#### 2024-10-24训练日志：

1. DCDR\_PPO simple

DCDR的训练参数

parser.add\_argument("--dr\_min\_ratio", type=float, default=0.8,

help="Dynamic rl min reward ratio in rl and gail reward")

parser.add\_argument("--dr\_max\_ratio", type=float, default=0.99,

help="Dynamic rl max reward ratio in rl and gail reward")

parser.add\_argument("--start\_episode", type=float, default=1000,

help="Start episode for dynamic rl reward ratio increasing")

parser.add\_argument("--end\_episode", type=float, default=2000,

help="End episode for dynamic rl reward ratio increasing")

parser.add\_argument("--train-episodes", type=int, default=3000,

help="number of time steps")（稀疏奖励）

average episode steps 25.0

agent average returns 9.9

... saving agent checkpoint ...

The best reward of agent is 9.95 when episode is 3000

2. DCDR\_PPO simple

DCDR的训练参数

parser.add\_argument("--dr\_min\_ratio", type=float, default=0.8,

help="Dynamic rl min reward ratio in rl and gail reward")

parser.add\_argument("--dr\_max\_ratio", type=float, default=1,

help="Dynamic rl max reward ratio in rl and gail reward")

parser.add\_argument("--start\_episode", type=float, default=1000,

help="Start episode for dynamic rl reward ratio increasing")

parser.add\_argument("--end\_episode", type=float, default=1500,

help="End episode for dynamic rl reward ratio increasing")

parser.add\_argument("--train-episodes", type=int, default=3000,

help="number of time steps")（稀疏奖励）

average episode steps 25.0

agent average returns 15.8

... saving agent checkpoint ...

The best reward of agent is 15.8 when episode is 3000

3. DCDR\_PPO simple

DCDR的训练参数

# DCDR的训练参数

parser.add\_argument("--dr\_min\_ratio", type=float, default=0.5,

help="Dynamic rl min reward ratio in rl and gail reward")

parser.add\_argument("--dr\_max\_ratio", type=float, default=0.5,

help="Dynamic rl max reward ratio in rl and gail reward")

parser.add\_argument("--start\_episode", type=float, default=1000,

help="Start episode for dynamic rl reward ratio increasing")

parser.add\_argument("--end\_episode", type=float, default=1500,

help="End episode for dynamic rl reward ratio increasing")

parser.add\_argument("--train-episodes", type=int, default=3000,

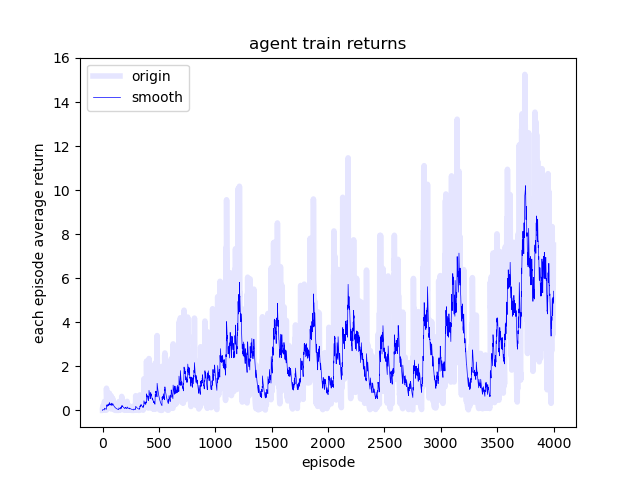
help="number of time steps")（稀疏奖励）

average episode steps 25.0

agent average returns 5.3

... saving agent checkpoint ...

