A picture is worth 1000 words, but how many emotions?

February 6 2015

Log on to Twitter, Facebook or other social media and you will find that much of the content shared with you comes in the form of images, not just words. Those images can convey a lot more than a sentence might, and will often provoke emotions in the viewer.

Jiebo Luo, professor of computer science at the University of Rochester, in collaboration with researchers at Adobe Research has come up with a more accurate way than currently possible to train computers to be able to digest data that comes in the form of images.

In a paper presented last week at the American Association for Artificial Intelligence (AAAI) conference in Austin, Texas, they describe what they refer to as a progressive training deep convolutional neural network (CNN).

The trained computer can then be used to determine what sentiments these images are likely to elicit. Luo says that this information could be useful for things as diverse as measuring economic indicators or predicting elections.

Sentiment analysis of text by computers is itself a challenging task. And in social media, sentiment analysis is more complicated because many people express themselves using images and videos, which are more difficult for a computer to understand.

For example, during a political campaign voters will often share their

views through pictures. Two different pictures might show the same candidate, but they might be making very different political statements. A human could recognize one as being a positive portrait of the candidate (e.g. the candidate smiling and raising his arms) and the other one being negative (e.g. a picture of the candidate looking defeated). But no human could look at every picture shared on social media - it is truly "big data." To be able to make informed guesses about a candidate's popularity, computers need to be trained to digest this data, which is what Luo and his collaborators' approach can do more accurately than was possible until now.

The researchers treat the task of extracting sentiments from images as an image classification problem. This means that somehow each picture needs to be analyzed and labels applied to it.

To begin the training process, Luo and his collaborators used a huge number of Flickr images that have been loosely labeled by a machine algorithm with specific sentiments, in an existing database known as SentiBank (developed by Professor Shih-Fu Chang's group at Columbia University). This gives the computer a starting point to begin understanding what some images can convey. But the machine-generated labels also include a likelihood of that label being true, that is, how sure is the computer that the label is correct? The key step of the training process comes next, when they discard any images for which the sentiment or sentiments with which they have been labeled might not be true. So they use only the "better" labeled images for further training in a progressively improving manner within the framework of the powerful convolutional neural network. They found that this extra step significantly improved the accuracy of the sentiments with which each picture is labeled.

They also adapted this sentiment analysis engine with some images extracted from Twitter. In this case they employed "crowd intelligence,"

with multiple people helping to categorize the images via the Amazon Mechanical Turk platform. They used only a small number of images for fine-tuning the computer and yet, by applying this domain-adaptation process, they showed they could improve on current state of the art methods for sentiment analysis of Twitter images. One surprising finding is that the accuracy of image sentiment classification has exceeded that of the text sentiment classification on the same Twitter messages.

Provided by University of Rochester