Detecting Depression Using K-Nearest Neighbors (KNN) Classification Technique

Md Rafiqul Islam[1], Abu Raihan M.Kamal[1], Naznin Sultana[2], Robiul Islam[1], Mohammad Ali Moni[3], Anwaar ulhaq[4]
[1]Department of CSE, Islamic University of Technology (IUT), Dhaka, Bangladesh
[2]Department of CSE, Daffodil Int. University, Dhaka, Bangladesh
[3]Department of Biomedical Science, University of Sydney, Sydney, Australia
[4]College of Engineering & Science, Victoria University, Melbourne, Australia
Corresponding author's E-mail: rafiqulislam.cse24@gmail.com

Abstract— **Social networks have developed as a promising point for everybody to communicate with their interested friend and share their opinions, photos, and videos. Also, it has been an upcoming research field and has picked an established position globally. In this paper, we considered depression problems among various Facebook users. Already, a number of researchers have studied and applied many techniques to detect depression, but still need to detect accurately from social network data. So, we investigate the possibility to utilize Facebook data and apply KNN (k-nearest neighbors) classification technique for detecting depressive emotions. We do believe that our investigation and approach might be helpful to raise consciousness in online social network users.**

Keywords— Emotions, depression, sentiment analysis, predictions, social media.

I. Introduction

Social networks (Facebook, Twitter, LinkedIn, WhatsApp, and Viber etc.) are becoming the preferred communication channel for people nowadays. People can easily communicate with their interested friend and share their information's, photos, and videos to each other [1] [16]. Moreover, they can express their feelings as a post and can comment on another post on the social media platform. But sometimes their posts and comments refer to as their emotional expressions, such as joy, sadness, fear, anger, and surprise. The analysis of emotions in the text is a necessary pre-processing step in many different fields such as natural language interfaces [2], sentiment analysis, e-learning and educational environment, opinion mining [3] and security applications [4]. In addition, sometimes these sort of feelings, for example, happiness, pity and surprise is valuable for various purposes, including recognizing online journals that express particular feelings towards the point of interest, distinguishing what feeling a daily paper feature is trying to inspire, and formulating programmed discourse frameworks that react appropriately to various emotional conditions of the users [5]. In general, diverse feelings are communicated through various words. For instance, delightful and yummy show the feeling of happiness, gloomy and cry are demonstrative of misery, shout and boiling are characteristic of anger and so on. [6].

Nowadays, detecting depression from social network data is an emerging research field in emotion processing [7]. In this paper, we aim at mining one of the essential deep level types of information, i.e. depression detection, which gives suitable information for public opinion mining. In emotion processing, the reason event and emotion relationship is a rich ground for extraction and entailment of new data. As an initial move towards completely automatic inference of cause feeling correlation, we applied data classification based approaches to detect depression in this paper. Note that we have taken only comment texts from the different post of Facebook users from a specific Facebook page. In this paper, we examine different linguistic prompts from those comments, which helped us to identify feeling cause events: emotional process like positive feeling (e.g. happy, nice), negative feeling (e.g. worth, loser), sadness (e.g. worry, crying, sad), anger (e.g. Stop, Shit) and anxiety (e.g. nervous, terrible). Some of the temporal processes are present focus (e.g. now, today, is), past focus (e.g. talked, did, ago) and future focus (e.g. shall, will, may). Linguistic words like articles (e.g. a, an, the), prepositions (e.g. for, of, to, with, above), auxiliary verbs (e.g. do, have, am, will), conjunctions (e.g. and, But, whereas), personal pronoun (e.g. I, them, her), impersonal pronouns (e.g. it's, those), verbs (e.g. Go, Good) and negation (e.g. Deny, Dishonest, no, not, never).

The goal of this article is to investigate the depressive emotion and the process of evolution in social network data by outlining research inquiries from existing exploration, giving cases of new procedures and applications, and lighting up future research bearings. As a major aspect of this, a portion of the particular research questions addresses the part of depression in feeling within social network public comments, using Facebook.

(a) How to extract data from Facebook users comment? and

(b) How to detect depression using classification techniques?