The best reward of agent is 7.32 when episode is 3800



4. DCDR\_PPO simple

DCDR的训练参数

parser.add\_argument("--dr\_min\_ratio", type=float, default=0.8,

help="Dynamic rl min reward ratio in rl and gail reward")

parser.add\_argument("--dr\_max\_ratio", type=float, default=1,

help="Dynamic rl max reward ratio in rl and gail reward")

parser.add\_argument("--start\_episode", type=float, default=1000,

help="Start episode for dynamic rl reward ratio increasing")

parser.add\_argument("--end\_episode", type=float, default=1001,

help="End episode for dynamic rl reward ratio increasing")

parser.add\_argument("--train-episodes", type=int, default=3000,

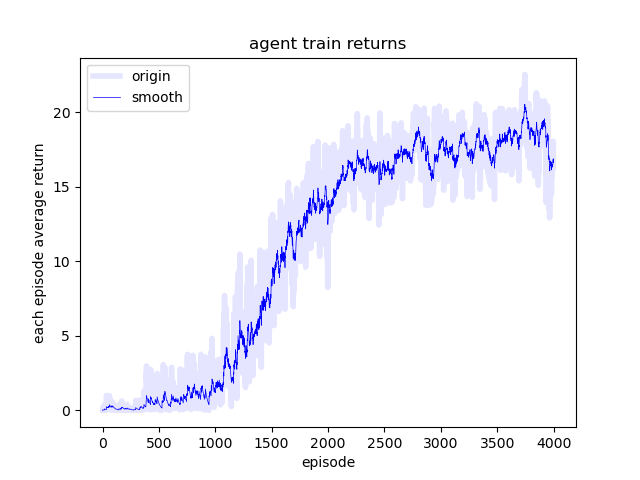
help="number of time steps")（稀疏奖励）

average episode steps 25.0

agent average returns 17.2

... saving agent checkpoint ...

The best reward of agent is 19.37 when episode is 3800



5. DCDR\_PPO simple

DCDR的训练参数

parser.add\_argument("--dr\_min\_ratio", type=float, default=0,

help="Dynamic rl min reward ratio in rl and gail reward")

parser.add\_argument("--dr\_max\_ratio", type=float, default=1,

help="Dynamic rl max reward ratio in rl and gail reward")

parser.add\_argument("--start\_episode", type=float, default=1000,

help="Start episode for dynamic rl reward ratio increasing")

parser.add\_argument("--end\_episode", type=float, default=1001,

help="End episode for dynamic rl reward ratio increasing")

parser.add\_argument("--train-episodes", type=int, default=3000,

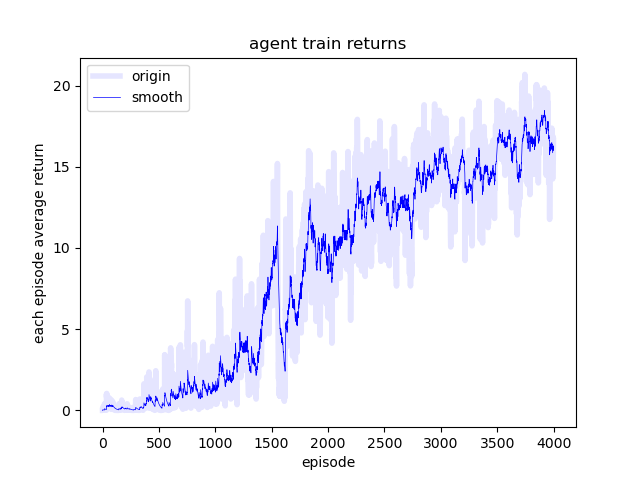
help="number of time steps")（稀疏奖励）

average episode steps 25.0

agent average returns 16.7

... saving agent checkpoint ...

The best reward of agent is 17.29 when episode is 3900



5. DCDR\_PPO simple

# DCDR的训练参数

parser.add\_argument("--dr\_min\_ratio", type=float, default=1,

help="Dynamic rl min reward ratio in rl and gail reward")

parser.add\_argument("--dr\_max\_ratio", type=float, default=1,

help="Dynamic rl max reward ratio in rl and gail reward")

parser.add\_argument("--start\_episode", type=float, default=1000,

help="Start episode for dynamic rl reward ratio increasing")

parser.add\_argument("--end\_episode", type=float, default=1001,

help="End episode for dynamic rl reward ratio increasing")

