TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 3074 - 3081

A Novel Control Scheme for Variable Load Drive Systems with Reinforcement Learning

Reynato Andal Gamboa, Aravind CV, Gowthamraj R.

Article Info
Volume 82
Page Number: 3074 - 3081
Publication Issue:
January-February 2020

Article History

Article Received: *14 March 2019*
Revised: *27 May 2019*
Accepted: *16 October 2019*
Publication: *19 January 2020*

Abstract
Abstract—In this paper, the authors proposed a novel Reinforcement Learning based controller for the optimal control of a variable frequency drive (VFD) system for an induction motor. Traditional research in this area concentrates on a different level of analysis, that of operational control – e.g. different PWM schemes to create an idealize output – while this research looked on operational optimization – to select the action given the state of the device. The agent is based on a class of algorithm known as Policy Gradient, and the entirety of the study was done through computer simulation. The agent was trained and tested on a series of complex spatial-temporal load sequences to demonstrate its robustness and generalizability. In addition, the agent does not need any explicit knowledge of the device model itself, it could learn just by observing a set of user defined observations of the environment and receiving a reward signal as feedback for its action. The novel control scheme outperforms an uncontrolled benchmark significantly in multiple areas such as power factor, slip performance, among others by a significant 25% margin.

Keywords: *Artificial Intelligence, Control Systems, Reinforcement Learning, VFD..*

I. INTRODUCTION

As the world enters into the late quarter of the 21st century, energy demand and consumption have sky-rocketed due to our ever-increasing population and activity. According to the International Energy Agency (IEA), the demand for electricity has increased by 3.1% from 2016 to 2017 and the consumption of energy across all sources in the same period have grown by 2.1%, twice more than the previously observed growth rate in the former years. A significant contributor to these consumptions is the usage of electric-motor driver systems (EMDS), whose applications can be found in all sectors and industries. For example, actuators for conveyors, pump and fan for processing plants in the industry; large-scale heating, ventilation, air-conditioning systems and elevators in commercial buildings; air-conditioners, refrigerators, washing machines and so on in domestic settings. Table 1 shows the breakdown and details of EMDS electricity consumption by sectors . It can be noticed from Table 1 that the industrial sector accounts for 63% of the system consumption, while the transport and agriculture sector has the least at 4%. It is therefore important if the efficiency in the industrial processes be managed to significantly reduce energy consumption.

Table 1. Breakdown of Energy Consumption By Sector

Sector	Energy Consumption (TWh/year)	Percent Consumption Share
Industrial Sector	4,488	63%
Commercial Sector	1,412	20%
Residential Sector	948	13%
Transport and Agriculture Sector	260	4%

There are several programs to reduce energy consumption, that is, to vary the device utilization with respect to the on-load demand. Two common energy efficiency programs for HVAC application are the adoption of dampers and inlet guide vanes and the variable frequency drives (VFD) systems for variable speed control . Fig. 1 shows a comparison of performances driving a common load with some of the control schemes as laid out above.

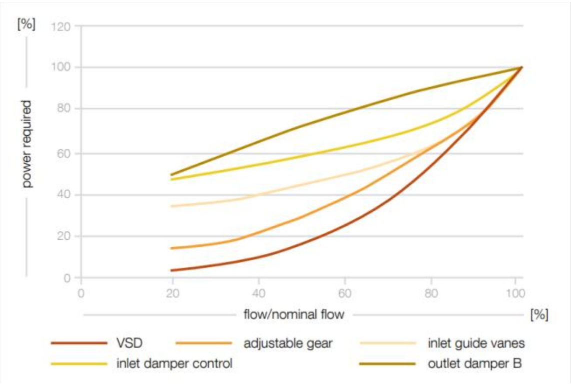

Fig. 1. Performance Comparison of Different Pump Motor Control Schemes

One of the control methodologies for the VFD itself is to vary its voltage-frequency ratio linearly across the range of operation as shown in EqThis particular approach is preferable over pure frequency or voltage augmentation due to its advantage such as robustness against flux saturation encountered in pure frequency control scheme.

$$a_{v/f} = \frac{v}{f}$$

This, however, introduces a new problem of optimal control, that is to determine the relationship between $a_{v/f}$ and other parameters of the systems such as power factor, current drawn by the motor, power consumed, resulting torque and stator-rotor slip from the adjustment with respect to the load, and to be able to derive an optimal value for $a_{v/f}$ with respect to all these parameters dynamically across range of different loads. It is important to note however, the relationship among the aforementioned parameters are non-linear in nature and thus they are non-trivial to optimize.

