TEST Engineering & Management

January-February 2020
ISSN: 0193-4120 Page No. 6573 - 6577

An Enhanced Optimization Technique to Implement Dino Game using Machine Learning Techniques

Amara Roshini Roy[1], M. Nalini[2], D. Shiny Irene[3]

[1]UG Scholar, Department of Computer Science and Engineering
Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences Chennai, India
[2]Assistant Professor, Department of Computer Science and Engineering
Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences Chennai, India
[3]Assistant Professor, Department of Computer Science and Engineering
Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences Chennai, India
Email: rramara21@gmail.com, nalini.tptwin @gmail.com, dshinyirene@gmail.com

Article Info
Volume 82
Page Number: 6573 - 6577
Publication Issue:
January-February 2020

Article History
Article Received: *18 May 2019*
Revised: *14 July 2019*
Accepted: *22 December 2019*
Publication: *01 February 2020*

Abstract
In this task, we put in force both characteristic-extraction based algorithms and a give up-to-cease deep reinforcement gaining knowledge of approach to learning to manipulate Chrome offline dinosaur recreation at once from excessive-dimensional recreation display screen input. Results show that as compared with the pixel feature primarily based algorithms, deep reinforcement learning is extra powerful and powerful. It leverages the high-dimensional sensory enter at once and avoids capacity errors in feature extraction. Finally, we advise unique schooling methods to address magnificence imbalance issues caused by the boom in-game speed. After education, our Deep-Q AI can outperform human specialists.

Keywords: *Deep Q-Learning; MLP; Feature Extraction with OpenCV.*

1. Introduction

Learning human-degree manage rules in real time from high- dimensional sensory statistics (inclusive of originative and prescient) could be a protracted-status task for dominant machine layout. Several classical manipulate algorithms area unit supported correct modeling or space info of the underlying dynamics at intervals the convenience. However, for structures with unknown dynamics and excessive dimensional input, these ways area unit unreasonable and exhausting to generalize.

In this assignment, we have a tendency to advocate Deep Q-studying, on-line learning Multi-Layered Perceptron (MLP) and rule-based total decision-making (Human Optimization) for aiming to grasp to control the sport agent in T-Rex, the classic game embedded in Chrome offline mode, in real time from excessive-dimensional image enter. All techniques area unit based totally on 2 methods of kingdom- space simplification. The primary method is to extract pixel-based capabilities and see things the usage of laptop originative and discerning processes for imitating human participants, like MLP. The various are to automatically examine patterns from resized raw recreation pictures while not manual feature extraction, like Deep Q-getting to understand. To be distinctive, the input to Deep Q-getting to understand a collection of rules could be a stack of four pictures. We have a tendency to then use reinforcement aiming to grasp to update the schooling samples in neural community and leverage a Convolutional Neural Network (CNN) to expect that movement to require at a lower place given instances. Similarly, we have a tendency to take the extracted constituent capabilities as input and use MLP or rule-based decision- creating (Human Optimization) for prediction. The consequences reveal that our approaches drastically exceed even the toughened human participant within the game.

TEST Engineering & Management

January-February 2020
ISSN: 0193-4120 Page No. 6573 - 6577

We applied the extraction of pixel-based capabilities from the game screenshot and also the MLP rule. We have a tendency to boot recognition at the implementation of the Q-studying framework, and also the comparison between hand-coded on-line gaining data of ways and deep reinforcement learning. We have a tendency to recognition our efforts on the deep Q-mastering network model, the picks of hyper parameter of coaching, the distinctive schooling approach, and analysis of our education technique. We have a tendency to boot analyze the impact of acceleration in T-Rex and examine our sport with alternative games.

2. Related Works

In recent years, Reinforcement Learning marketers have applied powerful overall performance in a number of the utmost exhausting classical video games. This was at first finished board game in 2002 and larger recently a set of ATARI video games and also the game of bear Alpha Go. Thanks to combined upgrades in Deep Learning and also the provision of reasonably-priced procedure energy, formerly an intractable problem is currently within earned.

