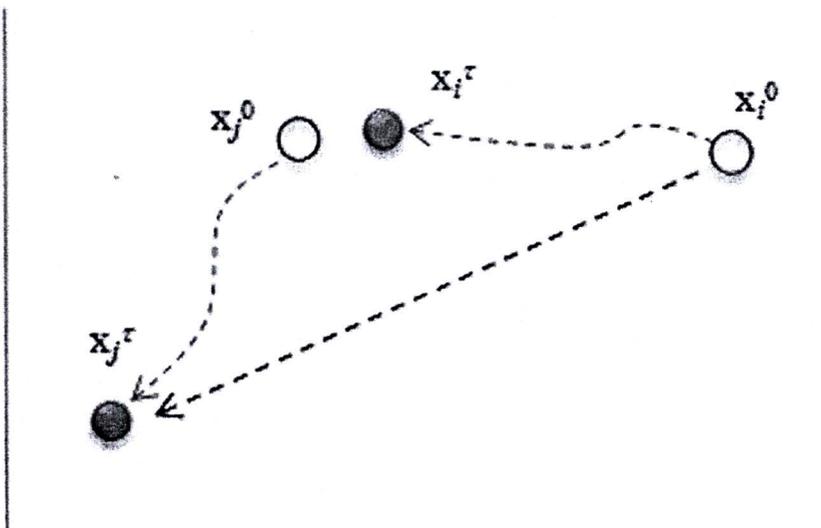


บทที่ 3

วิธีการที่นำเสนอและวิธีการดำเนินการวิจัย

วิธีการวิจัยครั้งนี้ได้นำเสนอขั้นตอนวิธีใหม่เพื่อลดการทำงานที่ซ้ำซ้อนและเพิ่มความเร็วขั้นตอนวิธีการย้ายตามค่าเฉลี่ย 2 ขั้นตอนวิธี ขั้นตอนวิธีแรกเรียกว่า ขั้นตอนวิธีการลำเลียงย้ายค่าเฉลี่ย (Transport Mean Shift Algorithm) มีแนวความคิดมาจากการลดการย้ายตำแหน่งที่ซ้ำซ้อนของขั้นตอนวิธีการย้ายตามค่าเฉลี่ยเดิม โดยการย้ายตำแหน่งนี้มีข้อมูลจำนวนมากที่ต้องย้ายตำแหน่งไปในเส้นทางเดียวกันหรือใกล้เคียง สุดท้ายตำแหน่งเหล่านี้ก็จะย้ายไปยังตำแหน่งที่มีความหนาแน่นสูงสุดตำแหน่งเดียวกัน ในขั้นตอนวิธีการลำเลียงย้ายค่าเฉลี่ยที่นำเสนอ ได้ทำการลดจุดเหล่านี้ โดยถือว่าเป็นจุดที่ต้องถูกลำเลียงที่อาศัยตัวแทนตัวเดียวกันจะช่วยให้การคำนวณในรอบต่อไปมีจำนวนตำแหน่งที่ต้องคำนวณน้อยลง และขั้นตอนวิธีที่สองเรียกว่า วิธีการลำเลียงย้ายค่าเฉลี่ยแบบครอบคลุม (Generalized Transport Mean Shift Algorithm) โดยจะนำการพิจารณาของเวกเตอร์มาใช้กับส่วนของการเลือกตัวแทนเพื่อลดความผิดพลาดของการจัดกลุ่มผิดในกรณีที่ข้อมูลอยู่ขอบและนำกราฟมาใช้เชื่อมต่อข้อมูลที่อาศัยตัวแทนกับตัวพ่วงหลังจากดำเนินการย้ายเสร็จสิ้นแล้วเพื่อลดการใช้หน่วยความจำ

1. ขั้นตอนวิธีการลำเลียงย้ายค่าเฉลี่ย (Transport Mean Shift Algorithm)



ภาพที่ 23 การย้ายตำแหน่งของจุดสองจุดในรอบที่ τ

มีการพิจารณาวิธีการเพื่อช่วยลดความซ้ำซ้อนของขั้นตอนวิธีการย้ายค่าเฉลี่ยในการคำนวณแต่ละรอบโดยการหาตัวแทนของตำแหน่งที่ลู่ออกไปยังจุดที่มีความหนาแน่นสูงสุดเดียวกัน ใช้แนวความคิดของจุดใกล้สุด และความสัมพันธ์ของการเชื่อมถึงกันได้ระหว่างแต่ละจุด เพื่อลดจุดเหล่านั้นแล้วใช้เฉพาะตัวแทนในการลำเลียงเพื่อคำนวณในรอบต่อไป โดยที่ตัวแทนยังคงใช้ขั้นตอนวิธีการย้ายตามค่าเฉลี่ย ส่งผลให้ตำแหน่งตัวแทนในแต่ละรอบ เป็นตำแหน่งเดียวกันกับตำแหน่งของขั้นตอนวิธีการย้ายตามค่าเฉลี่ยเดิม ดังแสดงตัวอย่างในภาพที่ 23 แสดงการย้ายในรอบที่ T ของตำแหน่ง i ที่ย้ายตำแหน่งไปยังจุดใหม่ที่ใกล้กับตำแหน่ง j ซึ่งเป็นตำแหน่งเริ่มต้นของข้อมูลเดิมที่ยังไม่มีการย้ายตำแหน่ง จะเห็นได้ว่าการคำนวณรอบต่อไปตำแหน่ง i จะย้ายไปตามลักษณะเดียวกับตำแหน่ง j ที่มีการย้ายตำแหน่งแล้วซึ่งคือตำแหน่ง x_j^T ตำแหน่ง i ควรพิจารณาตามการย้ายของตำแหน่ง j เสมือนกับว่าให้ตำแหน่ง j เป็นตัวแทน ดังนั้นตำแหน่ง i ก็ไม่จำเป็นต้องคำนวณอีกในรอบต่อไป

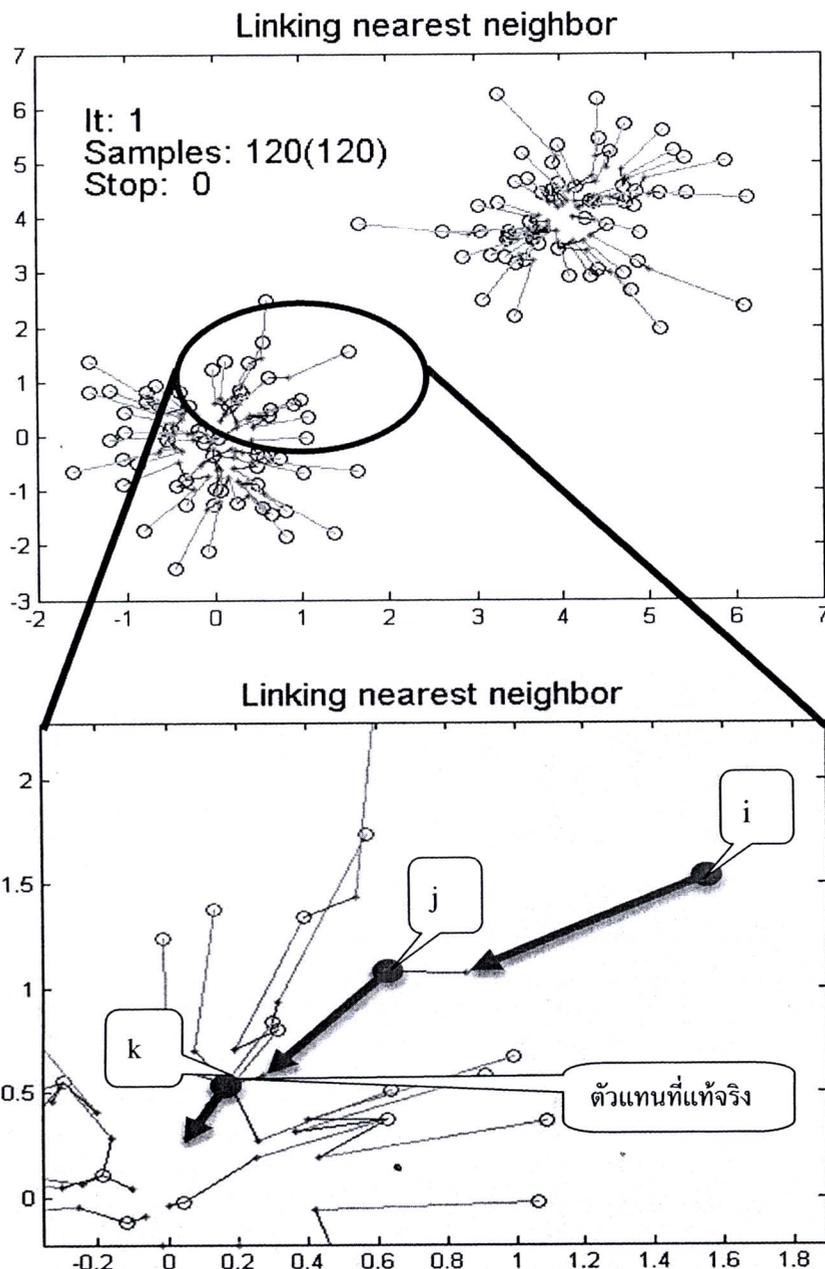
ดังนั้นผู้วิจัยจึงได้มีการนำเสนอขั้นตอนวิธีในการหาตำแหน่ง i เพื่อลดการคำนวณในรอบถัดไป และตำแหน่ง j เพื่อใช้เป็นตัวแทนในการลำเลียงสมาชิก ซึ่งตำแหน่ง j ที่ใช้เป็นตัวแทนนี้จะดำเนินการตามขั้นตอนวิธีการย้ายตามค่าเฉลี่ยแบบเดิม ดังนั้นตำแหน่งสุดท้ายของ j จะเป็นตำแหน่งเดียวกันกับขั้นตอนวิธีการย้ายตามค่าเฉลี่ยแบบเดิมเสมอ