Traditional control schemes such as PID requires extensive modelling of the entire system itself before any viable solution could be derived, which may be prohibitive in highly complex systems with many inter-dependencies.

Artificial Intelligence method such as Reinforcement Learning (RL) on the other hand thrives in scenarios where the internal state of a system is highly complex, and thus would be unfeasible to be explicitly define in addition to being non-trivial to optimize. It achieves this feat by learning how to act in an environment through a mixture of exploration and exploitation with the aid of a reward signal for each of its actions. With only modelling of the input, output and the evaluation of the particular output with respect to the environmental state, one would be able to achieve convergence to a sufficiently optimal solution. Thus, in this paper, the researchers adopted a novel control method to approximate an optimal control scheme for said drive systems with an Artificial Intelligence based approach, specifically, to control a VFD drive with a RL agent. In this paper, the researchers derived and prepared a custom environment with an approximate model of a VFD-driven inductor motor to act in said environment, and an agent based on the Policy Gradient (PG) algorithm – a class of RL method.

Previous work on control optimization for VFD have been attempted, but the topics revolve around the use of traditional classical schemes such as the Proportional – Integral - Derivative (PID) controller or method that require explicit modelling of the intended environment in its entirety, and a pseudo-optimal tuning procedure thereafter. Other works addresses the challenge from a different level of analysis, that of the optimal operational control – not of operational optimization – of the input signals to the machine based on a series of clean and noisy feedback signals to approximate an idealize input. On the other hand, this paper seeks to create an agent that could optimally choose what the input signal itself should be, or select the best action given a set of environmental states. Control schemes such as PID would not be ideal in the task of action selection of this kind without adopting a much more complex setting.

In previous advances in RL for the task of complex non-linear system optimization and control, there were a plethora of attempts as the practitioners of the subject itself have studied many of such problems. An example would be the cart pole problem, where an agent is required to swing-up and balance an upright pole hinged onto a cart by moving the cart itself. A more sophisticated version of this would be the double pendulum problem, where, instead of a single pole, an agent is required to balance a series of upright joint poles by

January - February 2020
ISSN: 0193 - 4120 Page No. 3074 - 3081

controlling the cart as shown in Fig. 2 and Fig. 3 respectively. Solving these types of chaotic non-linear problems by traditional control schemes are a significant challenge, especially in problems that are even more complex which could be found in abundance in the real world. The current task of control of a VFD induction motor is akin to the former application of these algorithms and agents on the aforementioned classical non-linear control problems.

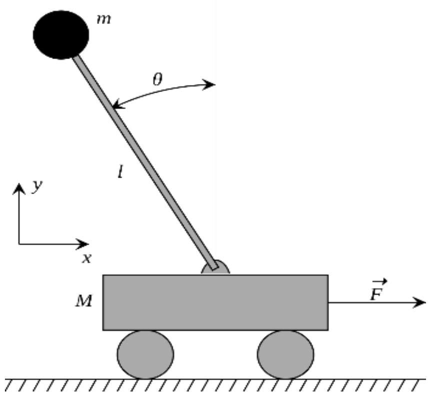

Fig. 2.The classic cartpole problem whereby the pole needs to be maintained upright by moving the cart horizontally.

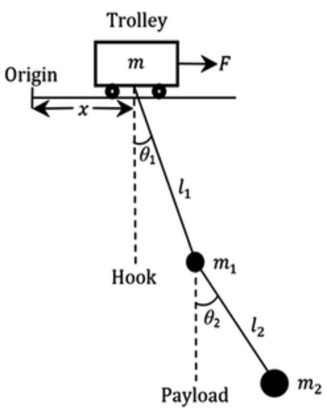

Fig. 3.The double-pendulum problem whereby both the conjoined poles need to be swung-up and maintained upright by moving the cart horizontally.

In addition, there were a series of significant breakthrough in this particular field of study in recent years where agents of these methods have been able to achieve superhuman level of performance in a multitude of tasks such as Go, playing competitive video games, and so on.

