mood and affect are found in literature and associated with emotion. Sometimes it is confusing to one who is not an expert in the emotion field. I find it worth briefly mentioning how we deal with the term in the rest of this thesis. Some authors consider feelings a general category that includes attitude, emotions, moods, and other affective states. Affects are subjective experienced emotional feelings that are difficult to describe but have been linked to bodily states such as homeostatic drives (hunger and thirst), and external stimuli (visual, auditory, taste, touch, smell) [169]. Affect, then, is the experience or feeling of emotion. The terms affect and emotion has been used interchangeably within the professional literature. For example, the words' affect' and' emotion' are used interchangeably in [52]: "Sin 2: Affect is subcortical. There is a tendency among some investigators to regard emotions as largely subcortical and to sometimes also assume that cognitions are cortical." [52].

Mood is defined as 'transient episodes of feeling or affects' [237]. There is an agreement in research literatures that the deferent between them is mood lasts longer than emotions [194] [237] [144]. Additionally, it has been reported that mood generally has low intensity and shows little response synchronization [194]. Mood has symptoms that patients may report, such as depression [102]. However, mood can be characterized by positive and negative moods shared by emotion. Generally, this thesis does not specialize in defining emotions or their related mechanism in neurosciences as it's not the main subject of my research. Therefore, I want to raise a question: does distinguishing affect, mood, and emotion make any sense for my work. In brief, fundamental valence (positive and negative) is the crucial characteristic shared by emotion, mood, and affect [225]. Therefore, I will go with the idea that 'The term emotion exemplifies the "umbrella" concept that includes affective, cognitive, behavioral, expressive and physiological changes' [225] for the rest of this thesis. Also, the term mental health will be used in exchange for emotion, as mental health includes mood, emotion, affect, and feeling.

2.1.1 Emotion modelling

Emotion modeling is representing of emotion. Therefore, emotion models are used to define how to describe emotion. Two fundamental viewpoints guided forms of expressing emotion in literature [148]: emotions are considered discrete and fundamentally different constructs as the first viewpoint, and emotions are characterized on a dimensional basis in a combination of two or three dimensions as the second point of view. Depending on these views, we have two different emotion models: The first category is discrete emotional models (DEM), The second category is affective dimensional models (ADM).

Discrete emotional models rely on Categorical descriptions of emotion theorize that there is a set of independent "basic" emotions. For example, James [104] identified fear, grief, love, and rage as basic emotions. Ekman et al. identified six basic emotions: Anger, Disgust, Fear, Happiness, Sadness, Surprise [64]. Orthony, Clore, and Collins (OCC) model is proposed by Ortony et al. [165] based on the theory of appraisal, which stated that emotion arose from individual perceiving of events. They added 16 emotions to the emotions Ekman suggested as basic: relief, envy, reproach, self-reproach, appreciation, shame, pity, disappointment, admiration, hope, fears-confirmed, grief, gratification, gloating, like, and dislike [165]. In total, they classified emotions to 22 types according to the degree of intensity of how humans perceive events [165]. Polarity classifying sentiment as positive, negative, and neutral is also applied as a discrete emotion model [112]

[229].

Dimensional emotion models assume that emotions are not independent and can be placed in a spatial set of values. Thus, they map emotion into a continuous spectrum. one of these representative models is proposed by Russell [191], which adopts the concept of Valence and Arousal, Arousal differentiating emotions by Activation and Deactivation. In contrast, Valence determines emotions by Pleasantness and Unpleasantness Figure 2.1. Another model presents emotion in 3 dimensions: Pleasure-Arousal Dominance (PAD) emotional model [147]. In addition to Arousal and Valence, as they represent how pleasant/unpleasant or active/inactive an emotion is, respectively, the third dimension of Dominance describes the degree to which experiencers had control over their emotions.

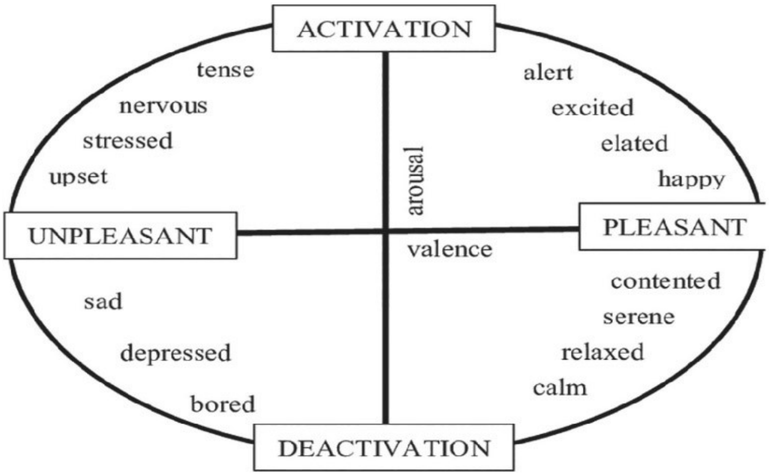


Figure 2.1: The model of Emotion by Russell [191]

Due to its simplicity, the categories models have been widely used for emotion classification in computational approaches [35]. However, compared with dimensional models, emotional categories models may not cover all emotions sufficiently because they are limited in categories, whereas dimensional models can capture wider emotion concepts that differ only slightly. However, a dimensional emotion model can be recommended for measuring the similarity between affective states [4]. As the dataset used by this project is labeled as discrete labels, and this thesis is not interested in measuring the similarity of emotions, will utilize a discrete emotion model.

2.1.2 Modelling user emotion from the text

According to [194], to measure the emotion, we need to measure:

- Changes in the evaluation process of the central nervous system.
- Response patterns in neuroendocrine, autonomic, and somatic nervous systems.
- Motivational changes derived from the appraisal result in a particular action.
- Facial and voice expression patterns and body movements.
- The nature of the feeling state of subjective experience reflects the changes of all these components.

This comprehensive emotional measurement is unlikely to become standard procedure [194]. Nevertheless, researchers in the different subjects used some of these measures. First, the first three measures regard Neurocognitive and Neurolinguistics. While computer science researchers mostly use the last two measures.

As it is possible to measure the emotion based on behaviour [194], and with the emerging field of human-computer, research in automatic emotion recognition has been attracted to extract or detect emotion (modelling emotions) on parts from people behaviour, including: text [242], audio [117], visual [108], and physiological data [201]. They aim to develop machines that can detect users' emotions and express different kinds of emotions.

Earlier research proved that emotions could be sensed in text-based computer-mediated communication [217] [90] [85]. Emotion appears to be detectable linguistically in text-based communication via word choice, word count, punctuation, and timing [217]. Experience of sadness in text-based CMC is related to fewer words, more disagreement, and slower response times than happiness [90]. There is numerous literature available in modeling emotion from the text. This thesis will focus on modeling emotion from the text. Therefore, another method will not be reviewed.

Many researches has been conducted to enable machines detect and understand human emotion states from user generated text. Linguistically observed patterns prove to be effective hints for emotion. For example, the incidence of abnormal language in writing of highly scored depression students was observed by psychologists and they found that students had more frequent usage of first person singular pronouns and more frequently expressed negative emotion words [47] [190]. In another case, it has been founded that people can observe diagnostic information on a wide range of psychiatric disorders used online, such as depression and post traumatic stress disorder (PTSD) [10] [190]. Moreover, the differences in language practitioners between disordered and control groups on social media have been noticed by applying language-analysis methods to social media data [183].

Social media and emotion modelling

Online social media are Internet-based services that enable express and exchange thoughts between people, recording and sharing their details of daily life, and making relationships [109]. There are various forms of social media like social networking (Facebook and LinkedIn), microblogging (Twitter), social search (Google and Ask.com), photo sharing (Flickr and Instagram), video sharing (YouTube), instant messaging (Skype and WhatsApp) and social gaming (World of Warcraft) [7]. With the revolution of internet and advances in technology and mobile phone, users of social media increased dramatically. Recently, it have been showed that over 3.6 billion people were using social media worldwide, a number projected to increase to almost 4.41 billion in 2025, and the average time a person spends on social media a day is 2 hours 24 minutes [62]. Consequently, huge data has been available which provides opportunities to observe people health problems on a large scale by studying people's thinking, emotions, concerns, activities and socialization based on user-generated data on social media [58] [54] [99] [198].

It is claimed that passively browsing others' photo of vacations might have triggered envy and loneliness [120]. Whilst social relationships are an important source of emotional support, they can also be stressful. A systematic review [25] of social media use in adolescents highlighted potential harms such as social isolation, depression and cyber-bullying. Potential negative effects on people using social media for healthcare have also been raised, including loss of privacy and being targeted for advertising [204]. In the other side, emerging research has begun to examine how social media can be used to improve people's mental wellbeing. Several studies have demonstrated the effectiveness of online social support for mental health through randomized controlled trials, where participants were recruited to engage with purpose-built social technology platforms. One study showed that after 3-weeks of use, participants, especially those with depression symptoms, reported a significant improvement in depression [161]. A 12-week study found a positive effect of online support groups and automated depression training programs in reducing depressive symptoms [81].

Use of social media as a tool for (mental) health monitoring

With the pervasiveness of social media, social media is used as key health information source. This provides opportunities to learn more about health problems on a large scale. A growing number of people are now describing their healthcare experience platforms such as Facebook and Twitter. The increasing availability of patients' accounts of their health online in a public forum presents an unprecedented opportunity to gather new forms of health data and to advance the patient-centred care agenda such as sleep issues [146], eating disorders [41], and smoking cessation [48]. Moreover, Twitter has been used to automatically measure the incidence of a set of health conditions [180]. Chunara et al analyzed cholera-related tweets published during the first 100 days of the 2010 Haitian cholera outbreak [46]. Posts were used to looking for regular language patterns to detect potential suicide attempts [9]. As well as food consumption and dietary choices have been investigated based on social media data [57]. A large-scale linguistic analysis on tweets has been done to the posts from top 100 the most populous counties in the U.S., and indicated significant correlation between Twitter information and health statistics, such as ingestion of healthy foods and obesity [51].

Regarding digital mental health monitoring, social media has been examined as a new tool for mental health measurement and surveillance [54], stress [131], and the severity of mental illness [40]. A ground body of research that detect and modelling mental health and mood from social networking text. A recent review by Chancellor and de Choudhury [39] highlighted the works has been done in mental health and it has been found that the number of studies using social networking: Twitter (30/75), Sina Weibo (13), Reddit (13), Facebook (6), Instagram (4), Tumblr (3), and ReachOut (2). Single papers inspect Flickr, PTT, mixi, LiveJournal, and TOBYO Toshoshitsu [39].

Behavioural and linguistic cues from social media data have been used by researchers to predict the presence of emotion and psychosocial disorders [21]. De Choudhury et al. [56] [55] utilized signals of social activity, emotion, and language from Twitter to propose analysis of population depression. Statistical methodology in a prediction framework was developed to derive different markers of shifts to suicidal ideation from modelling user of Reddit, Inc data [57]. The frequency of determiner my and me's use was used as a marker for depression in Twitter, Inc messages [162]. Machine learning in combination with clinical appraisals was utilised to proposed a method that identifying social media markers of schizophrenia [26].

Now, we can find an enormous experience and feeling states published by users in every social network. However, due to the vast amount of data generated by the different users in different times, detecting emotions and feeling from social networking is very challenging. Therefore, automatically detecting the emotion from text is crucial to save time and effort and take advantage of the huge data available in social network. Consequently, researchers have started investigating and developing approaches that can and can effectively detect and classify emotion, and they proved the effectiveness of these approaches.

Sentiment analysis a common approach to quantify people's attitudes by applies natural language processing (NLP) and text analysis. It is identifying the opinion or sentiment polarity of user towards some topics [133]. Applying sentiment analysis techniques in social media can detect health issues like adverse drug reactions [118] and predict the user feeling on political issues like election [223].

Sentiment analysis and text emotion recognition are closely related. Sentiment analysis is goal-oriented and aims to determine opinions or attitudes about topics or entities (such as products, movies). On the other hand, emotion recognition focuses on recognizing the emotion expressed in the text or caused by the text and does not depend on a specific objective [255]. In fact, there is relationship between these two terms emotion and sentiment. Moreover, emotions and sentiments can express experiences that be caused by the biological, the cognitive, and the social [255]. Stating the attitude of someone towards a particular target or topic can be the sentiment in general. Moreover, defining the polarity of a piece of text is the common aim of the sentiment analysis as a binary classification or as positive, negative, or neutral classification. In either case, emotional analysis will be considered as a comprehensive term of sentiment analysis. In overall, there is similarity in modelling sentiment analysis and emotion analysis in the NLP works. However, due to the complexity emotion representation models, there is minority of emotion analysis works compared to sentiment analysis which is polarity-based. Both terms (sentiment analysis and emotion analysis) will be used interchangeably in this thesis when there is no ambiguity.

Emotion detection (analysis, extraction, modelling) from the text can be done mainly at various levels: from words level (Aspect-level or features level); from sentences level (chat messages, post, and tweets) and from a whole document level. Even more, streams of texts such as tweets mentioning an entity over time.

At aspect-level classification, which is also also called feature-based level or entity-based sentiment analysis, features or aspects in a sentence (which is a user-generated review of an entity) is identified and then the emotion is determined regarding the features as positive or negative. At Sentence-level emotion such as positive, negative, or sad, happy is assigned to whole sentences. Summing the polarities of the words of the sentence can not considered as the emotion of a sentence but using models which can learn from labelled training data. Moreover, models use sentences' low-dimensional vector representations to represent the text. At document level, such as data comes from product reviews, and news comments, the emotion state assigned to the whole document as one unit. Some literature extract emotion and sentiment in level of streams of documents and tweets. Their aim is to detect aggregate trends in emotions over time. For example, Fraser, et al [71] analysed emotions in the tweets that mentioned a hitch-hiking Canadian robot. This thesis mainly focus on sentence level emotion analysis. Hence related works in sentence-level emotion analysis will be will specifically be introduced.

Emotion Resources

There are two common types of data resources for emotion analysis in text:

word-emotion lexicons and text corpora (datasets).

Emotion lexicons

Word-emotion lexicons were baseline for text-based emotion modelling methods. Emotion lexicons provide label or rating to individual words depending on the type of emotion model either categorical or dimensional. Either manual annotation or automatic annotation can do obtained the value for words.. Manual annotation for emotion effected word can be done by crowdsource or by researchers. Below are some most used manual annotated emotion lexicons.

- General Inquirer (GI) [207]: considered to be the first sentiment lexicons in English and contain 3,600 English terms annotated for associations.
- The Affective Norms for English Words (ANEW) [30]: valence, arousal, and dominance annotation for 1,000 English words.
- Multi-perspective Question Answering (MPQA) Subjectivity Lexicon [240]: contain 8,000 English terms classified by valence to strongly positive, weakly positive, strongly negative and weakly negative.
- AFINN [164]: This lexicon consists of 2,500 English terms rated for valence between -5 (most negative) to +5 (most positive).
- Linguistic Inquiry and Word Count (LIWC) [173]: consist of 1,400 English terms manually annotated to the affect categories: positive emotion, negative emotion, anxiety, anger, and sadness.
- NRC Emotion Lexicon (EmoLex) [160]: was the first crowd sourcing annotated emotion lexicon and it is contain14k English terms annotated by Amazon's Mechanical Turk. Each word has ten binary scores (0 or 1) indicating no association or association with anger, anticipation, disgust, fear, joy, sadness, surprise, trust, positive and negative.
- NRC Valence, Arousal, and Dominance Lexicon (NRC-VAD) [155]: it is crowdsourced emotion lexicons that include 20K English words and their valence, arousal, and dominance scores between 0 and 1.

- NRC Emotion Intensity Lexicon [156]: researchers annotated 10K English words for intensity scores corresponding to eight basic emotions including anger, anticipation, disgust, fear, joy, sadness, surprise, and trust.

Manually annotated Emotion lexicons is simple and effective to analyse emotion in text and it is used by many machine learning systems. If the training data is quite small, researchers use them to improve the accuracy of the prediction. However, it is costly in time and resources and limited in words entries. Thus, automatic annotation approaches to emotion lexicons have been considered.

Automatic annotation methods for sentiment lexicons manly rely on three basic methods: label propagation method, a large lexical resource method, and set of seed terms method.

In propagation methods, sentiment information propagated from a seed word to lexical resource other terms. For example, Mohammad et al. [160] used a simple word–seed co-occurrence in text to propagation. It is common to produce a word graph and then label propagation using graph-propagation algorithms in a semantic network like SENTIWORDNET 3.0 [19]. Hamilton et al [88] combined domain-specific word embeddings with a label propagation framework to induce sentiment lexicons using small sets of seed words.

The second approach relies on using commonly lexical resources such as WordNet [94], or text corpora such as collections of tweets [160] [2] to create a dense vector for the word. Then regression model or a classifier is used to mapping this vector to emotion value. Recently, word embeddings (word vectors generated from large text corpora) have been used as a primary lexical resource [68].

The third approach of automatic methods depend on seed words that are already labelled either manually or through existing manually created sentiment lexicons. By using statistical information, the target word will be labelled depending on the relation with the seed word. For example, hashtags (angry, sad, good, terrible, etc.) and emoticons (:), :() were used to determine and emotion scores [160]. likewise, Mohammad and Kiritchenko [159] used Point-wise Mutual Information (PMI) from twitter hashtags to annotate discrete emotion lexicon.

Here we also list some other popular resource for Emotion lexicons:

- WordNet [151]: an online English lexical database inspired by psycholinguistic theories of human lexical memory. It groups verbs, nouns, adjectives, and adverbs into sets of synonyms called synsets.

- WordNet-Affect [209]: a lexicon extend WordNet by annotating some of WordNet synsets, words.
- EmoSenticNet [177]: extends WordNet-Affect by using fuzzy c-means clustering and support-vector-machine classification to providing both emotion labels and polarity scores for a large set of natural language concepts.
- DepecheMood [206]: lexicon of 37 thousand terms annotated with emotion scores by regression and classification methods.

Text corpora

The requirement of large text corpora or datasets increased recently with the rise of machine learning and deep learning models to modeling emotion from the text. Recently, the availability of enormous text data on the internet and social media makes the existence of the dataset easier for researchers. However, building a large dataset is still a challenging task. A labelled corpus is essential to quality and accuracy of machine learning models, especially deep learning models. There is an increasing demand for the labelled dataset with the improving of deep learning methods and increasing of text data. As well as annotation of emotion lexicons, annotation of text corpora can be divided into three forms: manual annotation by experts or crowd-sourcing, semi-automatic and automatic In the first category is manual annotation, earlier work by Scherer and Wallbott [195] generated a labelled dataset from collection of sentences from questionnaires. Strapparava, Mihalcea [208] provided a labelled dataset with 1,250 news headlines from major newspapers such as New York Times, CNN, and BBC News, and from the Google News search engine. This dataset focuses on the classification of emotions and valence (emotion polarity) and labeled with the six Ekman emotion labels. Roberts et al. [187] annotated around 3,500 tweets manually for 7 emotions (anger, disgust, fear, joy, love, sadness and surprise) for 14 topics which they believe should include emotional tweets. Schuff et al. [196] annotated the SemEval 2016 Stance Data set [157] with emotion annotation by six annotators. 15,553 tweets with 28 emotions have been annotated by crowdsourcing through Amazon Mechanical Turk (AMT) [129].

The second and third category of building emotion datasets has been done though semi-automatic or automatic methods. Statistical and machine learning models have been used with text witch generally contains evidence for emotion categories. For example, labelling Twitter messages automatically through hashtags and emoticons [91]. An example of a semi-automatic approach is the work of Li et al. [127]. They produced an emotion corpus from micro-blogs through a three-stage method. They started with verifying hashtags automatically by a lexicon-based voting approach. Then, they used an SVM classifier to select the data with natural labels, and professional annotators annotated the remaining parts. Below are some of the annotated datasets that are available for research purposes.

- ISEAR Dataset [195]: dataset consisted of 7665 sentences labeled with emotions (joy, sadness, fear, anger, guilt, disgust, and shame) through questionnaires.
- SemEval "Affective Text" [208]: dataset contains 1250 labelled text from Arabic and English news headlines.
- IMDB [138]: dataset consisted of movie reviews labelled on a scale of 1–10.
- SSTb (Stanford Sentiment Treebank) [205]: dataset consisted of labelled reviews in 5 classes for multi-class classification.
- Yelp 2014 [15]: dataset consisted restaurant review labelled on a scale of 1–5 derived from Yelp Dataset Challenge.
- Emotion-Stimulus [75]: dataset consisted of 1594 sentences labelled with Ekman's basic emotion categories.
- EMOBANK [33]: dataset of 10k English sentences annotated with dimensional emotion in the Valence-Arousal Dominance (VAD) model.

- WASSA-2017 [158]: datasets consisted of tweets annotated for anger, fear, joy, and sadness emotions. the data were used for the WASSA-2017 shared task on detecting the intensity of emotion.

- DailyDialog [128]: manually labelled dataset with emotion information. It consisted of a total of 13118 sentences from daily communication dialogs annotated for neutral, anger, disgust, fear, happiness, sadness, surprise emotions labels.

Generally, the manual annotation dataset approach seems to have higher quality than automatic or semi-automatic methods. However, it is considered a time and effort-consuming method and seems to be limited scale. Moreover, it is not taking advantage of the huge text data available recently. However, the automatic method enables researchers to consider big datasets from social networking and another resource, which seems to increase the corpus's noise.On the other hand, both manually annotated and automatic annotated dataset seem to have lack of quality as the label is assigned by an annotator or automatic method which designed by annotator and both could wrongly decide what the writer feeling. This is a notable gap in the datasets that used by researchers in this area. Scherer believes that letting users report their feelings is the best way to reflect mental changes [194]. Consequently, the quality of training machine learning models will increase and will classify the text with more accuracy. Existing of huge dataset with self-labels of writers themselves is a gap in this field. This thesis will fill the gap by using a new dataset with millions of labelled posts. Moreover, this dataset is labelled by certain emotion labels, and this is an advantage as giving users certain emotion labels better than let them express in their words [194].

Emotion recognition from the text approaches

The general categorization for emotion analysis approaches presented in literature is: knowledge-based techniques, statistical techniques, and hybrid approaches [245].

knowledge-based techniques

Emotion analysis's early methods were based on predefined knowledge, and it can be categorized into two classes: keyword-based method and rule-based models.

In the keyword-based method, the sentence is classified into effect emotion based on emotion words like 'happy', 'sad', 'angry, and 'good'. These words are predefined in emotion lexicons such as WordNet [151], WordNet-Affect [209], EmoSenticNet [177], DepecheMood [206]. For example, Strapparava and Mihalcea [208] used WordNet-Affect lexicon to classify the text by inspecting the attendance of emotion words. Then the score of the frequency of the words is computed. The emotion words might be assigned to the emotion category by computing every word's intensity and taking the category with the highest intensity level as the final emotion label [206]. However, this straightforward approach depends on the context, and consequently, it leads to inaccuracies in the classification. For instance, if we have a sentence like 'I am happy because I see my teacher', the sentence will be labelled by 'happy' emotion. As well as the sentence 'I should be happy to see my teacher, but I am not will be labelled with the same emotion.

In rule-based models, emotions will be analysed based on the presence of manually defined logical rules, or linguistic patterns. For example, [43] developed a rule-based system based on linguistic features, using WordNet, SentiWordNet, and WordNet-Affect lexical resources to analyze news headlines' emotions. The authors also aimed to identify the expression that carries the title's main topic and make it weigh more. Another work by Ma et al. defined the emotional content of online textual messages based on the role of a verb in a sentence [137]. the authors used the word spotting technique to decide the emotion of the basic phrase type, and then the extracted emotion category is transferred to the upper layer phrase. However, this approach is considered consumer to time and computing during the rule definition process and is hard to generalize.Generally speaking, knowledge-based techniques need a comprehensive knowledge base that encompasses human knowledge. Consequently, it is hard for the machine to define the emotion associated with natural language. Furthermore, this method is affected by the limitation of keywords, the ambiguity of keywords, and the lack of linguistic information. Therefore, another limitation of knowledge-based approaches lies in the need for a reasonable number of predefined emotion categories. With the vast amounts of data we currently have, the need for a more intelligent system and not consuming time and predefined rules is increased. Thus, statistical and machine learning techniques is developed.

Statistical techniques

Statistical emotion recognition techniques is designed to learn from the data. Most of the statistical emotion methods are essentially machine learning methods. Machine learning systems learn a model from given some text and associated true emotion labels, and then the model predicts the emotion label of new previously unseen text. To assess the model's performance, the accuracy on a held-out test set for which emotion labels are available as well is measured. Basic machine learning-based approaches can be categorized into two methods unsupervised machine learning [251] and supervised learning, including Support Vector Machine (SVM) [54], Naïve Bayes (NB) [241], Maximum Entropy (ME), Random Forests [212], and regression [254]. However, according to a recent review by [245] the supervised learning method is mainly used by text-based emotion analysis researchers. Deep learning approaches will be discussed in the next section.

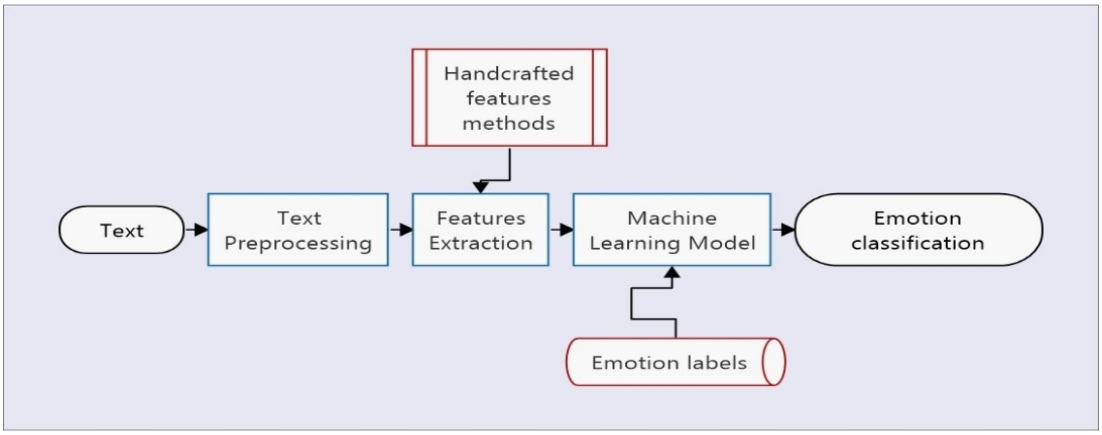


Figure 2.2: Illustration for basic machine learning approach

The procedure of classical machine learning-based approach shown in Figure 2.2. In this approach, given some text and associated true emotion labels, the basic framework starting from pre-processes input text. The pre-processing steps may include tokenization, stop words removal and lemmatization. Then the text will be converted into a feature vector representation, such as the Bag-of-Word (BOW) features (frequency of occurrence of each word), the Part-of-Speech (POS) features (labelling each word with parts of speech such as nouns, adverbs, verbs, adjectives, etc.) [167], the n-gram features (combination of a contiguous sequence of n words in the input text) [167]. Vectors can be a series of 0s and 1s or a series of real-valued numbers. In the initial researches representation, the sentences converted by hand-engineered vectors, such as, the presence of the word as a positive term in the sentiment lexicon, the presence a negation followed by of the positive word, the frequency of positive words in a sentence, and so on. Most of the vectors is represented by collection of zeroes and a few non-zero integer values since most of features were binary and sparse. Given the feature set and emotion labels, the system will be trained to produce an optimal hyperplane that can classify emotions. SVM has been proved to be the most effective algorithm on text-based emotion categorization tasks and robust on large feature space [115]. However, feature engineering and feature extraction are time-consuming processes in addition to the lack of ability to generalize well for other domains or areas. Hence, deep learning is introduced and will be reviewed in the next section. Due to the need for these approaches to huge of data to sufficiently train themselves in machine learning approaches, researchers observed hybrid approaches that combine knowledge-based techniques and statistical techniques to emotion recognition.

Hybrid approaches

Combination of machine learning models and knowledge-based techniques to model emotion from the text has been proposed and applied by researchers. For example, the joint use of knowledge bases from lexical resources and machine learning has been conducted by [31]. Gievska et al. [76] combined a lexicon of emotion words related to Ekman's six basic emotions and several classification algorithms, including naïve Bayes, SVM, and decision trees. He found that the lexicon approach improves when it is a combined SVM classifier and provides the best results. Due to resolve polarity disambiguation in contextual concept, Xia et al. [244] used SenticNet as a Knowledge approach and a Bayesian model.

2.1.3 Deep learning in text processing

Due to the challenges in traditional approaches of text-based emotion analysis such as the lack of generalization for other domains and time-consuming and tedious process of using handcrafted features, deep learning models have been applied recently. Deep learning automatically creates the required features for the classification process, in contrast with traditional machine learning models which need feature extraction process. Additionally, the recent rapid growth of available text shows that there is an increasing demand to methods other than traditional machine learning models to extract emotion effectively. At the same time, deep learning models outperform the machine learning approaches in fields that have massive features with large datasets like Computer Vision [110] and Speech Recognition [80]. Moreover, statistical text classifiers need a sufficiently large text input like a page- or paragraph-level to classify well as it is need features extractions and can not work individually. Hence, smaller text units such as sentences, or posts will be not classified appropriately. However, non-linear and complex patterns in the data can be learned effectively with architecture of deep learning which is multiple layers. Thus, I believe deep learning models improved performance more than traditional machine learning models, and this thesis will continue the work and utilise the state of the art deep learning models to the text based emotion modelling. Furthermore, deep learning take advantage of increasing of available computation and data as it does not need much engineering by hand.

Deep learning model is a machine learning model that a hierarchy of concepts is applied to learn from experience and historical data by where each concept is defined in terms of its relation to simpler concepts which learning complicated concepts by building them based on simpler ones [125]. Deep learning made up of artificial neural networks having multiple hidden layers between the input layer and the output layer. The back-propagation algorithm is used in deep learning to change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. In NLP domain, deep learning has produced extremely promising results for various tasks such as question answering [28], language translation [105] and sentiment analysis [245]. Deep learning is one of the fastest-growing areas, this section provides overview to some of the state of the art deep learning models that have been used in emotion modelling from the text.

