مجلة جامعة دمشق للعلوم الهندسية
المجلد40 العدد(3) – 2024: 350–359

Damascus University Journal for engineering sciences
Vo 40 number (3) -2024: 350-359

تحسين اداء روبوت ذكاء اصطناعي باستعمال خوارزميات التعلم المعزز العميق

أسامة ابراهيم[1*] سمير منهل كرمان[2] رؤوف مهنا حمدان[3]

[1*]. طالب دكتوراه في قسم هندسة الحواسيب والأتمتة جامعة دمشق.
Osamaibrahim@Damascusuniversity.edu.sy
[2]. استاذ في قسم هندسة الحواسيب والأتمتة جامعة دمشق.
.Samir.karaman@Damascusuniversety.edu.sy
[3]. مدرس في قسم هندسة الحواسيب والأتمتة جامعة دمشق.
Raouf.hamdan@Damascusuniversity.edu.sy

الملخص:

قدمت خوارزميات التعلم المعزز العميق حلا لروبوتات الذكاء الاصطناعي لاكتشاف المسار الافضل والاسرع لها في الوصول الى هدفها من خلال القدرة على التعلم من تجاربهم السابقة واكتساب المعرفة والتمثيل الصحيح لبيئاتهم للوصول لمستوى قريب من الانسان في بيئات العالم الحقيقي المعقدة. تم في هذه الدراسة تحليل اداء خوارزميات التعلم المعزز العميق مثل خوارزمية DQN وخوارزمية Policy gradient وذلك من اجل مساعده روبوت الذكاء الاصطناعي للوصول الى هدفه بأفضل واسرع مسار من خلال ضبط المعاملات العليا للشبكة العصبونية الملتفة العميقة المستخدمة ، ومن خلال مقارنة النتائج تبين تفوق خوارزمية policy gradient بنسبة 50 % تقريبا

الكلمات مفتاحية: التعلم المعزز – الذكاء الاصطناعي–الشبكة العصبونية الملتفة

تاريخ الايداع: 2022/9/7
تاريخ القبول:18 /2022/10

Improving the performance of an artificial intelligence robot using deep reinforcement learning algorithms

Osama Ibrahim[*1] Samir Manhal Karaman[2] Raouf Mhana Hamdan[3]

[*1.] PhD student in Computer and Automation Engineering Department Damascus University. Osamaibrahim@Damascusuniversity.edu.sy

[2.] prof in Computer and Automation Engineering Department, Damascus University. Samir.karaman@Damascusuniversety.edu.sy.

[3.] Lecturer in Computer and Automation Engineering Department, Damascus University. Raouf.hamdan@Damascusuniversity.edu.sy

Abstract:

The Deep reinforcement learning algorithms provided a solution for artificial intelligence robots to discover the best and fastest path to reach their goal through the ability to learn from their past experiences and gain knowledge and correct representation of their environments to reach a level close to human in complex real world environments. In this study, the performance of deep reinforcement learning algorithms such as the DQN algorithm and the Policy gradient algorithm was analyzed in order to help the AI robot reach its target with the best and fastest path by adjusting the higher parameters of the deep convolutional neural network used, and by comparing the results it was shown that the policy gradient algorithm was superior to 50% Approximately.

keywords: Reinforcement learning Artificial intelligence Convolutional Neural network

Received:7 /9/2022
Accepted: 18/10/2022

المقدمة :

