

“What’s on your Mind?” : Using Social Media to predict Mental Health

Khanh Xuan Nguyen, Andrew Pachulski, Darshan Pandit, Jinfeng Rao, Yogarshi Vyas

Mental health is one of the largest problems facing the United States today. The National Alliance on Mental Illness (NAMI) highlights some gloomy statistics: one in four adults suffers from some form of mental health disorder; 60% of them do not receive proper diagnosis or care; and the spending on serious mental illness ends up costing Americans nearly \$200 billion in lost earnings each year [1]. The difficulty of diagnosing mental illness lies partly in the negative stereotypes associated with receiving treatment for such disorders. Furthermore, traditional approaches which rely on patients answering questionnaires are plagued with serious issues. It is very common for participants to provide biased or dishonest answers that may seem appealing to them. Or very often, the questionnaires are lengthy enough to make participants answer erratically and they only capture a participant’s emotions and feelings for a limited time.

To battle all these shortcomings, Professor Philip Resnik and his colleagues at the University of Maryland are working hard to understand how use of language correlates with mental disorders such as depression and Post-Traumatic Stress Disorder (PTSD) and develop computational algorithms to better understand this relationship. They believe that people suffering from these disorders are more likely to use languages very differently from the rest of the populace. Armed with English Tweets and data from other social networks, Resnik and his group have been working towards identifying these differences. Taking cues from subtle observations such as frequent use of a particular word class, they believe they can successfully identify symptoms of these disorders very early in an individual. Using a statistical technique called ‘Topic Modeling’, they are able to extract meaningful themes from the plethora of text that is generated by users online [2]. This technique relies on co-occurrence statistics of words to cluster them into groups that exhibit certain topical relations, called *hidden topics*. These topics can then be further investigated to identify traits of mental health disorders. Topic modeling is computationally efficient, low-cost and can be applied to practically any collection of texts such as essays [3] or journal entries. To demonstrate the usefulness of their technique, Prof. Resnik and colleagues published a result showing that adding a topic model on top of a traditional rule-based system improved its accuracy on predicting neuroticism - a trait that correlates consistently with depression.

Darshan Pandit	- dmp478@nyu.edu
Jinfeng Rao	- jinfeng@cs.umd.edu
Yogarshi Vyas	- yogarshi@cs.umd.edu