

## การจัดตารางโหลดที่เหมาะสมที่สุดด้วยการเรียนรู้การเสริมแรงเชิงลึกเพื่อประสิทธิภาพการใช้พลังงานในฟาร์มหมู อัจฉริยะในจังหวัดพัทลุง

### Deep Reinforcement Learning-Based Optimal Load Scheduling for Energy Efficiency in Smart Pig Farms in Phatthalung Province

วิวรรณ์ สุหรินทร์<sup>1,2</sup>, ศุภชัย แก้วพวง<sup>2</sup>, ธวัช ชูชิต<sup>2</sup>, ธีรยุทธ ทองทวี<sup>3</sup> และ ทศพล รัตนนิมิตชัย<sup>1\*</sup>

<sup>1</sup>สาขาวิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีสุรนารี jamewiwat2542@gmail.com

<sup>2</sup>ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ มหาวิทยาลัยทักษิณ tawat.c@tsu.ac.th, supachai.ka@tsu.ac.th

<sup>3</sup>สาขาวิชาวิศวกรรมไฟฟ้าและปัญญาประดิษฐ์ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏนครศรีธรรมราช

<sup>1</sup>สาขาวิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีสุรนารี tosaphol@sut.ac.th\*

#### บทคัดย่อ

บทความนี้เสนอการจัดการจัดสรรโหลดไฟฟ้า (Load Management) อย่างมีประสิทธิภาพ ในฟาร์มหมู เนื่องจากการใช้ระบบอากาศ ระบบให้อาหาร อัตโนมัติ และแม้กระทั่งโหลดแสงสว่าง โดยอาศัยการเรียนรู้เชิงลึกแบบเสริมกำลัง (Deep Reinforcement Learning: DRL) เพื่อหาแนวทางการใช้พลังงานที่เหมาะสมที่สุดภายใต้โครงสร้างอัตราค่าไฟฟ้าแบบตามช่วงเวลา (Time-of-Use: TOU) ของการไฟฟ้าส่วนภูมิภาค (กฟภ.) โดยนำเสนอวิธีการจากการเรียนรู้เชิงลึกแบบเสริมกำลัง ซึ่งประกอบด้วยเทคนิค Deep Q-Network (DQN) และ Proximal Policy Optimization (PPO) ผลการทดลองชี้ให้เห็นว่า DRL สามารถปรับเปลี่ยนเวลาการทำงานของอุปกรณ์ในฟาร์มหมูได้ตรงตามเงื่อนไข ลดการใช้พลังงานในช่วงเวลาที่มีค่าไฟสูง และเพิ่มการใช้พลังงานในช่วงที่มีค่าไฟต่ำ ส่งผลให้ค่าไฟฟ้าเฉลี่ยลดลงถึง 69.43% และ 75.96% ต่อเดือนด้วยเทคนิค DQN และ PPO ตามลำดับ งานวิจัยนี้จึงแสดงให้เห็นถึงศักยภาพของ DRL ร่วมกับระบบ Internet of Things (IoT) ในการออกแบบฟาร์มหมูอัจฉริยะ (Smart Pig Farm) ที่ส่งเสริมการใช้พลังงานอย่างมีประสิทธิภาพและยั่งยืน

**คำสำคัญ:** การเรียนรู้เชิงลึกแบบเสริมกำลัง, การจัดตารางโหลดที่เหมาะสมที่สุด, ฟาร์มอัจฉริยะ

#### Abstract

This paper presents an Optimal load-scheduling for pig smart farms energy management, addressing the operation of ventilation systems, automated feeding systems, and lighting loads. The proposed framework leverages Deep Reinforcement Learning (DRL) to optimize energy consumption under the Time-of-Use (TOU) tariff scheme of the Provincial Electricity Authority (PEA). Two DRL techniques, namely Deep Q-Network (DQN) and Proximal Policy Optimization (PPO), are employed to derive adaptive energy scheduling policies. Experimental results

demonstrate that DRL effectively shifts the operational schedules of farm equipment by reducing energy usage during peak-price periods and increasing utilization during off-peak hours. As a result, the average monthly electricity cost is reduced by up to 69.43% with DQN and 75.96% with PPO. These findings highlight the potential of DRL, in conjunction with the Internet of Things (IoT), for enabling Smart Pig Farms that enhance energy efficiency and sustainability.

**Keywords:** Deep Reinforcement Learning, Optimal Load Scheduling, Smart Pig Farm

#### 1. บทนำ

อุตสาหกรรมการเลี้ยงหมูเป็นภาคส่วนที่มีการใช้พลังงานไฟฟ้าสูง โดยเฉพาะในระบบระบายอากาศ ระบบให้อาหาร รวมถึงระบบแสงสว่างภายในฟาร์ม เนื่องจากความต้องการพลังงานมีความผันผวนตามสภาพอากาศ เวลา และพฤติกรรมของสัตว์ การใช้พลังงานโดยขาดการจัดการอย่างเป็นระบบ ทำให้เกิดปัญหาการใช้ไฟฟ้าเกินช่วงเวลาที่มียอดราคาไฟฟ้าสูงสุด (Peak Time) ทำให้การจัดการโหลด (Load Management) อย่างมีประสิทธิภาพเป็นสิ่งสำคัญเพื่อลดต้นทุนด้านพลังงาน และเพิ่มความยั่งยืนในการผลิต