قبل بضعة أعوام كان الذكاء الاصطناعي والتعلم الآلي بمثابة مادة خصبة للخيال العلمي فقط الذي نشاهده في أفلام هوليوود[1]، لكن الأمر لم يعد كذلك خلال الألفية الثالثة، حيث تعدت تلك التقنيات نطاق الخيال العلمي لتصبح ذات وجود حقيقي ومتطور في حياتنا بجميع القطاعات . بالتفكير فقط في كثير من الأشياء التي لم تكن موجودة قبل بضع سنوات، مثل سهولة معالجة البيانات الضخمة Big Data، والترجمة الآلية الفورية، وروبوتات الدردشة التفاعلية Chatbots التي يمكنها إجراء محادثات شخصية مؤتمتة مع العملاء على نطاق واسع، وحتى تقنية التزييف الشهيرة (DeepFake) التي تُستخدم في تزييف مقاطع الفيديو بطريقة يصعب كشفها، كان سيبدو الأمر غريبًا قبل أعوام مضت، لكن الآن أصبحت هناك تطلعات لمزيد من تلك التقنيات. هنا ايضا يعتبر (التعلم العميق) Deep learning أحد أنواع (التعلم الآلي) Machine Learning، وهما من أهم تطبيقات الذكاء الاصطناعي، ويُعرف التعلم العميق بأنه عبارة عن تطبيق ذكاء اصطناعي يسمح للأنظمة بتحسين وظائفها تلقائيًا من خلال اكتساب المعرفة من التجربة ثم استخدام الشيء نفسه في معالجة البيانات والحسابات المعقدة، وبناء عليه لن تحتاج الآلات إلى برمجة بشكل منفصل لكل وظيفة حيث أصبح التعلم العميق ممكنًا بمساعدة الوصول إلى البيانات التي جمعتها الأجهزة، وبالتالي تعزيز قدرتها على التعلم، وتختار الشركات التعلم العميق لأنظمتها من أجل تحسين أدائها والحصول على نتائج دقيقة، وتحديد المخاطر التي قد تتعرض لها، والعمل على تجنبها. ونظراً لأن توجه الذكاء الاصطناعي يمكّن الآلات من اتخاذ قرارات سريعة، فإن أهم المجالات التي تستخدم فيها الشركات التعلم الآلي تشمل، أنظمة التنبؤ والتصنيف، والتعرف على الكلام، والرؤية الحاسوبية، والسيارات الذاتية القيادة. ومن ابرز فروع التعلم الالي العميق هو التعلم المعزز العميق،[2] وهو في الواقع فرع من التعلُّم الآلي ، ولكنه ليس مثل التعلم المشترك الخاضع للإشراف والتعلم غير الخاضع للإشراف. يهدف التعلم المعزز إلى اختيار القرار الأمثل ، ويؤكد أنه بموجب سلسلة من السيناريوهات ، فإن القرار المناسب متعدد الخطوات لتحقيق الهدف هو مشكلة صنع القرار المتسلسل متعدد الخطوات، ينصب التركيز على عمل التوقعات. يمكن أن تساعد خوارزميات التعلم المعزز العميق روبوتات الذكاء الاصطناعي في صياغة سلوك مشابه لسلوك الكائنات الحية بدافع المكافأة.

1- الأعمال السابقة:

تتميز صناعة روبوتات الذكاء الاصطناعي بخصائص كثافة التكنولوجيا العالية ، كما تحتل مكانة إرشادية مهمة في التطور العلمي والتكنولوجي الدولي. يدرس[3] تطور صناعة روبوتات الذكاء الاصطناعي في الصين ويدرس المشاكل والحلول التي تواجه هذه الصناعة من اجل تحسين الانتاجية والاداء. يمكن تصنيف خوارزميات التعلم المعزز الى خوارزميات غير المستندة الى نموذج model free ومنها مقاربات مستنده الى سياسة policy based مثل خوارزمية policy gradient في [4] يتم تعلم تابع السياسة policy function،هذا التابع هو طريقة ربط كل حالة والفعل المناسب لها من اجل تحسين اداء الخوارزمية في المركبات ذاتية القيادة. وهناك مقاربات مستنده الى قيمة value basedهنا يهدف العميل الى تحسين تابع القيمة .