การดำเนินการดังกล่าวข้างต้นสามารถทำได้โดยการใช้กราฟมาแทนความสัมพันธ์ระหว่างผู้ฟังกับผู้ขนส่ง โดยกำหนดให้ G เป็นกราฟไม่มีทิศทาง ที่มีจำนวนโหนดเท่ากับ n แต่ละโหนดแทนตำแหน่งของข้อมูลซึ่งมีทั้งหมด n ตำแหน่ง กราฟ G จะมีเส้นเชื่อม (i, j) ก็ต่อเมื่อในรอบที่ T ตำแหน่ง x_i^T อยู่ใกล้กับ x_j^0 มากที่สุด จะเรียก j เป็นตำแหน่งตัวแทนของ i หรือ x_j^T อยู่ใกล้กับ x_i^0 มากที่สุดจะเรียก i เป็นตำแหน่งตัวแทนของ j ดังสมการ

$$\begin{aligned} i &= \arg \min_k (\|x_i^T - x_k^0\|^2), \\ j &= \arg \min_k (\|x_j^T - x_k^0\|^2) \end{aligned} \quad (8)$$

จากนั้นจะทำการเก็บข้อมูลของผู้ขนส่งของข้อมูลแต่ละตัว ซึ่งในแต่ละรอบจะมีการหาผู้ขนส่งที่แท้จริง เช่น หากตำแหน่ง j เป็นผู้ขนส่งของตำแหน่ง i และตำแหน่ง k เป็นผู้ขนส่งของตำแหน่ง j ในรอบนั้น ดังนั้นตำแหน่ง k จะเป็นผู้ขนส่งที่แท้จริงของตำแหน่ง i โดยกำหนดให้ $h \in N^1$ ดังนั้น h_i จะต้องเป็นไปตามกฎดังต่อไปนี้

$$\left. \begin{aligned} 1. & h_i = \arg \min_j (\|x_i^T - x_j^0\|^2), \text{ กฎเบื้องต้น} \\ 2. & 1 \leq h_i \leq n \\ 3. & h_i = h_{h_i}, \text{ กฎแห่งความต่อเนื่อง} \end{aligned} \right\} \quad (9)$$



ภาพที่ 24 การย้ายตำแหน่งและความสัมพันธ์กับผู้อื่น

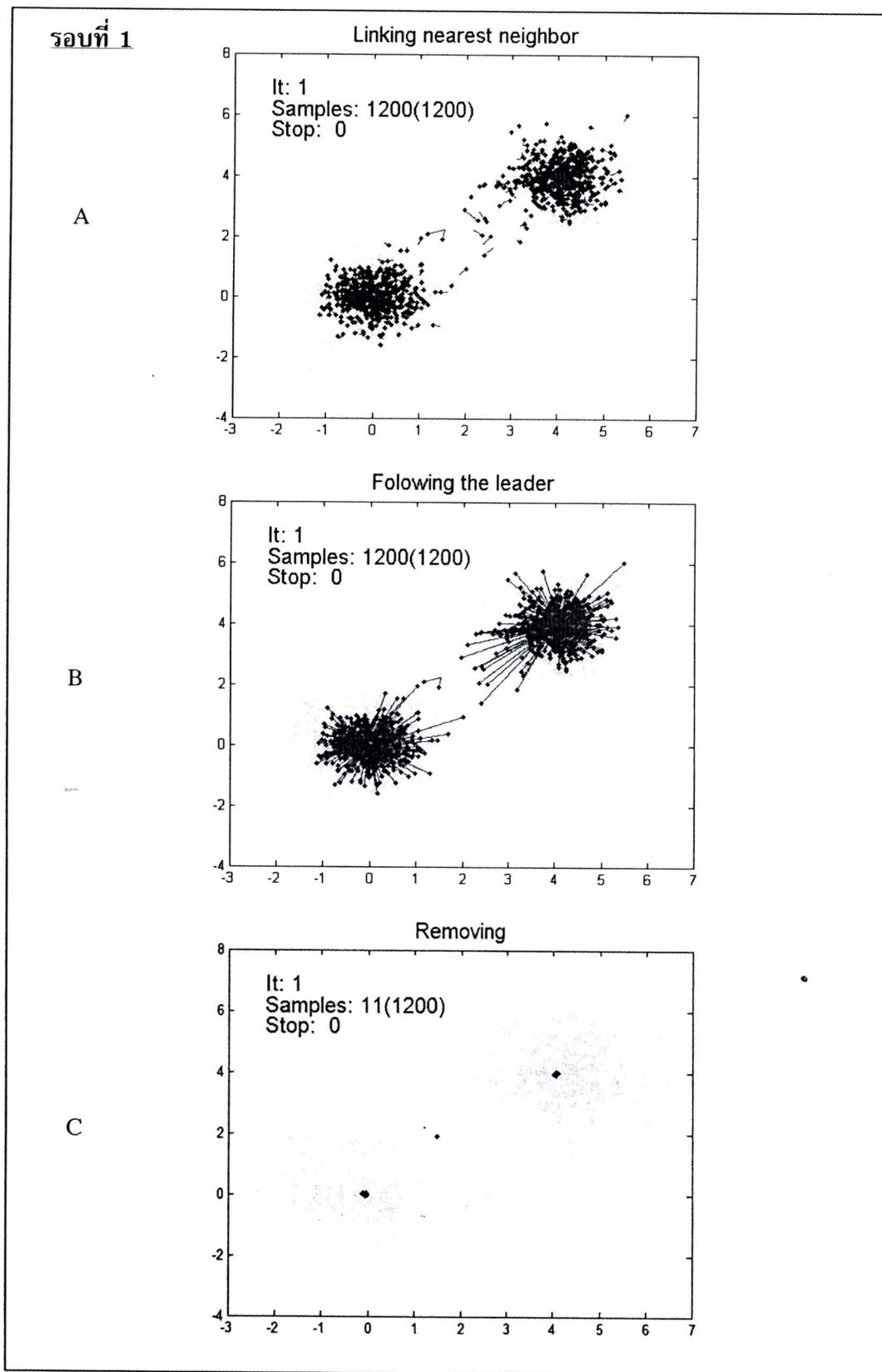
จากภาพที่ 24 จะเห็นว่า h_i ซึ่งเป็นตำแหน่งตัวแทนของตำแหน่ง i ที่ระบุโดยสมการที่ (9) เป็นไปตามกฎเบื้องต้นในที่นี้คือตำแหน่ง j ซึ่งจะระบุค่า h_i ในตอนแรก ตำแหน่งใดที่มีการเชื่อมไปถึงกันก็จะใช้ตัวแทนเดียวกันซึ่งเป็นไปตามกฎข้อที่ 3 คือ กฎแห่งความต่อเนื่อง ดังนั้นเมื่อพิจารณากฎข้อนี้แล้ว h_i ก็คือตำแหน่ง k ซึ่งไม่ใช่ตำแหน่งใกล้สุดตามกฎเบื้องต้น และกราฟ G จะต้องมีส่วนเชื่อมทุก ๆ (i, h_i) จากนั้นในแต่ละรอบจะพิจารณาการย้ายไปตำแหน่งที่ $\nabla p(x) = 0$ โดยใช้เงื่อนไขการหยุดเช่นเดียวกับ ขั้นตอนวิธีการย้ายตามค่าเฉลี่ยเดิมซึ่งเป็นไปตามสมการที่ (6)