Word embedding

In NLP, when the input is a text, words and sentences are represented using a real-valued vectors. These continuous word vectors called word embeddings. It is depending on the hypothesis that the word occur in similar context will have similar meaning. Thus, words that are closer in the vector space are expected to be similar in meaning. One-hot encoding technique to represent the word is not able to capture the relationship between words. Thus, word embedding become a trend as it can capture the semantic and syntactic representation of word and it has capability to represent the relation of a word with other words. Consequently, the accuracy of modelling the word will be increased. Word2vec [149] is a technique of NLP uses a neural network models to learn word embeddings from a large corpus of text. Two main learning algorithms in word2vec to obtain the word embedding: continuous bag-of-words (predicts the current word from surrounding context words) and continuous skip-gram (predicts the surrounding context words from the current word) [149]. Basically, mapping the words to a word matrix and words converted to vectors in an n-dimensional vector space. Global Vectors (GloVe) [174] as well is an unsupervised learning model to generate vector representations for words. It is trained in parallel thus it is trained quickly on more data. Char2vec [36] is a model generate embedding associated with each character of a word instead of learning the embedding of the full word.

Convolutional neural networks (CNNs)

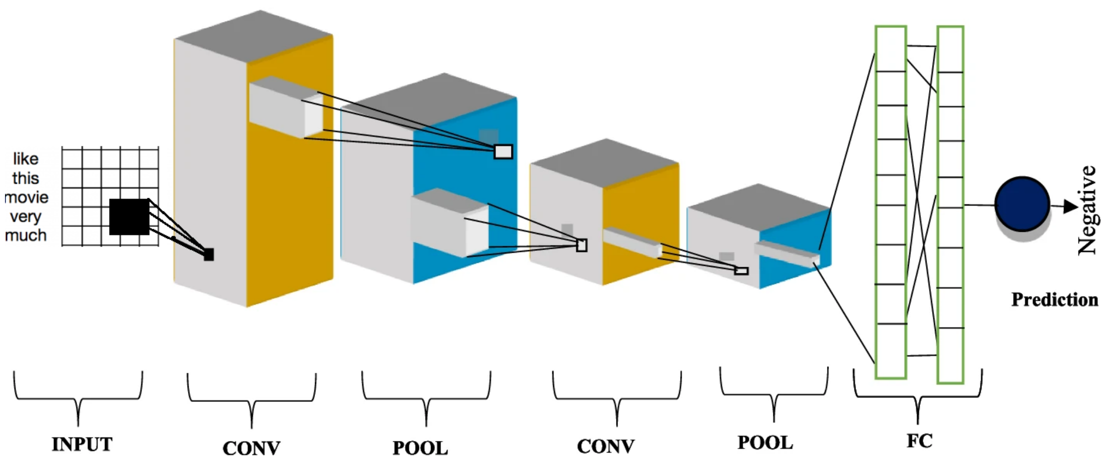


Figure 2.3: CNN architecture

CNN is a neural network initially used in Computer vision. It is depend in the math operation of a discrete convolution to filtering the feature map and that where the name come from. The architecture of a typical CNN is structured as Three main types of layers: Convolutional Layer, Pooling Layer, and Fully-Connected Layer, these layers stacked to form a full CNN architecture (Figure 2.3).

In more detail, In the convolutional layer, the input will be convolved and passed the result to the next layer. Convolutional layer contains feature maps of organized unites. units arranged in 3 dimensions: width, height, depth. Each unit is connected to local patches in the feature maps of the previous layer through a set of weights called a filter bank. ReLU activation function used to pass this local weighted sum. The filter bank is shared between all units in a feature map to enable the same convolution to be used to find the result, allowing filters that are invariant across the time dimension to be learned. Basically, the convolutional layer detect local conjunctions of features from the previous layer.

Pooling layers reduce the dimensions of data by merging semantically similar features into one. Pooling could be local pooling (a sliding window aggregates the feature map reducing it by length) or global pooling (acts on all the neurons of the feature map resulting in a single value). There are tow types of pooling functions: Average Pooling: Calculate the average value for each patch on the feature map and Maximum Pooling (or Max Pooling): takes the maximum value for each patch of the feature map. It is hard to manually choose the optimal pooling type; it is determined based on the empirical performance. In addition to pooling layers, normalisation layers can be used to help a network converge quicker.

Moreover, Two or three stages of convolution, non-linearity and pooling are stacked, followed by more convolutional and fully-connected layers. The final output layer would have one dimension all the features map will be reduced to a single vector of class scores.

CNNs have been used recently used by researchers in the field of sentiment analysis. Convolutional neural networks used to extract from character to sentence level features to predict sentiment analysis of short text [61]. Sentence classification using one-layer CNN has been discussed in the work of Zhang and

Wallace [260]. The authors explored how the performance of a model can be affected by changing its configuration (hyperparameters, filter size, regularization parameters, etc.). Convolutional neural networks also proposed the task of sentiment classification and the authors did experiments with three datasets and they show that employing consecutive convolutional layers is effective for relatively longer text [113]. Jianqiang et al. used a word embeddings with n-grams features and word sentiment polarity score features as feature set integrated into a deep convolution neural network for training and predicting sentiment classification labels of Twitter dataset [106].

Recurrent neural networks (RNNs)

RNN is used for modelling the sequential data. RNN considers the time factor for processing the elements in a sequence. The term "recurrent" based on performing the same computation over each token of the sequence and each time step the results of previous hidden state is an input to the next time step. Basically, a sequence is represented by a fixed-size vector by feeding tokens one by one to a recurrent unit [106]. Meaning, a "memory" over previous computations is presented in RNN to use it in current processing. RNNs increasingly become a popular for NLP tasks in recent years such as speech recognition [80], language modeling [150], machine translation [213]. In the area of text classification like emotion analysis, RNNs is popular more than CNNs in these areas due to its ability to capture the inherent sequential nature present in text, ability to model variable length of text and modeling such context dependencies in language and similar sequence modeling tasks [252].

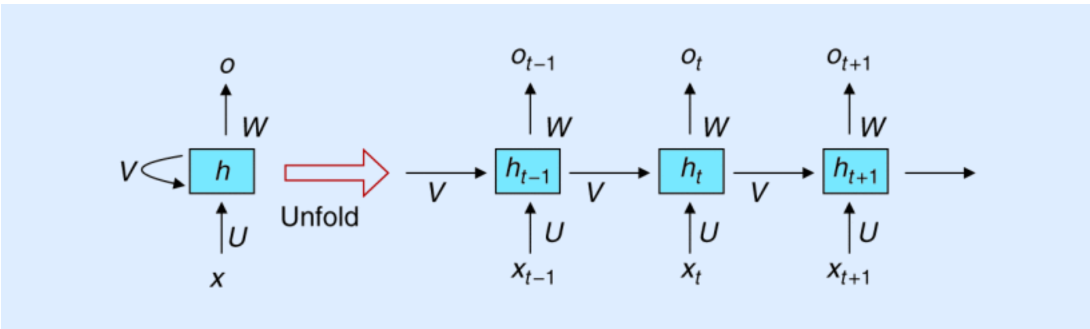


Figure 2.4: RNN architecture [252]

The architecture of RNN is illustrated in Figure 2.4 which is connected multiple copies of RNN, and they communicate by passing information from one state to another. If we consider x_t as the input to the network at time step t, s_t represents the hidden state at the same time step, function f is a non-linear transformation such as tanh, ReLU and U,V,W signifies shared weights across time, the current state is computed by using the previous state value and the current input based as on the equation [252]:

$$\mathbf{s}_t = f(U\mathbf{x_t} + W\mathbf{s_{t-1}}) \tag{2.1}$$

However, long-distance correlations in a sequence is difficult to be modelled with RNN because components of the gradient vector can grow or decay exponentially over long sequence [97]. This problem called exploding or vanishing gradients. Therefore, other types of RNN networks such as long short-term memory (LSTM), gated recurrent units (GRUs) used to overcome this problem.

Bidirectional RNNs

Bidirectional RNNs [197] are accomplished by running two RNNs, a left and a right networks allow to predict over a sequence with both the forward context and "backward" context. Every two RNNs together constitute a single bidirectional layer. The output of the two RNNs, is joined to form a single output vector either by summing the two vectors, concatenating, averaging, or another method. Furthermore, this approach can be extended to other forms of recurrent networks such as bidirectional LSTMs (BiLSTM). In text-based emotion modelling, there are many uses for this type of structure and have been stated bidirectional networks outperform forward-only RNNs [42] [263].

Long short-term memory (LSTM)

LSTM was first introduced by [98] to solve the problem of vanishing gradients of RNNs by having a memory cell that is able to preserve state over long periods of time. The hidden state of LSTM memory unit is calculated by utilizing three gates: input, forget and output gates. The architecture of LSTM is shown in Figure 2.5 and represents forget gate as f_t, input gate as i_t, output gate as o_t. h_{t-1} is previous state information and x_t is the current input. The LSTM transition functions are defined as follows: the forget gate controls how much is remembered from step to step by output a value of 0 or 1 where 0 signifies completely forget, and 1 signifies completely keep:

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \tag{2.2}$$

The input gate decides the values to updated and then a tanh layer creates a vector of values C_t:

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \tag{2.3}$$

$$C_t = \tanh(W_c.[h_{t-1}, x_t] + b_c) \tag{2.4}$$

The old cell state is multiplied with forget gate to forget the data previously determined and then i_t*C_t is added:

$$C_t = f_t * c_{t-1} + i_t * C_t \tag{2.5}$$

A sigmoid layer will be operated to decide the part of the cell state that will serve as an output:

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \tag{2.6}$$

Finally, the cell state is passed through a tanh layer to for restricting the value between -1 and 1:

$$h_t = o_t \odot \tanh(C_t) \tag{2.7}$$

It is seem to be that LSTM is the most used deep learning model in the filed text based emotion or sentiment analyses, either by only LSTM [235] [42] or with combine it with other lyres of deep learning networks such as LSTM and CNN [100] [92] [6], with multiplayer self-attention mechanism [24]. LSTM was utilized for twitter sentiment prediction and authors proved the effectiveness of LSTM to dealing with complex sentiment phrases such as negation expression [235]. Chatterjee et detected contextual emotions from textual dialogues and classified into one of four emotion classes as happy, sad, angry, and others using the long short term memory (LSTM) [42].

Gated recurrent units (GRUs)

A popular variant of RNN is GRU cells with simpler design and significantly

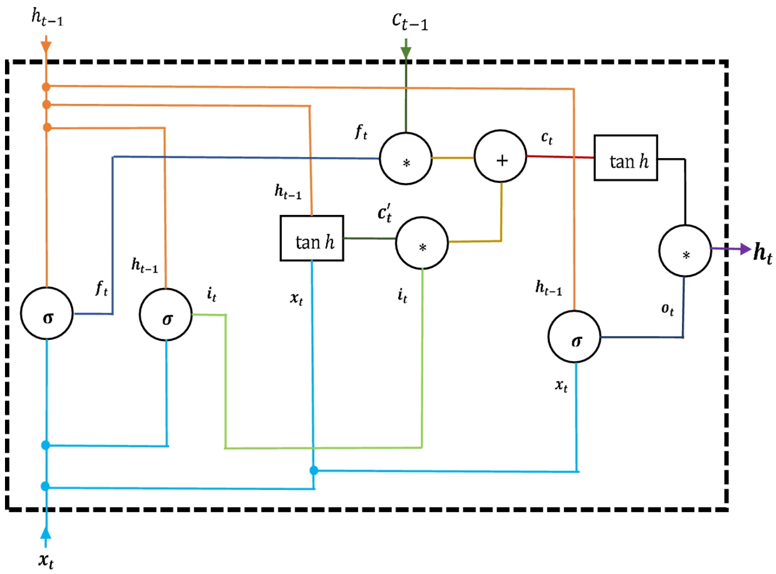


Figure 2.5: LSTM architecture [245]

reduce the time required to train models. As shown in Figure 2.6, only two gates (a reset gate to defines how much information (memory) needs to be kept for the future and an update gate to defines how much past information the network will forget) used in GRU rather than the three gates used by an LSTM. When performance is equivalent between the LSTM and GRU, GRU is chosen because it is uses fewer parameters. Furthermore, if the dataset is small, It has been stated GRU models out-perform LSTM networks [107] [266].

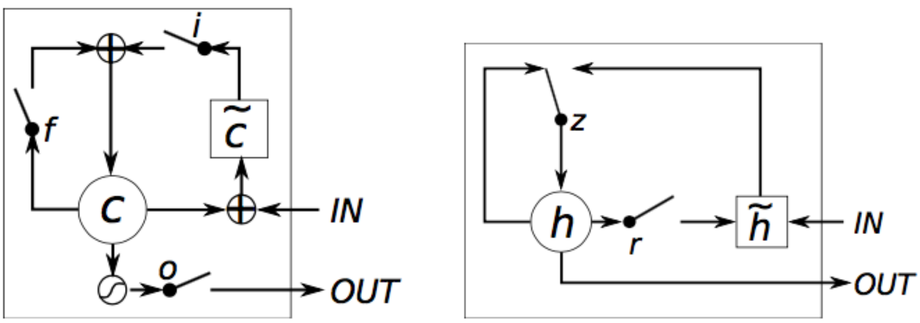


Figure 2.6: LSTM (left) and GRU (right) cells [233]

Attention Mechanism

It is unrealistic to expect a fixed-size vector in RNN networks to encode all information in a piece of text whose length can potentially be very long. Additionally, traditional RNN approaches capture irrelevant information in the piece of information-rich text and all the words are treated in equal weight. Hence, a recent trend in text classification inspired by the visual attention mechanism found in humans has been introduced which is the attention mechanism [20]. It is the process of focusing on certain parts of sequences or regions and ignore the remaining ones during the learning. In text classification, instead of encoding the full sentence, It decides which part of the text should be focused on. Moreover, attention provides a score for each time step and specify which inputs were most useful for the prediction. Thus, interpreting the output and inspecting the quality of a network will be easier. As shown in Figure 2.7 attention represented by a vector, often the outputs of dense layer through softmax function.

if we consider h_i is the word and h_i' is the the word representation obtained through the previous layer, the model calculate the similarity between them first:

$$sim_i = H_i\prime.H^T \tag{2.8}$$

then normalized by the softmax function to obtain the corresponding weight coefficient a_i:

$$a_i = \frac{e^{\mathrm{sim}_i}}{\sum_{j=1}^{L_h} e^{\mathrm{sim}_j}} \tag{2.9}$$

Finally, a weighted summation operation is performed to obtain the Attention value C_i:

$$C_i = \sum_{j=1}^{L_h} a_i \cdot h_j \tag{2.10}$$

There are two common attention mechanisms that produce state of the art results in NLP tasks: Self-attention [132] and Multi-Head Self-Attention [228]. Self-attention first introduced by Lin et al. and sometimes called intra-attention [132]. It is using additive attention to compute the score for each hidden state, then using all the hidden states to produce attention vector. After that, the final sentence vector is computed by several hops of attention are performed by using a matrix of vectors instead of a single vector that relating different positions of a single sequence and extract an attention matrix. Multi-Head Self-Attention will be reviewed in section of transformer.

In sentiment analysis field, Yuan et al proposed a domain attention model for multi-domain sentiment analysis where the attention is the domain representation and select the most domain-related features in each domain [253]. Yang et al. proposed Feature-enhanced Attention Network to improve the performance of target-dependent Sentiment classification [248].

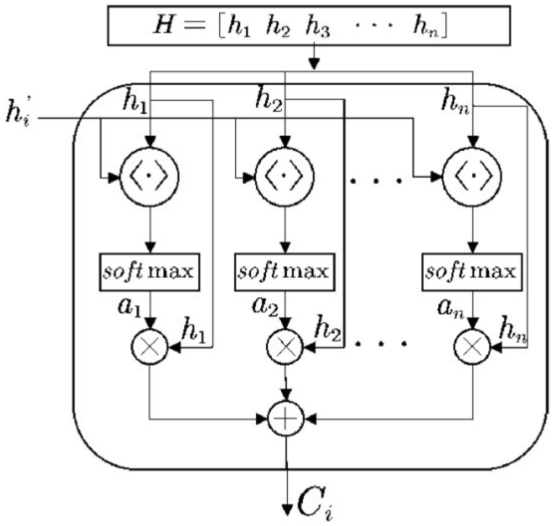


Figure 2.7: Illustration of Attention Mechanism [44]

Memory networks (MenNN)

MenNN was introduced by Weston et.al as a network with a memory m which is an array of vectors or an array of string indexed by M_i and four components I, G, O and R [239]. Input Feature Map I: converts the incoming input to the internal feature representation. generalization G: updates old memories given the new input. output feature map O: produces a new output (in the feature representation space), given the new input and the current memory state. response R: converts the output into the response format desired such as a textual response or an action. MenNN is extended by Sukhbaatar et al. to end-to-end memory networks or MemN2N [210]. MemN2N use attention mechanism in its memory to retrieve, thus enabling end-to-end training. MemN2N proposed good performance in answer a specific question and model for language modeling [210].

Transfer Learning (TL)

TL is a common approach in machine learning to overcome the problem of lack of data [166]. TL is based on the ability to learn new tasks quickly relying on similar tasks or domines by pre-training a model on a task which is similar to the target task and transfer the learned knowledge to the target task. This method gave rise to many successes in various domains of traditional machine learning algorithms. In the text emotion modelling and classification, TL used such as using unsupervised techniques for word embedding from a large corpus of data and employing it for various supervised tasks. It is possible to improve modelling emotion from the text by using a pre-trained model from a different domain while training using few samples. A common challenge of this field as well is lacking of labelled data for emotion classification, thus using transfer learning, would improve the performance of learning.

Transformer Networks

Vaswani et al. proposed a novel model that achieving the state of the art results in NLP tasks like the translation task and beating RNN and CNN networks and they called it transformer [228]. Transformers completely relying entirely on attention. It is the first model compute representations of its input and output without using sequence aligned RNNs or convolution. It is use "multiheaded" attention directly on the input embeddings Instead a memory of RNNs of previous states and that improve the computation by performing it in parallel. Similarly, transformer network improves the computation time by reduces the number of steps required for prediction.

In particular, the attention function that has been used in transformer is mapping a query and a set of key-value pairs to an output, where the query, keys, values, and output are all vectors. As shown in Figure 2.8, in the Scaled Dot-Product Attention, The input consists of queries and keys of dimension d_K, and values of dimension d_v. Where Q is a matrix of queries, K is matrix of keys and values also packed together into matric V, the matrix of outputs computed as:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2.11}$$

Instead of performing a single attention function, Multi-head attention performed

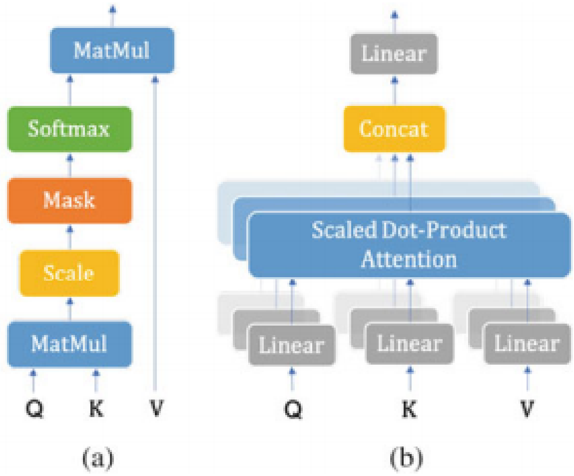


Figure 2.8: (a) Scaled dot-product attention, (b) multi-head attention [228]

as shown in figure 6 and applied as:

$$MultiHead(Q, K, V) = Concat(head_1, ..., head_h)W^o \tag{2.12}$$

Where

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \tag{2.13}$$

The parameters of all W matrices are projection matrices.

BERT: Pre-training of Deep Bidirectional Transformers

Building on the power of Transformers, recently, a popular state-of-the-art model has arisen in text classification which is BERT model [59]. It has been approved that BERT outperform other methods in NLP tasks [78]. BERT is a transformer-based, masked language model that is generating deep word embeddings. It is bidirectionally trained thus it is capture left-to-right and right-to-left contexts. The generated embeddings by BERT require very little fine-tuning to perform NLP tasks such as entailment or question-answering. Recently, in language representations, BERT has broken multiple performance records. Basically, there are two steps in its framework as shown in Figure 2.9: pre-training and fine-tuning. The model is trained on unlabelled large corpus in the pre-training phase and then in fine-tuning phase, the BERT model is first initialized with the pre-trained parameters and then all the parameters are fine-tuned using labelled data for specific tasks.

The input token sequence to may be a single sentence or two sentences packed together e.g., Question, Answer. The first token of every sequence is always a special classification token ([CLS]). for pair of sentences, they separated with a special token ([SEP]) and a learned embedding to every token is added to indicating whether it belongs to sentence A or sentence B.

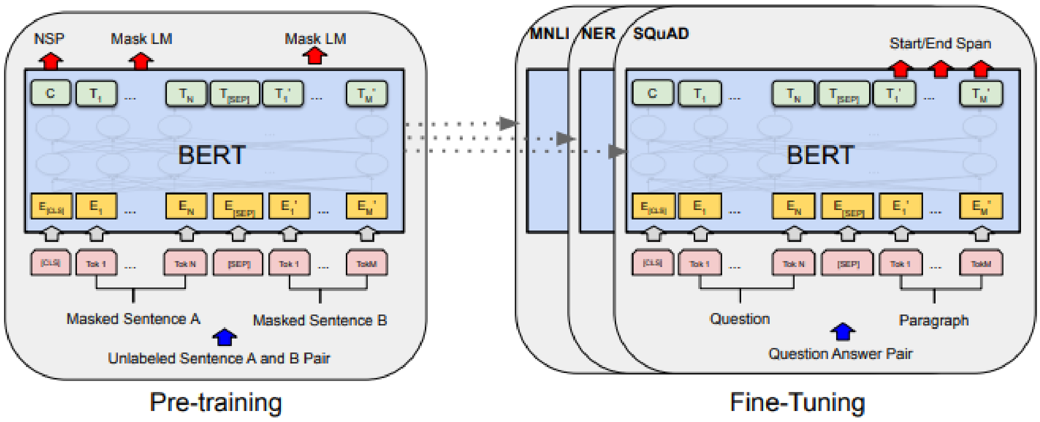


Figure 2.9: Pre-training and fine-tuning BERT [59]

The pretraining has been done by training BERT on the Masked Language Modeling (MLM) and the Next Sentence Prediction (NSP) mechanisms. In MLM, some percentage of the input tokens is masked at random using the "[MASK]" token, and then, predict those masked tokens. In NSP, given two sentences A and B, 50% of the time B is the actual next sentence that follows A (labeled as IsNext), and 50% of the time B is a random sentence from the corpus (labeled as NotNext) [59]. BERT was trained on 16GB texts from the BooksCorpus datasets (800M words) [265] and the English Wikipedia (2,500M words).

The Fine-tuning phase is straightforward since the self-attention mechanism in the Transformer allows BERT to model many downstream tasks. Basically, For each task, just simply plug in the specific inputs and outputs into BERT and fine-tune all the parameters.

There are two models of BERT, i.e., the BERT-base and BERT-large models. The BERT-base model comprises 12-layered transformer encoder blocks with each block containing 12-head self-attention layers and 768 hidden layers and producing 110 million parameters in total. On the other hand, the BERT-large is made up of 24-layered transformer encoder blocks with each block containing 24-head self-attention layers and producing a total of 340 340 million parameters.

However the BERT ability to trains faster due to its bi-directional ability; it is limited to monolingual classifications, and limited to the length of input sentences. Thus, researchers developed variants model from BERT to overcome the limitation of BERT. For example, The Generalized Autoregression Pre-training for Language Understanding (XLNet), [250] which is eradicates the fixed-length problem in BERT by using the positional encoding and recurrence mechanisms, whereas BERT masks the data and tries to predict the masked data using a bi-directional context. Similarly,A Robustly Optimized BERT Pretraining Approach (RoBERTa) [136] does not use the next sentence prediction task in the training mechanism and can handle large tokenized sentences with a maximum length of 512 words where each sentence is mapped to an embedding of 1024 floating points. Additionally, DistilBERT [216] a model is faster and lighter than the original BERT by reducing the number of layers in the BERT-base model by removes token embeddings.

BERT and its variants are being used actively in question answering (QA), natural language inference (NLI), text summarization (TS), and other NLP tasks. However, little attention has been seen in text based emotions modelling. Thus, This thesis will fill the gap by shed light on the efficacy of the BERT models in modelling emotions from social networking apps. There is no work has analyzed the efficacy of BERT models on TalkLife dataset to the best of our knowledge.

2.2 Modeling temporal dynamics (emotion change)

Emotions are dynamic by nature. They vary over time and transition from one to the next. As emotions are essentially responses or reactions to external stimuli, understanding emotion changes might help understanding the external environment, such as what is happening to the speaker, or what triggers these emotions. These demands motivate research aiming to detect emotion changes in time. Research on extracting emotions from text has focused primarily on steady-state emotion rather than on emotional transitions, the movement between emotion states. Can we predict if the user will change his/her emotion.

During social interactions, people do not necessarily stay in one steady emotional state; they can and do move between emotional states based both on their own intrapsychic changes and in response to environmental stimuli [72]. Hence, we need to model this change.

Previous research shows that when a person of a group experiences a particular emotion it can lead the other [93]. Moreover, when a person interacts with another, he will catch his emotion state via emotional contagion. Emotional state can also be transferred even by textual interaction. Previous researches show that people can catch emotional states they observe in others in text-based communication over time frames ranging from minutes to years [119]. Hancock et al. [90] observed emotional contagion in text-based communication by experiment the effect of the emotional contagion over a short interaction (approximately 15 minutes), suggesting that emotional contagion can operate quickly [90]. In other side, and for the long-time experiments Fowler and Christakis [70] observed data from social network collected over a 20-y period and the suggested that longer-lasting moods (e.g., depression, happiness) can be transferred through textual interaction as well [70].

Emotion transition is an research area where it can be defined as a movement between two or more affective, or emotional states [69]. Researcher has studied how people emotion change from one state to another [220] and the role of emotional transitions on social interactions [69]. Several studies have demonstrated that emotional changes can be observed from changes in body posture [60], facial expirations [23], and during the speech [227] [17].

2.2.1 Modelling User Change of Emotion

In general, modelling the change of emotion in literatures have been done in three ways: modelling the change of emotion intensity and its direction [65], modelling the transition from emotion state to another [38], or modelling the trend of emotion change [16], [23].

Modelling the direction of change of emotions state has been done by [220] when they modelling emotion transitions in four dimensions: valence (positive vs.

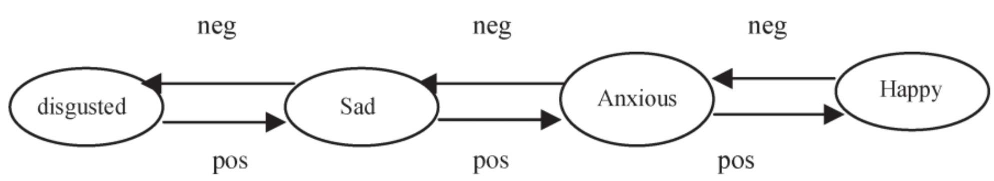


Figure 2.10: The proposed graph of emotions transition by [38]

negative), social impact (high arousal, social vs. low arousal, asocial), rationality (cognition vs. affect), and human mind (purely mental and human specific vs. bodily and shared with animals). Markov chains provide formal characterization of how emotions change over time [220].

An example of modelling transition of emotions from state to another state is the work done by [38]. In this work, they stated that a person's emotional state changes from happy to anxious on a negative input. On the other hand, with increasing positive influence (pos), the human emotion has a gradual transition from the disgusted state to the happy state Figure 2.10.

Modelling the intensity of change in emotions basically measuring the scales of each state, like what have done by [227] in analysing records of speakers and for each recording, the speakers rated 16 emotional states on a 10-point Llikert Scale. The Random Forest algorithm was applied to 207 speech features that were ex-tracted from recordings to qualify (classification) and quantify (regression) the changes in speaker's emotional state. They observed predicting the direction of change of emotions and the change of intensity.

Modelling the trend of emotions changes basically can be done by observing the patterns of emotions state by time. For example, [16] proposed a method to model the emotional changes of the utterance during the speech by using a statistical learning method. Their system classifies emotions of the utterance in six standard types including, anger, boredom, fear, disgust, sadness and neutral and then apply a classification approach based on statistical learning classifier to simulate changes trend of emotional states.

Another example of modelling the trend of changes is the work of [23], this work capturing the facial images of an individual using a visible and a thermal camera to decide the present affective state of the person and then the emotional state changes are observed using the trained HMM. The pattern of changes was used to classify the personality of an individual using trained HMM. The Idea of classify individuals depending on emotion and the change of emotion is motivate me to observe modelling user or user profiling as will discussed in next section.

2.2.2 Modelling emotion change from social network

Although there is recently rapid growth in emotion analysis from text are able to tell us what people are feeling, we also need to be able to know how people emotion change to be able to monitor the early signs of mental health problems or to suggest behaviour changes or early recommendations to users.

Social media provides online communication in real time at low cost, which allows researchers to monitor and track disease outbreak [203], sense the public's moods and responses to certain health issues [45], and through these identify target communities to disseminate health information [193]. Research on this field has focused primarily on steady-state emotion rather than on emotional transitions, the movement between emotion states. Since we can understand the user emotional state from help of understanding complex patterns in which an individual interacts with an online community, then that mean we can understand when the user mood change from his behaviour from social network. I believe that social media can be used as a tool for mental health monitoring, where population data can be captured, serving as a record and sensor for events in peoples' lives [54]. More specifically, analysing from linguistic, interaction patterns, and sentiment perspective of social media posts can offer a novel methodology for measuring and understanding mood changes when engaging with online peer-to-peer mental health support.

Text or social networking is considered as a tool to measure the change in user behaviour. For example, Pruksachatkun et. al. [182] proposed a model that can predict whether a conversation thread in an online mental health forum is associated with 'cognitive change' where people re-frame their negative belief into a positive one [182]. Likewise, researchers study the effect of posts in changing opinion.

2.2.3 NLP works in modelling emotion change

Although emotion change has been well studied by psycho linguistics and behaviour scientists, there are very little attention paid on predicting emotion changes from NLP perspective. A recent review illustrated that people change their emotion during one conversation [141]. Moreover, they stated that, presence of emotion change is a hard problem to solve, as the state-of-the-art Emotion recognition in conversation (ERC) model is more accurate in emotion detection for the utterances without or with less emotional change. It is interesting and challenging to track the dynamics of the users emotion by the time and how the change their mood.

The works that have been done in the area of modelling emotion change in social networking is done in two ways. Investigate the change of emotion in groups of users [65] [27] and investigate the change of emotion of users [54] [86].

The work done by Elgarroussi et al. [65] that analyse temporal emotion change by using tweets as the knowledge source for emotion analysis. They use the location and time of the tweets as inputs to their system and their corresponding emotion assessment score falling in the range [-1, +1], with +1 representing a very positive emotion and -1 representing a very negative emotion [65]. Then the author used VADER Sentiment Analysis Tool to assign an emotional score to each tweet. A set of unary and binary change predicates are evaluated with respect to the set of spatial clusters in batches uncovering the temporal variations in happiness. Next, an emotion change graph whose nodes are the spatial clusters and whose edges capture the temporal relationships of different types between spatial clusters is constructed. However, this work analyse the temporary change of emotions for groups or clusters. Moreover, they focusing in analysing the differences between clusters (how spatial clusters describing areas of positive and negative emotions appear, continue, disappear, intensify, grow and shrink in time and space). Additionally, they just use time and space as inputs without the consideration of text information.

