Problem Set #2: Quantitative Models of Behavior

Hsieh Yu-Guan

March 18, 2017

Introduction

In this report we will work on several classic behavior models. The topics that we will cover include conditional learning (classical conditioning and operant conditioning), decision making and reinforcement learning.

1   Rescola-Wagner model

1.1   Model description

Let's start with the Rescola-Wagner model. It's a model of classical conditioning in which learning signifies building associations between conditioned (CS) and unconditioned (UCS) stimuli. UCS are often represented in form of rewards like food (for an animal) or money (for a person) while CS are some kinds of neutral stimuli that may or may not allow us to predict the occurrence of this reward.

In a single trial, the presence or absence of the reward is respectively denoted by $r = 0$ or $r = 1$. More than one CS can be taken into account, then the presence of the ith stimulus is denoted by $u^{(i)} = 1$ and its absence by $u^{(i)} = 0$. The animal's prediction v is given by the formula

$$v = \sum_{i=1}^{m} w^{(i)} u^{(i)}$$

where m is the number of CS, and for each i, $w^{(i)}$ is the prediction parameter associated with $u^{(i)}$. If we note $u = (u_1, ..., u_m)$ and $w = (w_1, ..., w_m)$, it can also be written in the form

$$v = w \cdot u$$

where $\cdot$ denotes the scalar product. We can now calculate the predition error $\delta = r - v$ which, with the learning rate ϵ, allows us to write down the update rule for every $w^{(i)}$ after the trial at time t

$$w_{t+1}^{(i)} = w_t^{(i)} + \epsilon \delta_t u_t^{(i)}.$$

1.2   A simple test

To examine the reliability of the model, we'll first look at a very simple experiment. Only one stimulus is considered, so we identify for the moment u with $u^{(1)}$ and w with $w^{(1)}$. Let's assume that in the first 25 trials, both stimulus and reward are present, and during the next 25 trials, only the stimulus is present. The plot of r and u is shown in the figure below.
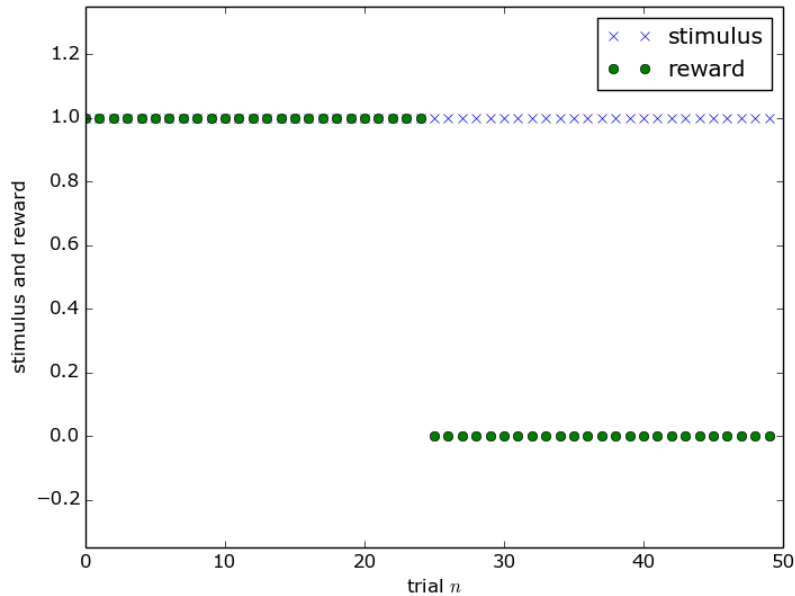

FIGURE 1: A very simple experiment

Now if we do the simulation with the learning rate $\epsilon = 0.1$, the value of w evolves as shown (since u equals always 1 here, we have also all the time $v = w$).
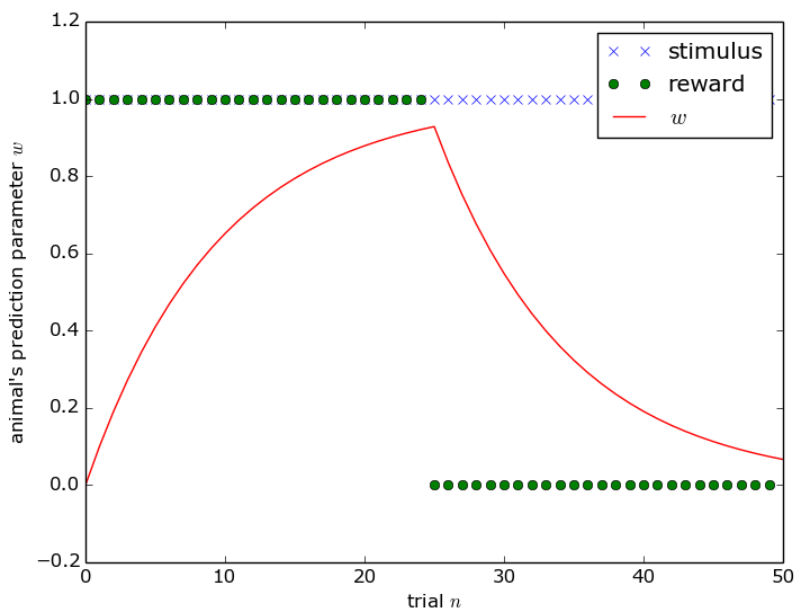

