Training computers to tease out the subtext behind the text

30 November 2021, by Brittany Steff

It is hard enough for humans to interpret the deeper meaning and context of social media and news articles. Asking computers to do it is a nearly impossible task. Even C-3PO, fluent in over 6 million forms of communication, misses the subtext much of the time.

Natural [language](#) processing, the subfield of [artificial intelligence](#) connecting computers with [human](#) languages, uses statistical methods to analyze language, often without incorporating the real-world context needed for understanding the shifts and currents of human society. To do that, you have to translate online communication, and the context from which it emerges, into something the computers can parse and reason over.

Dan Goldwasser, associate professor of computer science at Purdue University, and other members of his team strive to address that by developing new ways to model human language and allow computers to better understand us.

"The motivation of our work is to get a better understanding of public [discourse](#), how different issues are discussed, the arguments made and the perspectives underlying these arguments," Goldwasser said. "We would like to represent the points of view expressed by the thousands, or even more, of people describing their experiences online. Understanding the language used to discuss issues can help shed light on the different considerations behind decision-making processes, including both individual health and well-being choices and broader policy decisions."

Goldwasser emphasizes that part of the challenge is that so much of online communication relies on readers already knowing the context—whether it's shorthand on Twitter or the basis of understanding a meme. To analyze the communication, the context is a vital part of the message.

"In many of the scenarios we study, progress relies on finding new ways to conceptualize language understanding, by grounding it in a real-world context," he said. "Operationalizing it requires developing new technical solutions."

Goldwasser and his students use techniques distilled from the combined wisdom of [computer](#) science, artificial intelligence and computational social science.

Goldwasser's lab studies the language used on social media, traditional media stories and in legislative texts to understand the [context](#) and assumptions of the speakers and writers. In a world where the written word is flourishing and every person with an internet connection can act as a journalist, being able to study and analyze that writing in an unbiased manner is crucial to human understanding of our own society.

APA citation: Training computers to tease out the subtext behind the text (2021, November 30) retrieved 27 May 2022 from https://techxplore.com/news/2021-11-subtext-text.html