Natural Language Processing Applied to Mental Health

Emily Lopez
University of California, Berkeley
emilysarahi@berkeley.edu

Abstract

Over the past years, we have seen an increase in the use of NLP techniques within mental health research. Through NLP tasks, researchers can incorporate the real-world complexity of a patient's life through methods like deep learning, information extraction, sentiment analysis, emotion detection, and mental health surveillance. The datasets that have been widely used to accomplish this primarily include medical records and social media text. Research in this domain is making strides in identifying mental illnesses in patients earlier and improving treatment through personalized care. This survey discusses how these novel advancements in NLP are revolutionizing mental health care.

1 Introduction

A survey conducted by the Anxiety Disorders Association of America in 2006 highlights that 40% of participants reported experiencing "persistent stress or excessive anxiety in their daily lives," but only 9% of them reported having been diagnosed with an anxiety disorder (ADAA, 2021). This mental disorder is one of many that affect tens of millions of people each year (ADAA, 2021), but why has a stressful work-life relationship become the norm? People are not receiving appropriate help. Although therapy is used to diagnose and treat many mental health conditions, the effectiveness of treatment varies considerably depending on multiple factors.

Natural Language Processing (NLP) presents new opportunities to incorporate real-world complexity in mental health analysis through methods like deep-learning, information extraction, sentiment analysis, emotion detection, and mental health surveillance (Zhang et al., 2022). There has been an upward trend in NLP-driven mental illness detection research and deep learning-based methods have particularly gained popularity in the last couple of years (Dai et al., 2021; Rojas-Barahona et al., 2018; Zhang et al., 2022). Researchers and professionals believe that NLP techniques may be promising tools for identifying mental illnesses earlier in patients' lives and for properly treating their diagnoses through personalized care.

When conducting NLP research within the mental health field, the corpora that is used consists of clinical and non-clinical texts, particularly medical records and social media data (Harvey et al. 2022; Le Glaz, 2021). We will consider the different methodologies used with both forms of data as well as the means of evaluation for these methods.

1 NLP in Mental Health Using Non-Clinical Data

Currently, methods for assessing mental health at a population level are costly and require the collection of large samples of data through measures like surveys. Therefore, they are slow to reflect the present-day social conditions which are rapidly changing (Fine et al., 2020). Using publicly available social media data, such as data from Twitter, Reddit, etc., can provide real-time estimates of psychological distress in the population (Fine et al., 2020). Due to its ubiquitous nature, online health communities and social media outlets have become effective and popular means of information exchange and social support, particularly for individuals who may face stigmas related to mental health (Kulkarni, 2021). Social media text also includes implicit information about the author which is

Word Count: 2197

useful when carrying out NLP tasks (Benton et al., 2017). Digital life data—data from the interactions a person has with their digital devices—collected with consent may be beneficial for identifying mental health conditions and associated risks earlier (Coppersmith et al., 2018). Using this data may also help in surveillance efforts and provide decision-makers with helpful information about the mental state of the population (Skaik and Inkpen, 2020).

1.1 Twitter Data

One of the sources of data that has been widely used is Twitter. However, Twitter normally does not allow the texts of downloaded tweets to be shared publicly, only tweet identifiers (Ive et al., 2020). However, Twitter is still the most widely used data source since the collection of public data is relatively easy (Calvo et al., 2017).

Twitter data has been particularly used to find correlations between negative-emotion language and mental health illnesses or suicide risks (Conway and O'Connor, 2016). In their research, Fine et al. (2020) demonstrated that NLP applied to US-based Twitter data from healthcare providers was useful in providing real-time estimates of psychological distress in the population. They used classification models to score Tweets in the sample with an estimate of the probability that a Tweet was written by an individual with anxiety, depression or who had attempted suicide. Their plausibility and validity were confirmed using human annotators with clinical training. Logistic regression with character n-gram features was trained on three separate samples corresponding to anxiety, depression and suicide to distinguish users with a self-stated mental health diagnosis from control users reporting no diagnoses. They then employed the anxiety and depression models from Coppersmith et al. (2015) and the suicide model from Coppersmith et al. (2018). Their results show that baseline scores for each mental health variable were higher/more severe for healthcare providers than for the community population.

