To what extent does machine learning benefit in predicting personal mental issues through social media analysis?

Suihan Gao

Department of computer science, University of North Carolina, Chapel Hill, United States

sgao4@unc.edu

Abstract. Social media has become an increasingly popular platform for individuals to express their emotions and share personal experiences. This unique characteristic of social media has led to its recognition as a valuable resource for studying mental health trends at the individual level. This article investigates the application of machine learning techniques, such as sentiment analysis, emotion extraction, and topic modeling, in detecting signs of depressive symptoms within social media content. By examining and synthesizing findings from multiple studies, this review provides an analysis of the advantages and limitations of these machine learning techniques in this specific context. Furthermore, it highlights the ethical considerations that should be taken into account in future research endeavors. The integration of machine learning techniques with social media data presents a promising avenue for identifying and addressing mental health concerns, but it is essential to approach this research responsibly and ethically. The growing body of knowledge in this field can contribute to the development of effective interventions and support for individuals experiencing depressive symptoms.

Keywords: Machine learning, sentiment analysis, topic modeling, mental health, native language processing, machine learning.

1. Introduction

With the growing concern for mental well-being becoming a global issue in nowadays society, diagnosing mental health issues (MHI) requires technological improvements, which help to compensate for the insufficiency aspect of the traditional diagnoses. The traditional methodology for mental health diagnosis usually relies on a common practice short-termed as the "DSM-5." The diagnosis identifies aspects like patients's physical signs and symptoms combined with clinical expertise to determine responses to stress [1]. However, relying heavily on the diagnosis of the DSM-5 could be prone to human errors that present with wrong results: "The 'false positives problem' of mislabeling normal condition as mental disorders is the issue that most impact psychiatric epidemiology, given its heavy reliance on DSM criteria in community studies" [2]. Thus, seeking new psychological diagnosis methods that would benefit from perfecting existing methodology becomes necessary. Thankfully, the advancements of the digital era took advantage of the rapid expression of ideas. Platforms like Twitter, Facebook, and Instagram help users to exchange emotions and thoughts remotely. The vast number of shared posts and tweets have inadvertently become an embodiment of various human sentiments, whether it was positive or negative emotions expressed. It was found that there is a positive relationship between the form of depression and the use of social media. Excessive usage indicates disastrous results, leading to anxiety and further depression [3]. Such an environment, on the other side, has offered an excellent opportunity for scholars to analyze potential MHI within the users. Applying the advancement of Machine Learning through model training may be an efficient tool. Through the pattern and logical analysis of a massive load of context, examples like Natural Language Processing (NLP), the crossfield intersection between Computer Science and Linguistics has the capability to indicate an individual's psychological status based on their social media contents. This paper embarks on a review of modern research to explore to what extent machine learning technology can detect signals of personal MHI via social media content.

2. Sentimental Analysis

The application of Sentimental Analysis allows the interpretation of emotions through posted social media contexts by performing algorithms. The application of sentimental analysis is a pivotal aspect of Natural Language Processing (NLP) that can analyze and discern the subjective ideas from the text to interpret sentiments, attitudes, and emotions from social media contexts by performing methods like emotion extraction. The essence of sentimental analysis is to transform unstructured text into structured data for analysis, which figures sentiments conveyed by textual discourse.

Sentimental analysis can process unstructured text into different categories, for instance, positive, neutral, and negative, by classifying categories of sentiments. It has three levels of classification: document, sentence, and aspect level. At the document level of classification, its sentiment classification classifies the overall attitude of an opinionated document. Using semantic orientation, researchers can detect the polarity of the documents and use the average value to determine whether the document sentiment is positive. The sentence classification can classify the sentiment of the opinionated sentence since not all the sentence provides opinion. The aspect classification focuses on the emotions and opinions expressed by entities [4]. It is more important to focus on aspect classification than the general document level. The reason is that people might express an overall attitude or emotion in total, but they might express different emotions over distinct entities. For instance, people might share their daily routine over social media platforms, which is overall objective in statements. By recognizing entities in detail, we can utilize aspect classification to find their opinions and emotions on details that are not quickly revealed or expressed.