ในแต่ละรอบตำแหน่งที่หยุดแล้วจะไม่ถูกนำมาคำนวณหรือพิจารณาอีกในรอบต่อไป กำหนดให้ $S \in R^{m \times n'}$ แทนตำแหน่งที่หยุดแล้วเมื่อ n' คือจำนวนข้อมูลที่หยุดตามเงื่อนไขในสมการ (6) โดยเวกเตอร์ $s_i \in S$ ซึ่งเป็นตำแหน่งสุดท้ายของข้อมูล x_i ดังนั้น s_i เป็นตำแหน่งที่ $\nabla p(x_i) = 0$ ทุกจุดที่มี s_i เป็นตัวแทนก็จะถือว่าถูกลำเลียงมาหยุดที่ตำแหน่ง s_i และตำแหน่งสุดท้ายของจุดเหล่านั้นจะถูกแทนด้วยตำแหน่ง s_i นี้ ซึ่งหากนำไปใช้กับการจัดกลุ่มข้อมูลข้อมูลที่มีตัวแทนเดียวกันจะถือว่าเป็นข้อมูลที่อยู่ในกลุ่มเดียวกัน รวมทั้งมีความเป็นไปได้ว่า s_i กับ s_j จะอยู่ในกลุ่มเดียวกันหาก s_i กับ s_j อยู่ใกล้กันมากพอซึ่งถูกกำหนดโดยค่าคงที่ค่าหนึ่ง หมายความว่าแต่ละตัวนั้นจะเข้าสู่ความหนาแน่นสูงสุดเดียวกันนั่นเอง จากนั้นจะเพิ่มเส้นเชื่อมตำแหน่ง i กับ j นี้ในกราฟ G แล้วหาส่วนที่เชื่อมกันของกราฟเพื่อจัดกลุ่มข้อมูลทั้งหมดในขั้นตอนสุดท้าย ขั้นตอนวิธีการลำเลียงย้ายค่าเฉลี่ยสามารถดำเนินการได้ดังนี้

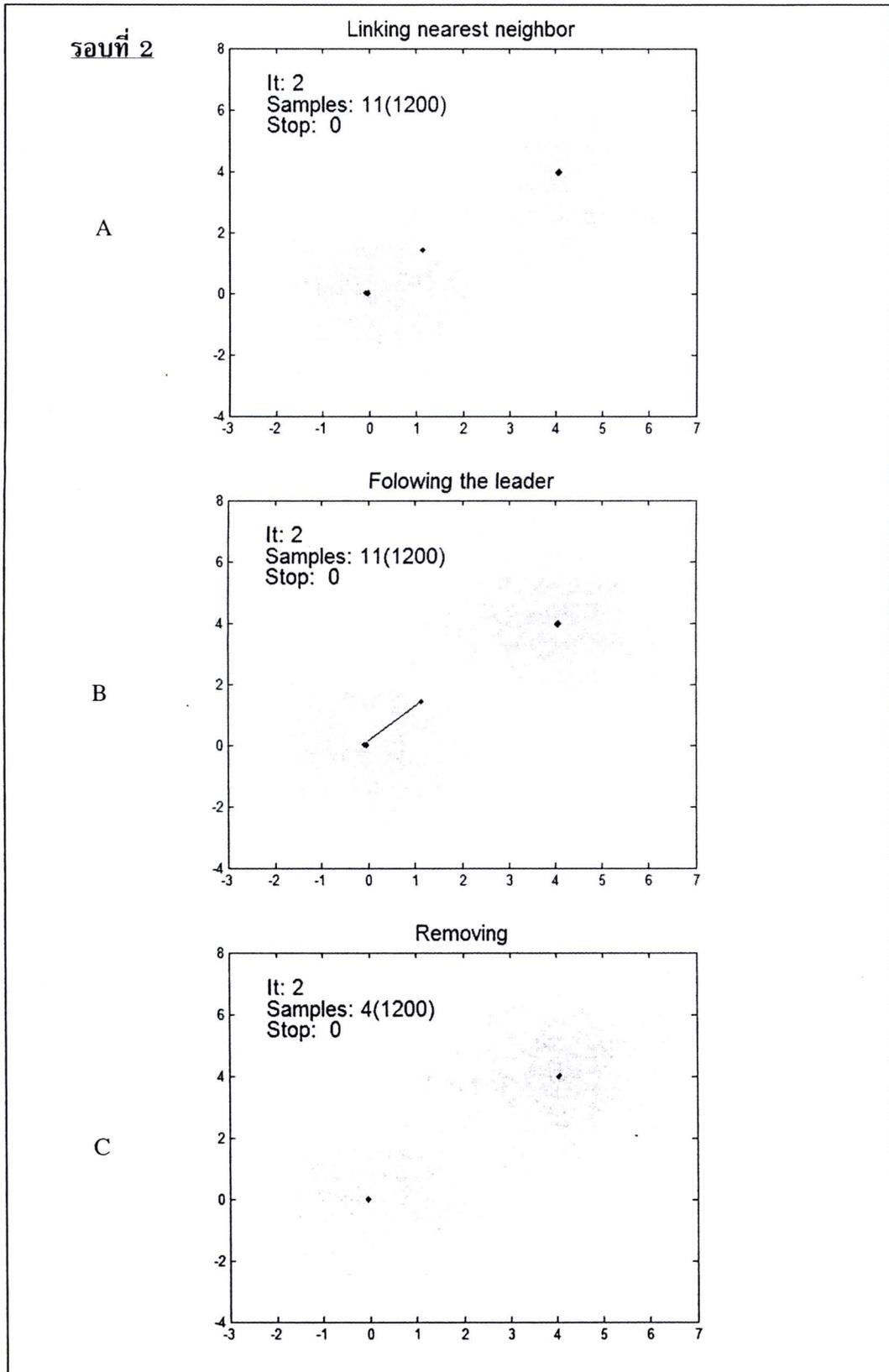
ขั้นตอนวิธีการลำเลียงย้ายค่าเฉลี่ย (Transport Mean Shift Algorithm)

1. ย้ายตำแหน่งตามขั้นตอนวิธีการย้ายค่าเฉลี่ยแบบเดิม
2. หาค่าตำแหน่งที่ใกล้ที่สุดระหว่างตำแหน่งที่ย้ายมากับตำแหน่งเริ่มต้นตามกฎเบื้องต้น
3. หาเมตริกซ์ตำแหน่งของตัวแทน (h) ซึ่งเป็นไปตามกฎแห่งความต่อเนื่อง
4. ในจำนวนของตัวแทนนี้ หาตัวแทนที่หยุดการคำนวณตามสมการที่ (6)
5. คงข้อมูลที่เป็นตัวแทนซึ่งยังไม่หยุดการคำนวณ เพื่อใช้ในการคำนวณในรอบต่อไป
6. ทำซ้ำขั้นตอน 1-5 จนกระทั่งไม่เหลือข้อมูลในการคำนวณ