Benton et al. (2017) also worked with Twitter user datasets, however, they worked to estimate mental health conditions using a deep-learning framework. Their team noticed that studies in this domain typically model social media author's characteristics in isolation, which does not address coinciding influence factors. NLP tasks with underlying commonalities like part-of-speech tagging and parsing have been shown to benefit from multi-task learning (MTL) since the learning implicitly leverages interactions between them (Benton et al., 2017). By modeling multiple conditions, the system learns to make predictions about mental health and suicide risk at a low false-positive rate. In their research, conditions are modeled as tasks in an MTL framework with gender prediction as an additional auxiliary task since modeling gender has been shown to improve accuracy in tasks using social media text (Benton et al., 2017).

1.2 Reddit Data

Another large data source that is used in this domain is Reddit user data. The main difference between Reddit and other social media sources is that posts are grouped into different subreddits based on topics (Zhang et al., 2022). Unlike Twitter, Reddit's open policy allows its datasets to be publicly available (Zhang et al., 2022).

In his research, Wolohan (2020), used Reddit data and state-of-science depression prediction models to quantify the impact that COVID-19 had on population depression. His research shows the effectiveness of a deep long short-term memory (LSTM) neural network with fastText embeddings—word embeddings—at predicting user-level depression and population depression considering the pandemic.

Biester et al. (2020) specifically examine discussions from mental health subreddits to understand how activity in these forums has changed during the pandemic. They do so by creating a time series for different metrics that may be affected by the pandemic and using time series intervention analysis techniques to determine if there are significant changes in their metrics during the pandemic. They also analyze how COVID-19 has influenced language and topics of discussion.

Kulkarni et al. (2021) also make use of subreddits. Their team noticed that contemporary research in the field focuses on mental illness prediction or classification models (Harvey et al., 2022; Zhang, 2022); thus, they draw their attention to the identification of discussion clusters by introducing contextualized word representations for topic and theme extraction from subreddits. Mining information from these clusters can provide a lens over the main discussion themes, discourse anatomy, and

dialogue structure in these online forums; mining information can also increase understanding of the current climate regarding mental health (Kulkarni et al., 2021). Their team utilized a novel data representation technique called "topic-infused deep contextualized representations" which combines deep contextual embeddings with topic information for generating robust document representations; they demonstrated how this method performed better for text clustering when compared to contextual embeddings generated by the pretrained RoBERTa model (Kulkarni et al., 2021)—the RoBERTA model is a robustly optimized BERT approach that trains for longer, on more data, with bigger batches and on longer sequences (Spruit, 2022).

2 NLP in Mental Health Using Clinical Data

Aside from social media datasets, various forms of clinical data have been used to implement NLP tasks within the mental health domain. The most common medical corpora are transcribed patient interviews and records, like electronic health records (EHRs), Psychological Evaluation Reports, and Coroner Reports (Le Glaz, 2021). The most frequently used data source EHRs is divided into two formats, structured and unstructured information (Wang and Preininger, 2019) —structured information uses existing lexicons while unstructured information refers to free-text documents. Research using clinical documentation has been particularly challenging since it relies heavily on free text that is difficult to de-identify completely (Ive et al., 2020). However, Ive at al. (2020) have tackled this problem by generating artificial medical data.

Additionally, the rise of online and text-message-based therapy services has generated new sources of information and there has been increasing NLP research that makes use of this textual data (Althoff, 2016; D'Alfonso, 2020; Sharma 2020). Lastly, the application of chatbots and conversational agents in mental health has shown to be useful for making diagnoses, classifying mental states, promoting health education, and providing emotional support (Madeira et al., 2020). Their increasing use has resulted in a new source of data that is helpful for mental health NLP research (Madeira et al., 2020).