FIGURE 2: Evolution of w in this simple experiment ($\epsilon = 0.1$)

We see an exponential rise of w to value 1 during the first 25 trials and its value decays exponentially to 0 for the rest of the experiment. It's rather reasonable intuitively, and from

a mathematical point of view, we write simply $w_{t+1} - w_t = \epsilon \delta_t u_t$. We know that $\delta_t = r_t - w_t$ (remember that $v_t = w_t$ here) and $u_t = 1$ for all t. For the first 25 trials, we get $w_{t+1} - w_t = \epsilon(1 - w_t)$, so if we put it in a continuous form, it becomes

$$\frac{dw}{dt} = \epsilon(1 - w).$$

We now just need to solve this differential equation to see that $w_t = 1 - e^{-\epsilon t}$. In the same way, for the trials 26 to 50, we can get $w_t = Ae^{-\epsilon(t-25)}$ where A is a constant that can be decided given a particular value of w_t.

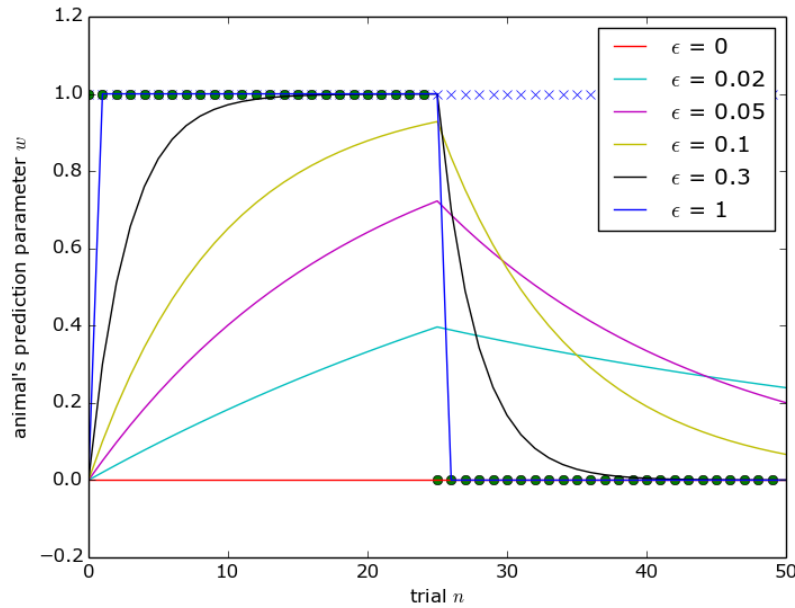The next thing to do is surely to study the impact of the learning rate, so we vary the value of ϵ.

FIGURE 3: Evolution of w in the same experiment but with different ϵ

As what our formula implies, greater the learning rate, faster the animal learns and unlearns the association. It's also interesting to notice that in most of the cases, it's more difficult to unlearn than to learn because the initial δ is smaller (in terms of the absolute value). However, even though one learns faster with a greater learning rate, it doesn't necessarily means that it's better. In fact, there are always noises in what is observed and a smaller learning rate indicates that one is somehow doing an average with past experiences.

Two extreme cases are shown in the figure, when $\epsilon = 0$, the animal can never learn, and when $\epsilon = 1$, the animal exploit only the information from the current trial to compute w and is hence very sensitive to noises. In cosequence, the best value of ϵ depends on the environment where one is.

