NLP applied Methods in Diagnosis of Mental Health Disorders

TA: Mohamed Gamal

Faculty of computer & artificial intelligence, Beni-Suef University

CS361: Medical informatics

Amany Adel

Esraa Mahmoud

Rofida Ehab

Nosiba Seif

Norhan Abdul Menaem

Sep 22, 2021

NLP applied Methods in Diagnosis of Mental Health Disorders

Summary

the purpose of this document is to summarize a piece of research papers methods and techniques associated to NLP role in diagnosis of mental health disorders.

Methods:

◆ A standardized questionnaire

like the CESD or the WHO-5 Well-Being Index (WHO-5), These questionnaires have basically the same schema:

1. Detecting depression symptoms (e.g., insomnia)
2. Determining the frequency of each symptom
3. Calculating a depression score by summing the frequencies of the symptoms
4. Using thresholds to classify the intensity of depression (none, mild, moderate, severe)

Lexicon generation:

1. It contained a hand-selected list of depression symptoms from conventional sources for psychiatrists.A good overall base list is assumed by merging the two major classification systems for depression,

   _the DSM-IV

   _the ICD-10

2. After the manual selection of symptoms from the above sources, the word base of each item was extended with the corresponding noun, verb and adjective whenever reasonable.
3. A synonym finder tool had been developed that parsed a publicly available web-based dictionary service: ***Thesaurus.com from Dictionary.com, LLC.*** It contained more than one million words, including derivatives and inflections.
4. the problem of ambiguous synonyms like homonyms (blue = color; mood) and antagonyms (bad = very good) as well as wrong or fuzzily interpreted synonyms had solved by creating individual blacklists.
5. the frequency gradations from ***the CES-D*** had adapted.

- Level 0: Rarely or none of the time
- Level 1: Some or a little of the time
- Level 2: Occasionally or a moderate amount of time
- Level 3: Most or all of the time

6. As a last step, the lexicon items were lemmatized. For example, the phrase this is depressing does not match the item depressed, but after lemmatization, this is depress does match depress.

◆ NLP processing

❖ To implement a grammar-oriented processing:

1. ***P reprocessing:*** removed unwanted text such as formatting, HTML links and quotations of earlier comments .
2. ***Boundary detection:*** using sentence detection by the Stanford Parser because it is a relatively reliable method in NLP.

3. ***Pronouns and negation:*** With the help of the part of speech (POS) tags generated by the Stanford Parser, various negation words were detected.
4. ***Matching:*** The phrases were matched with the lexicons. Each lemma in the phrase was compared with all lemmas in the three symptom lexicons (seed terms and first- and second degree synonyms) and the four frequency lexicons.

- ### a machine learning approach:

that uses a combination of text-based and ***TF-IDF*** based (term frequency–inverse document frequency) features is used to predict whether a message belongs to a positive or negative user. These features are fed to a ***SVM classifier***. If the classifier flags a message as positive, the user is classified as positive.

The model can be divided in three distinct stages:

1. Data p re-processing

1.1 To obtain clean text, contractions separated into their own words and removed hyperlinks, all punctuation characters, and decimal codes (words that started with # followed by numbers and ended with ;).

1.2 Tokens, stems and part-of-speech tags were obtained from clean text by first removing stop words, then tokenizing.

1.3 calculating part-of-speech tags and stemming.

1.4 Sliding window: For every new received message, the system combined its text with the previous *w* messages, where *w* is a configurable parameter, to form a window. The features were calculated on that . a size of 1 only uses the current message, while a size of “all” would use the whole user history.

1.5 User subsets: the first messages in the sequence from each user carry more information about the risk (or lack of risk) than the last ones. It’s tested by training and validating the model with only the first *m* messages of each user, where *m* was a configurable size. After testing different values for *m* (10, 100, 200, 500 and all messages), we found that using only the first 100 messages gave the best results.

2. Features

2.1 text-based

2.2 TF-IDF-based

2.1 Text features:

were normalized by the text length, and then discretized. The number of bins the features were discretized to was a parameter configurable to a *d* size.

• Grammar-based features: Table 2 shows the list of features.

• Special features: They were tailored to the self-harm dataset.

To choose the special features, previous work was done in analyzing the eRisk 2019 dataset. participation from the previous year shows details of this analysis. From this, the following features were developed:

1.1.1 Pronouns

There is evidence suggesting that people who use more first-person pronouns on average are more depressed than people who use the third person.We can track this difference by counting first-person pronouns.

2.1.2 Sentiment analysis

This feature shows the sentiment of the window as a numeric score normalized by the length of the texts. A negative score demonstrates a negative sentiment, while a positive score demonstrates a positive emotion.

2.1.3 NSSI words

using a word dictionary of terms related to non-suicidal self injury (NSSI words from now onwards). This dictionary is divided in several categories:

1. Methods of NSSI;

2. NSSI terms;
3. Instruments used;
4. Reasons for NSSI.

In this feature, we tracked the number of words from each category in a text, normalized by the length of that text. Each NSSI category became an independent feature.

