Edge Computing and Artificial Intelligence

Really brief dive for CCPS 721

Ilkka Kokkarinen, Ryerson University

Servers and terminals

- Before internet, all computations were performed locally
- Computer networks enabled remote computations
- Other extreme of shared central server used through dumb terminals
- What is the best division of labour between the terminal and server?
- Answer depends on network latency versus cost of computation

Cloud computing

- Sufficiently fast networks enable cloud computation in real time
- From human user POV, all computers act together as one
- Files and other resources available everywhere
- Still maintain conceptual distinction between "computing devices" versus other artificial things that are not "computing devices"
- Eventually reach ubiquitous computing where even this distinction vanishes

Headless ubiquity

- Internet of Things consists of things talking to other things
- Sometimes these things have to make decisions that radiate to far reaching consequences
- Artificial Intelligence as mechanization of applied Decision Theory
- Somebody has to make the decision based on the available data
- Some decisions can be made locally

Principle of Localism

- Every decision should be made by the lowest-level entity that is competent and has enough information to make that decision
- Example of security video camera that is supposed to do something when something significant changes in the image
- It would be silly to stream 24/7 hi-res video feed for the central server that will then simply throw away most of that feed
- Some decisions must be immediate, such as self-driving car

Yesterday's supercomputers

- Hardware improvements add more processing power on the edge
- Powerful enough to execute pre-trained neural network or other model
- Also powerful enough to perform new machine learning themselves
- Edge devices communicating with other edge devices create a form of swarm intelligence

Edge analytics

- Analytical computation performed on raw data at the sensor level
- Filter and preprocess data to make data sent to cloud less noisy
- Bandwidth and latency gains
- Computation distributed on the edges is by its nature perfectly scalable
- Analytics can be customized and trained for needs of each sensor

Deep learning... on the edge!

- Distributed training of ML models on the edge
- Federated learning allows each edge node to update its own model
- Suitable for traffic lights and traffic cameras in a smart city
- Local decisions propagate to neighbours
- Distributed training also alleviates some security and privacy issues