RA could be divided into a few broad categorizations of value-based learning, on-policy learning and model-based learning, though they are not mutually exclusive. Model-based learning agents seeks to learn by modelling the environment itself, which may prove to be intractable in complicated tasks and environments. Thus, the model-free methods are usually preferred over such an explicit approach. The two most popular category of learning algorithms are of the Q-learning and Policy Gradient approach, though the two approaches could be combined to form another class of robust algorithms known as Actor-Critic. Most recent methods that have enabled fundamental breakthroughs utilizes Policy Gradient algorithms instead, such as the Deep Deterministic Policy Gradient algorithm for continuous environment control systems and the Proximal Policy Optimization family of algorithms which is the current state of the art in videogames playing, robotic locomotion and so on.

II. METHODOLOGY

Fundamentally, a RA agent would need an environment to function, and a reward function that appraises the value or significance of its action towards a pre-determined end goal as shown in Fig. 4. In this case, it is to select the best $a_{v/f}$value for a series of differing loads that would optimize for power factor, current drawn, and stator-rotor slip.

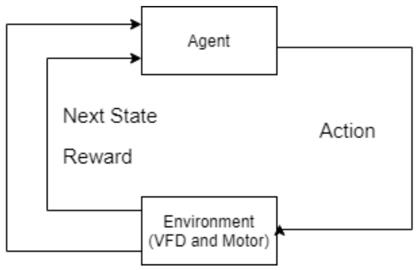

Fig. 4.The standard flow of a Reinforcement Learning setup.

A model of the VFD-driven induction motor was needed for the agent to learn, though the agent itself does not actually know of the internal mechanisms of the model. It is only allowed to observe a given set of environmental state information at each time step, S_t which is pre-determined by the user. In our case, the observable states are the current drawn by

TEST Engineering & Management
January - February 2020
ISSN: 0193 - 4120 Page No. 3074 - 3081

the machine, I_t, the current $a_{v/f}$ value, a pseudo-efficiency metric μ_{psu}, the power factor pf_t, and the current motor slip s_t. These metrics were chosen to because they could be easily gleamed from the standard auxiliary tools accompanying a motor, and they are descriptive enough to embody the necessary information pertaining to the functionality of the particular machine. Fig. 5 outlines the general flow of the test environments that we have used to train the agent. Most of the information pertaining to the setup and activity is defined in the Environment file such as the list of external loads to be exerted on the motor's shaft, the reward function itself, the criteria for program termination, details of the motor, calculation about the current state's information and fetching of subsequent states and so on. The VFD Model file contains a list of formulas that provides an approximate behavior of a VFD-driven induction motor. Finally, we have defined our agent – the Policy Gradient algorithm – and some auxiliary algorithms in the Agent file. The program interacts interdependently in a manner as laid out in Fig. 4.

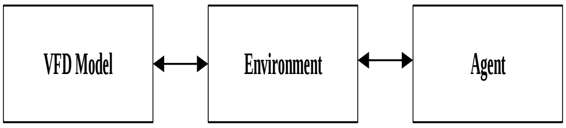

Fig 5. The general flow of our setup.

As for the actual algorithm itself, a vanilla Policy Gradient algorithm was used. The programs were written in Python together with Tensor flow and Sympy. The following subsections cover the aforementioned materials in more detail.

A. VFD and Motor Modelling

As mentioned, the VFD Model file contains a list of equations that enable us to model a VFD-driven induction motor. As it could be very complex to model the system in a forward dynamic manner as in the real world and due to the limitation on complexity, the researchers have taken an inverse dynamic approach to solving for the response of the motor to the various applied loads, and the response of the motor to the new $a_{v/f}$ assigned by the agent in response to the load. This allow us to vastly simplify the modelling of the motor and trims down on the computation time taken to simulate the model. A few of the core equations used are as follow.

Eq. (2) and Eq. (3) for the calculation of stator and motor speeds in RPM

$$N_{stator} = \frac{120 f a_{v/f}}{P} \quad (2)$$

$$N_{motor} = N_{stator}\,(1 - s) \quad (3)$$

Eq. (4), Eq. (5), Eq. (6) defines the torque, maximum slip and power factor respectively. A few variants of these equations were used to determine the torque, slip and $a_{v/f}$ of the motor respectively.

