**A draft of simulation implementation notes**

To implement q-learning, we need to define state space and action space

1. Experiment 1: SR Mechanism

Let’s consider the noiseless case, for simplicity. We need to define the q-matrix for buyers and sellers separately.

For the buyer, we first label the action by combining the buyer’s choices in two stages:

a = 1 first report truthfully, if it goes to arbitration reports 0

a = 2 first report truthfully, if it goes to arbitration reports 20

a = 3 first report truthfully, if it goes to arbitration reports 70

a = 4 first misreport, if it goes to arbitration reports 0

a = 5 first misreport, if it goes to arbitration reports 20

a = 6 first misreport, if it goes to arbitration reports 70

For the seller, we label the action by

b=1 report truthfully

b=2 misreport

Now we define the state space. Let the memory length m, the state vector at time t is defined as

Here are the true value, the action of the buyer, and the action of the seller at the time j. The state vector is the same for both the buyer and seller. For the initial trials, we may consider m = 0 and consider just a pair of a seller and a buyer playing many rounds of the game. The payoff function follows what is defined in the paper. We can check how the strategies of the buyer and the seller evolve and if the strategies converge and how fast they converge. This evaluation will be in comparison of the SPI mechanism

1. Experiment 1: SPI Mechanism

For the SPI mechanism, we will need three Q-matrices: the first one is for the buyer’s strategy for the initial report, the second is for the seller’s strategy given the buyer’s initial report, and the third is for the buyer’s strategy of accepting or rejecting the seller’s counteroffer in the arbitration phase.

For the buyer’s initial reporting, we have the following actions

a = 1 report truthfully

a = 2 misreport

For the seller’s strategy, the actions are

b = 1 trade

b = 2 arbitration

For the buyer’s strategy in the arbitration phase, the actions are

c = 1 accept the counteroffer

c = 2 reject the counteroffer

The state vectors used are different for different q matrices. For the buyer’s initial reporting, the state vector is

For the seller’s decision, the state vector is

Note that the seller’s decision depends on the buyer’s initial reporting at time t.

For the buyer’s decision in the arbitration phase (=2)

Here are the true value, the initial reporting of the buyer, the action of the seller, and the buyer’s decision in the arbitration phase at time j.

If the seller does not go to the arbitration phase, the q-matrix for the buyer’s decision in the arbitration phase will not be updated (for , we can simply fix for easy bookkeeping). The buyer’s final payoff can be used to update the q-matrices of both buyer’s decision in the initial reporting and the arbitration phase.

1. Experiment 2: SR Mechanism

We also need to define the q-matrix for buyers and sellers separately.

For the buyer, we label the action by combining the buyer’s actions (reporting both the value and the cost in the initial phase and the value in the arbitration phase) in two phases:

a = 1 first report (200,130), if it goes to arbitration reports 200

a = 2 first report (200,130), if it goes to arbitration reports 250

a = 3 first report (200,130), if it goes to arbitration reports 320

a = 4 first report (200, 80), if it goes to arbitration reports 200

…

a = 26 first report (320,10), if it goes to arbitration reports 250

a = 27 first report (320,10), if it goes to arbitration reports 320

For the seller, we label the action (cost reporting) by

b = 1 first report (200,130), if it goes to arbitration reports 130

b = 2 first report (200,130), if it goes to arbitration reports 80

b = 3 first report (200,130), if it goes to arbitration reports 10

b = 4 first report (200, 80), if it goes to arbitration reports 130

…

b = 26 first report (320,10), if it goes to arbitration reports 80

b = 27 first report (320,10), if it goes to arbitration reports 10

Now we define the state space. Let the memory length m, the state vector at time t is defined as

Here are the true value, true cost, the action of the buyer, and the action of the seller at the time j. The state vector is the same for both the buyer and seller. For the initial trials, we may consider m = 0 and consider just a pair of a seller and a buyer playing many rounds of the game. The payoff function follows what is defined in the paper. We can check how the strategies of the buyer and the seller evolve and if the strategies converge and how fast they converge.

Experiment 2: KTH treatment

The implementation is simpler.

For the buyer, we label the action (reporting both values and costs) as;

a = 1 report (200, 130)

a = 2 report (200, 80)

a = 3 report (200, 10)

a = 4 report (250, 130)

a = 5 report (250, 80)

a = 6 report (250, 10)

a = 7 report (320, 130)

a = 8 report (320, 80)

a = 9 report (320, 10)

For the seller, we label the action (reporting both values and costs) as

b = 1 report (200, 130)

b = 2 report (200, 80)

b = 3 report (200, 10)

b = 4 report (250, 130)

b = 5 report (250, 80)

b = 6 report (250, 10)

b = 7 report (320, 130)

b = 8 report (320, 80)

b = 9 report (320, 10)

The state vector at time t is defined as

Here are the true value, true cost, the action of the buyer, and the action of the seller at the time j.

1. Experiment 2: SPI Mechanism

For the SPI mechanism, we will need 6 Q-matrices: the first one is for the buyer’s the initial report, the second is for the seller’s initial report, the third is for the buyer’s decision on arbitration, the fourth is for the seller’s decision on arbitration, the fifth one is for the buyer’s decision on counteroffer, and the sixth one is for the seller’s decision on counteroffer.

For the buyer’s initial reporting, we have the following actions

a = 1 report 200

a = 2 first report 250

a = 3 first report 320

For the seller’s strategy, the actions are

b = 1 report 130

b = 2 report 80

b = 3 report 10

For the buyer’s decision on calling for arbitration

w = 1 call

w = 2 no call

For the seller’s decision on calling for arbitration

x = 1 call

x = 2 no call

For the buyer’s decision on counteroffer

y= 1 accept

y= 2 reject

For the seller’s decision on counteroffer

z = 1 accept

z = 2 reject

The state vectors used are also different for different q matrices. For the buyer’s and seller’s initial reportings, the state vector is

For the seller’s decision, the state vector is

For the buyer’s and seller’s decisions in calling arbitration

For the buyer’s and seller’s decisions in the arbitration phase