Reinforcement Learning Based Adaptive Energy Management on Diverse Applications for Embedded System

Cheng-Ting Liu, Roy Chaoming Hsu[*]
Dept. of Computer Science and Information Engineering
National Chiayi University
Chiayi, Taiwan
e-mail: rchsu@mail.ncyu.edu.tw[*]

Abstract—**Use of adaptive energy management strategies is essential in improving energy utilization, efficiency, and sustainable operation of embedded systems. Accordingly, this paper presents embedded system applications, which are suitable for employing adaptive energy management utilized reinforcement learning. We proposed rewarding function for embedded system applications in stimulating the learning agent to select the best strategy by learning from environment the agent situated. The proposed adaptive energy management applications are for battery-aware embedded systems of energy harvest wireless sensor network and human-electric hybrid bicycle. Future work of the extended study will focus on applying the reinforcement learning based adaptive energy management for hybrid electric vehicle, and designing a unified rewarding function for diverse applications.**

Keywords- reinforcement learning; adaptive energy manegement; embedded system; wirless sensor network; human-electric hybrid bicycle; hybrid electric vehicle

I. INTRODUCTION

Study of reinforcement learning (RL) is originally appeared in 1994 handbook of Brain Theory and Neural Networks by Barto [1]. In RL, as shown in Fig. 1, a decision-making agent takes actions in the environment the agent situated and receives reward for its actions in trying to solve a problem. After a set of trial-and-error runs, the agent gradually learns the best policy, which is the sequence of actions that maximize the total reward. Recently, Singh et al. pursued the RL study using an evolutionary perspective [2], which addressed that the intrinsic motivation of RL plays a key role in the human's intellectual growth and mental development. The framework of RL has been widely applied to many applications such as control problem, optimize problem, resource management problem, etc.
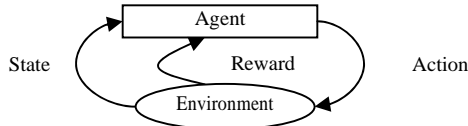

Fig. 1. Agent-environment interactions in reinforcement learning

For unknown environment, the agent would repeatedly count the reward during a serial of trials and the optimized policy would be obtained after a number of trials. In the beginning of the learning process, the agent is unable to make a for sure decision, the exploration strategy is, hence, required for retrieving the sufficient reward information on each state, while the exploitation is then adopted in deciding the actions with higher reward and higher probability on certain state in the later part of learning process. In such a case, the softmax function is utilized for the strategy of exploration-exploitation in most applications. In the case of nondeterministic environment or state transition model, where the transition probability is unknown or unsure, a certain action for the next state can not be precisely decided by observing the current state. However, the Q-learning algorithm can be utilized in recording the accumulative reward and supporting exploration-exploitation strategy to decide the near best policy or action.

In Q-learning, the accumulative reward is function of state, s, and action, a, and the learner iteratively updates the state-action value with the updating equation shown below,

$$Q(s_t,a_t)=(1-\eta)Q(s_t,a_t)+\eta[r_{t+1}+ \gamma \max_{\forall at+1 \in A}Q(s_{t+1},a_{t+1})] \quad (1)$$

where $Q(s_t, a_t)$ is the accumulative reward standing at state s and take an action a at step t, and the parameters η and γ, are the learning rate and discount rate, respectively , with value between 0 and 1. r_{t+1} is a reward value which is obtained by taking action a_t and then making transition from state s_t to state s_{t+1}. Normally the state-action values are stored in a table, named Q-table, as the reference for learner to take the next action. In this paper, how RL is applied as a mean of adaptive energy management for embedded system is discussed for energy harvesting wireless sensor network (WSN) and human-electric hybrid bicycle as below.

II. CASE STUDY

A. Manitaining Energy Neutrility for Energy Harvest Wireless Sensor Network

Study of the dynamic power management for energy harvesting sensor node has been proposed by [3]. We approach this problem by abstract architecture is organized by two parts as shown in Fig. 2. A reinforcement learning based dynamic power management for energy harvesting wireless sensor network is proposed by Hsu et al. [4], which was our previous work. The agent of the reinforcement learning is employed in the main management unit of Fig. 2 to learn from the environment information of e_{node}, $e_{harvest}$, and e_b, and to adaptively decide and execute the action

characterized by the desired operational duty-cycle. During the learning, the agent is encouraged to select the action with positive reward and a series of beneficial actions will be iteratively generated such that a better energy management performance is gradually achieved.
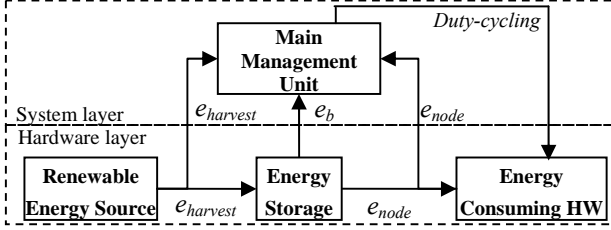

Fig. 2. System architecture for the sensor node

State space of the environment is defined as a set of state. The domination state is the state of "energy neutrality", and this term is first defined by Kansal et al. [5]. The state of energy neutrality (S_D)is defined as the difference between $e_{harvest}$ and e_{node} at the end of the same sensing period.

