Researchers train tech tool to find relationship clues from written conversations

April 24 2020, by Laurel Thomas

Credit: CC0 Public Domain

Social scientists have identified 10 dimensions to describe the nature of human relationships but little research has focused on how these concepts are expressed through written language, and what role they

have in shaping social interactions.

New research from the University of Michigan and Nokia Bell Labs has used crowdsourcing and a tech tool to detect how these characteristics are expressed in everyday language and how they shape social dynamics.

In particular, the researchers wanted to find out if conversations could provide insight into knowledge, wealth, education and mental illness, including suicide. By examining 160 million Reddit messages, 290,000 email messages from the defunct Enron Corp., and 300,000 lines of dialogue from movies, the researchers were able to identify the 10 characteristics in written communications.

Using natural language processing, the team predicted social dimensions, including the relationship between people, for example one of conflict or support, and the type of real-world communities they shape, (e.g., wealthy or deprived).

"We first demonstrate how we build those models for measuring the levels of each dimension from a given conversation. We then show that our models perform well in predicting not only the dimensions that exist within a conversation, but also at a higher level, between individuals," said Minje Choi, doctoral student at the School of Information who conducted the research while on an internship at Nokia Bell Labs.

"We also showed that levels of dimensions such as knowledge or social support can relate to societal outcomes such as how wealthy they are, or what the suicide rate is."

Choi, the study's first author, and colleagues used crowdsourcing to first identify messages according to the 10 characteristics: knowledge, power, status, trust, support, romance, similarity, identity, fun and conflict.

More than 900 crowdsourced annotators labeled 7,855 sentences from Reddit posts, 400 from movie lines and 436 from Enron emails, which demonstrated the presence of the 10 characteristics.

The researchers then trained a deep-learning classifying tool to look for those characteristics and the relationships they represented in all of the Reddit and Enron messages, and the movie dialogue.

They also used data from Tinghy.org, a gamified psychological test that measures Twitter users' perceptions of their online relationships using the 10 dimensions. They studied 1,772 relationships between 1,406 unique individuals.

In addition to identifying the known dimensions in the messages, the researchers found:

- Knowledge is the strongest significant predictor of education level and income.
- The presence of support and absence of trust are the two most important predictors of suicide rates.
- Population density impacts suicide rates, with urban areas that are richer and more educated showing fewer cases.
- Suicide rates are higher in states with fewer expressions of identity, in line with previous studies that found an association between lack of sense of belonging and risk of depression-related suicides among young people.
- States with higher education exhibit lower levels of conflict, consistent with studies finding that hate speech is fueled by low education levels.
- Wealth is associated with a reduced number of expressions that point out similarities between points of view, a possible sign of structurally and culturally diverse communities.

Choi said the team's hope is that others will use their model to continue to explore the connections between relationship dimensions and written communication.

"This can be used as an analysis tool for researchers who have these conversation data and would like to measure levels or changes in dimensions, such as social support or conflict from their data," Choi said. "It can be used to look for temporal changes (as we did in the Enron example) or community-wise differences (as we did with U.S. state-level Reddit comments).

Provided by University of Michigan