This new technology of agents combines the ability of supervised aiming to grasp reinforcement finding out and Deep neural networks give {a method|how|some way the way the simplest way} wont to educate retailers to maximise rewards by way of interacting with surroundings. Supervised aiming to grasp is employed with sport replay records (nation, moves, and reward tuples) to approximate capabilities employed in numerous RL algorithms.

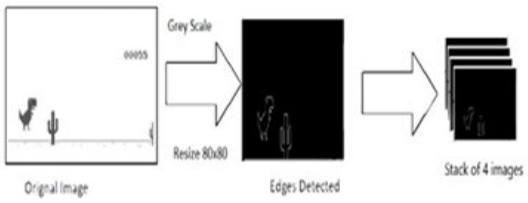

Figure 1: Process of the Image

In advanced video games that embody Go, researchers have relied extensively on skilled sports datasets for coaching - this outstanding data point is usually robust to attain, and not generalizable across problems.

3. Methodology

Datasets

This enterprise is applied within the PyCharm with the assist of the PyGame package, this package deal is employed to extract pictures. Entire paintings area unit dead within the PyCharm solely with the assist of a number of totally different packages like [1]Image Grab: wont to copy the content material of the prevailing show, ImageOps: wont to remap the photograph for the constituent calculations, pyautogui: this is often a user interface automation module. On the portable computer monitor, pictures area unit mounted with the 1200X300 set length of pixels. There are a unit *some ways to convert raw pixels* victimization information filtration or preprocessing.

Preprocessing

A fashionable preprocessing is employed within the Q-learning model, for the 80x80 grid of pixels conversion is achieved by the employment of the grid-scale once re-scaling the photographs. With this, the result's 80x80x4 enter is made to the deep Q-gaining data of method.

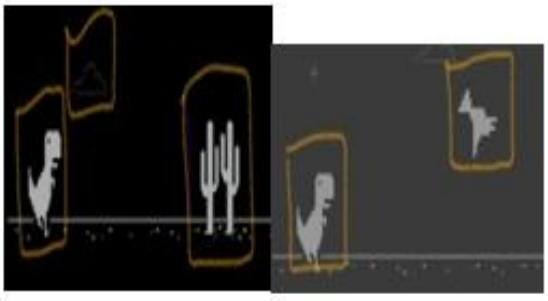

Figure 2: Image Filtering and object Detection

In-game there are a unit varied gadgets area unit there the primary issue is classification is dead per them in line with the article by method of extracting the pixels. This item is denoted victimization the sq. bins once the detection action is accomplished to testify the task objective.

Object detection in all probability turns into larger easier as a result of filtration is completed earlier. Fig.2. Denotes the gadgets finite with orange coloration packing containers. Objects area unit measured and known by taking the x, y coordinates of the item. In these Objects area unit arch saurian reptile, chicken, poles. Objects area unit labelled through the usage of the height and dimension of the bounding boxes. Object observance is that the toughest activity within the style thanks to the actual fact calculation has completed primarily based entirely on the rate and pace of the agent and its interaction with the environment at that specific time supported the movement and kind of the detected item it tracks the article. Menstruation the things frame by frame for observance whereas the bodysuits following is captured.

Training Models

Any version is trained victimization the suggestion of the consultants at the sector, right here professionals area unit human players as players aforementioned that jump the boundaries which are not positive at intervals the height and dimension. Speed and speed plays an important position to show the agent whereas

TEST Engineering & Management

January-February 2020
ISSN: 0193-4120 Page No. 6573 - 6577

the course is enclosed speed will increase thereupon step by step pace to boot goes higher.

Multi-layered Perceptron

A multilayer perceptron could be a deep artificial neural network. During this bedded methodology, signals area unit vital signals area unit reckoning on the enter layer interaction with the output layer by method of taking the alternatives ready to predict the hidden layers of the perceptron. Applying the multi-layered perceptron for the education of a set of enter-output pairs and capable of examining the model of correlation among the inputs and outputs of the one.
In the below table A is that the input "1", B is that the enter "2" and output is "C.