The action space is defined on the controllable variable of duty-cycle. Hence, an action with higher value means a higher operational duty cycle is assigned to the sensor node; where more energy would be consumed consequently.

In reinforcement learning, the reward value is utilized to evaluate the performance of decided action. The distance of energy neutrality is a good candidate of reward by measuring the degree it close to zero. The harvesting energy is stored in the energy storage such that the remaining energy storage must be taking into consider in defining the immediate reward. By considering the effect of the remained energy in the storage, the equation of immediate reward r'_D can be derived by (2)

$$r_D = -|S_D| / (e_{maxharvest} - e_{minharvest}) \qquad (2)$$
$$r'_D = -r_D (1 - 2 e_b / E_B)$$

The positive distance of energy neutrality would induce the positive reward, while the storage remained energy stands on higher level. However, the positive distance would induce the negative reward, while the remained energy stands on lower level.

B. Comfort of Riding for Human-electric Hybrid Bicycle

A reinforcement learning based power assisted method for pedelec, a human-electric hybrid bicycle, is proposed by Hsu et al. [6] with framework shown in Fig. 3. The RL agent can adaptively select and execute the action by the desired operational duty-cycle for the motor. The duty-cycle is defined as the operational percentage of the motor. By adaptively adjusting the assisted power after learning for pedelec, not only energy utilization is improved but comfort of riding for pedalist is also satisfied. Within the comfortable zone, the closer the pedelec's velocity is to the predetermined comfortable velocity and the larger the reward will be given. The main purpose of the power

assisted method is to satisfy the comfort of ride to pedelec's rider, such that a comfortable velocity, v_{comf}, on riding, and the tolerance of deviation, μ, to the pre-determined comfortable velocity should be first defined and arranged into the rewarding function as the following,

$$r = \begin{cases} \dfrac{k_1 \mu}{|v - v_{comf}|} & , if \, |v - v_{comf}| \le \mu \\ -k_2 |v - v_{comf}| & , otherwise \end{cases} \qquad (3)$$

where v is the current pedelec's velocity, and k_1 and k_2 are positive weighting factors.
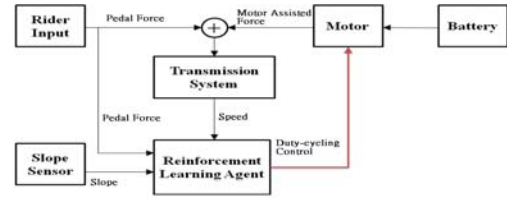

Fig. 3. System architecture of power assisted pedelec

III. CONSLUSION AND FUTURE WORK

In this paper, the reinforcement learning based adaptive energy management for energy harvest WSN and human-electric hybrid bicycle is discussed. Experiment results of above cases have been obtained in [4] and [6], respectively, which shown that RL is an effective mean in improving energy utilization, efficiency, and sustainable operation of embedded systems. The reinforcement learning approach can possibly be extended to hybrid electric vehicle by modifying the definition of environment and rewarding function. How to derive a unified rewarding function to fit various adaptive energy management for embedded system is another challenge and will be our future work.

REFERENCES

[1] A. G. Barto, *Reinforcement Learning*, in handbook of Brain Theory and Neural Networks, M.A. Arbib (Ed.), Cambridge: MIT Press, 1994.

[2] S. Singh, R. L. Lewis, A. G. Barto, and J. Sorg, "Intrinsically Motivated Reinforcement Learning: An Evolutionary Perspective," *IEEE Trans. on Autonomous Mental Development,* This article has been accepted for publication in a future issue of this journal, 2010

[3] C. Moser, L. Thiele, D. Brunelli, and L. Benini, "Adaptive Power Management in Energy Harvesting Systems," *in Proc. of Design, Automation & Test in Europe Conference & Exhibition*, pp. 1–6, 2007.

[4] R. C. Hsu, C. T. Liu, and W. M. Lee, "Reinforcement Learning-Based Dynamic Power Management for Energy Harvesting Wireless Sensor Network," IEA/AIE 2009, LNAI 5579, pp. 399–408, 2009.

[5] A. Kansal, J. Hsu, M. Srivastava, V. Raghunathan, "Harvesting Aware Power Management for Sensor Networks," *in Proc. of ACM/IEEE Design Automation Conference*, pp. 651–656, 2006.

[6] R. C. Hsu, C. T. Liu, W. M. Lee, and C. H. Chen, "A Reinforcement Learning Based Power Assisted Method with Comfort of Riding for Light Electric Vehicle," IEEE Proc. Of Inte. Conf. Vehicular Technology, pp. 1-5, 2010.