การจัดกลุ่มลักษณะการใช้ไฟฟ้าโดยใช้วิธีการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค
Electrical Load Profile Clustering Using Particle Swarm Optimization

ส่งเสริม ชัยเปรม และ พิสุทธิ์ รพีศักดิ์
ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์
E-mail: songsermnew@hotmail.com, fengpir@ku.ac.th

บทคัดย่อ

ปัจจุบันนี้พลังงานไฟฟ้าส่วนใหญ่ถูกผลิตขึ้นมาจากเชื้อเพลิงประเภทปิโตรเลียม แก๊ส ถ่านหิน ซึ่งเชื้อเพลิงเหล่านี้เป็นเชื้อเพลิงที่เหลือน้อยลงและเป็นพิษต่อสิ่งแวดล้อม ในขณะเดียวกันความต้องการใช้พลังงานไฟฟ้าก็เพิ่มสูงมากขึ้น จึงได้มีกลุ่มนักวิจัยค้นหาวิธีการจัดการด้านพลังงานอย่างมากมาย การจับกลุ่มข้อมูลการใช้ไฟฟ้าเป็นหนึ่งในวิธีการ การจัดการด้านพลังงาน ในบทความนี้ได้มีการนำเสนอวิธีการจัดกลุ่มข้อมูลการใช้ไฟฟ้าด้วยวิธีการจัดกลุ่มข้อมูลแบบกลุ่มอนุภาค (Particle Swarm Optimization Clustering, PSOC) ซึ่งเป็นการจัดกลุ่มข้อมูลที่มีประสิทธิภาพสูง แล้วนำผลการทดลองที่ได้มาเปรียบเทียบกับผลการจัดกลุ่มข้อมูลการใช้ไฟฟ้าแบบ K-Mean ซึ่งเป็นวิธีการจัดกลุ่มข้อมูลการใช้ไฟฟ้าที่นิยมใช้ทั่วไป โดยข้อมูลการใช้ไฟฟ้าที่ใช้ในการทดลองนี้มีผู้ใช้ไฟฟ้า 482 ราย เก็บค่าพลังงานการใช้ไฟฟ้าทุก 15 นาที ดังนั้นในช่วงเวลา 24 ชั่วโมงจะเก็บค่าพลังงานการใช้ไฟฟ้าได้ 96 ค่า ข้อมูลการใช้ไฟฟ้าได้ผ่านการทำความสะอาดข้อมูลและการทำบรรทัดฐานข้อมูล จากนั้นการจัดกลุ่มข้อมูลการใช้ไฟฟ้าด้วยวิธีการจัดกลุ่มแบบกลุ่มอนุภาคและวิธีการจัดกลุ่มแบบ K-Mean จำนวนกลุ่มข้อมูลที่ดีที่สุดถูกเลือกโดยใช้ด้วยวิธีการ Knee การเปรียบเทียบผลการจัดกลุ่มข้อมูลการใช้ไฟฟ้าที่ได้จากวิธีการจัดกลุ่มทั้ง 2 วิธี สามารถสรุปได้ว่าวิธีการจัดกลุ่มข้อมูลการใช้ไฟฟ้าแบบวิธีการจัดกลุ่มแบบกลุ่มอนุภาคมีประสิทธิภาพดีกว่าวิธีการจัดกลุ่มแบบ K-Mean และการจัดกลุ่มแบบกลุ่มอนุภาคลู่เข้าได้เร็วกว่าวิธีการจัดกลุ่มแบบ K-Mean

คำสำคัญ : ลักษณะการใช้ไฟฟ้า การจัดกลุ่มแบบวิธีการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค การจัดกลุ่มแบบ K-Mean

Abstract

Nowadays, most of the electricity is produced from petroleum, and coal. These fuels are depletable and produce toxic to the environment. In addition, the electricity demand is increasing. Many researches have been focusing in energy management. Load profile clustering is one of the energy management tools. In this paper, one of the most efficient clustering method, Particle Swarm Optimization Clustering (PSOC), was firstly explored in load profile clustering. The results were compared to the K-mean clustering which was a widely used method. It is a standard data clustering method. There were total of 482 load profiles. Each load profile was energy consumption recorded every 15 minutes for a 24-hour period. Therefore, there were total of 96 points in one load profile. The load profiles were cleansed and normalized. Then, they were clustered by particle swarm optimization and K-means clustering methods. The best number of groups was identified using the Knee method. The results could be concluded that the particle swarm optimization clustering method provided a better result than the K-means clustering method, and the particle swarm optimization clustering method converged faster than K-means clustering method.

Keywords :

Load Profile; Particle Swarm Optimization Clustering; K-Mean Clustering

1. บทนำ

ปัจจุบันนี้เทคโนโลยีระบบการวัดค่าทางไฟฟ้ามีมากมายหลากหลายชนิด เช่น ระบบการอ่านมิเตอร์อัตโนมัติ (Automatic Meter Reading, AMR) และระบบมิเตอร์อัจฉริยะ (Advance Metering Infrastructure, AMI) เป็นต้น ซึ่งระบบเหล่านี้สามารถอ่านและเก็บบันทึกค่าข้อมูลการใช้พลังงานไฟฟ้าของภาคครัวเรือนและภาคอุตสาหกรรมของประเทศได้ โดยข้อมูลที่ได้จากการอ่านและเก็บบันทึกค่าข้อมูลการใช้พลังงานไฟฟ้า เรียกว่าข้อมูลลักษณะการใช้ไฟฟ้า (Load Profile) ซึ่งข้อมูลลักษณะการใช้ไฟฟ้า (Load Profile) ของผู้ใช้ไฟฟ้าสามารถนำมาบริหารการจัดด้านการใช้พลังงานไฟฟ้าของผู้ใช้ไฟฟ้าได้ โดยวิธีการจัดการการใช้พลังงานไฟฟ้าของผู้ใช้ไฟฟ้าจะทำได้โดยการจำแนกลูกค้าบนพื้นฐานรูปร่างของลักษณะการใช้ไฟฟ้าและสามารถสร้างลักษณะการใช้ไฟฟ้าที่เป็นจุดศูนย์กลางกลุ่ม (Typical Load Profile) [1,2] ที่สามารถใช้เป็นรูปแบบภาระของการใช้ไฟฟ้า (Load for electrical) ของระบบจำหน่ายได้ [3]

ในขั้นตอนแรกของกระบวนการการจัดการด้านพลังงานไฟฟ้าจะต้องทำ การจัดกลุ่มผู้ใช้ไฟฟ้าจากรูปแบบภาระการใช้ไฟฟ้าของผู้ใช้ไฟฟ้า ผลที่ได้จะทำให้ทราบว่าผู้ใช้ไฟฟ้ามีรูปแบบการใช้ไฟฟ้ากี่แบบและมีลักษณะอย่างไรเพื่อจะได้นำไปหามาตรการที่เหมาะสมในการจัดการพลังงาน

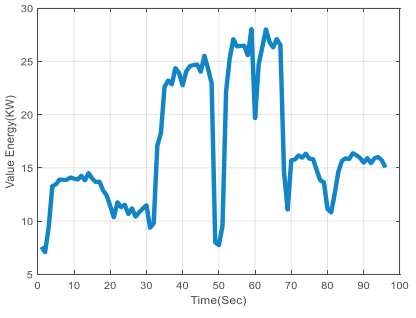
ความหลากหลายของวิธีการที่ได้รับการเสนอและผ่านการทดสอบโดยใช้ความแตกต่างของข้อมูลพื้นฐานลักษณะการใช้ไฟฟ้า เช่น K-Means Clustering [12,13] การจัดกลุ่มแบบเป็นลำดับขั้นตอน. (Hierarchical Clustering) [6,7] การแบ่งกลุ่มข้อมูลด้วย

โครงข่ายประสาทเทียมแบบแผนผังคุณลักษณะการจัดการตัวเอง (Self-Organizing Feature Map, SOFM) [4,5] โครงข่ายประสาทเทียมหรือข่ายงานประสาทเทียม (Neural Networks) [8] ตรรกศาสตร์คลุมเครือ (Fuzzy Systems) [9] แบบจำลองทางสถิต (Statistical Methods) [10] และเมื่อไม่นานมานี้ก็มีการคิดค้นขึ้นอีกหนึ่งวิธีคือ การจัดกลุ่มแบบวิธีซัพพอร์ตเวกเตอร์ (Support Vector Clustering) [11]

ในบทความนี้ผู้เขียนจะนำเสนอวิธีการจัดกลุ่มข้อมูลการใช้ไฟฟ้าโดยใช้วิธีการจัดกลุ่มแบบวิธีหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค (Particle Swarm Optimization Clustering, PSOC) [14,15] ซึ่งเป็นวิธีการจัดกลุ่มข้อมูลที่ง่ายมีความเสถียรภาพ และความน่าเชื่อถือสูง