Likewise, [27] observed tracking changes in the public mood state from the content of large-scale Twitter feeds. They used two mood tracking tools: OpinionFinder that measures positive vs. negative mood and Google-Profile of Mood

States (GPOMS) that measures mood in terms of 6 dimensions (Calm, Alert, Sure, Vital, Kind, and Happy). Then, they used causality analysis and a Self-Organizing Fuzzy Neural Network to analyse the public mood states [27].

In the study of [56], they model the change of emotion of new mothers around childbirth. They built predictive model about the influence of childbirth on the forthcoming behavior and mood of new mothers. They built predictive model to investigate how childbirth affects mothers' mood and behavior. They tracked the changes in Twitter posts, social engagement, emotion, social network, and linguistic style. Then, statistical models have been built by the author to predict substantial postpartum changes. These models achieved an accuracy of 71% with predicting the significant changes following childbirth and this accuracy increased to be 80-83% by adding more training data of 2-3 weeks of postnatal data. However, it is give good sight to track the change of emotion, they model the change of emotion around special events and not for long time series. Additionally, they use statistical models and may be more accurate with machine learning models for a short period.

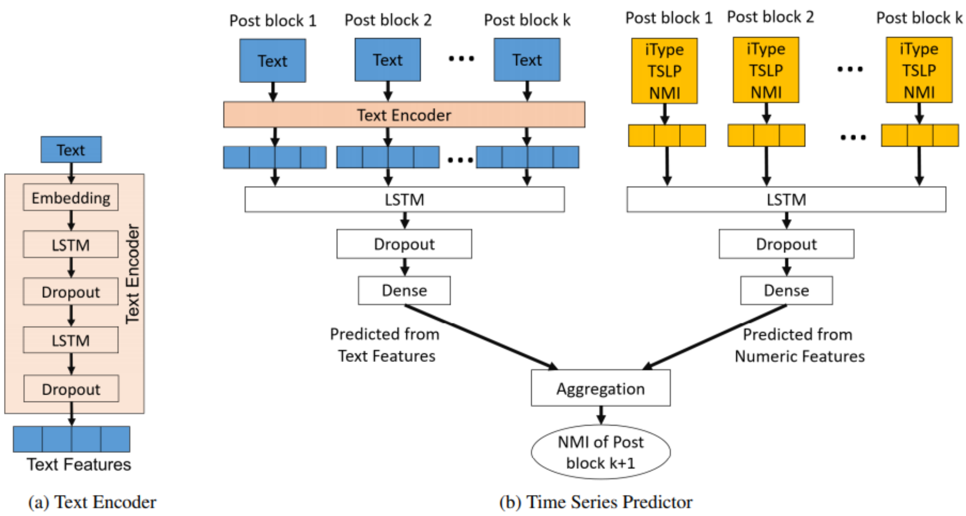


Figure 2.11: The proposed model by [86] to model the temporal progression of emotional status in online health forums

Research conducted by [86] is one of the very few works towards modelling the temporal progression of emotional status in online health forums. They used textual and forum engaging features to model the temporal progression of user mood using LSTM network as shown in Figure 2.11. To measure Temporal Progression of Emotional Status, They define a Negative Motion Index (NMI) of a post and ine case NMI < 0 points to those patients who are improving with time; > 0 is for those who are deteriorating; otherwise it denotes those patients who are stable. The prediction process is to predict the next NMI score given past k post details (text, and other participation metrics). A single post-block of time period to be 24 hours is constructed by combining the posts of the user that written within this period.

Table 2.1 summarise the work have been done in this area. Although there have been only a few studies that aim to detect the instant of emotion changes, there is a lack of modelling user emotion change. Moreover, there is a gap in modeling user emotion changes based on user's timeline and posted texts during a period of time.

2.3 User Modelling

User modeling is defined in Wikipedia as the process of building up and modifying a conceptual understanding of the user. With the rise of adaptive system, which can adapts itself to various circumstances, user model needed because this system are based on user's representation, knowledge, goals and preferences.

There are two common terms user profiling and user modelling to describe the process of modelling user and they are often used as synonyms or only one term is used by meaning of both in most of researchers community. However, user profile may be considered as a collection of personal information that stored without adding further description or interpreting whereas user modelling using this information to build up a model of the user [73]. In this thesis, these terms will be used in interchange as synonyms.

Since the early of 80's, researchers have been exploring the problem of user modelling [186]. Recently, with rapid growth of social media, vast amounts of data is becoming available, which has enabled the resurgence of machine learning in user modelling. Moreover, user modelling in social networking is used in different tasks in literatures, including news recommendation [3] [246], movie

Paper	predict	features	model	dataset
(Elgarroussi et al. [65]	temporal emotion change of groups	Time, location	graph	Twitter
de Choudhury et al. [54]	Postpartum Changes in Emotion	patterns of posting, linguistic style, and emotional expression	SVM, statistical	Twitter
Halder et al. [86]	temporal progression of a user	Time Since Last Post (TSLP), Interaction Type, NMI score, Glove word embdding	LSTM	Mental Health section of health-boards.com
(Bollen et al. [27]	public mood time series	Opinion Finder'sb Google-Profile of Mood States (GPOMS)	Neural Network model	Twitter

Table 2.1: Summarizing of the works of modelling the change of emotions in social networking

recommendations [77], place recommendations [126] [257] and recommendation smart mobility [12].

The key to model a user user profile is using dense vector representation user activities. Moreover, based on user archives and user activities, a representation of a certain user can be constructed. There are tow different types of user profiling: explicit user profiles and implicit user profiles. In explicit profile can present data like demographic attributes such as the user's address, marriage status, job status, birthday, gender and age. on the other hand, latent information implicitly expressed through the activities like user preferences. Collecting implicit information can be done through machine learning techniques that analyse user activity to predict the user's future needs and recommendations [154] [126] [257] [3] [246]. However, to create an accurate user profile, a large amount of data about interaction between the user and the content is required.

2.3.1 User modelling approaches

User modelling approaches can be categorized to neighbourhood-based approach, machine-learning approach, ontology-based approach, filtering approach, and statistical modeling approach [63]. This thesis will focus on machine learning approach thus it will be reviewed.

In the review of Eke et al. [63], it was stated that machine learning approach is the mostly used as profiling techniques. Additionally, machine learning models considered the standard method of user modelling in the recommendation systems [170]. Both forms of machine learning (the supervised and unsupervised learning) are used in user modelling. In supervised learning methods, K-Nearest Neighbour algorithm is utilized to user profiling based on personalization [29]. Likewise, Support vector machine classifier is used in user profiling to identify the relevant documents on the web [215]. On the other hand, in unsupervised Learning approach, multi-agent system is common used to user profiling as it used multiple agents that handle different personalization issues and phases [153]. [89] employed a fuzzy clustering technique to develop ontology-based user profiles. K-means clustering likewise used to profile users to recommendation production to smart mobility [12]. Recently, with deep neural networks and autoencoders evolution, neural network has been used to model users to enhance the recommendation performance [232] and for recommending music [226]. RNN was applied to model user's click behavior sequence [96]. users are modeled by LSTM and attention mechanism by modelling all the corresponding content, behavior and temporal information [163].

It has been classified user modeling into the following three classes: Behavioral modeling, Interest modeling and Intention modeling [74]. In social networking the behaviour of the user can be considered as two different types of activities, one and the rich valued is user composed content (including text like comments, posts, and personal status, uploaded videos and pictures) and another is ego network information such as connect with friends and followers, giving likes and link information. However, text that produced by the user often state strong perceptions about the user personality and the preferences on events and entities. It has been modelling users based on the textual context on social networking social network [14] [140] [66].

Depending on these behaviours, it has been profiling users of social network for Several tasks. Recently, use of a social graph of interaction between users and using network embedding to modelling users of social network become trending in literature [11] [152]. As social networking modelling suffer from dimensionality problem, lack of generalizability and overfittting problems with features engineering Without proper feature learning, using automatically encode features in low-dimensional embeddings become a critical tool to boost the performance of modelling users in social networking.

User Text Embedding was introduced in section 2.1 as a general text embedding but here will be discussed as a user text representation. In social networking, a user's posts will be mapped into a vector representation to capture the linguistic style of the user. Instead of word embedding, user generated text representation need to consider larger text units beyond words as a paragraph or document embedding model. Le and Mikolov initially proposed Learning embedded representations algorithm for larger text like documents and paragraphd which is Doc2vec as an extension to word2vec to learn document-level embeddings [124]. In this work, there are two Doc2Vec taraining models: Paragraph-Vector Distributed-Memory or PV-DM and Paragraph-Vector Distributed Bag-of-Words (PV-DBOW). In PV-DM, word order is incorporated while training and each sequence of words is mapped to a sequence vector as document vector and each word is mapped to a unique word vector. Moreover, document vector and word vectors will be aggregated to predict a target word in the context. On the other hand, in PV-DBOW, words are randomly sampled and only learns a vector for the entire document and creates Paragraph Vectors by training to predict words within a window of the paragraph. There are two typical methods for learning a user embedding from such vector representations: either concatenating all the posts from the same user (User-D2V), or simply deriving a user embedding from all the post vectors from the same person using some pooling methods (Post-D2V). Deep neural network models as well has been used to obtain user generated text embedding such as Long Short-Term Memory (LSTM) was used to to capture the sequential relations between words and posts during the learning of text representation vectors [259].

User Network Embedding basically is construction of a function that maps raw user features in a high dimensional space to dense vectors in a low dimensional embedding space. For example, the video recommender system of Youtube by embedding the users, videos and queries in the portal [50]. likewize, sarcasm detection in social media task has been done by modelling context with User Embeddings [11]. There are four most widely used network embedding methods for User Embedding DeepWalk [175], Node2vec [82], deep neural networks and Matrix Factorization.

In **Matrix Factorization** the connections between nodes represented by a matrix and then the matrix will be factorize to obtain the embedding. The matrices used to represent the connections include node adjacency matrix, Laplacian matrix, node transition probability matrix, and Katz similarity matrix [79]. As example, Wang et al. used a Non-Negative Matrix Factorization (NMF) which is matrix decomposition method that all the entries in all the matrices have only positive values to learn a network-based user embedding that preserves both the first and second-order proximity [234].However, matrix factorization are computationally expensive for large scale user data [79], Therefore, Other methods were proposed such as DeepWalk [175], Node2vec [82].

DeepWalk use truncated random walks to learn latent representations of vertices in a network. Firstly, it generates short random walks and each random walk is treated as the equivalent of sentences.Then, it performs the optimization over sum of log-likelihoods for each random walk. After that, SkipGram model (SG) is used in word2vec to learn the latent representation of a vertex [175].

In **node2vec**, a biased random walk is employed to nterpolate between Breadth-first Sampling (BFS) and Depth-first Sampling (DFS) graph searches. With balance between breadth-first approach and depth-first approach random walks, Node2vec can better preserve both the second-order and high-order proximity [82]. After generating neighboring vertices by biased random walk, the vertex representation will be learned using the SkipGram model (SG) [82]. For example, it has been used to represent users to Fake News Detection [87].

Deep neural networks based methods likewise applied to graphs to generate user embedding. For example, Wang et al. proposed Structural Deep Network Embedding (SDNE) [231] which is consisted of semi-supervised and unsupervised models to preserve the first and second order network proximities. An autoencoder with non-linear functions that able to capture the highly non-linear network structure to find an embedding for a node which can reconstruct its neighborhood. The unsupervised model to capture the global network structure. Another example is Deep neural networks for learning graph representations (DNGR) [37], which contain three components: random surfing to generate a probabilistic co-occurence matrix to the input graph, calculation of positive pointwise mutual information PPMI matrix, and feature reduction by applied stacked denoising autoencoder to obtain the embedding.Graph Convolutional Networkswas also used to generate User Embedding [114] to overcome the problem of computationally expensive for large sparse graphs by aggregating embedding of local neighborhoods of nodes. The general network architecture of Deep User Perception Network (DUPN) takes user behavior sequence as input and transfers each behavior into an embedded vector space and the apply LSTM and attention-based pooling to obtain a user representation vector [163]. Recently, Author2Vec proposed to generating User Embedding by incorporate sentence representations generated by BERT with a unsupervised pre-training objective, authorship classification [243].

Fusing multi type of information such as text, attributes, and image, has been applied to produce User Embedding. For example, User Profile Preserving Social Network Embedding (UPPSNE) was proposed to learn a joint vector representation of a user by combining user profiles and social network structures [256]. Community-enhanced Network Representation (CENE) [222] as well utilized both network link information and text information to learn embeddings of both vertices and communities. Text is modelled as a special kind of nodes, and then optimizes the probabilities of heterogeneous links.

2.3.2 User modelling based emotion analysis

As emotions are a cognitive process and largely subjective, user bias will play a significant role in emotion modeling. It is noticed that the user subjectivity has been ignored by most of emotion classification models. As well as user modelling-based emotion analysis is still in its infant stage.

The early stage of User profile based emotion analysis represented by using the social network data like twitter to utilize user-user relationships and user-tweet sentiment to improve the user-level sentiment analysis [214]. User intersubjectivity network (UserInter) is a network that proposed with linking review writers, terms they used, as well as the polarities of the terms to learn writer embeddings. Aftr that, the network use max-pooling layer of a CNNs to classify the writer embeddings [84]. Shen et al. proposed a dual user and product memory network (DUPMN) model to learn user profiles and product information for reviews classification using separate memory networks. The DUPMN model capture user profiles and product information by utilising two separate memory networks: a User Memory Network (UMN) and a Product Memory Network (PMN) [199]. Similarity, a novel User-Product gated LSTM network (UP-LSTM) is proposed which incorporates user and product information into LSTM cells at the same time of generating text representations and it can dynamically produce user- and product-aware contextual representations of texts [221]. For financial-related sentiment analysis, Wang et al. proposed hybrid user and topic embedding (UTE) approach and then applied a deep contextual neural network architecture [230].

However most of the work of User modelling based emotion analysis has been done for review of products or media or to recommendation.

Chapter 3

The Mental Health Social Network Dataset

3.1 Introduction

With the development of the Internet, researchers have examined social interactions which happened naturally "in the wild" on social media platforms to infer some understanding of health behaviour as expressed online. Social media provides easy-to-access peer-to-peer support online communication platforms in real-time at low-cost to support users financially, emotionally, physically ,or socially [179] and help them to describe their experiences in healthcare. Online peer-to-peer mental health support such as TALK-LIFE(talklife.com) and 7Cups(7cups.com) have provided new pathways for seeking social support and that allows users to engage in the transference of informational and emotional support and share their experiences with others globally [178].

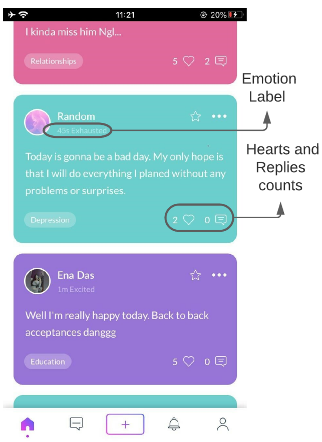
As discussed in the previous chapter, social networks has contributed to improve people's mental wellbeing and provide valuable opportunities for researchers to model users in different domains. More specifically, online peer-to-peer mental health support can offer new mechanisms for measuring and understanding mental health as users of these networks are more likely to express their thoughts emotionally. Moreover, users engagement via posts and other interaction patterns can reflect the internal emotional states of users.

To model users' well-being, their data is crucial to understand and predict their behavior or emotion. The availability of Social networks provides emotion and user modeling research communities with massive datasets. With the advances in machine learning models, these datasets made a distinguished improvement in modeling users and users' emotions. However, After reviewing most of the existing available datasets in the previous chapter, it is evident that most existing datasets are primarily small and not enough to train machine learning models efficiently or weakly labeled based on emotion-related hashtags or emoticons. Moreover, most datasets in modeling emotion are either manually annotated by professionals or automatically annotated by technique designed by annotators. These techniques could wrongly decide the feeling of the user. Therefore, the best way to model emotion precisely is to let user label their posts [194]. Unfortunately, there is a lack of massive data set with self-reflected labels in this area. The dataset provided by this thesis will fill this gap.

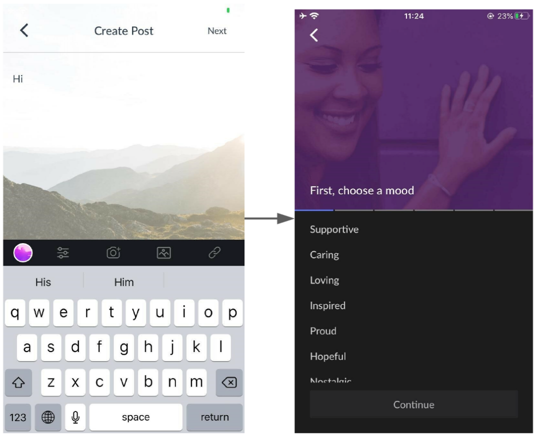
We have been got access to a new dataset of peer-support online mental health social networks with millions of labeled posts in TalkLife. This dataset is self-labeled by the users of this network. This makes it different than most social networks datasets as how humans generally express their emotions might not align with labels produced by distant supervision using hashtags or other methods. Moreover, specific emotion labels have been determined to users to choose between them instead of letting them express by their own words, and this way has been proven to be more efficient by researchers [194]. Throughout this chapter, this dataset will be discussed in detail.

Talklife, a peer support social network for mental health founded in 2012, has been designed as a safe space for people to be open about their feelings and seek emotional support from others. As a result, users of this platform discuss diverse topics from general social network discussion to depression and life distresses. Talklife forum usage is characterized by individuals posting a message, often asking for advice while dealing with deep emotions, and other users answering questions or participating in a discussion in the form of replies to comments on that initial post.

3.2 The scial Network App



(a) TalkLife posting Feed



(b) Posting with Emotion Picker

Figure 3.1: Screenshots of the interface of TalkLife App

Talklife platform is a free peer-to-peer network for mental health support that lets users share their feelings and frustrations without fearing a negative backlash. Users are required to register to use the app. However, verification is not essential, and there is no option for users to connect to other social accounts, thus leaving room for anonymity. The usage is characterized by individuals posting a message, using the interfaces shown in Figure 3.1b to share their thoughts or things that happen in their daily life or to ask for help or advice. After creating a post, they proceed to the following interface to choose one of 61 predefined Mood labels, including Heartbroken, Sad, lonely, Depressed, Stressed, Confused, and others. The label “Meh” expresses a lack of interest or enthusiasm, which is interpreted as a miscellaneous label for messages that the user couldn’t fit into another category.

Then users are required to select a category for their posts with a range in specificity from the narrower category “Parenting” to the all-encompassing “Others”. Users can browse through the feeds of other users (see Figure 3.1a), and interact with them by commenting or reacting to their vents via a set of emotions. Users can upload their status, send messages to the people they are connected with, follow other users, and set up a profile for themselves with photos and demographic information such as gender, age, and status.

The platform consists of discussion threads. Typically, the original user started the thread with a single post of comments with discussion or question and several community members' responses, which are generally relevant to the original post in the discussion thread. The original post is associated with two main features: category and mood. Responses in a typical network thread follow a hierarchical structure, the responses to the original post are called answers, and the responses to the answers are called replies. Responses posts have no affiliated emotion labels, and they are tagged in the same category as the original post within the thread. However, this thesis will consider both answers and replies under the umbrella term of “responses”.

3.3 Dataset

We extracted data from the TalkLife platform from August 2011 to August 2020. Access to the dataset has been done by logging into the TalkLife server with a unique Username and Password obtained from TalkLife with consent. SQL Management Server and Python have been used to collect the user data.

Privacy, Ethics and Disclosure: All data analyzed were sourced from the TalkLife platform. All personally identifiable information was removed, and formal license agreements have been signed-in accordance with ethical data usage and GDPR regulations. All direct identifiers of users have been removed, as well as shared URLs and usernames. In addition, all unique identifiers have been masked to prevent user linking. This work does not make any medical treatment recommendations or medical diagnostic claims.

Structure of the dataset

The provided dataset is structured in tables, each containing different entities such as questions, Answer, Reply, and users. Entities referenced via the unique identifiers (ID) are described as follows:

- User: the user table contains all user's information and profile. It includes User ID, Gender, Date of Birth, How many users are following this user, The amount of posts and comments the user has made, User account creation timestamp, User Type ((Volunteer, Moderator, Administrator, and Standard User), and other information.

- Question: the Question table is the all the posts in the network, and it contains all the information about the post such as Post ID, Post Content, Post Category ID, Post Creator ID (the poster), Post Time Stamp, Post View Count, The amount of hearts and gifts, flags a post has received, HashTags, Mood and other information.

- Answer: it contains comments to the post, which are marked with the ID, The content of a comment/post, The ID of the original post, Comment Creator ID, Time Stamp, and other information.

- Reply: similar to Answer, however these are replies to Answers rather than the original post directly.

- USER ACTIONS includes components such as Like, Gift, and Follow tables.

3.3.1 Dataset Exploration

Data	Statistics
Number of Threads	9,849,271
Number of all Posts	43,990,097
Number of Original post	12M
Number of labelled Original post	10M
Number of responses	34,140,826
Number of Users	1,106,836

Table 3.1: Statistics of TalkLife dataset

Table 3.3 shows summary statistics for TalkLife dataset. It contains more than 43M comments, 22% of them are original posts, and the rest are responses. Not all posts are labeled. The labeled posts make up around 81% of them.

There are over 1M users registered in the dataset. Male represent 20% of them and female are 40% of users while the rest specify their gender as ”others or prefer not to say”. For the user types, there are around 150 users who are categorized as

”Moderators”, ”Buddies” and ”Administrators” the rest of the users are standard users who use the network to discuss their wellbeing and other everyday activities and thoughts.

Figure 3.2 shows the distribution of users’ age, and it is clear that the highest density in the age is around 20 years, whereas age over 40 and less than 18 is scarce. The mean is 23 years, and 25 years represent (75-th percentile).

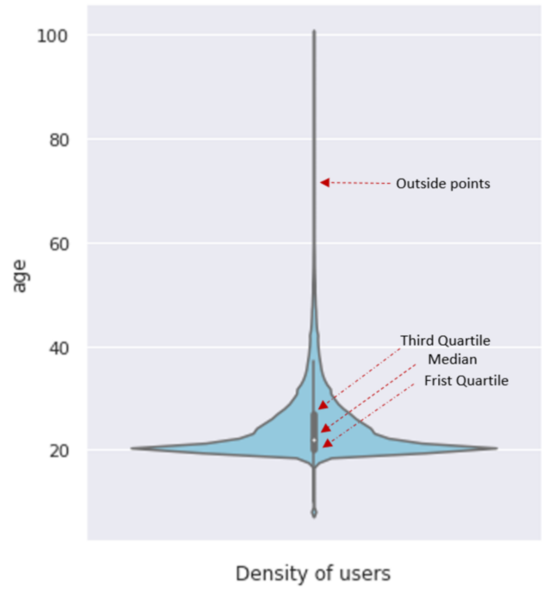


Figure 3.2: Distribution of TalkLife users’ age

Figure 3.3 shows the cumulative distribution function (CDF) of the number of posts and replies per user after filtering users without posts. Both are governed by heavy-tailed distributions that span four orders of magnitude, indicating that a vast majority of more than 80% of the users posted less than 10 posts while a small group of few users posted more than 100 posts. A similar pattern holds for the number of replies per user. In our dataset, around 106K of users posted without any replies activities. In addition, 90% of the rest with less than 100 replies. We noticed very few of them with more than 1000 replies.

The Social Network

Users can form social links to other users on the platform. However, There are 697K users without any following activities, and only 37% of the users follow. Therefore, they constructed a social network with approximately 400K of nodes

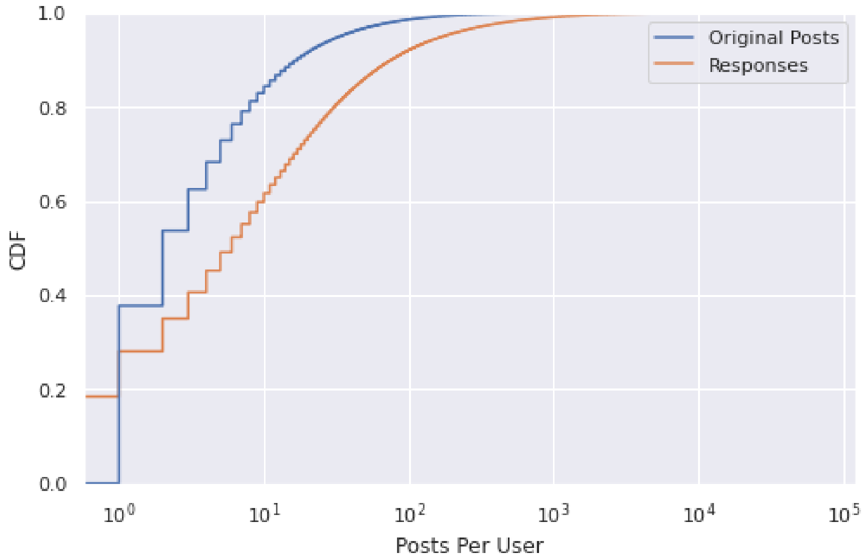


Figure 3.3: Cumulative distribution functions (CDFs) of the number of posts per user

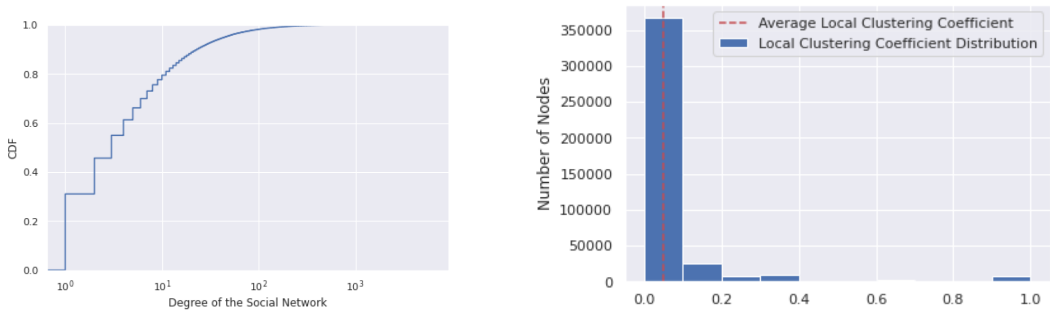
(users) and around 2.5 million edges (directional links between them). If we consider G as a directed graph of the network, it is a pair (V, E), where V is a finite set of users and E is a follow relation on V. The main network indicators of the graph listed as follows:

number of nodes V is 419,405, number of edges E is 2,518,335, Average degree which is the average number of edges per node in the graph is 12, density D is 0.000028 which is meaning the network is not a dense network at all.

Figure 3.4a shows the degree distribution of the social network. A typical heavy tail behavior is observed, with a vast majority of users linked to a small number of other users and a small minority of users having more than 100 links. Local clustering coefficient distribution is shown in Figure 3.4b, and it indicates the extent to which any given node is located within a tight 'cluster' of neighboring nodes. Most users have very small or no ties that exist between the neighbors. In contrast, a small minority of users have increased weighed coefficient, reflecting users' strong ties to neighbors that are themselves connected.

Temporal dynamics of the network activities

Figure 3.5 shows the number of posting activities per month, including original



(a) Degree distribution of the social network

(b) Local Clustering Coefficient of the network

Figure 3.4: Distribution of Degree and Clustering Coefficient

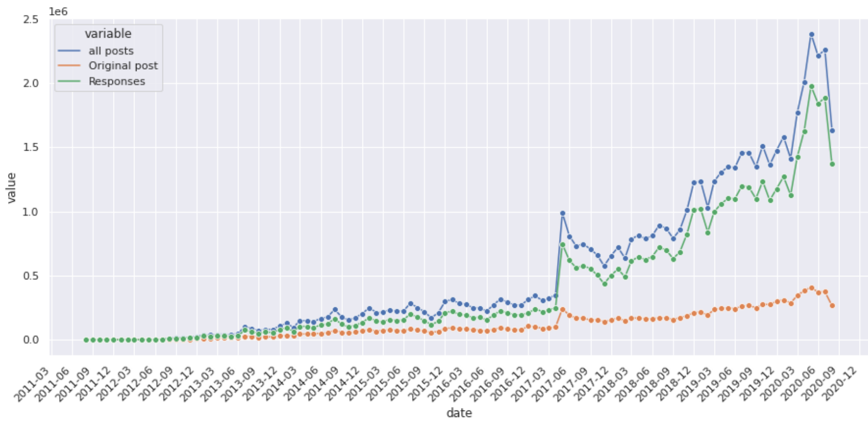


Figure 3.5: Aggregated monthly posts in the dataset

posts and responses. The activity shows a modest increase in the number of posts from September 2013 and continued to flip up and down until it made a sharp increase from 250K to 1M in May 2017. Then, it slowly decreased and increased until it is reached a million again in the end of 2018. At the end of 2019, the number of posting reached 1.5 M; after that, it sharply increased until a peak of activity was reached around May 2020. That maximum of activity comprised more than 2.4 million posts during May, June, and July. It seems to be a pandemic, and lockdown contributes to increasing engagement activities as the peak was reached during the pandemic. Then the engagement rate went down sharply to 1.6M in August when most countries have eased the restrictions.

No significant differences are observed in patterns of change in original posts and responses ($p > 0.05$). The notable change is in the quantities of increases as the answers made considerable variances in the number of increases. For example, during 2020 and the peak of activities, original posts increased from 300K to 400K, whereas responses increased from 1.1M to 2M.