v(s)يعرف تابع القيمة بانة التابع الذي يعطينا المكافاة المستقبلية المتوقعة الاكبر التي يمكن ان يحصل عليها العميل في حالة معينة مثل خوارزمية Q learning.اما التعلم المعزز المستند الى نموذج model based الهدف منه تحديد السياسة التي تعطينا افضل نتائج وتحسين مكافأة العميل حيث يتم الاعتماد على نموذج للانتقالات بين الحالات مثل نموذج ماركوف لصنع القرار MDP وخوارزمية مونت كارلو، في [5] تم تطبيق خوارزمية مونت كارلو في عده انظمة روبوتية حقيقية واوضحت النتائج فعالية الخوارزمية في تحسين الاداء .ان الفرق الاساسي بين التعلم المعزز غير المستند الى نموذج والتعلم المعزز المستند الى نموذج انه في التعلم المعزز المستند الى نموذج يتم نمذجه البيئة اي صنع نموذج لتصرف البيئة المحيطة بالعميل وبالتالي تحديد السياسة هذا النمط يتطلب نمذجه بيئة جديده عند دراسة كل حالة وهذا ما يصعب القيام به دائما، فتم التركيز في هذه الدراسة على مقارنه اداء خوارزمية Q من النمط غير المستند الى نموذج مع خوارزمية policy gradient بعد اضافة الشبكة العصبونية العميقة الملتفة CNN اليهما ، في [6] تم دراسة خوارزمية Q العميقة من اجل اتخاذ القرارات للمركبات ذاتية القيادة وبينت النتائج فعالية الشبكة العميقة وفي [7] تم استخدام شبكة عصبونية ملتفه عميقة من اجل استخراج الميزات لتصنيف الصور بأفضل طريقة وفي [8] تم استخدام الشبكة العصبونية الملتفة العميقة لقيادة روبوت متنقل بواسطة كاميرا وماسح ضوئي ليزري وبينت النتائج زياده متانة النظام بنسبة 35% وذلك من خلال ضبط بارا متراتها بشكل يهدف للوصول لأفضل اداء للروبوت. في الوقت الحالي تزايد الاعتماد على الشبكات الملتفة العميقة في الروبوتات بسبب التطور الواضح في تقنيات الابصار الحاسوبي في [9] تم دراسة اساليب قياس المسافات بواسطة كاميرا ستريو لنظام رؤية حاسوبية للتعرف على الوجوه وانواع الكاميرات المستخدمة مع الروبوتات لإدخال الصور الى الشبكات الملتفة بأفضل طريقه ممكنه.

2- ادوات البحث وطرائقه

تم في هذا البحث استخدام مجموعة من المكتبات: TensorFlow Keras,numpy التي تدعم التعلم العميق وتم استخدام مكتبة matplotlib لرسم المنحنيات البيانية وتم استخدام توابع توليد maze حيث توفر بيئة python عدده نماذج لاختبار النتائج كما في الشكل (1).

الشكل (1) بيئة maze

3- توصيف العمل في بيئة maze

عناصر التعلم المعزز هي العميل والبيئة التي يتحرك فيها والإجراءات التي يقوم بها والحالات والمكافئات والسياسة التي يعلم نفسه التحرك بها اعتمادا على المكافأة التي يحصل عليها وبالتالي تكون العناصر في بيئة maze هي:

التفاعل مع البيئة: يكون بالتحرك يمينا او يسارا او اعلى او اسفل.

الحالات : الصور المدخلة وموقع الروبوت في كل صوره.

المكافاة : 1+ عندما يصل الروبوت الى الخلية الهدف.

0.04 – التحرك في خلية حره 0.75 – الاصطدام بالجدار 0.25 – زيارة خلية زارها مسبقا .

السياسة : يحاول الروبوت التحرك بالاتجاهات الاربع من نقطه بداية يتم تعينها في المتاهة الى النقطة الهدف بأفضل واسرع مسار.

4-خوارزمية policy gradient

تعد خوارزمية policy gradient من اهم خوارزميات التعلم المعزز التي تعمل على تحسين هدف التعلم من خلال تحقيق نزول متدرج لبارامترات السياسة على عكس الاساليب القائمة على دالة القيمة كما في خوارزميات (SARSA, Q-learning) والصيغة الرياضية التي تصفها [10]

$$\nabla_\theta J(\theta) = \nabla_\theta \sum_{s\in S} d^\pi(s) \sum_{a\in A} Q^\pi(s,a)\pi_\theta(a|s)$$

