International Conference on Innovative Research in Engineering, Management and Sciences
International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 4 | Issue 9 | ISSN : 2456-3307

Mental Health Analysis using Natural Language Processing

Bindu V S*[1], Susmita N Gonkar[2], Celeste[3]

*[123]MTECH CSE Department, NHCE, Bangalore, India

ABSTRACT

Our increasingly digital life provides a wealth of data about our behavior, beliefs, mood, and well-being. This data provides some insight into the lives of patients outside the healthcare setting, and in aggregate can be insightful for the person's mental health and emotional crisis. Here, we introduce this community to some of the recent advancement in using natural language processing and machine learning to provide insight into mental health of both individuals and populations. We advocate using these linguistic signals as a supplement to those that are collected in the health care system, filling in some of the so-called "whitespace" between visits. Whitespace information provides a lens through which we can analyze psychological phenomena like emotional crisis, suicide attempts, and drug relapse.

Keywords : Natural Language Processing; Machine Learning, Mental Health, Whitespace; Psychology;

I. INTRODUCTION

Mental illness causes strain on human race at a very large scale physically, emotionally and financially. When Compared to other physical illness there is very less understanding about mental illness and ailments using natural language processing. The most important point of this issue is emotional crisis for example say the loss of sobriety (the state of being sober) or attempting suicide. These type of behaviors cause a high emotionally and financially troll and ill understood. Interactions with the health care systems will help to know more about emotional crisis and from where mental health comes from. As technology has taken a major part in our lives (by smart phones, smart devices, IOT devices etc.). We can easily analyze the behavior and mental illness of individual.

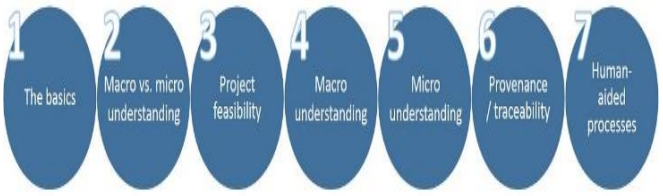
II. NATURAL LANGUAGE PROCESSING

The study of this paper highlight the work relevant to the personalization, idiosyncratic, prevention and scalable measure of the whitespace to the mentally affected community. The major cause of disability around the globe is depression. If depression is not treated at early stage it might lead to drug or alcohol addiction or other risks such as suicide. Rather then psychiatric specialist the person usually idealize his-self or might consult to the primary care physicians. Surveys say that depression has lead to more than the 91% people to spoil there self or commit suicide. With growth of social media large number data is shared by users voluntarily expressing their feelings, moods, and struggles and their daily problems with a mental health on various social media platform. This will helps other user to understand the communities. whenever any user post any such depressed thoughts it makes easy for studying the person behavior by

using natural language processing or unsupervised or supervised learning.

We collect the data from the various social media platforms and apply cross-sectional design to test the hypothesis that machine learning algorithms can classify suicide notes as well as or better than practicing mental health professionals and psychiatry physician trainees.

The 7 steps to extract information using NLP technique are as follows:



STEP 1: The Basics

The input to NLP will be a simple stream of Unicode characters (typically UTF-8). Then processing will be required to convert given character stream into a sequence of lexical items (as in words, phrases, and syntactic markers) which can then be used to better understand the content.

The basics include:

1. Structure extraction – To identify fields and blocks of content based on tagging. Identify and mark sentence, phrase, and paragraph boundaries

2. Language identification – This is to detect the human language for sentence or for each paragraph and for the entire document. Language detectors are critical to determine what linguistic algorithms and dictionaries to apply to the text.

3. Tokenization – To split character streams into tokens which is used for further processing and understanding. Tokens can be words, numbers, identifiers or punctuation (depending on the use case)

4. Acronym normalization and tagging – Acronyms can be specified as "I.B.M." or "IBM" so these should be tagged and normalized.

5. Lemmatization / Stemming – Reduces word variations to simpler forms that may help increase the coverage of NLP utilities.

6. De-compounding – For some languages (typically Germanic, Scandinavian, and Cyrillic languages), compound words are split into smaller parts so that we get accurate NLP.

7. Entity extraction – e.g. people, places, companies, etc.

1. Regex extraction – e.g. SSN, driver's licenses, etc.

2. Dictionary extraction – e.g. colors, units, sizes, employees, business groups, drug names, products, brands, and so on.

3. Complex pattern-based extraction – e.g. extract an item based on its context.

4. Statistical extraction – e.g. academic or journalistic text.

5. Phrase extraction – e.g. "Big Data" has a strong meaning which is independent of the words "big" and "data" when used separately.

STEP 2: Decide on Macro versus Micro Understanding

Macro Understanding – provides a general understanding of the document as a whole. Typically performed with statistical techniques. It is used for: clustering, categorization, similarity, topic analysis, word clouds, and summarization.

Micro Understanding – extracts understanding from each phrases or sentences. Typically performed with NLP techniques it is used for: extracting facts, entities (see above), entity relationships, actions, and metadata fields.

STEP 3: Project feasibility

Not all natural language understanding (NLP) projects are possible within a reasonable cost and time. After having done numerous NLP projects. Everything should be properly planned and understanding RAID is an important aspect.