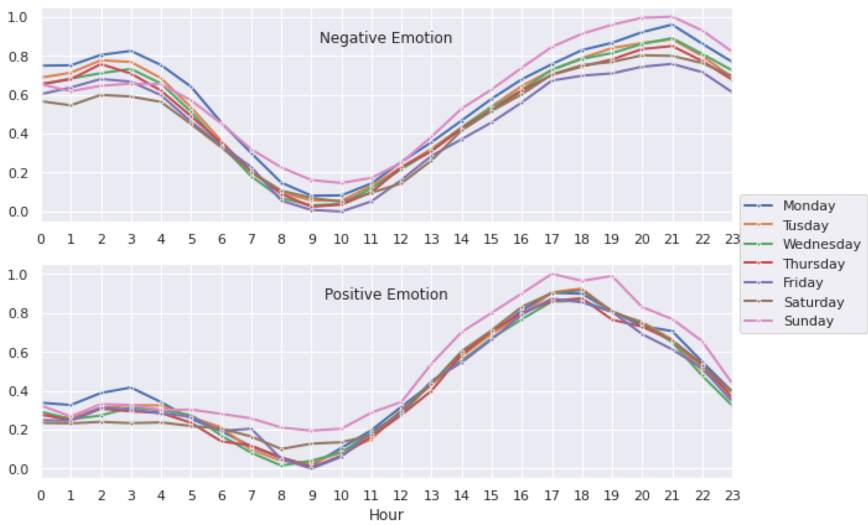


Figure 3.6: Aggregated posts per day of the week and per hour of the day

Figure 3.6 shows patterns of posting regarding positive and negative moods during the day of the week. There is no significant difference between the two groups. The general trend shows both negative and positive emotions posts appear to be low during 8-11 am and this is might be because the beginning of the day to work or study at the morning for most of people. People mostly busy preparing or travailing to work, eating breakfast or for example preparing their kids for schools. Therefor, we can see from 10 am, the posting trend start to grow. In addition, we can see in Sunday, which is a weekend day for most of people, the posting line more than other days during this time (8-11 am).

However, there are few variances in terms of the peak of posting and the number of posts during the early morning hours. The general trend of posting for all groups increases from 11 am until reaching the peak at 5 to 7 pm for positive moods and 8 to 10 for negative moods. Then decreasing until reach the bottom at 9 am. The most significant difference between groups is posting during the night. People with negative emotions seem to be posting during the night more than people with positive emotions.

Regarding days of the week, for all groups, posting on Sunday from morning until midnight is more than other days, whereas after midnight until 6 to 7 am, posting on Monday is the most. The opposite for Friday seems to be the day where posting is the least most of the time until midnight. Whereas, from midnight until 6 to 7 am Saturday is the least.

3.3.1.1 Emotion Categories and Their Usage

Users must select one of 61 emotion labels when posting. Talklife app divides these moods into six thematically related pages, each consisting of words that evoke similar feelings. These groups have been given descriptive names for ease of interpretation in the work of [121]. Groups' name and associated emotions are shown in Table 3.2. Figure 3.7 shows the distribution of emotion labels in term of number of posts per emotion. In addition, it shows the groups of emotions, each group within a color. Categories of Sadness, Inadequacy, and Frustration considered to be "negative" emotions where Support, Relief, and Positivity to "positive" classes [121].

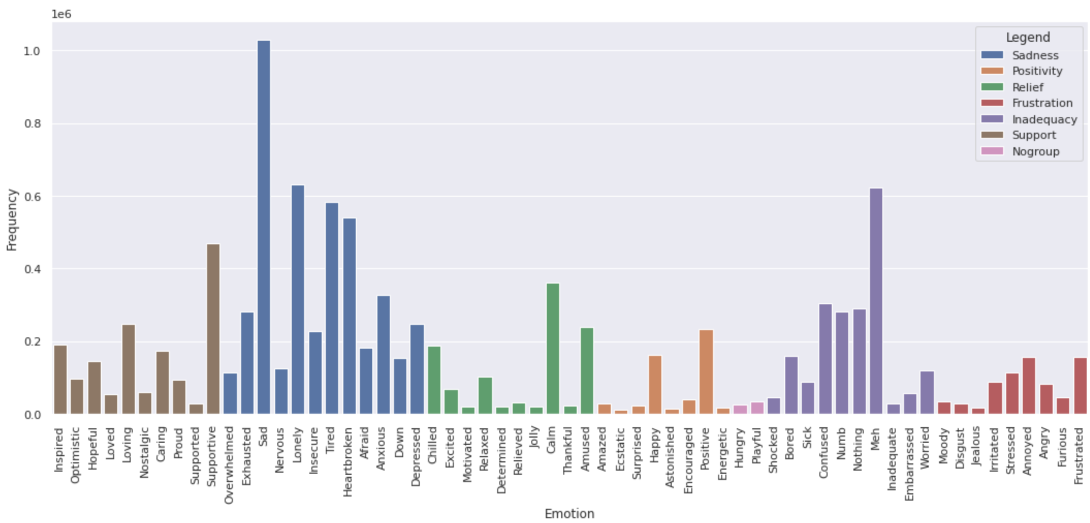


Figure 3.7: Histogram of number of posts per emotion label

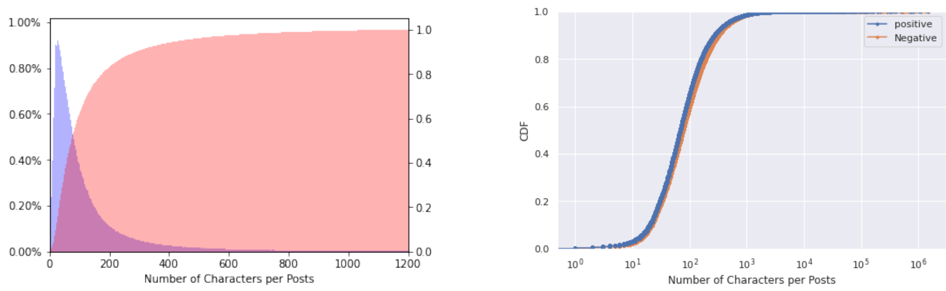
In Figure 3.7 we can see that the most frequent emotions posted were dominated by negative emotions, including: Sad, Meh, Lonely, Heartbroken and Tired

Group	Included Moods
Sadness	sad, heartbroken, depressed, anxious, nervous, down, lonely, tired, insecure, exhausted, overwhelmed, afraid
Inadequacy	Worried, meh, inadequate, numb, confused, embarrassed, shocked,sick, bored, nothing
Frustration	frustrated, annoyed, angry, furious, irritated, jealous, stressed, moody, disgust
Support	Supportive, hopeful, optimistic, loving, inspired, proud, nostalgic, caring, loved, supported
Relief	excited, amused, thankful, calm, relaxed, chilled, relieved, jolly, determined, motivated
Positivity	astonished, positive, surprised, encouraged, happy, amazed, ecstatic, energetic

Table 3.2: Mood groups of TalkLife grouping scheme with associated mood labels

. In addition, the label 'Meh' expresses a lack of interest or enthusiasm. In addition, there is a significant disparity in emotion frequencies (e.g., admiration is 30 times more frequent than grief).

3.3.1.2 Analysis of posts text



(a) Chars histogram and CDF for the number of chars per post

(b) Chars per post (CDF) of the number of Chars per Negative and Positive posts

Figure 3.8: Distribution of the number of chars per post

We take a preliminary analysis to examine some of the textual properties of the posts. Network on Talklife tend to be short and conversational, with an average length of 151 characters and 75-th percentile of posts within 150 characters. Figure 3.8a shows the distribution of the number of characters per post in general which have a well-defined typical length and indicates that a vast majority of posts(more than 80%) with less than 200 characters. Posts with more than 1000 lengths are infrequent, with around 148K. Figure 3.8b shows the cumulative distribution function of the length of the post grouped by emotion category. There are no huge variances. However, the positive emotions have more probability mass for short posts more than negative types. On average, Positive posts are 143 characters long (with 75-th percentile of 133), whereas Negative posts are 154 characters long (with 75-th percentile of 158).

There are few posts with extreme length values, such as 7 posts with more than a million characters and 30 posts with more than 500k length, whereas 130,493 empty posts and 26,041 with one character. Figure 3.9 shows a sample of the post with more than 10K in length, and it is clear that almost all posts like this are fake.

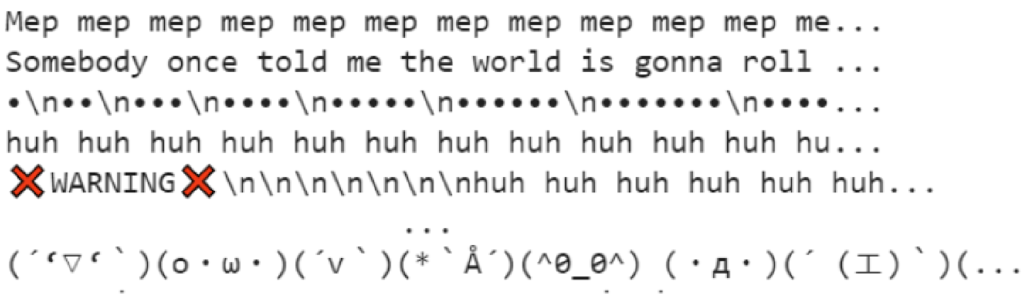


Figure 3.9: Sample of posts with more than 10K characters

Word clouds are also generated from the posts associated with two contrasting groups of emotions: negative and positive. Figure 3.10 shows these two-word clouds generated using Python package. It is clear that the word 'Like' seem to be the most word used in the both categories with equal frequencies. This is might be because the 'like' word give positively meaning with positive group, and it might be used in the phrase 'I would like' to express needs in negative group. In additions, it might be used to give examples in both groups.

Positive words 'love', 'life', 'good' and 'happy' are presented in the positive map with great frequencies more than negative map. These words express positive feeling more than negative feeling.

On the other hand, 'want', 'help' and 'feel' words have been stated in the negative group more than positive group and it might be because people with negative emotion talk more about their feeling and ask for help. In the same time, there are negative feeling words like 'hate' and 'bad' noted in only the negative map. Likewise, swear words and harm words like 'kill' have been stated in the negative map.

(a) Positive Emotions (b) Negative Emotions

Figure 3.10: Word clouds for positive and negative categories

3.4 Dataset constructing

The use of 61 emotion labels to model and classify will be expensive and difficult and reduce the model's accuracy. Also, the variation between using labels has been found clearly as in Figure 3.7. Therefore, Simplifying the labels are necessary. The classes or groups of emotions that seem to be not used in most literature. In addition, we the need for labeling with scores to model the difference of change between emotions. Using polarity labels like positive, negative as a discrete emotion model has been applied in literature as the work of [112] [229]. Thus, the labels are transferred to their polarity to 5 categories (very positive,

positive, neutral, negative, very negative).

Group	Included Moods
very positive	Proud, Thankful, Amused, Loving, Positive, Happy, Inspired, Excited, Supportive, Amazed, Playful, Supported, Loved, Optimistic, Motivated, Ecstatic, Energetic, Jolly, Hopeful
Positive	Encouraged, Calm, Astonished, Relaxed, Caring, Relieved, Determined, Chilled
Neutral	Empty, Surprised, Meh, Embarrassed, Nostalgic, Down
negative	Hungry, Tired, Bored, Jealous, Confused, Numb, Overwhelmed, Depressed
very negative	Stressed , Shocked , Frustrated , Anxious , Annoyed, Irritated, Exhausted, Sick, Afraid, Lonely, Insecure, Worried, Nervous , Furious, Sad, Negative, Angry, Heartbroken, Nothing, inadequate, Moody, Disgust, Inadequate

Table 3.3: Mood groups with polarity grouping scheme and associated moods

Each emotion label is mapped to it's sentimental score. Then, we map the continuous label values onto a small set of 5 discrete categories: 1=very negative, 2=negative, 3=neutral, and 4=positive, very positive=5 as in Table 3.3. This labelling method will be summarized to be 3 classes (Negative, Positive and Natural) when it is needed. The sentimental scores are calculated using Stanford NLP Core [142] and also verified with other available online tools such as Twinword [1].

Figure 3.11 shows category densities as each color represents an individual category of emotions, with its total size being the percentage of the emotion across all posts in that month. It is clear that the very negative class dominates other groups most of the time, and it might be because the network cares of people with mental health problems. Moreover, people seem to be posting when they are in negative mood more than when they are feeling positively to ask for help or seek for advice or just to talk.

Furthermore, regarding the very positive category, from the end of 2019 until end of 2020 it increased until becomes equal or slightly more than the very positive

[1] $https://www.twinword.com/api/sentiment-analysis.php$

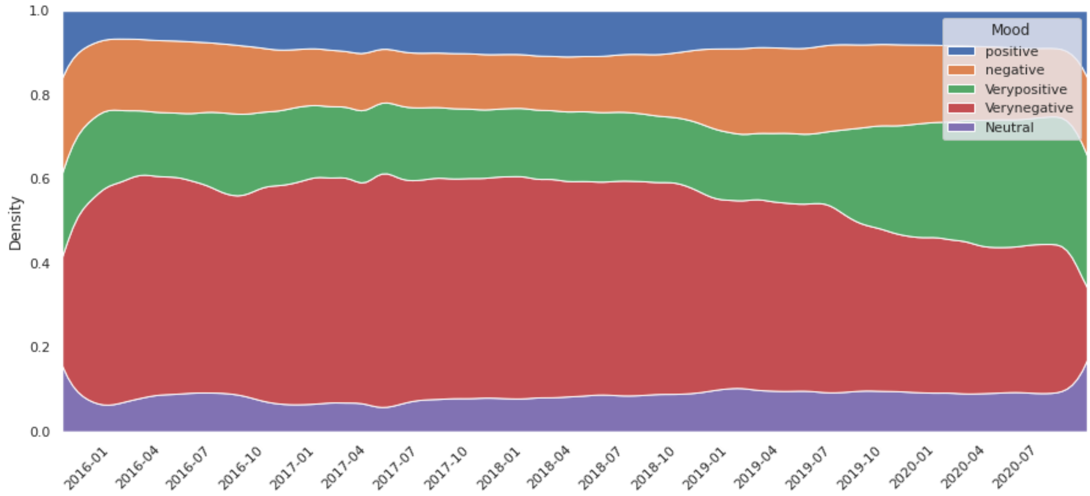


Figure 3.11: Category-Level densities of emotions on a month-by-month basis

group. This time was during the lockdown due to the COVID-19 pandemic. Thus, people with positive mood seem to be posting more during the lockdown to help others with negative emotion or to encourage other users to be in good feeling during the pandemic.

3.4.1 Dataset Filtering

The dataset is extensive and with missing and nosing data. Thus, filtering will be applied in different levels like users, posts and date.

Time filtering

The network dataset posts start from August 2011 until August 2020. It has noticed that before 2016, the posting activities are too low and most of mood labels have not used by users whereas after January 2016 the regularity of using emotions has been noticed. Likewise, it is shown in Figure 3.11 categories of emotions become defined and relatively stable from 2016. Therefor, posts before 2016 will be filtered.

Users filtering

Some users placed an unreasonable date of birth values for themselves; thus, their age comes to be a negative value or over 200 years. Therefor, users who claimed that their age over 100 years and below 8 years have been removed.

Moreover, As there are different users' type in the original dataset, some of them "Moderators", "Buddies" and "Administrators", we will focus on "Standard users". Thus, non-standard users have been removed from the dataset. Furthermore, in the dataset, 47% of the users without posting activities; therefore, they are removed from the dataset.

Posts filtering

Posts with extreme length values, such posts with more than 500k characters length have been removed. Likewise empty posts and posts with one character have been removed. Then, the probability distribution of the number of words (tokens) per post has been calculated and the 75th percentile value was 32. Therefore, it has been used 32 as the maximum number of tokens per post in alignment with the 75th percentile in the length distribution.

3.5 Conclusion

This chapter has provided a comprehensive overview of the dataset, which will be utilized as the primary data source throughout this thesis. With millions of labeled posts and many users, this dataset provides a unique opportunity to delve deep into users' behaviors and the link to their emotions based on their regular interaction on this platform, hoping that this data set encourages further research on emotion analysis. Therefore, extensive studies on this large data set have been done.

The next chapter will utilize users' posts to model emotions using a popular transformer model (BERT).

Chapter 4

Exploring Language Models for Text-based emotion modeling

Abstract Text-based emotion detection is one of the most challenging problems in natural language understanding. Therefore, building a system that understands the context of the sentences and differentiates between emotions has recently motivated the machine learning community. In this chapter, we explore Bidirectional Encoder Representations from Transformers (BERT), one of the most popular pre-trained language models based on Transformer, on both the tasks of sentiment analysis and emotion detection. We have evaluated the BERT model on the network dataset. And we observed that the BERT system shows substantial improvements (F1-Score= 0.67) over the baseline models provided.

4.1 Introduction

With the introduction of BERT, using pre-training models to solve text classification problems has become popular. Initially, pre-trained language models have been applied as encoders to enhance performance and address problems related to natural language understanding [59]. In addition, Many studies have showed that pre-training Language Models are highly successful in different NLP tasks, including text summarizing [135] [134], Text Generation [258] , translation [264], question answering [249], and text classification [211]. However very recently, there have been few attempts to apply pre-trained models to text classification tasks such as predicting whether any two sentences are in an entailment relationship; or determining the completion of a sentence among four alternative sentences, text emotion classification is still in its infancy and requires further research which is very challenging due to its complexity and subjectivity. Thus, the influence of language model pretraining on text emotion analysis will be examined in this chapter.

Although language models is pretrained, they could be fine-tuning for a specific task and their parameters can be jointly fine-tuned with additional task specific parameters. In contrast to other deep learning models where the parameters is specify in the task. However, there is few research to enhance BERT to improve the performance on text based emotion classification tasks. Therefore, investigating how to maximize the utilization of BERT for the emotion modelling task has been attractive. In this chapter, experiments will be design to make a detailed analysis of BERT and its variants. The utilization of BERT and other classical NLP approaches will be empirically examined in the network dataset for first time to emotion analysis. Furthermore, the behaviour of BERT against traditional NLP models will be utilized. Different experiments about text based emotion analysis to adding empirical evidence to support the use of BERT on NLP tasks.

With utilizing Language Models for a large scale social network dataset and comparing BERT against other NLP approaches for emotion analysis in this dataset, the contributions of this chapter is to;

- explore ML models and NLP approaches with large social network dataset,
- analyze the efficacy of utilizing transformers and other NLP approaches on the social network dataset for modelling emotions,
- demonstrate the performance of BERT by fine-tuning on the dataset to emotion analysis,
- compare BERT model with a traditional machine learning NLP approaches.

The chapter organization is as follows; section 2 discusses related works; section 3 highlights the emotion detection pipeline and the model architecture. The experimental setup of the study, the dataset used, and how the model was evaluated are presented in section 4. The results obtained and their corresponding analyses are presented in section 5. In section 6, the conclusion and future works are highlighted.

4.2 Related works

Various neural network models have been proposed to perform textual emotion analysis, and provide promising performance in this classification task. For example, SVM has been utilised to classify emotions from blogs with F1 ranging from 0.493 to 0.751 for each class [1]. The features were corpus-based UniGram. Among deep learning models, LSTM have been proved to be useful to modelling long and sort time dependencies in text emotion classification [235] [42] and give more improvement when combine it with other lyres of deep learning networks such as LSTM and CNN [100] [92] [6] [235] [42]. Also, GRU models some times out-perform LSTM networks [107] [266].

Recent works have explored the use of deep representations for text for emotion analysis in contrast to low level representations. Rule based was compared with TF-IDF features and the best accuracy is 0.83, obtained by using SVM algorithm and unigram with TF-IDF features [18]. GloVe is deep learning technique which has been proved that it plays a vital role in improving the accuracy of sentiment classification with deep learning model [261] [236].

To classify text with BERT, normally there are tow different ways, as word-embedding or fine tuning approach. In the first approach, pre-trained BERT used to extract word embeddings from text data and then word embeddings can be used in ML model for specific tasks like classification, topic modeling, summarisation ..etc. For example, Kazameini et al. [111] after extracting contextualized word embeddings from the text, bagged-SVM classifier was used to predict the authors' personality traits automatically. They predict the personalty from the writer essays by dividing them into sub-documents, preprocessed, and fed into a BERT base model. Word embeddings vectors for the document then were fed into ten SVM classifiers to produce a prediction. They show that BERT increased performance of 1.04% in comparison with baseline methods. In the fine tuning approach, a dense layer to be added on top of the last layer of the pretrained BERT model and then train the whole model with a task specific dataset. Sun et al. investigated fine-tuning BERT for the text classification task and they achieved state-of-the-art performances on eight widely studied text classification datasets [211].

Regarding modelling emotion, BERT achieved state-of-the-art results in based Contextual Emotion Classification. For example, in the EmotionX shared task, which is to capture contextual information, to understand informal text dialogues outperforms the previous state-of-the-art model and shows competitive performance in the challenge [247]. Huang et al. investigated the emotion recognition ability of BERT on continues dialogue emotion prediction tasks, and they achieved the best results in the testing dataset [101]. However still these works on the dialogue level which depend on various aspects of language such as utterance and different background of the conversation, tone of speaking or personality. Whereas, in this chapter, BERT will be investigated in social network post level.

Using BERT for emotion modelling has been explored by adding a classifier on the top of BERT for predicting emotion classes on the ISEAR dataset, and it is outperforming that of the state-of-the-art results in ISEAR dataset [5]. However, ISEAR dataset contain just 7666 sentences and BERT approach need huge amounts of texts to deliver proper results.

Santiago and Eduardo appear to be the first and only work that compared BERT against traditional machine learning text classification methods [78]. In simple terms, they empirically test BERT against Term Frequency - Inverse Document Frequency (TF-IDF) and they proved superiority of BERT in four experiments. However, this work just examine BERT with (TF-IDF) whereas in this chapter, multiple of NLP approach will be utilized. In addition, their experiments with different datasets for different text classification purposes except of emotion analysis and most of them in level of documents level.

4.3 Methodologies

After prepossessing data, Different encoding techniques have been explored to convert the text into vector representation such as using TF-IDF, static word embedding (Glove), or contextualized word embeddings. These techniques have been used in combine with different machine learning models. After that, the results has been analysed to compare the contextualized word embedding like BERT with other models.

4.3.1 Text Representation Approaches

To represent the text statistical and embedded representations method were applied, the following 4 different algorithms were selected as baselines:

- Bag-of-Words (with Unigram): this approach is a simple and effective representation model for natural language. Each post is represented as a sparse vector containing the counts of every word. We limit the vocabulary size after filtering English stop words using Scikit-learn's list [172] but without applying any transformations like stemming.

- TF-IDF:]TF-IDF algorithm is used to weigh a keyword in any content and assign importance to that keyword based on the number of times it appears in the document [188], the following equation used to compute inverse document frequency as:

  $$IDF(t_i) = log\frac{N}{n_i} \tag{4.1}$$

  Where N is documents in the collection, and the term t_i occurs in n_i of these documents. Then, TF-IDF is defined for a given term t_i in a given document as follows:

  $$TF - IDF(t_i) = TF_i \cdot IDF(t_i) \tag{4.2}$$

  Where TF_i the frequency of the term t_i in the document.

- GloVe [174]: a popular static word embedding model based on two prominent model families: local context window and global matrix factorization methods. Rather than training on the entire sparse, it only trains the nonzero elements in a word co-occurrence matrix. Each element X_{ij} in the matrix represents the frequency of the word w_i, and the word w_j co-occur in a particular context window. As synonyms and similar words usually have similar contexts, GloVe mapped them to feature vectors close to each other. GloVe embedding consists of a 300-dimensional vector. Pre-trained model of GloVe with 300 dimensions computed in this experiment[1].

- BERT: BERT, which stands for Bidirectional Encoder Representations from Transformers. Unlike recent language representation models, BERT is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers. As a result, the pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, including the text classification.

4.3.2 Models

Five distinct classification models are applied:

- Conventional Machine learning models: two conventional machine learning models have been explored. The first mode, Random Forest(RF) is a powerful and popular machine learning algorithm. It creates several decision trees to classify a new object based on attributes. Each tree will predict an output, and the voting is done for the prediction to get the highest vote for a particular prediction class. The control over the model is less as there are fewer parameters to tune. The second model is SVM, as it has been proved to be effective on text categorization tasks and robust on large feature space [115]. It produces a hyper-plane near the extreme points of the dataset, which has one dimension less than the dimension of the data points. Support Vectors are the data points closer to the other classes, and they are pushing the boundary farther to make better predictions. These models have experimented with Bag-of-Words (Unigram) and TF-IDF features. RAPIDS cuML Python library used to perform fast modelling on a GPU. cuML can provide a speedup of 500x relative to scikit-learn [171].

[1] $https://nlp.stanford.edu/projects/glove/$

- Deep learning models: two deep learning models are selected as baselines: LSTM and GRU. They have been introduced in detail in chapter 2. They experimented with GloVe embedding as a basic word representation.

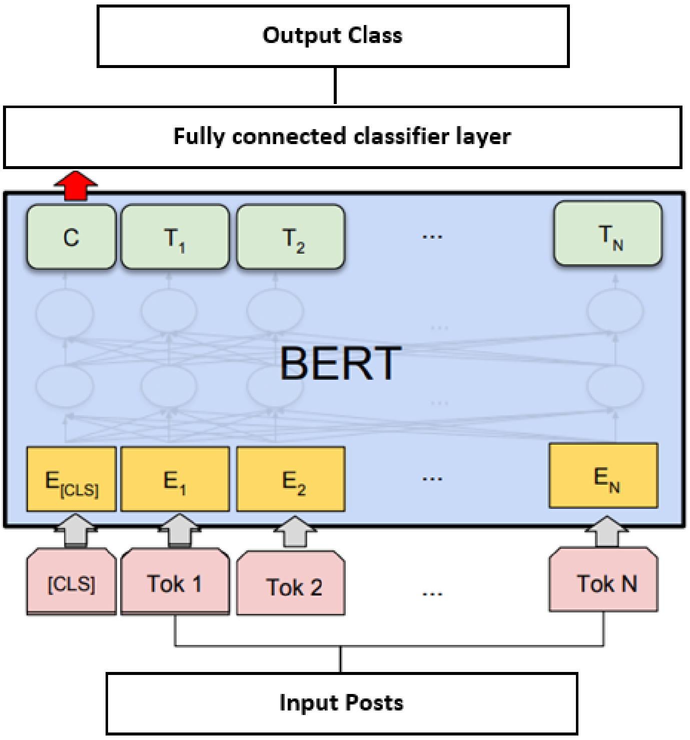


Figure 4.1: The architecture of the proposed BERT model

- Transformer model: BERT has been detailed explained in chapter 2. Although the BERT model is already pre-trained, it needs to be fine-tuned on the dataset to a specific task to be used to classify the text. As it is stated in the BERT paper [59], fine-tuning is straightforward since the self-attention mechanism in the Transformer allows BERT to model many downstream tasks. Therefore, as it is shown in Figure 4.1 a dense output layer has been added on top of the pre-trained BERT model from TensorFlow Hub[1] for fine-tuning, with a cross-entropy loss function. [2]

[1] *https : //www.TensorFlow.org/hub*
[2] In the experiments of this chapter, BERT-base model with a hidden size of 768, 12 Trans-

4.3.3 Evaluation metrics

To evaluate the performance of individual classifiers, the following evaluation metrics were employed:

- Accuracy: consider $P=\#$ of positive instances, $N=\#$ of negative instances, $TP=$ number of instances correctly labeled as belong to positive class, and $FP=$ items incorrectly labeled as belonging to the class. The accuracy is computed like:

$$Accuracy = \frac{TP + TN}{T + N} \tag{4.3}$$

- Precision:

$$Precision = \frac{TP}{TP + FP} \tag{4.4}$$

- Recall:

$$Recall = \frac{TP}{TP + FN} \tag{4.5}$$

- F1 score:

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{4.6}$$

4.4 Experiments

4.4.1 Data pre-processing

We have using bellow pre-processing methods:

- Normalization text by removing http, https, RT and username as they do not contain the sentiment information.
- Convert emoticon and emoji to the text describes expression of feeling as thay considers user moods expression in the form of icons.
- Tokenised the post data on white space, converted all words to lower case, Removal all non-ASCII and non-English characters.

former blocks and 12 self-attention heads has been used.

- Spell correction: This is a process of checking for the text's spelling to correct the wrong spelt text. A PyEnchant spell checker python library was employed to correct all the misspelt words.

- Removal all numbers as they do not significantly affect sentiment information.

- Remove one character words, punctuation, multiple spaces, apostrophe, and misspells,

- Replace contraction, and some noticed words

- Removal stopwords: Stop words usually refer to the most common words in a language. removed using NLTK libtary in python[1].

- Stemming and lemmatizing.

Standard pre-processing steps like stemming and removing the stopwords have not been used in BERT because it can get more features from the raw data. This is because stop words provide some context and as we also indirectly learn the sentence representation while feeding the context as the output or input. removing them can have a negative impact.

After the prepossessing and filtering phase, around 5.9 million posts with moods have been left. The dataset was split using random sampling technique into a training set of 76% data instances and evaluation set to be used during the training of 20%, and the rest 20% is test set.

Figure 4.2 shows the distributions of emotion classes on the data set. As discussed in the previous chapter, in Talklife App, the emotion labels are grouped into 6 classes.

4.4.2 Results

Experiments settings proposed in details in appendix .1. All the experiments were replicated 10 times (using the 10-folds method) for each model, with recording

[1] $https://www.nltk.org/$

Figure 4.2: Pie charts showing the distributions of the dataset with 3 classes scheme and with 6 classes scheme

the average performance for each replication, the mean average accuracy, F1-Measure, precision and recall values observed over 10 folds were reported.

Experiments with grouping of emotions

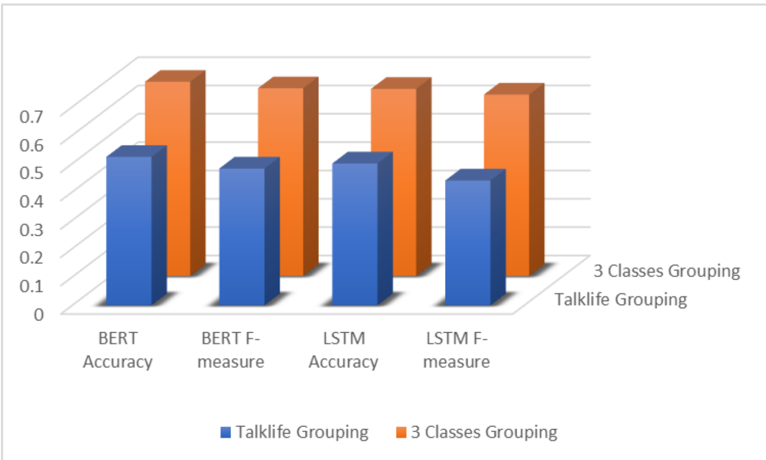


Figure 4.3: Results of experiments with two types of labelling methods of the dataset (3 classes method and 6 classes method of TalkLife app)

Figure 4.3 shows the results for a Talklife 6 grouped model (F1-score = 0.48 with BERT and 0.44 with LSTM) and 3 grouped model (F1-score = 0.66 with BERT and 0.64 with LSTM), respectively. The significant difference in performance for deep learning models between these models is perceptible.