To address the above research questions, we applied NCapture to collect data and KNN approach for extracting paraphrases to detect emotions from user comments. Basically, we focused on three basic emotions based on emotional process, temporal process, and linguistic analysis. We used LIWC tool for analyzing user comments to detect emotion level. For this purpose, we used a large collection of Facebook comments data from bipolar, depression and anxiety Facebook page, in which the accuracy is measured and which are described in Section IV.

The major contributions of this paper are as follows:

We discuss many of the relevant works that have shown how various emotion detection techniques of the social network have already been used to detect depression.

Manually we organize and justify the ground truth result of our dataset for testing the classification technique.

With these ideas established, we applied KNN classification technique to detect a set of emotions that can detect depression.

The remainder of the paper is organized as follows: Section II presents the related work of detecting depression analysis of social network data. Discussion about data is explained in Section III. The classification and experimental analysis are presented in Section IV, and its discussion in Section V. Finally, the conclusion and future work are provided in Section VI.

II. RELATED WORK

Choudhury et al. [6] examined the possibility to use the online social network to identify and analyze the significant depressive issues in people. Through their web-based social networking postings, they quantified behavioral credits identifying/identified? with/from? social engagement, feeling, dialect and semantic styles, sense of the self-system, and notices of antidepressant medications.

Choudhury et al. [8] considered online networking as a promising instrument for public health, concentrating on the utilization of Twitter presence on fabricating predictive models about the forthcoming impact of childbirth on the conduct and disposition of new mothers. Utilizing Twitter posts, they measured postpartum changes in 376 mothers along measurements of social engagement, feeling, informal community, and phonetic style.

B. O'Dea et al.[9, 10] has examined that Twitter is progressively researched as methods for recognizing psychological well-being status, including depression and suicidality in the population and shown to recognize the level of depression/anxiety among suicide-related tweets.

On the other hand, L. Zhang et al. [10] have shown that if individuals with a high danger of suicide can be recognized through online networking like microblog, it is conceivable to actualize a dynamic intervention system to save their lives.

Many researchers have demonstrated that utilizing user-created/generated content (UGC) accurately may help decide individuals' psychological wellness levels. M. M. Aldarwish and H. F. Ahmad [11] examined that the utilization of Social Network Sites (SNS) is expanding these days, particularly by the more youthful eras. The accessibility of SNS enables clients to express their interests, sentiments and offer day by day schedule.

Nowadays a number of researchers using online groups to examine psychological well-being issues. T. Nguyen et al. [12] utilized machine learning and statistical strategies to separate online messages amongst depression and control groups utilizing temperament, psycholinguistic procedures and substance subjects removed from the posts created by individuals from these groups.

From our point of view, this work is the first attempt to specify depression problem from Facebook user comments because none of the above exiting work used Facebook user comments for detecting depression. With our current work, we analyze the scope of social media-based depression measures, describing the different features of Facebook user comments. We applied machine-learning approaches that can use those measures to detect individuals who are suffering from depression.

III. DATA

In this section, we describe about the data collection procedure and how we constructed our dataset in details.

We used NCapture for collecting data from Facebook because it is a powerful software for qualitative data analysis in the world today [18]. After collecting the data from Facebook, we processed our dataset by using LIWC2015. Our data record includes 5 emotional variables (positive, negative, sad, anger, anxiety), 3 temporal categories (present focus, past focus and future focus), and 9 standard linguistic dimensions (e.g., articles, prepositions, auxiliary verb, adverbs, conjunctions, pronoun, verbs and negations).

We now discuss how we constructed our dataset with ground truth label information (on whether or not the comments is depression indicative). We used facebook user comments divided by two sets (1) for the positive (YES) class (depression indicative comments) and (2) for the negative (NO) class (non-depression indicative comments). For all of the comments of each set, we justified this by two experts manually. We obtained total 7145 comments where 58% obtained YES for depression indicative comments and 42% obtained for No for non-depressive indicative comments.

IV. CLASSIFICATION AND EXPERIMENTAL FINDINGS

In this section, we investigate the performance of different KNN classifiers in detecting depression in a shorter time. The experiment is conducted using MATLAB 2016b. We applied KNN classifiers: Fine KNN, Medium KNN, Coarse KNN, Cosine KNN, Cubic KNN and Weighted KNN.