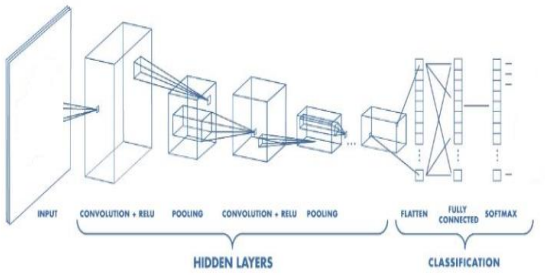
حيث ان $\sum_{s\in S} d^\pi(s)$ هو مجموع توزع ماركوف الاحتمالي للحالات الممكنة, وان $\sum_{a\in A} Q^\pi(s,a)\pi_\theta(a|s)$ هو مجموع توابع خوارزمية Q تحت تدرج السياسة ، حيث ان خوارزمية Q هي احد خوارزميات التعلم المعزز غير المستند الى نموذج التي تسمح للعميل بالعمل في بيئة غير معروفة بفاعلية كبيره بحيث يراكم المكافئات ليصل للهدف بأفضل طريقة والصيغة الرياضية لها بالاعتماد على معادلة بيلمان[11].

$$\boldsymbol{Q(s_t,a_t)} \leftarrow \boldsymbol{Q(s_t,a_t)} + \boldsymbol{\alpha(R_{t+1} + \gamma \max_{at+1} Q(s_{t+1},a_{t+1}) - Q(s_t,a_t))}$$

حيث ان $\boldsymbol{\alpha}$ نسبة التعلم ، $\boldsymbol{\gamma}$ عامل الخصم ، R المكافاة

5-الابصار الحاسوبي والشبكة العصبونية الملتفة العميقة:

يمكننا التفكير ان الشبكة العصبونية الملتفة العميقة [12] لها تأثير جيد على استخراج الميزات المعقدة من الصور وعند دمج تقنيات الابصار الحاسوبي والشبكات العصبونية الملتفة والتعلم العميق نحصل على افضل نتائج لعمليات التصنيف وكما نرى في الشكل (2) تتألف الشبكة العصبونية الملتفة من عده طبقات ملتفه تتم فيا مقارنه الصورة المراد دراستها مع مرشح (كاشف ميزه) بعملية التفاف الهدف من العملية الحصول على اهم الميزات من الصورة ويمكن تغيير نوع وابعاد المرشح حسب الميزة المراد دراستها. بعد الطبقات الملتفة تكون هناك طبقات تجميع الهدف منها تقليل عدد الابعاد لتقليل البارامترات وزمن التدريب في الشبكة ولها نوعين max pooling ,average pooling ، واخيرا طبقات تامه الاتصال مهمتها تسوية النتائج السابقة واستخدام داله التنشيط softmax لاختيار المخارج حسب التطبيق المدروس.

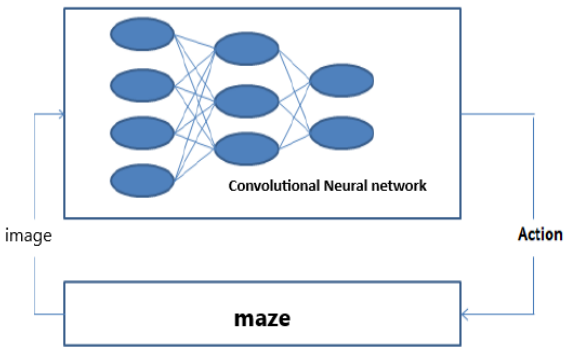


الشكل (2) مخطط الشبكة العصبونية الملتفة العميقة.

في هذه الدراسة تم استخدام خوارزميات التعلم المعزز policy gradient و Q learning في تدريب واختبار الشبكة العصبونية الملتفة العميقة ،حيث ان مداخل الشبكة هي صوره البيئة (maze) التي تتضمن موضع الروبوت في كل خطوه ومخارج الشبكة هي اربع مخارج تمثل التفاعلات الممكنة في البيئة اي تحرك الروبوت يمينا او يسارا او اعلى او اسفل .

