Machine Learning by Watching and Listening

October 5 2009

(PhysOrg.com) -- To expand the boundaries of machine intelligence, Ben Taskar is using television shows with large fan bases like CSI, Alias, and Lost to teach computers how to be smarter about what they see, hear and read.

Ben Taskar is teaching computers how to watch television. Not, as you may think, because they need to relax after reading all that code, but because through this research, Taskar, the Magerman Term Assistant Professor in the Department of Computer and Information Science, is taking machine learning to the next level. Using novel learning algorithms combining video, sound and text streams, his team has shown that computers can be taught to associate what is in a video clip with existing descriptions of characters and actions and then infer information about new material and categorize it according to what it has already learned.

Currently, to categorize videos, photos and other electronic media, computers are “told” through assigned tags the contents of an image. Even new “self-tagging” technologies rely on existing labels to tag new media as it is saved. This is not at all similar to how a human learns and infers. For example, when we watch an episode of our favorite show and we hear one character say to the other, “Joe is coming over soon,” we are able to infer when a new character arrives that his name is Joe. We do not need an explicit label of “Joe” over his face or a subtitle “Joe” at the bottom of the screen.

To expand the boundaries of machine intelligence, Taskar and a team comprised of graduate students Timothee Cour and Benjamin Sapp, along with undergraduate Chris Jordan, are using television shows with large fan bases like CSI, Alias, and Lost to teach computers how to be smarter about what they see, hear and read. Take, for example, the show Lost. Hundreds of thousands of viewers enjoy spending hours of their time writing and posting scripts of episodes on fan sites, video clips on YouTube, and information in discussion boards. Taskar is taking this collective "wisdom of the crowds" and entering the massive quantities of digitized knowledge and the associated scenes and clips into computers.

From there, computers are given specialized algorithms to be able to combine the information with the video and "learn" which person is which character, what each character is doing, and with whom. At no time does anyone in the research team tag anything. This is known as "unsupervised" or "weakly" supervised learning.

Once this learning has taken place, researchers can ask the computer, "show all scenes where Kate is talking to Jack," or "produce a montage of all scenes with swimming," and the computer will generate the sequence. By checking on what is produced, the team then looks for patterns containing errors that suggest the algorithms and models need fine-tuning. Once the algorithm is perfected, the computer can then watch new material and add to the already known information, using its past learning to amass more knowledge.

As you can imagine, using algorithms to teach a computer to learn the nuances of language and parts of speech in written data, along with different camera angles, lighting and other filming conditions, is a daunting task. Taskar compares it to how children learn about their environments. At first, a young child may call all moving vehicles with four wheels "cars," and later learns to distinguish "trucks" or "vans" from the group. Similarly, computers are given simpler distinctions and tasks

at the beginning of learning and more and more complicated ones as patterns to “teach” are better identified.

Future applications of this research go far beyond the “cool” factor of being able to get a computer to show all the scenes in which a favorite character appears. Two areas that will likely benefit are general image and audio search. In order to develop more accurate technologies that can robustly recognize and correctly analyze immense collections of images, videos and spoken language, computers will need to learn to identify hundreds of thousands of different concepts. By tapping into contributions of millions of people on the web and burgeoning data from multiple modalities, the research of Ben and his team will push the field of machine learning towards unsupervised techniques to make computers learn about our complex world.
Computers can only take us so far. We still can’t figure out where those Lost writers are going with that island.

Provided by University of Pennsylvania (news : web) Original story can be found here.