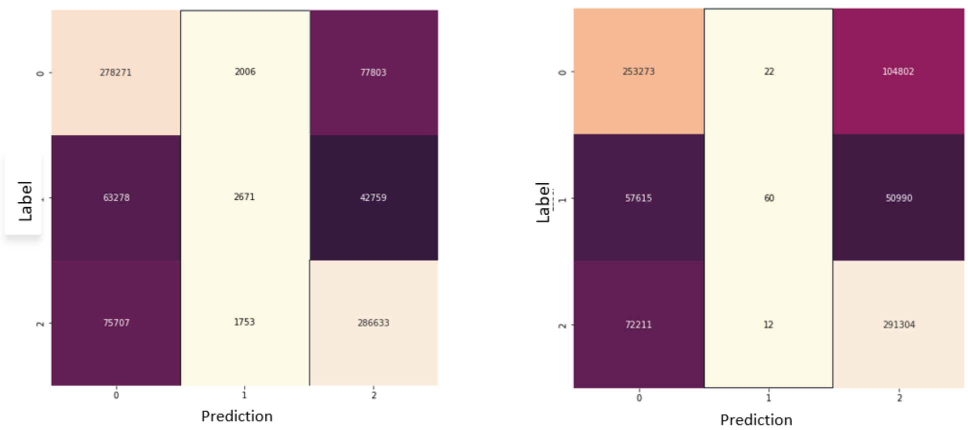
Experiment in comparing models with fine-tuned BERT model

Table 4.1 summarizes the five models' performance (RF, SVM, GRU, biLSTM and BERT). The model that achieved the highest scours in all metrics was BERT, whereas the model that obtained the worst result was RF +TF-IDF—although using BoW text representation improved the performance of RF and SVM.

Conventional machine learning models (SVM and RF) did not perform well on the dataset, with an average accuracy lower than 0.51 and an F-score lower than 0.45. On the other hand, in terms of deep learning models (LSTM and GRU), it is clear that they improved the performance by around 16% of accuracy and by 19% of F1 score. Moreover, biLSTM presents a slightly better performance than GRU. Figure 4.4 illustrates the confusion matrices yielded by the LSTM and BERT models, and it shows an improved classification with BERT in all classes.

Models	Accuracy	Precision	Recall	F1
RF(BoW)	0.4955	0.3400	0.4288	0.3663
RF(TF-IDF)	0.4922	0.3364	0.4260	0.3639
SVM(BoW)	0.5090	0.4600	0.4590	0.4500
SVM(TF-IDF)	0.5067	0.4558	0.4586	0.4453
biLSTM(GloVe)	0.6600	0.7086	0.6050	0.6475
GRU(GloVe)	0.6600	0.7000	0.6000	0.6400
BERT	0.6900	0.7200	0.6300	0.6700

Table 4.1: Results of the experiments with five models (RF, SVM, GRU, biLSTM and BERT)



(a) BERT model (b) LSTM model

Figure 4.4: Confusion matrices of LSTM and BERT models

Personlised experiments

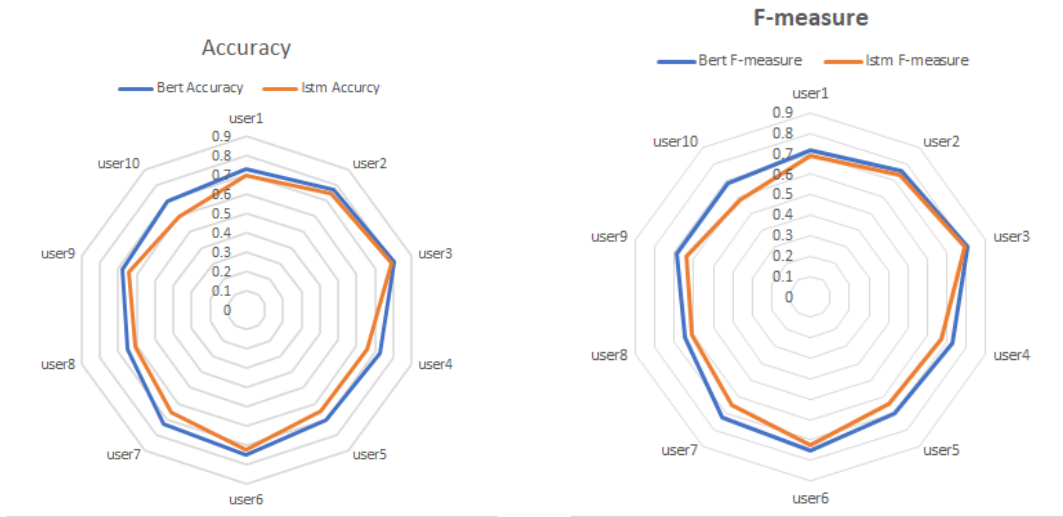


Figure 4.5: Radar charts showing the accuracy and F1-score of BERT and LSTM models based on ten users data

To examine the performance of BERT and LSTM on the personals dataset, we select posts written by ten most posting users. We numbering users was by the number of posts as User 1 is the most and User 10 is the least between them. Radar chart in 4.5 shows the difference in accuracy and f measures levels of 10 users experiments. There is a high variation between users in two of the models, and biLSTM resulted in the highest divergence between users with accuracy levels ranging between 0.60 to 0.79 and between 0.59 to 0.79 in the F1 score.

4.5 Discussion

The objectives of this chapter were to (i) evaluate deep learning as a computational model for emotion recognition from our dataset following state-of-the-art methodologies and (ii) to assess the overall power of BERT by fine-tuning on the dataset to emotion analysis. This is the only study looking into emotional classification on social media data using multiple NLP methods and comparing models with BERT. The results have demonstrated that the BERT model outperformed other deep learning models. Moreover, in general, results have suggested that deep learning methodologies are appropriate for modeling emotion from social media text, and more importantly, Language models can be even the most appropriate.

Although BERT is very large, complicated, and has millions of parameters, the state-of-the-art performance has been achieved by adding a simple one-hidden-layer neural network classifier on top of BERT and fine-tuning BERT. BERT obtains these results because BERT was trained on a considerable amount and already encoded a lot of information about the language contexts.

Regarding personalized modeling for users, it was evident that personalizing the model by training it in the user text is better than generalizing it. In addition, the more data we have to train BERT, the more performance to give us, and that is clear when the last three users were worse performance.

Finally, in the experiments of comparing the two groupings types of labels, the significant difference in performance for the BERT model between these two grouping classes is perceptible. It might indicate that the six classes grouping mitigates confusion among inner group lower-level categories.

4.6 Conclusion

This chapter has showcased how pre-trained contextualized language models can be usefully applied in text-based emotion modeling. In addition, the comparison between pre-trained LMs and some other NLPs ML models has been introduced. Furthermore, experimental results across these models show how BERT has outperformed the traditional NLP approach, adding empirical evidence of its superiority in emotion classification. Although it has been mainly focused on text embedding for classification, in the next chapter, I would like to take advantage of the capabilities of BERT by fusion other features with BERT features for emotion modeling.

Chapter 5

Feature Fusion with BERT for Emotion analysis of unstructured social media text

5.1 Introduction

Emotion can be presented in and predicted from the text. The people were writing text in different styles and genres, such as formal styles like news text and informal styles like social media text. Therefore, emotions suppose to be presented differently depending on the text type. Social media texts are typically rich in emotion because they are often written as immediate, unfiltered reactions to breaking events or expressions of personal feelings. However, social media text is indeed noisy, less grammatical than edited text and with more misspellings [22]. Therefore, using only word vectors to classify the text might be less effective than the formal text as some of these words might be misspelled or used grammatically wrong. Also, the social media text is usually short; thus, most words will be represented in different classes simultaneously. These reasons motivate researchers to fusion other features (semantic) with word embeddings to classify text of social networking [115].

Integrating multiple sources of data has gathered significant interest and has been successful in various applications of NLP tasks, especially text classification

[200] [67] [54]. The main goal for integrating sentiment data and other sources of data to form a unified picture or make a better decision in text classification with deep learning models, and it is achieved this goal with many pieces of literature such as LSTM [184] and convolution neural network [106].

However, even though BERT is one of the most potent NLP models available and it achieved ground-breaking performance on text classification tasks, there is a gap in exploring BERT integration with other sources of features. Therefore, the focus of this chapter is to examine the potential for BERT in emotion modeling from social media (casual language) text under a general framework encompassing both sentiment and other features paradigms.

To better handle social media text, this chapter approaches emotion analysis from a novel perspective by taking advantage of pre-trained language models that have been produced the state of the art results in emotion classifying. Language models make semantic vectors that capture the semantic meaning of the text. By incorporating linguistics and grammatical features associated with user posts, including punctuation or parts of the speech, Mental health language-based such as Anxiety, Sadness, and Anger. Likewise, some features regarding the user post, such as Word counts, the number of hashtags, and the count of emoticons and emojis in the post. These features with other features will be discussed in section 5.3 and fused with the semantic vectors produced from the pre-trained BERT model. The proposed model incorporates these fused features to classify emotion from social network posts. The proposed system combines a fully connected neural network architecture and BERT model, and it is evaluated by our large dataset used in this thesis to test how the classification will be improved with social media text. The performance results of the proposed model show substantial improvements over the baseline models that have been used to compare in the same dataset.

5.2 Related works

Integrating semantic, and sentiment features has been successful in modelling emotion and sentiment analysis form short text. The early exploring of integrating these tow source of features has been done by Kiritchenko et al [115], and the system is based on a supervised statistical text classification approach leveraging semantic and sentiment features. Moreover, sentiment features were derived from sentiment lexicon elongated, and semantic features were emoticons,punctuations and hashtags counts, occurrences of each part-of-speech tag and the number of negated contexts. This system obtaind state-of-the-art performance on the SemEval-2013 dataset. Later, with deep learning models and advances in word embedding, word embeddings were combined with other features and integrated into a deep learning models for predicting sentiment classification in social media such as convolution neural network [106].

LIWC [173] is a tool for text analysis that counts words in psychologically meaningful categories. LIWC reads raw text files in batches and counts the percentage of words that belong to each category, which can be grouped as linguistic, punctuation, psychological, and summary features. It utilises over 60 categories, and dictionaries of words related to each category. It is has been used widely in literature to represent text features for several NLP tasks and empirical results. Such wide use of LIWC demonstrate its ability, for example, predicting depressed and non depressed social media users from text classification [200], profiling social media users by inferring age, gender and personality traits of users [67] and predicting the emotion of the user [54]. It has been proved the ability of LIWC to detect meaning and to show attentional focus, emotionality, social relationships, thinking styles, and individual differences [219]. Moreover, it has been proved that LIWC can be reliably used to identify emotional state in the user posts [49].

Combining BERT with other deep learning models in the infant stage recently and even if it is very few studies exploring it, it is achieved the state of the art results in emotion classification from text. Recent work by Adoma et al [5] presented a model for predicting emotion with tow stages, the first stage has (BERT) model, the second stage consisting of a Bi-LSTM classifier. The results become the new state-of-the-art in detecting emotions on the ISEAR dataset.

The only work that has been found in literature that has been combined BERT Embeddings and a set of other features is the work that integrated psycholinguistic features from AffectiveTweets Weka-package with BERT [8]. In this work, they proposed a system (EmoDet2) that combining BERT and BiLSTM neural network and they obtained performance results that show substantial improvements (F1-Score 0.748) over the baseline model provided by Semeval-2019 (F1-score 0.58) [8]. However, the system was for English textual dialogue and the input to the system are utterances along with three turns of context which is completely different from the goal of this chapter which is modelling informal social media text.

5.3 Methods

This section will introduce the proposed system by presenting the proposed features and model.

5.3.1 Proposed Features

Several features have been included for analyzing the cognitive, affective, and grammatical processes in the text, which helps examine the difference between the writing style of posts among emotion states and how they will contribute to predicting the current emotion. In addition, most of the extracted features from LIWC as all the LIWC features have been used. Thus, in addition to BERT features, every post will be represented into a vector of features as follows:

-**BERT embeddings** will be of dimension 768. These embeddings have been treated as features of the post itself.

-**Post summary features**: Word counts, the number of hashtags, URLs and other users mentions, Words, sentences, Words > 6 letters, How many emoticons and emojis. Also, the frequency of words that reflect the writer's thoughts, perspective, and honesty. There are a total of 12 features under this category.

- **Sentiment Analysis of the post**: the polarity value and the subjectivity were computed by TextBlob package[1].

-**Punctuation-based**: Features derived from the punctuation used in the forum, such as exclamation points and question marks. There are 11 features under this category, all of them extract from LIWC.

-**Mental health language-based**: Features derived from the occurrence of mental health-related words in posts include Positive emotion, Negative emotion,

[1] *https : //textblob.readthedocs.io/en/dev/*

Anxiety, Sadness, and Anger.

-**Linguistics and grammatical features**: refer to features that represent the functionality of text, such as negations as well as parts of the speech (adjective, noun, verb, conjunction) frequencies. Also, Grammar words like adjectives and Comparisons. There are 20 features in this category.

-The use of **informal language** such as swear words.

-**Personal concerns features** such as job, work, family, and money.

- Other Linguistics features can be used to scrutinize the emotional part of the posts. There are a total of 28 features under this category, including the following features:

-**Time orientations words** like Past focus(did), or Future focus(will).

-**Perceptual words** such as see and hear.

-**Social features** such as friends and family.

-**Biological words** used like body, health, and sex.

5.3.2 Features extraction

Following methods have been used to extract the proposed features.

- Direct extraction based on the specific symbols: This method can be applied to all, Emoji (E-mark),the number of hashtags, URLs and other users mentions.

- Extracting based on processing tools: LIWC has been used to extract the rest of features. TextBlob[1] is a Python library for processing textual data was used to compute the Sentiment Analysis value.

The Emojis were extracted from the notifications retrieved and processed with a python script. According to [218], today's lexicon-based approaches typically do not consider Emoji. Conversely, one of the first steps in most current work is to remove the Unicode symbols, including Emojs, ignoring that Emoji may express emotions when regular text fails to do so and be classified as anger, disgust, fear, happiness, sad, and surprise.

[1] *https : //textblob.readthedocs.io/en/dev/*

To scale features, the Min-Max normalization method is utilized to scale every feature vector into [0,1] by obtaining the values 0 and 1 at the minimum and maximum points, respectively, for the features that had not been scaled by LIWC such as Word counts, the number of hashtags and emoticons and emojis. Lastly, the number of proposed features was 95.

5.3.3 The proposed model

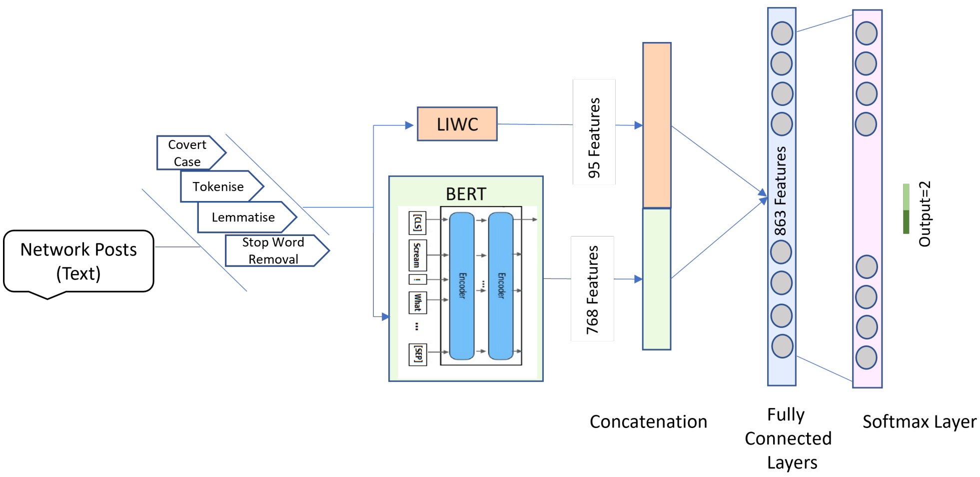


Figure 5.1: The proposed model architecture of fused features to classify emotions from social network posts

Figure 5.1 shows the architecture of the proposed model. Each user post is fed into the LIWC tool to extract the features and combined with other features which results in 95 features as total. Layers of the model as alligned as follows:

- The first layer: is the input layer to set the inputs to the model, which is two sets of information: text input(the post itself) will feed the BERT model and the second set is the proposed features introduced in the previous section.
- The BERT layer, which will produce embedding for every post with 768 dimensions as BERT base, has been used in this layer. These embeddings can be treated as features of the post.

- Concatenation layer: The main task of the connection layer is to combine the results from the BERT layer (semantic features) and the proposed features from LIWC (syntax features) to construct the full represented feature of the post. The concatenation of the embeddings is defined in Equation (5.1).
- Fully Connected layer (FC) with ReLU activation and dropout after the BERT layer.

$$a = LIWC(X_Syn), b = BERT(X_{embeddings}), x_{fuse} = [a, b] \tag{5.1}$$

Finally, the Fully Connected layer is fed into the Softmax layer to classify the deep-extracted features. The Softmax function transforms the numerical results of the neural network to the output of probability and decides the predicted class with the max probability.

In the model training process, the classification task employs the Cross-entropy loss function (is defined in Equation (5.2) to calculate the error between the prediction and the true value to achieve better results:

$$L = -\sum_{i=1}^{2} t_i log(p_i) \tag{5.2}$$

5.4 Experiments

5.4.1 Data pre-processing

The data set used in this thesis was introduced in chapter 3. The Similar steps that have been represented in chapter 4 to prepossessing the data were used in this chapter, but before data cleaning, the number of hashtags and URLs have been counted as they are variables that we need. Then it is followed by feeding data to the LIWC tool to extract the features. After that, data cleaning steps have been utilized to the text and spell correction. After the data cleaning, we have around 6 million labeled posts with three classes(%47 Negative, % 12 Natural, %41 Positive).

5.4.2 Baseline models

The baseline models utilized in this chapter utilised the raw text features to represent the text information and the proposed features in section 5.3 to the baseline experiment.

Two baseline Machine learning models have been explored: Random Forest(RF) and Support Vectors Machine (SVM). These models experiment with Bag-of-Words (Unigram) as a representation of the text as this method outperformed Term Frequency times inverse document frequency (tf-idf) in this dataset (chapter 4). The proposed features have been added to these models by pipeline in the RAPIDS cuML Python library.

Another two Deep learning models: LSTM and GRU, have been utilized with GloVe embedding as text representation. Finally, the proposed features are concatenated with GloVe embedding features to classify emotion.

Lastly, the pre-trained BERT model has been compared with the proposed model.

5.4.3 Experimental Setup

Experiment data is divided into a training set, validation set, and a test set at a ratio of 70:15:15. The training set has a total of around 3.7 million of labeled posts, and the test set and validation set includes 900K of data for every set. All the experimental is based on the Python language using GPU as the data size is huge. The experiments utilize the following metrics(the precision, recall, and F1 score) to evaluate the model's performance.

In search of the optimal model, we explored the following hyperparameters: batch size has been set as 32 and learning rate to 3e-5 and Number of epochs is four after preliminary experiments with various values. A Regularization lambda of λ=0.01 along with drop rate of 0.1, and batch normalization were applied on the fully connected layers in the network to reduce overfitting. The Max sequence length of the set to 40 represents 75% of dataset length. Warm-up steps set to 10,000 and Training steps to 100,000.

In experiments with training the model in a limited set of data of users posts, and similarly to avoid overfitting the model, Cross-validation with 10 folds has been used. Cross-validation allows for using every observation in both training and test sets as the dataset is small for the complicated model.

5.4.4 Results and Discussion

Models	Accuracy	Precision	Recall	F1
RF(BoW)	0.49	0.34	0.43	0.36
RF (BoW+ Features)	0.48	0.34	0.43	0.37
SVM(BoW)	0.50	0.46	0.46	0.45
SVM(BoW + Features)	0.52	0.46	0.49	0.47
biLSTM(GloVe)	0.66	0.71	0.60	0.65
biLSTM(GloVe+Features)	0.66	0.70	0.62	0.65
GRU(GloVe)	0.66	0.70	0.60	0.64
GRU(GloVe+Features)	0.66	0.70	0.61	0.65
BERT	0.69	0.72	0.63	0.67
The proposed model	0.69	0.72	0.64	0.68

Table 5.1: Results of the experiments with all Models compared to the proposed model

The results are compared with the current mainstream models to verify the feasibility and effectiveness of the model. The comparative experimental models include RF, SVM, LSTM, GRU, and BERT. All of these models have been experimented with and without the proposed features. The experimental evaluation results are shown in Table 5.1.

In general, integrating the proposed features with text represented features improves the performance of all models as F1 value has been improved by (0.90% to 2%) among all models. However, it is slightly improved. It is good evidence that modeling emotion from social media text can be improved by integrating more than one source of features. Additionally, the comparison results of the ten models show that the proposed text classification model achieves a better classification effect than all other models in 3 evaluation metrics (Accuracy, Recall, and F1), whereas, in the Precision value, the BERT model outperform the proposed model.

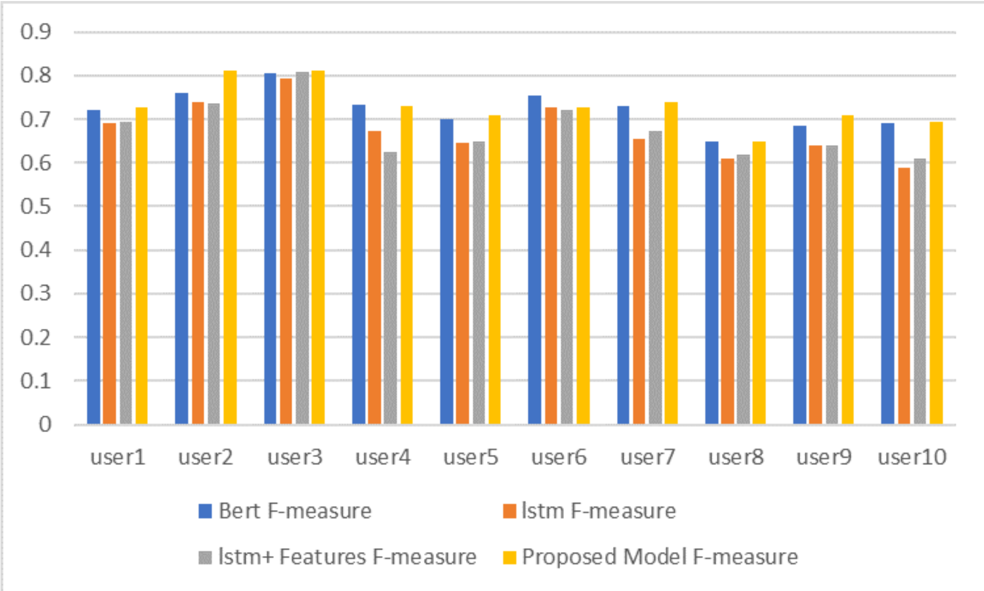


Figure 5.2: Comparison of F measures levels of models for 10 users

	0	1	2
Precision	0.66	0.53	0.66
Recall	0.73	0.04	0.79
F1-score	0.69	0.09	0.72

Table 5.2: The classification report for the three classes in BERT model

	0	1	2
Precision	0.67	0.41	0.70
Recall	0.78	0.02	0.79
F1-score	0.72	0.05	0.74

Table 5.3: The classification report for the three classes in the proposed model

Furthermore, Even though adding features to deep learning models has improved the classification performance, the pre-trained BERT model still has achieved better accuracy of nearly %69 and a higher F1 score (0.67) by a reasonable difference of around 2% from other scores. This indicates the power of the BERT model in understanding the context and capturing complementary facets of word meaning better than other deep learning models. Our results outperform other similar models in the literature. For example, [143] has achieved F1 scores between 0.30 to 0.67 for 6 different classes using the basic BERT uncased model after fine-tuning.

Figure 5.2 presents charts of 10 users models for LSTM, LSTM+ Features BERT model, and the proposed model as they the best performance in the results. There are variations in the Comparison of the performance between the proposed model and whether it improved the classification's performance. The results indicate the proposed model improved the performance in 6 users models, did not improve remained in the same results for 3 users whereas decrease the values of scores for one user (user 6). This variation in the performance in the proposed model among users might be the result of the cased text for different personalities. Some of them give extra features to the model to understand the context better. Additionally, The under-performance on the user dataset might be related to over-fitting.

Tables 5.2 and 5.3 show the classification reports for the three classes in the BERT model and the proposed method respectivly. Classes in the table as 0 is Negative, 1 is Neutral, and 2 is Positive. The proposed BERT fusion method performs better than the BERT-only approach in F1-score in three classes. The two models did not predict Neutral emotion and other emotion classes, and this is expected as this class is the lowest number of posts in the dataset. However, Integrating features with the BERT model improved the classification of this class and the classification performance for all classes.

5.5 Conclusion

In this chapter, the potential of fusing BERT Language Model features and syntax features, the popular linguistic features in social psychology features, have been investigated. Moreover, integrating different sets of features with BERT for emotion analysis based on users' posts in social media. The system's performance (F1-Score 0.6795) surpasses the performance of the baseline models indicating that this approach for modeling informal text is promising, adding empirical evidence of its superiority in average social networks classification problems. In the proposed model, word embedding from BERT with feature vectors extracted using the LIWC and other features from user posts have been integrated to obtain the predictions for user emotion.

Chapter 6

Modelling the Temporal Dynamics of User Mood Change

6.1 Introduction

Considering the importance of social media, recent studies looking into the impact of online communication and social media on mental well-being have found potential benefits to understanding users' mental health based on social media. For example, using Twitter data to propose analysis population depression [56] [55] and using Instagram data to predict people's suicidal ideation [32]. Meanwhile, social interactions and text analysis have been examined widely on social media to infer some understanding of mood behavior expressed in these networks (chapter 2). However, most works on emotion modeling treat individual emotions independently without considering user profiles.

With the growing peer support networks, people with mental health issues turn to these online forums to seek advice and who will listen and emotional support. Furthermore, studies proved that using these networks helps to track the changes in user's health [203] [45]. Thus, the dataset adopted in this thesis can be used as a benchmark for monitoring mental health change.

By analyzing a real-world dataset from an extensive Mental Health social network, this chapter will bridge the gap in modeling the change in mood of social networks. It will model the changes of users' mood states based on how they interact and respond to posts over time and predict future change states. We propose a novel model to find that future mood change states can be predicted and in which direction. The Mood Change model predictive model has achieved a good accuracy level using three different Deep Learning unite (2 BERT units and LSTM units).

In the current study using the mental health social network dataset, the following contributions have been made in this chapter :

- Prove to what extent the mood change can be modeled and predicted (i.e., do people feel more positive or negative over time?).
- Propose a novel predictive model and framework to model the emotion through time.
- Analyse and explore what factors/features are driving this change.

The rest of the chapter is organized as follows; section 2 discusses related works; section 3 present the dataset. The proposed framework to predict the change in the mood and the model architecture is presented in section 4. The Experiments, the results obtained, and their corresponding analyses are presented in section 5. Finally, in section 6, Peer influence has been discussed, and the chapter conclusion in section 7.

6.2 Related works

It seem to be only two works in literature have been done in the change in cognitive state in Talk life user. One of them is the work of Pruksachatkun et al. [182] which is analyzing support comments on a thread. They built a model for predicting whether the thread leads to a positive cognitive change for the individual. Furthermore, they used a dictionary of phrases where people self-reported their change to check if there was a positive cognitive change. However, the focus of this study is the change of the user's cognitive state in each topic, whereas the work in this chapter aims to model user mood change in general.

The other one is done by [122] to explore temporal patterns of the user posting in Online Mental Health Forums. They found that user activity follows a distinct pattern of high activity(bursts) periods with interleaving periods (breaks) of no activity, and they proposed a method for identifying these periods based on activity. They provide statistical analysis to examine characteristics of user activity that lead some users to find support and help while others fall short. Furthermore, they use metrics of the previous study [182] which is (moments of cognitive change) to robustly test some features such as Engagement with Others and Linguistic Features and if it might contribute to the change of the cognitive state. Similar features in this work will be used with other features to predict a moment of change. In addition, their analysis was focused on a short period, by considering the first original posts within a burst to identify a general mood for a burst, and then see the change of the mood in the next burst, a user's mood across a period may vary. While this is ignored in their analysis, the presented work will consider every reported mood. [54]

6.3 Dataset

A part of the dataset will be used in this chapter. The goal is to monitor users' mood progressions through time and how it is changed, the focus will be extracting users' data instead of extracting posts. Users of the dataset were sorted by the number of posts, Then the first 100 users with the most posting activities have been selected to this chapter to model their mood change. For a given user, their timeline activities will include:

- All original posts.
- Reply threads for each of their original posts.
- Replies written by the user on other's posts.
- Personal information such as age and gender.

In addition, every user generates a time series of their activities to employ measures/features extraction and labeling.

6.4 Model Framework

Mainly, the user posts a post with a mood, such as a negative mood. Then, the user may engage with others in the same thread or in other threads, after which the user will post a new post with a new mood. Can we predict whether the individual new post will include a change of mood from Negative to Positive or from Positive to Negative, or the new mood will not change?.

To this end will pursue the use of supervised learning to build a classifier trained to predict the user's emotional changes depending on their reported mood.

6.4.1 Data Labeling

Given observed data for the user during the period from the last mood reported until the current post, the prediction framing will be a classification problem per measure of change. The progression of the mood among the categories of mood will be in three directions (see figure 6.1):

- Mood changes from negative to positive.
- Mood changes from positive to negative.
- Mood remains the same.

Consequently, we model this change as a three classes classification problem: Positive change, Negative change, and No change.

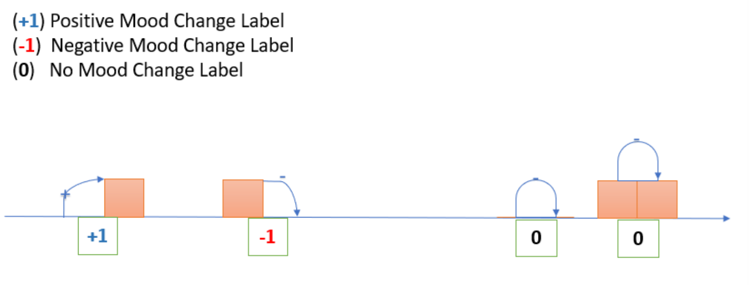


Figure 6.1: Illustrating of mood change labeling: change from negative to positive, change from positive to negative or remaining in the same mood

Moreover, The self-reported mood is adopted in this chapter to reflect how the user feels. Therefore, instead of using manually annotated labeling to the mood change like (moments of cognitive change) [182], we use the self-reported labels to categorize an individual's change in mood. Change in Mood is defined as the difference in the Mood of the last post from the user to the current post. Quantitatively defining the change in the Mood as follows:

If the Positive class is assigned to = 1, and the Negative class is assigned to = 0, the following equation will give the label of change:

$$\begin{cases} if L_t - L_{t-1} > 0, X = 1 \\ if L_t - L_{t-1} < 0, X = -1 \\ Otherwise, X = 0 \end{cases} \tag{6.1}$$

Where L_t is the current mood, L_{t-1} is the previous mood, and X is the label of the change. The result of this equation will be one of the three values -1, 0, 1. Meaning, the scenarios of changes will be the following: If the current mood is Positive (1) and the last mood is negative (0), then the change is equal to 1, which is a positive change. On the other hand, if the current mood is one and the previous mood was one, then the result is 0 representing no change—a similar calculation for Negative mood. As a result, the three classes to classify the change of mood: -1 will represent the change in a negative direction, change of mood in a positive direction yields = 1, and No change in the mood =0. Figure 6.2 shows an example of a random user timeline with labels indicating positive, negative, and neutral mood changes during the time.