เพื่อแก้ไขปัญหาดังกล่าว จึงได้มีการพัฒนาระบบ การจัดการพลังงานในฟาร์มหมูอัจฉริยะ โดยประยุกต์แนวคิดจาก Smart Home Energy Management (SHEM) [1-2] ซึ่งมุ่งเน้นการควบคุมและจัดตารางการใช้งานโหลดไฟฟ้าให้มีประสิทธิภาพสูงสุด โดยจะอาศัยวิธีการ Load Shifting การเลื่อนการใช้งานโหลดไฟฟ้าบางส่วนจากช่วง Peak ไปยังช่วง Off-peak ซึ่งอัตราค่าไฟฟ้าลดลงเกินกว่าครึ่งของช่วง Peak โดยอาศัยวิธีการเรียนรู้เครื่อง (Machine Learning) โดยเฉพาะการเรียนรู้เชิงลึกแบบเสริมกำลัง (Deep Reinforcement Learning) มาประยุกต์ใช้

การเรียนรู้เชิงลึกแบบเสริมกำลังเป็นการเรียนรู้ที่อาศัยการทดลองและข้อผิดพลาด (Trial-and-Error) โดยมีตัวแทน (Agent) ทำการตัดสินใจ

เลือกการกระทำ (Action) จากสถานะแวดล้อม (State) และได้รับผลตอบแทน (Reward) เพื่อนำไปปรับปรุงแนวทางการตัดสินใจ (Policy) ให้เหมาะสมที่สุด งานวิจัยด้านการจัดการพลังงานได้พิสูจน์แล้วว่า DRL สามารถให้ผลลัพธ์ที่มีประสิทธิภาพสูงกว่าวิธีแบบดั้งเดิม โดยเทคนิคที่นิยม ได้แก่ Deep Q-Network (DQN) และ Proximal Policy Optimization (PPO) [3-4]

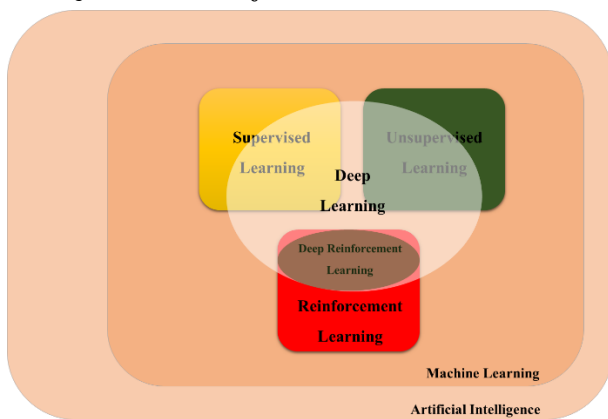
งานวิจัยนี้จึงมุ่งเน้นการพัฒนาระบบ การจัดโหลดพลังงานในฟาร์ม หมู้อจรรย์ด้วย DRL เพื่อหา นโยบายการใช้พลังงานที่เหมาะสมที่สุด ลด ค่าไฟฟ้าในอัตรา TOU ทำให้ลดต้นทุน และสร้างความยั่งยืนให้แก่ อุตสาหกรรมการเลี้ยงหมูในอนาคต

## 2. ทฤษฎีที่เกี่ยวข้อง

### 2.1 การเรียนรู้แบบเสริมกำลัง

การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning : RL) เป็น สาขาหนึ่งของการเรียนรู้ของเครื่อง (Machine Learning) ดังรูปที่ 1 โดยมี จุดประสงค์ในการพัฒนาตัวแทน (Agent) ให้สามารถตัดสินใจทำงานและ ตอบโต้กับสภาพแวดล้อม (Environment) โดยอาศัยหลักการทดลองซ้ำ และเรียนรู้จากผลลัพธ์ (trial-and-error) และรับรางวัล (Reward) หรือ โทษ (Penalty) จากการกระทำ โดยพื้นฐานของการเรียนรู้แบบเสริมกำลัง มาจากกระบวนการตัดสินใจแบบมาร์คอฟ (Markov Decision Process : MDP) [5]

ในปัจจุบันการเรียนรู้แบบเสริมกำลังทำให้มีความซับซ้อนมากขึ้น จากการพัฒนาอย่างรวดเร็วของเทคโนโลยีการเรียนรู้เชิงลึก (Deep Learning) เพื่อใช้ในการแก้ปัญหาที่มีความซับซ้อน และแม่นยำมากขึ้น



รูปที่ 1 ลำดับความเป็นมาของอัลกอริทึม

#### 2.1.1 กระบวนการตัดสินใจแบบมาร์คอฟ

กระบวนการตัดสินใจแบบมาร์คอฟ คือ แบบจำลองทางคณิตศาสตร์ ที่ใช้ในแก้ปัญหาสำหรับการเรียนรู้แบบเสริมกำลัง ประกอบด้วย 5 ส่วน สำคัญดังนี้

1. สถานะ (State space:  $S_t$ ) เป็นชุดของสถานะทั้งหมดที่ระบบ สามารถตอบโต้ได้ในสภาพแวดล้อม และมีการส่งค่าพารามิเตอร์ต่อไป เพื่อ วิเคราะห์การกระทำในสถานะต่อไป

2. การกระทำ (Action:  $a_t$ ) ชุดข้อมูลของตัวแทนที่สามารถเลือก การกระทำต่อสภาพแวดล้อมในแต่ละสถานะนั้น ๆ

