

A STUDY ON USE OF NLP IN SENTIMENT AND MENTAL ILLNESS DETECTION

Ms.Shraddha Doshi

Assistant Professor - Smt.C.Z.M.G.BCA College - Jamnagar.

Email: Sadoshi84@gmail.com

Abstract

In today's world people are used to communicating using different social media platforms and with some digital mediums. They share their memories, events, views, thoughts, opinion and sentiments. In this paper we have reviewed different researches based on natural language processing to analyze sentiments and mental health. Natural language processing with machine learning techniques gives good results in identifying mental illness like depression, anxiety and sentiments through posts of users over micro blogging sites. This study has also found that improved Lexicon dictionaries, word embedding and stop word removal also needed while processing data for research. NLP and machine learning techniques like SVM, naive bayes, logical regression can automate sentiment and mental illness identification by analyzing social media communication of users.

Keywords:

Natural language processing, machine learning, sentiment analysis, mental illness

Introduction

Natural language processing – NLP is defined as an automatic manipulation of natural language like text and speech it is also known as automatic computational processing of human languages

Natural language processing has given good outcomes in analyzing sentiments, mood, and views of people from their text. Medical field has a great hope on NLP and machine learning in diagnosis of mental health. 'joy', 'sadness', 'fear', 'anger', or 'surprise' like moods are expressed in users posts. (Nigam et al., 2018)

NLP is a useful technique to find suicidal thoughts, feeling of worthlessness, loss of interest and depressed mood (Deshpande & Rao, 2018)

Natural language process generally provides some features to work on users' text based posts like text processing and tokenization, semantic analysis, lexical analysis and syntactic analysis. (Rajput, 2019)

Suicidal Ideation Detection is possible using natural language processing and applications of machine learning by analyzing users text based communication over the social media platforms.(Ji et al., 2021)

Related review

M. kanakaraj has used NLP to improve sentiment based classification.(Kanakaraj & Guddeti, 2015) For these purposes, sentiment feature vector and beg of word are used to improve accuracy in outcomes. In this study it is found that various machine learning algorithms like naive bayes, SVM and maximum entropy are used mostly and Word Sense Disambiguation is also useful. To increase accuracy of classification, NLP and WSD are used.

This paper applies methodology that first gathers data from twitter social media, data then after cleaned by some techniques. From training set of data, feature vectors are extracted and then different machine learning classifiers are applied to get classification. Python 2.7.3 with NLTK library for NLP Techniques are also used in this study.

Results from this research shows that by using synsets 3% to 6% improvement can be achieved. Author has suggested future work may focus on minimizing cost and enhancing methods at paragraph level for this kind of research.

A.calvo hse used NLP to develop common language which draws together the NLP, HCI and mental health to give psychological assistance. Author has found that NLP can be used to create marketing applications based on customers' reviews and sentiments; it is also used to develop AI Applications and can be extended to design mental health applications.

This research has proposed models based on Data, labeling and interventions. They used NLP to analyze data and interventions. Author has reviewed that different kinds of data can be used for designing such mental health applications using NLP(Calvo et al., 2017). They reviewed that twitter, facebook, suicidal notes; blogs can be used for this purpose
Author has suggested to improve ethical implications of being able to identify people in need, also that this information can be relayed sensitively and ethically back to participants.

Li-Chen Cheng proposes a deep learning framework using NLP to extract sentiments from various social media.(Cheng & Tsai, 2019) This research also focuses upon opinion mining advanced to traditional NLP which finds positive, negative and amplitude in message. NLP based deep learning framework first crawls through the different web platforms to extract reviews. Then review is preprocessed at second stage.at the third step, sentiment analytics is applied using Microsoft analytics API, NLTK text Blob. Data also labeled in this stage of the process. Word embedding is performed at this stage using word2vc and GloVe. At the last step, a deep learning model LSTM, BiLSTM and GRU are applied on the data. This study found that BiLSM gives superior results comparing with LSTM and GRU.