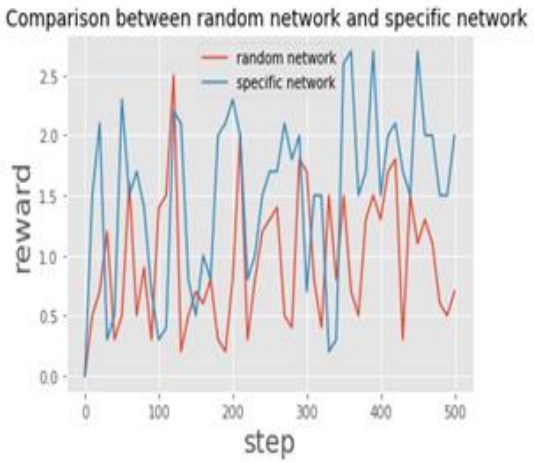
6- التطبيق العملي و النتائج:

6-1 في المرحلة الاولى طبقنا خوارزمية policy gradient على بيئة maze الموجودة في الشكل (1) في بيئة python كما في النظام المقترح التالي :



الشكل (3) المخطط الصندوقي للنظام المقترح.

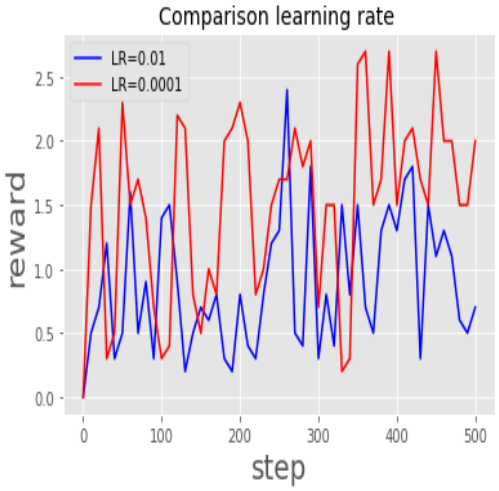
حيث استخدمنا شبكة عصبونية ملتفة عميقة مؤلفة من طبقتين تلافيفيتين وطبقتي تجميع average pooling وثلاث طبقات تامة الاتصال تتألف الطبقة الاولى من 64 نيورون حيث ان مداخلها هي الصورة بعد تطبيق اول مرحلتين عليها ثم القيام بعملية flatten من اجل توافق ميزات الصورة مع مداخل هذه الطبقة والثانية 32 نيورون والثالثة طبقة المخارج بأربع نيورونات تمثل الاجراءات الاربعة للروبوت وتم التحديد بهذه الطريقة لضمان الحصول على افضل اداء للشبكة حسب البروفيسور اندرو [13] وحصلنا على تغير مكافاة العميل خلال 500 دوره تدريب وتمت المقارنة مع اداء الخوارزمية مع شبكة عصبونية ملتفة عميقة مبنية بشكل عشوائي كما في الشكل(4).



الشكل (4) مقارنة اداء خوارزمية policy gradient مع الشبكة العصبونية الملتفة العميقة المحددة والعشوائية.

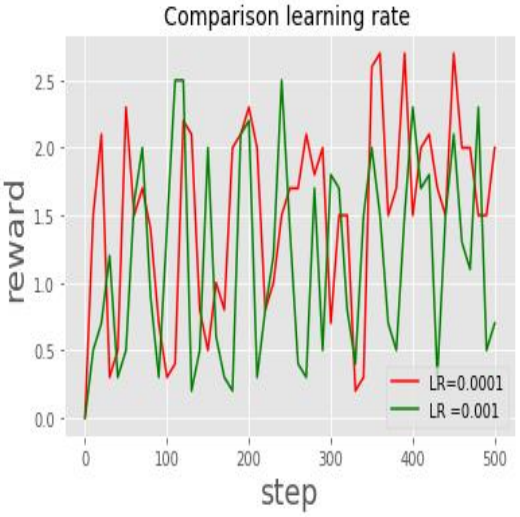
من الشكل (4) نلاحظ التفوق الواضح للشبكة المحددة وذلك بسبب تمثيل النيورونات بشكل قوى العدد 2 المتناقصة المماثلة لتوزيع مساري المعالج وعناوين الذاكرة مما يضمن افضل اداء للحاسب.