3. ฟังก์ชันการเปลี่ยนสถานะ (State transition probability:  $P$ ) ฟังก์ชันความน่าจะเป็นที่บอกว่าเมื่ออยู่ในสถานะ  $S_t$  และทำการกระทำ  $a$  จะไปยังสถานะ  $S_{t+1}$  ได้ด้วยความน่าจะเป็นเท่าไร ดังสมการที่ 1

$$P(S_{t+1}|S_t, a_t) \quad (1)$$

4. ฟังก์ชันรางวัล (Reward function:  $R$ ) ฟังก์ชันที่บอกว่าหลังจาก ทำการกระทำ  $a_t$  ในสถานะ  $S_t$  และได้สถานะ  $S_{t+1}$  มีค่ารางวัลเท่าใด ในสถานะนั้น ๆ สามารถหาได้ ดังสมการที่ 2

$$R(S_t, a_t, S_{t+1}) \quad (2)$$

5. ค่าลดทอน (Discount Factor:  $\gamma$ ) เป็นค่าที่กำหนดคุณค่าของ รางวัล เพื่อให้เหมาะสมสำหรับการเรียนรู้ โดยจะมีค่าเป็นจำนวนจริง ระหว่าง 0 ถึง 1 หากค่าลดทอนไม่อยู่ในช่วงดังกล่าว โดยถ้ามีค่ามากกว่า 1 ส่งผลให้คุณค่าของรางวัลในรอบถัดไปมีค่าเกินจริง และหากมีค่าต่ำกว่า 0 ส่งผลให้คุณค่าของรางวัลที่ได้ในรอบถัดไปทำให้สลับสัญญาณ(บวก,ลบ)

### 2.2 DEEP Q-NETWORK

Deep Q-Network (DQN) เป็นอัลกอริทึมการเรียนรู้แบบเสริม กำลังที่ได้รับการพัฒนาต่อยอดมาจากวิธีการ Q-Learning โดยมีจุดเด่น คือ การนำเครือข่ายประสาทเทียมแบบลึก (Deep Neural Network) มา ใช้สำหรับประมาณค่าฟังก์ชัน Q (Q-function) ซึ่งมีการประเมินค่ารางวัล ที่ได้จากการกระทำในสถานะที่กำหนด ทำให้สามารถเพิ่มมิติและความ ซับซ้อนของสภาพแวดล้อมได้

นอกจากนี้ DQN ยังบูรณาการกลไกสำคัญสองประการ ได้แก่ Experience Replay และ Target Network เพื่อเพิ่มเสถียรภาพในการ เรียนรู้ โดย Experience Replay ทำหน้าที่เก็บประสบการณ์การ ปฏิสัมพันธ์ระหว่างตัวแทน (Agent) กับสภาพแวดล้อมในรูปของ  $(s_t, a_t, r_t, s_{t+1})$  และสุ่มตัวอย่างจากหน่วยความจำดังกล่าวมาใช้ในการฝึกฝน เพื่อลด ความสัมพันธ์ของข้อมูลที่เกิดขึ้นตามลำดับเวลา ส่วน Target Network เป็นเครือข่ายที่ถูกอัปเดตค่าพารามิเตอร์เป็นระยะเพื่อช่วยป้องกันความไม่ เสถียรจากการเปลี่ยนแปลงค่าที่คาดการณ์ในแต่ละรอบการเรียนรู้ ด้วย กลไกดังกล่าว DQN จึงสามารถประยุกต์ใช้ได้อย่างมีประสิทธิภาพใน สภาพแวดล้อมที่มีความซับซ้อนสูง [1][4] ซึ่งมีรหัสจำลองของ DQN อัลกอริทึม ดังรูปที่ 2 ในการใช้งานการเรียนรู้แบบเสริมกำลังด้วย DQN เหมาะสมกับ Action Space แบบ discrete และขนาดไม่ใหญ่มาก เช่น เกม Atari, CartPole

### 2.3 PROXIMAL POLICY OPTIMIZATION

Proximal Policy Optimization (PPO) เป็นอัลกอริทึมการเรียนรู้ แบบเสริมแรงขั้นสูงที่พัฒนามาเพื่อแก้ไขข้อจำกัดด้านความซับซ้อนและ ความไม่เสถียรของวิธีการเดิม เช่น Policy Gradient และ Trust Region

Policy Optimization (TRPO) โดย PPO มีจุดเด่นคือการปรับปรุงนโยบาย (Policy Improvement) อย่างค่อยเป็นค่อยไป โดยใช้ฟังก์ชัน Clipped Objective Function เพื่อจำกัดไม่ให้อัปเดตนโยบายในแต่ละครั้งแตกต่างจากนโยบายเดิมมากเกินไป [1][4] รหัสจำลองของ PPO อัลกอริทึม ดังรูปที่ 3 ในการใช้งานการเรียนรู้แบบเสริมกำลังด้วย PPO เหมาะสมกับ Action Space แบบต่อเนื่องและ งานควบคุมซับซ้อน เช่น หุ่นยนต์, ระบบพลังงาน

