[Finished] Framework for Negotiation with DRL method under gym environment

`https://github.com/YueNing/tn_source_code.git`

[Finished] Learned acceptance network and offer network

1. [IMRPOVEMENT]: Acceptance Strategy [Finshed] [Single Issue, Multi issues]
    - observation_space = [opponent_offer, time]
    - action_space = Box.Discrete(3)
        - ResponseType.ACCEPT,
        - ResponseType.Wait
        - ResponseType.REJECT_OFFER
2. [IMPROEMENT]: Offer Strategy [Finshed] [Single Issue and Multi issues]
    - observation_space = [opponent_offer, time] normalization between [-1, 1]
    - action_space = all of outcomes, normalization [-1, 1]

Code example:

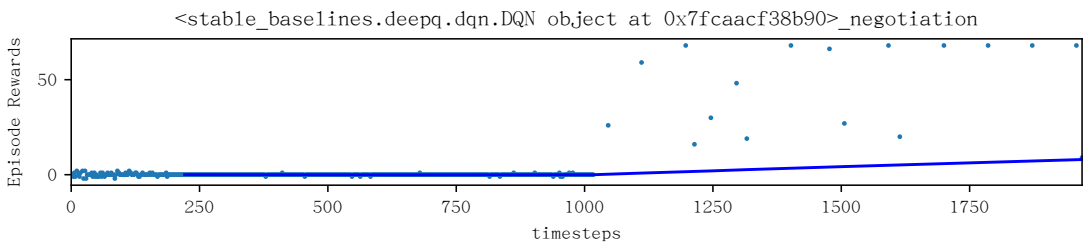
In [ ]:
```
!pip install negmas==0.7.0
!pip install gym==0.17.2
# Before install stable_baselines Need to install libopenmpi-dev:
command 'sudo apt install libopenmpi-dev'
!pip install mpi4py==3.0.3
!pip install stable_baselines==2.10.0
!pip install tensorflow==1.15.3
!pip install -i https://test.pypi.org/simple/ drl-negotiation
```

Using DQN to train the Acceptance Strategy of MyDRLNegotiator, competite with the MyOpponentNegotiator

In [4]:

```
from drl_negotiation.env import NegotiationEnv
from drl_negotiation.utils import generate_config, genearate_obse
rvation_space
from drl_negotiation.game import NegotiationGame
from drl_negotiation.negotiator import MyDRLNegotiator, MyOpponen
tNegotiator
from drl_negotiation.utility_functions import MyUtilityFunction
from drl_negotiation.train import train_negotiation

config = generate_config(n_issues=1)

game = NegotiationGame(
    name="negotiation_game",
    game_type="DRLNegotiation",
    issues=config.get("issues"),
    competitors=[
        MyDRLNegotiator(
            name="my_drl_negotiator",
            ufun=MyUtilityFunction(weights=config.get("weight
s")[0]),
            init_proposal=False,
        ),
        MyOpponentNegotiator(
            name="my_opponent_negotiator",
            ufun=MyUtilityFunction(weights=config.get("weight
s")[1])
        )
    ],
    n_steps=config.get("n_steps")
)
env = NegotiationEnv(
        name="negotiation_env_ac_s",
        strategy="ac_s",
        game=game,
        observation_space=genearate_observation_space(config),
        action_space=3
    )

plot = True

game.set_env(env=env)
model = "DQN"
done, _ = train_negotiation(plot=plot, model=model, num_timesteps
=2000, env=env, monitor=False)
assert done,  f'train false by the model {model}'

print('Finished!')
```

```
Logging to train_negotiation_DQN
--------------------------------------
| % time spent exploring  | 2        |
| episodes                | 100      |
| mean 100 episode reward | 0.1      |
| steps                   | 219      |
--------------------------------------
--------------------------------------
| % time spent exploring  | 2        |
| episodes                | 200      |
| mean 100 episode reward | 0        |
| steps                   | 419      |
--------------------------------------
--------------------------------------
| % time spent exploring  | 2        |
| episodes                | 300      |
| mean 100 episode reward | -0       |
| steps                   | 619      |
--------------------------------------
--------------------------------------
| % time spent exploring  | 2        |
| episodes                | 400      |
| mean 100 episode reward | 0        |
| steps                   | 819      |
--------------------------------------
--------------------------------------
| % time spent exploring  | 2        |
| episodes                | 500      |
| mean 100 episode reward | 1.6      |
| steps                   | 1197     |
--------------------------------------
Finished!
```