2-6 في المرحلة الثانية تم اختبار اداء الخوارزمية مع الشبكة العصبونية الملتفة العميقة من اجل نسب تعلم مختلفة وحصلنا على تغير مكافاة العميل خلال 500 دوره تدريب كما في الشكل(5).



الشكل (5)مقارنة اداء خوارزمية policy gradient مع الشبكة العصبونية الملتفة العميقة لنسب تعلم مختلفة.

من خلال الشكل (5) نلاحظ تحسن مكافاة العميل بشكل واضح من خلال تقليل نسبة التعلم والشكل (6) يؤكد الفكرة لقيم اخرى من نسبة التعلم والجدول التالي يبين زمن دورات التدريب والاختبار لكل نسبة.



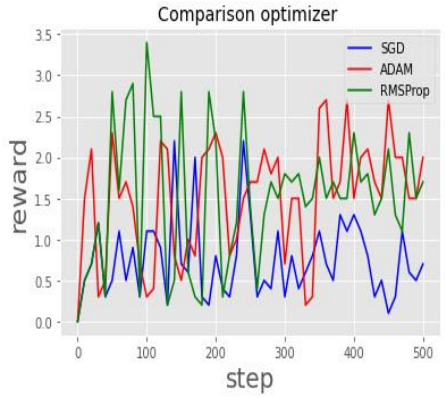
الشكل (6) مقارنة اخرى لأداء خوارزمية policy gradient مع الشبكة العصبونية الملتفة العميقة لنسب تعلم مختلفة.

الجدول (1) زمن حل ونسبة تحسن اداء العميل مع تغير نسبة التعلم.

المرحلة	نسبة التعلم	عـــــدد الدورات	زمـن تـدريب الشبكة	نسبة التحسن
1	0.01	500	1 hours	
2	0.001	500	1.4 hours	20%
3	0.0001	500	1.8 hours	55%

من الجدول (1) نلاحظ تحسن قيمة المكافاة للروبوت عند تقليل نسبة التعلم على حساب زمن تدريب الشبكة ولكن كون عملية الاختبار هي الاهم نهتم بتحسن المكافأة.

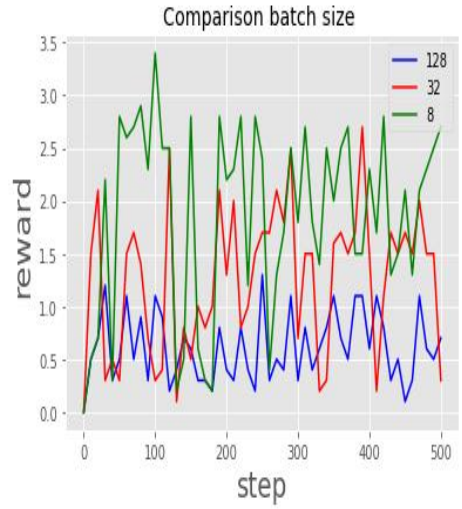
3-6 في المرحلة الثالثة تم اختبار اداء خوارزمية policy gradient مع شبكة عصبونية ملتفة عميقة من اجل عدة انواع من المحسنات وحصلنا على تغير مكافاة العميل خلال 500 دوره تدريب كما في الشكل(7).



الشكل (7) مقارنه اداء خوارزمية policy gradient مع شبكة عصبونية ملتفة عميقة محددة من خلال استخدام انواع مختلفة من المحسنات

في الشكل (7) استخدمنا ثلاث انواع مختلفة من المحسنات,ونلاحظ تحسن مكافاة العميل بشكل واضح عند استخدام المحسن RSMProp على النوعين الاخرين وايضا تفوق المحسن ADAM على اداء المحسن SGD

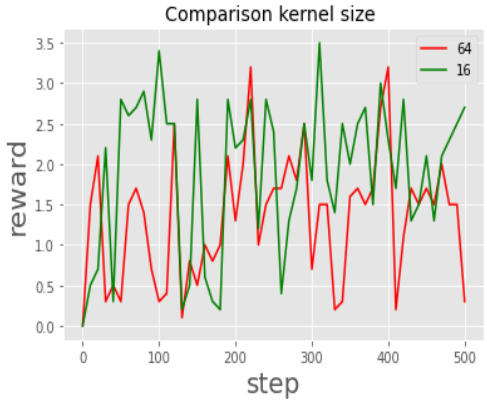
4-6 في المرحلة الرابعة تم اختبار اداء خوارزمية policy gradient مع شبكة عصبونية ملتفه عميقة من اجل قيم مختلفة من batch size وحصلنا على تغير مكافاة العميل خلال 500 دوره تدريب كما في الشكل (8).