Table 1: XOR Table

A	B	C
0	0	0
0	1	1
1	0	1
1	1	0

XOR characteristic is employed to allow the neural network for the assumption of non-linear sort. XOR perform suggests that when 2 inputs area unit identical the motion is do nothing and inputs area unit unequal means that it should be acted per educated beneath the version.

Deep Reinforcement Learning

In reinforcement finding out technique selecting the Q-studying set of rules improves the potency of the method toward the outcomes.

In this complete methodology, agent needs to manipulate itself as a result of, because the rate will increase at constant time as jump the pole or object landing are often varied once in a very whereas at that point agent needs to managed via itself.

Q-learning Modules

When the agent gets practiced with the help of following the positive rules {and strategies and *methodologies* and techniques} there upon ready to take movement con to the agent the usage of the Q-Learning method. Obtained observations from the sport are: By the usage of the Bellman's equation transition for the state, space is Tuple by method of the usage of the values no heritable at intervals the on top of observations in step with the specified movement.

$$Q_\pi(s,a) = \sum_{s'} T(s,a,s')[Reward(s,a,s') + \gamma \max_{a'} Q_\pi(s',a')]$$

Tuple must be up so far in real time once each statement is up so far.

$$Q_{opt}(s,a) \leftarrow Q_\pi(s,a) - \eta[Q_{opt}(s,a) - u]$$
$$= Q_{opt}(s,a) - \eta[Q_{opt}(s,a) - (r + \max_{a'} Q_{opt}(s',a'))]$$

In CNN to feature the letter fee of each country it needs the pictures with pixels improvement for the reading of these method have to be compelled to be optimized to attain the observations:

$$\min_w \sum_{(s,a,r,s')} (Q_{opt}(s,a) - (r + \max_{a'} Q_{opt}(s',a')))^2$$

The backward propagation technique is employed to optimize the procedure neural community. For the education reason, there is a need of excessive reminiscence to avoid wasting the remark values no heritable at every body. To show neural community reminiscence can every which way choose the records helpful for coaching.

Objects area unit measured and known by taking the x, y coordinates of the item. In these Objects area unit arch saurian reptile, chicken, poles. We have a tendency to recognition our efforts on the deep Q-mastering network model, the picks of hyper parameter of coaching, the distinctive schooling approach, and analysis of our education technique.

CNN Design

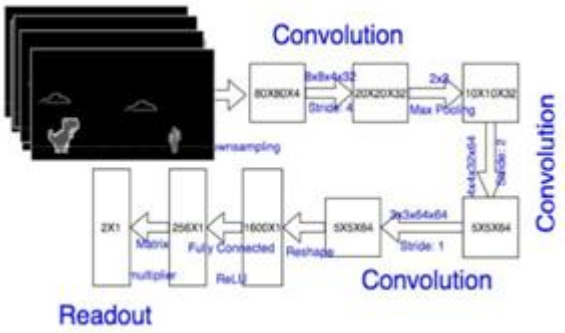

Figure 3: CNN Architecture

Neural Network: The convolutional neural community used at intervals the reinforcement mastering of consists of three layers and one goop pooling layer and 2 associated layers. Taking the stack of 4 pictures and associate with CNN layer the usage of thirty two filters of size 8x8 in the course of victimization the CNN layer. Finally, the planate tensor can bear 2 utterly coupled layers and outputs the letter values for two moves.

4. Results and Discussion

In this section, we examine the effects of MLP model. Given a list of obstacles to examine the optimized jumping role, we notice the average rankings stop to growth over 900 rounds new release. In addition to the deviation of picture processor, we agree with the overpowering computation over parameters in a fully-related network undermines the overall performance

January-February 2020
ISSN: 0193-4120 Page No. 6573 - 6577

as well. To assist it, we behavior a comparable manner, elegance imbalance occurs with some distance fewer education items in high speed instances comparatively.

Table 2: Comparison between different model

Algorithm	*Average 20*	*Max*	*Std*
Keep-jump	*41*	*111*	*23*
Human-expert	*910*	*1500*	*420*
Human-optimized	*196*	*4670*	*134*
MLP	*469*	*1335*	*150*
Deep Q-learning	*1216*	*2501*	*678*

The imbalance might make the agent even less in all likelihood to respond to high velocity mode after a thousand ratings. In addition, G-grasping prohibited the agent from getting higher rankings. It is because random jump could result in surprising crash into the obstacles. However, we should not cancel the G-grasping under excessive-velocity mode completely. Otherwise, we can play to the excessive pace mode to cowl the imbalance instances however no exploration takes place beneath this case.

