

Editorial

Natural language processing in psychiatry: the promises and perils of a transformative approach

Neguine Rezaii, Phillip Wolff and Bruce H. Price



A person's everyday language can indicate patterns of thought and emotion predictive of mental illness. Here, we discuss how natural language processing methods can be used to extract indicators of mental health from language to help address long-standing problems in psychiatry, along with the potential hazards of this new technology.

Keywords

Natural language processing; machine learning; privacy; precision medicine; prediction and classification of diseases.

Copyright and usage

© The Author(s), 2022. Published by Cambridge University Press on behalf of the Royal College of Psychiatrists.

Neguine Rezaii is an Instructor of Neurology and Physician Investigator in Neuropsychiatry and Behavioral Neurology at Harvard Medical School. Her research focuses on extracting the indicators of mental health from natural language through advances in artificial intelligence. **Phillip Wolff** is a Professor of Psychology at Emory University. He is an expert in the use of Machine Learning and Natural Language Processing in the analysis of language semantics and human cognition. **Bruce H Price** is a senior clinical associate in neurology at Massachusetts General Hospital, with over 140 original articles, critical reviews, commentaries and text book chapters in neuropsychiatry.

By expressing how we feel, what we think and where it hurts, we provide an account of the functional status of our bodies. The information contained in these expressions often extends beyond what is directly said. Use of a particular phrase, word order, conversational maxim or even acoustic feature can provide clues to a person's physiological or pathological state. Though rich in diagnostic value, these signals are often too subtle to be heard with the naked ear. Crucially, what is concealed to the typical human observer can be revealed through computational tools made available through machine learning and natural language processing (NLP).

NLP is a subdivision of computer science that focuses on the learning, comprehension and production of everyday human languages. State-of-the-art NLP models acquire their knowledge of syntax, semantics and pragmatics from large amounts of text, on the order of billions of words, and store that knowledge in layers of artificial neural networks. The resulting knowledge can be harnessed to address multiple long-standing problems in psychiatry.

Making the subjective objective

Clinicians learn to pick up on the key indicators of illness in people's communication, including their spoken language, gestures, facial expressions, silences, eye contact and posture. Typically, these aspects of communication are evaluated in a qualitative manner, making them vulnerable to bias and inconsistency across clinicians. Tools from NLP can be used to help make these impressions more objective (path 2a in Fig. 1). For example, a graph theoretic approach has been developed to measure incoherent language, a cardinal symptom of many neuropsychiatric diseases. Through an analysis of the linkages between words, this approach can measure incoherence and identify its causes, such as whether the incoherence was a result of schizophrenia or mania.¹ Developments in other areas of artificial intelligence are rapidly emerging that will help quantify the nonverbal aspects of communication, such as the analysis of facial expressions.

Discovering symptoms that are easily missed

Methods from NLP can go beyond what is overtly stated by revealing hidden dimensions of meaning, thereby addressing a long-sought goal in psychiatric disciplines such as psychoanalysis. A recent study using statistical models of NLP found that indirect reference to auditory concepts, such as whisper, chant or voice, could predict later development of psychosis during a 2-year follow-up. The implicit reference to auditory concepts could only be detected by an NLP model because the participants did not directly use these words, but rather implied their concepts. Subjective ratings by clinicians of the same language samples at this early stage were insensitive to the increased reference to auditory concepts and could not predict the onset of psychosis.²

Facilitating the assignment of patients to disease categories

Sometimes the disease categories in psychiatry are clear, but the symptoms determining the assignment of a patient to a particular disease category are not. For example, suicide is an unambiguous disease outcome, but predicting the likely occurrence of suicide has been a challenge. However, recent work in NLP demonstrates that the occurrence of suicide is well-predicted by an individual's word choices and vocal characteristics.³ Crucially, the models for predicting a particular disease integrate hundreds of linguistic and acoustic features, likely more than can be simultaneously considered by a human observer.

Discovering new categories of disease

Sometimes, on the other hand, the symptoms are clear, but the disease categories are not well understood. NLP methods can be used to help discover and define natural categories of disease (path 2b in Fig. 1). This can be achieved by measuring similarities in language production across individuals and determining the number of ways these individuals tend to cluster in terms of this data. For example, natural clusters in primary progressive aphasia have been identified on the basis of the language features of syntactic complexity, hesitation rate and naming.⁴ The validity of these clusters can be assessed by determining whether the clusters indicate discontinuities in other kinds of biomarkers, such as imaging or pathological data. Validity can also be assessed by determining the generalisability of the clustering solution across different groups of individuals, using cross-validation methods. These methods for assessing