parser.add\_argument("--train-episodes", type=int, default=3000,

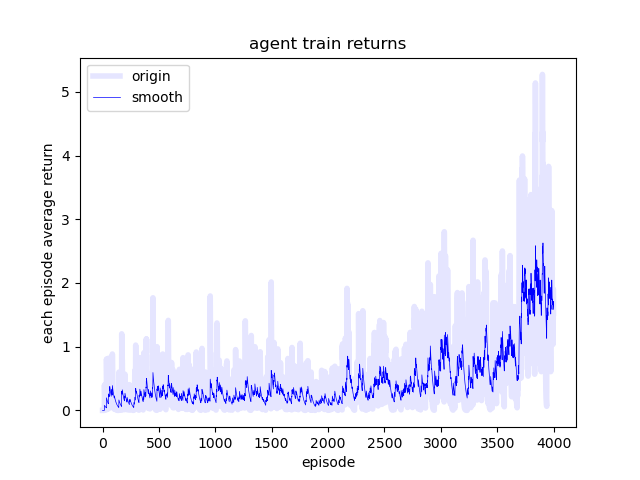
help="number of time steps")（稀疏奖励）

average episode steps 25.0

agent average returns 1.7

... saving agent checkpoint ...

The best reward of agent is 1.97 when episode is 3900



6. DCDR\_PPO simple

# DCDR的训练参数

combined\_ratio = 0.6

if self.episode\_num % 100 == 0 and self.episode\_num > 100:

agent\_returns = np.array(pd.read\_csv(self.data\_save\_path + "/agent\_returns.csv"))

if np.mean(agent\_returns[-self.display\_episodes:]) > np.mean(agent\_returns[-1000:]) \* 0.99:

self.factor += 1

else:

self.factor -= 1

if np.abs(self.factor) >= 3:

combined\_ratio = 1 / (1 + np.exp(-self.factor))

else:

combined\_ratio = 0.8

print(np.mean(agent\_returns[-self.display\_episodes:]), np.mean(agent\_returns[-1000:]))

print("combined\_ratio", combined\_ratio)

parser.add\_argument("--train-episodes", type=int, default=3000,

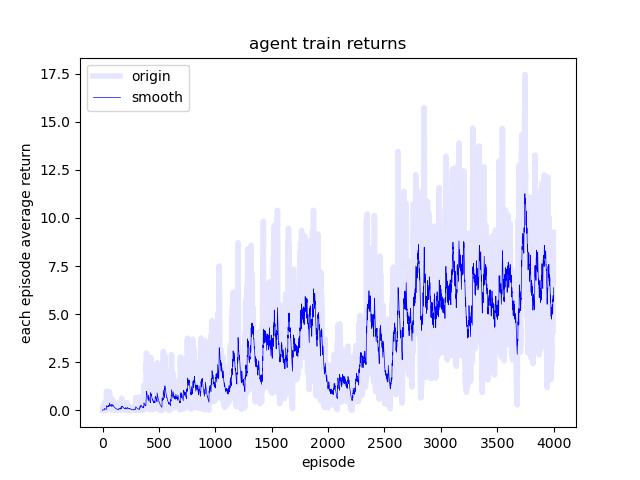
help="number of time steps")（稀疏奖励）

average episode steps 25.0

agent average returns 6.5

... saving agent checkpoint ...

The best reward of agent is 8.3 when episode is 3800



7. DCDR\_PPO simple

if self.episode\_num % 100 == 0 and self.episode\_num > 100:

agent\_returns = np.array(pd.read\_csv(self.data\_save\_path + "/agent\_returns.csv"))

if np.mean(agent\_returns[-self.display\_episodes:]) > np.mean(agent\_returns[-1000:]) \* 0.99:

self.factor += 1 / 2

else:

self.factor -= 1 / 2

self.factor = np.clip(self.factor, 0, 6)

if self.factor > 3:

combined\_ratio = 1

else:

combined\_ratio = 0.8

print(np.mean(agent\_returns[-1000:]))