الشكل (8) مقارنه اداء خوارزمية policy gradient مع شبكة عصبونية ملتفة عميقة محددة من خلال استخدام قيم مختلفة لل batch size

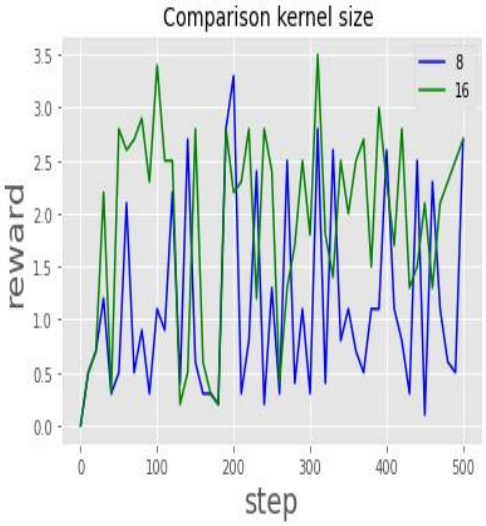
في الشكل (8) نلاحظ تحسن قيمة المكافأة عند تصغير قيمه ال batch size.

6-5 في المرحلة الخامسة تم اختبار اداء خوارزمية policy gradient بشبكة عصبونية ملتفة عميقة من اجل قيم مختلفة من ال kernel size وحصلنا على تغير مكافأة العميل خلال 500 دوره تدريب كما في الشكل (9) .



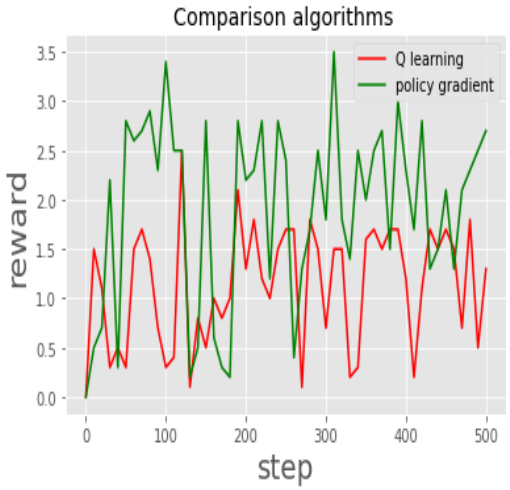
الشكل (9) مقارنه اداء خوارزمية policy gradient مع شبكة عصبونية ملتفة عميقة محددة من خلال استخدام قيم مختلفة من kernel size

من الشكل (9) نرى اهمية اختيار قيمة وسطية لأبعاد المرشح المستخدم في الطبقات الملتفة للشبكة كما نرى في الشكلين (9) و (10).



الشكل (10) مقارنه اخرى لأداء خوارزمية policy gradient مع شبكة عصبونية ملتفة عميقة محددة من خلال استخدام قيم مختلفة من ال kernel size

6-6 في المرحلة الاخيرة وبعد ضبط اهم البارامترات العليا للشبكة العصبونية الملتفة العميقة المستخدمة مع الخوارزميتين تمت مقارنة اداء الخوارزميتين policygradient , Q learning باستخدام نفس البارامترات اي بعد تحسين كلا الخوارزميتين باستخدام نفس البارامترات السابقة ونفس بنيه الشبكة في بيئة maze وحصلنا على تغير مكافاة العميل خلال 500 دوره تدريب كما في الشكل (11).



الشكل (11) مقارنه اداء خوارزمية policy gradient مع شبكة عصبونية ملتفة عميقة مع خوارزمية Q learning لنفس الشبكة.

