TD and Markov Chain Modelling for Time-series Prediction

COMP 767 Final Project Presentation

Charles C Onu

260663256

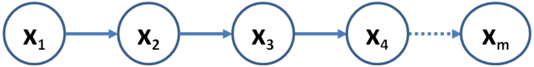
20 April, 2017

Objectives

1. To formulate and apply TD to the task of predicting an outcome from discrete time-series data

$$x_1^1, x_2^1, \dots, x_m^1, y^1$$
$$x_1^2, x_2^2, \dots, x_m^2, y^2$$
$$x_1^n, x_2^n, \dots, x_m^n, y^n$$

2. To investigate the relationship between TD and Markov models in the context of time-series/sequence prediction tasks



3. Experiment on
   - a synthetic example and
   - real-world data

Relation to Model-free RL

Sequence prediction is similar to the model-free reinforcement learning setting

- Similarity
  - Dynamics of underlying system is not known
  - We have actual example trajectories $x_1, x_2, ..., x_m$, and the final outcome y
- Difference
  - In RL: trying to figure how best to act in the system to accumulate 'good' outcomes (active learning)
  - In Supervised Learning: Given a trajectory already taken, what is the final outcome/reward - win or lose? rain or sun, survive or die

TD Learning

Learning - TD(0)

- For t = 1 to m:

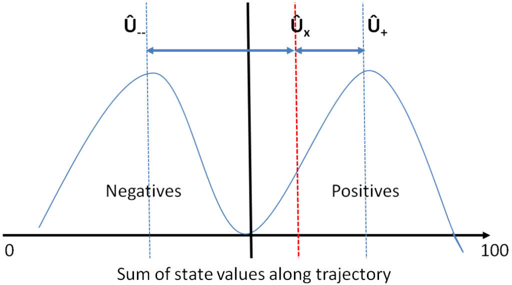
$$V(x_t) = V(x_t) + \alpha(r + \gamma V(x_{t+1}) - V(x_t))$$

- At all non-terminal time steps t reward is 0. At terminal time step m reward is -1 or 1 depending on outcome.

Prediction

- For a new sequence C, A, C, B, C, we define a quantity U, the sum of the state values: $U = 3V(C) + V(A) + V(B)$
- How do we predict the class from this real number? What is the threshold?

TD Prediction



- Goal is to find the threshold that maximises the distance between the centres of the 2 distributions.
- Given a new sequence x, take the absolute distances:

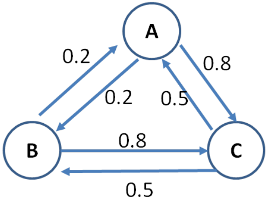
$$d_{+} = |U_x - \hat{U}_{+}|$$

$$d_{-} = |U_x - \hat{U}_{-}|$$

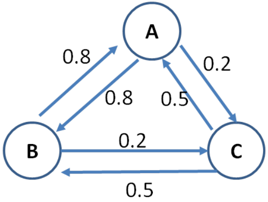
- Select the class c for which

$$\arg\min_{c} d_c$$

Synthetic Experiment: an ergodic Markov chain



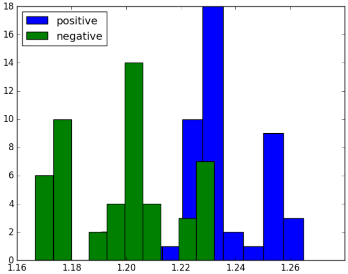
positive



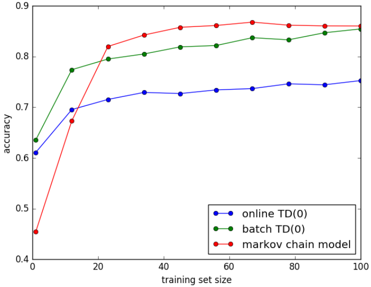
negative

- Generates sequences of 5 time steps. E.g., A, C, B, C, A - 1
- Positive and negative examples are generated by different transition probabilities
- No dwell time in states (always transitions to another state)
- 100 examples generated for training. 50 generated for test.

Synthetic Experiment: Results



(a) Histogram of cumulative values of 100 training examples



(b) Accuracy vs training set size

"batch TD(0) always finds the estimates that would be exactly correct for the maximum likelihood model of the Markov process" - Sutton and Barto, 2016

Real-world Experiment: Prediction extubation readiness

Respiratory States

1. Pause (PAU)
2. Asynchronous breathing (ASB)
3. Movement Artifact (MVT)
4. Synchronous breathing (SYB)
5. Unknown (UNK)

Table 1: Learned values for each respiratory state using $\gamma = 1$, $\alpha = 0.01$

States	PAU	ASB	MVT	SYB	UNK
Learned Values	0.61	0.62	0.60	0.63	0.61

TD: Sensitivity = 0.60; Specificity = 0.52
MLE Markov chain: Sensitivity = 0.71; Specificity = 0.52

Real-world Experiment: Results

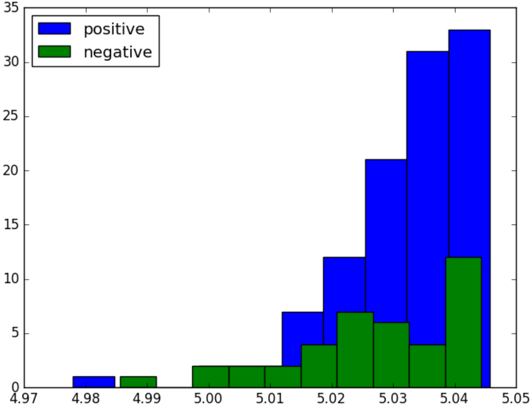


Figure 2: Histogram of cumulative values of 100 training examples

Summary

- Goal was to determine how well TD would perform in a truly Markov system.
- We compare model-based (Markov chain modelling) vs model-free TD(0) learning.
- Batch TD(0) performed nearly as well as explicit Markov chain modelling. It may do just as well:
  - if we tune α (alpha was simply set to 0.01),
  - repeat presentation of each example more times (currently 1000 repeats)
  - learn better discriminator for the returns
- This suggests that indeed TD is approximating the Markov chain that defines the underlying system.
- TD (both online and batch) does better with little data than explicit Markov chain modelling. For more complex systems (e.g., larger state space) the training set size over which TD performs better could be larger.

References

- R. S. Sutton, "Learning to Predict by the Method of Temporal Differences" in Machine Learning 3, Boston:Kluwer Academic Publishers, 1988.
- E. Barnard, "Temporal-difference methods and Markov models" IEEE Transactions on Systems, Man, and Cybernetics, 23(2), 357–365, 1993.
- R.S. Sutton, A. G. Barto,"Introduction to Reinforcement Learning," MIT Press, Cambridge, MA, USA, 1998.