Using twitter data pre-processed data framework based on NLP to filter tweets is proposed by Md.Rakibul Hasan.(Hasan et al., 2019)this model is based on BoW and TF-IDF based for the classification of positive and negative tweets. This study produced accuracy of 82.25% in sentiment analysis.

K. Katchapakirin has used Posts of Facebook users from Thai are used to identify depression.(Katchapakirin et al., 2018) This research employs NLP to generate algorithms for depression detection. This research uses TMHQ mental health questionnaire and extracted data from micro blogging sites. SVM with Weak and random forest techniques with rapid miners used to

further classifications. Data are separated in two groups one with positive polarity and second with negative polarity

This research proves that people who post negative polarity with privacy only me are more depressed than people who share memory every one and from morning to noon time.

Author has studied the opinion type in a predetermined text using tokenization sentiment and elimination of stop words. Lexicon dictionary is used in this research with natural language process. (Kusrini & Mashuri, 2019) In this research, the first web crawler extracts data from twitter. Text processing is performed next to clean data and to make data meaningful for further research work. Use of lexicon and evaluator is the last step to analyze sentiments. Naive Bayes classifier gave better results amongst all others. It is also suggested that lexicon dictionary can be improved to get more accuracy.

Sanjana Mendu has developed a generalized framework which extracts features from digital text related to mental health. (Mendu et al., 2019) This research also creates a bridge between qualitative and quantitative research. Frame work in this research first grabs data from social media and separates it based on time of posts. Later on it finds whether the post is incoming, outgoing or bilateral. Role relation of the actor is also identified in the next step. Symentical, lexical, temporal and topological feature based clusters are developed and then analyzed.

Wei Yen Chong has done experiments to relate sentiment lexicon and subject grammatically. Proposed system in this research detects sentiments on particular subject using natural processing language. (Chong et al., 2014)

For this research 1513 tweets were extracted from twitter. First data is processed to make it useful for the research and then sentiment classification done using NLP. Results are compared with alchemy API proving this system has more accuracy over alchemy API. Raw data and processed data are also

classified using naive Bayes and SVM. SVM proved more efficiency over naive Bayes.

Using NLP, Author Sandeep Nigam has shown taxonomy of various sentiment analysis methods and proved that logistic regression has higher accuracy while comparing it with other techniques. (Nigam et al., 2018) This research is based on twitter data and supervised machine learning approach. Proposed framework first cleans data by removing HTML decoding, URL, symbols etc. Then visualize words in negative and positive occurrence. Again it visualizes words in negative and positive occurrence but after removing stop words. After extracting features, different machine learning techniques are applied to classify and analyze data and results.

Methodology

Research to identify mental illness needs much more data which collected data from digital platforms and micro blogging sites and Data cleaning and feature extracting tasks performed on that data. In various researches, they also used LEXICON, LIWC, Semantic and Syntactic Analysis to improve results.

NLP applied to identify suicidal thoughts, depressive terms and some habits which depict mental illness from the users' posts on social media in such research.

Generally negative or positive polarity of words used in users' messages and replies, frequency of negative words in posted text and time of post is analyzed to find mental health of the users.

Naive Bayes classifiers, SVM and random forest classifiers are used to compare results on filtered train and test data sets in most of the experiments which we reviewed. Some have also used logistic regression techniques.

Data

Users' data from their social media sites and micro blogging sites are used in such research. Some researchers have collected twitter users' tweets as

research data. Twitter also provides data for research purpose in a way that identity does not reveal.

Kaggle and other online sources also provide user sentiment data of tweets like reviews on movies, opinion on elections, leaders and parties, and customer review on ecommerce sites.

Results

In the review we did, we found that naive Bayes and SVM have more improved results in comparison with other techniques like random forest and others. Use of improved LEXICON has also given improved results.

Limitations

Only text based data is observed for such research but now a days there are lot of searches over internet are based on voice search. And there are lots of post containing voice and video.