print("self.factor", self.factor)

parser.add\_argument("--train-episodes", type=int, default=3000,

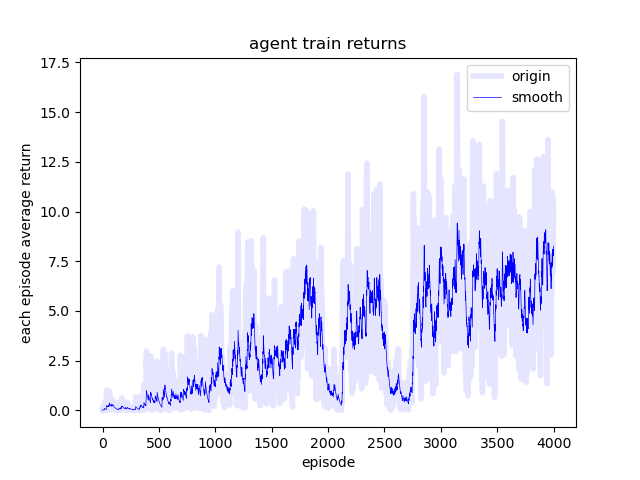
help="number of time steps")（稀疏奖励）

average episode steps 25.0

agent average returns 7.9

... saving agent checkpoint ...

The best reward of agent is 7.94 when episode is 4000



8. DCDR\_PPO simple

parser.add\_argument("--dr\_min\_ratio", type=float, default=0,

help="Dynamic rl min reward ratio in rl and gail reward")

parser.add\_argument("--dr\_max\_ratio", type=float, default=1,

help="Dynamic rl max reward ratio in rl and gail reward")

parser.add\_argument("--start\_episode", type=float, default=1000,

help="Start episode for dynamic rl reward ratio increasing")

parser.add\_argument("--end\_episode", type=float, default=1200,

help="End episode for dynamic rl reward ratio increasing")parser.add\_argument("--train-episodes", type=int, default=3000,

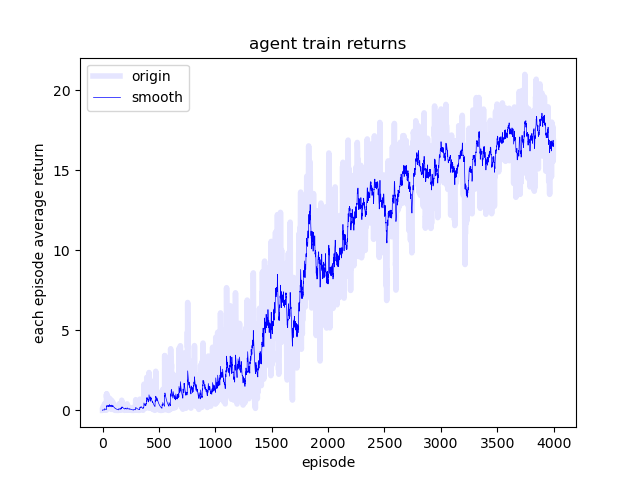
help="number of time steps")（稀疏奖励）

average episode steps 25.0

agent average returns 16.8

... saving agent checkpoint ...

The best reward of agent is 17.87 when episode is 3900



8. DCDR\_PPO simple

# DCDR的训练参数

parser.add\_argument("--dr\_min\_ratio", type=float, default=0.5,

help="Dynamic rl min reward ratio in rl and gail reward")

parser.add\_argument("--dr\_max\_ratio", type=float, default=1,

help="Dynamic rl max reward ratio in rl and gail reward")

parser.add\_argument("--start\_episode", type=float, default=400,

help="Start episode for dynamic rl reward ratio increasing")

parser.add\_argument("--end\_episode", type=float, default=1000,

help="End episode for dynamic rl reward ratio increasing")

