Getting your TV to understand you better

September 5 2018

Credit: CC0 Public Domain

New research out of the University of Waterloo has found a way to improve the voice query understanding capabilities of home entertainment platforms.

The research, in collaboration with the University of Maryland and Comcast Applied AI Research Lab, uses artificial intelligence (AI) technology to achieve the most natural speech-based interactions with TVs to date.

"Today, we have become accustomed to talking to intelligent agents that do our bidding—from Siri on a mobile phone to Alexa at home. Why shouldn't we be able to do the same with TVs?" asked Jimmy Lin, a professor at the University of Waterloo and David R. Cheriton Chair in the David R. Cheriton School of Computer Science.

"Comcast's Xfinity X1 aims to do exactly that—the platform comes with a 'voice remote' that accepts spoken queries. Your wish is its command—tell your TV to change channels, ask it about free kids' movies, and even about the weather forecast."

In tackling the complex problem of understanding voice queries, the researchers had the idea to take advantage of the latest AI technology—a technique known as hierarchical recurrent neural networks—to better model context and improve the system's accuracy.

In January 2018, the researchers' new neural network model was deployed in production to answer queries from real live users. Unlike the previous system, which was confused by approximately eight per cent of queries, the new model handles most of the very complicated queries appropriately, greatly enhancing user experience.

"If a viewer asks for 'Chicago Fire,' which refers to both a drama series and a soccer team, the system is able to decipher what you really want," said Lin. "What's special about this approach is that we take advantage of context—such as previously watched shows and favourite channels—to personalize results, thereby increasing accuracy."

The researchers have started work on developing an even richer model. The intuition is that by analyzing queries from multiple perspectives, the system can better understand what the viewer is saying.

The paper, Multi-Task Learning with Neural Networks for Voice Query Understanding Entertainment Platform, was presented at the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining held recently in the United Kingdom. The research was undertaken by Jinfeng Rao, a Ph.D. graduate from the University of Maryland, his advisor Lin, and mentor Ferhan Ture, a researcher at Comcast Applied AI Research Lab.

Provided by University of Waterloo