DS 760

Machine Learning Systems - Changing Rights Landscape

Rick W. Lentz, M.S., M.S.

5 November, 2017

Introduction

We have had a solid introduction to software intellectual property rights in this course. This background has helped make clear a divide between what is currently understood and protected and what is not due to rapid technological advances.

The starkest contrast and most pressing need are to understand how the advances in machine learning, specifically deep-learning impact our society and intellectual property rights. The paper first covers recent developments, meant to illustrate the contrast that exists between the current rules and the current deep-learning systems. I attempt to apply the tools we learned through our intellectual property readings to these technologies. While I usually stand for open data and source code, I'm questioning this belief due to new awareness and practical understanding of these technologies ability to change our current socio-technical systems.

Beyond Encoding to Remixing

We have systems in place that protect digital media rights. The FBI has done an excellent job helping enforce DMCA rules. Gone are the days of Pirate Bay and torrent sharing of movies or Napster MP3 downloads. Detection technology on YouTube has codified fair use and recognizes a copyrighted audio track placed over a home movie.

With the advent of Generative Adversarial Networks (GAN), we have an entirely new way to encode information that can produce a limitless number of different digital representations that for most people are not discernable from what we believe is real.

This presents a challenge at the moment because nothing about a latent space knowledge representation can be said to represent the original art. (Without reading a paper, skip through some of these computer generated celebrity images:).

I'm curious to know how Richard Stallman will react when open source algorithms gobble up copyrighted media and can produce derivative works at the individual level? Under current copyright, protection is limited to the original form. With this technology, every form can be different, meaning any two digital representations may be different. A hit movie could be remixed with different celebrities, sounds, music, and scenes. How different would it have to be before it is viewed as the original? Could the technology save House of Cards (if Kevin Spacey was 'remixed' out of all prior episodes)?

A Path Beyond Intellectual Property Extensions

So far I have covered a form of encoding to show that visual images and sounds can be reproduced and altered using GANs and this presents challenges to enforcement due to the significant loss of uniqueness in comparison to the original work. But there is something more fundamental that is up for grabs - the right to do a task. Up until recently, metaheuristics or real-world data were used to train deep-learning networks to get superhuman task proficiency.

Google's Deep Mind group released this Nature publication describing how they made two improvements to make the successor to Alpha Go, Alpha Go Zero. Instead of using millions of example games and using hundreds of servers over six months, they used reinforcement learning to have the machine play against itself and trained it on one machine in under 40 days.

Some balked at Google paying $500M for DeepMind for an algorithm that could play many different Atari games well. I am doing cognitive backflips over the awesomeness of this advancement, and I see how truly valuable this improvement is when looking at its application across many different fields.

The entire intellectual property system is set up to allow for the capital return on investment. This technology can accelerate the timeline one would typically expect to have a beneficial advantage in the marketplace. Twenty years ago, several AI experts thought it would take at least a century for a system to beat a human Go player. With a reduction of this benefit of exclusivity protection, companies may opt to leverage this technology for increased competitive advantage.

Right to not be Swamped

We discussed the individual's rights regarding information protections during this section of the course. We specifically learned that what one reveals about themselves to third parties, is not going to be protected. There's a lot of this information, and it seems there may be people interested in using AI to aid in analyzing all this data.

The critical point of interest in this topic was the scope of data required for behavior mining. Can what you reveal about yourself lead to the prediction of criminal behavior? To infer behavior; one would need full-scope access to the same information the individual experiences. It seems these vast stores of data are what we want to collect and mine for all sorts of interesting questions.

Monte Carlo Tree Search and Bayesian Optimization have shown high value in these structured, moderate dimension problem spaces. With today's high flows of information, updating Bayesian networks to calculate new expected behavior probabilities can be done each time the distribution of a factor improves. This may help systems not only detect, but also defuse criminal behavior that results in loss of life.

Summary

I discussed how GANs could encode and transform media to the point where it will be difficult to match to a copyrighted source using traditional methods.

There have been advances in applications of reinforcement learning that will likely challenge our intellectual property rights system further than we may imagine, including to the point where we have the right to perform a task. Obtaining superhuman skill at the Chinese game Go was achieved in 1/10th of the time using only 2% of the computing resources that were required was just a few months earlier.

I found sources that may be looking to use AI for behavior mining. Monte Carlo Tree Search and Bayesian optimization methods are useful for structured problems with moderate complexity.