In [7]:

```
from drl_negotiation.utils import generate_action_space
env =  NegotiationEnv(
     name="negotiation_env_of_s",
     strategy="of_s",
     game=game,
     observation_space=genearate_observation_space(config),
     # action_space=[[config.get("issues")[0].values[0], ], [conf
ig.get("issues")[0].values[1], ]]
     # action_space=[[-1, ], [1, ]]
     action_space = generate_action_space(config)
)
game.set_env(env=env)
model = "PPO1"
done, _ = train_negotiation(plot=plot, model=model, env=env, moni
tor=False)
assert done,  f'train false by the model {model}'
```

```
Logging to train_negotiation_PPO1
WARNING:tensorflow:From /home/nauen/anaconda3/envs/tn/lib/python
3.7/site-packages/stable_baselines/common/distributions.py:418: T
he name tf.random_normal is deprecated. Please use tf.random.norm
al instead.

WARNING:tensorflow:From /home/nauen/anaconda3/envs/tn/lib/python
3.7/site-packages/stable_baselines/ppo1/pposgd_simple.py:162: The
name tf.assign is deprecated. Please use tf.compat.v1.assign inst
ead.

********** Iteration 0 ************
Optimizing...
     pol_surr |    pol_entpen |       vf_loss |            kl |
ent
     -0.00450 |      -0.01419 |       3.49917 |       0.00029 |
1.41919
     -0.01926 |      -0.01417 |       2.81515 |       0.00292 |
1.41699
     -0.03233 |      -0.01415 |       2.36316 |       0.01067 |
1.41506
     -0.04270 |      -0.01413 |       2.17231 |       0.02609 |
1.41327
Evaluating losses...
     -0.04192 |      -0.01412 |       2.09893 |       0.03788 |
1.41228
----------------------------------
| EpLenMean       | 6.68         |
| EpRewMean       | 1.4          |
| EpThisIter      | 38           |
| EpisodesSoFar   | 38           |
| TimeElapsed     | 0.884        |
| TimestepsSoFar  | 256          |
| ev_tdlam_before | 0.00494      |
| loss_ent        | 1.412278     |
| loss_kl         | 0.037876707  |
| loss_pol_entpen | -0.01412278  |
| loss_pol_surr   | -0.041923054 |
| loss_vf_loss    | 2.0989327    |
----------------------------------
********** Iteration 1 ************
Optimizing...
     pol_surr |    pol_entpen |       vf_loss |            kl |
ent
     -0.00743 |      -0.01412 |       2.33849 |       0.00058 |
1.41209
     -0.02620 |      -0.01411 |       2.33709 |       0.00715 |
1.41084
     -0.03752 |      -0.01409 |       2.33942 |       0.02285 |
1.40932
     -0.03954 |      -0.01408 |       2.34301 |       0.04478 |
1.40804
Evaluating losses...
     -0.03912 |      -0.01408 |       2.33689 |       0.05831 |
1.40756
----------------------------------
| EpLenMean       | 7.24         |
| EpRewMean       | 1.63         |
| EpThisIter      | 32           |
```