To make the agent now not best see the high pace mode however additionally hold the exploration, we use any other education approach in the TRAINING PHASE III. We train our agent with large acceleration for zero.5 million iterations and preserve the G- grasping on the same time. During those iterations, the agent will observe a lot of high velocity samples. Subsequently, the model is retrained for an extra ten thousand iterations below normal acceleration mode to get used to the original sport. This training technique in large part will increase the overall performance of our algorithm and make the average rating a 25% increment. The maximum score additionally reaches the 2500 factors with an increment of 500 factors. As a end result, the overall performance of our set of rules after 3 training stages is higher than the human professionals.

5. Future Work

In this project, all observations area unit glad and got outcomes. At the time of version schooling, the matter is speed and pace management for that model schooling has become further advanced. So, in destiny by victimization further realistic and customary ways this downside may well be solved thereupon no want for any advanced fashions. In our challenge, neither MLP nor Deep Q-mastering handles pace nicely sufficient. In Deep Q-gaining knowledge of, we observe velocity makes first rate impact over the jumping role choice.

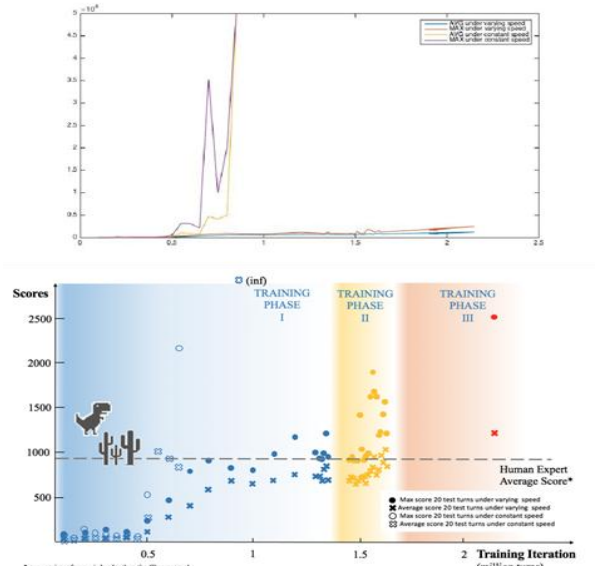

Figure 4: Training curve of DQL

Firstly, we simplify the computation in neural networks with the aid of resizing the uncooked sport picture into 80 pixels. In this manner, the threshold of the boundaries could be difficult to understand which negatively inspired the prediction.

6. Conclusion

In this undertaking by using the usage of the many techniques, technique and procedures consequences are received with an impeccable document to defeat any human participant this all executed due to schooling of the neural networks by CNN. Feature extraction algorithms are cautiously designed to achieve the frames and states of the agent. Finally, bot inside the T-Rex sport will interact with the gaming interface parallelly with the PyCharm display such that bot will get fulfillment automatically in the game.

Reference

[1] G. Bradski et al. The opencv library. Doctor Dobbs Journal, 25(eleven):a hundred and twenty–126, 2000.

[2] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller. Playing atari with deep reinforcement getting to know. ArXiv preprint arXiv: 1312.5602, 2013.

[3] V. Padmanaban and Nalini, M. , Adaptive Fuel Optimal and Strategy for vehicle Design and Monitoring Pilot Performance, Proceedings of the 2019 international IEEE Conference on Innovations in Information and Communication Technology, Apr 2019. [DOI>10.1109/ICIICT1.2019.8741361]

[4] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Belle mare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et

January-February 2020
ISSN: 0193-4120 Page No. 6573 - 6577

al. Human-stage control via deep reinforcement mastering. Nature, 518(7540):529– 533, 2015.