#### Algorithm 1: Deep Q Learning Algorithm

```
Initialize replay memory D with capacity N
Initialize action-value function Q with random weights
for episode = 1 to M do
    Initialize sequence  $s_1 = \{x_1\}$  Stack initial frame k times
    for t = 1 to T do
        With probability  $\epsilon$  select a random action  $a_t$  Otherwise select action:
         $a_t = \max_a Q^*(s_t, a|\theta)$ 
        Execute action  $a_t$  in the environment. Observe reward  $r_t$  and next frame  $x_{t+1}$ 
        Construct next state  $s_{t+1}$  by stacking  $(x_{t-k+1}, \dots, x_t, x_{t+1})$ 
        Store transition  $s_t, a_t, r_t, s_{t+1}$  in D
        Sample random minibatch of transitions  $(s_j, a_j, r_j, s_{j+1})$  from D
        Set target  $y_j = \begin{cases} r_j & \text{for terminal } s_{j+1} \\ \gamma \max_{a'} Q(s_{j+1}, a'|\theta) & \text{for non-terminal } s_{j+1} \end{cases}$ 
        Perform a gradient descent step on  $(y_j - Q(s_j, a_j|\theta))^2$  according to equation
        
$$\nabla_{\theta} J(s) = N^{-1} \sum \nabla_{\theta} Q_{\theta}(s_j, a_j) |_{a=\pi(s_j)} \nabla_{\theta} \pi_{\theta}(s_j)$$

    end for
end for
```

รูปที่ 2 รหัสจำลองของ DQN อัลกอริทึม

#### Algorithm 2: PPO Algorithm

```
Randomly initialize policy  $\pi$  and value function  $V$  with parameters  $\theta, \nu$  respectively
Create  $N_e$  environments
n = number of steps to calculate TD over
 $\theta \leftarrow \theta$ 
repeat
    for actor 1..... $N_e$  do
        for time t.....t+n do
            Sample action at  $\sim \pi(a|s_t, \theta)$ 
            Execute actions  $a_t$  and observe rewards  $r_t$  and next states  $s_{t+1}$ 
        end for
    end for
    Compute Advantages A and Returns R
    for epoch 1.....K do
        for minibatch 1.....N do
            sample minibatch transitions randomly without replacement  $(s_j, a_j, A_j, R_j)$ 
            
$$r(\theta) = \frac{\pi(a_j|s_j, \theta)}{\pi(a_j|s_j, \theta_{old})}$$

            
$$\nabla_{\theta} L(\theta) = E[\nabla_{\theta} \min(r(\theta) A_j, \text{clip}(r(\theta), 1 - \epsilon, 1 + \epsilon) A_j) + \beta_H \nabla_{\theta} H(\pi(s, \theta))]$$

            
$$\nabla_{\theta} L(\theta_{\nu}) = E_{\pi}[\frac{\partial (R^{\pi} - V(s|\theta_{\nu}))}{\partial \theta_{\nu}}]$$

        end for
    end for
     $\theta_{old} \leftarrow \theta$ 
until  $t_{max}$ 
```

รูปที่ 3 รหัสจำลองของ PPO อัลกอริทึม

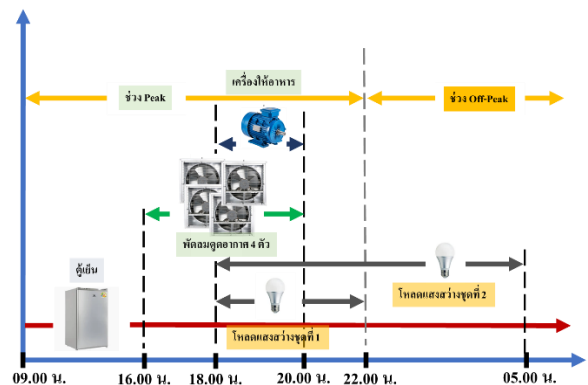
### 3. ระเบียบวิธีการ

#### 3.1 การจัดการพลังงานไฟฟ้าในฟาร์มหมู

ในการจัดการพลังงานไฟฟ้าในฟาร์มหมู สามารถจัดชนิดของโหลดได้ 2 รูปแบบ 1. โหลดปรับเปลี่ยนได้ คือ โหลดที่สามารถปรับเปลี่ยนได้ตามเงื่อนไข เช่น พัดลมดูดอากาศ เครื่องให้อาหาร 2. โหลดคงที่ คือ โหลดที่ไม่สามารถปรับเปลี่ยนได้และอาจขึ้นกับความต้องการของผู้ใช้งาน เช่น แสงสว่าง ตู้เย็น ดังตารางที่ 1 และการใช้งานโหลดไฟฟ้าในฟาร์มหมูก่อนการจัดการโหลดด้วย DRL ดังรูปที่ 4 นอกจากนี้ในฟาร์มหมูมีการใช้พลังงานไฟฟ้าจากการไฟฟ้าและสามารถผลิตเองได้จากโซลาร์เซลล์ ดังรูปที่ 5 และมีค่าไฟย้อนหลังภายในฟาร์มหมู 31 เดือน ดังรูปที่ 6 โดยในการควบคุมอุปกรณ์ต่าง ๆ ในฟาร์มหมู จำเป็นต้องยกระดับให้เป็นฟาร์มอัจฉริยะด้วยการใช้ระบบ IoT ร่วมด้วย โดยมีรูปแบบการเชื่อมต่อ DRL และอุปกรณ์ IoT ดังรูปที่ 7 และรูปแบบการทำงานของระบบที่นำเสนอ ดังรูปที่ 8 โดยมี การเก็บข้อมูลจากอุปกรณ์ IoT เพื่อนำไปใช้สำหรับการฝึกสอนโมเดล DRL จากนั้นให้โมเดลตัดสินใจในการควบคุมอุปกรณ์ไฟฟ้า และบันทึกผลเพื่อเปรียบเทียบความคุ้มค่าในการลดต้นทุน