In the next step of data processing for sentimental analysis, the probabilistic classifier is a typical mathematical model that estimates the probability of a given instance belonging to each possible class. The instance could conceptually be a data point or a single item under examination, while classes refer to various categories. In modern data analysis, probabilistic classifiers such as the Maximum Entropy classifier (ME), Bayesian Network (BN), and Naïve Bayes Classifier (NB) are often used. And the Naïve Bayes classifier is the most prevalently employed among all [5]. For instance, a Native Bayes classifier can be trained and labeled sentiment data to learn the likelihood of words and phrases occurring with positive or negative sentiment contexts. Therefore, the probabilistic classifier technique can better represent and analyze the uncertainties in data and classification processed from social media content.

Another approach can be based on the lexicon-based methods, which use sentiment lexicon to analyze the strength of the expressed feelings and emotions. There are twin approaches to building a sentiment lexicon: dictionary-based and corpus-based [5]. Since it can not only produce positive, neutral, and negative output, it can provide a more precise and detailed output based on a range. In this scenario, researchers can figure out various levels of negative emotions to detect the most negative and extreme instances. For instance, in the research searching extreme opinions by Sattam Almatarneh and Pablo Gamallo, they first created a corpus-based lexicon with two values: not very negative and very negative, and the same logic for the positive side. Then, they process sentimental classification based on lexical resources from websites showing public data. Then, they can generate the lexicons by any corpus with labels of stars from one to N stars, meaning from most negative to most positive. Calculating the relative frequency (RF) can represent each word by (word, tag) pairs. The next phase involves computing the average RF values across various ranges: specifically comparing not most negative (NMN) to most negative (MN) and not most positive (NMP) to most positive (MP). Also, they define an extreme borderline to evaluate extreme opinions and use weights to categorize the result [6]. With this experience, it is easier for researchers to have a clearer and more detailed definition of different sentimental scales, which benefits them during sentimental analysis.

3. Topic Modeling

Through the gathering of topics and word terminology, the topic modeling strategy spots recurring words into stress-related themes in such a way as to examine signs of MHI. As a general overview of the topic modeling, it uses a Bag of Word (BOW) approach (Figure 1) [7]. There are very specific rules to the approach, which state that a set should only contain items that only appeared once in a set, and a bag may have items that can appear more than once [7]. These rules are in the purpose of preventing not to overweight words by allowing them to repeatedly occur which might inflate the importance of certain terminologies. Topic modeling helps to be sufficient at identifying similarities within the context because it doesn't require knowing everything about the actual documents [7]. Such a strategy indicates convincing results in the detection of the MHI. There are many ways in which topic modeling could be performed. For instance, one research team utilized the *eRisk 2018* dataset, a cluster of written production from Reddit, a social media platform [8].

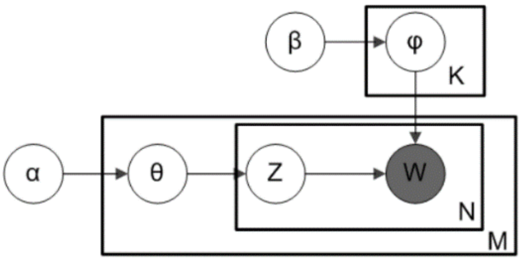
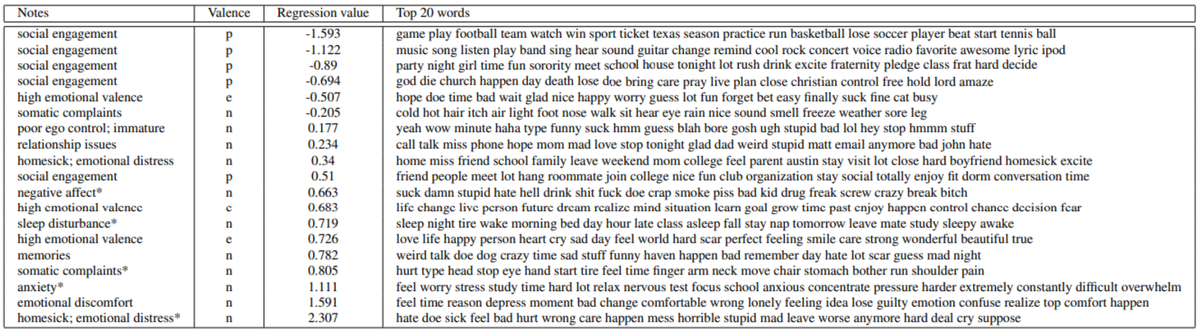


Figure 1. Topic Modeling [7]

Datasets are then divided into two groups training and testing. There are ten chunks of the dataset for each group. The users and RISK users then categorize the datasets, RISK writings, and no-risk writings. The researchers then underwent topic extractions where the one discussing could perspectively reflect his or her mental stability. To extract the topic, the research team constructed a simple model using the Latent Dirichlet Allocation (LDA) model. LDA is a statistical model that can determine similarities between the two variables. In this case, the words and topic terms are presented in a Dirichlet prior distribution under the input parameters, which are the general range of the depression-related topic and words. LDA then gives out a range from which these terminologies have similarities. In such a way, LDA helps to categorize one's writing as a portion belonging to broader topics and thus maps the user's posted context into the prediction of mental risk.