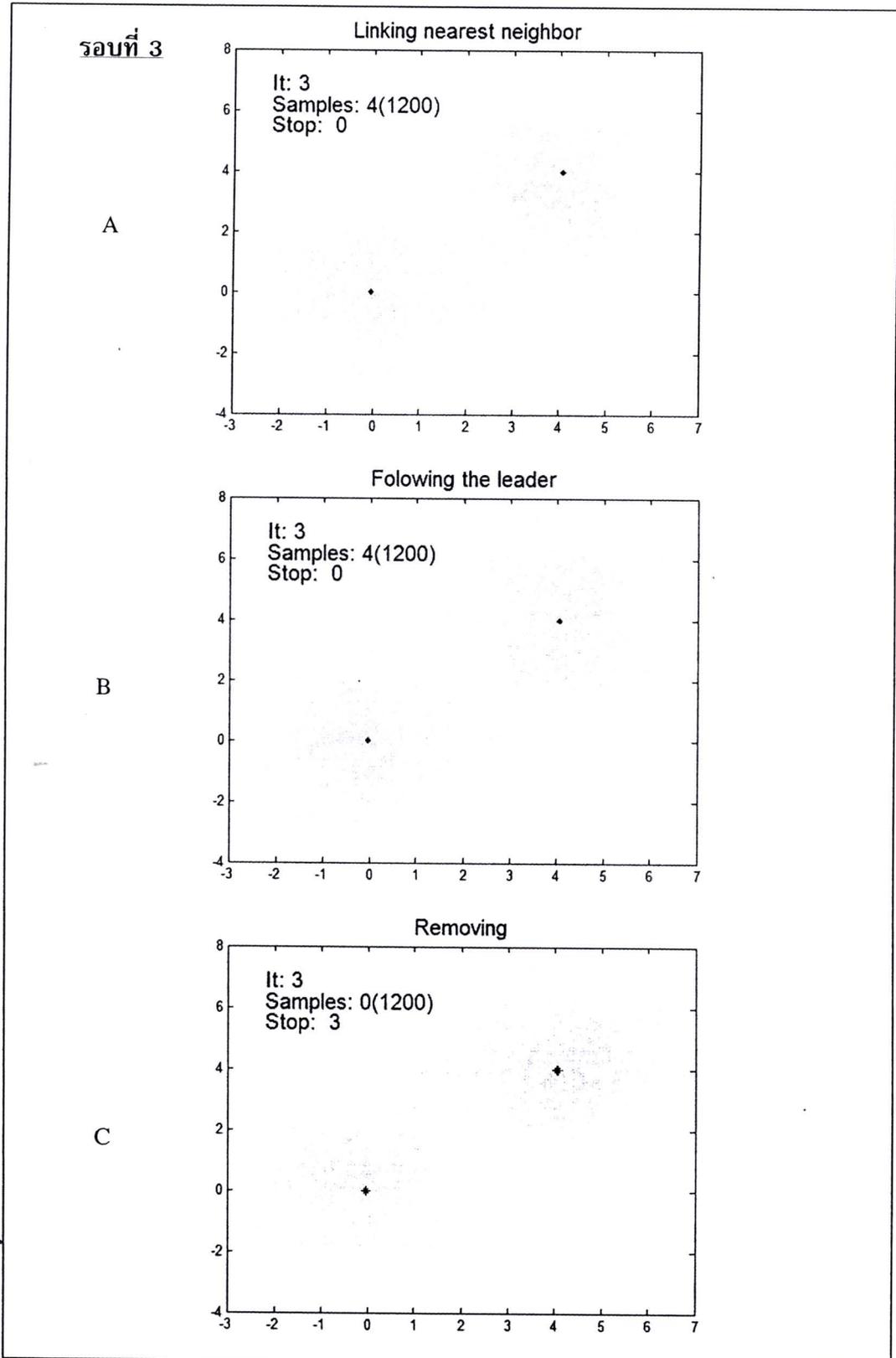
เพื่อให้เห็นการทำงานของขั้นตอนวิธีได้อย่างชัดเจน ผู้วิจัยได้ทำการสุ่มข้อมูล 2 มิติ จำนวน 1200 ตำแหน่งแบ่งเป็น 2 กลุ่ม เป็นข้อมูลแบบกระจายตัวปกติแล้วดำเนินการตามขั้นตอนวิธีการลำเลียงย้ายค่าเฉลี่ย (TMS) ดังแสดงในภาพที่ 25 โดยจะแสดงผลภาพผลลัพธ์ในแต่ละรอบของการคำนวณซึ่งข้อมูลที่ทำการสุ่มขึ้นมาจะมีการคำนวณทั้งหมด 3 รอบ ซึ่งจะแสดงผลลัพธ์บางช่วงของขั้นตอนวิธีการลำเลียงย้ายค่าเฉลี่ย ดังนี้คือ แถว (A) เป็นภาพของจุดที่ย้ายแล้วกับจุดที่อยู่ใกล้ที่สุดตามขั้นตอนที่ 1 และ 2 ของขั้นตอนวิธีการลำเลียงย้ายค่าเฉลี่ยแสดงให้เห็นจุดทั้งสองโดยเส้นเชื่อม แถว (B) แสดงเส้นเชื่อมระหว่างจุดที่ย้ายแล้วกับตัวแทน ตามขั้นตอนที่ 3 และแถว (C) แสดงจุดที่ยังเหลืออยู่หลังจากขั้นตอนที่ 5 ของขั้นตอนวิธีการลำเลียงย้ายค่าเฉลี่ย ซึ่งจะเห็นว่าจำนวนจุดที่เหลือในการคำนวณในรอบต่อไปจะเหลือน้อยลง โดยในรอบที่ 1, 2 และ 3 จะเหลืออยู่เพียง 11, 4 และ 0 จุดตามลำดับ จะเห็นว่าขั้นตอนวิธีการลำเลียงย้ายค่าเฉลี่ยสามารถลดจำนวนจุดที่ใช้คำนวณลงได้เป็นจำนวนมาก



ภาพที่ 25 แสดงผลการจัดกลุ่มบางขั้นตอนของข้อมูลทดสอบโดยใช้ขั้นตอนวิธีการลำเลียงย้ายค่าเฉลี่ย



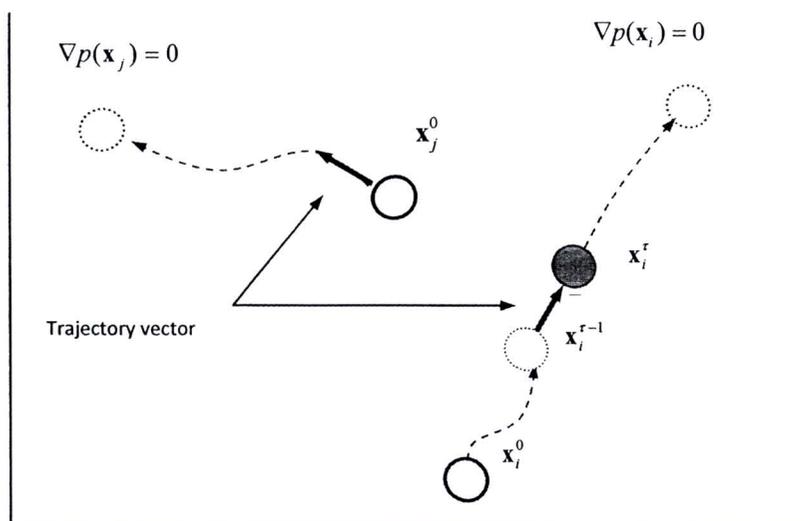
ภาพที่ 25 แสดงผลการจัดกลุ่มบางขั้นตอนของข้อมูลทดสอบโดยใช้ขั้นตอนวิธีการลำเลียงย้ายค่าเฉลี่ย (ต่อ)



ภาพที่ 25 แสดงผลการจัดกลุ่มบางขั้นตอนของข้อมูลทดสอบโดยใช้ขั้นตอนวิธีการลำเสียงย้ายค่าเฉลี่ย (ต่อ)

2. ขั้นตอนวิธีการลำเลียงย้ายค่าเฉลี่ยแบบครอบคลุม (Generalized Transport Mean Shift Algorithm)