ตารางที่ 1 ประเภทของโหลด

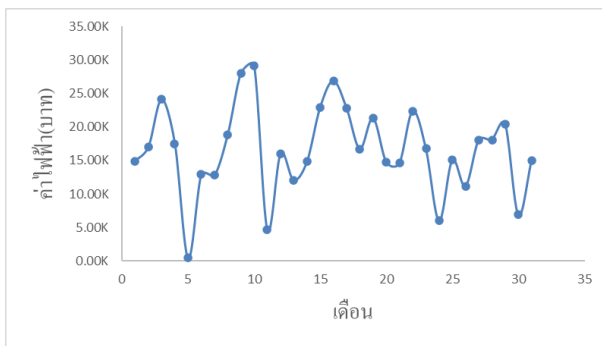
| ลำดับ | รายการอุปกรณ์ไฟฟ้า |                               |                    |       |
|-------|--------------------|-------------------------------|--------------------|-------|
|       | อุปกรณ์            | กำลังไฟฟ้า<br>(Kilo Watt: kW) | ชนิดของอุปกรณ์     | จำนวน |
| 1     | พัดลมดูดอากาศ      | 0.75                          | โหลดปรับเปลี่ยนได้ | 6     |
| 2     | เครื่องให้อาหาร    | 3                             | โหลดปรับเปลี่ยนได้ | 1     |
| 3     | โหลดแสงสว่าง       | 0.1                           | โหลดคงที่          | 2     |
| 4     | ตู้เย็น            | 0.3                           | โหลดคงที่          | 1     |



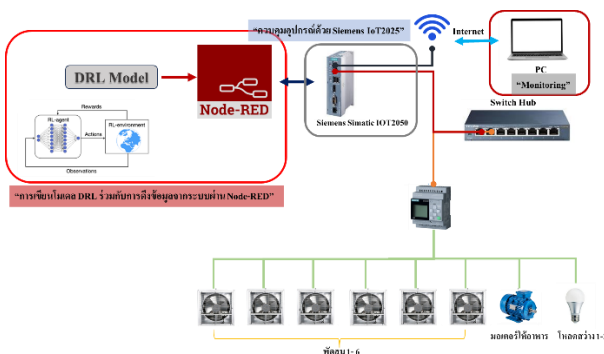
รูปที่ 4 การใช้งานอุปกรณ์ไฟฟ้าในฟาร์มหมู



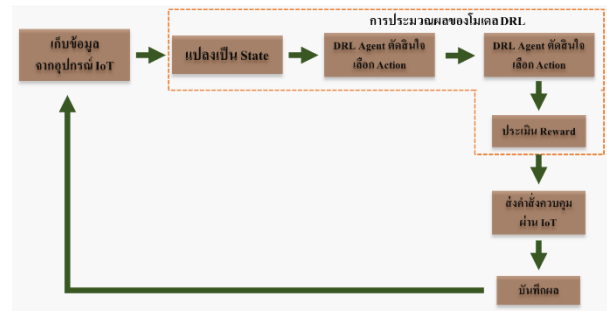
รูปที่ 5 รูปแบบการใช้พลังงานของฟาร์มหมู



รูปที่ 6 กราฟค่าไฟฟ้า 31 เดือนย้อนหลัง



รูปที่ 7 การใช้ระบบ IoT ร่วมกับ DRL



รูปที่ 8 ผังงานการทำงานของระบบ

ในการจัดการโหลดให้เหมาะสมกับผู้ใช้และการจัดการฟาร์มหมูอย่างมีประสิทธิภาพ ต้องให้ความสำคัญที่โหลดที่สามารถปรับเปลี่ยนได้เนื่องจากยังสามารถปรับเปลี่ยนได้ตามเงื่อนไขในแต่ละวัน โดยในฟาร์มหมูตัวอย่าง มีการติดตั้งโหลดแสงสว่างที่ทำงานในช่วง 18.00 – 6.00 น. เครื่องให้อาหารหมูอัตโนมัติที่ทำงาน 1 ครั้งใน 1 วัน ช่วงเวลาใดก็ได้ และโหลดพัดลม 6 ตัว ที่ต้องทำงานอย่างน้อย 4 ตัว ภายใน 1 วัน เพื่อระบายอากาศในฟาร์มหมู และในการจัดการโหลดที่สามารถปรับได้จะมีการใช้ข้อมูลจากการไฟฟ้า ซึ่งแบ่งการใช้ไฟฟ้า 3 ช่วง คือ สำหรับวันจันทร์ - วันศุกร์เวลา 9.00 – 22.00 น. จัดเป็นช่วง Peak เวลา 22.00 – 9.00 น. จัดเป็นช่วง Off- Peak และสำหรับวันเสาร์ - วันอาทิตย์ ช่วงเวลา 0.00 – 24.00 น. จัดเป็นช่วง Off- Peak ดังตารางที่ 2 [6]

ตารางที่ 2 อัตราค่าไฟฟ้าจากการไฟฟ้าส่วนภูมิภาค ประเภทที่ 1.2

| รูปแบบ   | ช่วงวัน         | ช่วงเวลา        | อัตราค่าไฟฟ้าต่อหน่วย (บาท) |
|----------|-----------------|-----------------|-----------------------------|
| Peak     | จันทร์ - ศุกร์  | 9.00 – 22.00 น. | 5.7982                      |
| Off Peak | จันทร์ - ศุกร์  | 22.00 – 9.00 น. | 2.6369                      |
| Off Peak | เสาร์ - อาทิตย์ | 0.00 – 24.00 น. | 2.6369                      |