2. ทฤษฎีการจัดกลุ่มข้อมูลลักษณะการใช้ไฟฟ้า

2.1 ลักษณะการใช้ไฟฟ้า (Load profile หรือ LP) ลักษณะการใช้ไฟฟ้าเป็นข้อมูลการใช้ไฟฟ้าที่วัดและบันทึกโดยมิเตอร์อิเล็กทรอนิกส์ ค่ากำลังไฟฟ้า (kW) จะถูกอ่านค่าและเฉลี่ยทุกๆ 15 นาที [19,20,21] การบันทึกเริ่มตั้งแต่เวลา 00:00 น. ค่าแรกที่บันทึกของวันคือกำลังไฟฟ้าเฉลี่ยในช่วง 00:00 ถึง 00:15 น. ดังนั้นหนึ่งวันจะบันทึกทั้งหมด 96 ค่า ตัวอย่าง LP แสดงอยู่ในภาพที่ 1



ภาพที่ 1 ตัวอย่างลักษณะการใช้ไฟฟ้า

2.2 การบรรทัดฐานข้อมูล (Data normalization) การบรรทัดฐานข้อมูลเป็นขั้นตอนการปรับ LP ให้มีค่าสูงสุดระดับเดียวกันเท่ากับ 1 เนื่องจากการวิเคราะห์ลุ่มข้อมูลของ LP จะพิจารณาเฉพาะรูปร่างของ LP เป็นหลัก การบรรทัดฐาน LP คำนวณได้จากสมการที่ 1 [19]

$$p_d = \frac{lp_d}{lp_{max}} \qquad (1)$$

เมื่อ lp_d คือ ลักษณะการใช้ไฟฟ้าที่ d โดย $lp_d = [lp_{d1}, lp_{d2}, \ldots, lp_{dm}, \ldots, lp_{d96}]$

lp_{dm} คือ ค่า kW เฉลี่ย 15 นาที ช่วงที่ m

lp_{max} คือ ค่า kW สูงสุดจากภายใน 1 วัน

p_d คือ ลักษณะการใช้ไฟฟ้า lp_d ที่ได้ทำบรรทัดฐานข้อมูลแล้ว

2.3 วิธีการจัดกลุ่ม K-Mean

K-Mean เป็นวิธีพื้นฐานในการจัดกลุ่มข้อมูลซึ่งมีขั้นตอนไม่ซับซ้อน วิธีนี้จะแบ่งกลุ่มข้อมูลตามจำนวนกลุ่ม (k) ที่ผู้ใช้งานกำหนด แต่ละกลุ่มจะมีจุดศูนย์กลางของกลุ่ม หากข้อมูลอยู่ใกล้จุดศูนย์กลางไหนก็จะเป็นสมาชิกของกลุ่มนั้น ดังนั้นเส้นแบ่งกลุ่มระหว่าง 2 กลุ่มที่ใกล้กันคือเส้นที่มีระยะห่างจากจุดศูนย์กลางทั้งสองกลุ่มเท่ากัน (Equidistance line) ขั้นตอนการจัดกลุ่มเริ่มจากสุ่มจุดศูนย์กลางของกลุ่มตามจำนวนกลุ่ม (k) ที่กำหนด ขั้นที่ 2 กำหนดสมาชิกให้ข้อมูลแต่ละตัวตามจุดศูนย์กลางกลุ่มที่ใกล้ที่สุด ขั้นที่ 3 คำนวณจุดศูนย์กลางกลุ่มใหม่โดยการหาค่าเฉลี่ยของสมาชิกในกลุ่ม ทำขั้นตอนที่ 2 และ 3 จุดศูนย์กลางกลุ่มเปลี่ยนแปลงน้อยกว่าค่ากำหนดหรือสมาชิกกลุ่มไม่เปลี่ยนแปลง

2.4 การจัดกลุ่มด้วยวิธีการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค (Particle Swarm Optimization Clustering)

วิธีการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค (Particle Swarm Optimization หรือ PSO) คิดค้นโดยนักวิทยาศาสตร์สองท่านคือ James Kennedy เป็นนักจิตวิทยา และ Russell C. Eberhart เป็นวิศวกรไฟฟ้า โดย PSO เป็นวิธีขั้นการหาคำตอบที่ดีที่สุดซึ่งสามารถนำมาประยุกต์ใช้ในการแก้ไขปัญหาทางคณิตศาสตร์และปัญหาทางวิศวกรรมอย่างมากมาย เทคนิคนี้พัฒนาโดยอาศัยการเลียนแบบพฤติกรรมทางสังคมของสิ่งมีชีวิต เช่นลักษณะการเคลื่อนที่ของฝูงนก ฝูงปลา โดยนก หรือปลาแต่ละตัวที่อยู่ในกลุ่ม ในที่นี้เรียกว่าอนุภาค (Particle) ในกรณีฝูงนกนกแต่ละตัวจะทำหน้าที่บินคนหาอาหารและอาจย้ายที่อยู่ไปพร้อมกัน จุดใดที่มีอาหารอยู่สามารถพักอาศัยได้ หรือเป็นจุดที่มีอันตรายนกจะสื่อสารกันในกลุ่มเพื่อแลกเปลี่ยนข้อมูล (Sharing Information) ระหว่างกัน การบินของนกหากมองเป็นพารามิเตอร์จะประกอบด้วยตำแหน่งที่นกบินอยู่ (Position) และความเร็วของการเคลื่อนที่ (Velocity) [16,17]

การจัดกลุ่มแบบกลุ่มอนุภาค จะแบ่งกลุ่มข้อมูลตามจำนวนกลุ่ม (k) ที่ผู้ใช้งานกำหนดขั้นตอนวิธีการเริ่มต้นด้วยการสุ่มอนุภาค จำนวน N_a ซึ่งประกอบด้วยค่าตำแหน่งของอนุภาค X และความเร็วของอนุภาค V โดย

$$X_i = [x_{i1}, ..., x_{ij}, ..., x_{ik}] \quad (2)$$

$$X_{ij} = [x_{ij1}, x_{ij2}, ..., x_{ij96}]^T \quad (3)$$

$$V_i = [v_{i1}, ..., v_{ij}, ..., v_{ik}] \quad (4)$$

$$V_{ij} = [v_{ij1}, v_{ij2}, ..., v_{ij96}]^T \quad (5)$$

เมื่อ X_i คือตำแหน่งของอนุภาค i ประกอบด้วยเวคเตอร์ x_{ij} จำนวน k ตัว

X_{ij} คือตำแหน่งของจุดศูนย์กลางกลุ่มข้อมูล j ที่ค้นหา โดยอนุภาค i

V_i คือความเร็วของอนุภาค i ประกอบด้วยเวคเตอร์ v_{ij} จำนวน k ตัว

v_{ij} คืออัตราการปรับตำแหน่งจุดศูนย์กลางกลุ่มข้อมูล j ที่ค้นหาโดยอนุภาค i

$i \in I$ โดย $I = [1, ..., N_a]$

$j \in J$ โดย $J = [1, ..., k]$

การค้นหาตำแหน่งจุดศูนย์กลางกลุ่มที่เหมาะสมอนุภาคจะเคลื่อนที่เป็นกลุ่มโดยแต่ละอนุภาคปรับตำแหน่งของตนเองจากความเร็วของอนุภาคเอง เพื่อไปยังตำแหน่งที่ดีที่สุดซึ่งประเมินโดยใช้ฟังก์ชั่นความเหมาะสม (Fitness function) ในสมการที่ 6 แต่ละลักษณะการใช้ไฟฟ้า p_d เป็นสมาชิกของกลุ่มข้อมูลซึ่งกำหนดด้วยค่า c_{id} ตามสมการที่ 7 นั่นคือ LP อยู่ใกล้จุดศูนย์กลางของกลุ่มใดเป็นสมาชิกของกลุ่มนั้น [22]

$$f(\mathbf{X}_i) = \sum_{j=1}^{k} \sum_{c_{id}=j, j\in J} \left\|\mathbf{p}_d - \mathbf{x}_{ij}\right\|^2 \quad (6)$$

$$c_{id} = \underset{j}{argmin} \left\|\mathbf{p}_d - \mathbf{x}_{ij}\right\|^2 \quad (7)$$

เมื่อ $f(X_i)$ คือค่าความเหมาะสมของอนุภาค i ค่า $f(X_i)$ ยิ่งน้อยยิ่งมีความเหมาะสม

c_{id} ทำหน้าที่เก็บหมายเลขกลุ่มของลักษณะการใช้ไฟฟ้า p_d สำหรับอนุภาค i

ตำแหน่งของอนุภาคจะมีการปรับเป็นรอบๆ (Iteration) เพื่อใช้ค้นหาตำแหน่งจุดศูนย์กลางกลุ่มที่ดีที่สุด โดยตำแหน่งของอนุภาคในรอบที่ $t + 1$ จะคำนวณจากตำแหน่งในรอบที่ t และความเร็วของ