ผู้วิจัยได้นำเสนอขั้นตอนวิธีในการลำเลียงย้ายค่าเฉลี่ยแบบครอบคลุม (GTMS) ขึ้นโดยได้นำแนวความคิดพื้นฐานขั้นตอนการย้ายค่าเฉลี่ยมาใช้ ซึ่งก็คือการย้ายตำแหน่งอย่างอัตโนมัติของแต่ละจุดไปยังจุดที่มีความหนาแน่นสูงสุด อย่างไรก็ตามในการหาจุดที่มีความหนาแน่นสูงสุด (Mode) ของทุก ๆ จุดของข้อมูล GTMS ใช้ข้อมูลที่นำมาคำนวณเพื่อหาความหนาแน่นสูงสุดจำนวนน้อย โดยจะใช้ผู้ขนส่ง (transporters) ที่เป็นตัวแทนของจุดอื่น ๆ ที่เป็นผู้พ่วง (trailer) ซึ่งในการหาผู้ขนส่งไม่ได้เป็นการคำนวณเพิ่มขึ้น เนื่องจาก GTMS ใช้ระยะห่างของแต่ละจุดซึ่งต้องถูกคำนวณอยู่แล้วในขั้นตอนของการย้ายตำแหน่ง และผู้พ่วงคือตำแหน่งที่จะถูกย้ายไปยังความหนาแน่นสูงสุดเดียวกันกับผู้ขนส่ง ความสัมพันธ์นี้เรียกว่า “ผู้ขนส่ง - ผู้พ่วง” โดยมี การพิจารณาทิศทางการย้ายตำแหน่งของผู้พ่วงและทิศทางเดินของเส้นทางของผู้ขนส่ง กำหนดโดยเวกเตอร์ทิศทางเดินเริ่มต้น การพิจารณาทิศทางของผู้ขนส่ง และผู้พ่วง โดยค่าที่ใช้ตัดสินใจความขนานกันของทิศทางทั้งสอง ข้อมูลมีความยืดหยุ่นในการกำหนดค่า โดยสามารถมีค่าเปลี่ยนแปลงตามความเหมาะสมของธรรมชาติของขั้นตอนวิธีการย้ายค่าเฉลี่ยเมื่อจำนวนรอบมากขึ้น ต่อจากนั้นผู้พ่วงก็จะไม่ถูกนำมาคำนวณในรอบต่อไป ซึ่งจะเหลือเพียงผู้ขนส่งเท่านั้นที่ถูกนำมาคำนวณ ยิ่งไปกว่านั้นผู้ขนส่งในรอบปัจจุบันอาจจะกลายเป็นผู้พ่วงในรอบต่อไป ซึ่งจะทำให้ในแต่ละรอบจะมีจำนวนผู้ขนส่งที่ต้องถูกนำมาคำนวณน้อยลง ซึ่งหมายความว่าผู้ขนส่งบางตัวมีจำนวนผู้พ่วงเพิ่มขึ้นนั่นเอง วิธีการนี้เสมือนกับเป็นการจัดกลุ่มข้อมูลไปพร้อม ๆ กับการย้ายตำแหน่งเพราะว่าผู้พ่วงทั้งหมดที่พ่วงไปกับผู้ขนส่งเดียวกันเปรียบเสมือนข้อมูลนี้ควรจะอยู่ในกลุ่มเดียวกันนั่นเอง และเมื่อสิ้นสุดขั้นตอนวิธีการแล้ว ในการจัดกลุ่มข้อมูลจะพิจารณาผู้ขนส่งที่อยู่ใกล้กัน เพื่อพิจารณาว่ามีโอกาสเป็นกลุ่มเดียวกันหรือไม่ เนื่องจากจุดที่มีความหนาแน่นสูงสุดของข้อมูลมีโอกาสที่จะมีหลายผู้ขนส่งได้ในรอบสุดท้าย



ภาพที่ 26 แสดงการย้ายตำแหน่งของ x_i ในรอบที่ r

จากภาพที่ 26 จะเห็นว่า การย้ายตำแหน่งของตำแหน่ง i ในรอบที่ r นี้ จะย้ายมาใกล้ตำแหน่ง j ซึ่งหากเราให้ j เป็นผู้ขนส่งสุดท้ายแล้ว i จะถูกระบุให้มีตำแหน่งเดียวกับ j ซึ่งไม่ใช่จุดยอดที่ตำแหน่งสุดท้ายของ i ควรจะอยู่ หากพิจารณาทิศทางของเส้นทางของตำแหน่งที่ i ในรอบปัจจุบันเทียบกับในรอบก่อนหน้า จะเห็นว่ามีความ

แตกต่างกับทิศของเส้นทางของจุด j ในตอนเริ่มต้น ดังนั้นก่อนที่จะระบุให้ j เป็นตัวแทนจึงมีการตรวจสอบทิศทางดังที่กล่าวนี้ ในการตรวจสอบที่นำเสนอในขั้นตอนวิธีนี้จะใช้ข้อมูลที่จำเป็นจะต้องคำนวณอยู่แล้วในขั้นตอนวิธีการย้ายค่าเฉลี่ยเดิม ทำให้ไม่เสียเวลาในการคำนวณใหม่ ในแต่ละรอบ และในแต่ละรอบจะไม่ใช้กราฟในการเชื่อมผู้ขนส่งและผู้พ่วง เพื่อลดเวลาและลดการใช้พื้นที่ในการเก็บข้อมูลระหว่างการทำงาน

เพื่อแก้ปัญหาดังที่กล่าวมาแล้ว ผู้วิจัยจึงได้นำเสนอการใช้เมตริกซ์ 4 เมตริกซ์ โดยที่เมตริกซ์แรก (U) เก็บเวกเตอร์ทิศทางเดินเริ่มต้นของผู้ขนส่งตัวที่ i เมตริกซ์ที่สอง (T) ใช้เก็บผู้ขนส่งของข้อมูลตัวที่ i โดยระบุเป็นดัชนีของผู้ขนส่งนั้น เมตริกซ์ที่สาม (C) เก็บข้อมูลที่แสดงถึงการตัดสินใจแบบตรรกะของข้อมูลตัวที่ i ที่ระบุว่าข้อมูลเข้าสู่ความหนาแน่นสูงสุดแล้วหรือไม่ตามสมการ (6) เมตริกซ์สุดท้าย (A) เก็บข้อมูลที่แสดงถึงการตัดสินใจแบบตรรกะของข้อมูลตัวที่ i ที่ระบุว่าในรอบต่อไปจะต้องคำนวณ i หรือไม่โดยข้อมูลที่จำเป็นจะต้องคำนวณต่อคือข้อมูลที่เข้าสู่ความหนาแน่นสูงสุดแล้ว หรือเป็นข้อมูลที่ได้ระบุว่าเป็นผู้พ่วงแล้ว ซึ่งรายละเอียดของแต่ละเมตริกซ์เป็นดังนี้

ให้ $U \in R^{m \times n}$, $u_i \in U$ เป็นเมตริกซ์ขนาดหนึ่งหน่วยที่ใช้ระบุทิศทางของเมตริกซ์ผู้ขนส่งของข้อมูลตัวที่ i ในรอบแรก สามารถคำนวณได้จากสมการ

$$u_i = \frac{x_i^r - x_i^0}{\|x_i^r - x_i^0\|} \quad (10)$$

ให้ $T \in I^{1 \times n}$ คือเมตริกซ์ผู้ขนส่ง โดยที่ $t_i \in T$ หมายถึง ผู้ขนส่งของข้อมูลตัวที่ i ในรอบปัจจุบัน ดังสมการ

$$t_i = \begin{cases} \arg \min_j (\|x_i^r - x_j^r\|^2) & \text{if } \delta_{ij}^r \leq \alpha \\ i & \text{otherwise} \end{cases} \quad (11)$$

ค่า δ_{ij}^r เป็นมุมที่ครอบคลุม ระหว่างเมตริกซ์ขนาดหนึ่งหน่วยที่ใช้ระบุทิศทางของเมตริกซ์ผู้ขนส่งของข้อมูลตัวที่ i (u_i) และ เวกเตอร์ของการย้าย (v_i^r) ค่า $\delta \subset [0 \ 1]$ ซึ่งสามารถพิจารณาค่าได้จากสมการต่อไปนี้

$$\delta_{ij}^r = \frac{1}{2} \left(1 - \frac{v_i^r \cdot u_j}{|v_i^r| \cdot |u_j|} \right)$$