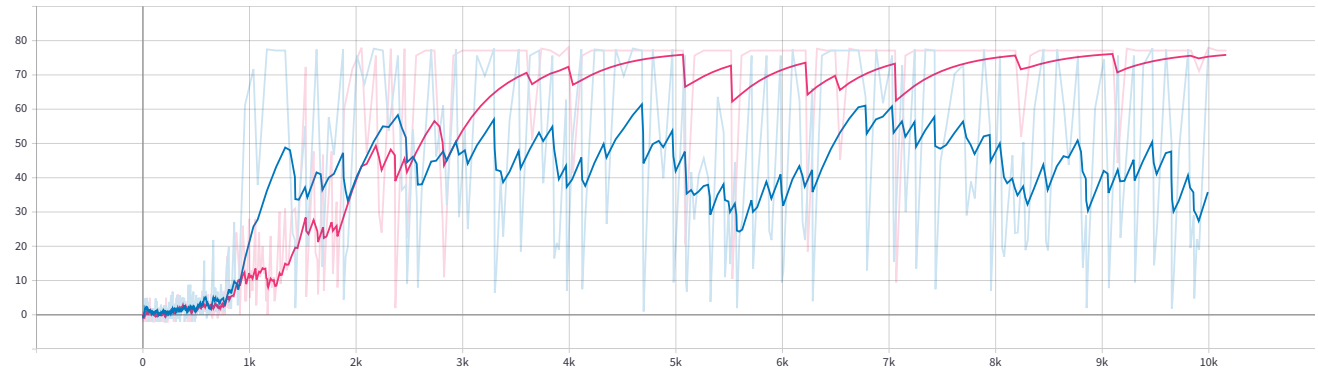
```
| EpisodesSoFar   | 70           |
| TimeElapsed     | 1.52         |
| TimestepsSoFar  | 512          |
| ev_tdlam_before | -0.00713     |
| loss_ent        | 1.4075621    |
| loss_kl         | 0.058307104  |
| loss_pol_entpen | -0.014075621 |
| loss_pol_surr   | -0.039115265 |
| loss_vf_loss    | 2.336886     |
----------------------------------
********** Iteration 2 ************
Optimizing...
     pol_surr |    pol_entpen |       vf_loss |            kl |
ent
     -0.00426 |      -0.01407 |       8.54860 |       0.00030 |
1.40718
     -0.00902 |      -0.01406 |       8.02638 |       0.00371 |
1.40591
     -0.00869 |      -0.01404 |       7.33204 |       0.00939 |
1.40448
     -0.00952 |      -0.01403 |       6.66217 |       0.01020 |
1.40285
Evaluating losses...
     -0.01027 |      -0.01402 |       6.23898 |       0.00843 |
1.40179
----------------------------------
| EpLenMean       | 8.78         |
| EpRewMean       | 2.32         |
| EpThisIter      | 16           |
| EpisodesSoFar   | 86           |
| TimeElapsed     | 2.14         |
| TimestepsSoFar  | 768          |
| ev_tdlam_before | 0.0602       |
| loss_ent        | 1.401785     |
| loss_kl         | 0.008433653  |
| loss_pol_entpen | -0.01401785  |
| loss_pol_surr   | -0.010266896 |
| loss_vf_loss    | 6.238983     |
----------------------------------
********** Iteration 3 ************
Optimizing...
     pol_surr |    pol_entpen |       vf_loss |            kl |
ent
     -0.00014 |      -0.01401 |      13.70640 |      1.81e-06 |
1.40150
     -0.00221 |      -0.01401 |      13.19557 |      9.35e-05 |
1.40065
     -0.00443 |      -0.01400 |      12.63545 |       0.00057 |
1.39981
     -0.00439 |      -0.01399 |      12.10211 |       0.00128 |
1.39912
Evaluating losses...
     -0.00427 |      -0.01399 |      11.79006 |       0.00174 |
1.39874
----------------------------------
| EpLenMean       | 10.3         |
| EpRewMean       | 3.33         |
| EpThisIter      | 13           |
| EpisodesSoFar   | 99           |
| TimeElapsed     | 2.78         |
```

```
| TimestepsSoFar   | 1024         |
| ev_tdlam_before | 0.133        |
| loss_ent         | 1.3987445    |
| loss_kl          | 0.0017395527 |
| loss_pol_entpen | -0.013987444 |
| loss_pol_surr    | -0.004270671 |
| loss_vf_loss     | 11.790057    |
```

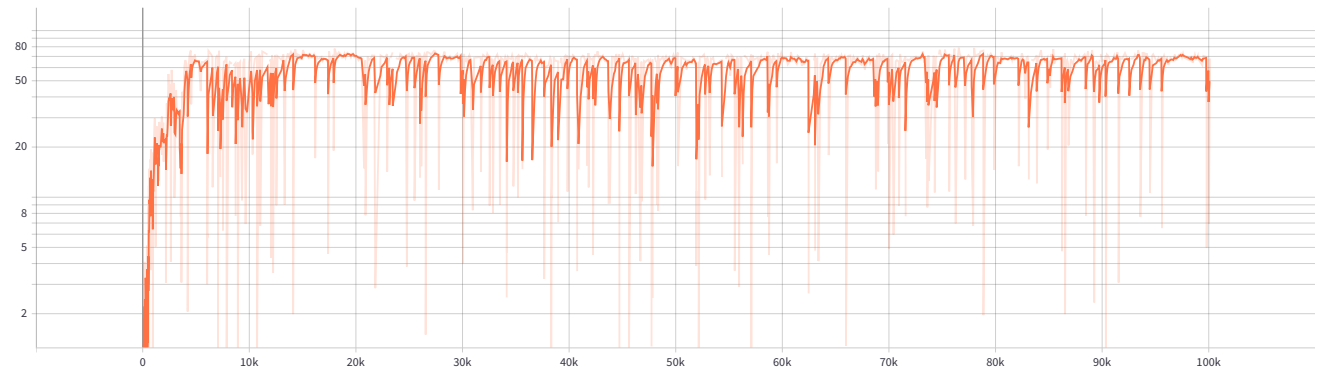
Example result shown in tensorboard

Episode reward: The reward of Acceptance strategy and offer/bidding strategy is increasing.

SIgnle issue, acceptance strategy, episode reward, dqn(blue line) and ppo1(pink line)



SIgnle issue, offer/bidding strategy, episode reward, ppo1



The basic environment of negotiation with method deep reinforcement learning has been implemented!

The future work is about implementing a scml environment as similar as negotiation environment.

Due to the many ideas of improvement of agents in scml. can not try all parts. The future work (myagent in scml) is mainly in **negotiation manager** and **negotiation algorithm**.