Text based analysis cannot cover tone of words and cannot analyze facial expression while conveying message linguistic

Conclusion and future scope

In recent research NLP has proved efficiency in text analysis and sentiment analysis. NLP is used with AI to make AI powerful in natural language processing. NLP has proved its usefulness in mood detection, anxiety detection, depression detection and other mental illness detection. Natural language processing is used in analyzing negative or positive polarity of words in text. Natural language processing is promising to the medical field in mental health diagnosis.

Future scope of NLP is that it has capability to improve AI and also useful to automate the detection of mental illness from non-medical text and messages of users

References

- Calvo, R. A., Milne, D. N., Hussain, M. S., & Christensen, H. (2017). Natural language processing in mental health applications using non-clinical texts. *Natural Language Engineering*, 23(5), 649–685. <https://doi.org/10.1017/S1351324916000383>
- Cheng, L. C., & Tsai, S. L. (2019). Deep learning for automated sentiment analysis of social media. *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2019*, 1001–1004. <https://doi.org/10.1145/3341161.3344821>
- Chong, W. Y., Selvaretnam, B., & Soon, L. K. (2014). Natural Language Processing for Sentiment Analysis: An Exploratory Analysis on Tweets. *Proceedings - 2014 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology, ICAIET 2014*, 212–217. <https://doi.org/10.1109/ICAIET.2014.43>
- Deshpande, M., & Rao, V. (2018). Depression detection using emotion artificial intelligence. *Proceedings of the International Conference on Intelligent Sustainable Systems, ICISS 2017, Iciss*, 858–862. <https://doi.org/10.1109/ISS1.2017.8389299>
- Hasan, M. R., Maliha, M., & Arifuzzaman, M. (2019). Sentiment Analysis with NLP on Twitter Data. *5th International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering, IC4ME2 2019*, 1–4. <https://doi.org/10.1109/IC4ME247184.2019.9036670>
- Ji, S., Pan, S., Li, X., Cambria, E., Long, G., & Huang, Z. (2021). Suicidal Ideation Detection: A Review of Machine Learning Methods and Applications. *IEEE Transactions on Computational Social Systems*, 8(1), 214–226. <https://doi.org/10.1109/TCSS.2020.3021467>
- Kanakaraj, M., & Guddeti, R. M. R. (2015). NLP based sentiment analysis on Twitter data using ensemble classifiers. *2015 3rd International Conference on Signal Processing, Communication and Networking, ICSCN 2015*. <https://doi.org/10.1109/ICSCN.2015.7219856>
- Katchapakirin, K., Wongpatikaseree, K., Yomaboot, P., & Kaewpitakkun, Y. (2018). Facebook Social Media for Depression Detection in the Thai Community. *Proceeding of 2018 15th International Joint Conference on Computer Science and Software Engineering, JCSSE 2018*, 1–6.

<https://doi.org/10.1109/JCSSE.2018.8457362>

Kusrini, & Mashuri, M. (2019). Sentiment analysis in twitter using lexicon based and polarity multiplication. *Proceeding - 2019 International Conference of Artificial Intelligence and Information Technology, ICAIT 2019*, 365–368.

<https://doi.org/10.1109/ICAIT.2019.8834477>

Mendu, S., Boukhechba, M., Baglione, A., Bae, S., Wu, C., & Barnes, L. (2019). SocialText: A Framework for Understanding the Relationship between Digital Communication Patterns and Mental Health. *Proceedings - 13th IEEE International Conference on Semantic Computing, ICSC 2019*, 428–433.

<https://doi.org/10.1109/ICOSC.2019.8665567>

Nigam, S., Das, A. K., & Chandra, R. (2018). Machine Learning Based Approach to Sentiment Analysis. *Proceedings - IEEE 2018 International Conference on Advances in Computing, Communication Control and Networking, ICACCCN 2018*, 157–161. <https://doi.org/10.1109/ICACCCN.2018.8748848>

Rajput, A. (2019). Natural language processing, sentiment analysis, and clinical analytics. *Innovation in Health Informatics: A Smart Healthcare Primer*, 79–97.

<https://doi.org/10.1016/B978-0-12-819043-2.00003-4>