INTISARI

Multi Agent Deep Deterministic Policy Gradient
terhadap Robot Lengan Enam Sumbu
Oleh
Fahtul Ikhsan
19/445637/PA/19461

Integrasi antara robot dan manusia dalam suatu proses manufaktur telah berjalan sejak lama. Pengintegrasian robot mampu membantu meningkatkan produktivitas dalam bidang manufaktur. Produktivitas dapat ditingkatkan dengan mengoptimalkan aksi yang akan dilakukan robot dalam proses manufaktur. Algoritma MADDPG merupakan salah satu algoritma yang berpotensi untuk dapat digunakan dalam mengoptimalkan suatu robot dalam menjalankan aksi-aksi demi mencapai suatu tujuan tertentu. Algoritma MADDPG adalah pengaplikasian *multi agent* pada algoritma DDPG yang akan mempelajari *policy* dengan memaksimalkan nilai Q.

Pada penelitian ini, pengujian terhadap kemampuan MADDPG untuk melatih robot dalam menjalankan aksi-aksi demi mencapai suatu tujuan tertentu dilakukan dengan mengaplikasikan algoritma MADDPG pada suatu robot enam sumbu. Setiap sumbu pada robot lengan akan menjadi suatu agen tersendiri dan harus berkooperasi dengan kelima sumbu lainnya. Robot ini kemudian dilatih untuk meraih satu titik menggunakan *end effector* dari suatu posisi awal. *Prioritized Experience Replay* diaplikasikan pada *replay buffer* sebagai metode untuk membantu mempercepat pelatihan. Algoritma DDPG juga diuji dengan kondisi yang sama sebagai perbandingan.

Hasil dari penelitian yang didapatkan dari 10 sampel masing-masing pada algoritma MADDPG dan DDPG menunjukkan bahwa robot mampu meraih titik objektif dengan tingkat keberhasilan 100% pada kedua algoritma tersebut. Pada kasus ini, performa *policy* yang dihasilkan oleh algoritma MADDPG tidak memiliki perbedaan signifikan dibandingkan dengan algoritma DDPG, dimana pada 100 episode terakhir pelatihan nilai minimal dari rata-rata jumlah *frame* per episode pada algoritma MADDPG ialah 2,1 *frame* dan algoritma DDPG ialah 2,3 *frame.* Namun pelatihan menggunakan algoritma MADDPG memerlukan lebih banyak langkah atau *step* dibandingkan algoritma DDPG, dimana algoritma DDPG mampu mencapai rata-rata *reward* total dengan nilai >30 pada episode 253 sedangkan algoritma MADDPG baru mencapainya pada episode 292. Penelitian lebih lanjut dengan lingkungan yang lebih dinamis perlu dilakukan untuk melihat kelayakan dan performa algoritma MADDPG untuk melatih robot lengan enam sumbu, dengan suatu sumbu menjadi agen tersendiri pada lingkungan dinamis.

Kata kunci : Robot lengan enam sumbu, *reinforcement learning*, MADDPG.

ABSTRACT

Multi Agent Deep Deterministic Policy Gradient
on Six Axis Arm Robot
By
Fahtul Ikhsan
19/445637/PA/19461

The integration of robots and humans in manufacturing processes has been ongoing for a long time. Robot integration can help increase productivity in manufacturing. Productivity can be increased by optimizing the actions taken by robots in the manufacturing process. The MADDPG algorithm is one algorithm that has the potential to be used to optimize a robot's actions to achieve a specific goal. The MADDPG algorithm is a multi-agent application of the DDPG algorithm that learns policies by maximizing the Q value.

In this study, the ability of MADDPG to train a robot to perform actions to achieve a specific goal was tested by applying the MADDPG algorithm to a six-axis robot. Each axis of the robot arm becomes a separate agent and must cooperate with the other five axes. The robot is then trained to reach a point using an end effector from a starting position. Prioritized Experience Replay is applied to the replay buffer as a method to help accelerate training. The DDPG algorithm was also tested under the same conditions for comparison.

The results of the study, obtained from 10 samples each for the MADDPG and DDPG algorithms, show that the robot is capable of reaching the objective point with a 100% success rate in both algorithms. In this case, the policy performance generated by the MADDPG algorithm does not have a significant difference compared to the DDPG algorithm, where on the last 100 episodes of training the minimum value of the average number of frames per episode in the MADDPG algorithm is 2,1 frames and the DDPG algorithm is 2,3 frames. However, training using the MADDPG algorithm requires more steps than the DDPG algorithm, where the DDPG algorithm is able to achieve an average total reward with a value of >30 in episode 253 while the MADDPG algorithm only reaches it in episode 292. Further research with a more dynamic environment is needed to see the feasibility and performance of the MADDPG algorithm for training a six-axis arm robot, with an axis being a separate agent in a dynamic environment.

Keywords : Six axis robot, reinforcement learning, MADDPG.