

Personalized machine-learning models capture subtle variations in facial expressions to better gauge emotion

25 July 2018, by Rob Matheson



General differences can be seen among cultures, genders, and age groups. But other differences are even more fine-grained: The time of day, how much you slept, or even your level of familiarity with a conversation partner leads to subtle variations in the way you express, say, happiness or sadness in a given moment.

Human brains instinctively catch these deviations, but machines struggle. Deep-learning techniques were developed in recent years to help catch the subtleties, but they're still not as accurate or as adaptable across different populations as they could be.

MIT Media Lab researchers have developed a machine-learning model that takes computers a step closer to interpreting our emotions as naturally as humans do. The model better captures subtle facial expression variations to better gauge moods. By using extra training data, the model can also be adapted to an entirely new group of people, with the same efficacy. Credit: Massachusetts Institute of Technology

MIT Media Lab researchers have developed a machine-learning model that takes computers a step closer to interpreting our emotions as naturally as humans do.

In the growing field of "affective computing," robots and computers are being developed to analyze facial expressions, interpret our emotions, and respond accordingly. Applications include, for instance, monitoring an individual's health and well-being, gauging student interest in classrooms, helping diagnose signs of certain diseases, and developing helpful robot companions.

A challenge, however, is people express emotions quite differently, depending on many factors.

The Media Lab researchers have developed a [machine-learning model](#) that outperforms traditional systems in capturing these small facial expression variations, to better gauge mood while training on thousands of images of faces. Moreover, by using a little extra training [data](#), the model can be adapted to an entirely new group of people, with the same efficacy. The aim is to improve existing affective-computing technologies.

"This is an unobtrusive way to monitor our moods," says Oggi Rudovic, a Media Lab researcher and co-author on a paper describing the model, which was presented last week at the Conference on Machine Learning and Data Mining. "If you want robots with social intelligence, you have to make them intelligently and naturally respond to our moods and emotions, more like humans."

Co-authors on the paper are: first author Michael Feffer, an undergraduate student in electrical engineering and computer science; and Rosalind Picard, a professor of media arts and sciences and founding director of the Affective Computing research group.

Personalized experts

Traditional affective-computing models use a "one-size-fits-all" concept. They train on one set of images depicting various facial expressions, optimizing features—such as how a lip curls when smiling—and mapping those general feature optimizations across an entire set of new images.

The researchers, instead, combined a technique, called "mixture of experts" (MoE), with model personalization techniques, which helped mine more fine-grained facial-expression data from individuals. This is the first time these two techniques have been combined for affective computing, Rudovic says.

In MoEs, a number of neural network models, called "experts," are each trained to specialize in a separate processing task and produce one output. The researchers also incorporated a "gating network," which calculates probabilities of which expert will best detect moods of unseen subjects. "Basically the network can discern between individuals and say, 'This is the right expert for the given image,'" Feffer says.

For their model, the researchers personalized the MoEs by matching each expert to one of 18 individual video recordings in the RECOLA database, a public database of people conversing on a video-chat platform designed for affective-computing applications. They trained the model using nine subjects and evaluated them on the other nine, with all videos broken down into individual frames.

Each expert, and the gating network, tracked facial expressions of each individual, with the help of a residual network ("ResNet"), a neural network used for object classification. In doing so, the model scored each frame based on level of valence (pleasant or unpleasant) and arousal (excitement)—commonly used metrics to encode different emotional states. Separately, six human experts labeled each frame for valence and arousal, based on a scale of -1 (low levels) to 1 (high levels), which the model also used to train.

The researchers then performed further model

personalization, where they fed the trained model data from some frames of the remaining videos of subjects, and then tested the model on all unseen frames from those videos. Results showed that, with just 5 to 10 percent of data from the new population, the model outperformed traditional models by a large margin—meaning it scored valence and arousal on unseen images much closer to the interpretations of human experts.

This shows the potential of the models to adapt from population to population, or individual to individual, with very few data, Rudovic says. "That's key," he says. "When you have a new population, you have to have a way to account for shifting of data distribution [subtle facial variations]. Imagine a model set to analyze facial expressions in one culture that needs to be adapted for a different culture. Without accounting for this data shift, those models will underperform. But if you just sample a bit from a new culture to adapt our model, these models can do much better, especially on the individual level. This is where the importance of the model personalization can best be seen."

Currently available data for such affective-computing research isn't very diverse in skin colors, so the researchers' training data were limited. But when such data become available, the model can be trained for use on more diverse populations. The next step, Feffer says, is to train the model on "a much bigger dataset with more diverse cultures."

Better machine-human interactions

Another goal is to train the [model](#) to help computers and robots automatically learn from small amounts of changing data to more naturally detect how we feel and better serve human needs, the researchers say.

It could, for example, run in the background of a computer or mobile device to track a user's video-based conversations and learn subtle facial expression changes under different contexts. "You can have things like smartphone apps or websites be able to tell how people are feeling and recommend ways to cope with stress or pain, and other things that are impacting their lives negatively," Feffer says.

This could also be helpful in monitoring, say, depression or dementia, as people's facial expressions tend to subtly change due to those conditions. "Being able to passively monitor our [facial expressions](#)," Rudovic says, "we could over time be able to personalize these models to users and monitor how much deviations they have on daily basis—deviating from the average level of facial expressiveness—and use it for indicators of well-being and health."

A promising application, Rudovic says, is human-robotic interactions, such as for personal robotics or robots used for educational purposes, where the robots need to adapt to assess the emotional states of many different people. One version, for instance, has been used in helping robots better interpret the moods of children with autism.

Roddy Cowie, professor emeritus of psychology at the Queen's University Belfast and an affective computing scholar, says the MIT work "illustrates where we really are" in the field. "We are edging toward systems that can roughly place, from pictures of people's faces, where they lie on scales from very positive to very negative, and very active to very passive," he says. "It seems intuitive that the emotional signs one person gives are not the same as the signs another gives, and so it makes a lot of sense that emotion recognition works better when it is personalized. The method of personalizing reflects another intriguing point, that it is more effective to train multiple 'experts,' and aggregate their judgments, than to train a single super-expert. The two together make a satisfying package."

More information: A Mixture of Personalized Experts for Human Affect Estimation: dam-prod.media.mit.edu/x/2018/...ed_final_tYWcW0Y.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

APA citation: Personalized machine-learning models capture subtle variations in facial expressions to better gauge emotion (2018, July 25) retrieved 27 September 2022 from <https://techxplore.com/news/2018-07-personalized-machine-learning-capture-subtle-variations.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.