$$T = \frac{3P}{4\pi f a_{v/f}} \cdot \frac{(\alpha v/_f V)^2}{\left[(r_1+\frac{r_2}{s})^2 + (x_1+x_2)^2 (a_{v/f}{}^2)\right]} \cdot \frac{r_2}{s} \quad (4)$$

$$s_{max} = \frac{r_2}{\sqrt{(r_1)^2 + (x_1+x_2)^2 (a_{v/f}{}^2)}} \quad (5)$$

$$pf = \frac{r_1+\frac{r_2}{s}}{\sqrt{(r_1+\frac{r_2}{s})^2 + (x_1+x_2)^2 (a_{v/f}{}^2)}} \quad (6)$$

B. Environment Modelling

The environment is the main backbone of the entire program as it facilitates the flow of information between the model and the agent, in addition to housing the auxiliary functions needed for training the agent such as the reward function, the state calculation and updating functions, logging facilities and so on. Table 2 below outlines some of the main functions defined in the Environment file.

Table 2. Functions Defined in the Environment

Function	Description
Initializer	Initializes the motor and environment parameters.
Reset	Reset the environment at the end of every training episode.
Step	Takes the agents action and returns the reward and fetches the next state.
Reward	Calculates the agent's reward based on the current environment's state and action taken.

For every action taken, the agent returns the 5 observations of the state, the current drawn by the machine, I_t, the current $a_{v/f}$ value, a pseudo-efficiency metric μ_{psu}, the power factor pf_t, and the current motor slip s_t. In return the agent is allowed to agent is allowed to take 10 possible discrete

January - February 2020
ISSN: 0193 - 4120 Page No. 3074 - 3081

actions – setting $a_{v/f}$ to a value in the range of 0.1 to 1.0 in 0.1 steps increment. The series of loads exerted on the shaft varies randomly in terms of amplitude and periodicity within a predefined limit – so as to not exert an over-rated load. Thus, the agent would be trained to react complex spatial-temporal scenarios as opposed to fixed amplitude or periodicity loads only.

C. The Agent

A vanilla Policy Gradient algorithm was used for this particular control task. This algorithm was chosen due to its relative adaptability to a range of complex tasks, which is harder to achieve with Q-learning based method. The fundamental idea underlying Policy Gradient method is to find a policy that would enable the agent to obtain the maximum cumulative discounted reward from the task. A policy is defined as the rule or scheme which dictates the action taken by the agent given a certain state. The maximum cumulative discounted reward is the total sum of reward returned by the environment with respect to the actions taken by the agent, increasingly discounted the further it lies in the future; this means that the agent will learn to prioritise its action to the current state first rather than the future state. Eq. (7) and Eq. (8) are the formal mathematical expression of these concepts.

$$\pi_\theta(a, \text{action}|s, \text{state}) = P[a, \text{action}|s, \text{state}] \quad (7)$$

$$J(\theta) = E_\pi[R_1 + \gamma^1 R_2 + \gamma^2 R_3 + \gamma^2 R_4 + \cdots] \quad (8)$$

Specifically, Eq. (7) defines that given the state, take action under policy π, which outputs a probability distribution across the range of possible actions. One would note that the implication is that the algorithm is non-deterministic in nature as opposed to traditional control schemes, which may not be a desirable characteristic in mission-critical tasks. Equally, Eq. (8) gives the expected cumulative discounted rewards under policy π.

As for the policy itself, a neural network was employed to approximate it. The advantage of such an approach lies in the fact that deep neural nets are a universal function approximate which could be used to parametrise the policy in a back-propagation trainable manner [16]; back - propagation training is the current go-to method in the Artificial Intelligence community, and as such is very well optimised in many software packages including Tensorflow. Fundamentally, it is a method that utilizes a class of gradient-based optimisation method to learn such as such Eq. (9) and Eq. (10) which were used to compute the policy gradient and the corresponding update rule used for said training. Note that α is the learning rate of our agent – a manually tuned parameter – and $R(\tau)$ is our reward function.

$$g = E_\pi[\nabla_\theta(\log \pi(s, a, \theta)) \cdot R(\tau)] (9)$$

$$\Delta\theta = \alpha \cdot \nabla_\theta(\log \pi(s, a, \theta)) \cdot R(\tau) \quad (10)$$

III. RESULTS AND DISCUSSION

The results below are as obtained after training our tuned agent for a few thousand episodes, though better results could be obtained with more training episodes. Fig. 6 shows a sample of the load that was exerted on the motor shaft. Fig. 7 in turns is the mean reward across episodes obtained by our agent during its training.