อนุภาคในรอบที่ t ดังแสดงในสมการที่ 8 ในการค้นหารอบแรก t = 1 จะใช้ค่า $X_i(t = 0)$ และ $V_i(t = 0)$ ที่ได้จากการสุ่มเป็นค่าเริ่มต้น ตำแหน่งของแต่ละอนุภาคจะถูกประเมินค่าความเหมาะสมในแต่ละรอบ $f(X_i)$ และบันทึกตำแหน่งที่ดีที่สุด (ให้ค่า $f(X_i)$ น้อยที่สุด) ของตนเองเท่าที่ผ่านมาตั้งแต่รอบแรกถึงรอบที่ t ในตัวแปร $P_i(t)$ (Personal best position) ดังแสดงในสมการที่ 9 และบันทึกตำแหน่งที่ดีที่สุดเทียบกันทุกอนุภาคตั้งแต่รอบแรกถึงรอบที่ t ในตัวแปร $G_i(t)$ (Global best position) ดังแสดงในสมการที่ 10 ค่า $P_i(t)$ และ $G_i(t)$ ใช้คำนวณความเร็วของอนุภาคในรอบถัดไป ดังแสดงในสมการที่ 11

$$\mathbf{X}_i(t+1) = \mathbf{X}_i(t) + \mathbf{V}_i(t) \quad (8)$$

$$\mathbf{P}_i(t) = \begin{cases} \mathbf{P}_i(t-1), if\ f(\mathbf{X}_i(t)) > f(\mathbf{P}_i(t-1)) \\ \mathbf{X}_i(t), if\ f(\mathbf{X}_i(t)) \le f(\mathbf{P}_i(t-1)) \end{cases} \quad (9)$$

$$\mathbf{G}(t) = \underset{\mathbf{P}_i(t), i \in I}{argmin}\{f(\mathbf{P}_i(t))\} \quad (10)$$

$$\mathbf{V}_i(t+1) = \omega\mathbf{V}_i(t) + \alpha_1\beta_1(\mathbf{P}_i(t) - \mathbf{X}_i(t)) + \alpha_2\beta_2(\mathbf{G}(t) - \mathbf{X}_i(t)) \quad (11)$$

เมื่อ ω คือน้ำหนักแรงเฉื่อย (Inertia weight) ของอนุภาค

α_1 คือค่าคงที่อัตราเร่งของตำแหน่งที่ดีที่สุดของแต่ละอนุภาค

α_2 คือค่าคงที่อัตราเร่งของตำแหน่งที่ดีที่สุดของกลุ่มอนุภาค

β_1 และ β_2 คือค่าที่เกิดจากการสุ่มระหว่าง 0 ถึง 1

การค้นหาตำแหน่งจุดศูนย์กลางกลุ่มข้อมูลจะหยุดเมื่อจำนวนรอบการค้นหาถึงจำนวนรอบสูงสุดที่ผู้ใช้งานกำหนดไว้ตั้งแต่เริ่มต้น หรือ $f(G(t))$ ไม่มีค่าที่ดีขึ้นติดต่อกันถึงจำนวนรอบที่กำหนด เมื่อการค้นหาเสร็จสิ้น กลุ่มข้อมูลลักษณะการใช้ฟ้าระบุโดยตามสมการที่ 7 โดยจุดศูนย์กลางกลุ่มได้จาก $G(t)$ ของรอบสุดท้าย

2.5 ลักษณะการใช้ไฟฟ้าทั่วไป (Typical Load profile หรือ TLP)

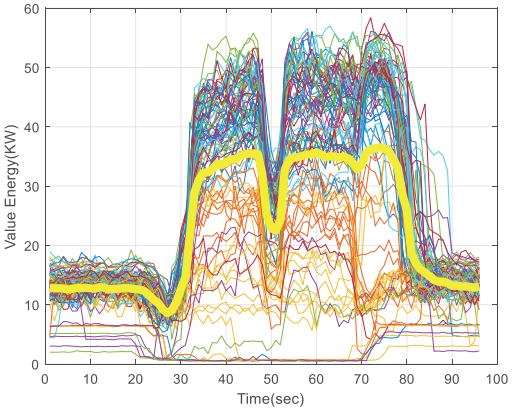
ลักษณะการใช้ไฟฟ้าทั่วไป (TLP) คือเซนทรอยด์ (Centroid) ของกลุ่มลักษณะการใช้ไฟฟ้า สามารถคำนวณโดยเฉลี่ยลักษณะการใช้ไฟฟ้าทั้งหมดในกลุ่ม ดังแสดงในสมการที่ 12 TLP ถูกใช้เป็นตัวแทนของกลุ่มเพื่อแสดงรูปร่างร่วมของ LP ในกลุ่ม เซ็นทรอยด์เป็นจุดศูนย์กลางที่แท้จริงของกลุ่ม สำหรับจุดศูนย์กลางกลุ่มที่ได้จาก $G(t)$ เป็นจุดศูนย์กลางกลุ่มที่ได้จากการค้นหาโดย PSO ซึ่งนำมาใช้ระบุสมาชิกของแต่ละกลุ่ม ตัวอย่างของ TLP แสดงในภาพที่ 2

$$\text{TLP} = \frac{\sum_{i=1, i \in I}^{N_k} M_{ik}}{N_k} \quad (12)$$

$i \in I$ โดย I = [1, 2, 3, ..., N_k]

N_k คือ จำนวนสมาชิกในกลุ่มที่ k

M_{ik} คือ สมาชิกตัวที่ i ในกลุ่มที่ k



ภาพที่ 2 ตัวอย่างลักษณะการใช้ไฟฟ้าทั่วไป (เส้นสีเหลือง) เทียบกับลักษณะการใช้ไฟฟ้าในกลุ่ม

3. การตรวจสอบผลการวิเคราะห์การจัดกลุ่มข้อมูลการใช้ไฟฟ้า (Cluster Validation)

3.1 การวัดระยะห่างระหว่างข้อมูล ซึ่งมีวิธีการวัดระยะห่างระหว่างข้อมูลดังต่อไปนี้

3.1.1 การวัดระยะห่างระหว่างสมาชิกของเวกเตอร์ x_s และสมาชิกของเวกเตอร์ x_t โดยใช้วิธีการวัดแบบ Euclidean distance เฉลี่ย ซึ่งสามารถคำนวณได้ดังสมการที่ 13

$$d_e(x_s, x_t) = \sqrt{\frac{1}{D}\sum_{i=1}^{D}(x_{s,i} - x_{t,i})^2} \quad (13)$$

เมื่อ D คือ จำนวนสมาชิกทั้งหมดในเวกเตอร์ x_s และเวกเตอร์ x_t

3.1.2 การวัดระยะห่างระหว่างสมาชิกของเวกเตอร์ x_s และสมาชิกของเวกเตอร์ x_t โดยใช้วิธีการวัดแบบ Cityblock distance ซึ่งสามารถคำนวณได้ดังสมการที่ 14

$$d_{cb}(x_s, x_t) = \sum_{i=1}^{D}|x_{s,i} - x_{t,i}| \quad (14)$$

เมื่อ D คือ จำนวนสมาชิกทั้งหมดในเวกเตอร์ x_s และเวกเตอร์ x_t

3.1.3 การวัดระยะห่างระหว่างสมาชิกของเวกเตอร์ x_s และสมาชิกของเวกเตอร์ x_t โดยใช้วิธีการวัดแบบ Chebychev distance ซึ่งสามารถคำนวณได้ดังสมการที่ 15

$$d_{ch}(x_s, x_t) = max_i|x_{s,i} - x_{t,i}| \quad (15)$$

กำหนดให้ $d(x, y)$ เป็นระยะห่าง (Distance) ระหว่างเวกเตอร์ x และเวกเตอร์ y ซึ่งในบทความนี้จะเลือกใช้ d_e, d_{cb} และ d_{ch} เพื่อนำผลการทดสอบมาเปรียบเทียบ

3.1.4 การวัดระยะห่างระหว่างสมาชิกในกลุ่มกับจุดศูนย์กลางกลุ่ม ซึ่งสามารถคำนวณได้ดังสมการที่ 16

$$g(c_k, S_k) = \sqrt{\frac{1}{N_k}\sum_{i=1}^{N_k}(x_i \in S_k)d^2(c_k, x_i)} \quad (16)$$

เมื่อ c_k คือ จุดศูนย์กลางของกลุ่มที่ k

S_k คือ เซตสมาชิกของกลุ่มที่ k

N_k คือ จำนวนสมาชิกของกลุ่มที่ k

3.1.5 การวัดระยะห่างเฉลี่ยระหว่างสมาชิกในกลุ่ม ซึ่งสามารถคำนวณได้ดังสมการที่ 17

$$h(S_k) = \sqrt{\frac{1}{2N_k}\sum_{i=1}^{N_k}(x_i \in S_k)d^2(x_i, x_k)} \quad (17)$$

เมื่อ S_k คือ เซตของสมาชิกในกลุ่มที่ k

X_i, X_k คือ สมาชิกของกลุ่มที่ k

3.2 การตรวจสอบผลการจัดกลุ่ม (Cluster Validation) โดยใช้การตรวจสอบด้วยการคำนวณค่าประสิทธิภาพการจัดกลุ่มด้วยดัชนีดังต่อไปนี้

