

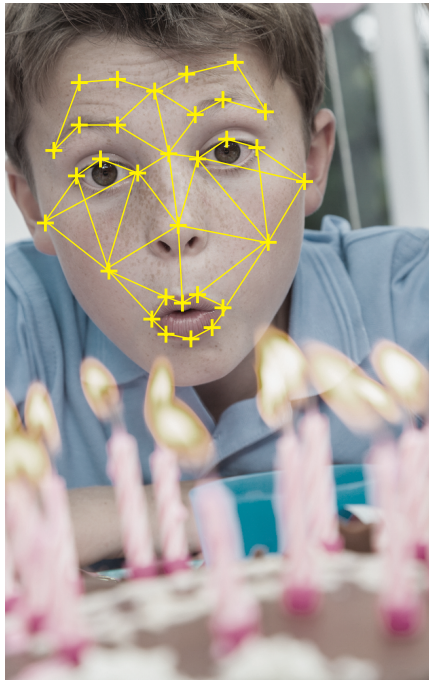
Artificial (Emotional) Intelligence

Enabled by advances in computing power and neural networks, machines are getting better at recognizing and dealing with human emotions.

ANYONE WHO HAS BEEN frustrated asking questions of Siri or Alexa—and then annoyed at the digital assistant’s tone-deaf responses—knows how dumb these supposedly intelligent assistants are, at least when it comes to emotional intelligence. “Even your dog knows when you’re getting frustrated with it,” says Rosalind Picard, director of Affective Computing Research at the Massachusetts Institute of Technology (MIT) Media Lab. “Siri doesn’t yet have the intelligence of a dog,” she says.

Yet developing that kind of intelligence—in particular, the ability to recognize human emotions and then respond appropriately—is essential to the true success of digital assistants and the many other artificial intelligences (AIs) we interact with every day. Whether we’re giving voice commands to a GPS navigator, trying to get help from an automated phone support line, or working with a robot or chatbot, we need them to really understand us if we’re to take these AIs seriously. “People won’t see an AI as smart unless it can interact with them with some emotional *savoir faire*,” says Picard, a pioneer in the field of affective computing.

One of the biggest obstacles has been the need for context: the fact that emotions can’t be understood in isolation. “It’s like in speech,” says Pedro Domingos, a professor of computer science and engineering at the University of Washington and author of *The Master Algorithm*, a popular book about machine learning. “It’s very hard to recognize speech from just the sounds, because they’re too ambiguous,” he points out. Without context, “ice cream” and “I scream” sound



identical, “but from the context you can figure it out.”

The same is true of emotional expression. “If I zoom in on a Facebook photo and you see a little boy’s eyes and mouth,” says Picard, “you might say he looks surprised. And then if we zoom out a little bit you might say, ‘Oh, he’s blowing out a candle on his cake—he’s probably happy and excited to eat his cake.’” Getting the necessary context requires the massive amounts of data that computers have only recently been able to process. Aiding that processing, of course, are today’s powerful deep-learning algorithms.

These advances have led to major breakthroughs in emotion detection just in the last couple of years, such as learning from the raw signal, says Björn Schuller, editor-in-chief of *IEEE Transactions on Affective Computing* and head of Imperial College London’s Group on Language, Audio & Music. Such “end-

to-end” learning, which Schuller himself has helped develop, means a neural network can use just the raw material (such as audio or a social media feed) and the labels representing different emotions to “learn all by itself to recognize the emotion inside,” with minimal labeling by humans.

The latest algorithms are also enabling what scientists call “multimodal processing,” or the integration of signals from multiple channels (“modalities”), such as facial expressions, body language, tone of voice, and physiological signals like heart rate and galvanic skin response. “That’s very important because the congruence of those channels is very telling,” says Maja Matarić, a professor of computer science, neuroscience, and pediatrics at the University of Southern California. Given people’s tendency to mask their emotions, information from only a single channel, such as the face, can mislead; a more accurate picture emerges by piecing together multiple modalities. “The ability to put together all the pieces is getting a lot more powerful than it ever has been,” adds Picard, whose research group has used the multimodal approach to not only discern a person’s current mood, but even to predict their mood the next day, with the goal of using the information to improve future moods.

Along similar lines, several researchers have engineered ways to mine multiple streams of data to detect severity of depression, thus potentially predicting the suicide risk of callers on a mental-health helpline.

Although recognizing emotions might seem like a uniquely human strength, experts point out that emotions can be distilled to sets of signals that can be measured like any other phenomenon. In fact, the power of

a multimodal approach for emotion recognition suggests that computers actually have an edge over humans. In a recent study, Yale University social psychologist Michael Kraus found that when people try to guess another's emotion, they're more accurate when they only hear the person's voice than when they're able to use all their senses, which tend to distract. In other words, for humans, less is more. Kraus attributes this effect to our limited bandwidth, a processing constraint computers are increasingly shedding. "I think computers will surpass humans in emotion detection because of the bandwidth advantage," says Kraus, an assistant professor of organizational behavior.