โดย

$$\frac{v_i^r \cdot u_j}{|v_i^r| \cdot |u_j|} = \cos \theta$$

ดังนั้น

$$\delta_{ij}^r = \frac{1}{2} (1 - \cos \theta) \quad (12)$$

เนื่องมาจาก $\cos\theta$ มีค่าอยู่ในช่วง 1 ถึง -1 คือ ทำมุมตั้งแต่ 0 ถึง 180 องศา โดยที่ค่า -1 คือเวกเตอร์ทำมุมกัน 180 องศา ส่วนค่า 1 คือเวกเตอร์ทำมุมกัน 0 องศา ทำให้ค่า $\delta = 0$ หมายถึง เวกเตอร์ทั้งสองขนานกันและมีทิศทางเดียวกัน และหาก $\delta = 1$ หมายถึงเวกเตอร์ทั้งสองขนานกันแต่มีทิศทางตรงกันข้าม

สองเมตริกซ์สุดท้ายคือ $C_{1 \times n}$, $c_i \in C$ และ $A_{1 \times n}$, $a_i \in A$ ระบุได้ดังนี้

$$c_i = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ data convert} \\ 0 & \text{otherwise} \end{cases} \quad (13)$$



$$a_i = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ data should be present} \\ & \text{to the next iteration} \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

จากขั้นตอนวิธีการลำเลียงย้ายค่าเฉลี่ยแบบคlobคลุม (GTMS) แสดงได้ดัง Algorithm1 (GTMS) ในขั้นตอนที่ 1-4 นั้นเป็นขั้นตอนวิธีการย้ายตามค่าเฉลี่ยแบบเดิม ยกเว้นขั้นตอนที่ 2 เป็นการเพิ่มการคำนวณเวกเตอร์ทิศทางเดินเริ่มต้นของการย้ายตำแหน่งในรอบแรกของทุกจุดข้อมูลเก็บไว้ในเมตริกซ์ U ซึ่งการคำนวณนี้เป็นการใช้ระยะทางที่ต้องคำนวณอยู่แล้วในขั้นตอนวิธีการย้ายตามค่าเฉลี่ยเดิมดังสมการ (10) ในขั้นตอนที่ 3 เป็นการพิจารณาระยะทางที่ทำการคำนวณในขั้นตอนที่ 2 ว่ามีระยะน้อยกว่า threshold หรือไม่เพื่อพิจารณาการหยุดการคำนวณโดยเก็บไว้ในเมตริกซ์ C จากนั้นขั้นตอนที่ 5 เป็นการเก็บ $U=V$ และในขั้นตอนที่ 6, 9 และ 10 เป็นขั้นตอนที่ต้องดำเนินการอยู่แล้วในขั้นตอนวิธีการย้ายตามค่าเฉลี่ยแบบเดิม ส่วนขั้นตอนที่ 7 เป็นการหาระยะทางใกล้สุดของจุดที่ย้ายมาเทียบกับจุดเดิม และขั้นตอนที่ 8 เป็นการพิจารณาผู้ขนส่งซึ่งเป็นไปตามการกำหนดค่า α โดยถ้าหากผู้ขนส่งถูกพบในขั้นที่ 8 ก็ลดการคำนวณในขั้นตอนที่ 9 อีก ทำให้ไม่ต้องคำนวณ x_i ในรอบที่ T ซึ่งเป็นการหาค่าเฉลี่ยโดยกำหนดน้ำหนักของทุก ๆ จุดใน X ตามสมการ (5) โดยทั่วไปแล้วนิยมใช้ฟังก์ชันเคอร์เนลเป็นฟังก์ชันเกาส์เซียนในการกำหนดน้ำหนัก ซึ่งเป็นการคำนวณฟังก์ชันยกกำลังของระยะห่างระหว่างทุก ๆ ตำแหน่งเมื่อเทียบกับตำแหน่งปัจจุบันที่ได้จากขั้นตอนที่ 6 ที่คำนวณตามสมการ (6) จะเห็นได้ว่าถ้ามีการพบผู้ขนส่งจำนวนมากก็จะเหลือจุดพิจารณาน้อยสำหรับการคำนวณระยะห่างต่อในขั้นตอนที่ 6 ของรอบต่อไป รวมทั้งการลดการคำนวณในขั้นตอนที่ 9 ดังที่กล่าวมาแล้ว ดังนั้น GTMS จะลดการทำงานที่ซ้ำซ้อนของขั้นตอนวิธีการย้ายตามค่าเฉลี่ยแบบเดิมเป็นอย่างมากซึ่งจะแสดงให้เห็นในผลการทดลองการจัดกลุ่มข้อมูล และการแยกส่วนประกอบของภาพต่อไป

Algorithm1 (GTMS)

Initialization:

C is initialized to false.

A is initialized to true.

Initialize the value of parameter α .

Set $\tau = 1$

for each $x_i \in X$ **do**

1. Compute x_i^τ

2. Calculate the shift distance and the trajectory vector

Set $z = x_i^\tau - x_i^{\tau-1}$

Set $s = \|z\|^2$

Set $v_i = z/s^{0.5}$

3. Considering the convergence

if $s \leq \text{threshold}$ **then**

$c_i = \text{true}$, $a_i = \text{false}$

end if

4. Set i^{th} to be a transporter itself, $t_i = i$

end for

5. $U=V$ // u_i is a unit trajectory vector

while these are $a_i = \text{true}$ **do**

6. Calculate the distance (d_k) from x_i^τ to $x_k \in X, 1 \leq k \leq n$

7. Find the j^{th} data that nearest to the i^{th} data

$j = \text{argmin}_k d_k$

8. Investigate the transporter

Compute $\delta = \frac{1}{2}(1 - v_i^\tau \cdot u_j / \|v_i^\tau\| \|u_j\|)$

if $i \neq j$ & $t_i \neq i$ & $\delta \leq \alpha$ **then**

set $t_i = j$, $a_i = \text{false}$ // assign transporter

and inactive the trailer

else

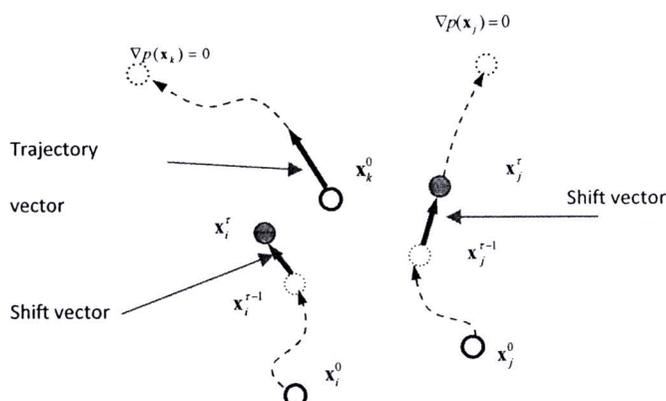
9. Compute $x_i^{\tau+1}$

10. Follow step 2 and 3

end if

Increase the number of iteration by setting $\tau = \tau + 1$.

endwhile



ภาพที่ 27 ทิศทางเดินของเส้นทาง x_k และทิศทางการย้ายของ x_i และ x_j ในรอบที่ τ

เพื่อให้เห็นภาพชัดเจนในการพิจารณาทิศทางเส้นทางเดินของผู้ขนส่ง และผู้พ่วง โดยค่าที่ใช้ตัดสินใจความขนานกันของทิศทางทั้งสองคือ α นั้นมีแนวความคิดจากทั่วไปแล้วมีจุดจำนวนมากที่จะถูกย้ายไปในเส้นทางเดียวกันไปยังจุดยอดของตัวเอง ดังตัวอย่างที่เห็นในภาพที่ 18 อย่างไรก็ตามเมื่อพิจารณาการย้ายตำแหน่งของจุดในภาพที่ 27 ในรอบที่ τ ข้อมูลตัวที่ i ถูกย้ายไปที่ตำแหน่งซึ่งใกล้กับตำแหน่งเริ่มต้นของข้อมูลตัวที่ k ซึ่งทิศทางกรย้ายของข้อมูลตัวที่ i ขนานกับทิศทางเส้นทางเดินเริ่มต้นของข้อมูลตัวที่ k ดังนั้นข้อมูลตัวที่ i จะถูกพิจารณาเสมือนกับเป็นผู้พ่วงของข้อมูลตัวที่ k ซึ่งข้อมูลตัวที่ k สมมุติให้เป็นผู้ขนส่งของข้อมูลตัวที่ i ดังนั้นในรอบต่อไปก็ไม่จำเป็นต้องคำนวณการย้ายตำแหน่งของตัวที่ i อีก และในการพิจารณาข้อมูลตัวที่ j แม้ว่าการย้ายตำแหน่งของข้อมูลตัวที่ j ในรอบที่ τ นี้ถูกย้ายไปใกล้กับตำแหน่งเริ่มต้นของข้อมูลตัวที่ k ด้วยเหมือนกัน แต่เมื่อพิจารณาทิศทางเส้นทางเดินของข้อมูลตัวที่ j กับ ทิศทางเส้นทางเดินเริ่มต้นของข้อมูลตัวที่ k แล้วมีทิศทางที่แตกต่างกัน ซึ่งในความเป็นจริงแล้ว จุดยอดของ j อยู่คนละตำแหน่งกับจุดยอดของ k ซึ่ง GTMS สามารถช่วยพิจารณาจุดที่มีลักษณะเช่นนี้ได้