3.2.1 Mean Square Error หรือ Error Function J ซึ่งสามารถคำนวณได้ดังสมการที่ 18

$$J = \frac{1}{N}\sum_{i=1}^{N} d^2(x_i, c_i) \ , x_i \in s_k \quad (18)$$

เมื่อ N คือ จำนวนสมาชิกทั้งหมดในกลุ่มที่ k

จากสมการที่ 18 สามารถอธิบายได้ว่าเป็นการหาระยะทางเฉลี่ยโดยรวมระหว่างสมาชิกในกลุ่มกับจุดศูนย์กลางกลุ่ม

3.2.2 Clustering Dispersion Indicator (CDI).ซึ่งสามารถคำนวณได้ดังสมการที่ 19

$$CDI = \frac{\sqrt{\frac{1}{K}\sum_{k=1}^{K} h^2(S_k)}}{\sqrt{\frac{1}{2K}\sum_{k=1}^{K} d^2(c^k, C_k)}} \qquad (19)$$

3.2.3 Mean Index Adequacy (MIA). ซึ่งสามารถคำนวณได้ดังสมการที่ 20

$$MIA = \sqrt{\frac{1}{K}\sum_{k=1}^{K} g^2(c_k, S_k)} \qquad (20)$$

3.2.4 Within Cluster Sum of Squares to Between Cluster Variation (WCBCR). ซึ่งสามารถคำนวณได้ดังสมการที่ 21

$$WCBCR = \frac{\sum_{k=1}^{K}\sum_{i=1}^{N} d^2(c_k, x_i)}{\sum_{1\le s<t}^{N} d^2(c_s, c_t)} \qquad (21)$$

3.2.5 Similarity Matrix Indicator (SMI). ซึ่งสามารถคำนวณได้ดังสมการที่ 22

$$SMI = max_{s>t}\left\{\left(1 - \frac{1}{\ln[d(c_s, c_t)]}\right)^{-1}\right\} \qquad (22)$$
$$; s, t = 1, \dots, N$$

3.2.6 Similarity Matrix Indicator 1 (SMI1). ซึ่งสามารถคำนวณได้ดังสมการที่ 23

$$SMI1 = \sqrt{max_{s>t}\left\{\left(1 - \frac{1}{\ln[d(c_s, c_t)]}\right)^{-1}\right\}} \qquad (23)$$
$$; s, t = 1, \dots, N$$

3.2.7 Modified Dunn Index (MDI). ซึ่งสามารถคำนวณได้ดังสมการที่ 24

$$MDI = max_{1\le q<K}\{\hat{d}(x_q)\}(min_{s\neq t}\{d(c_s, c_t)\})^{-1} \qquad (24)$$

3.2.8 Davies-Bouldin Index (DBI). ซึ่งสามารถคำนวณได้ดังสมการที่ 25

$$DBI = \frac{1}{K}\sum_{s,t=1}^{K} max_{s\neq t}\left\{\frac{\hat{d}(X_s) - \hat{d}(X_t)}{d(c_s, c_t)}\right\} \qquad (25)$$

3.2.9 Intra Cluster Index (IAI). ซึ่งสามารถคำนวณได้ดังสมการที่ 26

$$IAI = \sum_{i,k=1}^{K}\sum_{x_i\in X}(x_k - c_k).(x_k - c_k)^t \qquad (26)$$

3.2.10 Inter Cluster Index (IEI). ซึ่งสามารถคำนวณได้ดังสมการที่ 27

$$IEI = \sum_{k=1}^{K} N_k.(c_k - p).(c_k - p)^t \qquad (27)$$

เมื่อ $p = \frac{1}{N}\sum_{i=1}^{N} x_i$ คือค่าเฉลี่ยของข้อมูลทั้งหมด

3.2.11 Variance Criterion (VRC). ซึ่งสามารถคำนวณได้ดังสมการที่ 28

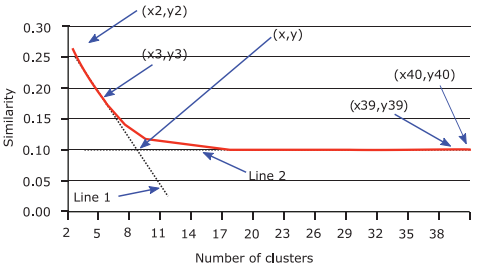
$$VRC = \frac{N-K}{K-1} \cdot \frac{IEI}{IAI} \qquad (28)$$

3.2.12 Scatter Index (SI). ซึ่งสามารถคำนวณได้ดังสมการที่ 29

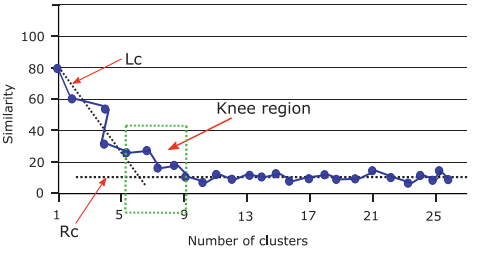
$$SI = \frac{\sum_{i=1}^{N} d^2(x_i, p)}{\sum_{k=1}^{K} d^2(c_k, p)} \qquad (29)$$

4. การเลือกจำนวนกลุ่ม (SELECTION OF CLUSTERS)

ในการเลือกจำนวนกลุ่มมีหลากหลายวิธี ในงานวิจัยนี้ขอใช้วิธีการ Knee โดยวิธีนี้จะทำการสร้างเส้นความชันที่สัมผัสกับส่วนโค้งที่กำลังหักศอกของกราฟดัชนีแล้วมาตัดกับเส้นสัมผัสกับเส้นกราฟที่วางตัวในแนวนอนดังภาพที่ 3 (ก) ซึ่งจุดที่เส้นสัมผัส 2 เส้นตัดกันคือ จุดที่มีจำนวนกลุ่มที่ดีที่สุด หลังนั้นก็ทำการตรวจสอบว่าจำนวนกลุ่มที่เลือกถูกต้องหรือไม่ โดยวิธีการแสดงเส้นกราฟเฉลี่ยของสมาชิกในกลุ่มหรือเส้นกราฟลักษณะการใช้ไฟฟ้าทั่วไป ทุกกลุ่ม ถ้าจำนวนกลุ่มถูกต้องเส้นกราฟนี้ควรมีค่าใกล้เคียงหรือเท่ากับ 1 ในกรณีที่เส้นกราฟดัชนีมีลักษณะความชันระหว่างจุดไม่ลดลงอย่างเดียวจะต้องมีการประมาณเส้นความชันเฉลี่ยสองเส้นมาตัดกันดังภาพที่ 3 (ข) แล้วหาช่วงพื้นที่ Knee region โดยขอบเขตพื้นที่ Knee region มีลักษณะเส้นกราฟที่เพิ่มขึ้นเล็กน้อยหลังจากลดลง แล้วค่อยๆ คงที่วางตัวตามแนวนอน หลังจากช่วงพื้นที่ Knee region แล้วทำการเลือกจำนวนกลุ่มที่ดีที่สุดในเขตช่วงพื้นที่ Knee region โดยจะเลือกจุดพื้นที่ตรงไหนก็ได้ในเขตช่วงพื้นที่ Knee region ซึ่งถ้าเลือกไปในทิศทางด้านขวาของพื้นที่ จำนวนที่ดีที่สุดก็จะมากขึ้น [30,31]



(ก)



(ข)

ภาพที่ 3 (ก) ตัวอย่างการเลือกจำนวนกลุ่มที่ดีที่สุดจากกราฟดัชนี Similarity ที่มีลักษณะเส้นกราฟที่มีความชันลดลงอย่างเดียว (ข) ตัวอย่างการเลือกจำนวนกลุ่มที่ดีที่สุดจากกราฟดัชนี Similarity ที่มีลักษณะเส้นกราฟที่มีความชันระหว่างจุดไม่ลดลงอย่างเดียว