1.3   Partial conditioning

In this paragraph we are interested in the partial conditioning experiment. We modify slightly the experiment of the last paragraph, there is always only one stimulus and it's all the time present, but the presence of the reward is now a random event with a fixed probability p. We simulate this experiment with $p = 0.4$ and the learning rate $\epsilon = 0.1$ over 160 trials.
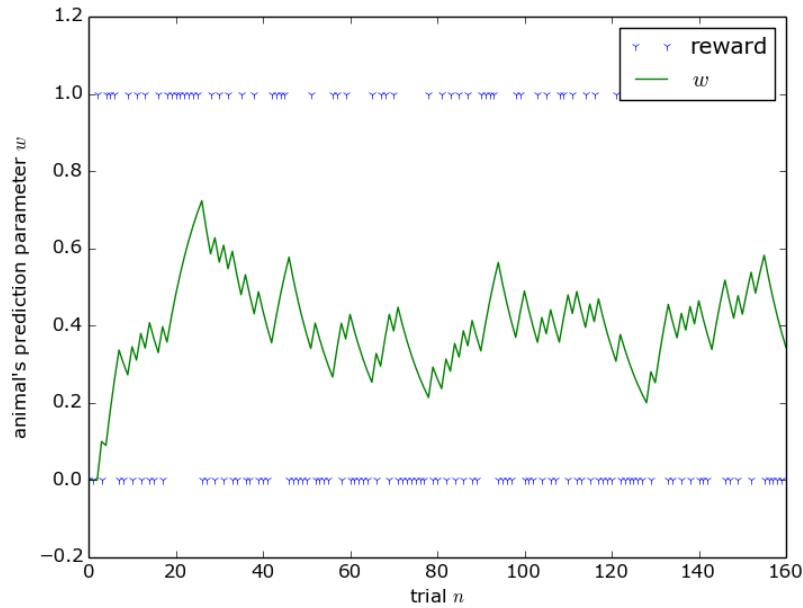
FIGURE 4: Evolution of w in the partial conditioning experiment ($p = 0.4$, $\epsilon = 0.1$)

We can see that the curve becomes quite noisy because the experiment is not deterministic anymore, but roughly speaking, w tends to increase at the beginning and then oscillates around 0.4. However, its value will never converge. As a matter of fact, the learning rate $\epsilon = 0.1$ is too high and a small number of trials can affect the animal's prediction parameter. We can redo the same simulation but with other ϵ values and over more trials to be able to see the evolution of w for smaller learning rate.
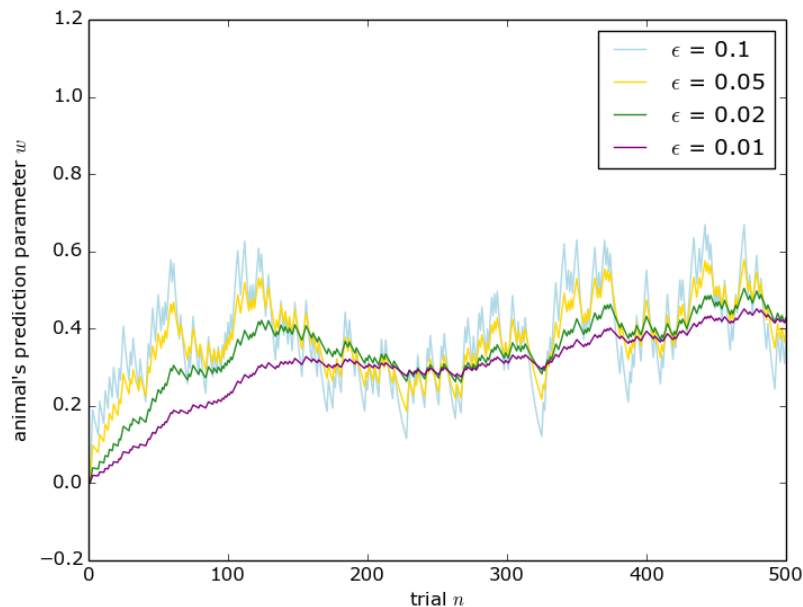
FIGURE 5: Evolution of w in the partial conditioning experiment with different ϵ ($p = 0.4$)

As predicted, when the learning rate decreases, in particular when it equals 0.01, it takes more time for the animal to learn, but the final value is also more stable and doesn't deviate very much from 0.4. On the contrary, we can imagine that if ϵ is bigger than 0.1, the curve

becomes even noisier and one can never learn the probability value $p = 0.4$, which confirms what is said before.

1.4   Blocking effect

Another advantage of the Rescola-Wagner model is that it allows us to explain the blocking effect. It means that the conditioning of an association between a CS and an US can be impaired if during the conditioning process, the CS is presented together with another CS that is already associated with the US. Therefore, in the newer experiment, we need two stimuli CS1 and CS2. In the first 25 trials, only CS1 and the reward (US) are present, and during the next 25 trials, CS1, CS2 and the reward are all present. We choose as usual $\epsilon = 0.1$.
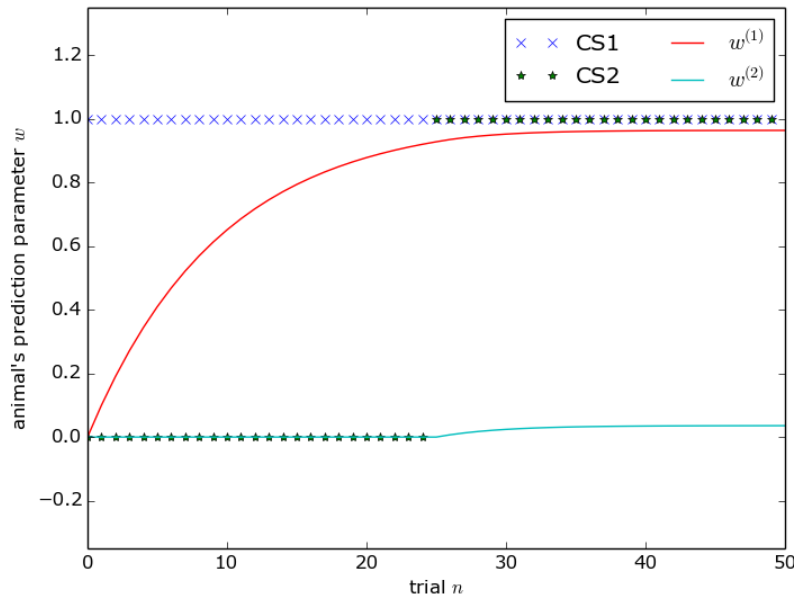