من الشكل (11) نلاحظ التفوق النسبي لخوارزمية policy gradient بمقدار 50%.

7-الاستنتاجات و التوصيات:

من خلال الدراسة السابقة لبيئة maze تبين تفوق خوارزمية policy gradient على خوارزمية Q learning لروبوت الذكاء الاصطناعي بعد ضبط قيم البارامترات العليا للشبكة العصبونية الملتفة العميقة بنسبة 50% وبقي لنا ان نذكر انه يمكن تطوير اداء هدة الخوارزميات مستقبلا للتغلب على

مشكلة الزمن الكبير لعملية تدريب الشبكة العصبونية العميقة بإحدى الطرائق التالية

1- استخدام تقنية الانحدار الاشتقاقي للمجاميع الصغيرة mini batch gradient descent في المحسن المستخدم لتقليل زمن تدريب الشبكة .

2- تطبيق النظام المقترح على عميل حقيقي وبالتالي استخدام كاميرا بالا ضافة لتقنيات الليزر في تحديد الموقع لتحسين اداء الروبوت .

التمويل: هذا البحث ممول من جامعة دمشق وفق رقم التمويل (501100020595).

References:

[1] Ongsulee ,P. (2017). Artificial Intelligence, Machine Learning andDeep Learning Reinforcement Learning: Siam UniversityBangkok, Thailand , IEEE.

[2] 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.

[3] Lin,W.,(2022) Development of artificial intelligence robot industry in the era of big data: IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC).

[4] Aradi,S. , Becsi,T., Gaspar,P.(2018). Policy Gradient based Reinforcement Learning Approach for Autonomous Highway Driving: IEEE Conference on Control Technology and Applications (CCTA) Copenhagen, Denmark, August 21-24, 2018.

[5] Amadio,F., Dalla,A.,Antonello,R.,Nikovski,D.,Carli,R., Romeres,D.(2021) Model-Based Policy Search Using Monte Carlo Gradient Estimation with Real Systems Application: IEEE Transactions on Robotics. MC-PILCO arXiv:2101.12115v4.

[6] Ronecker,M., Zhu ,Y.(2019).Deep Q-Network Based Decision Making for Autonomous Driving :2019 3rd IEEE International Conference on Robotics and Automation Sciences.

[7] Saini,A. , Gupta,T. , Kumar,R. , Gupta ,A., Panwar,M. , Mittal,A.(2017). Image based Indian Monument Recognition using Convoluted Neural Networks: 2017 International Conference on Big Data, IoT and Data Science (BID) Vishwakarma Institute of Technology, Pune, Dec 20-22, 2017.

[8] Sadeghi Esfahlani 1,S. , Sanaei,A., Ghorabian ,M., Shirvani,H.(2022). The Deep Convolutional Neural Network Role in the Autonomous Navigation of Mobile Robots (SROBO). Remote Sens. 2022, 14, 3324.

[9] DANDIL,E., Kürşat,K..,(2019). Computer Vision Based Distance Measurement System using Stereo Camera View, 978-1-7281-3789-6/19/$31.00 ©2019 IEEE.

[10] Jan Peters, J. Bagnell, Policy gradient methods Published in Scholarpedia 26 3698Corpus ID: 1988822(2010)،November, .2010DOI:10.4249/scholarpedia

[11] Donoghue,B., Osband,I., Munos,R., Mnih,V.(2018). The Uncertainty Bellman Equation and Exploration: arXiv:1709.05380v4 [cs.AI] 22 Oct 2018.

[12] Akshaya, B., Prof. Kala M,T.(2020) . Convolutional Neural Network Based Image 978-1-7281-7590-4/20/$31.00 ©2020 IEEE.

[13] Andrew, Ng.(2017). Machine Learning: Stanford university , http://cnx.org/content/col11500/1.4/.

جدول المصطلحات

المصطلح العربي	المعنى الانكليزي
الشبكة العصبونية الملتفة	convolutional neural network
الذكاء الاصطناعي	artificial intelligence
التعلم المعزز	reinforcement learning
التعلم الآلي	Machine learning
الابصار الحاسوبي	Computer vision
المكافأة	reward