Background, methods and goals:

Predicting Depression in Social Media Texts

Liisa Jullinen

Behavioral and linguistic cues from social media data have been used to predict various psychological problems using machine learning methods: major depression, suicidality, eating disorders, schizophrenia and many other conditions, in addition to related symptomatology, such as self-harm and stress, with high accuracy (80–90%) (see Chancellor and Choudhury, 2020).

Data for this research is usually taken from various social media platforms, such as Twitter/X, Reddit, Facebook and both the texts and metadata have been analysed: for example in predicting depression the features have been demographic, language, social media activity (for example Twitter activity – Tsugawa et al., 2015), temporal characteristics and even the color schemas of social media photos (Reece et al., 2017).

Regarding depression prediction, the main symptoms (not all are necessary for a diagnosis) of depression can be found in DSM-5: depressed mood, diminished interest or pleasure in what used to be pleasurable, significant changes in appetite or weight, insomnia or hypersomnia, psychomotor agitation/ retardation, fatigue or loss of energy, feelings of worthlessness and guilt, diminished ability to concentrate, recurrent thoughts of death or suicide (American Psychiatric Association, 2022). Some of these, but not all, can be detected in language use of depressed persons. Linguistically speaking, it has been found that depression and first-person singular pronoun use ("I-talk") and negative emotion word use are correlated positively with major depression and positive emotion word use has been correlated negatively to major depression (Tackman, et al., 2019). There is also a correlation between severity of symptoms and the aforementioned linguistic features found in a larger meta-analysis (Tølbøll, 2019).

Thus, not all text analysis tools can be used to predict depression - the tools must be chosen in order to guarantee construct validity. The main caveat in much of the research so far has been using weak labels (depression = posting in a subforum for depression) and analysing the occurrence of specific words (reflected mostly by TF-IDF features). This means that some of the models, that have a high accuracy, are actually not able to predict depression, but rather predict the subject matter: they capture WHAT is being said, rather than HOW. If HOW depressed people use language differently could be captured, models could differentiate depressed subjects from non-depressed subjects no matter what they were talking about (be it depression, weather or a football match).

To understand style, structure, sentiment, pragmatics, discourse (how language is used) rather than topics, entities, keywords (what is being said), methods and features that capture *linguistic form, style, and function* should be used, rather than semantics alone. Methods that could be used include: stylometry, syntactic analysis, discourse and pragmatic analysis, sentiment / emotion and affect modeling, linguistic complexity and readability,

Among other text analyses tools, some of the tools are: Linguistic Inquiry and Word Count LIWC (Tausczik and Pennebaker, 2010) and DepecheMood (Staiano and Guerini, 2014), Empath (open-source LIWC alternative), SEANCE (Sentiment & Affect wordlists), TAALES / TAACO / TAASSC, stylo (R package), JGAAP (Java Graphical Authorship Attribution Program), Writeprints, Burrows’ Delta, Function word frequency analysis, Character n-grams, MTLD, HD-D (Vocabulary richness metrics), Coh-Metrix, readability metrics (like Flesch–Kincaid, Gunning Fog, Dale–Chall), etc.

A short table of the most often used methods for assessing style and form:

Method	**What is measures**
LIWC	psychological + stylistic profiling
Coh-Metrix	discourse cohesion & complexity
TAALES + TAACO + TAASSC + SÉANCE	lexical / discourse / syntax / emotion
Biber’s Multidimensional Analysis (MAT)	register & style
spaCy / Stanza	POS + syntax structure
NRC Emotion Lexicon / VADER	tone & affect
Stylometry tools (stylo, JGAAP)	author style
Readability metrics (textstat)	complexity

Writing style and word usage has been linked to depression in several studies so far: for example absolutism and the ‘I-talk’ (more 1st person pronouns) have been linked to depression (Zulkarnain, et al., 2020) and in another study the style of writing could predict if a writer was suicidal later in life or not (Agurto et al., 2018).

In much of the literature so far, however, predicting depression via social media posts and machine learning methods is pitted against traditional clinical methods of diagnosing depression, saying the former being a better option in regards to money, patient availability (claiming people don't want to seek help), even accuracy. This should be taken with extreme caution and machine learning and traditional clinical methods, however accurate and valid the results of the former, should not compete and they should also serve entirely different purposes. Predicting depression via social media posts should never be used as a clinical diagnosis for an enormous number of reasons: including legal (informed consent among many), adequacy (diagnosis are based on several scientific methods, available only in person), availability and effectiveness of therapy/medication, etc.

Having said that, predicting depression via social media posts does have a potentially very important role in prevention and support for mental illness. Machine learning algorithms are used to nudge in various ways useful for large stakeholders - however they could be used for some greater good, for example depression prevention. The most important question, of course, is the usability of the results. Can we actually do something if depressive symptoms (potentially a suicidal or depressive user) are detected? Could we somehow make the platforms nudge these users toward help? Could we actually interfere if a life is threatened and how does GDPR relate to this? The results could be used in a separate app, that users can voluntarily download, agreeing to user terms (if the GDPR allows this). From another perspective, the results could also provide for new data to supplement clinical care, assessing developing conditions (Chancellor and Choudhory, 2020).

References

Agurto, C., Pataranutaporn, P., Eyigoz, E., Stolovitzky, G., Cecchi, G. (2018). Predictive Linguistic Markers of Suicidality in Poets. 282-285. 10.1109/ICSC.2018.00051.
American Psychiatric Association. (2022). Diagnostic and statistical manual of mental disorders (5th ed., text rev.).

Chancellor, S., De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: a critical review. npj Digit. Med. 3, 43. https://doi.org/10.1038/s41746-020-0233-7

Reece, A. G. & Danforth, C. M. Instagram photos reveal predictive markers of depression. EPJ Data Science 6, 1–34 (2017).

Staiano, J. and Guerini, M. (2014). Depeche Mood: a Lexicon for Emotion Analysis from Crowd Annotated News. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 427–433, Baltimore, Maryland. Association for Computational Linguistics.

Tackman, A. M., Sbarra, D. A., Carey, A. L., Donnellan, M. B., Horn, A. B., Holtzman, N. S., Edwards, T. S., Pennebaker, J. W., & Mehl, M. R. (2019). Depression, negative emotionality, and self-referential language: A multi-lab, multi-measure, and multi-language-task research synthesis. Journal of personality and social psychology, 116(5), 817–834. https://doi.org/10.1037/pspp0000187

Tausczik, Yla & Pennebaker, James. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. Journal of Language and Social Psychology. 29. 24-54. 10.1177/0261927X09351676.

Tølbøll, K. B. (2019). Linguistic features in depression: a meta-analysis. Journal of Language Works - Sprogvidenskabeligt Studentertidsskrift, 4(2), 39–59. Retrieved from https://tidsskrift.dk/lwo/article/view/117798

Tsugawa, S. et al. Recognizing depression from twitter activity. In Proc. ACM Conference on Human Factors in Computing Systems (CHI). 3187–3196 (ACM, 2015).

Zulkarnain, Nur Zareen and Abdullah, Norida and Basiron, Halizah (2020) Writing style and word usage in detecting depression in social media: A review. Journal of Theoretical and Applied Information Technology, 98 (1). 124 - 135. ISSN 1992-8645