جامعة دمشق
كلية الهندسة الميكانيكية والكهربائية
قسم هندسة الحواسيب والأتمتة

أمثلة مسار روبوت استكشاف باستخدام التعلم المعزز العميق

Optimizing Path of Exploration Robot via Deep Reinforcement Learning

م. أسامة ابراهيم

المشرف العلمي :أ.د.م. سمير كرمان المشرف المشارك : د.م. رؤوف حمدان

الملخص

قدمت خوارزميات التعلم المعزّز العميق حلّاً لروبوتات الذكاء الاصطناعي؛ وذلك لاكتشاف المسار الأفضل والأسرع لها في الوصول إلى هدفها من خلال: القدرة على التعلم من تجاربهم السابقة، واكتساب المعرفة، والتمثيل الصحيح لبيئاتهم للوصول إلى مستوى قريب من الإنسان في بيئات العالم الحقيقي المعقدة. جرى في هذه البحث تحليل أداء خوارزميات التعلم المعزّز العميق، مثل خوارزمية Q learning وخوارزمية Policy gradient،وذلك لمساعده الروبوت للوصول إلى هدفه بأفضل وأسرع مسار وذلك باستخدام أحدث التقنيات في ضبط المعاملات العليا للشبكات العصبونية العميقة المستخدمة

القسم النظري

تعد خوارزميات التعلم المعزّز من أهم الخوارزميات التي تساعد الروبوتات على الملاحة في البيئات غير المعروفة سابقا، ومع التحسن الكبير الذي طرأ على هذه الخوارزميات ماتزال تعاني من مشاكلات في الدقة والسرعة، ولاسيما عند إضافة الشبكات العصبونية العميقة إليها، يدرس هذا البحث إمكانية تحسين أداء هذه الخوارزميات عبر ضبط معاملات الشبكات العصبونية العميقة المستخدمة بأفضل طريقة، وتقليل زمن التدريب عبر إضافة أحدث التقنيات المستخدمة في هذا المجال.

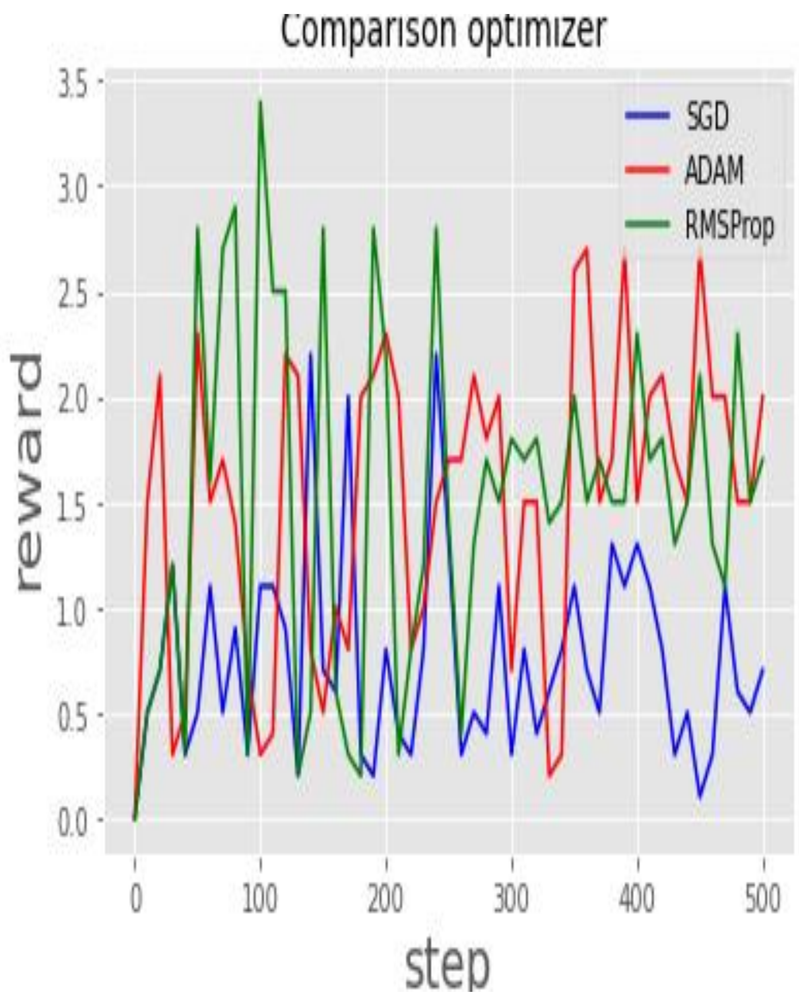
القسم العملي

في المرحلة الأولى تُحدد مداخل الشبكة وهي الحالات الحالية الممكنة للعميل حسب البيئة (CartPole-v1) وهي أربع حالات، ومخارج الشبكة تمثل التفاعلين الممكنين في البيئة، أي تحرك العربة يمينا أو يسارا .في المرحلة الثانية من خوارزمية Q learning Deep يتم اختيار الإجراء المناسب في كل مرحلة حسب خوارزمية epsilon greedy، وفي المرحلة الأخيرة يتم تحديث أوزان الشبكة العصبونية العميقة حسب معادلة بيلمان.

يستعرض السيناريو الثاني من البحث استخدام الشبكات العصبونية التلافيفية العميقة(DCNN) لتحسين أداء خوارزميات التعلم المعزز العميق، واختبرت النتائج على بيئة MAZ بأبعاد (6*6) تأتي أهمية هذا السيناريو من أن مدخلات الشبكة العصبونية عبارة عن صور لبيئة المحاكاة لرصد مكان الروبوت وهذا السيناريو أقرب للواقع من حيث إن أغلب الروبوتات تعمل بالواقع بواسطة كاميرات مثبته بها لإدخال البيانات

النتائج والمناقشة

في هذا البحث جرى اختبار أداء خوارزمية policy gradient مع شبكة عصبونية ملتفة عميقة من أجل أنواع عدة من المحسنات، وحصلنا على تغير مكافاة العميل خلال 500 دورة تدريب كما في الشكل التالي



المراجع

[1] Tan ,R. , Zhou,J. , Du, H. , Shang ,S., Dai ,L.(2019) An modeling processing method for video games based on deep reinforcement learning: University of TechnologyHefei, Anhui, China,IEEE.

[2] Mnih ,V., Kavukcuoglu,K., Silver,D., Graves,A., Antonoglou,L., Wierstra,D., Riedmiller,M.(2013) . Playing Atari with Deep Reinforcement Learning: DeepMind Technologies. deepmind.com.

[3] Doltsinis, S., Ferreira, P., Lohse,N.(2014) .An MDP Model-Based Reinforcement Learning Approach for Production Station Ramp-Up Optimization: Q-Learning Analysis: IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, VOL. SEPTEMBER.