2.1 EHRs, Clinical Reports, Records, and Interviews

Although data in EHRs can be useful for retrospective clinical studies, much of this data is stored as unstructured text which cannot be directly used in computation (Viani et al., 2018). However, NLP methods can be useful in extracting this data to identify symptoms and treatments. Viani et al. (2018) are specifically developing an EHR corpus annotated with time expressions, clinical entities, and their relations to be used for NLP development. They notice that relevant temporal information in mental health records, such as symptom onset, is not always well represented by current temporal models, so they developed a novel gold standard corpus—manually annotated collection of text—, compared it to other related corpora in terms of content and time expression prevalence, and adapted two NLP models for extracting time expressions. To assess the quality of their corpus, they calculated inter-annotator agreement (IAA) for each annotated batch using average precision and recall and F1 score (Viani et al., 2018).

Mukherjee et al. (2020) also address the challenges and opportunities of unstructured EHR data by transforming status assessment data generated by physicians into binary vector representations. These vectors represent patients' symptoms, functional states, and emotional states and are in a structured and quantifiable format that allows for intra- and interpatient quantifications. To assess the generalization of their model, they calculate accuracy, precision, recall, F1 score and AUROC.

Psychotherapy note text has also been used to evaluate whether NLP of this data provides additional accuracy over currently used prediction models (Levis et al., 2021). Notes in Levis et al.'s (2021) study were assessed using Sentiment Analysis and Cognition Engine, a Python-based NLP package. The output was evaluated using machine-learning algorithms and the area under curve (AUC) was calculated to determine the models' predictive accuracy. Their findings suggest that NLP derived variables from psychotherapy notes in fact offer additional predictive value over current state-of-the-art prediction models (Levis et al., 2021).

2.2 Online Mediated Counseling

Recent advances in technology and increased demand for counseling services have resulted in increased large-scale data on online-mediated

counseling services (Althoff et al., 2016). Althoff et al. (2016) particularly worked with data from SMS texting-based crisis counseling services to develop a set of novel computational discourse analysis methods to measure how different linguistic aspects of conversations are related to conversation outcomes. Among the methods they developed was a novel psycholinguistics-inspired word frequency analysis approach which measured perspective change; it demonstrated how perspective change results in better counseling conversation outcomes. They evaluated their model with 10-fold-cross validation and compared models using the area under the ROC curve.

Sharma et al. 2020 also used text messaging-based counseling data to understanding how empathy is expressed in online mental health platforms. They developed a multi-task RoBERTa-based bi-encoder model for identifying empathy in conversations and for extracting rationale underlying its predictions (Sharma et al., 2020). Using manual evaluation methods to check for accuracy, their work demonstrates that their approach can effectively classify empathic conversations.

2.3 Chatbots and Conversational Agents

Another NLP application of clinical data is seen in chatbots. Chatbots are programs that mimic conversations with users via a chat interface. The history of chatbots is closely linked with psychology, as seen by the first well-established chatbot ELIZA in 1966 which was programmed to simulate a known psychotherapist (Conway and O'Connor., 2016; D'Alfonso., 2020).

Chatbots are more likely to advance health and behavioral change if integrated with human support (Madeira et al., 2020). Madeira et al. (2020) found that this intersection has been overlooked and worked to create an open-source framework to assist chat operators of mental health services. Their framework implements a classification model to classify messages from chat users. These classifications are then used to provide a list of suggestions for counseling session topics to the chat operator. They evaluate their framework using reports from users and counselors.

Additionally, like Wolohan (2020), Rojas-Barahona et al. (2018) identified that deep-learning models combined with word embeddings significantly outperform non-deep learning models in mental health-related NLP tasks. Rojas-Barahona et al. (2018) however applied this approach to define a mental health ontology—a central element of a dialogue system that defines the concepts this system can understand and talk about—based on Cognitive Behavioral Therapy (CBT) principles. They then annotated a large corpus where this relationship is exhibited and performed understanding using deep-learning and distributed representations; they evaluated their models using inter-annotator agreement. They anticipate that this understanding module will be a key feature of statistical dialogue systems delivering therapy.