parser.add\_argument("--train-episodes", type=int, default=3000,

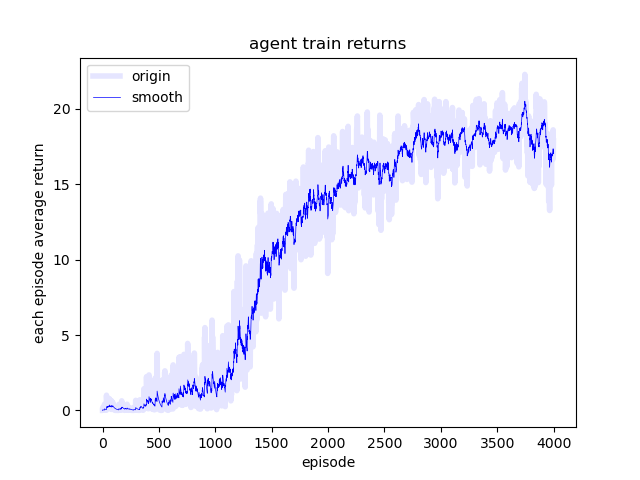
help="number of time steps")（稀疏奖励）

average episode steps 25.0

agent average returns 17.3

... saving agent checkpoint ...

The best reward of agent is 19.05 when episode is 3800



8. DCDR\_PPO simple

# DCDR的训练参数

parser.add\_argument("--dr\_min\_ratio", type=float, default=1-1e-3,

help="Dynamic rl min reward ratio in rl and gail reward")

parser.add\_argument("--dr\_max\_ratio", type=float, default=1-1e-3,

help="Dynamic rl max reward ratio in rl and gail reward")

parser.add\_argument("--start\_episode", type=float, default=1000,

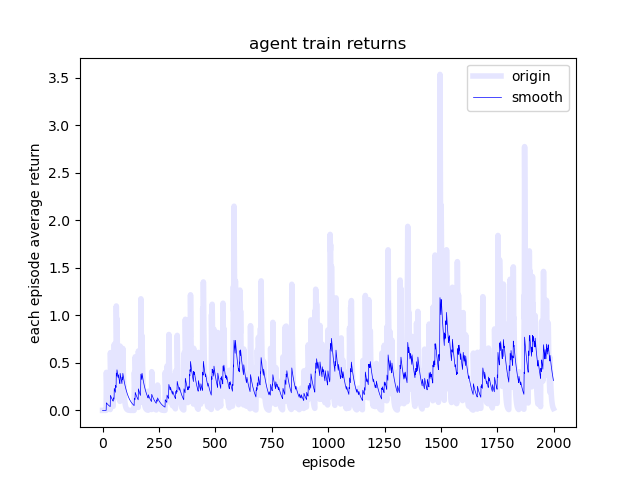
help="Start episode for dynamic rl reward ratio increasing")

parser.add\_argument("--end\_episode", type=float, default=1001,

help="End episode for dynamic rl reward ratio increasing")

parser.add\_argument("--train-episodes", type=int, default=3000,

help="number of time steps")（稀疏奖励）



8. DCDR\_PPO simple

# DCDR的训练参数

parser.add\_argument("--dr\_min\_ratio", type=float, default=1-1e-1,

help="Dynamic rl min reward ratio in rl and gail reward")

parser.add\_argument("--dr\_max\_ratio", type=float, default=1-1e-3,

help="Dynamic rl max reward ratio in rl and gail reward")

parser.add\_argument("--start\_episode", type=float, default=1000,

help="Start episode for dynamic rl reward ratio increasing")

parser.add\_argument("--end\_episode", type=float, default=1001,

help="End episode for dynamic rl reward ratio increasing")

