# Time-series Prediction via the method of Temporal Differences

**Charles C Onu** charles.onu@mail.mcgill.ca
*School of Computer Science*
*McGill University*

## Abstract

This work applies temporal difference (TD) learning to the task of predicting an outcome from sequence data or a time series of observations in time. We specifically formulate a scheme for classifying sequences given some learned state values of a discrete state system. We applied this to a synthetic, truly Markov system and show that batch TD learning gives almost exactly the same performance as prediction from the maximum likelihood estimate of an underlying Markov chain. We further implemented the prediction algorithm to the task of classifying preterm newborns as succeeding or failing extubation based on 5-minute sequence of respiratory patterns.

## 1. Introduction

The task of classifying time series data is both similar to and different from the model-free reinforcement learning (RL) problem. They are different in that in RL the goal is to teach an agent to act optimally in an environment, whereas in the supervised learning task of predicting from time series data, the objective is to determine a final outcome given a trajectory already taken. Both cases are similar in that the dynamics of the underlying system are not known but can be implicitly learned through methods such as temporal difference learning. Over the years, several researchers have shown that TD learning is not restricted to this specific case of "learning to act" but can be a general method for learning from temporally changing sequence of observations (aka time-series data) (Sutton, 1988).

We approach the problem of predicting extubation readiness in preterm newborns. Infants who are born before 28 weeks usually require breathing support via *intubation* and mechanical ventilation. The newborn is monitored under this system until a time when an attending clinician deems it fit for *extubation*. Extubation (removal of the tube) is a critical decision for clinicians. Early extubation could lead to the suffocation and death (or disability) of the newborn. Whereas delayed extubation could result to the development of broncho-pulmonary dysplasia (BPD) or chronic lung disease.

In this work, we explore the application of TD learning to predict the readiness of an infant for extubation based on a discrete sequence of breathing patterns (or respiratory states). The 5 breathing patterns considered were: Pause, Asynchronous breathing, Movement artifact, Synchronous breathing and Unknown. Due to the complex nature of the actual data, we first designed a synthetic experiment to design and test the algorithms for classification. The experiment not only showed the data efficiency of TD in classifying sequences but also confirmed previous finding that "batch TD(0) always finds the estimates that would be exactly correct for the maximum likelihood model of the Markov process" (Sutton and Barto, 1998).

Our experiments applying TD to the actual problem showed that the two groups of patients may be indistinguishable based solely on the sequence of respiratory patterns in this 5 minute period of SBT. This is consistent which previous work which makes prediction based on maximum likelihood estimates of a Markov chain modeled to the data.

The subsequent sections are broken down as follows: Section 2 provides brief details about the data used in the work. Section 3 describes related work. Section 4 describes our methodology. Description of experiments and results are shown in Section 5. In the final Section 6, a discussion on the significance of our results is provided as well as paths for future directions of this work.

## 2. Cohort

Data from a total of 186 infants was available for this study. The data was collected from sites in Canada and the USA: Royal Victoria Hospital, Montreal Childrens Hospital, Glen site of the Montreal Childrens Hospital, Quebec, Detroit Medical Center, MI, and Women and Infants Hospital of Rhode Island, RI. Ethical approval was obtained from the boards at all institutions, and informed parental consent was obtained before recruiting each patient.

In the following sub-sections we provide brief details about the patient and data collected that is relevant to this present work. Readers interested in full data collection procedure for this study are referred to Precup et al. (2012) and Robles-Rubio et al. (2015).

### 2.1 Cohort Selection

Eligible infants were of gestational age $\leq 28$ weeks, birth weight $\leq 1250$g, and undergoing endo-tracheal tube with mechanical ventilation (ETT-MV) at time of recruitment. Infants were excluded if they had any major congenital anomalies such as heart disease, or were receiving any vasopressor or sedative drugs at the time of extubation.

### 2.2 Data Acquisition

Respiratory signals were measured using Respiratory Inductance Plethysmography (RIP) bands placed around the infant's ribcage and abdomen. Signals were acquired for a 5-minute period of spontaneous breathing trial prior to extubation at a sampling frequency of 1000Hz.

### 2.3 Respiratory Patterns

RIP signals were analyzed (to 50Hz) using Automated Unsupervised Respiratory Event Analysis (AUREA), which extracts sample-by-sample metrics of respiratory power, synchrony between RCG and ABD, and movement artifact (Robles-Rubio et al., 2011). AUREA uses k-means clustering to assign each sample to one of the following respiratory patterns:

- Pause (PAU): A cessation of breathing.
- Synchronous Breathing (SYB): RCG and ABD are in phase.
- Asynchronous Breathing (ASB): RCG and ABD are out of phase.