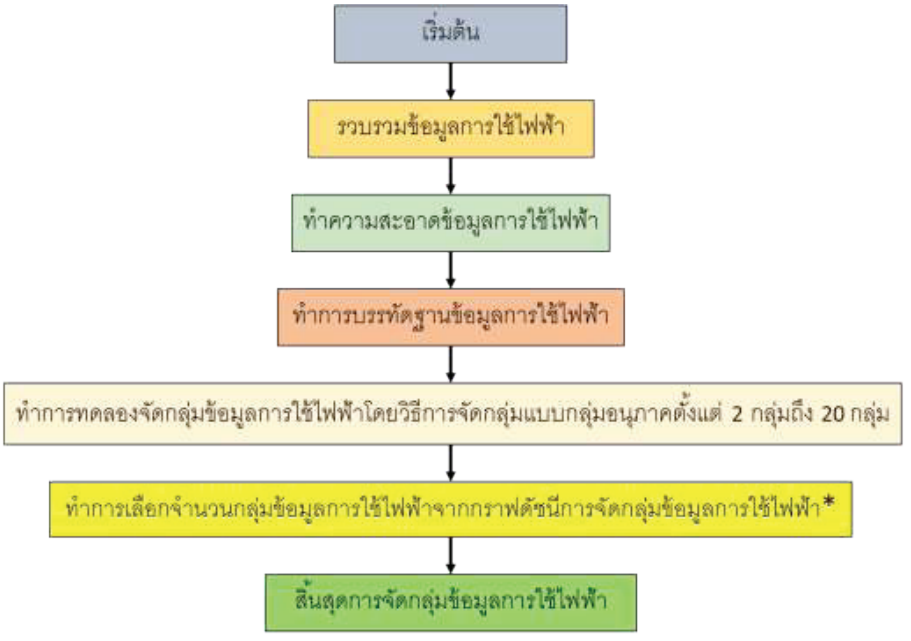
จากภาพที่ 3 (ก) จะเห็นว่าความชันของเส้นกราฟมีลักษณะที่ลดลงอย่างเดียวเพราะฉะนั้นเส้นในแนวสัมผัสกับเส้นโค้งสามารถหาได้จากความชันเส้นโค้ง แต่จากภาพที่ 3 (ข) จะได้ว่าความชันระหว่างจุดของเส้นกราฟมีลักษณะไม่ลดลงอย่างเดียว กล่าวคือ มีความชันที่เพิ่มขึ้นและลดลง เพราะฉะนั้นเส้นที่สัมผัสกับส่วนโค้งสามารถหาได้จากการประมาณค่าเฉลี่ยความชันเส้นกราฟ [3,23]

เมื่อพิจารณาที่ภาพที่ 3 (ข) จะพบว่าเส้นกราฟมีลักษณะความชันเพิ่มขึ้นและลดลงไม่คงที่ เพราะฉะนั้นการจำนวนกลุ่มที่ดีที่สุดหาได้จากช่วงพื้นที่ที่เรียก Knee region ซึ่งเป็นกรอบเส้นไข่ปลาสี่เหลี่ยมสีเขียวดังภาพที่ 3 (ข)

จากที่กล่าวมาข้างต้นในหัวข้อนี้เป็นวิธีการเลือกจำนวนกลุ่มที่เหมาะสม ซึ่งสามารถนำไปใช้ในการจัดกลุ่มข้อมูลได้หลายวิธี (K-Mean, PSO) โดยพิจารณาค่าดัชนีความคล้ายคลึง (Similarity index) เช่น J, CDI เป็นต้น

5. กรณีศึกษา (CASE STUDY)

ในการทดลองนี้จะนำข้อมูลการใช้ไฟฟ้าจากการไฟฟ้าแห่งประเทศไทยปีพุทธศักราช 2559 ของประเทศไทยมาใช้ในการทดลองซึ่งประกอบไปด้วยข้อมูลการไฟฟ้าของกลุ่มอุตสาหกรรมจำนวน 482 ราย แต่ละรายจะประกอบไปด้วยการบันทึกข้อมูลทุก 15 นาทีจนครบ 24 ชั่วโมง จะได้ค่าข้อมูล 96 ค่าใน 24 ชั่วโมง โดยทำการทดลองตามแผนภาพดังภาพที่ 4.

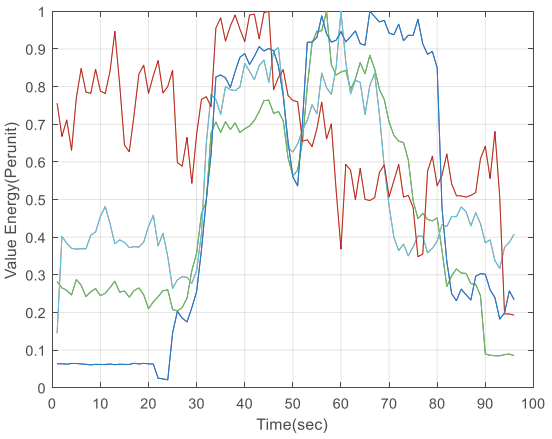


ภาพที่ 4 ขั้นตอนการจัดกลุ่มข้อมูลการใช้ไฟฟ้าที่ใช้ในการทดลองนี้

*ในการเลือกจำนวนกลุ่มจากกราฟดัชนีการจัดกลุ่มข้อมูลการใช้ไฟฟ้าสามารถทำได้โดยการแสดงกราฟค่าดัชนี J, MIA, CDI, WCBCR, SI และ VRC ของการจัดกลุ่มข้อมูลการใช้ไฟฟ้า โดยกราฟค่าดัชนีทุกดัชนีควรมีค่าลดลงและคงที่ แล้วทำการพิจารณาเลือกจำนวนกลุ่มโดยวิธีการ Knee

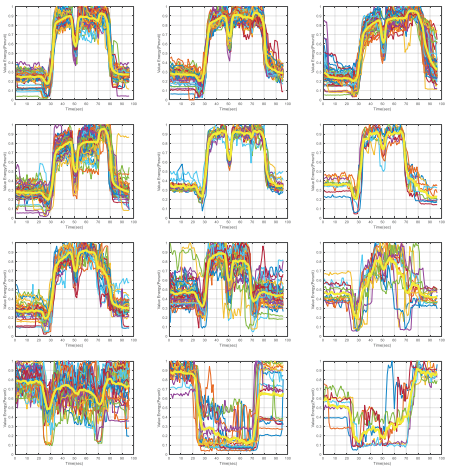
จากขั้นตอนการจัดกลุ่มข้อมูลการใช้ไฟฟ้าดังภาพที่ 4 สามารถอธิบายได้ว่าก่อนการทดลองจับกลุ่มข้อมูลการใช้ไฟฟ้าจะต้องมีการรวบรวมข้อมูลทำความสะอาดข้อมูลโดยกำจัดข้อมูลที่ไม่สมบูรณ์และข้อมูลที่ไม่เกี่ยวข้องออก ขั้นตอนต่อไปก็ทำการบรรทัดฐานข้อมูลเพื่อให้ข้อมูลมีค่าอยู่ในระดับเดียวกัน

ทั้งหมด ขั้นตอนสุดท้ายคือนำข้อมูลการใช้ไฟฟ้าที่ทำการบรรทัดฐานข้อมูลแล้วเข้าสู่กระบวนการการจัดกลุ่มแบบวิธีหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค ซึ่งจะได้ผลการทดลองดังต่อไปนี้

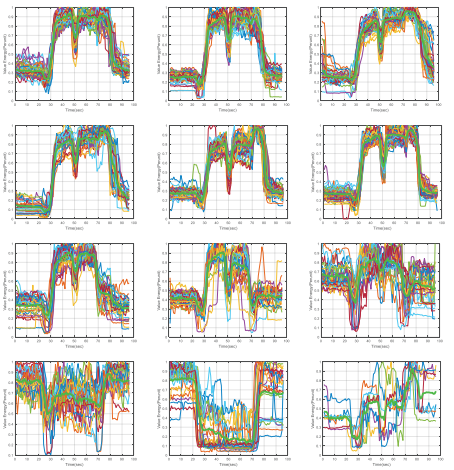


ภาพที่ 5 ข้อมูลการใช้ไฟฟ้า (Load Profile)

จากภาพที่ 5 เป็นตัวอย่างข้อมูลการใช้ไฟฟ้าของผู้ใช้ไฟฟ้า 482 รายที่นำมาทำบรรทัดฐานข้อมูลแล้ว ซึ่งการทำบรรทัดฐานข้อมูลได้กล่าวไว้ในหัวข้อที่ 2.2 จากภาพที่ 5 จะเห็นได้ว่าค่าสูงสุดของข้อมูลมีค่าเท่ากับ 1 และมีค่าสูงสุดต่างกันทุกๆ 15 นาที โดยในข้อมูลการใช้ไฟฟ้าค่าของพลังงานปกติจะมีมูลค่าสูงในเวลาเที่ยงคืนถึงห้านาฬิกาและเก้านาฬิกาถึงยี่สิบสี่นาฬิกา และจะมีค่าต่ำสุดในช่วงตั้งแต่ห้านาฬิกาถึงเก้านาฬิกา