FIGURE 6: Simulation of the blocking effect ($\epsilon = 0.1$)

We get the expected result: the conditioning of CS2 is blocking by the presence of CS1. In the R-W model, the blocking effect can be explained by the fact that $\delta = r - v = r - w^{(1)} - w^{(2)}$ is small during the last 25 trials and using the update rule $w_{t+1}^{(2)} = w_t^{(2)} + \epsilon \delta_t u_t^{(2)}$ we can hardly change the value of $w^{(2)}$.

1.5   Overshadowing

In reality, there is no reason to assume that ϵ is the same for all the stimuli. We should replace the global ϵ by some individual $\epsilon^{(i)}$ for each i. Then, in order to compare different $\epsilon^{(i)}$, a simple experiment can be considered: all the stimuli and the reward are all the time present and we just need to see which particular stimulus is the most associated with the reward after a certain number of trials. For example, if we use two stimuli CS1 and CS2 with respective learning rate 0.2 and 0.1, we plot the result as below.
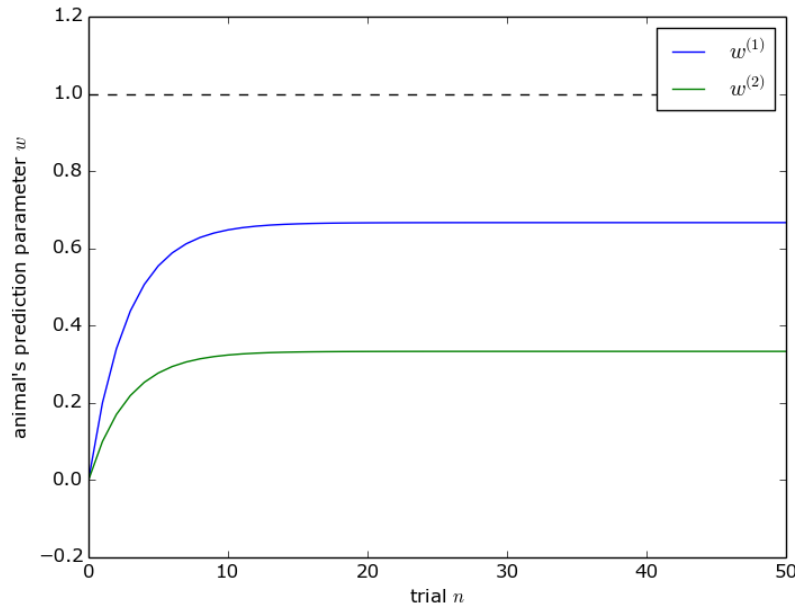
FIGURE 7: Simulation of the overshadowing effect ($\epsilon^{(1)} = 0.2$, $\epsilon^{(2)} = 0.1$)

The stimulus with a higher learning rate (in this case CS1) is more associated with the reward. And if the difference between two learning rates become even larger, it can result in a very weak conditioning of CS2. This is known as the overshadowing effect. We note that neither of the stimuli is fully associated with the reward, but if we add $w^{(1)}$ and $w^{(2)}$, we get something very near to 1.

1.6   Conclusion

In the Rescola-Wagner model conditioning means learning association between conditioned and unconditioned stimuli. At the beginning, we saw that this model could help us understand how one gets conditioned by a stimulus and how this conditioning can again disappear. The most appropriate value of the learning rate may vary from case to case. Next, it turned out that even in a non-deterministic experiment the model is able to find the key probability value. Blocking and overshadowing can also be explained.

Of course this very simple model has its limit. For instance, high-order conditioning requires us to take into consideration the time factor, which is not done yet for the time being. Nonetheless, this model is not totally absurd either. In fact, studies have suggested that the activity of some dopamine neurons in the brain encodes effectively the prediction error δ.