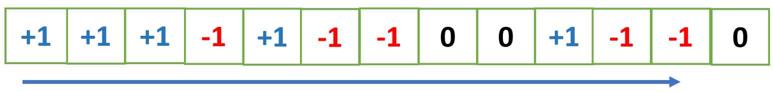


Figure 6.2: Example of user timeline with labels indicating positive, negative and neutral mood changes.

6.4.2 Proposed Features

Training a classifier from labeled data will represent each post into a vector of features to capture features describing the emotion change between the current and last posts. Feature selection plays an important part in the effectiveness of the classification process. For this study, we explore the usage of different features:

User Post text representation: To represent the text used in the user post, the pre-trained language model BERT will be utilized as the BERT proved its ability to understand the language and classify the emotion in this dataset (chapter 4 and 5). Thus, the post will be represented in the vector of 512 dimensions. We use BERT to classify the change in mood as the novelty of this chapter as the previous works use LIWC [54] [122] or bigram and trigrams [182].

Post's Replies text representation: To utilize the supports given to the user, the replies that deceived the previous mood until the current reported mood will be aggregated and fed to BERT to understand the kind of support if there is support from other users. Furthermore, only the replies from other users will be included in this set of features as sometimes the poster user reply in the same thread of his post. Whereas, In the study of [182], they include all the threads in a trigrams representation.

Linguistic Features: Following [122] [182] [54], as they used LIWC Features as indicate the change in user cognitive or depression. For example, LIWC features are used as indicators of questions and exclamations that often occur during cognitive restructuring [182]. In addition, in the analysis of [122], users who experience a Moment of change are more likely to use positive affective language and receive replies from others containing positive affective words. In this category of features, two sets of Features will be included: User's Linguistic Feature's set to capture the change in the user language and Replies Linguistic Feature set to capture the effect of replies language.

Time Since Last Post (TSLP): The time between the last mood and the current mood is important to the predictive system to analyze the effects of social networks in this change. For example, if the user posts with a negative mood and after 5 days posts with a negative mood, this second mood is not necessary because of the interaction with the network. Consequently, it will confuse the prediction process. Thus, the time between two moods will be a trainable feature in the model. Furthermore, the time gap between a user's posts can represent her diminishing social interactions [192]. In the study of [122], they used Activity

Burst where such that the time between any two consecutive posts remains within this period to be included in their analysis.

Mental health and emotion language: These features will be derived from LIWC to observe the emotions or support in replies which might help the poster user to change their mood. It will include Positive emotion, Negative emotion, Anxiety, Sadness, Anger, and Sentiment Analysis.

Post Engagement: As it is stated that the attraction of the post in a social community and getting attention from the community, more likely to make this post persuasive [238]. Similarly, in mood change, the more posts attract people to engage in it, the more likely these people will affect the mood of the poster. Post Engagement here means how the post has engaged with users in the network, such as the number of views, likes, replies, gifts, supports, etc. Following the formula given by to [238] to quantify the engagement of posts on Twitter, the formula 6.2 will be used with modifications to calculate the Post Engagement.

$$E = \frac{\sum (2 \times Nofreplies) + hugsrank + likesrank + flagrank + giftsrank}{ViewRanks} \times 100 \tag{6.2}$$

Engagement with others: Sometimes, the user replies to people who engage in their post. This might lead to a change in their mood. Moreover, the way that the user engages, in addition to the frequency, can be indicative of her emotional health [192]. Thus, following the formula that has been used in [122], the engagement of the user in the previous mood (E) will be calculated as following:

$$E = \frac{numberofrepliestoothersinthepost}{totalnumberofreplies} \tag{6.3}$$

6.4.3 Mood Change Predictive Proposed model

To characterize the behavior of the mood and in any direction of change will be, the following model is proposed as it is shown in Figure 6.3. After generating the user activities and the activities that happened between the last and current post features, it will be fed to the model, which consists of layers as follows:

- Input layer: consists of three types of inputs: i)the user's current post, ii)the replies received from the last post until this post, and iii)the other features extracted from the thread.

- BERTs encoding layers: 2 layers from pre-trained BERT model, one for the post input, the other one for replies text inputs. These layers will produce 2 vectors with 512 dimensions.

- LSTM layer: the input to this layer is the proposed features in this chapter to be trained.

- Concatenation layer: the three outputs from BERTs layers and LSTM layer will be concatenated in this layer to construct the full represented feature of the post.

- Fully Connected layer (FC) with ReLU activation and dropout after the BERT layer.

- Softmax layer as the final layer to classify the concatenated vectors. The Softmax function transforms the neural network's numerical results to the output of probability and decides the predicted class with the max probability.

6.5 Experiments

6.5.1 Feature extraction

We follow a multi-level feature extraction approach as follows:

- Hand-Crafted features: such as user engagement and the time differences.

- Linguistic and Sentimental, and word counts Features using LIWC and TextBlob tools.

- Text embeddings from BERT model.

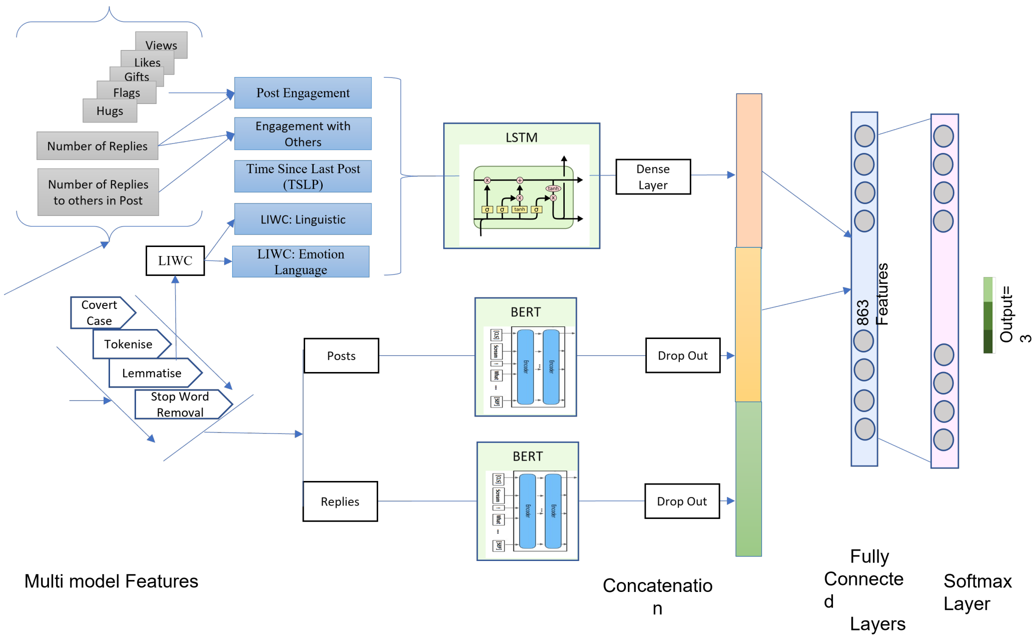


Figure 6.3: Mood Change Model based on multi model features and two BERT units and LSTM unit

To this end, we used the Min-Max normalization method, which scales every feature vector into [0,1] by obtaining the values 0 and 1 at the minimum and maximum points, respectively. If the features are categorical, it will be ideal for converting to one-hot vectors and concatenating them.

6.5.2 Experimental Setup

The dataset is divided into three parts as 60:20:20. The total size of the data is 120k records. There is a noticeable imbalance in the data as the class. No change(0) represents around half of the data (54% of the data). The rest of the labels (positive and negative change) are nearly equal. Thus, the class (0) data has been down-sampled t nearly 10%.

In the experiments, small pre-trained BERT from TensorFlow Hub[1] has been chosen to reduce the dimensions that represent the posts to avoid the over-fitting problem in the training process. To evaluate the proposed model, in addition

[1] $https://www.tensorflow.org/hub$

to the metrics that have been used in the previous chapters, Area Under Curve (AUC) has been utilized [182]. Furthermore, we compare the model's performance using different features combinations to cover various scenarios. The model performance will be evaluated on five sets of features added one after the other to evaluate the impact of these features (Posts, Replies, TSLP, Engagement, Emotion language, Linguistic).

6.5.3 Classification Results and Discussion

Features	Accuracy	Precision	Recall	F1	AUC
Posts	0.48	0.42	0.32	0.36	0.60
Posts+Replies	0.56	0.56	0.50	0.53	0.73
Posts+Replies+TSLP	0.55	0.56	0.50	0.54	0.74
Posts+Replies+TSLP +Engage	0.55	0.56	0.49	0.53	0.74
Posts+Replies+TSLP +Engage+Emotion	0.56	0.59	0.51	0.56	0.75
Posts+Replies+TSLP +Engage+Emotion+ Linguistic	0.58	0.60	0.51	0.57	0.76

Table 6.1: Comparison of model performance based on different features sets using the following metrics (Accuracy, Precision, Recall, F-Score and AUC).

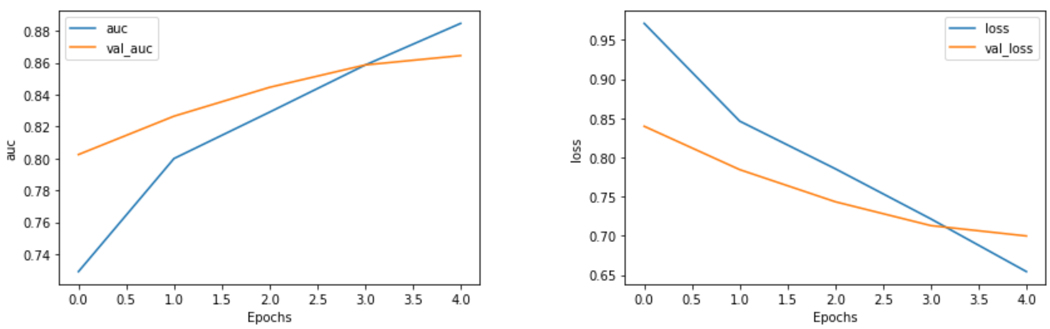


Figure 6.4: The model training performance for user4

Evaluation results across all experiments are illustrated in Table 6.1, based on five standard performance evaluation metrics: Precision, Recall, F-Score, Accuracy, and AUC. The accuracy levels of the results are also compared between the model with just the user post text as the only input and the combined modalities across all the sets of the features by adding one set by one. Mood Change model did not perform well on user posts only with an average accuracy=%48 and AUC=0.60. Moreover, this score was the worst score among the rest of the features in all the evaluation measures, and it can be considered an indication that the user posts are only insufficient to predict changes in the user mood. By adding the replies text features, the performance improved significantly and achieved F-score=0.53 and AUC=0.73. This improvement presents the importance of **influence of peers** in predicting changes in users' mood while interacting on social media. The model has slightly improved by adding the time between last mood and current (TSLP) and achieved AUC=0.74.

Moreover surprisingly, adding Post and User Engagement features set decreases the model's performance by decreasing F1-score by 0.0090 and AUC by 0.0030. It was expected from these features to improve the performance of prediction, but the results show that these features are non-Robust Features in this dataset. However, Adding Emotion and other Linguistic Features have well influenced the model by increasing AUC to be 0.75 and 0.76, respectively.

	0	1	2
Precision	0.62	0.50	0.50
Recall	0.76	0.3798	0.35
F1-score	0.68	0.43	0.41

Table 6.2: Classification model performance for the Mood Change model (for each class)

In general, The best performance for the model was with including all the features as the model achieved nearly F-score=0.60. Furthermore, Replies BERT features' encouraging performance indicates that the respondents' language is the most associated with mood change. Table 6.2, illustrate the confusion matrices yielded by the model based on the dataset and represent classes as follows: 0 is No-change, 1 is Negative change, and 2 is the positive change. The best performance of the model was with class No-change (0), which is as expected as most of the data labeled as No-change.

Users	Accuracy	Precision	Recall	F1	AUC
User1	0.54	0.56	0.41	0.48	0.74
User2	0.60	0.64	0.63	0.64	0.77
User3	0.54	0.57	0.39	0.46	0.73
User4	0.70	0.71	0.65	0.68	0.85
User5	0.63	0.67	0.54	0.61	0.82

Table 6.3: Results of the model performance on 5 selected users.

Table 6.3 shows the difference in measures levels of 5 users experiments which the most posting users select with accuracy levels ranging between 0.54 to 0.70 and AUC levels ranging between 0.73 to 0.85. It is clear from the table that the proposed model resulted in the highest variation between users, which might be because of the personality aspects in language and how they express their mood. In addition, it might be because of the difference in personality in being affected by others. Based on best performance to the model and it was in User4 data, Figure 6.4 shows the model training performance (lose and AUC values) and performed well based on training data for this user.

Comparing the proposed framework to the classification of the mood change with the framework of Pruksachatkun et al.. The mood change model did well. They measured the classification in AUC, and they did not show other measures. Their classification achieved high AUC scores (> 0.8) for predicting a moment of change. However, their dataset is just 6,396 threads, and it is divided 50:50 between change or not. To check how the mood change model do in 2 classes labeling (change or No change), table 6.4 present the results with a high AUC score (> 0.8), and it shows good performance of the model in predicting the change of the mood with 0.65 in F1 measure.

	Accuracy	Precision	Recall	F1	AUC
2 classes classification	0.64	0.66	0.63	0.65	0.83

Table 6.4: Results for Mood Change model with 2 classes (Change and no change)

6.6 Analysis

The novel proposed framework to model the change in health social mood performs well Talklife social dataset and predicts the change with reasonable measures. Moreover, it is proved that the change of user emotion through time can be predictable from user activities and the interaction of peers in the network and the direction of the change (Positive or Negative). Although neither post and user engagement features were robust to the mood-change, there might be a justification from the Talklife dataset. In the previous analysis on [122], it has been shown there is difficulty in observing differences in engagement between positive and negative emotion on the platform, and they almost receive the same engagement.

	WC	posemo	TSLP	negemo	informal
Change and No-change	0.12	$0.28e^{-48}$	$0.23e^{-39}$	$0.26e^{-30}$	0.04
Positive and Negative	0.03	$0.13e^{-22}$	0.48	$0.21e^{-20}$	0.05

Table 6.5: P-value calculated from the distribution of groups in selected features

Going beyond features to the analysis of the influence of peers was gave some indicators of peer's impact on mood change. Table 6.5 represent P-value between changed and not changed mood groups and positive and negative changed groups. P-values were calculated from the distribution of groups using python.

Regarding how long replies the user receives, Users who experience a change in the mood are more likely to have increased in the numbers of the words they received (mean 39.861472 vs. 18.559634 , p = 0.12). This is an important indicator of the influence of peers on the user mood change whatever this change to negative or positive mood (p= 0.03).

In the case of the positive or negative emotion received from peers, it make a difference with positive and negative change more than being effective in mood change. Moreover, the similar behavior for the time since the last mood. Lastly, the amount of informal language from peers, such as swear words, seems to affect people who changed their mood by mean 9.39069 vs. 4.563515 who do not change and with $p = 0.04$.

6.7 Conclusion

To the best of our knowledge, this is the first work to model the change in the Talklife dataset by providing a novel framework for analysis of user mood change in online mental health communities. It has been found that predicting changes in user emotion can be done with reasonable accuracy from modeling user behavior and peer supports in the network. Moreover, the direction of change to positive or negative emotion can be predicted as well with the proposed model. Furthermore, the influence of peers on the user mood has been discussed, providing some insight into what impacts users' mood.

Chapter 7

Social Networking User Modelling Based on their text

7.1 Introduction

The results from the experiments in Chapters 4,5 and 6 show that the personalised of modelling of emotion did well with social media text. This results give the motivation to explore modelling (profiling) users of mental health social network. User profiling have been done through behavioural-based approaches such as ego networks [139], social ties [87] personal reviews [123] and user Features [87]. The most popular methods used for user modelling in literatures are clustering [123] [139] [130] [63] and Graph Embedding [87] [262] [200]. However, in our dataset, the primary behavior of users is posting or replies, as discussed in chapter 2, when the majority of users have very small or, no ties exist between the neighbors. In addition, In chapter 6, when it is proved that the text features were the most indicators for user behavior. Therefore, this chapter will explore the modeling of the users based on their posting text. Moreover, finding similarities in the behavior of users and clustering them into groups, these similarities could be used later in the downstream task such as personalized emotion modeling. Moreover, having patterns of user posting activities could be highly indicative of user personality and can be directly leveraged to generate type labels for users.

For this purpose, this chapter will propose an approach to user inference over social media that entails learning unsupervised user embeddings from user post histories to be used as features in downstream clustering models such as K-Means and other clustering algorithms find groups of similar users. The contribution of this chapter lay on following:

- Exploring to what extant users of mental health social networks can be modeled.
- Proposing a model (User BERT model) using state-of-the-art embeddings to find a pattern of Mental Health social network users (clustering) by clustering models.

7.2 Clustering Framework (User BERT model)

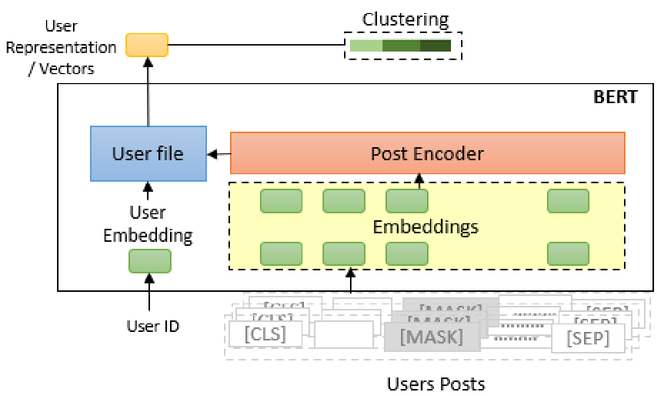


Figure 7.1: User Clustering model using BERT

Clustering is applied to group user data objects depending on the data's dimensions that specify the objects and their association. The group behavior will determine the clustering of users into separate groups using a clustering model. Cluster analysis can also be helpful when there is no clear group structure in the data, such as the used dataset.

To this end, representing users by features is required to cluster social network users. In our dataset, the most important user information in their text will be used to represent the users in the framework. The proposed model includes two main phases shown in figure 7.1.

1) User embedding stage: The most successful representation of the text in emotion modeling in the previous chapters was text embedding by pre-trained language models, as they are supposed to contain more accurate representations of words and sentences. Therefore, this method was the trigger to represent users by embeddings. To represent users by embedding vectors, all user posts in the network are aggregated to one document. Then the user text is fed to BERT pre-trained model to produce one vector for the user with 768 dimensions and so on for the rest of the users. To this end, we have a vector of features for every user. Therefore we can cluster them.

2) The clustering stage: In this stage, user embeddings are fed to cluster models to explore to what extent they can be clustered and show a visual representation of clusters. By exploring different models to explore the ability of clustering users, firstly, applying Hierarchical Density-Based Spatial Clustering (HDBSCAN), a density-based algorithm that works quite well and maintains a lot of local structure even in lower-dimensional space. Moreover, HDBSCAN does not force data points to clusters as it considers outliers. In addition, it is not required for us to determine the number of clusters; thus it is good at the start to explore if the embeddings can be clusters. In addition, Hierarchical clustering has been utilized, which is a cluster analysis method that seeks to build a cluster tree (a dendrogram) to represent data; therefore, it is good to visualize. The groups in the dendrogram are nested and organized as a tree, which ideally ends up as a meaningful classification scheme. Each node in the cluster tree contains a group of similar data.

The k-means clustering algorithm is the most common unsupervised method used in user modeling because it is easy for its implementation, simplicity, efficiency, and empirical success [63]. In addition, it is more systematic than other models, and it is required to specify the nuber of clusters in advance. Thus, the elbow method was utilized to find the proper value of the number of clusters. The idea of the elbow method is to run k-means clustering on the dataset for a range of values and for each value of k to calculate the sum of squared errors (SSE). Then, plot a line chart of the SSE for each value of k. After some value, the line will make a sharp ”elbow”, elbow’s shape will reveal the mean value of k that is best to use. After applying K mean clustering, the results will be visualized by T-distributed Stochastic Neighbor Embedding TSNE[1] which is a tool to visualize high-dimensional data.

7.3 Experiments

7.3.1 Experimental Setup

With choosing randomly 200k users from the network, all their posts aggregated to documents, and the BERT utilized in this chapter was from sentence-transformers package[2] as the resulting embeddings have shown to be of high quality and typically work immensely well for document-level embeddings [185]. The rest of the clustering models are from the scikit-learn library.

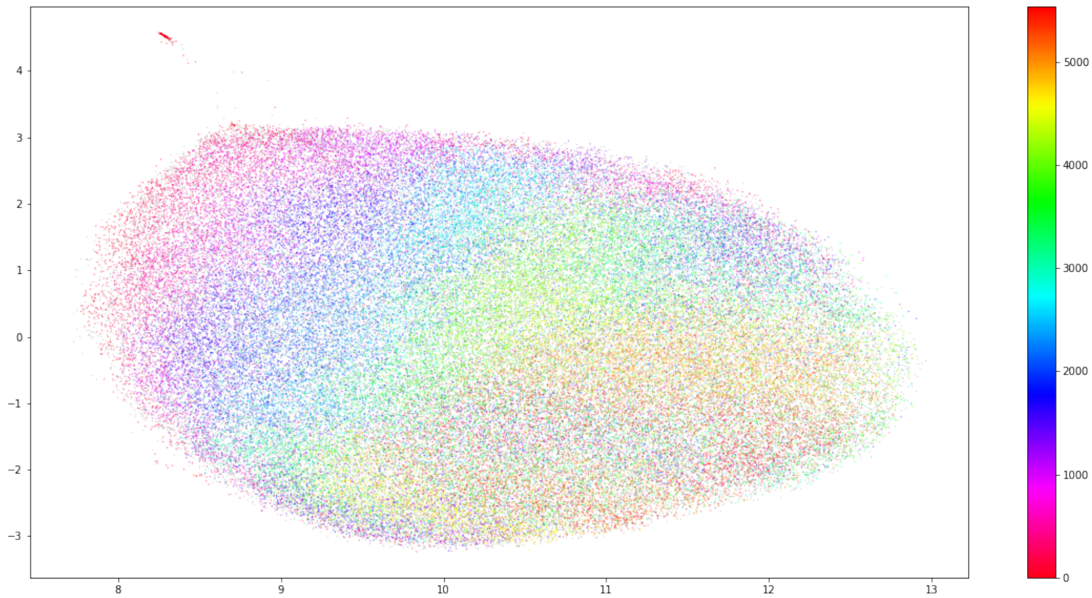


Figure 7.2: 200k Users embeddings clustering

[1] $https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html$
[2] $https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2$

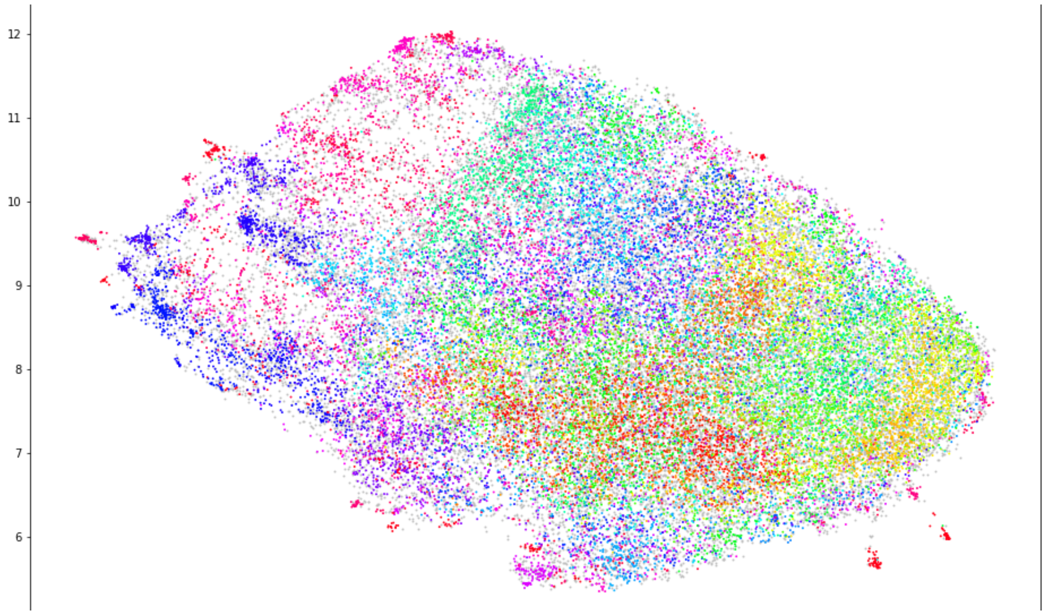


Figure 7.3: 100K Users text embeddings clustering based on Positive emotions

7.3.2 Results and Discussion

Figure 7.2 shows the clusters of 200k users embeddings after clustering with HDBSCAN. The constructed users vectors have too much dimensionality, therefore before visualizing them, dimensionality reduction algorithms (UMAP) [145] have been applied to reduce the dimensionality of the vectors to 2 dimensions. Likewise, Figure 7.3 represents the clusters of users in a positive mood. Even though it is challenging to visualize the individual clusters that have 768 dimensions with just 2 of them, we can see that some local structure is kept. Even when users post in the same class of mood, there are different clusters of them. This is a good indicator of that based on their content, whether user different from another in the behavior. Moreover, Figure 7.4 shows the dendrogram of Hierarchical clustering results of clustering 50k users, and it is clear that different groups have been constructed with numbers of users between (195 to 4500).

Figure 7.5 rely on the elbow method to identify the optimal number of clusters, and it is clear that the elbow method suggests the optimal number of clusters is somewhere around 7. Therefore, the K means algorithm applied with the number of classes to be 7.

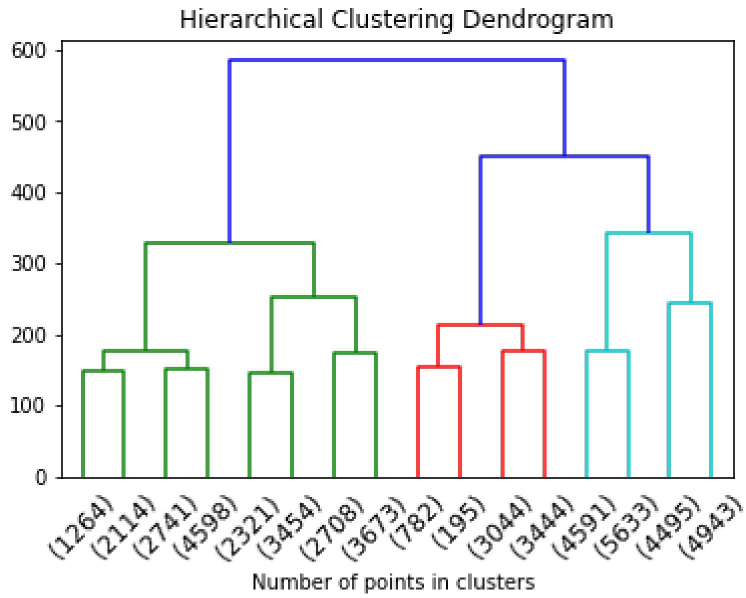


Figure 7.4: The dendrogram represented by Hierarchical Clustering Analysis of 50k users

Once this is identified, t-SNE (t-distributed Stochastic Neighbor Embedding) algorithm was applied. It is dimensionality reduction algorithms like PCA but it is nonlinear dimensionality reduction and it allows us to separate data that cannot be separated by any straight line. As a results, Figure 7.6 shows the map that has been produced using sklearn Python library to represent the clusters in a lower dimensionality (3 dimensions). The clusters with three dimensions visualization seem to be in a better structure.

7.3.3 Discussion

This chapter aimed to explore modeling users of mental health social network dataset based on clustering them into groups. By proposing the User BERT framework, the results show good indicators of the ability of modeling users.

However, clustering is just a first step to modeling users. It could be completing the process by labeling the clusters and assigning users to typical patterns in expressing feelings, seeking help, or giving support. Then, these clusters can be used to improve the emotion modelling by personalised modelling of emotion.

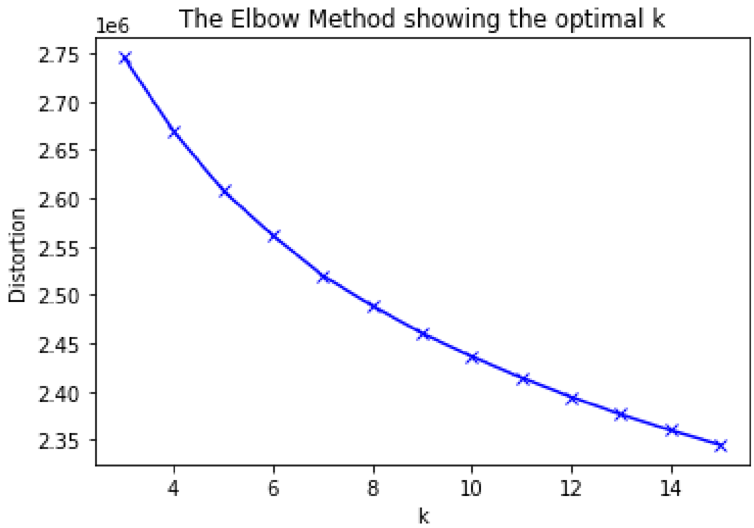


Figure 7.5: Representation of the elbow method to suggest the optimal number of clusters

Moreover, user embedding can be more comprehensive by taking into account the behavior profile of each user, other activities in the network such as follow, likes, replies to others, and user post histories to be used as features in specialized downstream models like emotion prediction.

7.4 Conclusion

In this chapter, the text of posts in mental health social networks is used to find groups of similar users by proposing a User BERT framework. The model uses the BERT model to construct user embedding and uses clustering algorithms to define groups of users and it is show the ability to cluster users when it is visualized.