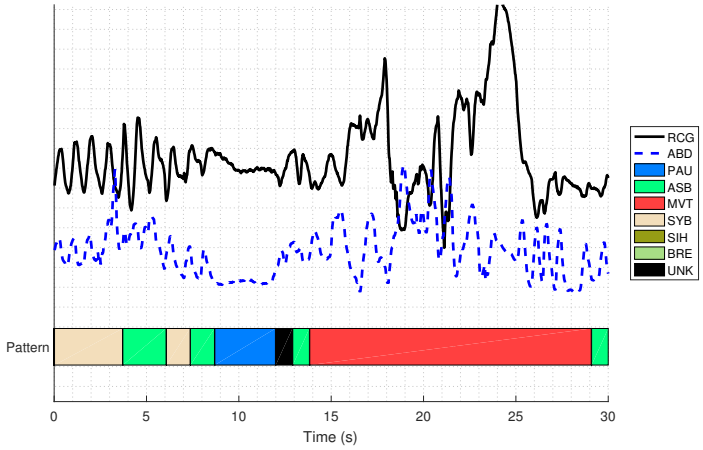


Figure 1: Example of a RCG and ABD signal segment and the corresponding respiratory patterns computed by AUREA.

- Movement Artifact (MVT): Associated with infant moving or nurse handling.
- Unknown (UNK): Ambiguous patterns not belonging to any other pattern category.

AUREA provides repeatable results with no human intervention. An example of RIP signals and corresponding patterns assigned by AUREA to the different samples is shown in Fig. 1.

## 3. Related Work

Sutton (1988) first presented TD learning as method for making predictions from sequence data. In the work, the relationship between TD and Markov models was also introduced, demonstrating that the faster learning and convergence of TD learning over the standard supervised learning paradigm are guaranteed only when the underlying system generating the observed data is Markov. In Sutton and Barto (1998) it was shown that "batch TD(0) always finds the estimates that would be exactly correct for the maximum likelihood model of the Markov process". Barnard (1993) not only showed the correlation between temporal difference methods and first-order Markov but illustrated a generalisation to higher-order Markov models.

## 4. Problem Formulation

### 4.1 Temporal Difference (TD) Learning

Consider an observation sequence $x = x_1, x_2, ..., x_t, ..., x_m$, which culminates in some outcome $y$ (-1 or 1) and where each $x_t$ is some discrete state representation of the observation at time step $t$. The value of each state can be learned using TD as follows:

For t = 1 to m:

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

where $r$ is the reward obtained at each time step (typically 0 at all time steps and $y$ at the last time step), $\gamma$ is the discount factor $\in [0\ 1]$ and $V(x_t)$ is the value of the state observed in $x_t$. In the batch TD(0) algorithm, equation (1) is repeatedly presented with every sequence in the dataset until the values $V$ for each state do not change beyond a small threshold.

### 4.2 Classification from Learned State Values

After the state values have been learned, there is the question of how to use them to classify a new example sequence. In the case of function approximation with continuous state values, it has been proposed (Sutton, 1988) to use the prediction on $x_m$ i.e., the feature at the last time step to make a classification. Whereas this makes sense for continuous states, it is not quite an adequate discriminatory measure for discrete state sequences, since a lot of information about a sequence's history will thrown away. Moreover, when sequences of 2 classes are differentiated not by the last state they visit before termination, but by relative duration spent in certain 'good' states, then classifying based on last visited state will prove futile.

We designed a new method for classifying sequences of discrete states using the learned state values. Fist, we defined the value of a sequence, $U$ as the average of the values of all states encountered along the sequence:

$$U_x = \frac{1}{m}[V(x_1) + V(x_2) + V(x_3) + ... + V(x_m)] \tag{2}$$

Given the sequence value $U_x$ of a new sequence $x$, how then do we classify it as belonging to the positive or negative class. This is another learning problem. We can calculate the sequence values of all positive examples, the sequence values of all negative examples and then learn a discriminating (linear or non-linear) threshold. In this example, we define a linear threshold which maximises the distance between the centers of the 2 distributions. Given $\hat{U}_+$, the center of the sequence value distribution for positive examples and $\hat{U}_-$ for negative examples, the absolute distances from $x$ can be obtained:

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

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

To classify, one simply selects the class $c$ which minimises the distance:

$$\arg\min_c d_c$$

This method is illustrated in Figure 2. Note that the centers could be taken as mean or median depending on the spread of the distribution.