parser.add\_argument("--train-episodes", type=int, default=3000,

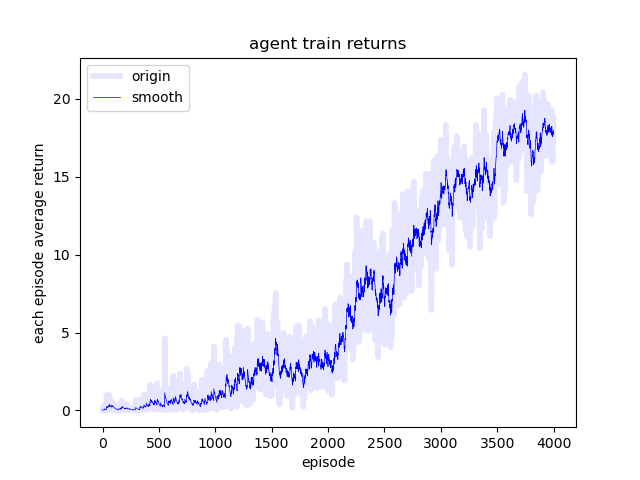
help="number of time steps")（稀疏奖励）

average episode steps 25.0

agent average returns 18.1

... saving agent checkpoint ...

The best reward of agent is 18.1 when episode is 4000



8. DCDR\_PPO simple

# DCDR的训练参数

parser.add\_argument("--dr\_min\_ratio", type=float, default=0.8,

help="Dynamic rl min reward ratio in rl and gail reward")

parser.add\_argument("--dr\_max\_ratio", type=float, default=1,

help="Dynamic rl max reward ratio in rl and gail reward")

parser.add\_argument("--start\_episode", type=float, default=800,

help="Start episode for dynamic rl reward ratio increasing")

parser.add\_argument("--end\_episode", type=float, default=1200,

help="End episode for dynamic rl reward ratio increasing")

parser.add\_argument("--train-episodes", type=int, default=3000,

help="number of time steps")（稀疏奖励）

combined\_ratio = self.dr\_min\_ratio + (self.dr\_max\_ratio - self.dr\_min\_ratio) \* (

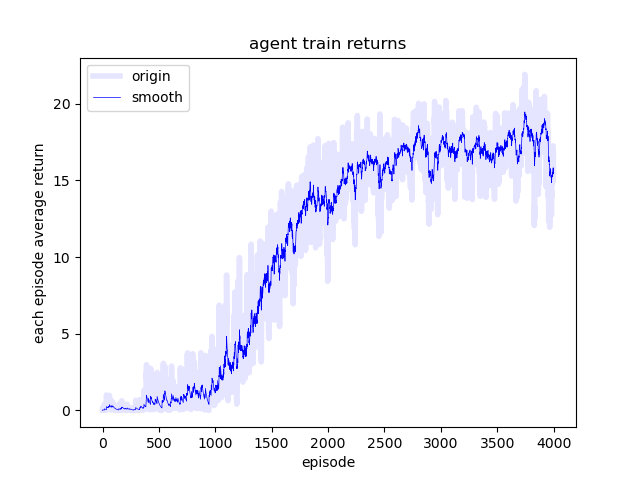
1 / (1 + np.exp(-0.04 \* (self.episode\_num - (self.start\_episode + self.end\_episode) / 2))))

average episode steps 25.0

agent average returns 16.4

... saving agent checkpoint ...

The best reward of agent is 18.44 when episode is 3800



8. DCDR\_PPO simple

# DCDR的训练参数

parser.add\_argument("--dr\_min\_ratio", type=float, default=0.8,

help="Dynamic rl min reward ratio in rl and gail reward")

parser.add\_argument("--dr\_max\_ratio", type=float, default=1,

help="Dynamic rl max reward ratio in rl and gail reward")

parser.add\_argument("--start\_episode", type=float, default=800,

help="Start episode for dynamic rl reward ratio increasing")

parser.add\_argument("--end\_episode", type=float, default=1000,

help="End episode for dynamic rl reward ratio increasing")

parser.add\_argument("--train-episodes", type=int, default=3000,

help="number of time steps")（稀疏奖励）

combined\_ratio = self.dr\_min\_ratio + (self.dr\_max\_ratio - self.dr\_min\_ratio) \* (

1 / (1 + np.exp(-0.05 \* (self.episode\_num - (self.start\_episode + self.end\_episode) / 2))))

average episode steps 25.0

agent average returns 18.0

... saving agent checkpoint ...

The best reward of agent is 19.91 when episode is 3800

