

Neural network learns speech patterns that predict depression in clinical interviews

August 30 2018, by Rob Matheson



MIT researchers have developed a neural-network model that can analyze raw text and audio data from interviews to discover speech patterns indicative of depression. This method could be used to develop diagnostic aids for clinicians that can detect signs of depression in natural conversation. Credit: Massachusetts Institute of Technology

To diagnose depression, clinicians interview patients, asking specific questions—about, say, past mental illnesses, lifestyle, and mood—and identify the condition based on the patient's responses.

In recent years, machine learning has been championed as a useful aid for diagnostics. Machine-learning models, for instance, have been developed that can detect words and intonations of speech that may indicate depression. But these models tend to predict that a person is depressed or not, based on the person's specific answers to specific questions. These methods are accurate, but their reliance on the type of question being asked limits how and where they can be used.

In a paper being presented at the Interspeech conference, MIT researchers detail a [neural-network](#) model that can be unleashed on raw text and audio data from interviews to discover speech patterns indicative of depression. Given a new subject, it can accurately predict if the individual is depressed, without needing any other information about the questions and answers.

The researchers hope this method can be used to develop tools to detect signs of depression in natural conversation. In the future, the model could, for instance, power mobile apps that monitor a user's text and voice for mental distress and send alerts. This could be especially useful for those who can't get to a clinician for an initial diagnosis, due to distance, cost, or a lack of awareness that something may be wrong.

"The first hints we have that a person is happy, excited, sad, or has some serious cognitive condition, such as depression, is through their speech," says first author Tuka Alhanai, a researcher in the Computer Science and Artificial Intelligence Laboratory (CSAIL). "If you want to deploy [depression-detection] models in scalable way ... you want to minimize the amount of constraints you have on the data you're using. You want to deploy it in any regular conversation and have the model pick up, from

the natural interaction, the state of the individual."

The technology could still, of course, be used for identifying mental distress in casual conversations in clinical offices, adds co-author James Glass, a senior research scientist in CSAIL. "Every patient will talk differently, and if the model sees changes maybe it will be a flag to the doctors," he says. "This is a step forward in seeing if we can do something assistive to help clinicians."

The other co-author on the paper is Mohammad Ghassemi, a member of the Institute for Medical Engineering and Science (IMES).

Context-free modeling

The key innovation of the model lies in its ability to detect patterns indicative of depression, and then map those patterns to new individuals, with no additional information. "We call it 'context-free,' because you're not putting any constraints into the types of questions you're looking for and the type of responses to those questions," Alhanai says.

Other models are provided with a specific set of questions, and then given examples of how a person without depression responds and examples of how a person with depression responds—for example, the straightforward inquiry, "Do you have a history of depression?" It uses those exact responses to then determine if a new individual is depressed when asked the exact same question. "But that's not how natural conversations work," Alhanai says.

The researchers, on the other hand, used a technique called sequence modeling, often used for speech processing. With this technique, they fed the model sequences of text and audio data from questions and answers, from both depressed and non-depressed individuals, one by one. As the sequences accumulated, the model extracted [speech patterns](#)

that emerged for people with or without depression. Words such as, say, "sad," "low," or "down," may be paired with audio signals that are flatter and more monotone. Individuals with depression may also speak slower and use longer pauses between words. These text and audio identifiers for [mental distress](#) have been explored in previous research. It was ultimately up to the model to determine if any patterns were predictive of depression or not.

"The model sees sequences of words or speaking style, and determines that these patterns are more likely to be seen in people who are depressed or not depressed," Alhanai says. "Then, if it sees the same sequences in new subjects, it can predict if they're depressed too."

This sequencing technique also helps the model look at the conversation as a whole and note differences between how people with and without depression speak over time.

Detecting depression

The researchers trained and tested their model on a dataset of 142 interactions from the Distress Analysis Interview Corpus that contains audio, text, and video interviews of patients with mental-health issues and virtual agents controlled by humans. Each subject is rated in terms of depression on a scale between 0 to 27, using the Personal Health Questionnaire. Scores above a cutoff between moderate (10 to 14) and moderately severe (15 to 19) are considered depressed, while all others below that threshold are considered not depressed. Out of all the subjects in the dataset, 28 (20 percent) are labeled as depressed.

In experiments, the model was evaluated using metrics of precision and recall. Precision measures which of the depressed subjects identified by the model were diagnosed as depressed. Recall measures the accuracy of the model in detecting all subjects who were diagnosed as depressed in

the entire dataset. In precision, the model scored 71 percent and, on recall, scored 83 percent. The averaged combined score for those metrics, considering any errors, was 77 percent. In the majority of tests, the researchers' model outperformed nearly all other models.

One key insight from the research, Alhanai notes, is that, during experiments, the model needed much more data to predict depression from audio than text. With text, the model can accurately detect depression using an average of seven question-answer sequences. With audio, the model needed around 30 sequences. "That implies that the patterns in words people use that are predictive of depression happen in shorter time span in text than in audio," Alhanai says. Such insights could help the MIT researchers, and others, further refine their models.

This work represents a "very encouraging" pilot, Glass says. But now the researchers seek to discover what specific patterns the [model](#) identifies across scores of raw data. "Right now it's a bit of a black box," Glass says. "These systems, however, are more believable when you have an explanation of what they're picking up. ... The next challenge is finding out what data it's seized upon."

The researchers also aim to test these methods on additional data from many more subjects with other cognitive conditions, such as dementia. "It's not so much detecting [depression](#), but it's a similar concept of evaluating, from an everyday signal in speech, if someone has cognitive impairment or not," Alhanai says.

More information: "Detecting Depression with Audio/Text Sequence Modeling of Interviews" groups.csail.mit.edu/sls/publications/Interspeech-2018.pdf

This story is republished courtesy of MIT News

(web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: Neural network learns speech patterns that predict depression in clinical interviews (2018, August 30) retrieved 26 April 2024 from <https://techxplore.com/news/2018-08-neural-network-speech-patterns-depression.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.