ภาพที่ 6 PSO clustering euclidean distance 12 กลุ่ม



ภาพที่ 7 K-Mean clustering euclidean distance 1 กลุ่ม

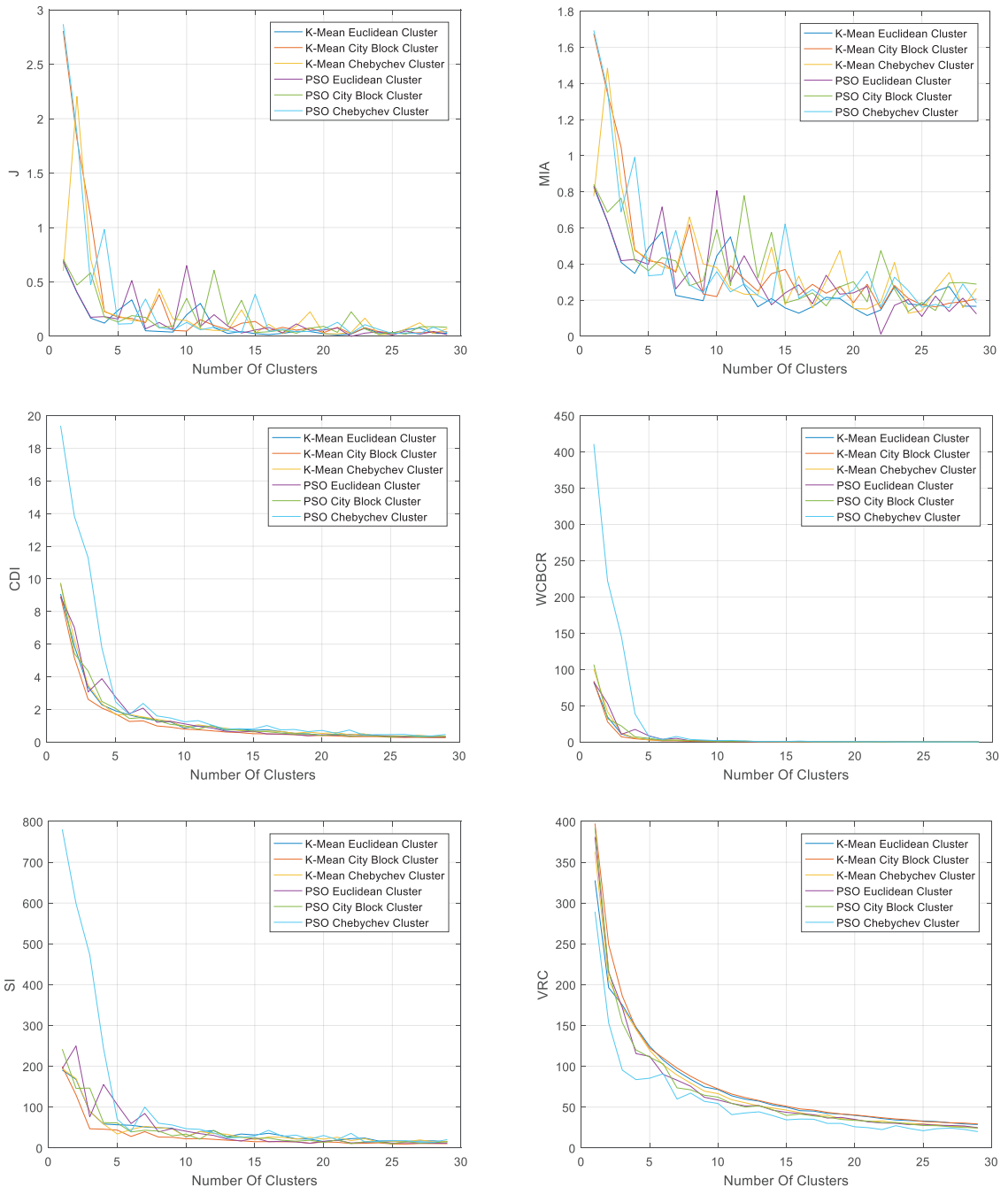
จากภาพที่ 6 เป็นการแสดงผลการจัดกลุ่มข้อมูลการใช้ไฟฟ้า 12 กลุ่มโดยใช้วิธีการการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค (Particle Swarm Optimization Clustering, PSOC) โดยใช้การวัดระยะห่างแบบ Euclidean Distance ซึ่งได้กำหนดให้เส้นกราฟใหญ่สีเหลืองเป็นจุดศูนย์กลางกลุ่ม โดยได้ทำการแสดงเส้นกราฟของจุดศูนย์กลางกลุ่มกับสมาชิกของกลุ่มร่วมกัน ซึ่งเป็นการแสดงให้เห็นว่ารูปร่างเส้นกราฟของจุดศูนย์กลางกลุ่มกับสมาชิกในกลุ่มมีรูปร่างเหมือนหรือคล้ายคลึงกัน และภาพที่ 7 เป็นการแสดงผลการจัดกลุ่มข้อมูลการใช้ไฟฟ้า 12 กลุ่มโดยใช้วิธีการจัดกลุ่มแบบ K-Mean โดยใช้การวัดระยะห่างแบบ Euclidean Distance ซึ่งได้กำหนดให้เส้นกราฟใหญ่สีเขียวเป็นจุดศูนย์กลางกลุ่ม โดยได้ทำการแสดงเส้นกราฟของจุดศูนย์กลางกลุ่มกับสมาชิกของกลุ่มร่วมกัน ซึ่งเป็นการแสดงให้เห็นว่ารูปร่างเส้นกราฟของจุดศูนย์กลางกลุ่มกับสมาชิกในกลุ่มมีรูปร่างเหมือนหรือคล้ายคลึงกัน จากภาพที่ 6 และภาพที่ 7 เป็นการเรียงรูปร่างของจุดศูนย์กลางกลุ่มโดยเรียงจากรูปร่างที่เหมือนกันไปสู่รูปร่างที่ต่างกัน รวมกันแล้วจะได้ 12 กลุ่ม และเมื่อทำการเปรียบเทียบวิธีการจัดกลุ่มทั้ง 2 วิธี จะพบว่าวิธีการการจัดกลุ่มแบบการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค มีความถูกต้องแม่นยำในการจัดกลุ่มมากกว่าวิธีการจัดกลุ่มแบบ K-Mean

เมื่อพิจาณาภาพที่ 6 และภาพที่ 7 จะพิจารณาได้ว่ามีการจัดกลุ่มข้อมูลการใช้ไฟฟ้าได้หลากหลาย และในแต่ละกลุ่มนั้นจะมีรูปร่างเส้นกราฟของจุดศูนย์กลางกับสมาชิกในกลุ่มที่มีความคล้ายคลึงกันมาก แต่ก็มีสมาชิกบางตัวในกลุ่มที่รูปร่างเส้นกราฟไม่เหมือนกับจุดศูนย์กลางกลุ่ม ซึ่งนั้นเป็นผลมาจากธรรมชาติของการจัดกลุ่มข้อมูล

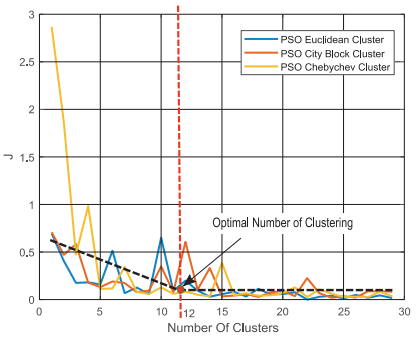
ในบทความนี้ผู้วิจัยยังได้มีการนำวิธีการวัดระยะห่างระหว่างข้อมูลแบบ Cityblock distance และ Chebychev distance มาเข้าร่วมพิจารณากับ Euclidean distance เพื่อค้นหาจำนวนกลุ่มที่เหมาะสมดังภาพที่ 8

เมื่อพิจารณาภาพที่ 8 จะได้ว่าทั้งสามวิธีของการวัดระยะห่างระหว่างข้อมูล คือวิธี Euclidean distance วิธี Cityblock distance และวิธี Chebychev distance ให้ผลการทดลองที่ไปในทิศทางเดียวกัน เพราะฉะนั้นในบทความนี้จะนำ Euclidean distance มาใช้ในการพิจารณาเลือกจำนวนกลุ่มที่เหมาะสม

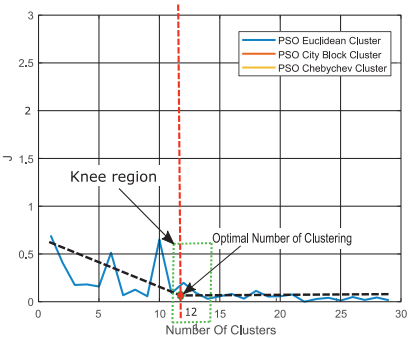
จากการทดลองการจัดกลุ่มข้อมูลการใช้ไฟฟ้าของผู้ใช้ไฟฟ้า 482 ราย ทั้ง 2 วิธี คือ การจัดกลุ่มโดยใช้วิธีการการจัดกลุ่มแบบการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค และการจัดกลุ่มโดยวิธี K-Mean จะได้ว่าเมื่อทำการพิจาณาดัชนีการจัดกลุ่มทั้ง 2 วิธี ดังภาพที่ 8 จะพบว่าการจัดกลุ่มทั้งสองวิธีมีค่าดัชนีทุกดัชนีลดลงแล้วก็คงที่ตามจำนวนกลุ่มที่เพิ่มขึ้น และเมื่อทำการเปรียบเทียบความแตกต่างระหว่างวิธีการจัดกลุ่มข้อมูลการใช้ไฟฟ้าทั้งสองวิธีจะพบว่าการจัดกลุ่มแบบวิธีหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาคมีค่าดัชนีลงและคงที่รวดเร็วกว่าการจัดกลุ่มแบบ K-Mean ซึ่งแสดงให้เห็นว่าการจัดกลุ่มแบบวิธีหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค มีประสิทธิภาพมากว่าวิธีการจัดกลุ่มแบบ K-Mean นอกจากนี้ชนิดวิธีการวัดระยะห่างระหว่างสมาชิกในกลุ่มด้วยกัน ระหว่าง สมาชิกในกลุ่มกับจุดศูนย์กลางกลุ่ม ระหว่างจุดศูนย์กลางกลุ่มกับจุดศูนย์กลางกลุ่มที่อยู่ต่างกลุ่มกัน และระหว่างจุดศูนย์กลางกลุ่มกับสมาชิกต่างกลุ่มกันของแต่ละวิธีการจัดกลุ่มข้อมูลการใช้ไฟฟ้าก็มีผลต่อความเร็วของการลดลงและการคงที่ของค่าดัชนีการจัดกลุ่มแต่ละค่าดัชนี ดังภาพที่ 9