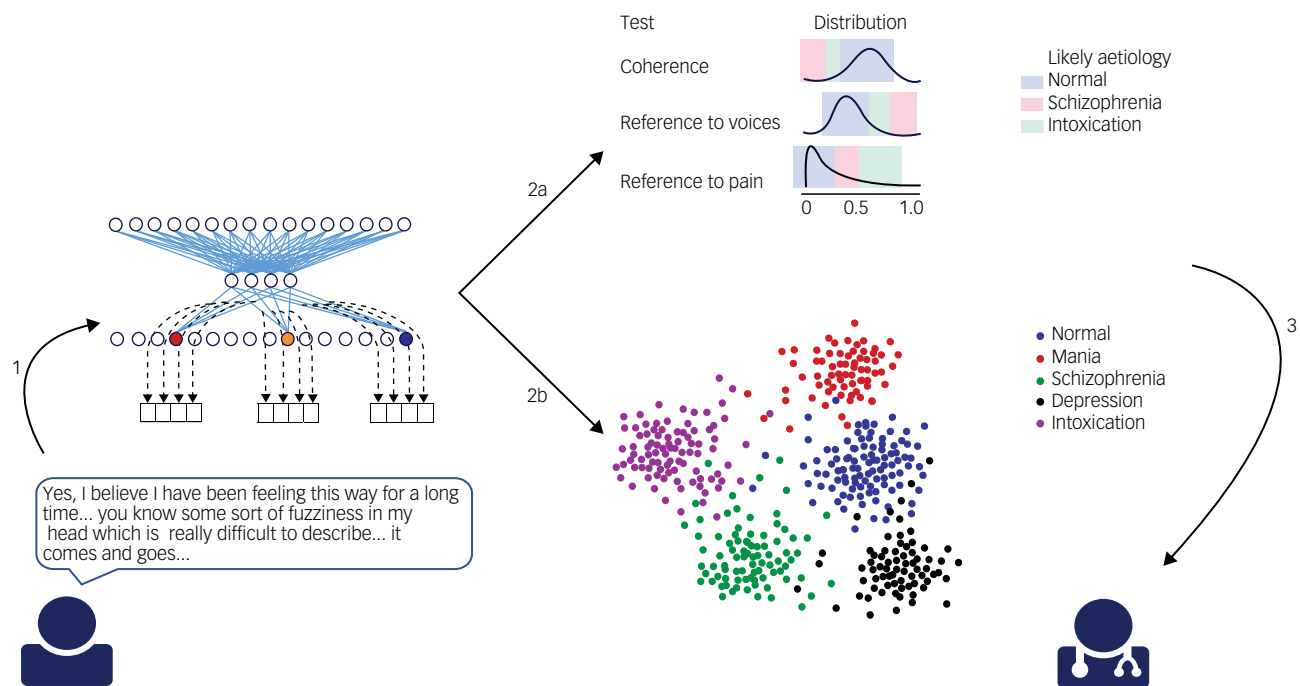


Fig. 1 The schematic figure of two NLP approaches to diagnosis. Path 1: A patient's language sample is used as an input to a trained NLP model. Path 2a: A model can be used to quantify various signs and symptoms of a disease in an objective way. Test values and their distributions can then provide insight into the likely aetiology. Path 2b: Alternatively, without knowing the explicit signs or symptoms of a disease, an NLP model can reveal natural partitionings in language samples of various aetiologies. Each dot is a language sample of an individual. Path 3: The outputs of these models can assist the provider in establishing the diagnosis. NLP, natural language processing.

naturalness are regularly included in the pipeline of selecting one clustering solution over another.


Machine learning and NLP methods have traditionally involved large amounts of patient training data to make reasonable inferences. This common practice need not imply that such models can only be used when there is a large amount of data. Recent NLP models can capitalise on prior training in other domains of knowledge, to be able to identify and classify even rare neuropsychiatric diseases. For example, the language model RoBERTa was used to obtain clusters of individuals with primary progressive aphasia on the basis of natural speech samples. The solution agreed highly with the canonical subtypes of the disease.⁵ Crucially, the model was able to discover these subtypes without specific training on the disease.

Amidst the multitude of promises lie perils. Because such models are trained on texts generated by humans, they can acquire some of the toxic stereotypes, biases and beliefs that might be inherent in these texts. They can also be vehicles for misinformation. Recent research has developed methods for detoxifying these models. However, large corporations have shown a limited ability or willingness to implement such protective measures. Although artificial intelligence may often be the source of damaging and inaccurate content, it may also be part of the solution. Language models may be used to detect misinformation by assessing the conceptual consistency of this information with prior knowledge. As the reasoning capabilities of these models increase, they may also be able to recognise damaging biases and stereotypes, which would be a first step in their removal.