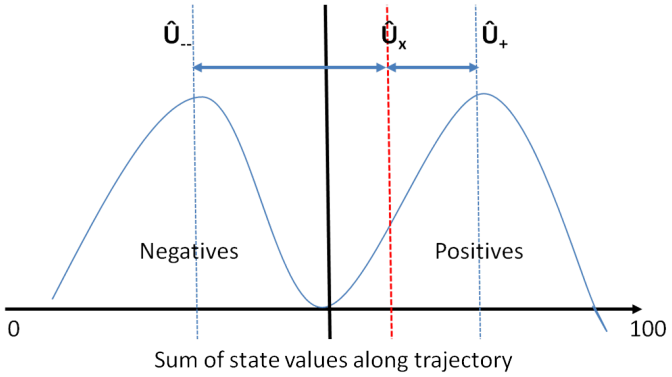


Figure 2: Theoretical probability density function of the sequence values of a set of example sequences. A new example $x$ with sequence value $U_x$, can be assigned to the class whose center it is closest to. This is equivalent to learning a maximal margin linear decision boundary.

## 5. Results

In the next 2 sub-sections, we show results on experiments applying TD to a synthetic example and then to classify newborns that may succeed or fail extubation based on respiratory states.

### 5.1 TD and Markov models: Toy Experiment

Consider a system, illustrated by the state transition diagrams in Figure 3, which generates sequences of two classes. Sequences classified as positive examples are generated by the state machine on the left while negative examples are generated by the one on the right.The system generates sequences of 5 time steps. Every time step results to a change of state (i.e., no dwell time). The primary difference between the two state machines is that the probability of going to state $C$ is 0.8 for positive examples and 0.2 for negative example.

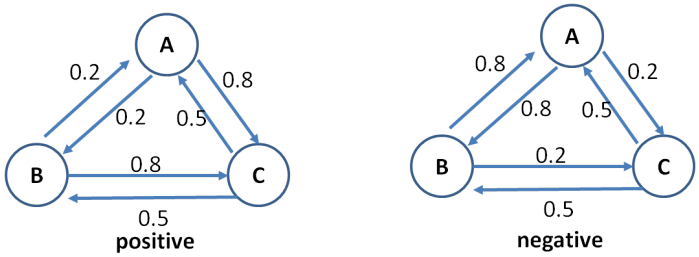


Figure 3: State transition diagrams for a system which generates sequences of 5 time steps which may be of a positive or negative class

As a practical motivation for this example, one could consider the 2 state diagrams as approximations of the respiratory system of infants with 2 different clinical outcomes. And

Table 1: Learned values for each state in toy experiment after using batch TD(0) to learn on a training set size of 100 sequences.

| **States** | A | B | C |
|---|---|---|---|
| **Learned Values** | 0.48 | 0.49 | 0.53 |

that the regular visitation of a certain respiratory state $C$ is an indication of those who will have positive outcome.

To execute the experiment, 50 examples (25 from each class) were generated and kept aside for testing. Training was conducted using progressively increasing training set sizes from 1 to 100 in increments of 10. For each size, 1) Batch TD was used to learn the values and make predictions on the test set (as described in sections 4.1 and 4.2), and 2) the maximum likelihood estimate (MLE) of the Markov chain was obtained and used to make predictions on the test set by assigning the class that maximises the posterior probability of a sequence.

The result (averaged over 100 repeats of the experiment) is shown in Figure 4. Online TD(0) was also run and is shown for reference. It can be seen that Batch TD(0) performs just as well as explicit modeling of the Markov chain, and performs better for training set sizes between 1 and 20. In Figure 5, we show a histogram of the average (learned) values $U$ for the positive and negative training examples in the TD experiment. The linear separation of the values of both classes can been in the bi-modal distribution formed. An examination of the state values learned (Table 1) also indicates the intuitive nature of the learned state values. The state $C$ is accorded the highest value and $A$ and $B$ are roughly equal.