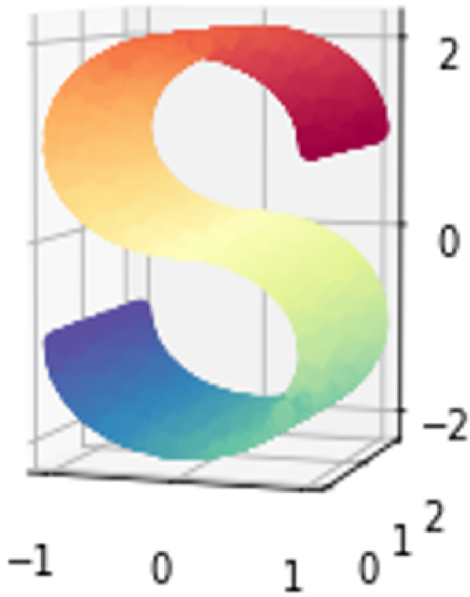


Figure 7.6: Map of 7 clusters created by K-mean algorithem and displayed in a lower dimensionality (3 dimensions) with the help of t-SNE algorithm

Chapter 8

Conclusion and Future Work

This thesis has presented a framework to model user profiles and user-generated data in a large-scale mental health social network. Most of our works explore how to model user emotion and the temporal dynamics of user emotion. This final chapter provides a general discussion of the work, a summary of our contributions, potential areas of improvement, the limitation of works and future work directions.

8.1 Summary of contributions of this thesis

This thesis has contributed to developing methods to improve the modeling of social network users by using a large-scale real-world mental health dataset and seeking to use these data to explore personalise user modeling. In particular, this thesis has studied how advanced deep learning models can help predict mental state and mood change over time. The conclusions for the various aspects of the project are presented below:

8.1.1 Data curation of the dataset

The data is the foundation of downstream user modeling and emotion analysis models. Obtaining a large-scale, well-annotated dataset is a challenging task. This thesis has been given access to a new peer-support online mental health social network dataset with millions of labeled posts (using TalkLife platform).

The dataset is self-labeled by the network users, the users choosing the label of emotion when they post in the network. Moreover, this type of labeling is advantageous and makes this dataset better than other available social network datasets when they are labeled by hand annotation or by hashtags. In chapter 3, the mental health social network dataset with its millions of users and user-generated contents has been introduced and analyzed in depth. This work ends up filtering the data to construct the TalkLife dataset that has been used for other chapters of this thesis.

8.1.2 Deep learning based emotion modelling

Text-based emotion modeling is one of the most challenging problems in natural language understanding. Therefore, with the availability of a user annotated large-scale dataset, in chapter 4, exploring wide range of machine learning models to classify the emotion of posts and applying the state-of-the-art pre-trained transformer model, Bidirectional Encoder Representations from Transformers (BERT) to the dataset. By evaluating and comparing the performance of the BERT model, it has been shown BERT model achieved the best score in evaluation metrics (F1-Score= 0.67). Moreover, the BERT system offers substantial improvements over the baseline models and outperforms the traditional NLP approach. This is proved that pre-trained contextualized language models, set by this chapter as a baseline for other researchers, can be usefully applied in text-based emotion modeling.

8.1.3 Fusing features with BERT emotion analysis model

The superiority of BERT model in emotion classification motivated me to fuse other features for further improvements in performance and interpretability, Therefore, chapter 5 proposed a novel perspective for modelling emotion based on social media text by integrating other features paradigms with semantic vectors that capture the semantic meaning from BERT model. By incorporating linguistics, grammatical features, mental health language-based features and some features regarding the user post. The performance results of the proposed model show substantial improvements over the vanila BERT model in the same task. Our works shows potential applications for the use of domain specific feature information to semantic processing in a wide range of current AI applications as well as in the potential benefit of leveraging mental health grounded linguistic features in a wide range of NLP tasks.

8.1.4 Modeling user mood changes

In chapter 6, to bridge the gap in modeling the change in mood of social networks, a novel predictive model and framework have been proposed to model the temporal dynamics of mental health social networks. The mood change model aims to find that future emotion change states can be predicted and the direction of the change in the mood. The proposed framework started by presenting the method for labeling the dataset, then extracting the proposed features. After that, we apply the model of mood change to predict if there is a change in the mood and if it is a negative or positive change. The Mood Change used predictive model consists of three different neural network units (2 BERT units and an LSTM unit), and it has achieved a good score of AUC to be 0.76. The proposed model was evaluated by adding sets of features every time and comparing the results. By analyzing the proposed features, we found a strong indicator of the influence of peers on the user's mood. This work prove to the effectiveness of the mental health social media in people's mental wellbeing by modelling the change of users mood through the time when they interacting with the network. Moreover, develop a novel predictive model and framework to model the change of the user mood and the direction of the change as well. This work also set up a novel user mood change modeling task for the research community.

8.1.5 User profile modelling of social media data

To explore profiling users of mental health social networks, the chapter 7 proposed a model (based on BERT) to learn unsupervised user embeddings from user post histories and then apply clustering models like K-Means for finding groups of similar users. These groups will define users with similar patterns to be used later in dawn stream tasks such as emotion modeling. Moreover, visualizing the resulting groups in 2 and 3 dimensions shows the ability to cluster users in groups via transformer-generated user profile embedding.

8.2 Limitations and Future Works

There are several lines of research arising from this work which could be pursued. Some suggestions are listed below:

- Although deep learning algorithms have achieved promising performance in canonical emotion analysis, knowledge-enhanced features still improve their performance, as we proved in chapter 5. However, these features are, in principle, limited by the availability of knowledge resources. Thus future directions are linked to acquiring them more efficiently and making the best use of them. Furthermore, feature acquisition usually requires a jump start through some form of manually acquired Knowledge to ensure quality for a specific domain. Hence automatically acquired part of the sentiment features can vastly extend the scale to improve coverage.

- Modelling the change of mood through time could be improved by utilizing more effective features that can capture the behavior of users in the network more precisely. In addition, the patterns of change behavior for every user could be modeled and used to classify users in the pattern of change to improve the prediction of change. Furthermore, more investigation can be conducted in model training refinement in future studies, e.g., Knowledge-driven BERT and affective-driven BERT through effective emotion vectors. Finally, mining moods and mood changes may be used as separate training targets for a multi-task learning process.

- User profile modeling is just explored briefly in chapter 7. I would improve user modeling by using deep user representation using Graph Embedding. I would use node2vec [83] and graph neural network to create user embedding and learn user profiles.

- Finally, I believe that such predictive models could be valuable tools in public health and could be used one day in designing and deploying new applications that can automatically identify those users who need help in the community.

References

[1] *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-II*, 2008. URL: https://aclanthology.org/I08-2000. 68

[2] Muhammad Abdul-Mageed and Lyle Ungar. Emonet: Fine-grained emotion detection with gated recurrent neural networks. In *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: Long papers)*, pages 718–728, 2017. 18

[3] Fabian Abel, Qi Gao, Geert-Jan Houben, and Ke Tao. Twitter-based user modeling for news recommendations. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, IJCAI '13, page 2962–2966. AAAI Press, 2013. 43, 45

[4] Francisca Adoma Acheampong, Chen Wenyu, and Henry Nunoo-Mensah. Text-based emotion detection: Advances, challenges, and opportunities. *Engineering Reports*, 2(7):e12189, 2020. 11

[5] Acheampong Francisca Adoma, Nunoo-Mensah Henry, Wenyu Chen, and Niyongabo Rubungo Andre. Recognizing emotions from texts using a bert-based approach. In *2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, pages 62–66, 2020. doi:10.1109/ICCWAMTIP51612.2020.9317523. 69, 81

[6] Zishan Ahmad, Raghav Jindal, Asif Ekbal, and Pushpak Bhattachharyya. Borrow from rich cousin: transfer learning for emotion detection using cross

lingual embedding. *Expert Systems with Applications*, 139:112851, 2020. 30, 68

[7] Thomas Aichner and Frank Jacob. Measuring the degree of corporate social media use. *International Journal of market research*, 57(2):257–276, 2015. 13

[8] Hani Al-Omari, Malak A. Abdullah, and Samira Shaikh. Emodet2: Emotion detection in english textual dialogue using bert and bilstm models. In *2020 11th International Conference on Information and Communication Systems (ICICS)*, pages 226–232, 2020. doi:10.1109/ICICS49469.2020.239539. 81, 82

[9] Ahmet Emre Aladağ, Serra Muderrisoglu, Naz Berfu Akbas, Oguzhan Zahmacioglu, and Haluk O Bingol. Detecting suicidal ideation on forums: proof-of-concept study. *Journal of medical Internet research*, 20(6):e9840, 2018. 14

[10] Jennifer Alvarez-Conrad, Lori A Zoellner, and Edna B Foa. Linguistic predictors of trauma pathology and physical health. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 15(7):S159–S170, 2001. 13

[11] Silvio Amir, Byron C. Wallace, Hao Lyu, and Paula Carvalho Mário J. Silva. Modelling context with user embeddings for sarcasm detection in social media, 2016. arXiv:1607.00976. 46, 47

[12] Michele Amoretti, Laura Belli, and Francesco Zanichelli. Utravel: Smart mobility with a novel user profiling and recommendation approach. *Pervasive and Mobile Computing*, 38:474–489, 2017. Special Issue IEEE International Conference on Pervasive Computing and Communications (PerCom) 2016. URL: https://www.sciencedirect.com/science/article/pii/S1574119216301341, doi:https://doi.org/10.1016/j.pmcj.2016.08.008. 44, 45

[13] L. ANDERSSON. Loneliness research and interventions: A review of the literature. *Aging & Mental Health*, 2(4):264–274, 1998. arXiv:https://doi.org/10.1080/13607869856506, doi:10.1080/13607869856506. 3

[14] Shlomo Argamon, Sushant Dhawle, Moshe Koppel, and James W Pennebaker. Lexical predictors of personality type. In *Proceedings of the 2005 Joint Annual Meeting of the Interface and the Classification Society of North America*, pages 1–16, 2005. 46

[15] Nabiha Asghar. Yelp dataset challenge: Review rating prediction. *arXiv preprint arXiv:1605.05362*, 2016. 20

[16] Reza Ashrafidoost and Saeed Setayeshi. A method for modelling and simulation the changes trend of emotions in human speech. In *Proceedings of The 9th EUROSIM Congress on Modelling and Simulation, EUROSIM 2016, The 57th SIMS Conference on Simulation and Modelling SIMS 2016*, number 142, pages 479–486. Linköping University Electronic Press, 2018. 38, 39

[17] Reza Ashrafidoost, Saeed Setayeshi, and Arash Sharifi. Recognizing the emotional state changes in human utterance by a learning statistical method based on gaussian mixture model. *Journal of Advances in Computer Engineering and Technology*, 3(2):113–124, 2017. 38

[18] Aldy Rialdy Atmadja and Ayu Purwarianti. Comparison on the rule based method and statistical based method on emotion classification for indonesian twitter text. In *2015 International Conference on Information Technology Systems and Innovation (ICITSI)*, pages 1–6, 2015. `doi:10.1109/ICITSI.2015.7437692`. 68

[19] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In *Lrec*, volume 10, pages 2200–2204, 2010. 18

[20] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate, 2016. `arXiv:1409.0473`. 32

[21] Sairam Balani and Munmun De Choudhury. Detecting and characterizing mental health related self-disclosure in social media. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1373–1378, 2015. 15

[22] Timothy Baldwin, Paul Cook, Marco Lui, Andrew MacKinlay, and Li Wang. How noisy social media text, how diffrnt social media sources? In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 356–364, 2013. 79

[23] Anushree Basu, Anirban Dasgupta, Anirud Thyagharajan, Aurobinda Routray, Rajlakshmi Guha, and Pabitra Mitra. A portable personality recognizer based on affective state classification using spectral fusion of features. *IEEE Transactions on Affective Computing*, 9(3):330–342, 2018. 38, 39

[24] Christos Baziotis, Nikos Athanasiou, Alexandra Chronopoulou, Athanasia Kolovou, Georgios Paraskevopoulos, Nikolaos Ellinas, Shrikanth Narayanan, and Alexandros Potamianos. Ntua-slp at semeval-2018 task 1: Predicting affective content in tweets with deep attentive rnns and transfer learning. *arXiv preprint arXiv:1804.06658*, 2018. 30

[25] Paul Best, Roger Manktelow, and Brian Taylor. Online communication, social media and adolescent wellbeing: A systematic narrative review. *Children and Youth Services Review*, 41:27–36, 2014. 13

[26] Rebecca Birnbaum and Daniel R Weinberger. Genetic insights into the neurodevelopmental origins of schizophrenia. *Nature Reviews Neuroscience*, 18(12):727–740, 2017. 15

[27] Johan Bollen, Huina Mao, and Alberto Pepe. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *Proceedings of the international AAAI conference on web and social media*, volume 5, 2011. 41, 42, 44

[28] Antoine Bordes, Sumit Chopra, and Jason Weston. Question answering with subgraph embeddings. *arXiv preprint arXiv:1406.3676*, 2014. 25

[29] Keith Bradley, Rachael Rafter, and Barry Smyth. Case-based user profiling for content personalisation. In Peter Brusilovsky, Oliviero Stock, and Carlo Strapparava, editors, *Adaptive Hypermedia and Adaptive Web-Based Systems*, pages 62–72, Berlin, Heidelberg, 2000. Springer Berlin Heidelberg. 45

[30] Margaret M Bradley and Peter J Lang. Affective norms for english words (anew): Instruction manual and affective ratings. Technical report, Technical report C-1, the center for research in psychophysiology ..., 1999. 17

[31] Felipe Bravo-Marquez, Marcelo Mendoza, and Barbara Poblete. Meta-level sentiment models for big social data analysis. *Knowledge-based systems*, 69:86–99, 2014. 24

[32] Rebecca C Brown, Eileen Bendig, Tin Fischer, A David Goldwich, Harald Baumeister, and Paul L Plener. Can acute suicidality be predicted by instagram data? results from qualitative and quantitative language analyses. *PloS one*, 14(9):e0220623, 2019. 91

[33] Sven Buechel and Udo Hahn. Emobank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 578–585, 2017. 20

[34] Robert J. Calsyn and Joel P. Winter. Social support, psychiatric symptoms, and housing: A causal analysis. *Journal of Community Psychology*, 30(3):247–259, 2002. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/jcop.10004`, `arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcop.10004`, `doi:https://doi.org/10.1002/jcop.10004`. 3

[35] Lea Canales and Patricio Martínez-Barco. Emotion detection from text: A survey. In *Proceedings of the workshop on natural language processing in the 5th information systems research working days (JISIC)*, pages 37–43, 2014. 11

[36] Kris Cao and Marek Rei. A joint model for word embedding and word morphology. *arXiv preprint arXiv:1606.02601*, 2016. 26

[37] Shaosheng Cao, Wei Lu, and Qiongkai Xu. Deep neural networks for learning graph representations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016. 48

[38] Aruna Chakraborty and Amit Konar. Fuzzy models for facial expression-based emotion recognition and control. In *Emotional Intelligence*, pages 133–173. Springer, 2009. ix, 38, 39

[39] Stevie Chancellor and Munmun De Choudhury. Methods in predictive techniques for mental health status on social media: a critical review. *NPJ digital medicine*, 3(1):1–11, 2020. 14, 15

[40] Stevie Chancellor, Zhiyuan Lin, Erica L Goodman, Stephanie Zerwas, and Munmun De Choudhury. Quantifying and predicting mental illness severity in online pro-eating disorder communities. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pages 1171–1184, 2016. 14

[41] Stevie Chancellor, Tanushree Mitra, and Munmun De Choudhury. Recovery amid pro-anorexia: Analysis of recovery in social media. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 2111–2123, 2016. 14

[42] Ankush Chatterjee, Kedhar Nath Narahari, Meghana Joshi, and Puneet Agrawal. Semeval-2019 task 3: Emocontext contextual emotion detection in text. In *Proceedings of the 13th international workshop on semantic evaluation*, pages 39–48, 2019. 29, 30, 68

[43] François-Régis Chaumartin. Upar7: A knowledge-based system for headline sentiment tagging. In *SemEval (ACL Workshop)*, pages pp–422, 2007. 22

[44] Dejun Chen, Congcong Xiong, and Ming Zhong. Improved lstm based on attention mechanism for short-term traffic flow prediction. In *2020 10th International Conference on Information Science and Technology (ICIST)*, pages 71–76, 2020. `doi:10.1109/ICIST49303.2020.9202045`. ix, 33

[45] Cynthia Chew and Gunther Eysenbach. Pandemics in the age of twitter: content analysis of tweets during the 2009 h1n1 outbreak. *PloS one*, 5(11):e14118, 2010. 40, 91

[46] Rumi Chunara, Jason R Andrews, and John S Brownstein. Social and news media enable estimation of epidemiological patterns early in the 2010

haitian cholera outbreak. *The American journal of tropical medicine and hygiene*, 86(1):39, 2012. 14

[47] Cindy Chung and James W Pennebaker. The psychological functions of function words. *Social communication*, 1:343–359, 2007. 13

[48] Ashley H Clawson, Belinda Borrelli, Elizabeth L McQuaid, and Shira Dunsiger. The role of caregiver social support, depressed mood, and perceived stress in changes in pediatric secondhand smoke exposure and asthma functional morbidity following an asthma exacerbation. *Health Psychology*, 35(6):541, 2016. 14

[49] Glen Coppersmith, Mark Dredze, and Craig Harman. Quantifying mental health signals in twitter. In *Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality*, pages 51–60, 2014. 81

[50] Paul Covington, Jay Adams, and Emre Sargin. Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*, RecSys '16, page 191–198, New York, NY, USA, 2016. Association for Computing Machinery. doi:10.1145/2959100.2959190. 47

[51] Aron Culotta. Estimating county health statistics with twitter. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1335–1344, 2014. 14

[52] Richard J Davidson. Affective neuroscience and psychophysiology: Toward a synthesis. *Psychophysiology*, 40(5):655–665, 2003. 9

[53] Munmun De Choudhury, Scott Counts, and Eric Horvitz. Predicting postpartum changes in emotion and behavior via social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, page 3267–3276, New York, NY, USA, 2013. Association for Computing Machinery. doi:10.1145/2470654.2466447. 3

[54] Munmun De Choudhury, Scott Counts, and Eric Horvitz. Predicting postpartum changes in emotion and behavior via social media. In *Proceedings*

of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13, page 3267–3276, New York, NY, USA, 2013. Association for Computing Machinery. doi:10.1145/2470654.2466447. 13, 14, 23, 40, 41, 44, 80, 81, 93, 96

[55] Munmun De Choudhury, Scott Counts, and Eric Horvitz. Social media as a measurement tool of depression in populations. In *Proceedings of the 5th annual ACM web science conference*, pages 47–56, 2013. 15, 91

[56] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. Predicting depression via social media. In *Seventh international AAAI conference on weblogs and social media*, 2013. 15, 42, 91

[57] Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 2098–2110, 2016. 14, 15

[58] Munmun De Choudhury, Meredith Ringel Morris, and Ryen W White. Seeking and sharing health information online: comparing search engines and social media. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1365–1376, 2014. 13

[59] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019. arXiv:1810.04805. ix, 35, 36, 66, 72

[60] Meili Ding, Robinson W Flaig, Hai-Long Jiang, and Omar M Yaghi. Carbon capture and conversion using metal–organic frameworks and mof-based materials. *Chemical Society Reviews*, 48(10):2783–2828, 2019. 38

[61] Cicero Dos Santos and Maira Gatti. Deep convolutional neural networks for sentiment analysis of short texts. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 69–78, 2014. 27

[62] Yogesh K Dwivedi, Nripendra P Rana, Anand Jeyaraj, Marc Clement, and Michael D Williams. Re-examining the unified theory of acceptance and use

of technology (utaut): Towards a revised theoretical model. *Information Systems Frontiers*, 21(3):719–734, 2019. 13

[63] Christopher Ifeanyi Eke, Azah Anir Norman, Liyana Shuib, and Henry Friday Nweke. A survey of user profiling: State-of-the-art, challenges, and solutions. *IEEE Access*, 7:144907–144924, 2019. `doi:10.1109/ACCESS.2019.2944243`. 45, 105, 107

[64] Paul Ekman. Basic emotions. *Handbook of cognition and emotion*, 98(45-60):16, 1999. 10

[65] Karima Elgarroussi, Sujing Wang, Romita Banerjee, and Christoph F Eick. Aconcagua: A novel spatiotemporal emotion change analysis framework. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*, pages 54–61, 2018. 38, 41, 44

[66] Golnoosh Farnadi, Geetha Sitaraman, Shanu Sushmita, Fabio Celli, Michal Kosinski, David Stillwell, Sergio Davalos, Marie-Francine Moens, and Martine De Cock. Computational personality recognition in social media. *User modeling and user-adapted interaction*, 26(2):109–142, 2016. 46

[67] Golnoosh Farnadi, Jie Tang, Martine De Cock, and Marie-Francine Moens. User profiling through deep multimodal fusion. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, WSDM '18, page 171–179, New York, NY, USA, 2018. Association for Computing Machinery. `doi:10.1145/3159652.3159691`. 80, 81

[68] Manaal Faruqui, Jesse Dodge, Sujay K Jauhar, Chris Dyer, Eduard Hovy, and Noah A Smith. Retrofitting word vectors to semantic lexicons. *arXiv preprint arXiv:1411.4166*, 2014. 18

[69] Allan Filipowicz, Sigal Barsade, and Shimul Melwani. Understanding emotional transitions: the interpersonal consequences of changing emotions in negotiations. *Journal of personality and social psychology*, 101(3):541, 2011. 38

[70] James H Fowler and Nicholas A Christakis. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study. *Bmj*, 337, 2008. 38

[71] Kathleen C Fraser, Frauke Zeller, David Harris Smith, Saif Mohammad, and Frank Rudzicz. How do we feel when a robot dies? emotions expressed on twitter before and after hitchbot's destruction. In *Proceedings of the Tenth Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 62–71, 2019. 16

[72] Nico H Frijda. Moods, emotion episodes, and emotions. 1993. 38

[73] Christoph Froschl. User modeling and user profiling in adaptive e-learning systems. *Graz, Austria: Master Thesis*, 2005. 43

[74] Min Gao, Kecheng Liu, and Zhongfu Wu. Personalisation in web computing and informatics: Theories, techniques, applications, and future research. *Information Systems Frontiers*, 12(5):607–629, 2010. 45

[75] Diman Ghazi, Diana Inkpen, and Stan Szpakowicz. Detecting emotion stimuli in emotion-bearing sentences. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 152–165. Springer, 2015. 20

[76] Sonja Gievska, Kiril Koroveshovski, and Natasha Tagasovska. Bimodal feature-based fusion for real-time emotion recognition in a mobile context. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 401–407. IEEE, 2015. 24

[77] J. Golbeck and J. Hendler. Filmtrust: movie recommendations using trust in web-based social networks. In *CCNC 2006. 2006 3rd IEEE Consumer Communications and Networking Conference, 2006.*, volume 1, pages 282–286, 2006. doi:10.1109/CCNC.2006.1593032. 44

[78] Santiago González-Carvajal and Eduardo C. Garrido-Merchán. Comparing BERT against traditional machine learning text classification. *CoRR*, abs/2005.13012, 2020. URL: https://arxiv.org/abs/2005.13012, arXiv:2005.13012. 35, 69

[79] Palash Goyal and Emilio Ferrara. Graph embedding techniques, applications, and performance: A survey. *Knowledge-Based Systems*, 151:78–94, 2018. URL: https://www.sciencedirect.com/science/article/pii/S0950705118301540, doi:https://doi.org/10.1016/j.knosys.2018.03.022. 47

[80] 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. 25, 28

[81] Kathleen M Griffiths, Andrew J Mackinnon, Dimity A Crisp, Helen Christensen, Kylie Bennett, and Louise Farrer. The effectiveness of an online support group for members of the community with depression: a randomised controlled trial. *PloS one*, 7(12):e53244, 2012. 14

[82] Aditya Grover and Jure Leskovec. Node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 855–864, New York, NY, USA, 2016. Association for Computing Machinery. doi:10.1145/2939672.2939754. 47, 48

[83] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864, 2016. 116

[84] Lin Gui, Ruifeng Xu, Yulan He, Qin Lu, and Zhongyu Wei. Intersubjectivity and sentiment: from language to knowledge. In *Social Media Content Analysis: Natural Language Processing and Beyond*, pages 129–144. World Scientific, 2018. 49

[85] Jamie Guillory, Jason Spiegel, Molly Drislane, Benjamin Weiss, Walter Donner, and Jeffrey Hancock. Upset now? emotion contagion in distributed groups. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 745–748, 2011. 12

[86] Kishaloy Halder, Lahari Poddar, and Min-Yen Kan. Modeling temporal progression of emotional status in mental health forum: A recurrent neural net approach. In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 127–135, 2017. ix, 41, 42, 44

[87] Tarek Hamdi, Hamda Slimi, Ibrahim Bounhas, and Yahya Slimani. A hybrid approach for fake news detection in twitter based on user features and graph embedding. In Dang Van Hung and Meenakshi D´Souza, editors, *Distributed Computing and Internet Technology*, pages 266–280, Cham, 2020. Springer International Publishing. 48, 105

[88] William L Hamilton, Jure Leskovec, and Dan Jurafsky. Cultural shift or linguistic drift? comparing two computational measures of semantic change. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing*, volume 2016, page 2116. NIH Public Access, 2016. 18

[89] Lixin Han and Guihai Chen. A fuzzy clustering method of construction of ontology-based user profiles. *Advances in Engineering Software*, 40(7):535–540, 2009. URL: https://www.sciencedirect.com/science/article/pii/S0965997808001762, doi:https://doi.org/10.1016/j.advengsoft.2008.10.006. 45

[90] Jeffrey T. Hancock, Kailyn Gee, Kevin Ciaccio, and Jennifer Mae-Hwah Lin. I'm sad you're sad: Emotional contagion in cmc. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, CSCW '08, page 295–298, New York, NY, USA, 2008. Association for Computing Machinery. doi:10.1145/1460563.1460611. 12, 38

[91] Maryam Hasan, Emmanuel Agu, and Elke Rundensteiner. Using hashtags as labels for supervised learning of emotions in twitter messages. In *ACM SIGKDD workshop on health informatics, New York, USA*, volume 34, page 100, 2014. 20

[92] Abdalraouf Hassan and Ausif Mahmood. Deep learning approach for sentiment analysis of short texts. In *2017 3rd international conference on*

control, automation and robotics (ICCAR), pages 705–710. IEEE, 2017. 30, 68

[93] Elaine Hatfield, John T Cacioppo, and Richard L Rapson. Emotional contagion. *Current directions in psychological science*, 2(3):96–100, 1993. 38

[94] Vasileios Hatzivassiloglou and Kathleen McKeown. Predicting the semantic orientation of adjectives. In *35th annual meeting of the association for computational linguistics and 8th conference of the european chapter of the association for computational linguistics*, pages 174–181, 1997. 18

[95] Tristan Henderson and Saleem Bhatti. Modelling user behaviour in networked games. In *Proceedings of the ninth ACM international conference on Multimedia*, pages 212–220, 2001. 1

[96] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. Session-based recommendations with recurrent neural networks, 2016. arXiv:1511.06939. 45

[97] Sepp Hochreiter. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(02):107–116, 1998. 29

[98] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997. doi:10.1162/neco.1997.9.8.1735. 29

[99] Christopher Homan, Ravdeep Johar, Tong Liu, Megan Lytle, Vincent Silenzio, and Cecilia Ovesdotter Alm. Toward macro-insights for suicide prevention: Analyzing fine-grained distress at scale. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 107–117, 2014. 13

[100] Qiongxia Huang, Riqing Chen, Xianghan Zheng, and Zhenxing Dong. Deep sentiment representation based on cnn and lstm. In *2017 International Conference on Green Informatics (ICGI)*, pages 30–33. IEEE, 2017. 30, 68

[101] Yen-Hao Huang, Ssu-Rui Lee, Mau-Yun Ma, Yi-Hsin Chen, Ya-Wen Yu, and Yi-Shin Chen. Emotionx-idea: Emotion BERT - an affectional model for conversation. *CoRR*, abs/1908.06264, 2019. URL: http://arxiv.org/abs/1908.06264, arXiv:1908.06264. 69

[102] Claire Hughes. Changes and challenges in 20 years of research into the development of executive functions. *Infant and child Development*, 20(3):251–271, 2011. 9

[103] Carroll E Izard. Emotion theory and research: Highlights, unanswered questions, and emerging issues. *Annual review of psychology*, 60:1–25, 2009. 9

[104] William James. What is emotion? 1884. 1948. 10

[105] Sébastien Jean, Kyunghyun Cho, Roland Memisevic, and Yoshua Bengio. On using very large target vocabulary for neural machine translation. *arXiv preprint arXiv:1412.2007*, 2014. 25

[106] Zhao Jianqiang, Gui Xiaolin, and Zhang Xuejun. Deep convolution neural networks for twitter sentiment analysis. *IEEE Access*, 6:23253–23260, 2018. doi:10.1109/ACCESS.2017.2776930. 28, 80, 81

[107] Rafal Jozefowicz, Wojciech Zaremba, and Ilya Sutskever. An empirical exploration of recurrent network architectures. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 2342–2350, Lille, France, 07–09 Jul 2015. PMLR. URL: https://proceedings.mlr.press/v37/jozefowicz15.html. 31, 68

[108] Samira Ebrahimi Kahou, Xavier Bouthillier, Pascal Lamblin, Caglar Gulcehre, Vincent Michalski, Kishore Konda, Sébastien Jean, Pierre Froumenty, Yann Dauphin, Nicolas Boulanger-Lewandowski, et al. Emonets: Multimodal deep learning approaches for emotion recognition in video. *Journal on Multimodal User Interfaces*, 10(2):99–111, 2016. 12

[109] Andreas M Kaplan and Michael Haenlein. Users of the world, unite! the challenges and opportunities of social media. *Business horizons*, 53(1):59–68, 2010. 13

[110] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1725–1732, 2014. 25

[111] Amirmohammad Kazameini, Samin Fatehi, Yash Mehta, Sauleh Eetemadi, and Erik Cambria. Personality trait detection using bagged svm over bert word embedding ensembles, 2020. `arXiv:2010.01309`. 68

[112] Geunyoung Kim, Tedra Walden, Vicki Harris, Jan Karrass, and Thomas Catron. Positive emotion, negative emotion, and emotion control in the externalizing problems of school-aged children. *Child psychiatry and human development*, 37(3):221–239, 2007. 10, 62

[113] Hannah Kim and Young-Seob Jeong. Sentiment classification using convolutional neural networks. *Applied Sciences*, 9(11):2347, 2019. 28

[114] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks, 2017. `arXiv:1609.02907`. 48

[115] Svetlana Kiritchenko, Xiaodan Zhu, and Saif M Mohammad. Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research*, 50:723–762, 2014. 24, 71, 79, 80