This computational advantage also shows itself in recent advances in recognizing "microexpressions," involuntary facial expressions so fleeting they are difficult for anyone but trained professionals to spot. "Modeling facial expressions is nothing new," says James Z. Wang, a professor of Information Sciences and Technology at Pennsylvania State University. But microexpressions—which might reveal a telltale glimpse of sadness behind a lingering smile, for example—are trickier. "These are so subtle and so fast"—flashing for less than a fifth of a second—that they require the use of high-speed video recordings to capture and new techniques to model computationally. "We take the differences from one frame to another, and in the end we are able to classify and identify these very subtle, very spontaneous expressions," says Wang.

What happens after a computer correctly recognizes an emotion? Responding appropriately is a separate challenge, and the progress in this area has been less revolutionary in recent years, according to Schuller. Most dialogue systems, for example, still follow hand-crafted rules, he says. A phone help line might be just smart enough to transfer you to a human operator if it senses you are angry, but still lacks the sophistication to calm you down by itself.

On the other hand, some researchers are already designing robots that can respond intelligently enough to influence human behavior in positive ways. Consider the "socially assistive robots" that Mataric designs to help

autistic children learn to recognize and express emotions. "All robots are autistic to a degree, as are children who are autistic," Mataric points out, "and that's something to leverage."

Autistic children find robots easier to interact with than humans, yet more engaging than a disembodied computer, making them ideal learning companions. If a kid-friendly robot seeing a child act in a socially appropriate way during a training session makes the child happy by blowing bubbles, for example, the child will be more motivated to improve. But while the robots must recognize and reward appropriate behavior, "the idea is not to just reward," Mataric explains, "but rather to serve as a peer in an interaction that gives children with autism the opportunity to learn and practice social skills."

Mataric and her colleagues are using similar approaches in designing robots that work with stroke patients and with obese teens, with robots understanding how much they can push each user to exercise more.

These robots are not meant to replace human caregivers, Mataric says, but to complement them. That's also the goal of Jesse Hoey, an associate professor of health informatics and AI at the University of Waterloo, who is developing an emotionally aware system to guide Alzheimer's disease patients through hand-washing and other common household tasks.

"At first glance it seems like a straightforward problem," Hoey says: just use sensors to track where the patients are in the task and use a recorded voice to prompt them with the next step when they forget what they've done. The mechanics of the system work just fine—but too often, people with Alzheimer's ignore the voice prompt. "They don't listen to the prompt, they don't like it, they react negatively to it, and the reason they react in all these different ways, we started to understand, was largely to do with their emotional state at a fairly deep level; their sense of themselves and who they are and how they like to be treated."

Hoey's starkest example is of a World War II veteran who grew very distressed because he thought the voice was a call to arms; for this user, a female voice might have been more effective. Another patient, who had once

been a lawyer, shifts his self-image from day to day, sometimes doling out legal advice and other days acting more in line with his current identity. "The human caregivers are good at picking this up," Hoey says, and that sensitivity enables them to treat the patient with the appropriate level of deference. So the challenge is to create a computer system that picks up on signals of power (such as body posture and speech volume), as well as signals of other aspects of emotion.

Applying artificial emotional intelligence to help those struggling with Alzheimer's, autism, and the like certainly seems noble—but not all applications of such technology are as admirable. "What every company wants to do, and won't necessarily admit it, is to know what your emotional state is second by second as you're using their products," says Domingos. "The state of the art of manipulating people's emotions is less advanced than detecting them, but the point at the end of the day is to manipulate them. The manipulations could be good or bad, and we as consumers need to be aware of these things in self-defense." **C**

Editor's Note: For more information on speech emotion recognition, look for Björn Schuller's article in the May 2018 issue.

Further Reading

Jaques, N., Taylor, S., Nosakhare, E., Sano, A., and Picard, R.

Multi-task Learning for Predicting Health, Stress, and Happiness, NIPS Workshop on Machine Learning for Healthcare, December 2016, Barcelona, Spain <http://affect.media.mit.edu/pdfs/16.Jaques-Taylor-et-al-PredictingHealthStressHappiness.pdf>

Tzirakis, P., Trigeorgis, G., Nicolaou, M.A., Schuller, B., and Zafeiriou, S. End-to-End Multimodal Emotion Recognition using Deep Neural Networks, *Journal of LaTeX Class Files*, Vol. 14 No.8, August 2015 <https://arxiv.org/abs/1704.08619>

Xu, F., Zhang, J., and Wang, J. Z. Microexpression Identification and Categorization using a Facial Dynamics Map, *IEEE Transactions on Affective Computing*, vol. 8, no. 2, pp. 254-267, 2017. <http://infolab.stanford.edu/~wangz/project/imsearch/Aesthetics/TAC16/>

Based in San Francisco, CA, USA, Marina Krakovsky is the author of *The Middleman Economy: How Brokers, Agents, Dealers, and Everyday Matchmakers Create Value and Profit* (Palgrave Macmillan).

© 2018 ACM 0001-0782/18/4 \$15.00