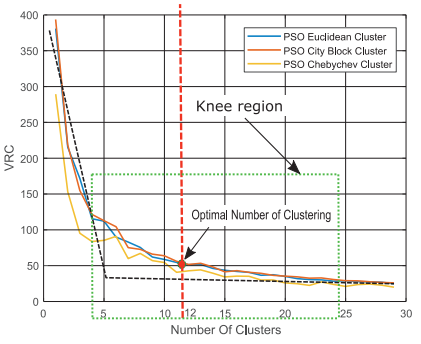
ภาพที่ 8 การเปรียบเทียบเครื่องมือที่ใช้ในการวัดการจัดกลุ่ม



(ก)



(ข)



(ค)

ภาพที่ 9 (ก) แสดงความแตกต่างของวิธีการวัดระยะห่างของดัชนี J ของการจัดกลุ่มลักษณะการใช้ไฟฟ้าแบบวิธีการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค (Particle Swarm Optimization Clustering, PSOC) (ข) แสดงการเลือกจำนวนกลุ่มที่ดีที่สุดโดยใช้การวัดระยะห่างแบบ Euclidean distance ของดัชนี J ของการจัดกลุ่มลักษณะการใช้ไฟฟ้า แบบวิธีการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค (Particle Swarm Optimization Clustering, PSOC) (ค) แสดงความแตกต่างของวิธีการวัดระยะห่างของดัชนี VRC ของการจัดกลุ่มลักษณะการใช้ไฟฟ้า แบบวิธีการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค (Particle Swarm Optimization Clustering, PSOC)

จากภาพที่ 9 จะได้ว่าการวัดระยะห่างแบบ Chebyshev จะทำให้ค่าดัชนีมีค่าลดลงอย่างรวดเร็วและคงที่อย่างรวดเร็วกว่าวิธีวัดระยะห่างแบบอื่นๆ ส่วนการวัดระยะห่างแบบ Euclidean กับแบบ City Block มักจะมีการเคลื่อนที่ของเส้นกราฟไปในทิศทางเดียวกัน อย่างไรก็ตามเมื่อทำการเปรียบวิธีการจัดกลุ่มข้อมูลลักษณะการใช้ไฟฟ้าทั้ง 2 วิธีจะพบว่าการจัดกลุ่มโดยใช้วิธีการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค มีค่าดัชนีลดลงและคงที่รวดเร็วกว่าวิธีการจัดกลุ่มข้อมูลลักษณะการใช้ไฟฟ้าแบบ K-Mean

เมื่อนำภาพที่ 8 มาวิเคราะห์ร่วมกับภาพที่ 9 จะได้ว่าค่าดัชนีทุกชนิดของการจัดกลุ่มข้อมูลการใช้ไฟฟ้าแบบวิธีการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค มีลักษณะลดลงและเริ่มจะคงที่ และจากภาพที่ 9 (ข) จะได้ว่าผู้วิจัยได้ทำการเลือกดัชนี J ที่ได้มาจากการจัดกลุ่มข้อมูลการใช้ไฟฟ้าแบบวิธีการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค โดยใช้การวัดระยะห่างระหว่างข้อมูลแบบ Euclidean มาพิจารณาในการเลือกจำนวนกลุ่มเนื่องจากเป็นดัชนีที่นิยมใช้กันและเป็นดัชนีพื้นฐานของการจัดกลุ่มข้อมูล โดยเมื่อวิเคราะห์

ภาพที่ 9 (ข) จะได้ว่าผู้วิจัยได้ทำการประมาณค่าเส้นความชันเฉลี่ย หาช่วงจำนวนกลุ่มที่เหมาะสมจาก Knee region และได้เลือกจุดที่หักศอกเป็นจำนวนกลุ่มที่เหมาะสมที่สุด ซึ่งจะได้จำนวนกลุ่มเท่ากับ 12 กลุ่ม และผู้วิจัยได้นำดัชนี VRC มาร่วมพิจารณาในการจัดกลุ่มข้อมูลการใช้ไฟฟ้าดังภาพที่ 9 (ค) ซึ่งจะได้ว่าจำนวนกลุ่มที่เหมาะสมที่สุดเท่ากับ 12 กลุ่มที่ได้จากดัชนี J อยู่ในช่วง Knee region ของดัชนี VRC เช่นกัน เพราะฉะนั้นในการทดลองนี้ผู้ทำการทดลองจึงทำการทดลองจัดกลุ่มทั้งสองวิธีเป็น 12 กลุ่ม ดังภาพที่ 6 และภาพที่ 7

จากภาพที่ 6 และภาพที่ 7 จะได้ว่าเป็นภาพผลการจัดกลุ่มข้อมูลลักษณะการใช้ไฟฟ้า 12 กลุ่ม โดยเรียงภาพกลุ่มที่เหมือนหรือคล้ายกันจากซ้ายไปขวา จากบนลงล่าง ของแต่ละวิธีการจัดกลุ่ม ซึ่งเมื่อนำภาพผลการทดลองทั้งสองวิธีมาเปรียบเทียบหาความแตกต่างต่างกันจะได้ว่า การจัดกลุ่มข้อมูลลักษณะการใช้ไฟฟ้า แบบวิธีการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค มีความละเอียด ความถูกต้องมากกว่าวิธีการจัดกลุ่มแบบ K-Mean กล่าวคือ การจัดกลุ่มข้อมูลลักษณะการใช้ไฟฟ้า แบบการจัดกลุ่มแบบวิธีการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาคจะมีรูปร่างกราฟจุดศูนย์กลางกลุ่มกับสมาชิกในกลุ่มที่เหมือนกันทั้งรอยยัก ทิศทางการเคลื่อนที่ของเส้นกราฟ แต่ในขณะที่วิธีการ การจัดกลุ่มของ K-Mean จะมีรูปร่างกราฟโดยรวมของสมาชิกในกลุ่มเหมือนกับรูปร่างกราฟจุดศูนย์กลางกลุ่ม แต่รอยยัก ทิศทางการเคลื่อนที่ของเส้นกราฟของจุดศูนย์กลางและของสมาชิกในกลุ่มไม่เหมือนกัน

จากวิธีการจัดกลุ่มข้อมูลแบบวิธีการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค และวิธีการจัดกลุ่มแบบ K-Mean ทั้งสองวิธีมีขั้นตอนการจัดกลุ่มข้อมูลลักษณะการใช้ไฟฟ้าที่คล้ายกันแต่แตกต่างกันตรงที่วิธีการจัดกลุ่มข้อมูลแบบวิธีการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค มีขั้นตอนการหาจุดศูนย์กลางกลุ่มแบบมีการสุ่มและการเสนอตำแหน่งจุดศูนย์กลางกลุ่มหลายชุดในจำนวนจุดศูนย์กลางเดียวกัน และยังมีการถ่วงน้ำหนักของตัวจุดศูนย์กลาง แต่ในขณะเดียวกันวิธีการจัดกลุ่มแบบ K-Mean มีการสุ่มหาจุดศูนย์กลางเพียงครั้งเดียวในขั้นตอนแรกต่อจากนั้นขั้นตอนที่สอง สาม ไปจนจบขั้นตอนจะเป็นการหาจุดศูนย์กลางโดยการหาค่าเฉลี่ยระยะห่างจากขั้นตอนแรก เพราะฉะนั้นวิธีการจัดกลุ่มข้อมูลลักษณะการใช้ไฟฟ้าแบบวิธีการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาคจะมีความละเอียดมากกว่าทั้งรอยยัก และทิศทางการเคลื่อนที่ของเส้นกราฟดังภาพที่ 6

ตารางที่ 1 ค่าดัชนีการจับกลุ่มข้อมูลการใช้ไฟฟ้า 12 กลุ่ม

Measure	PSO Clustering			K-Mean Clustering		
	Eucli-dean	City-block	Cheby-chev	Eucli-dean	City-block	Cheby-chev
J	0.199	0.607	0.084	0.08	0.100	0.055
MIA	0.446	0.779	0.2896	0.282	0.3169	0.233
CDI	0.864	1.022	1.0369	0.959	0.6921	0.938
WCB-CR	0.944	1.460	1.5457	0.920	0.5045	0.944
VRC	51.43	50.08	42.876	59.70	61.692	55.296
SI	30.570	42.805	36.851	42.39	20.0957	35.366