2.2 TF-IDF features

A TF-IDF featurizer was trained on the positive users of the train data. This featurizer was then used to obtain TF-IDF features for each window. We obtain with this the TF-IDF-based features for each message window. This is then passed on to the classifier. We experimented with single word and n-gram based features, but word-based features worked best.

❖ popular VPA for interactive conversations:

NLP was implemented for processing the conversations between the person and the VPA, termed as ‘VPA-DR (virtual personal assistant for depression recognition).

Intent Recognition Phase:

1. In the proposed system, four categories of user intents were implemented for speech analysis using the machine learning model. These emotional states were: ‘Happy’, ‘Neutral’, ‘Depressive’, and ‘Suicidal’.
2. Dialog flow-ML returned a predicted value indicating the emotional/ sentimental state through the built-in natural language processing.
3. Dialog flow Agent identified depression in a user by matching against the predefined intents when the ML threshold for ‘Depressive’/ ‘Suicidal’ intent was detected.

Emotion Nurture Phase:

4. For the supportive role, VPA-DR, the Google Home mini would continue conversations for personal mental assistance in accordance with the emotional state being identified.
5. If the machine learning model identified or mapped ‘Suicidal’ state in a user, Dialog flow triggered a web hook service to Twilio API.
6. A client with the SID and access token was created to connect phone numbers to call or send text messages.
7. Twilio API would sent a text message to a local suicide prevention center, sending a distress signal that requested professional mental health assistance for the user.

❖ Predict Suicidal Ideation and Psychiatric Symptoms in a Text-Based Mental Health Intervention:

Natural language processing (NLP) and machine learning were used to predict suicidal ideation and heightened psychiatric symptoms among adults recently discharged from psychiatric inpatient or emergency room settings in Madrid, Spain. Outcome variables of interest were suicidal ideation and psychiatric symptoms (GHQ-12).

1. Study Sample:

Participation in the study was offered to all suicidal adults referred to the psychiatric ED meeting . Adults who attempted suicide were admitted to the general ED and evaluated by an ED psychiatrist who decided patients’ discharge or hospitalization.

2. Description of Texting Intervention:

2.1 The intervention was comprised of therapeutic reminders delivered by SMS messages that were sent out two days, seven days, 15 days, and monthly after hospital discharge.

2.2 Each text message also provided a link to a mobile application that contained a questionnaire eliciting responses related to the patients' sources of help.

2.3 Participants were also asked one unstructured, open-ended question related to their current mental state, "how are you feeling today?,"

2.4 and were encouraged to report on their progress since the hospitalization.

3. Outcome Variables of Interest

3.1 The first outcome of interest was suicidal ideation measured by the question "Have you felt that you do not have the will to live?" Six possible responses were (never), (sometimes), (less than half the time), (more than half the time), (most of the time), and (all of the time). This variable was dichotomized as suicidal ideation, *yes* (endorsed suicidal ideation at any point in the study), or *no* (never endorsed suicidal ideation).

3.2 The second outcome of interest is heightened psychiatric symptoms as measured by the General Health Questionnaire (GHQ-12).

3.3 Participants responded to items such as "Have you been able to enjoy your normal day to day activities?" and "Have you been feeling unhappy or depressed?" by providing answers such as "more than usual" or "much less than usual" on a 0–3 Likert scale. Total scores range from 0 to 12.

3.4 Other Covariates Taken from the Mobile Application Participants also used the cell phone application to report hours of sleep of the previous night, sleep quality of the previous night, appetite, anger/aggression, treatment adherence to medications, and the WHO-5 screening variable.

3.5 The WHO-5 is a five-item questionnaire measuring subjective well being derived from longer ratings scales used by the WHO in a multi center, multi country study in Europe . Participants' responses to questions were recorded using a sliding bar for each question.

3.6 The range for the sliding bar varied by measure: sleep quantity (the bar could be slid between 0 and 12 hours), sleep quality (bad, regular, or good), appetite (less, no change, or bigger appetite), conflicts or fights ("never" to "all of the time"), and medication treatment adherence ("never take medications" to "take all medications").

❖ Predicting Suicidal Ideation and Psychiatric Problems Using Free Text:

Software developed by Wired Informatics was used to conduct higher level semantic processing of the complete text of respondents' answers to the open-ended question, "How are you feeling today?" This software uses the clinical Text Analysis Knowledge Extract System (cTAKES) to generate NLP - based algorithms.

1. the "*n*-grams" feature was used, meaning that predictions were based upon a contiguous sequence of *n* words (rather than single words with no linguistic context).
2. A number of settings for *n*-gram size (i.e., number of words included in a contiguous string) were tested in order to identify the greatest positive predictive value (PPV) and to increase the number of true positives (increase specificity),

recognizing that there was a trade off between higher specificity and lower sensitivity.

3. *n*-grams were then codified and used as inputs to a machine learning (ML) algorithm to predict patients' probabilities of suicidal ideation or heightened psychiatric symptoms (GHQ-12 $\geq$ 4).

4. The machine learning program uses data from patients with and without these outcomes to train the model using a LIBLINEAR machine learning protocol.