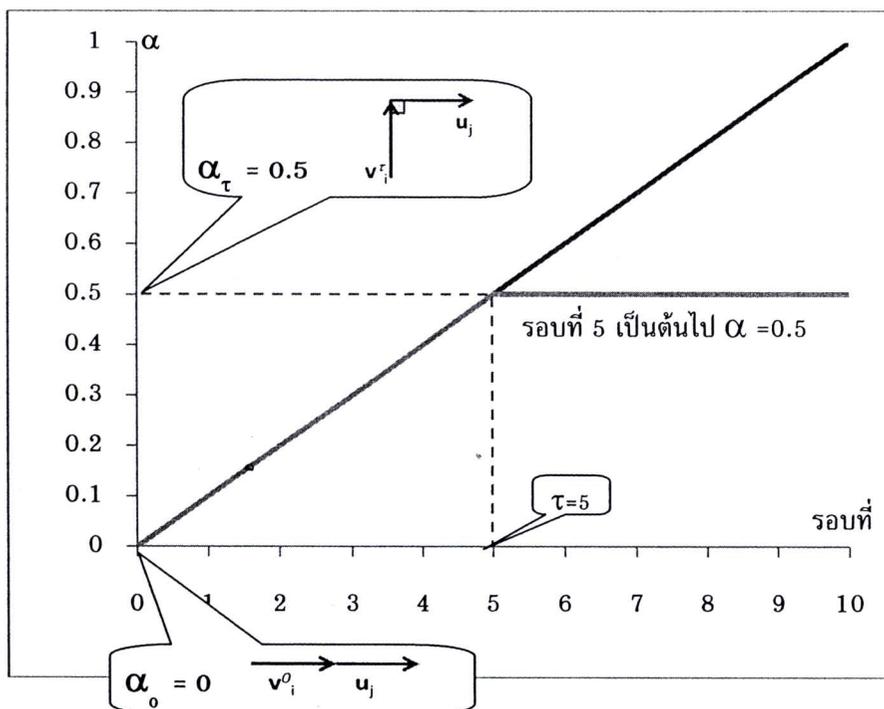
แม้ว่า GTMS จะมีตัวแปรที่ต้องกำหนดเพิ่มคือ α ในขั้นตอนที่ 8 ที่แสดงไว้ใน Algorithm 1 (GTMS) ซึ่งเป็นการพิจารณาผู้ขนส่ง แต่ตัวแปรนี้ทำให้ขั้นตอนวิธีมีความยืดหยุ่นมากขึ้น ทำให้ GTMS สามารถทำงานเหมือนกับ MS รวมทั้ง TMS ได้ ซึ่งง่ายต่อการเลือกค่าที่เหมาะสมกับการนำไปใช้งานเพื่อให้เหมาะกับงานแต่ละด้าน โดย α จะถูกใช้ในการตัดสินใจเลือกผู้ขนส่งที่มีทิศทางขนานกับทิศทางของการย้ายของข้อมูลที่พิจารณาในรอบปัจจุบัน โดย ถ้า $\alpha < 0$ หมายความว่า GTMS จะให้ผลเหมือนกับ MS ทุกประการเพราะว่าจะไม่มีการทำงานตามขั้นตอนที่ 8 โดยจะไม่มีการระบุผู้ขนส่งให้ผู้พ่วง เงื่อนไขในการหยุดมีเพียงขั้นตอนตาม MS เดิม แล้วถ้ากำหนดค่า $\alpha > 1$ ทุกจุดที่พิจารณาจะมีการทำงานในขั้นตอนที่ 8 เสมอ คือในกรณีนี้ตัวที่อยู่ใกล้ข้อมูลที่ย้ายมาในรอบปัจจุบันจะถูกระบุเป็นผู้ขนส่งเสมอ เปรียบเสมือนเป็นการทำงานแบบ TMS แล้วถ้า $0 \leq \alpha \leq 1$ ซึ่งเป็นค่าที่ตรงกับช่วง δ ที่สามารถพิจารณาความขนานกันของผู้ขนส่งและผู้พ่วงดังที่กล่าวมาแล้วข้างต้น อย่างไรก็ตามถ้าเราพิจารณาธรรมชาติของขั้นตอนวิธีกรย้ายตามค่าเฉลี่ยแบบเดิม ค่า α นี้ควรจะมีความเพิ่มขึ้นเมื่อจำนวนรอบเพิ่มขึ้น เนื่องจากเมื่อจำนวนรอบเพิ่มขึ้นข้อมูลจะถูกย้ายไปยังตำแหน่งที่ใกล้กับจุดยอด ซึ่งจะอยู่ไกลกว่าข้อมูลอื่นที่ไม่ได้อยู่ภายใต้จุดยอดเดียวกัน ในรอบที่เพิ่มขึ้นเหล่านี้ผู้ขนส่งมีโอกาสสูงที่จะอยู่

ภายใต้จุดยอดเดียวกันกับข้อมูลที่เหลือ ดังนั้นควรที่จะผ่อนคลายเงื่อนไขของการขนานกัน ในงานวิจัยนี้ได้เสนอวิธีการผ่อนคลายเงื่อนไขของการขนานกันของคุณสมบัตินี้ดังนี้

กำหนดให้ α_0 เป็นค่าความขนานเริ่มต้น ในรอบแรก α_τ เป็นค่าความขนานที่มากที่สุดที่จะยอมรับได้ในรอบที่ τ โดย α สามารถกำหนดได้ดังนี้

$$\alpha = \min(\alpha_\tau, \alpha_0 + (\frac{\alpha_\tau - \alpha_0}{\tau})i) \quad (15)$$

เมื่อ i เป็นจำนวนรอบ, α เป็นสมการเชิงเส้นที่เพิ่มขึ้นจาก α_0 จนถึง α_τ เมื่อขั้นตอนวิธีเพิ่มขึ้นจนถึงรอบที่ τ ตัวอย่างการกำหนดค่า α ($\alpha_0 = 0$, $\alpha_\tau = 0.5$, $\tau = 5$) มีความหมายว่า เริ่มต้นยอมรับเฉพาะผู้ขนส่งที่ขนานกันกับผู้พ่วง ($\alpha_0 = 0$) และมีการผ่อนคลายข้อจำกัดโดยยอมรับเพิ่มขึ้นเรื่อย ๆ จน $\alpha = 0.5$ ในรอบที่ 5 ($\alpha_\tau = 0.5$, $\tau = 5$ คือยอมรับได้เพียงเวกเตอร์ทั้งสองที่มีทิศทางตั้งฉากกัน) หลังจากรอบที่ 5 เป็นต้นไปค่า α จะถูกกำหนดให้ 0.5 แสดงตัวอย่างนี้ดังภาพด้านล่าง



ภาพที่ 28 ตัวอย่างการกำหนดค่า α ($\alpha_0 = 0$, $\alpha_\tau = 0.5$, $\tau = 5$) ตามสมการที่ (15)