### 3.2 Objective Function

ในงานวิจัยนี้ มีจุดประสงค์เพื่อคำนวณหาค่าต้นทุนการใช้ไฟฟ้า โดยหาค่าต้นทุนที่ต่ำที่สุด จากสมการ (3) และ (4)

$$\min (\sum_{i=1}^{74} P_i (\sum_{t=1}^{24} S_i^t TOU^t)) \quad (3)$$

เมื่อ

$$S_i^t = \begin{cases} L^P \in [0,1] \\ L^{OP} \in [0,1] \end{cases} \quad (4)$$

โดย  $P_i$  คือ กำลังไฟฟ้าของอุปกรณ์

$S_i^t$  คือ สถานการณ์ทำงานของอุปกรณ์ไฟฟ้า

โดยแบ่งเป็น 2 สถานะ คือ  $L^P$  ช่วงที่อุปกรณ์ทำงานช่วง Peak เวลา 9.00 – 22.00 น. และ  $L^{OP}$  ช่วงที่อุปกรณ์ทำงานช่วง Off Peak เวลา 22.00 – 9.00 น.

$TOU^t$  คือ อัตราค่าไฟฟ้าในแต่ละช่วง

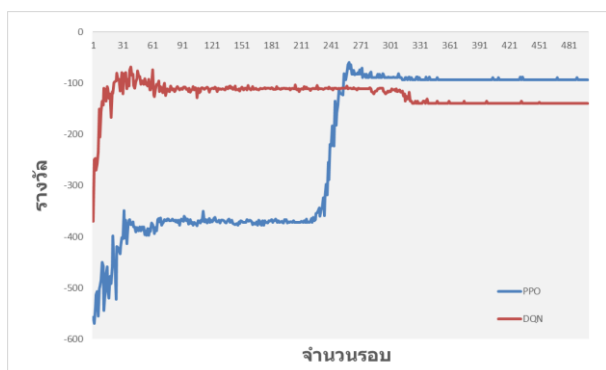
นอกจากนี้ในการคำนวณค่ากำลังไฟฟ้าที่ต้องใช้งานจากการไฟฟ้า ( $P_{Grid}$ ) จำเป็นต้องหักลบกำลังไฟฟ้าจากโซลาร์เซลล์ ( $P_{PV}$ ) ออกจากกำลังไฟฟ้าทั้งหมดของฟาร์มหมู ( $P_{USER}$ ) ดังสมการที่ (5) เนื่องจากฟาร์มหมูมีการสร้างพลังงานไฟฟ้าได้เองจากระบบโซลาร์เซลล์บางส่วน

$$P_{Grid} = P_{USER} - P_{PV} \quad (5)$$

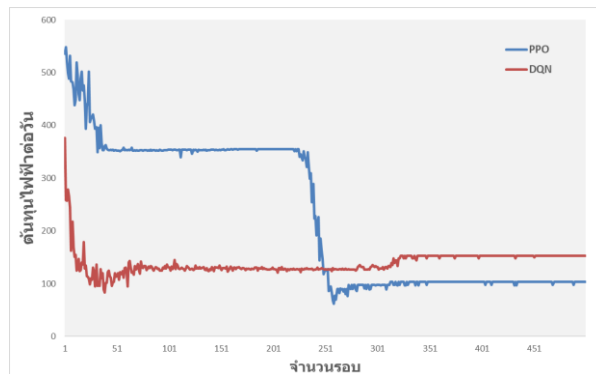
#### 4. ผลการทดลอง

จากการปรับใช้การเรียนรู้แบบเสริมกำลังด้วยการเรียนรู้จากการลองผิดลองถูก โดยต้องการการลดต้นทุนภายในฟาร์มหมูได้มากที่สุด ซึ่งในการใช้การเรียนรู้แบบเสริมกำลังในงานวิจัยนี้มีการใช้อุปกรณ์ไฟฟ้าภายในฟาร์มหมูเป็นตัวแทน โดยมีการโหลดตามเวลาที่ให้ลดต้นทุนมากที่สุด และตอบสนองตามเงื่อนไข พบว่าในการเรียนรู้แบบเสริมกำลังมีการตั้งค่าไฮเปอร์พารามิเตอร์ต่าง ๆ ในการเทรนโมเดลการเรียนรู้เชิงลึกทั้งสองรูปแบบดังตารางที่ 3 และการฝึกฝนโมเดลพบว่าค่ารางวัลมีแนวโน้มที่จะสูงขึ้นเรื่อย ๆ ดังรูปที่ 9 นอกจากนี้กราฟยังแสดงให้เห็นว่า PPO มีประสิทธิภาพดีกว่า DQN ในการจัดการตารางโหลด เนื่องจากสามารถปรับปรุงค่ารางวัลได้ต่อเนื่องและมีความเสถียรมากกว่า แม้ว่าในช่วงเริ่มต้นจะเรียนรู้ช้ากว่า DQN ก็ตาม ในส่วนของต้นทุนการใช้ไฟฟ้า มีแนวโน้มลดลงตามจำนวนรอบการเรียนรู้ด้วยเช่นเดียวกัน ดังรูปที่ 10 ตารางที่ 3 การตั้งค่าไฮเปอร์พารามิเตอร์ต่าง ๆ ในโมเดล

| พารามิเตอร์              | ค่าพารามิเตอร์ |           |
|--------------------------|----------------|-----------|
|                          | DQN            | PPO       |
| จำนวนรอบการเทรน          | 500            | 500       |
| Discount factor          | 0.99           | 0.99      |
| learning rate            | 0.001          | 0.001     |
| clipping for policy loss | -              | 0.2       |
| GAE Lambda               | -              | 0.95      |
| Hidden Layer Sizes       | [128,128]      | [128,128] |



