Applying TD Learning to Machine Learning

Reinforcement Learning (COMP 767 - McGill - Winter 2017)
April 20, 2017 -- Philippe Lacaille

Shameless self-promotion -> placailleblog.wordpress.com

The setup

Supervised ML

- ➔ Features / target pairs are the norm
- ➔ Error back propagated based on prediction and target
- ➔ Efficient in *single-step prediction problems*

TD learning in RL

- ➔ Series of episodes experienced by agent
- ➔ Error for updating V/Q values based on TD error
- ➔ Typically *multi-step prediction problems*

TD learning in ML?

- ➔ Consider ML *multi-step prediction problems*
- ➔ Sequence of features observed
- ➔ Could make use of the TD error in supervised learning?

Typical ML approach

Applied to multi-step predictions

$$w \leftarrow w + \sum_{t=1}^{T} \Delta w_t$$

$$\Delta w_t = \alpha(y - P_t)\nabla_w P_t$$

- ➔ Observation at each time step in sequence is a set of features
- ➔ Outcome is only known after sequence
- ➔ MSE gradient update rule

Typical ML approach

$$\Delta w_t = \alpha(y - P_t)\nabla_w P_t$$

- At each time step
  - Store the gradient
  - Store the prediction
- Computations pushed to end of sequence
  - Compute error for each time step
  - Compute weight updates
  - Make update

TD learning approach

Applied to multi-step predictions

$$w \leftarrow w + \sum_{t=1}^{T} \Delta w_t$$

$$\Delta w_t = \alpha(P_{t+1} - P_t) \sum_{k=1}^{t} \nabla_w P_k$$

- ➔ Typically used to predict value of state and/or action
- ➔ Equivalent weight updates to traditional ML approach (derived with dark magic)
- ➔ No need to store the predictions, replace by the sum of gradients
- ➔ Computation load is spread during sequence

What problems could this be applied to?

- Monthly predictions for the end-of-year financial results of a company
- Hourly predictions for rain level on the upcoming Saturday
- Rating prediction of a movie during watch time

My experiments based on MNIST dataset

  - Consider MNIST as sequence of pixels
  - Empirically confirmed same updates under both methods
  - Same results with intra-sequence updating
  - No promising results with TD(0)