4. Supervised LDA

Another similar experiment was conducted, in which researchers also applied LDA as the model for analyzing the depression-related language contents on the Twitter platform. However, the researcher adopted a supervised LDA that substitutes the original, as it is believed that the original LDA is just the starting point when it comes to characterizing topics [9]. An extension to the LDA method is required. The advanced sLDA model displays higher accuracy, whereas LDA might only indicate opinion words like great, enjoy, and dislike. The modified sLDA model has the capability to separate them out into positive opinion topic words associated with higher scores versus negative opinion topics such as "dislike," "sucks," and "bad." A lower rating would be given. The figure 2 of the attached graph shows the sLDA topics that the researcher was able to acquire by running sLDA on the Pennebaker stream-of-consciousness dataset [9]. The determination of risk factors relates to degrees of neuroticism, one of the psychological Big-5 personality traits. Showing on the table are the top 20 words of each level of intensity. For example, the words “hate,” “bad,” “stupid,” “suck,” and “mad” were given with a valence of “N,” indicating the negatively correlated score in the topic category of negative affect. On the other hand, words like the game, “football,” “team,” “win, “ticket,” and “excite" are organized in the valence of “P,” informing that no signs of depression were present. As the value of a topic’s regression valence increases, it is less connected with neuroticism, the negative trait of a person’s negative emotions and lousy self-regulation. The presented value through topic models is indicated with professional judgments on its accuracy. In fact, the sLDA values are in accordance with clinician judgments. It was shown that this dataset, where the interpreted values are drawn from, is able to produce clean and interpretable topic classification where clinicians are referenceable under the assessment of depression and physical discomfort symptoms.



Notes	Valence	Regression value	Top 20 words
social engagement	p	-1.593	game play football team watch win sport ticket texas season practice run basketball lose soccer player beat start tennis ball
social engagement	p	-1.122	music song listen play band sing hear sound guitar change remind cool rock concert voice radio favorite awesome lyric ipod
social engagement	p	-0.89	party night girl time fun sorority meet school house tonight lot rush drink excite fraternity pledge class frat hard decide
social engagement	p	-0.694	god die church happen day death lose doe bring care pray live plan close christian control free hold lord amaze
high emotional valence	e	-0.507	hope doe time bad wait glad nice happy worry guess lot fun forget bet easy finally suck fine cat busy
somatic complaints	n	-0.205	cold hot hair itch air light foot nose walk sit hear eye rain nice sound smell freeze weather sore leg
poor ego control; immature	n	0.177	yeah wow minute haha type funny suck hmm guess blah bore gosh ugh stupid bad lol hey stop hmmm stuff
relationship issues	n	0.234	call talk miss phone hope mom mad love stop tonight glad dad weird stupid matt email anymore bad john hate
homesick; emotional distress	n	0.34	home miss friend school family leave weekend mom college feel parent austin stay visit lot close hard boyfriend homesick excite
social engagement	p	0.51	friend people meet lot hang roommate join college nice fun club organization stay social totally enjoy fit dorm conversation time
negative affect*	n	0.663	suck damn stupid hate hell drink shit fuck doe crap smoke piss bad kid drug freak screw crazy break bitch
high emotional valence	e	0.683	life change live person future dream realize mind situation learn goal grow time past enjoy happen control chance decision fear
sleep disturbance*	n	0.719	sleep night tire wake morning bed day hour late class asleep fall stay nap tomorrow leave mate study sleepy awake
high emotional valence	e	0.726	love life happy person heart cry sad day feel world hard scar perfect feeling smile care strong wonderful beautiful true
memories	n	0.782	weird talk doe dog crazy time sad stuff funny haven happen bad remember day hate lot scar guess mad night
somatic complaints*	n	0.805	hurt type head stop eye hand start tire feel time finger arm neck move chair stomach bother run shoulder pain
anxiety*	n	1.111	feel worry stress study time hard lot relax nervous test focus school anxious concentrate pressure harder extremely constantly difficult overwhelm
emotional discomfort	n	1.591	feel time reason depress moment bad change comfortable wrong lonely feeling idea lose guilty emotion confuse realize top comfort happen
homesick; emotional distress*	n	2.307	hate doe sick feel bad hurt wrong care happen mess horrible stupid mad leave worse anymore hard deal cry suppose