รูปที่ 9 กราฟแสดงความสัมพันธ์ของค่ารางวัลและรอบ



รูปที่ 10 กราฟแสดงความสัมพันธ์ของค่ารางวัลและรอบ

จากผลการใช้ไฟฟ้าย้อนหลัง 31 เดือนของฟาร์มหมูพบว่าต้นทุนการใช้ไฟฟ้าเฉลี่ย 16,727 บาทต่อเดือน โดยหลังการจัดการโหลดด้วยการเรียนรู้แบบเสริมกำลังแบบ DQN และ PPO พบว่าสามารถหาต้นทุนไฟฟ้าเฉลี่ยต่อเดือนสำหรับแบบ DQN 5113.50 บาท และรูปแบบ PPO 4,019.70 บาท ซึ่งจากผลการทดสอบการจัดโหลดด้วยวิธี DQN ยังมีการจัดโหลดให้พัดลมดูดอากาศทำงานช่วงเวลาที่คิดค่าไฟฟ้าแบบ Peak อยู่บางส่วนทำให้ค่าไฟฟ้าที่ได้ยังไม่ถูกที่สุด ต่างจากวิธี PPO ที่สามารถจัดโหลดพัดลมและเครื่องให้อาหารให้อยู่ในช่วง Off-Peak ได้ทั้งหมด การจัดโหลดพัดลมดูดอากาศด้วยวิธี PPO มีการเวลาการทำงาน ดังตารางที่ 4 สำหรับการจัดโหลดการทำงานของเครื่องให้อาหารและโหลดแสงสว่าง ดังตารางที่ 5 และตารางที่ 6 แสดงรายละเอียดที่ได้หลังจากการจัดโหลดด้วยการเรียนรู้แบบเสริมกำลังแบบ DQN และ PPO

ตารางที่ 4 การจัดโหลดพัดลมดูดอากาศด้วยวิธี PPO (ช่วงเวลา Off-Peak)

| เวลา  | พัดลมตัวที่ 1 | พัดลมตัวที่ 2 | พัดลมตัวที่ 3 | พัดลมตัวที่ 4 | พัดลมตัวที่ 5 | พัดลมตัวที่ 6 |
|-------|---------------|---------------|---------------|---------------|---------------|---------------|
| 22.00 | 1             | 1             | 1             | 1             | 0             | 0             |
| 23.00 | 1             | 1             | 0             | 0             | 1             | 1             |
| 00.00 | 0             | 0             | 1             | 1             | 1             | 1             |
| 01.00 | 1             | 1             | 0             | 0             | 1             | 1             |
| 02.00 | 0             | 0             | 0             | 0             | 0             | 0             |
| 03.00 | 0             | 0             | 0             | 0             | 0             | 0             |
| 04.00 | 0             | 0             | 0             | 0             | 0             | 0             |
| 05.00 | 0             | 0             | 0             | 0             | 0             | 0             |
| 06.00 | 0             | 0             | 0             | 0             | 0             | 0             |
| 07.00 | 0             | 0             | 0             | 0             | 0             | 0             |
| 08.00 | 0             | 0             | 0             | 0             | 0             | 0             |
| 09.00 | 0             | 0             | 0             | 0             | 0             | 0             |

(1 = สถานะการเปิดทำงาน ,0 = สถานะการปิดทำงาน )



ตารางที่ 5 การจัดโหลดเครื่องให้อาหารและแสงสว่างด้วยวิธี PPO

| เวลา  | ช่วง     | เครื่องให้อาหาร | แสงสว่าง โหลดที่ 1 | แสงสว่าง โหลดที่ 2 |
|-------|----------|-----------------|--------------------|--------------------|
| 18.00 | Peak     | 0               | 1                  | 1                  |
| 19.00 | Peak     | 0               | 1                  | 1                  |
| 20.00 | Peak     | 0               | 1                  | 1                  |
| 21.00 | Peak     | 0               | 1                  | 1                  |
| 22.00 | Off-Peak | 1               | 1                  | 1                  |
| 23.00 | Off-Peak | 1               | 1                  | 1                  |
| 00.00 | Off-Peak | 0               | 1                  | 1                  |
| 01.00 | Off-Peak | 0               | 1                  | 1                  |
| 02.00 | Off-Peak | 0               | 1                  | 1                  |
| 03.00 | Off-Peak | 0               | 1                  | 1                  |
| 04.00 | Off-Peak | 0               | 1                  | 1                  |
| 05.00 | Off-Peak | 0               | 1                  | 1                  |

(1 = สถานะการเปิดทำงาน , 0 = สถานะการปิดทำงาน )

ตารางที่ 6 ผลการจัดโหลดด้วย DQN และ PPO

| โมเดล | กำลังไฟฟ้า ช่วง Off-Peak ต่อวัน (kWh) | กำลังไฟฟ้า ช่วง Peak ต่อวัน (kWh) | ต้นทุนไฟฟ้าเฉลี่ยต่อวัน (บาท) | การลดต้นทุน |
|-------|---------------------------------------|-----------------------------------|-------------------------------|-------------|
| DQN   | 10.55                                 | 24.47                             | 170.45                        | 69.43%      |
| PPO   | 16.12                                 | 15.65                             | 133.99                        | 75.96%      |

