Introduction to Reinforcement Learning and Contextual Bandits

Rajan Chari
Microsoft Research

Agenda

Adapting to the changing world

Introduction to reinforcement learning and contextual bandits

Personalizer overview

Apprentice mode

Making contextual bandits work in practice

Q&A

Supervised Learning

- Labeled examples
- Supervised Learning Algorithm
  - Classification
    - DNN
    - Logistic Regression
    - Boosted Trees
    - Support Vector Machines
  - Regression
- Trained Model
- Deploy Model & Inference

Supervised Learning

Problems:

- No labeled data
- No right answer
- Non-stationary world

Reinforcement Learning

- World provides a context
- Interact with the world (explore/exploit)
- Get feedback (reward)
- Learn from it (context, action, reward, …)
- Repeat

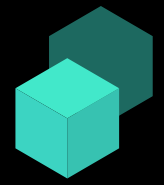
Reinforcement learning with Contextual Bandits

- Contextual Bandit
  - Choose among n actions
  - Cannot know the reward for all actions. Only the chosen action

  vw -d train.dat --cb <k actions>

  (demo)

Reinforcement learning with Contextual Bandits



- Contextual Bandit w/ Exploration
- Simple exploration strategy – epsilon greedy
- Others
  - bag
  - cover
  - softmax

  vw -d train.dat --cb_explore <k actions> --epsilon 0.2
  (demo)

Reinforcement learning with Contextual Bandits

- Contextual Bandit w/ Exploration, Context and variable actions
  - Present context about the world
  - No need to specify number of actions
  - Still must specify exploration strategy

  vw -d train.dat --cb_explore_adf --epsilon 0.2

  (demo)

Evaluating different RL methods

- Supervised learning – ROC
- Reinforcement Learning - CFE

Questions?

- https://github.com/VowpalWabbit/vowpal_wabbit
- https://vowpalwabbit.org/tutorials/contextual_bandits.html