[5] Shanmuga Sai, R., Priyadarsini, U., and Nalini, M, Cooperative Quality Choice and Categorization for Multi label Soak Up Process, Proceedings of the 2019 international IEEE Conference on Innovations in Information and Communication Technology, Apr 2019. [DOI > 10.1109/ICIICT1.2019. 8741469]

[6] G. Tesauro. Temporal distinction studying and td-gammon. Communications of the ACM, 38(three):58–sixty eight, 1995.

[7] Uma Priyadarsini and Nalini, M, Transient Factor- Mindful Video Affective Analysis- A Proposal for Internet Based Application, Proceedings of the 2019 international IEEE Conference on Innovations in Information and Communication Technology, Apr 2019. [DOI >10.1109/ICIICT1.2019.8741466]

[8] Gerald Tesauro. Programming backgammon the usage of self-coaching neural nets. Artificial Intelligence, 134(1- 2):181–199, 2002.

[9] Shiny Irene D., G. Vamsi Krishna and Nalini, M., "Era of quantum computing- An intelligent and evaluation based on quantum computers", Published in International Journal of Recent Technology and Engineering (IJRTE), Vol. 8, Issue no.3S, pp. 615- 619, October 2019.[DOI> 10.35940/ijrte.C1123.1083S19]

[10] VolodymyrMnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing Atari with Deep Reinforcement Learning. Pages 1– nine, 2013.

[11] G. Tesauro. Temporal distinction gaining knowledge of and td-gammon. Communications of the ACM, 38(three): fifty eight–68, 1995.

[12] Nalini, M. and Uma Priyadarsini, To Improve the Performance of Wireless Networks for Resizing the Buffer, Proceedings of the 2019 international IEEE Conference on Innovations in Information and Communication Technology, Apr 2019.[DOI >10.1109/ICIICT1.2019.8741406]

[13] Y. Bengio. Learning deep architectures for AI. Foundations and trends in Machine Learning, 2(1):1–127, 2009.

[14] Nalini, M. and Abu, S., "Anomaly Detection Via Eliminating Data Redundancy and Rectifying Data Error in Uncertain Data Streams", Published in International Journal of Applied Engineering Research (IJAER), Vol. 9, no. 24, 2014.

[15] M. Kearns, Y. Mansour, and A. Y. Ng. A sparse sampling algorithm for close to-most efficient making plans in big Markov selection methods. Machine Learning, 49(2-three):193–208, 2002.

[16] L. Kocsis and C. Szepesvari. Bandit based Monte-Carlo making plans. In Machine Learning: European Conference on Machine Learning 2006, pages282–293. Springer, 2006.

[17] J. Rene Beulah and D. Shalini Punithavathani (2017). "A Hybrid Feature Selection Method for Improved Detection of Wired/Wireless Network Intrusions", Wireless Personal Communications, vol. 98, no. 2, pp. 1853-1869 (Springer).

[18] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller. Playing Atari with deep reinforcement getting to know. In Deep Learning, Neural Information Processing Systems Workshop, 2013.

[19] G. Tesauro. Temporal difference gaining knowledge of and TD-gammon. Communications of the ACM, 38(3):fifty eight– 68, 1995.

[20] J. Rene Beulah and Dr. D. Shalini Punitha vathani (2015). "Simple Hybrid Feature Selection (SHFS) for Enhancing Network Intrusion Detection with NSL-KDD Dataset", International Journal of Applied Engineering Research, Vol. 10, No. 19, pp. 40498-40505

[21] Russel, S. And Norvig, P. (2003). Machine Learing: A Modern Approach. Second Edition. New York: Prentice-Hall.

[22] Baldi, P., Frasconi, P., Smyth, P. (2003). Modeling the Internet and the Web - Probabilistic Methods and Algorithms. New York: Wiley.

[23] Nalini, M. and Anvesh Chakram, "Digital Risk Management for Data Attacks against State Evaluation", Published in International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol. 8, Issue no. 9S4, pp. 197-201, July 2019.[DOI:10.35940/ijitee.I1130.0789S419]

[24] J. Schmidhuber. Deep gaining knowledge of in neural networks: An evaluate. Neural Networks, 2014.