2 Conclusion

The application of NLP techniques in mental health has been growing over the past decades. NLP tasks allow researchers to incorporate the complexity of individuals' lives through methods like deep-learning, information extraction, sentiment analysis, emotion detection, and mental health surveillance (Zhang et al., 2022). The datasets that have been primarily used in this research are medical records and social media datasets (Harvey et al., 2022; Le Glaz, 2021). We have seen how NLP tools applied to these data sources may be beneficial in identifying mental illnesses in patients earlier and for improving treatment through personalized care.

References

ADAA. 2021. *Facts & Statistics | Anxiety and Depression Association of America, ADAA.* https://adaa.org/understanding-anxiety/facts-statistics.

ADAA. 2021. *Highlights: Workplace Stress & Anxiety Disorders Survey | Anxiety and Depression Association of America, ADAA.* https://adaa.org/workplace-stress-anxiety-disorders-survey.

Adrian Benton, Margaret Mitchell, and Dirk Hovy. 2017. Multitask Learning for Mental Health Conditions with Limited Social Media Data. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 152–162, Valencia, Spain. Association for Computational Linguistics.

Alex Fine, Patrick Crutchley, Jenny Blase, Joshua Carroll, and Glen Coppersmith. 2020. Assessing population-level symptoms of anxiety, depression,

and suicide risk in real time using NLP applied to social media data. In *Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science*, pages 50–54, Online. Association for Computational Linguistics.

Ashish Sharma, Adam Miner, David Atkins, and Tim Althoff. 2020. A Computational Approach to Understanding Empathy Expressed in Text-Based Mental Health Support. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5263–5276, Online. Association for Computational Linguistics.

Atharva Kulkarni, Amey Hengle, Pradnya Kulkarni, and Manisha Marathe. 2021. Cluster Analysis of Online Mental Health Discourse using Topic-Infused Deep Contextualized Representations. In *Proceedings of the 12th International Workshop on Health Text Mining and Information Analysis*, pages 83–93, online. Association for Computational Linguistics.

Calvo, R., Milne, D., Hussain, M., & Christensen, H. 2017. Natural language processing in mental health applications using non-clinical texts. *Natural Language Engineering, 23*(5), 649-685. doi:10.1017/S1351324916000383

Dai H-J, Su C-H, Lee Y-Q, Zhang Y-C, Wang C-K, Kuo C-J and Wu C-S. 2021. Deep Learning-Based Natural Language Processing for Screening Psychiatric Patients. *Front. Psychiatry* 11:533949. doi: 10.3389/fpsyt.2020.533949

Glen Coppersmith, Mark Dredze, Craig Harman, and Kristy Hollingshead. 2015. From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. In Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, Denver, Colorado, USA. North American Chapter of the Association for Computational Linguistics.

Glen Coppersmith, Ryan Leary, Patrick Crutchley, and Alex Fine. 2018. Natural language processing of social media as screening for suicide risk. Biomedical informatics insights, 10:1178222618792860.

Harvey D, Lobban F, Rayson P, Warner A, Jones S Natural Language Processing Methods and Bipolar Disorder: Scoping Review. 2022. JMIR Ment Health 2022;9(4):e35928. URL: https://mental.jmir.org/2022/4/e35928. DOI: 10.2196/35928

Ive, J., Viani, N., Kam, J. *et al.* 2020. Generation and evaluation of artificial mental health records for Natural Language Processing. *npj Digit. Med.* 3, 69. https://doi.org/10.1038/s41746-020-0267-x

JT Wolohan. 2020. Estimating the effect of COVID-19 on mental health: Linguistic indicators of depression during a global pandemic. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, Online. Association for Computational Linguistics.

Laura Biester, Katie Matton, Janarthanan Rajendran, Emily Mower Provost, and Rada Mihalcea. 2020. Quantifying the Effects of COVID-19 on Mental Health Support Forums. In *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*, Online. Association for Computational Linguistics.