Statistical Methods:

1. Multivariate logistic regression models were estimated on the two dependent variables (suicidality and GHQ-12 $\geq$ 4), conditional on age, sex, nightly sleep hours, sleep quality, anger, appetite, medication adherence, and the WHO-5 scale.

2. To calculate sensitivity, specificity, and PPV of structured data predictors, logistic regression models are repeated using half of our sample (the same randomly selected "training" dataset used for unstructured data)

3. then used the predicted probabilities from this logistic regression to characterize the remaining half of our sample as having suicidal ideation or heightened psychiatric symptoms (GHQ-12 $\geq$ 4) based on the values of their structured data predictors (covariates).

4. Statistical analyses and logistic regression-based predictions were performed using STATA 14 software.

❖ Natural language processing to extract symptoms of severe mental illness from clinical text:

Methods: Definitions of SMI symptoms

A keyword lexicon of SMI symptoms was defined by a team of psychiatrists, based on pragmatic criteria.

1. First, the potential salience of symptoms for research applications was considered, such as the Positive and Negative Symptoms Scale (PANSS) and Young Mania Rating Scale (YMRS) which were used as templates for guidance.
2. Second, the language used in routine clinical records was taken into consideration in choosing symptoms, focusing particularly on those which were likely to be recorded in the most consistent and tractable language, based on clinical experience.
3. Third, we sought a priori to extract sufficient numbers of symptom types to generate scales for further evaluation within the following five domains:
   (1) positive symptoms;
   (2) negative symptoms;
   (3) disorganisation symptoms;
   (4) manic symptoms;
   (5) catatonic symptoms.
4. The first four of these followed the findings of Demjaha, although we had not at this point attempted to define depressive symptoms.
5. Catatonic symptoms were further added to improve consistency with the study of Cuesta and Peralta, and as a symptom group of interest, which is often not adequately captured in dimensional studies because of its relative rarity in recruited clinical samples.
6. We defined the NLP task as a sentence classification problem, with a classifiable instance as a sentence containing a symptom keyword or the general constructs of 'negative symptoms' or 'catatonic syndrome, clinically relevant modifier terms were also defined for some concepts, in order to produce sub classifications of symptoms where appropriate .

Feature extraction:

Natural language processing (NLP) and its sub discipline of Information Extraction (IE) are commonly employed within clinical records to process large quantities of unstructured (human authored) text and return structured information about its meaning.

- A large number of tools and frameworks exist for general purpose information extraction from clinical dictionaries, such as cTAKES, NOBLE and MedLee.
- we introduce the CRIS-CODE project, which has the long-term objective of offering comprehensive NLP models for mental health constructs.
- The focus of the initial programme of work described here was to develop sentence classification models for a substantial range of SMI symptomatology, to allow automatic extraction for many of the most informative symptoms from the patient narrative.

❖ (NLP) techniques to make inferences about peoples' mental states from their text on social media:

These inferences can then be used to create online pathways to direct people to health information and assistance and also to generate personalized interventions.

1. In Data, we focus exclusively on textual data (as opposed to physiologicalsignals, activity, etc.) that has been analyzed for mental health applications.We also focus on texts that have been written by users (e.g. mostly consumers,occasionally patients) rather than doctors or researchers.
2. In Intervention, we considered studies that showed ways in which NLP could be used in computer interventions to support mental health.

2.1 Hypothetical uses of NLP in psycho-education are discussed but we excluded those psycho-education interventions where texts or multimedia are presented to clients without personalization or where they did not utilize NLP.

2.2 The area of affective computing and its literature on emotion detection from texts was also useful, particularly research. These literature reviews focus on the computational aspects and do not generally consider the differences between data sources or the interventions.

2.3 Furthermore, although we found several reviews of e-Health interventions, we have limited this discussion to include only those with NLP features, such as the one by Barak and Grohol.

3. In Automated Labeling, we focus on applications of NLP to the textual data collected in the previous section.

3.1 We searched for papers that used the different components of a classification system: feature extraction, feature selection and classification and mental health and NLP.

◆ References:

1. https://www.researchgate.net/publication/274404702_Screening_Internet_forum_participants_for_depression_symptoms_by_assembling_and_enhancing_multiple_NLP_methods
2. http://ceur-ws.org/Vol-2936/paper-78.pdf
3. https://www.researchgate.net/publication/346497486_Recognizing_Suicidal_Intent_in_Depressed_Population_using_NLP_A_Pilot_Study?enrichId=rgreq-d40b4a23cabe453bcfa0e82cbf55c504-XXX&enrichSource=Y292ZXJQYWdlOzM0NjQ5NzQ4NjtBUzo5OTkxNjE5MDYzMzU3NDRAMTYxNTIzMDE3

4. https://www.cambridge.org/core/journals/natural-language-engineering/article/natural-language-processing-in-mental-health-applications-using-nonclinical-texts/32645FFCFD37C67DA62CA06DB66EB2F4
5. https://www.hindawi.com/journals/cmmm/2016/8708434/
6. https://bmjopen.bmj.com/content/7/1/e012012?cpetoc=&utm_source=trendmd&utm_medium=cpc&utm_campaign=oem&trendmd-shared=1&utm_term=TrendMDPhase4&utm_content=Journalcontent