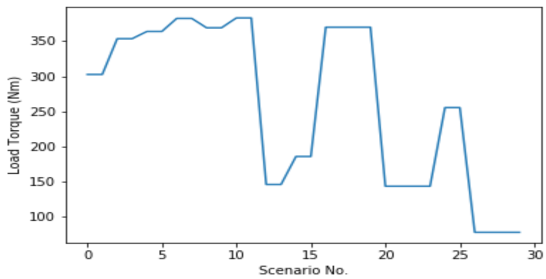

Fig. 6. Sample of load torque on motor shaft for 30 time steps

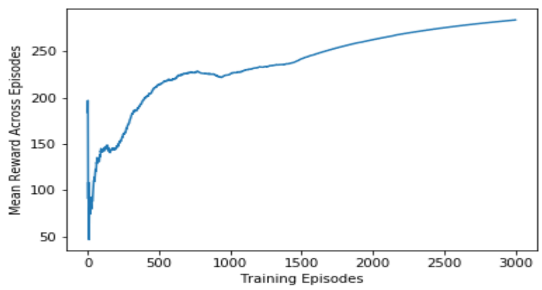

Fig. 7. Mean reward across episodes.

Note that certain load persists for different period of time and the load amplitude differs in non-uniform behaviour. This is to subject the agent to a complex and erratic environment as a form of worst-

January - February 2020
ISSN: 0193 - 4120 Page No. 3074 - 3081

case scenario assessment. In addition, the mean reward across episodes though slowing, is still increasing which indicates that the agent could still further learn given more training steps. Fig. 8 to Fig. 11 as laid out below in turn highlight the performance of our agent against a controller which maintains a constant $a_{v/f}$ at unity as a baseline comparison.

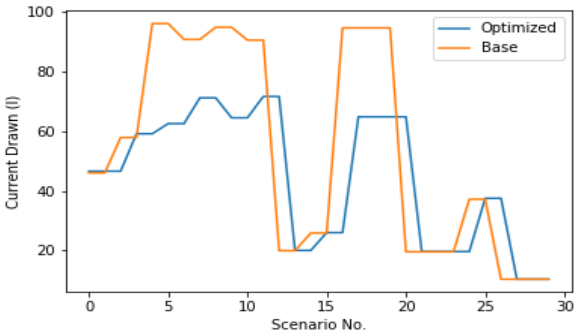

Fig 8. Current drawn by system across varying loads.

As shown Fig. 8 to Fig. 11, the agent outperforms the base controller by a significant margin in both parameters of current drawn and power factor of drive system. Significant improvement could be seen in terms of power factor with improvements as much as 25% for some load scenarios. Note that the slight shift in the plots are merely a consequence of the logger due to the way that the program was written and could be ignored as the artefact is not actually present in the actual simulation time steps itself.

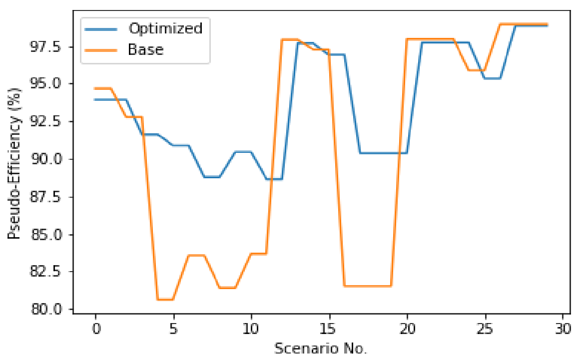

Fig 9. Power factor of system across varying loads.

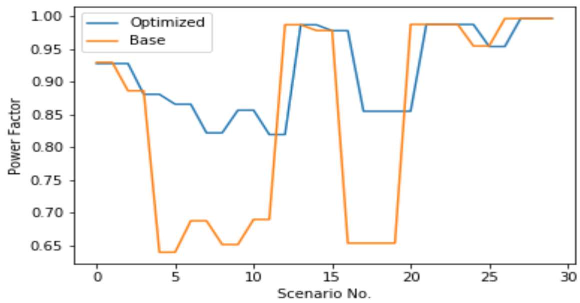

Fig 10. Measure of pseudo-efficiency across varying loads

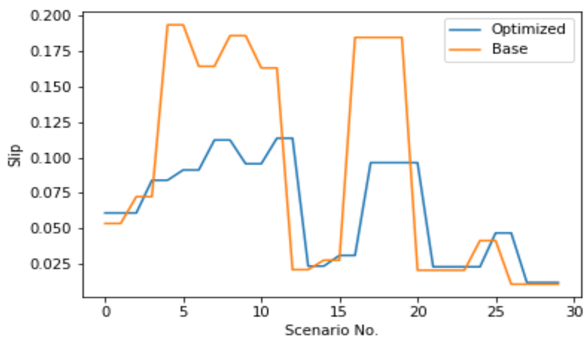

Fig 11. Measure of motor slip across varying loads.