STEP 4: Macro Understanding

To understand the below mention records we need a complete understanding of whole document.

1.Classifying / categorizing / organizing records

2.Clustering records

3.Extracting topics

4. General sentiment analysis

Record similarity, including finding similarities between different types of records. For instance:

1. Job descriptions to résumés / CVs
2. Keyword / key phrase extraction
3. Duplicate and near-duplicate detection
4. Summarization / key sentence extraction

STEP 5: Micro Understanding

Micro understanding is the extracting of individual entities, facts or relationships from the text.

There are three approaches that is used to perform extraction that provides micro understanding:

1. Top Down – Determine Part of Speech, then understand and extract the sentence into clauses, nouns, verbs, object and subject, modifying adjectives and adverbs, etc., then traverse this structure to identify structures of interest.

Advantages – This can handle complex, never-seen-before structures and patterns.

Disadvantages – It's hard to construct rules, brittle, often fails with variant input, and may still require substantial pattern matching even after parsing.

2. Bottoms Up – Create lots of patterns, then match the patterns to the text and extract the necessary facts. Patterns may be manually entered or may be computed using text mining.

Advantages – Easy to create patterns, it can be done by business users, does not require programming, easy to debug and fix, run, matches directly to desired outputs.

Disadvantages – Requires on-going pattern maintenance, cannot match on newly invented constructs.

3. Statistical – Similar to bottoms-up, but matches patterns against a statistically weighted database of patterns that is generated from tagged training data.

Advantages – Patterns are created automatically, built-in statistical trade-offs.

Disadvantages – requires generating extensive training, data has to be periodically retrained for best accuracy, cannot match on newly invented constructs, harder to debug.

STEP 6: Maintain Attribution

In accommodating the data from the internet and going through the content by extracting it involves several steps in it and it also has to pass across various stages. It is always more important to include tracing through the data for all the outputs generated so that the backtracking process also goes well through we can identify how the information is generated and where did it come from.

This usually involves:

1. Saving of the web pages which has the data concerned.
2. Saving the first and the last letter positions of the blocks of data from web site.
3. Savind the first and last letter positions of all elements, and the element id and the element type must be matched.
4. Even Identifying, normalization functions are put to the data content.

STEP 7: Human Supportive Process

Every process can't be done by itself untill human activity is involved in it. We always have to note that human intervention is need for following:

1. For generating or purging or picking lists of known elements.
2. For generating results accuracy
3. In discovering new outlines.
4. To estimate and correct results obtained
5. In creating training information

Most of all these processes can be frequent repeatedly. In a huge scale systems, we need to take human elements into picture and map it with the natural language processing system models. All the content obtained is been build as per the natural language processes and linked with the real time activities.

III. CONCLUSION

Various problems faced in the society like mental tension, drug addiction, distress, suicide attempts and many more can be removed by taking social media into centre. . Which can come accordance with the person's behaviour. This can be achieved by bring

natural language processing into picture. Here the various methods that can resolve the content into understandable format is given and through the past history and other technique its determined.

IV. REFERENCES

[1] Glen Coppersmith1, Casey Hilland1, Ophir Frieder2, Ryan Leary1, "Scalable Mental Health Analysis in the Clinical Whitespace via Natural Language Processing" 2017

[2] SAMHSA, "Substance abuse and mental health services administration," in Results from the 2013 National Survey on Drug Use and Health: Mental Health Findings, NSDUH Series H-49, HHS Publication No. (SMA) 14-4887, Rockville, MD, 2014.

[3] John pestian "Suicide note classification using Natural Language Processing: A Content Analysis"2011

[4] S. C. Curtin, M. Warner, and H. Hedegaard, "Increase in suicide in the united states, 1999-2014." NCHS data brief, no. 241, pp. 1–8, 2016.

[5] Glen Coppersmith1, Casey Hilland1, Ophir Frieder2, Ryan Leary1, "Scalable Mental Health Analysis in the Clinical Whitespace via Natural Language Processing" 2017

[6] SAMHSA, "Substance abuse and mental health services administration," in Results from the 2013 National Survey on Drug Use and Health: Mental Health Findings, NSDUH Series H-49, HHS Publication No. (SMA) 14-4887, Rockville, MD, 2014.

[7] John pestian "Suicide note classification using Natural Language Processing: A Content Analysis"2011

[8] S. C. Curtin, M. Warner, and H. Hedegaard, "Increase in suicide in the united states, 1999-2014." NCHS data brief, no. 241, pp. 1–8, 2016.

[9] J. W. Pennebaker, C. K. Chung, M. Ireland, A. Gonzales, and R. J. Booth, The development and psychometric properties of LIWC2007. Austin, TX: LIWC.net, 2007.

[10] C. Chung and J. Pennebaker, "The psychological functions of function words," Social Communication, 2007.

[11] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media," in Proceedings of the 7th International AAAI Conference on Weblogs and Social Media (ICWSM), 2013.

[12] G. Coppersmith, M. Dredze, C. Harman, and K. Hollingshead, "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, June 2015.