เมื่อพิจารณาตารางที่ 1 จะพบว่าการจัดกลุ่มข้อมูลการใช้ไฟฟ้าที่ 12 กลุ่มด้วยวิธีการจับกลุ่มแบบวิธีการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาคด้วยการวัดระยะห่างแบบ Euclidean distance จะให้ค่าดัชนี CDI และ SI น้อยกว่าค่า CDI และ SI ของการวิธีการจับกลุ่มแบบ K-Mean ด้วยการวัดระยะห่างแบบ Euclidean distance และในขณะเดียวกันวิธีการจับกลุ่มแบบวิธีการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาคของทุกเครื่องมือในการวัดระยะห่างก็ให้ค่าดัชนี VRC ที่มีค่าน้อยกว่าค่าดัชนี VRC ของวิธีการจับกลุ่มแบบ K-Mean ด้วยทุกเครื่องมือของการวัดระยะห่าง

จากตารางที่ 1 ให้ทราบว่าวิธีการ การจัดกลุ่มข้อมูลการใช้ไฟฟ้าแบบวิธีการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาคนั้นสามารถให้ค่าดัชนีน้อยกว่าค่าดัชนีของวิธีการจัดกลุ่มแบบ K-Mean ซึ่งเป็นการแสดง

ประสิทธิภาพที่ดีของวิธีการจัดกลุ่มแบบกลุ่มอนุภาค แต่ก็มีดัชนีบางค่าที่แสดงว่าวิธีการจัดกลุ่มแบบ K-Mean ดีกว่าวิธีการจัดกลุ่มแบบวิธีการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค

6. สรุปผล

งานวิจัยนี้มีจุดประสงค์เพื่อเปรียบเทียบประสิทธิภาพของวิธีการจัดกลุ่มข้อมูลการใช้ไฟฟ้าด้วย 2 วิธี คือ วิธีการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค และวิธีการ K-Mean เมื่อนำข้อมูลการใช้ไฟฟ้ามาทำการจัดกลุ่มด้วยทั้งสองวิธีแล้ว ต่อจากนั้นก็นำเครื่องมือวัดประสิทธิภาพการจัดกลุ่มข้อมูลที่ทั่วโลกยอมรับมาทำการวัดประสิทธิผลของการจัดกลุ่มข้อมูลทั้งสองวิธี โดยการจัดกลุ่มข้อมูลการใช้ไฟฟ้าทั้งสองวิธีในงานวิจัยนี้จะใช้เครื่องมือวัดระยะห่างระหว่างข้อมูลดังนี้ 1. Euclidean distance 2. Cityblock distance และ 3. Chebychev distance หลังจากที่ได้ทำการจัดกลุ่มข้อมูลการใช้ไฟฟ้าและทำการตรวจสอบการจัดกลุ่มข้อมูลการใช้ไฟฟ้าด้วยดัชนีการชี้วัดทั้ง 12 ชนิดแล้ว จะพบว่าการจัดกลุ่มข้อมูลการใช้ไฟฟ้าด้วยวิธีการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาคมีประสิทธิภาพดีกว่าการจัดกลุ่มด้วย K-Mean ซึ่งนั้นเพราะการจัดกลุ่มแบบวิธีการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาคมีการลู่เข้าได้เร็วกว่าการจัดกลุ่มแบบ K-Mean และประกอบกับสมาชิกในกลุ่มกับจุดศูนย์กลางกลุ่มมีรูปร่างเส้นกราฟที่คล้ายคลึงกันมากกว่าวิธีการการจัดกลุ่มแบบ K-Mean

7. เอกสารอ้างอิง

[1] Panapakidis, I.P., Christoforidis, G.C. and Papagiannis, G.K. (2013). Modifications of the Clustering Validity Indicators for the Assessment of the Load Profiling Procedure, the 4 th International Conference on Power Engineering, Energy and Electrical Drive 2013, 13-17 May 2013, Istanbul,Turkey.

[2] Chicco, G. (2012). Overview and performance assessment of the clustering methods for electrical load pattern grouping, journal homepage, Contents lists available at Sciverse Science Direct, Energy 42(2012) 68-80.

[3] Gavrilas, M., Gavrilas, G. and Sfintes, C.V. (2010). Application of Honey Bee Mating Optimization Algorithm to Load Profile Clustering, IEEE 2010.

[4] Gavrilas, M., Ivanov, O. and Gavrilas, G. (2008). Load Profiling with Fuzzy Self-Organizing, the 9th Symposium on Neural Network Applications in Electrical Engineering NEUREL 2008, Belgrade, Sept. 25-27, 2008, CD-ROM, ISBN:978-1-4244-2903-S.

[5] Verdu, S.V., Garcia, M.O., Senabre, C., Marin, A.G. and Franco, F.J.G. (2006). Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps, IEEE Trans.Poweryst., vol.21, no.4, pp. 1672-1682, Now.2006

[6] Figueiredo, V., Rodrigues, F., Vale, Z. and Gouveia, J.B. (2005). An electric energy consumer characterization framework based on data mining techniques, IEEE Trans. Power Syst., Vol.20.No.2,2005,pp. 596-602.

[7] Chicco, G., Napoli, R. and Piglione, F. (2006). Comparisons among clustering techniques for electricity customer classification, IEEE Trans. Power Syst., vol. 21,no. 2,pp.933-940, May 2006.

[8] Gerbec, D., Gasperic, S., Smon, I. and Gubina, F. (2004). Determining the load profiles of consumers based on fuzzy logic and probability neuralnetworks, Elect. Eng., Gener., Transm., Distrib., vol. 151, no. 3, pp. 395-400, May 2004.

[9] Özveren, C. S., Vechakanjana, C. and Birch, A.P. (2002). Fuzzy classification of electrical load demand profiles—a case study, IEE Power System Management Control, Apr. 17–19, 2002, pp. 353–358.

[10] Chen, C.S., Kang, M. S., Hwang, J. C. and Huang, C.W. (2000). Synthesis of power system load profiles by class load study, Elect. Power Energy Syst., vol. 22, pp. 325–330, 2000.

[11] G. Chicco, M. S. Ilie, "Support Vector Clustering of Electrical Load Pattern Data", IEEE Trans. Power Syst., Vol.24, No.3, 2009, pp.1619-1628.

[12] Forgy, E. (1965). Cluster Analysis of Multivariate Data: Efficiency versus Interpretability of Classification, Biometrics, Vol. 21, pp 768-769, 1965.

[13] Hartigan, J.A. (1975). Clustering Algorithms, John Wiley & Sons, New York, 1975.

[14] Kennedy, J. and Eberhart, R.C. (1995). Particle Swarm Optimization, Proceedings of the IEEE International Joint Conference on Neural Networks, Vol. 4, pp 1942–1948, 1995.

[15] Kennedy, J., Eberhart, R.C. and Shi, Y. (2002). Swarm Intelligence, Morgan Kaufmann, 2002.

[16] Kennedy, J. and Eberhart, R.C. (1995). Particle swarm optimization, In Proceedings of IEEE International Conference on Neural Networks, pages 1942-1948, Piscataway (USA), 1995.

[17] Kennedy, J. and Eberhart, R.C. (1997). A discrete binary version of the particle swarm algorithm, the 1997 IEEE International Conference on Systems, Man, and Cybernetics, volume 5, pages 4104–4108, Orlando (USA), October 1997.

[18] Van Der Merwe, D.W and Engelhrecht A.P. (2003). Data Clustering using Particle Swarm Optimization, Conference Paper January 2003, Conference, Proceedings of the IEEE Congress on Evolutionary Computation, CEC 2003, 8-12 December 2003, Canberra, Australia.

[19] กรัณย์ ตันทวีวงศ์ (2556). วิธีการจัดกลุ่มใหม่ (วิธีความน่าจะเป็นฟัซซี่ซีมีน สองขั้นตอน) สำหรับจัดกลุ่มโพรไฟล์ภาระ. ปริญญานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์.

[20] Load Profiles and their use in Electricity Settlement [Online],Available:https://www.elexon.co.uk/wp-content/uploads/2013/11/load_profiles_v2.0_cgi.pdf.

[21] BOBRIC1, E.C., CARTINA2, G. and GRIGORAŞ2 1 Stefan cel Mare, G. (2009). Clustering Techniques in Load Profile Analysis for Distribution Stations, University of Suceava str.Universitatii nr.13, RO-720229 Suceava, Romania 2 Gheorghe Asachi Technical University of Iaşi Bd. D. Mangeron nr.67, RO-700050 Iaşi, Romania.

[22] K Means [Online], Available:http://stanford.edu/~cpiech/cs221/handouts/kmeans.html.

[23] Azad, S.A., Ali, S. and Wolfsl, P. (2014). Identification of Typical Load Profiles using K-Means Clustering Algorithm, IEEE, 4-5 Nov. 2014.