Towards Automated and Trustworthy Machine Learning

Minhao Cheng
Assistant Professor, CSE @ HKUST
minhaocheng@ust.hk
Machine Learning Pipeline

SQL DB
Web Data
Cloud sourcing
Datawarehouse
...

Prepare Data
Build & Train
Deploy
The devil is in the details

• What feature
  - Constraint/Rule
  - Budget
  - Efficiency

• Which model
  - Linear model
  - Boosting model
  - Neural network

• Which parameter
  - Hyperparameter
  - Optimizer
Automated Machine Learning

- **AutoML** simplifies each step in the machine learning process, from handling a raw dataset to deploying a practical machine learning model.
Model Matters!

![Graph showing the impact of different models on accuracy and operations]

- **Top 1 accuracy [%]**
  - AlexNet
  - BN-AlexNet
  - ENet
  - GoogleNet
  - ResNet-18
  - VGG-16
  - VGG-19
  - ResNet-34
  - ResNet-50
  - ResNet-101
  - ResNet-152
  - Inception-v3
  - Inception-v4

- **Operations [G-Ops]**
  - 5M
  - 35M
  - 65M
  - 95M
  - 125M
  - 155M

- **Networks**
  - Inception-v3
  - Inception-v4
  - ResNet-50
  - ResNet-101
  - ResNet-152
  - VGG-16
  - VGG-19

- **Models**
  - BN-NIN
  - ENet
  - GoogLeNet
  - BN-AlexNet
  - AlexNet
Automated Machine Learning

- AutoML: simplifies each step in the machine learning process, from handling a raw dataset to deploying a practical machine learning model.
  - Neural Architecture Search (NAS) [ICLR 21’ECCV 21’]
Automated Machine Learning

• AutoML:
  • Neural Architecture Search (NAS)
  • Hyperparameter optimization (HPO)
  • Meta learning and Learning to learn
  • Automated Reinforcement learning
  • AutoML in Physical World
  • ...
Beyond Accuracy

Researchers trick Tesla Autopilot into steering into oncoming traffic

Syrian hackers claim AP hack that tipped stock market by $136 billion. Is it terrorism?

Microsoft silences its new A.I. bot Tay, after Twitter users teach it racism [Updated]
Trustworthy ML

• Not alchemy
  • Explainability
  • Robustness
• Security
• Privacy
• Fairness
• …

• Establish model understanding
Accuracy ≠ Robustness

• Solely pursuing for high-accuracy AI model may get us in trouble…

• We have established a toolbox to evaluate the robustness of machine learning models [ICRL 19’20’ AAAI 19’20’]
Model understanding

• Debugging
• Bias detection
• Provide recourse to individuals who are adversely affected by model predictions
• Assess if and when to trust model predictions
Thank you!