Once again, the plots of the motor pseudo-efficiency and slip indicates that our agent manages to outperform the base controller by a non-trivial margin. Equally significant margins are seen for both of these parameters. The gains are largely proportional, which is to be expected since these parameters are highly correlated in nature.

To the effect, it was seen that the RL based controller manages to learn of a sufficiently optimal solution to drive a VFD-driven induction motor across a sequence of complex load conditions, without any explicit modelling of the environment in the design of said controller; as opposed to traditional control schemes. Application of RL to such systems may prove to be viable in situations where any explicit modelling of a given system is too prohibitive in complexity or otherwise. In addition, learning agents such as these are known to discover non-instinctive and non-obvious optimal solutions that may have been previously unconceived of, making them useful in unconventional solution exploration too [17]. However, it must be stated that there are fundamental disadvantages to RL methods and

January - February 2020
ISSN: 0193 - 4120 Page No. 3074 - 3081

others that utilizes similar black-box heuristic-based approaches. The actions of the agent are not deterministically calculable and understanding neural networks is still a major topic of research. Thus, it may not desirable in industrial applications where the response of the controller under every possible scenario be map able – be known explicitly ahead of time. As an alternative, one could apply these type of control methods and agents to non-critical missions only, where minute chances of error are permissible. In return, one could leverage on the flexibility and the powerful generalisability of these RL agents across range of complex tasks that may prove to be unassailable through conventional means.

IV. CONCLUSION

This research had demonstrated the viability of a novel controller design using an Artificial Intelligence based agent in the task of controlling a VFD-driven induction motor with non-linear parameters optimisation. The Policy Gradient based agent does not need any explicit knowledge of the subject matter or the model of the device itself. It has learned the sufficiently optimal control scheme for the unseen before device and environment through a series of brief user-defined observations and a corresponding feedback signal. The resulting agent surpassed the uncontrolled benchmark by a significant margin in all performance criteria, with significant increase in power factor and slip performance – 25% improvement. However, still a few possible areas of improvements are suggested for future research as follows:

- Train an agent with a wider scope of performance optimisation; e.g. to be maximally stable in both steady-state and transient conditions in extreme load variation environment.
- Provide a framework for the implementation of such black-box methods of control in industrial applications; e.g. by statistical performance analysis to determine the percentage reliability of the controller across a wide range of possible actions.
- Apply more advance state-of-the-art algorithm such as the Deep Deterministic Policy Gradient (DDPG) algorithm. DDPG has a stochastic exploration behaviour, while being deterministic in its target policy estimation.

V. ACKNOWLEDGMENT

Authors would like to acknowledge the work done by graduate students Ang Zhe Han, Lou Wei Je, Lee for conceiving and developing the paper.

This work is partially part of the Taylor's University under its Flagship Research Project grant TUFR/2017/001/01.

VI. REFERENCES

[1] International Energy Agency, "Global Energy & CO2 Status Report 2017," IEA, 2018.
[2] P. Waide and C. U. Brunner, "Energy-Efficiency Policy Opportunities for Electric Motor-Driven Systems," IEA, Paris, 2011.
[3] ABB AG Power Systems Division, "Energy efficiency assessment Improving fan system efficiency using Variable Speed Drives," 2013. [Online]. Available: https://library.e.abb.com/public/deafbb6fff338a8ec1257cc800484065/DEABB%201904%2013%20en%20Variable%20Speed%20Drive%20final.pdf. [Accessed 9 December 2018].
[4] D. Silver and e. al, "Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm," Science, vol. 362, no. 6419, pp. 1140-1144, 2018.
[5] G. C. Archana and N. Reema, "Implementation of V/F Control of Induction Motor using PI and PID controller," International Journal of Industrial Electronics and Electrical Engineering, vol. 4, no. 8, pp. 135-138, 2016.
[6] W. Chen, D. Xu, R. Yang, Y. Yu and Z. Xu, "A novel stator voltage oriented V/F control method capable of high output torque at low speed," 2009 International Conference on Power Electronics and Drive Systems (PEDS), Taipei, 2009, pp. 228-233. doi: 10.1109/PEDS.2009.5385840
[7] A. Munoz-Garcia, T. A. Lipo and D. W. Novotny, "A new induction motor V/f control method capable of high-performance regulation at low speeds," in IEEE Transactions on Industry Applications, vol. 34, no. 4, pp. 813-821, July-Aug,1998. doi: 10.1109/28.703982