Another hazard of using NLP tools to analyse people's day-to-day language is the potential impact on privacy. It is human nature to communicate freely, making it easy to collect unsolicited language samples that can be mined for highly sensitive information

about a person's health, personality, race, and cognitive capabilities and limitations. This knowledge could be exploited to target an individual, predict their decisions and even mimic their behaviour. Therefore, unknown to most people, far more information about their beliefs, feelings and vulnerabilities may be revealed than they ever intended. As a countermeasure, language profiling programs may emerge to flag and conceal sensitive health information, much like word-processing programs that can be used to automatically identify and correct grammar or spelling errors.

Although it may be possible to manage damaging biases and invasion of privacy, an even deeper problem might emerge with respect to the way we view one another. NLP techniques might potentially lead to an objectifying perspective that reduces the patient to a matrix of numbers. Although the dangers of reductionism are certainly present, they are not inevitable. Such methods likely work best when they are based on language samples drawn from substantive clinical interactions and conversations that can only come about from a strong patient–provider connection. Once the results from the NLP methods are obtained, the provider will be in possession of an additional layer of information that deepens their understanding of the patient's personal experience, a type of insight that could lead to a more comprehensive level of care.

Neguine Rezaei , MD, Instructor of Neurology, Physician Investigator in Neuropsychiatry and Behavioral Neurology, Massachusetts General Hospital, Harvard Medical School, USA; **Phillip Wolff**, PhD, Professor of Psychology, Department of Psychology, Emory University, USA; **Bruce H. Price**, MD, Associate Professor of Neurology, Department of Neurology, Massachusetts General Hospital, Harvard Medical School, USA

Correspondence: Neguine Rezaei. Email: nrezaei@mgh.harvard.edu

First received 20 Jan 2021, final revision 26 Oct 2021, accepted 21 Nov 2021

Data availability

Data availability is not applicable to this article as no new data were created or analysed in this study.

Author contributions

Each individual author contributed to the development of the main ideas, writing and revision of the final manuscript.

Funding

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Declaration of interest

None.

References

- 1 Mota NB, Vasconcelos NAP, Lemos N, Pieretti AC, Kinouchi O, Cecchi GA, et al. Speech graphs provide a quantitative measure of thought disorder in psychosis. *PLoS One* 2012; **7**(4): e34928.
- 2 Rezaei N, Walker E, Wolff P. A machine learning approach to predicting psychosis using semantic density and latent content analysis. *NPJ Schizophr* 2019; **5**(1): 9.
- 3 Pestian JP, Sorter M, Connolly B, Bretonnel Cohen K, McCullumsmith C, Gee JT, et al. A machine learning approach to identifying the thought markers of suicidal subjects: a prospective multicenter trial. *Suicide Life Threat Behav* 2017; **47**(1): 112–21.
- 4 Hoffman P, Sajjadi SA, Patterson K, Nestor PJ. Data-driven classification of patients with primary progressive aphasia. *Brain Lang* 2017; **174**: 86–93.
- 5 Rezaei N, Carvalho N, Brickhouse M, Loyer E, Wolff P, Touroutoglou A, Wong B, Quimby M, Dickerson B. Neuroanatomical mapping of artificial intelligence-based classification of language in PPA. *Alzheimer's Association International Conference (Denver, 26–30 Jul 2021)*. Alzheimer's Association, 2021.

Extra

Auschwitz: dreaming the nightmare of day – Professor Viktor Frankl (119,104)

Greg Wilkinson 

'What did the prisoner dream about most frequently? Of bread, cake, cigarettes, and nice warm baths. The lack of having these simple desires satisfied led him to seek wish-fulfilment in dreams. Whether these dreams did any good is another matter; the dreamer had to wake from them to the reality of camp life, and to the terrible contrast between that and his dream illusions.

I shall never forget how I was roused one night by the groans of a fellow prisoner, who threw himself about in his sleep, obviously having a horrible nightmare. Since I had always been especially sorry for people who suffered from fearful dreams or deliria, I wanted to wake the poor man. Suddenly I drew back the hand which was ready to shake him, frightened at the thing I was about to do. At that moment I became intensely conscious of the fact that no dream, no matter how horrible, could be as bad as the reality of the camp which surrounded us, and to which I was about to recall him.¹

Reference

- 1 Frankl V. *Man's Search For Meaning*. Rider, 2004.

© The Author(s), 2022. Published by Cambridge University Press on behalf of the Royal College of Psychiatrists

The British Journal of Psychiatry (2022)
220, 253. doi: 10.1192/bjp.2021.174