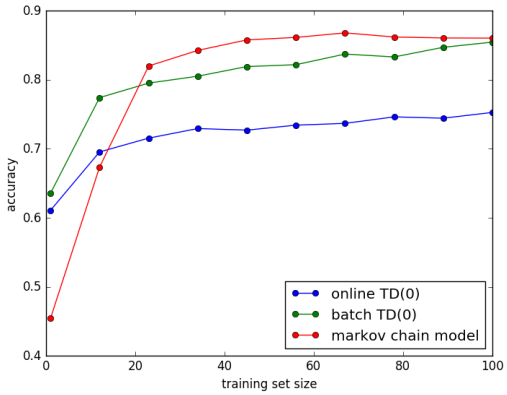


Figure 4: Learning curve showing accuracy as training set size is increased. Result is shown for training with online TD(0), batch TD(0) and maximum likelihood Markov chain modeling

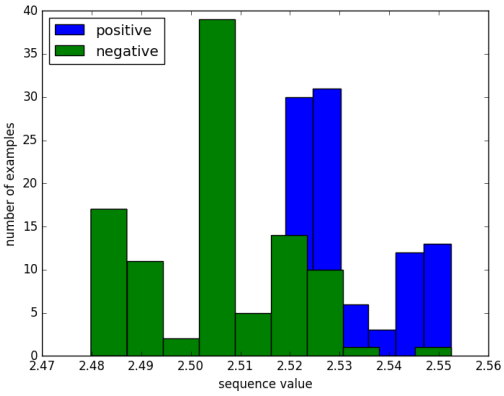


Figure 5: Histogram of sequence values of positive and negative examples based on learned state values, in synthetic experiment

### 5.2 TD for Predicting Extubation Readiness

The 186 patients' data obtained for this study contained 136 successes and 50 failures. Each of the 5 respiratory patterns were taken as a state. We applied TD as described in sections 4.1 and 4.2 for learning state values (from training set) and classifying (on the test set). The final reward for success and failure sequences was set to 1 and -1, respectively. Also, because each sequence was quite long ( 15000 samples = 5mins * 60 * 50Hz), a small reward of 1e-5 was used at every time step to ensure that terminal state reward could be propagated efficiently down the chain. The learning rate $\alpha$ was fixed at a sufficiently small value 0.01.

#### 5.2.1 Result Using Full Markov Chain

The discount factor $\gamma$ was varied from 0 to 1. Leave-one-out cross-validation was applied. Due to the imbalance in the dataset, sensitivity (fraction of failure patients correctly identified) and specificity (fraction of success patients correctly identified) measures were reported to provide a complementary perspective of performance on the 2 groups of patients. All values of $\gamma$ gave the exact same performance on sensitivity and specificity **0.53** and **0.54**, respectively. This was because rewards at non-terminal states was 0 (1e-5), so discounting simply scaled the learned values but had no impact on their relative proportions. The learned values for each state for $\gamma = 1$ under batch TD are shown in Table 2. The values are roughly equal with the Unknown (UNK) state having the highest value.

Table 2: Learned values for each respiratory state after training using Batch TD(0) on full Markov chain

| **States** | PAU | ASB | MVT | SYB | UNK |
|---|---|---|---|---|---|
| **Learned Values** | 0.6239 | 0.6292 | 0.6294 | 0.6257 | 0.6314 |

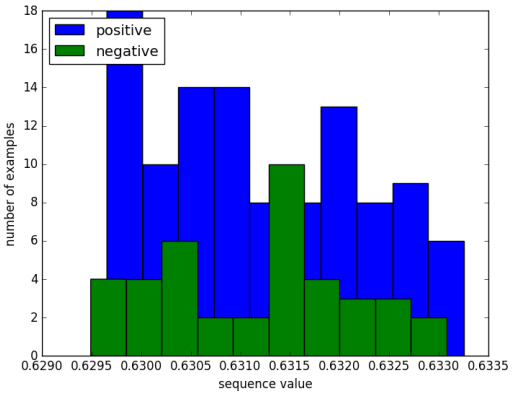


Figure 6: Histogram of sequence values of positive (success) and negative(failure) examples based on learned state values, in real experiment using full Markov chain

In Figure 6, the distribution of learned sequence values of the examples in the training set are shown. It can be seen that the sequence values for success and failure examples are distributed similarly, explaining the relatively low classification performance.