Using the different types of KNN techniques, we examined depression detection performance of Facebook data. So as to comprehend the significance of different feature types, we applied different types of KNN techniques each utilizing: Emotional process, linguistic style, temporal process and all features. The results of these techniques are shown in table 1 and table 2. The outcome of our research approach shows that, in our test set, the best performing model is Course KNN. Performances of these classifiers have been computed by using the evaluation matrices parameters such as precision recall and F-measure. The experiment is carried out by using 10-fold cross-validation on all testing datasets. For every classifier, we showed the value of its sub-classifier which persists to high F-measure? (see figure 1 and 2).

Table 1: Precision and Recall corresponding to the best F-measure of the emotional process and linguistic style using different KNN classifier.

Feature	Emotional Process			Linguistic Style		
K Nearest Neighbors (KNN)						
Algorithm	Pr.	Re.	Fm.	Pr.	Re.	Fm.
Fine KNN	.59	.59	.59	.58	.58	.58
Medium KNN	.59	.59	.59	.59	.55	.57
Coarse KNN	.59	.88	.71	.59	.80	.70
Cosine KNN	.58	.59	.58	.59	.60	.60
Cubic KNN	.59	.59	.59	.60	.54	.57
Weighted KNN	.58	.62	.60	.59	.65	.62

Table 2: Precision and Recall corresponding to the best F-measure of the temporal process and all features using different KNN classifier.

Feature	Temporal process			All features		
K Nearest Neighbors (KNN)						
Algorithm	Pr.	Re.	Fm.	Pr.	Re.	Fm.
Fine KNN	.57	.58	.58	.59	.57	.58
Medium KNN	.58	.57	.57	.59	.53	.56
Coarse KNN	.58	.89	.70	.59	.77	.67
Cosine KNN	.59	.58	.59	.60	.60	.60
Cubic KNN	.58	.57	.57	.59	.52	.55
Weighted KNN	.57	.59	.58	.59	.64	.61

V. DISCUSSION

For a better understanding of the general intuition behind depression, in this paper, we applied different types of KNN classification techniques for depression detection from emotional expression. We showed that these classification techniques based on linguistic style, emotional process, temporal process and all features are able to successfully extract the depressive emotional result. Tables 1 and 2 demonstrate the results of various characterizations with various proportions of four features. We examine all of the four features. It can be observed that Coarse KNN gives the better outcome. We believe that the present study would lay the ground for future research on deductions and revelation of new data in view of cause-event connection, for example, discovery of understood feeling or cause, and additionally expectation of popular conclusion in view of cause occasions, and so forth.

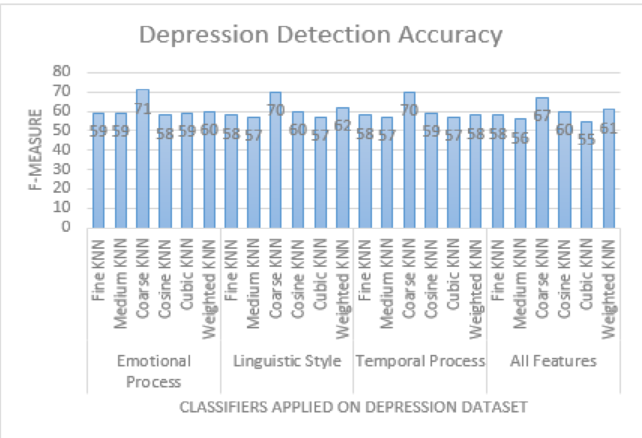

Figure 1: Depression detection accuracy of emotional process, linguistic style, temporal process and all features based on different classifier.