January - February 2020
ISSN: 0193 - 4120 Page No. 3074 - 3081

[8] Y. Duan and e. al, "Benchmarking Deep Reinforcement Learning for Continuous Control," in Proceedings of the 33rd International Conference on Machine Learning, New York, 2016.

[9] Krishnavedala, "Schematic drawing of an inverted perdulum on a cart," 2012.

[10] H. Jaafar and Z. Mohamed, "PSO-Tuned PID Controller for a Nonlinear Double-Pendulum Crane System," Communications in Computer and Information Science, vol. 752, pp. 203-215, 2017.

[11] V. Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra and Martin Riedmiller, "Playing Atari with Deep Reinforcement Learning," in Conference on Neural Information Processing Systems, California, 2013.

[12] Volodymyr Mnih Koray Kavukcuoglu, David Silver,Andrei A. Rusu,Joel Veness,Marc G. Bellemare et.al "Human-level Control through Deep Reinforcement Learning," Nature, vol. 518, pp. 529–533, 2015.

[13] I. Grondman, L. Busoniu, G. A. D. Lopes and R. Babuska, "A Survey of Actor-Critic Reinforcement Learning: Standard and Natural Policy Gradients," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 6, pp. 1291-1307, Nov. 2012.

[14] Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, Daan Wierstra, "Continuous control with deep reinforcement learning," Sep 2015, [Online]. [Available: https://arxiv.org/abs/1509.02971] [Accessed 9 Dec 2018].

[15] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, Oleg Klimov, "Proximal Policy Optimization Algorithms," 20 July 2017. [Online]. Available: https://arxiv.org/pdf/1707.06347.pdf. [Accessed 9 Dec 2018].

[16] A. N. Michael, Neural Networks and Deep Learning, Determination Press, 2015. [Available: http://static.latexstudio.net/article/2018/0912/neuralnetworksanddeeplearning.pdf]

[17] C. Metz, "How Google's AI Viewed the Move No Human Could Understand," Wired, 2016. [Available: https://cacm.acm.org/news/199693-how-googles-ai-viewed-the-move-no-human-could-understand/fulltext?mobile=false]

AUTHORS PROFILE

Dr. Reynato Andal Gamboa.This author is a Senior Lecturer at the School of Engineering, Taylor's University, Malaysia. His educational details include Bachelor of Science in Electrical Engineering (BSEE, Philippines), Master of Science in Electrical Engineering (MSEE, Philippines) major in Power Engineering and Doctor of Philosophy (Engineering, Malaysia). He is a registered Chartered Engineer (CEng) from the Institution of Engineering and Technology (IET, UK) and a registered Electrical Engineer (REE) in the Philippines. Organizational memberships include Institute of Integrated Electrical Engineers (IIEE) from Philippines, and a member of the Institution of Engineering and Technology (MIET,UK). He won several best paper awards in international conferences and medal awards in academic competitions for teaching and learning. He also have several publications in Electrical and Electronic Engineering (Power Engineering) and Engineering Education (Outcome-based Education assessments). His area of interests are on Power Engineering and Engineering Education.

Aravind CV is currently working as Associate Professor with the Faculty of Innovation and Technology, Taylor's University, Malaysia since 2011. He holds a doctoral degree in the field of electrical machines. His research interests includes sustainable energy and its applications for future living. s.

Gowthamraj rajendran. This author is a research scholar at the school of engineering, taylor's university, malaysia. His educational details include bachelor of engineering in electrical and electronics engineering (be-eee, india), master of engineering (me-eee, india), major in power electronics and drives. His area of research is power converters and drives. His area of interest are on power converters and controllers..