[116] Paul R Kleinginna and Anne M Kleinginna. A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and emotion*, 5(4):345–379, 1981. 8

[117] Shashidhar G Koolagudi and K Sreenivasa Rao. Emotion recognition from speech: a review. *International journal of speech technology*, 15(2):99–117, 2012. 12

[118] Ioannis Korkontzelos, Azadeh Nikfarjam, Matthew Shardlow, Abeed Sarker, Sophia Ananiadou, and Graciela H Gonzalez. Analysis of the effect

of sentiment analysis on extracting adverse drug reactions from tweets and forum posts. *Journal of biomedical informatics*, 62:148–158, 2016. 15

[119] Adam DI Kramer, Jamie E Guillory, and Jeffrey T Hancock. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24):8788–8790, 2014. 38

[120] Hanna Krasnova, Helena Wenninger, Thomas Widjaja, and Peter Buxmann. Envy on facebook: a hidden threat to users' life satisfaction? 2013. 13

[121] Taisa Kushner and Amit Sharma. Bursts of activity: Temporal patterns of help-seeking and support in online mental health forums. In *Proceedings of The Web Conference 2020*, WWW '20, page 2906–2912, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3366423.3380056. 59

[122] Taisa Kushner and Amit Sharma. Bursts of activity: Temporal patterns of help-seeking and support in online mental health forums. In *Proceedings of The Web Conference 2020*, pages 2906–2912, 2020. 93, 96, 97, 103

[123] Kleanthi Lakiotaki, Nikolaos F Matsatsinis, and Alexis Tsoukias. Multicriteria user modeling in recommender systems. *IEEE Intelligent Systems*, 26(2):64–76, 2011. 105

[124] Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. In Eric P. Xing and Tony Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 1188–1196, Bejing, China, 22–24 Jun 2014. PMLR. URL: https://proceedings.mlr.press/v32/le14.html. 46

[125] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015. 25

[126] Huayu Li, Yong Ge, Richang Hong, and Hengshu Zhu. Point-of-interest recommendations: Learning potential check-ins from friends. In *Proceedings of*

the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, page 975–984, New York, NY, USA, 2016. Association for Computing Machinery. doi:10.1145/2939672.2939767. 44, 45

[127] Minglei Li, Yunfei Long, Lu Qin, and Wenjie Li. Emotion corpus construction based on selection from hashtags. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1845–1849, 2016. 20

[128] Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. Dailydialog: A manually labelled multi-turn dialogue dataset. *arXiv preprint arXiv:1710.03957*, 2017. 21

[129] Jasy Suet Yan Liew and Howard R Turtle. Exploring fine-grained emotion detection in tweets. In *Proceedings of the NAACL student research workshop*, pages 73–80, 2016. 20

[130] Aristidis Likas, Nikos Vlassis, and Jakob J. Verbeek. The global k-means clustering algorithm. *Pattern Recognition*, 36(2):451–461, 2003. Biometrics. URL: https://www.sciencedirect.com/science/article/pii/S0031320302000602, doi:https://doi.org/10.1016/S0031-3203(02)00060-2. 105

[131] Chung-Ying Lin, Anders Broström, Per Nilsen, Mark D Griffiths, and Amir H Pakpour. Psychometric validation of the persian bergen social media addiction scale using classic test theory and rasch models. *Journal of behavioral addictions*, 6(4):620–629, 2017. 14

[132] Zhouhan Lin, Minwei Feng, Cicero Nogueira dos Santos, Mo Yu, Bing Xiang, Bowen Zhou, and Yoshua Bengio. A structured self-attentive sentence embedding, 2017. arXiv:1703.03130. 32

[133] Bing Liu and Lei Zhang. A survey of opinion mining and sentiment analysis. In *Mining text data*, pages 415–463. Springer, 2012. 15

[134] Yang Liu. Fine-tune BERT for extractive summarization. *CoRR*, abs/1903.10318, 2019. URL: http://arxiv.org/abs/1903.10318, arXiv:1903.10318. 66

[135] Yang Liu and Mirella Lapata. Text summarization with pretrained encoders. *CoRR*, abs/1908.08345, 2019. URL: http://arxiv.org/abs/1908.08345, arXiv:1908.08345. 66

[136] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach, 2019. arXiv:1907.11692. 37

[137] Zhonghua Ma and Themis J Michailides. Advances in understanding molecular mechanisms of fungicide resistance and molecular detection of resistant genotypes in phytopathogenic fungi. *Crop Protection*, 24(10):853–863, 2005. 22

[138] Andrew Maas, Raymond E Daly, Peter T Pham, Dan Huang, Andrew Y Ng, and Christopher Potts. Learning word vectors for sentiment analysis. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies*, pages 142–150, 2011. 20

[139] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967. 105

[140] François Mairesse, Marilyn A Walker, Matthias R Mehl, and Roger K Moore. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of artificial intelligence research*, 30:457–500, 2007. 46

[141] Navonil Majumder, Soujanya Poria, Devamanyu Hazarika, Rada Mihalcea, Alexander Gelbukh, and Erik Cambria. Dialoguernn: An attentive rnn for emotion detection in conversations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6818–6825, 2019. 41

[142] Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60, 2014. 63

[143] Viera Maslej-Krešňáková, Martin Sarnovský, Peter Butka, and Kristína Machová. Comparison of deep learning models and various text preprocessing techniques for the toxic comments classification. *Applied Sciences*, 10(23), 2020. URL: https://www.mdpi.com/2076-3417/10/23/8631, doi:10.3390/app10238631. 89

[144] Iris B Mauss, Robert W Levenson, Loren McCarter, Frank H Wilhelm, and James J Gross. The tie that binds? coherence among emotion experience, behavior, and physiology. *Emotion*, 5(2):175, 2005. 9

[145] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Grossberger. Umap: Uniform manifold approximation and projection. *The Journal of Open Source Software*, 3(29):861, 2018. 109

[146] David J McIver, Jared B Hawkins, Rumi Chunara, Arnaub K Chatterjee, Aman Bhandari, Timothy P Fitzgerald, Sachin H Jain, and John S Brownstein. Characterizing sleep issues using twitter. *Journal of medical Internet research*, 17(6):e140, 2015. 14

[147] Albert Mehrabian. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4):261–292, 1996. 11

[148] Herbert L Meiselman. *Emotion measurement*. Woodhead publishing, 2016. 10

[149] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013. arXiv:1301.3781. 26

[150] Tomáš Mikolov, Stefan Kombrink, Lukáš Burget, Jan Černocký, and Sanjeev Khudanpur. Extensions of recurrent neural network language model. In

2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 5528–5531, 2011. doi:10.1109/ICASSP.2011.5947611. 28

[151] George A Miller. *WordNet: An electronic lexical database*. MIT press, 1998. 18, 22

[152] Rohan Mishra, Pradyumn Prakhar Sinha, Ramit Sawhney, Debanjan Mahata, Puneet Mathur, and Rajiv Ratn Shah. Snap-batnet: Cascading author profiling and social network graphs for suicide ideation detection on social media. In *Proceedings of the 2019 conference of the North American Chapter of the association for computational linguistics: student research workshop*, pages 147–156, 2019. 46

[153] Ibrahim F. Moawad, Hanaa Talha, Ehab Hosny, and Mohamed Hashim. Agent-based web search personalization approach using dynamic user profile. *Egyptian Informatics Journal*, 13(3):191–198, 2012. URL: https://www.sciencedirect.com/science/article/pii/S1110866512000382, doi:https://doi.org/10.1016/j.eij.2012.09.002. 45

[154] Bamshad Mobasher. Data mining for web personalization. In *The adaptive web*, pages 90–135. Springer, 2007. 45

[155] Saif Mohammad. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 174–184, 2018. 17

[156] Saif Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. Semeval-2018 task 1: Affect in tweets. In *Proceedings of the 12th international workshop on semantic evaluation*, pages 1–17, 2018. 18

[157] Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, pages 31–41, 2016. 19

[158] Saif M Mohammad and Felipe Bravo-Marquez. Emotion intensities in tweets. *arXiv preprint arXiv:1708.03696*, 2017. 21

[159] Saif M Mohammad and Svetlana Kiritchenko. Using hashtags to capture fine emotion categories from tweets. *Computational Intelligence*, 31(2):301–326, 2015. 18

[160] Saif M Mohammad and Peter D Turney. Crowdsourcing a word–emotion association lexicon. *Computational intelligence*, 29(3):436–465, 2013. 17, 18

[161] Robert R Morris, Stephen M Schueller, and Rosalind W Picard. Efficacy of a web-based, crowdsourced peer-to-peer cognitive reappraisal platform for depression: randomized controlled trial. *Journal of medical Internet research*, 17(3):e72, 2015. 14

[162] Moin Nadeem. Identifying depression on twitter. *arXiv preprint arXiv:1607.07384*, 2016. 15

[163] Yabo Ni, Dan Ou, Shichen Liu, Xiang Li, Wenwu Ou, Anxiang Zeng, and Luo Si. Perceive your users in depth: Learning universal user representations from multiple e-commerce tasks, 2018. arXiv:1805.10727. 45, 48

[164] Finn Årup Nielsen. Afinn. *Richard Petersens Plads, Building*, 321, 2011. 17

[165] Andrew Ortony, Gerald L Clore, and Allan Collins. *The cognitive structure of emotions*. Cambridge university press, 1990. 10

[166] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010. doi:10.1109/TKDE.2009.191. 34

[167] Bo Pang and Lillian Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *arXiv preprint cs/0409058*, 2004. 24

[168] Jaak Panksepp. The periconscious substrates of consciousness: Affective states and the evolutionary origins of the self. *Journal of consciousness studies*, 5(5-6):566–582, 1998. 8

[169] Jaak Panksepp. Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and cognition*, 14(1):30–80, 2005. 9

[170] Michael J Pazzani. A framework for collaborative, content-based and demographic filtering. *Artificial intelligence review*, 13(5):393–408, 1999. 45

[171] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. 71

[172] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011. 70

[173] James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. The development and psychometric properties of liwc2015. Technical report, 2015. 17, 81

[174] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014. 26, 70

[175] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, page 701–710, New York, NY, USA, 2014. Association for Computing Machinery. `doi:10.1145/2623330.2623732`. 47

[176] Rosalind W Picard and Jennifer Healey. Affective wearables. *Personal technologies*, 1(4):231–240, 1997. 9

[177] Soujanya Poria, Alexander Gelbukh, Erik Cambria, Amir Hussain, and Guang-Bin Huang. Emosenticspace: A novel framework for affective common-sense reasoning. *Knowledge-Based Systems*, 69:108–123, 2014. 19, 22

[178] Julie Prescott, Terry Hanley, and Katalin Ujhelyi Gomez. Why do young people use online forums for mental health and emotional support? benefits and challenges. *British Journal of Guidance & Counselling*, 47(3):317–327, 2019. arXiv:https://doi.org/10.1080/03069885.2019.1619169, doi:10.1080/03069885.2019.1619169. 50

[179] Julie Prescott, Amy Leigh Rathbone, and Gill Brown. Online peer to peer support: Qualitative analysis of uk and us open mental health facebook groups. *DIGITAL HEALTH*, 6:2055207620979209, 2020. PMID: 33354335. arXiv:https://doi.org/10.1177/2055207620979209, doi:10.1177/2055207620979209. 50

[180] Victor M Prieto, Sergio Matos, Manuel Alvarez, Fidel Cacheda, and Jose Luis Oliveira. Twitter: a good place to detect health conditions. *PloS one*, 9(1):e86191, 2014. 14

[181] Yada Pruksachatkun, Sachin R. Pendse, and Amit Sharma. *Moments of Change: Analyzing Peer-Based Cognitive Support in Online Mental Health Forums*, page 1–13. Association for Computing Machinery, New York, NY, USA, 2019. 3

[182] Yada Pruksachatkun, Sachin R. Pendse, and Amit Sharma. *Moments of Change: Analyzing Peer-Based Cognitive Support in Online Mental Health Forums*, page 1–13. Association for Computing Machinery, New York, NY, USA, 2019. URL: https://doi.org/10.1145/3290605.3300294. 40, 92, 93, 95, 96, 100

[183] Nairan Ramirez-Esparza, Cindy K Chung, Ewa Kacewicz, and James W Pennebaker. The psychology of word use in depression forums in english and in spanish: Texting two text analytic approaches. In *ICWSM*, 2008. 13

[184] Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 2931–2937, 2017. 80

[185] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 11 2019. URL: https://arxiv.org/abs/1908.10084. 108

[186] Elaine Rich. User modeling via stereotypes. *Cognitive Science*, 3(4):329–354, 1979. URL: https://www.sciencedirect.com/science/article/pii/S0364021379800129, doi:https://doi.org/10.1016/S0364-0213(79)80012-9. 43

[187] Kirk Roberts, Michael A Roach, Joseph Johnson, Josh Guthrie, and Sanda Harabagiu. Empatweet: Annotating and detecting emotions on twitter. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 3806–3813, 2012. 19

[188] Stephen Robertson. Understanding inverse document frequency: on theoretical arguments for idf. *Journal of documentation*, 2004. 70

[189] Werner Rothengatter. Integration or disintegration of infrastructure and transport companies in the european railway sector–the pro's and con's. In *11th World Conference on Transport ResearchWorld Conference on Transport Research Society*, 2007. 1

[190] Stephanie Rude, Eva-Maria Gortner, and James Pennebaker. Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, 18(8):1121–1133, 2004. 13

[191] James A Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980. ix, 11

[192] Farig Sadeque, Ted Pedersen, Thamar Solorio, Prasha Shrestha, Nicolas Rey-Villamizar, and Steven Bethard. Why do they leave: Modeling partic-

ipation in online depression forums. In *Proceedings of The Fourth International Workshop on Natural Language Processing for Social Media*, pages 14–19, 2016. 96, 97

[193] Daniel Scanfeld, Vanessa Scanfeld, and Elaine L Larson. Dissemination of health information through social networks: Twitter and antibiotics. *American journal of infection control*, 38(3):182–188, 2010. 40

[194] Klaus R. Scherer. What are emotions? and how can they be measured? *Social Science Information*, 44(4):695–729, 2005. arXiv:https://doi.org/10.1177/0539018405058216, doi:10.1177/0539018405058216. 9, 12, 21, 51

[195] Klaus R Scherer and Harald G Wallbott. Evidence for universality and cultural variation of differential emotion response patterning. *Journal of personality and social psychology*, 66(2):310, 1994. 19, 20

[196] Hendrik Schuff, Jeremy Barnes, Julian Mohme, Sebastian Padó, and Roman Klinger. Annotation, modelling and analysis of fine-grained emotions on a stance and sentiment detection corpus. In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 13–23, 2017. 19

[197] M. Schuster and K.K. Paliwal. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681, 1997. doi:10.1109/78.650093. 29

[198] Benjamin I Schwartz. *Chinese communism and the rise of Mao*. Harvard University Press, 2014. 13

[199] JiaXing Shen, Mingyu Derek Ma, Rong Xiang, Qin Lu, Elvira Perez Vallejos, Ge Xu, Chu-Ren Huang, and Yunfei Long. Dual memory network model for sentiment analysis of review text. *Knowledge-Based Systems*, 188:105004, 2020. URL: https://www.sciencedirect.com/science/article/pii/S0950705119304198, doi:https://doi.org/10.1016/j.knosys.2019.105004. 49

[200] Anu Shrestha, Edoardo Serra, and Francesca Spezzano. Multi-modal social and psycho-linguistic embedding via recurrent neural networks to identify depressed users in online forums. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 9(1):1–11, 2020. 80, 81, 105

[201] Lin Shu, Jinyan Xie, Mingyue Yang, Ziyi Li, Zhenqi Li, Dan Liao, Xiangmin Xu, and Xinyi Yang. A review of emotion recognition using physiological signals. *Sensors*, 18(7):2074, 2018. 12

[202] Alessio Signorini, Alberto Maria Segre, and Philip M. Polgreen. The use of twitter to track levels of disease activity and public concern in the u.s. during the influenza a h1n1 pandemic. *PLOS ONE*, 6(5):1–10, 05 2011. doi:10.1371/journal.pone.0019467. 3

[203] Alessio Signorini, Alberto Maria Segre, and Philip M Polgreen. The use of twitter to track levels of disease activity and public concern in the us during the influenza a h1n1 pandemic. *PloS one*, 6(5):e19467, 2011. 40, 91

[204] Edin Smailhodzic, Wyanda Hooijsma, Albert Boonstra, and David J Langley. Social media use in healthcare: A systematic review of effects on patients and on their relationship with healthcare professionals. *BMC health services research*, 16(1):1–14, 2016. 14

[205] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1631–1642, 2013. 20

[206] Jacopo Staiano and Marco Guerini. Depechemood: a lexicon for emotion analysis from crowd-annotated news. *arXiv preprint arXiv:1405.1605*, 2014. 19, 22

[207] Philip J Stone, Dexter C Dunphy, and Marshall S Smith. The general inquirer: A computer approach to content analysis. 1966. 17

[208] Carlo Strapparava and Rada Mihalcea. Learning to identify emotions in text. In *Proceedings of the 2008 ACM symposium on Applied computing*, pages 1556–1560, 2008. 19, 20, 22

[209] Carlo Strapparava, Alessandro Valitutti, et al. Wordnet affect: an affective extension of wordnet. In *Lrec*, volume 4, page 40. Lisbon, 2004. 19, 22

[210] Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, and Rob Fergus. End-to-end memory networks, 2015. `arXiv:1503.08895`. 33

[211] Chi Sun, Xipeng Qiu, Yige Xu, and Xuanjing Huang. How to fine-tune bert for text classification? In Maosong Sun, Xuanjing Huang, Heng Ji, Zhiyuan Liu, and Yang Liu, editors, *Chinese Computational Linguistics*, pages 194–206, Cham, 2019. Springer International Publishing. 66, 69

[212] Deliang Sun, Haijia Wen, Danzhou Wang, and Jiahui Xu. A random forest model of landslide susceptibility mapping based on hyperparameter optimization using bayes algorithm. *Geomorphology*, 362:107201, 2020. 23

[213] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112, 2014. 28

[214] Chenhao Tan, Lillian Lee, Jie Tang, Long Jiang, Ming Zhou, and Ping Li. User-level sentiment analysis incorporating social networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11, page 1397–1405, New York, NY, USA, 2011. Association for Computing Machinery. `doi:10.1145/2020408.2020614`. 49

[215] Jie Tang, Limin Yao, Duo Zhang, and Jing Zhang. A combination approach to web user profiling. *ACM Trans. Knowl. Discov. Data*, 5(1), December 2010. `doi:10.1145/1870096.1870098`. 45

[216] Raphael Tang, Yao Lu, Linqing Liu, Lili Mou, Olga Vechtomova, and Jimmy Lin. Distilling task-specific knowledge from bert into simple neural networks, 2019. `arXiv:1903.12136`. 37

[217] Jianhua Tao and Tieniu Tan. Affective computing: A review. In *International Conference on Affective computing and intelligent interaction*, pages 981–995. Springer, 2005. 2, 12

[218] Channary Tauch and Eiman Kanjo. The roles of emojis in mobile phone notifications. UbiComp '16, page 1560–1565, New York, NY, USA, 2016. Association for Computing Machinery. doi:10.1145/2968219.2968549. 83

[219] Yla R. Tausczik and James W. Pennebaker. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54, 2010. arXiv:https://doi.org/10.1177/0261927X09351676, doi:10.1177/0261927X09351676. 81

[220] Mark A Thornton and Diana I Tamir. Mental models accurately predict emotion transitions. *Proceedings of the National Academy of Sciences*, 114(23):5982–5987, 2017. 38, 39

[221] Bing Tian, Yong Zhang, and Chunxiao Xing. Improving document-level sentiment classification with user-product gated network. In Leong Hou U, Marc Spaniol, Yasushi Sakurai, and Junying Chen, editors, *Web and Big Data*, pages 397–412, Cham, 2021. Springer International Publishing. 49

[222] Cunchao Tu, Hao Wang, Xiangkai Zeng, Zhiyuan Liu, and Maosong Sun. Community-enhanced network representation learning for network analysis. *arXiv preprint arXiv:1611.06645*, 2016. 48

[223] Andranik Tumasjan, Timm Sprenger, Philipp Sandner, and Isabell Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 4, 2010. 15

[224] Matthew Turk. Multimodal interaction: A review. *Pattern recognition letters*, 36:189–195, 2014. 2

[225] Chai M Tyng, Hafeez U Amin, Mohamad NM Saad, and Aamir S Malik. The influences of emotion on learning and memory. *Frontiers in psychology*, 8:1454, 2017. 10

[226] van den Oord, Aäron and Dieleman, Sander and Schrauwen, Benjamin. Deep content-based music recommendation. In Burges, Christopher and Bottou, Léon and Welling, Max and Ghahramani, Zoubin and Weinberger, Kilian, editor, *Advances in Neural Information Processing Systems 26 (2013)*, volume 26, page 9. Neural Information Processing Systems Foundation (NIPS), 2013. URL: http://papers.nips.cc/paper/5004-deep-content-based-music-recommendation.pdf. 45

[227] C Natalie van der Wal and Wojtek Kowalczyk. Detecting changing emotions in human speech by machine and humans. *Applied intelligence*, 39(4):675–691, 2013. 38, 39

[228] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017. ix, 32, 34, 35

[229] Tedra A Walden, Vicki S Harris, and Thomas F Catron. How i feel: a self-report measure of emotional arousal and regulation for children. *Psychological assessment*, 15(3):399, 2003. 11, 62

[230] Chenyu Wang, Zhongchen Miao, Yuefeng Lin, and Jian Gao. User and topic hybrid context embedding for finance-related text data mining. In *2019 International Conference on Data Mining Workshops (ICDMW)*, pages 751–760, 2019. doi:10.1109/ICDMW.2019.00112. 49

[231] Daixin Wang, Peng Cui, and Wenwu Zhu. Structural deep network embedding. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 1225–1234, New York, NY, USA, 2016. Association for Computing Machinery. doi:10.1145/2939672.2939753. 48

[232] Hao Wang, Naiyan Wang, and Dit-Yan Yeung. *Collaborative Deep Learning for Recommender Systems*, page 1235–1244. Association for Computing Machinery, New York, NY, USA, 2015. URL: https://doi.org/10.1145/2783258.2783273. 45

[233] Ri Wang, Maysum Panju, and Mahmood Gohari. Classification-based rnn machine translation using grus. *arXiv preprint arXiv:1703.07841*, 2017. ix, 31

[234] Xiao Wang, Peng Cui, Jing Wang, Jian Pei, Wenwu Zhu, and Shiqiang Yang. Community preserving network embedding. In *Thirty-first AAAI conference on artificial intelligence*, 2017. 47

[235] Xin Wang, Yuanchao Liu, Cheng-Jie Sun, Baoxun Wang, and Xiaolong Wang. Predicting polarities of tweets by composing word embeddings with long short-term memory. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1343–1353, 2015. 30, 68

[236] Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. Attention-based LSTM for aspect-level sentiment classification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 606–615, Austin, Texas, November 2016. Association for Computational Linguistics. URL: `https://aclanthology.org/D16-1058`, `doi:10.18653/v1/D16-1058`. 68

[237] David Watson. *Mood and temperament.* Guilford Press, 2000. 9

[238] Zhongyu Wei, Yang Liu, and Yi Li. Is this post persuasive? ranking argumentative comments in online forum. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 195–200, 2016. 97

[239] Jason Weston, Sumit Chopra, and Antoine Bordes. Memory networks, 2015. `arXiv:1410.3916`. 33

[240] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing*, pages 347–354, 2005. 17

[241] Meylan Wongkar and Apriandy Angdresey. Sentiment analysis using naive bayes algorithm of the data crawler: Twitter. In *2019 Fourth International Conference on Informatics and Computing (ICIC)*, pages 1–5. IEEE, 2019. 23

[242] Chung-Hsien Wu, Ze-Jing Chuang, and Yu-Chung Lin. Emotion recognition from text using semantic labels and separable mixture models. *ACM transactions on Asian language information processing (TALIP)*, 5(2):165–183, 2006. 12

[243] Xiaodong Wu, Weizhe Lin, Zhilin Wang, and Elena Rastorgueva. Author2vec: A framework for generating user embedding, 2020. arXiv:2003.11627. 48

[244] Jianguo Xia, Igor V Sinelnikov, Beomsoo Han, and David S Wishart. Metaboanalyst 3.0—making metabolomics more meaningful. *Nucleic acids research*, 43(W1):W251–W257, 2015. 24

[245] Ashima Yadav and Dinesh Kumar Vishwakarma. Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review*, 53(6):4335–4385, 2020. ix, 21, 23, 25, 31

[246] Carl Yang, Lin Zhong, Li-Jia Li, and Luo Jie. Bi-directional joint inference for user links and attributes on large social graphs. In *Proceedings of the 26th International Conference on World Wide Web Companion*, WWW '17 Companion, page 564–573, Republic and Canton of Geneva, CHE, 2017. International World Wide Web Conferences Steering Committee. doi:10.1145/3041021.3054181. 43, 45

[247] Kisu Yang, Dongyub Lee, Taesun Whang, Seolhwa Lee, and Heuiseok Lim. Emotionx-ku: Bert-max based contextual emotion classifier. *CoRR*, abs/1906.11565, 2019. URL: http://arxiv.org/abs/1906.11565, arXiv:1906.11565. 69

[248] Min Yang, Qiang Qu, Xiaojun Chen, Chaoxue Guo, Ying Shen, and Kai Lei. Feature-enhanced attention network for target-dependent sentiment classification. *Neurocomputing*, 307:91–97, 2018. URL: https://www.

sciencedirect.com/science/article/pii/S0925231218304764, doi: https://doi.org/10.1016/j.neucom.2018.04.042. 33

[249] Wei Yang, Yuqing Xie, Aileen Lin, Xingyu Li, Luchen Tan, Kun Xiong, Ming Li, and Jimmy Lin. End-to-end open-domain question answering with. *Proceedings of the 2019 Conference of the North*, 2019. URL: http://dx.doi.org/10.18653/v1/N19-4013, doi:10.18653/v1/n19-4013. 66

[250] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. Xlnet: Generalized autoregressive pretraining for language understanding. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL: https://proceedings.neurips.cc/paper/2019/file/dc6a7e655d7e5840e66733e9ee67cc69-Paper.pdf. 37

[251] Louerrad Yasmina and Kaid-Harche Meriem. Evaluation of phenolic compounds of two lygeum spartum l. cytotypes. *African Journal of Biotechnology*, 15(36):1991–1994, 2016. 23

[252] Tom Young, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. Recent trends in deep learning based natural language processing [review article]. *IEEE Computational Intelligence Magazine*, 13(3):55–75, 2018. doi:10.1109/MCI.2018.2840738. ix, 28, 29

[253] Zhigang Yuan, Sixing Wu, Fangzhao Wu, Junxin Liu, and Yongfeng Huang. Domain attention model for multi-domain sentiment classification. *Knowledge-Based Systems*, 155:1–10, 2018. URL: https://www.sciencedirect.com/science/article/pii/S0950705118302144, doi: https://doi.org/10.1016/j.knosys.2018.05.004. 33

[254] Nikolaos Zacharakis, Harshini Chinnasamy, Mary Black, Hui Xu, Yong-Chen Lu, Zhili Zheng, Anna Pasetto, Michelle Langhan, Thomas Shelton, Todd Prickett, et al. Immune recognition of somatic mutations leading to complete durable regression in metastatic breast cancer. *Nature medicine*, 24(6):724–730, 2018. 23

[255] Biqiao Zhang and Emily Mower Provost. Automatic recognition of self-reported and perceived emotions. In *Multimodal Behavior Analysis in the Wild*, pages 443–470. Elsevier, 2019. 15, 16

[256] Daokun Zhang, Jie Yin, Xingquan Zhu, and Chengqi Zhang. User profile preserving social network embedding. In *IJCAI International Joint Conference on Artificial Intelligence*, 2017. 48

[257] Jia-Dong Zhang and Chi-Yin Chow. Igslr: Personalized geo-social location recommendation: A kernel density estimation approach. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, SIGSPATIAL'13, page 334–343, New York, NY, USA, 2013. Association for Computing Machinery. doi:10.1145/2525314.2525339. 44, 45

[258] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. Bertscore: Evaluating text generation with bert, 2020. arXiv:1904.09675. 66

[259] Wei Zhang, Wen Wang, Jun Wang, and Hongyuan Zha. User-guided hierarchical attention network for multi-modal social image popularity prediction. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, page 1277–1286, Republic and Canton of Geneva, CHE, 2018. International World Wide Web Conferences Steering Committee. doi:10.1145/3178876.3186026. 47

[260] Ye Zhang and Byron Wallace. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. *arXiv preprint arXiv:1510.03820*, 2015. 28

[261] Ye Zhang and Byron C. Wallace. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. *CoRR*, abs/1510.03820, 2015. URL: http://arxiv.org/abs/1510.03820, arXiv:1510.03820. 68

[262] Hao Zhou, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. Emotional chatting machine: Emotional conversation generation with in-

ternal and external memory. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018. 105

[263] Peng Zhou, Zhenyu Qi, Suncong Zheng, Jiaming Xu, Hongyun Bao, and Bo Xu. Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling. *CoRR*, abs/1611.06639, 2016. URL: http://arxiv.org/abs/1611.06639, arXiv:1611.06639. 29

[264] Jinhua Zhu, Yingce Xia, Lijun Wu, Di He, Tao Qin, Wengang Zhou, Houqiang Li, and Tie-Yan Liu. Incorporating bert into neural machine translation, 2020. arXiv:2002.06823. 66

[265] Yukun Zhu, Ryan Kiros, Richard Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books, 2015. arXiv:1506.06724. 36

[266] Łukasz Kaiser and Ilya Sutskever. Neural gpus learn algorithms, 2016. arXiv:1511.08228. 31, 68

Appendix

.1 Experimental settings

Grid search has been applied to all models to get the best set of weights for each model.

For biLSTM and GRU models, the hidden layer dimensionality has been experimented with 64, 120 and 256, the learning rate to 0.1, 0.001 and 0.0001 the dropout of 0.1 and 0.2.

For GloVe, the parameters and its values are context windows size $W = 5, 7, 9, 11, 13, 15$, dimensionality $D = 100, 200, 300, 400, 500$, and iteration $I = 5, 10, 15, 20, 25, 30$.

The pre-trained BERT has been fine tuned using GPU with the following values as recommended in the original BERT paper:

- Batch size: 16, 32
- Learning rate (Adam): 5e-5, 3e-5, 2e-5
- Number of epochs: 2, 3, 4
- Max sequence length of 40
- Training steps of 100,000
- Warm-up steps of 10,000

The dropout probability is always kept at 0.1, and the warm-up proportion is 0.1 and saves the best model on the validation set for testing.