Le Glaz A, Haralambous Y, Kim-Dufor D, Lenca P, Billot R, Ryan T, Marsh J, DeVylder J, Walter M, Berrouiguet S, Lemey C. 2021. Machine Learning and Natural Language Processing in Mental Health: Systematic Review. J Med Internet Res 2021;23(5):e15708. URL: https://www.jmir.org/2021/5/e15708. DOI: 10.2196/15708

Levis, M., Leonard Westgate, C., Gui, J., Watts, B., & Shiner, B. 2021. Natural language processing of clinical mental health notes may add predictive value to existing suicide risk models. *Psychological Medicine, 51*(8), 1382-1391. doi:10.1017/S0033291720000173

Lina M. Rojas-Barahona, Bo-Hsiang Tseng, Yinpei Dai, Clare Mansfield, Osman Ramadan, Stefan Ultes, Michael Crawford, and Milica Gašić. 2018. Deep learning for language understanding of mental health concepts derived from Cognitive Behavioural Therapy. In *Proceedings of the Ninth International Workshop on Health Text Mining and Information Analysis*, pages 44–54, Brussels, Belgium. Association for Computational Linguistics.

Mike Conway, Daniel O'Connor. 2016. Social media, big data, and mental health: current advances and ethical implications, Current Opinion in Psychology, Volume 9, Pages 77-82, ISSN 2352-250X, https://doi.org/10.1016/j.copsyc.2016.01.004. https://www.sciencedirect.com/science/article/pii/S2352250X16000063

Mukherjee, S. S., Yu, J., Won, Y., McClay, M. J., Wang, L., Rush, A. J., & Sarkar, J. 2020. Natural Language Processing-Based Quantification of the Mental State of Psychiatric Patients. *Computational Psychiatry*, *4*, 76–106. DOI: http://doi.org/10.1162/cpsy_a_00030

Natalia Viani, Lucia Yin, Joyce Kam, Ayunni Alawi, André Bittar, Rina Dutta, Rashmi Patel, Robert Stewart, and Sumithra Velupillai. 2018. Time Expressions in Mental Health Records for Symptom

Onset Extraction. In *Proceedings of the Ninth International Workshop on Health Text Mining and Information Analysis*, pages 183–192, Brussels, Belgium. Association for Computational Linguistics.

Ruba Skaik and Diana Inkpen. 2020. Using Social Media for Mental Health Surveillance: A Review. ACM Comput. Surv. 53, 6, Article 129 (December 2020), 31 pages. https://doi.org/10.1145/3422824

Simon D'Alfonso. 2020. AI in mental health, Current Opinion in Psychology, Volume 36, Pages 112-117, ISSN 2352-250X, https://doi.org/10.1016/j.copsyc.2020.04.005. https://www.sciencedirect.com/science/article/pii/S2352250X2030049X

Spruit, M.; Verkleij, S.; de Schepper, K. 2022. Scheepers, F. Exploring Language Markers of Mental Health in Psychiatric Stories. Appl. Sci. https://doi.org/10.3390/ app12042179

Thiago Madeira, Heder Bernardino, Jairo Francisco De Souza, Henrique Gomide, Nathália Munck Machado, Bruno Marcos Pinheiro da Silva, and Alexandre Vieira Pereira Pacelli. 2020. A Framework to Assist Chat Operators of Mental Healthcare Services. In *Proceedings of Second Workshop for NLP Open Source Software (NLP-OSS)*, pages 1–7, Online. Association for Computational Linguistics.

Tim Althoff, Kevin Clark, and Jure Leskovec. 2016. Large-scale Analysis of Counseling Conversations: An Application of Natural Language Processing to Mental Health. *Transactions of the Association for Computational Linguistics*, 4:463–476.

Wang, Fei, and Anita Preininger. 2019. "Ai in Health: State of the Art, Challenges, and Future Directions." *Yearbook of Medical Informatics*, vol. 28, no. 01, pp. 016–026., https://doi.org/10.1055/s-0039-1677908.

Zhang, T., Schoene, A.M., Ji, S. *et al.* 2022. Natural language processing applied to mental illness detection: a narrative review. *npj Digit. Med.* 5, 46. https://doi.org/10.1038/s41746-022-00589-7