ในการใช้งานการเรียนรู้เชิงลึกแบบเสริมกำลังทั้ง 2 รูปแบบ ในการจัดตารางการทำงานของโหลดไฟฟ้าในฟาร์มสุกร พบว่าโมเดล PPO สามารถกระจายภาระโหลดไปยังช่วงเวลา Off-Peak ได้อย่างมีประสิทธิภาพมากกว่าโมเดล DQN โดยตารางการทำงานของพัดลม เครื่องให้อาหาร และแสงสว่างถูกปรับให้เหมาะสมกับโครงสร้างอัตราค่าไฟฟ้า TOU

ผลการเปรียบเทียบระหว่างโมเดล DQN และ PPO แสดงให้เห็นว่า การใช้พลังงานไฟฟ้าในช่วง Peak ของ PPO ลดลงเหลือ 15.65 kWh ต่อวัน ในขณะที่ DQN อยู่ที่ 24.47 kWh ขณะเดียวกัน การใช้ไฟฟ้าในช่วง Off-Peak ของ PPO สูงขึ้นเป็น 16.12 kWh ต่อวัน เมื่อเทียบกับ DQN ที่มีค่า 10.55 kWh ส่งผลให้ ต้นทุนค่าไฟฟ้าเฉลี่ยต่อวันของ PPO ลดลงเหลือ 133.99 บาท ในขณะที่ DQN มีค่า 170.45 บาท แสดงให้เห็นว่า PPO มีประสิทธิภาพในการจัดการโหลดมากกว่า DQN

## 5. สรุป

จากการผลการทดสอบการใช้การเรียนรู้แบบเสริมกำลังในการลดต้นทุนค่าไฟฟ้าของฟาร์มสุกร พบว่าสามารถลดต้นทุนเฉลี่ยต่อเดือน 69.43% และ 75.96% สำหรับรูปแบบ DQN และ PPO โดยวิธีการ PPO สามารถจัดโหลดปรับเปลี่ยนได้ให้อยู่ในช่วง Off-Peak ได้ทั้งหมด ทำให้อัตราค่าไฟฟ้าถูกลงมากกว่าครึ่งจากการใช้ไฟฟ้าแบบเดิมในฟาร์มสุกร ต่างกับรูปแบบ DQN ที่พบว่ายังมีอุปกรณ์บางส่วนที่ยังทำงานในช่วงเวลา

Peak ทำให้ลดต้นทุนค่าไฟฟ้าได้ไม่เท่ากับรูปแบบ PPO เนื่องจากการเรียนรู้เชิงลึกแบบเสริมกำลังแบบ DQN ไม่เหมาะกับการใช้งานที่ซับซ้อนมากนักทำให้มีประสิทธิภาพไม่เท่ารูปแบบ PPO ที่มีความต่อเนื่องและสามารถทำงานในสถานะการณ์ที่ซับซ้อนได้มากกว่า

อย่างไรก็ตามในการติดตามแนวโน้มค่ารางวัลต่อการการเรียนรู้พบว่าโมเดลทั้งสองมีการพัฒนาและปรับตัวอย่างต่อเนื่อง โดยค่ารางวัลเฉลี่ยมีแนวโน้มเพิ่มขึ้นในแต่ละรอบ ซึ่งสะท้อนให้เห็นว่าโมเดลเกิดการเรียนรู้อย่างแท้จริง สามารถเข้าใจรูปแบบการใช้พลังงานที่เหมาะสมต่อเงื่อนไขการใช้งานจริงของฟาร์มสุกร ทั้งนี้เมื่อพิจารณาผลลัพธ์โดยรวมทำให้เห็นว่าการเรียนรู้เชิงลึกแบบเสริมกำลังมีศักยภาพที่จะเป็นเครื่องมือในการจัดการระบบพลังงานภายในฟาร์มสุกร ช่วยให้ลดต้นทุน บริหารจัดการโหลด และอาจต่อยอดเพื่อนำไปใช้กับระบบฟาร์มอัจฉริยะ (Smart Farm) หรือระบบจัดการพลังงานในภาคส่วนอื่น ๆ ได้ในอนาคต

## 6. กิตติกรรมประกาศ

ขอขอบคุณ หจก.ศุภากรคอนโทรล ที่เอื้อเฟื้อข้อมูลการใช้ไฟฟ้าและสถานที่ในการวิจัย

## เอกสารอ้างอิง

- [1] A. Deansekeaw, N. Khortsriwong, P. Boonraksa, T. Boonraksa and B. Marungsri, "Optimal Load Scheduling for Smart Home Energy Management using Deep Reinforcement Learning," 2024 12th International Electrical Engineering Congress (iEECON), Pattaya, Thailand, 2024, pp. 1-4, doi: 10.1109/iEECON60677.2024.10537863.
- [2] P. Palensky and D. Dietrich, "Demand Side Management: Demand Response, Intelligent Energy Systems, and Smart Loads," IEEE Transactions on Industrial Informatics, 2011, Vol. 7, No. 3, pp. 381-388.
- [3] K. Yu, K. Jin and X. Deng, "Review of Deep Reinforcement Learning," 2022 IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China, 2022, pp. 41-48, doi: 10.1109/IMCEC55388.2022.10020015.
- [4] J. Hare, "Dealing with sparse rewards in reinforcement learning." arXiv preprint arXiv:1910.09281 (2019).
- [5] J. Jia and W. Wang, "Review of reinforcement learning research," 2020 35th Youth Academic Annual Conference of Chinese Association of Automation (YAC), Zhanjiang, China, 2020, pp. 186-191, doi: 10.1109/YAC51587.2020.9337653.
- [6] "Electricity Tariff," Provincial Electricity Authority(PEA), 2018