Figure 2. The top 20 words of each level of intensity [9]

5. Discussion

Though there are several approaches to analyzing data from social media platforms, many challenges still exist in processing the data. In sentimental analysis, building a sentimental lexicon to analyze feelings and emotions has become a frequently used method, and this lexicon-based method belongs to the unsupervised learning methods, which do not need any training. Although the advantage of the method is to process data and provide analysis results quickly, when it comes to sarcasm, the apparent effect of emotional words will significantly weaken or even have the opposite effect. Therefore, sarcasm detection is needed—for instance, the article from Joshi et al. examined the use features of word embeddings in the context of sarcasm detection, and they also conducted experiments with four different algorithms with augmented word embeddings [10]. Besides unsupervised lexicon-based methods, methods related to machine learning are considered supervised methods. For instance, SVM and Naïve Bayes are commonly used models. However, “Naïve Bayes is successful when applied to well-formed text corpus” [11] does not work well on social media platforms since content published by people has random length and plenty of spelling typos, making the training significantly harder and lowering the quality of the analysis [11]. Therefore, combining these two types of methods could complement each other, improving the analysis output.

6. Conclusion

There are many ways in which ML could apply its usefulness in analyzing internet context to extract and interpret emotions to formulate assessments on depression. The methods of sentiment analysis process the data with different levels and apply the probabilistic model that is able to seek within different levels of phrase paragraphs for possibility classes and trace possibilities under given instances.

The topic modeling technique is able to approach in such a way that categories terminologies into approximate topics and thus concluding words with most presented in this article are showcasing some of the strategies to indicate that these existing technologies could adapt onto the social platforms as a new medium to improve and modify upon the existing method of psychological diagnose on personal mental well-being.

References

[1] American Psychiatric Association. (2022). Introduction. In Diagnostic and Statistical Manual of Mental Disorders fifth edition text revision: DSM-5-TR (5th ed.). essay, American Psychiatric Association Publishing.

[2] Wakefield J. C. (2015). DSM-5, psychiatric epidemiology and the false positives problem. Epidemiology and psychiatric sciences, 24(3), 188–196.

[3] Bashir, Hilal & Bhat, Shabir. (2016). Effects of Social Media on Mental Health: A Review. The International Journal of Indian Psychology. 4. 125 - 131. 10.25215/0403.134.

[4] Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. WIREs Data Mining and Knowledge Discovery, 8(4). https://doi.org/10.1002/widm.1253

[5] Wang, Y., Guo, J., Yuan, C., & Li, B. (2022). Sentiment Analysis of Twitter Data. Applied Sciences, 12(22), 11775. MDPI AG. Retrieved from http://dx.doi.org/10.3390/app122211775

[6] Almatarneh, S., & Gamallo, P. (2018). A lexicon based method to search for extreme opinions. PLoS One, 13(5)

[7] Snyder, R.M. (2015). An Introduction to Topic Modeling as an Unsupervised Machine Learning Way to Organize Text Information.

[8] Maupomé, D., & Meurs, M. (2018). Using Topic Extraction on Social Media Content for the Early Detection of Depression. Conference and Labs of the Evaluation Forum.

[9] Resnik, Ps & Armstrong, William & Claudino, Leonardo & Nguyen, Thang & Nguyen, Viet-An & Boyd-Graber, Jordan. (2015). Beyond LDA: Exploring Supervised Topic Modeling for Depression-Related Language in Twitter. 99-107. 10.3115/v1/W15-1212.

[10] Joshi, A., Tripathi, V., Patel, K., Bhattacharyya, P., & Carman, M. (2016). Are word embedding-based features useful for sarcasm detection? Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. https://doi.org/10.18653/v1/d16-1104

[11] Drus, Z., & Khalid, H. (2019). Sentiment Analysis in social media and its application: Systematic Literature Review. Procedia Computer Science, 161, 707–714.