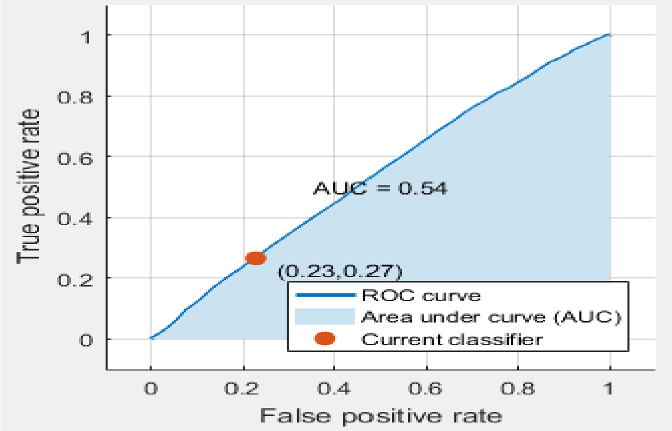

Figure 2: ROC curve showing the sensitivity of course KNN for all features.

VI. CONCLUSION AND FUTURE WORK

We have exhibited the capability of utilizing Facebook data as a source for measuring and detecting major depression among users. We studied four types of factors (emotional process, temporal process, linguistic style, and all features) and trained a model to utilize each type independently and jointly. Our findings show that the ground truth dataset result and different types of KNN technique results vary between 60-70% in terms of different metrics level. For future work, we have a plan to use another technique to extract paraphrases from more types of emotional features.

VII. REFERENCES

[1] R. Zafarani, M. A. Abbasi, and H. Liu, *Social media mining: an introduction*: Cambridge University Press, 2014.

[2] J. Yin, A. Lampert, M. Cameron, B. Robinson, and R. Power, "Using social media to enhance emergency situation awareness," *IEEE Intelligent Systems,* vol. 27, pp. 52-59, 2012.

[3] N. Ofek, G. Katz, B. Shapira, and Y. Bar-Zev, "Sentiment analysis in transcribed utterances," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2015, pp. 27-38.

[4] Y. Yang, P. Cui, W. Zhu, and S. Yang, "User interest and social influence based emotion prediction for individuals," in *Proceedings of the 21st ACM international conference on Multimedia*, 2013, pp. 785-788.

[5] Y. Wang and A. Pal, "Detecting Emotions in Social Media: A Constrained Optimization Approach," in *IJCAI*, 2015, pp. 996-1002.

[6] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting Depression via Social Media," *ICWSM,* vol. 13, pp. 1-10, 2013.

[7] E. Gilbert and K. Karahalios, "Predicting tie strength with social media," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2009, pp. 211-220.

[8] M. De Choudhury, S. Counts, and E. Horvitz, "Predicting postpartum changes in emotion and behavior via social media," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2013, pp. 3267-3276.

[9] B. O'Dea, S. Wan, P. J. Batterham, A. L. Calear, C. Paris, and H. Christensen, "Detecting suicidality on Twitter," *Internet Interventions,* vol. 2, pp. 183-188, 2015.

[10] L. Zhang, X. Huang, T. Liu, A. Li, Z. Chen, and T. Zhu, "Using linguistic features to estimate suicide probability of Chinese microblog users," in *International Conference on Human Centered Computing*, 2014, pp. 549-559.

[11] M. M. Aldarwish and H. F. Ahmad, "Predicting Depression Levels Using Social Media Posts," in *Autonomous Decentralized System (ISADS), 2017 IEEE 13th International Symposium on*, 2017, pp. 277-280.

[12] T. Nguyen, D. Phung, B. Dao, S. Venkatesh, and M. Berk, "Affective and content analysis of online depression communities," *IEEE Transactions on Affective Computing,* vol. 5, pp. 217-226, 2014.

[13] M. E. Seligman, *Helplessness: On depression, development, and death*: WH Freeman/Times Books/Henry Holt & Co, 1975.

[14] L. A. Jelenchick, J. C. Eickhoff, and M. A. Moreno, “Facebook depression?” Social networking site use and depression in older adolescents," *Journal of Adolescent Health,* vol. 52, pp. 128-130, 2013.

[15] P. Bazeley and K. Jackson, *Qualitative data analysis with NVivo*: Sage Publications Limited, 2013.

[16] Md Rafiqul Islam, Naznin Sultana, Mohammad Ali Moni, Prohollad Chandra Sarkar and Bushra Rahman. A Comprehensive Survey of Time Series Anomaly Detection in Online Social Network Data. *International Journal of Computer Applications* 180(3):13-22, December 2017.