#### 5.2.2 Result Using Semi-Markov Chain

The experiment described in the previous section was repeated, this time using a semi-Markov chain. In particular, each example sequence was processed to remove dwell times lasting for more than one time step. This ensured that all transitions along a sequence was to a state different from the previous. The Learned state values for each respiratory state are shown in Table 3. Generalisation evaluated using leave-one-out cross validation gave a sensitivity of **0.60** and specificity of **0.60**. This was a increase in performance from the use of the full Markov chain. The distribution of sequence values (Figure 7) was found to take a more Gaussian form. Though, the sequence values for both success and failure examples were still distributed within the same range.

## 6. Discussion

First, we designed a synthetic experiment. Our goal was to define a good prediction method from TD-learned state values and to determine how well TD would perform in a truly Markov system. We used as benchmark predictions obtained by fitting the maximum likeli-

Table 3: Learned values for each respiratory state after training using Batch TD(0) on semi-Markov chain

| **States** | PAU | ASB | MVT | SYB | UNK |
|---|---|---|---|---|---|
| **Learned Values** | 0.4849 | 0.4913 | 0.4946 | 0.5067 | 0.4927 |

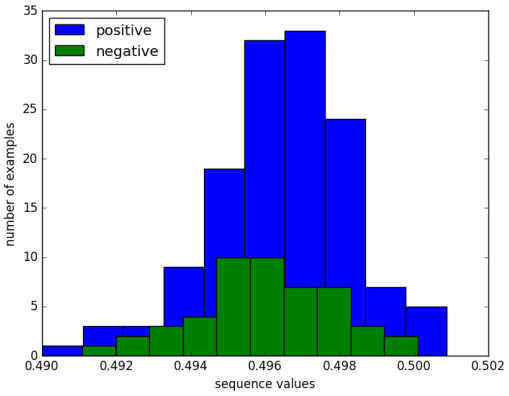


Figure 7: Histogram of sequence values of positive (success) and negative(failure) examples based on learned state values, in real experiment using *semi-Markov chain*

hood model of the Markov chain. Results show that batch TD(0) performed almost as well as explicit Markov chain modeling on larger training set sizes. Performance may improve if we tune the learning rate $\alpha$ (was simply set to 0.01) and/or repeat presentation of each example more times (currently fixed at 1000 repeats). The results nevertheless suggest that indeed TD is approximating the Markov chain that defines the underlying system. Additionally, both online and batch TD perform better with little data than explicit Markov chain modeling. This buttresses the data efficiency property of TD methods. We suspect that for more complex systems (e.g., larger state space) the training set size over which TD performs better would be larger.

Secondly, we applied TD to our problem of classifying infants as ready or not for extubation based on respiratory state sequences. The performance was quite low and almost at the level of a random classifier. A look at the distribution of sequence values (Figure 6) shows that the learned values for both success and failure sequences are distributed within the same range.

When a semi-Markov chain of the sequences was used as input to the TD algorithm, the performance improved to a level equivalent to previous work in which we performed classification based on maximum likelihood models of the Markov chain. The distribution of sequence values for both success and failure patients was found to take a near-Gaussian form. It is not immediately clear why this was the case. The sequence values were however still distributed across the same range.

Future work may involve experimenting with multi-step algorithms like TD($\lambda$). However, present results suggest that these 2 groups of newborns are not distinguishable solely on the basis of the respiratory pattern sequences during the 5-minute period of spontaneous breathing trial.

## References

Etienne Barnard. Temporal-difference methods and markov models. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(2):357–365, 1993.

D. Precup, C. A. Robles-Rubio, K. A. Brown, L. Kanbar, J. Kaczmarek, S. Chawla, G. M. Sant'Anna, and R. E. Kearney. Prediction of extubation readiness in extreme preterm infants based on measures of cardiorespiratory variability. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5630–5633, Aug 2012. doi: 10.1109/EMBC.2012.6347271.

C. A. Robles-Rubio, K. A. Brown, and R. E. Kearney. Automated unsupervised respiratory event analysis. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3201–3204, Aug 2011. doi: 10.1109/IEMBS.2011.6090871.

CA Robles-Rubio, J Kaczmarek, S Chawla, L Kovacs, KA Brown, RE Kearney, and GM Sant Anna. Automated analysis of respiratory behavior in extremely preterm infants and extubation readiness. *Pediatric Pulmonology*, 50(5):479–486, 2015. ISSN 1099-0496. doi: 10.1002/ppul.23151. URL `http://dx.doi.org/10.1002/ppul.23151`.

Richard S Sutton. Learning to predict by the methods of temporal differences. *Machine learning*, 3(1):9–44, 1988.

Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge, 1998.