3. การวิเคราะห์ประสิทธิภาพของขั้นตอนวิธี

ในการประมวลผลสิ่งที่สำคัญที่สุดที่ต้องคำนึงถึง คือ ผลลัพธ์ของการประมวลผลที่จะต้องมีความถูกต้องแม่นยำ แก้ปัญหาได้ตรงกับจุดประสงค์ที่ต้องการ ถ้าหากผลลัพธ์ที่ได้ไม่ถูกต้องแล้ว ขั้นตอนวิธีนั้นก็ไม่มีประโยชน์ ลักษณะที่ดีของขั้นตอนวิธี คือใช้เวลาในการประมวลผลไม่นานจนเกินไปเพราะฉะนั้นจึงต้องออกแบบขั้นตอนวิธีให้มีประสิทธิภาพมากที่สุด [24] ซึ่งประสิทธิภาพของขั้นตอนวิธีจะถูกพิจารณา 2 ส่วนคือ หน่วยความจำ (Memory) ที่จะต้องใช้ในการประมวลผล และเวลา (Time) ที่จะต้องใช้ในการประมวลผล ส่งผลให้การวิเคราะห์ประสิทธิภาพของขั้นตอนวิธีแบ่งออกได้เป็น 2 ส่วน คือ ส่วนแรกเป็นการวิเคราะห์หน่วยความจำที่ต้องใช้ในการประมวลผล (Space Complexity) และ ส่วนที่สองเป็นการวิเคราะห์เวลาที่ต้องใช้ในการประมวลผล (Time Complexity) ผู้วิจัยได้ทำการวิเคราะห์ประสิทธิภาพของขั้นตอนวิธีการย้ายตามค่าเฉลี่ย (MS) กับขั้นตอนวิธีที่เสนอขึ้นใหม่คือ ขั้นตอนวิธีการลำเลียงย้ายตามค่าเฉลี่ย (TMS) และขั้นตอนวิธีการลำเลียงย้ายค่าเฉลี่ยแบบครอบคลุม (GTMS) ดังต่อไปนี้

ตารางที่ 1 การวิเคราะห์ประสิทธิภาพของขั้นตอนวิธีการย้ายตามค่าเฉลี่ย (MS)

| ขั้นตอนวิธีการทำงานของ MS | space complexity | Time complexity |
|--|------------------|-----------------|
| 1. ทหาระยะทาง 1 ตัวกับทุกจุด | n | mn |
| 2. คำนวณสูตร $K(x) = e^{-\beta \ x\ ^2}$ | 0 | n |
| 3. คำนวณหา | | |
| 3.1) $\sum_{x_i=x} K(x_i - x)x_i$ | m | mn |
| 3.2) $\sum_{x_i=x} K(x_i - x)$ | 1 | n |
| 3.3) $m(x) = \frac{\sum_{x_i=x} K(x_i - x)x_i}{\sum_{x_i=x} K(x_i - x)}$ | 0 | m |
| 4. ตรวจสอบว่าหยุดหรือไม่ พิจารณา $m(x) \leq \text{threshold}$ | 1 | m |
| 5. ทำจนครบทุกตัว(ครบ 1 รอบ) | $2nm + n + m$ | $n^2 m$ |

ดังนั้นในการดำเนินการตามขั้นตอนวิธีการย้ายตามค่าเฉลี่ยนี้จะได้การวิเคราะห์หน่วยความจำที่ต้องใช้ในการประมวลผล (Space Complexity) คือ $O(nm)$ และส่วนที่สองเป็นการวิเคราะห์เวลาที่ต้องใช้ในการประมวลผล (Time Complexity) ทำครบ k รอบ คือ $O(kn^2 m)$

ตารางที่ 2 การวิเคราะห์ประสิทธิภาพของขั้นตอนวิธีการลำเลียงย้ายตามค่าเฉลี่ย (TMS)

| ขั้นตอนวิธีการทำงานของ TMS | space complexity | Time complexity |
|--|------------------|--|
| 1. ย้ายตำแหน่งตามขั้นตอนวิธีการย้ายตามค่าเฉลี่ยแบบเดิม | $2nm + n + m$ | $n^2 m$ |
| 2. หาดำแหน่งที่ใกล้ที่สุดระหว่างตำแหน่งที่ย้ายมากับตำแหน่งเริ่มต้นตามกฎเบื้องต้น | 1 | n |
| 3. หาเมตริกซ์ตำแหน่งของตัวแทน (h) ซึ่งเป็นตามกฎแห่งความต่อเนื่องและเชื่อมเอจระหว่างผู้ขนส่งและผู้พ่วงกับกราฟ G | n^2 | n |
| 4. ในจำนวนตัวแทนนี้ หาตัวแทนที่หยุดการคำนวณตามเงื่อนไขอสมการ $\ m(x) - x\ ^2 = < \text{threshold}$ | 1 | m |
| 5. คงข้อมูลเป็นตัวแทน ซึ่งยังไม่หยุดการคำนวณเพื่อใช้คำนวณในรอบต่อไป | 0 | m |
| 6. ทำซ้ำขั้นตอน 1- 5 จนไม่เหลือข้อมูลในการคำนวณ | $n^2 + mn + n$ | $\sum_{i=1}^p Q_i nm$ $, Q_p \leq \dots \leq Q_p \ll Q_1 = N$ |

ดังนั้นในการดำเนินการตามขั้นตอนวิธีการย้ายตามค่าเฉลี่ยนี้จะได้รับการวิเคราะห์หน่วยความจำที่ต้องใช้ในการประมวลผล (Space Complexity) คือ $O(n^2)$ และส่วนที่สองเป็นการวิเคราะห์เวลาที่ต้องใช้ในการประมวลผล (Time Complexity) คือ $O(\sum_{i=1}^p Q_i nm)$ โดยที่ $Q_p \leq \dots \leq Q_p \ll Q_1 = N$

ตารางที่ 3 การวิเคราะห์ประสิทธิภาพของขั้นตอนวิธีการลำเลียงย้ายค่าเฉลี่ยแบบครอบคลุม (GTMS)

| ขั้นตอนวิธีการทำงานของ GTMS | space complexity | Time complexity |
|--|------------------|-----------------|
| 1. คำนวณขั้นตอนวิธีการย้ายตามค่าเฉลี่ยแบบเดิม 1 รอบ | $2nm + n + m$ | $n^2 m$ |
| 2. หาเวกเตอร์ 1 หน่วยของตำแหน่งเริ่มต้นกับตำแหน่งที่ได้ย้ายไป | nm | nm |
| 3. ตรวจสอบว่าหยุดหรือไม่ พิจารณา ตามเงื่อนไขอสมการ $\ m(x) - x\ ^2 = < \text{threshold}$ | 1 | m |

ตารางที่ 3 การวิเคราะห์ประสิทธิภาพของขั้นตอนวิธีการลำเลียงย้ายค่าเฉลี่ยแบบครอบคลุม (GTMS) (ต่อ)

| ขั้นตอนวิธีการทำงานของ GTMS | space complexity | Time complexity |
|---|------------------|--|
| 1. คำนวณขั้นตอนวิธีการย้ายตามค่าเฉลี่ยแบบเดิม 1 รอบ | $2nm + n + m$ | $n^2 m$ |
| 2. หาเวกเตอร์ 1 หน่วยของตำแหน่งเริ่มต้นกับตำแหน่งที่ได้ย้ายไป | nm | nm |
| 3. ตรวจสอบว่าหยุดหรือไม่ พิจารณา ตามเงื่อนไขข้อสมการ $\ m(x) - x\ ^2 = < \text{threshold}$ | 1 | m |
| 4. กำหนดทุกตัวเป็นผู้ขนส่ง | n | n |
| 5. $U = V$ | 0 | nm |
| 6. คำนวณระยะทางใหม่ของจุดที่ยังไม่หยุด | n | n |
| 7. หาดำแหน่งที่ใกล้สุด | 1 | n |
| 8. กำหนด ถ้า $i \neq j$ และผู้พ่วงไม่ใช่ตัวมันและทิศทางที่กำหนด $< \alpha$ แล้วกำหนดให้เป็นผู้ถูกขนส่งแล้วไม่ต้องคำนวณต่อ | 0 | 1 |
| 9. คำนวณตำแหน่งที่ยังไม่หยุดคำนวณ | 0 | nm |
| 10. ทำตามขั้นตอน 2-3 | 0 | m |
| 11. คำนวณจนไม่เหลือตัวคำนวณ | $2nm + 3n + m$ | $\sum_{i=1}^p Q_i nm$ $, Q_p \leq \dots \leq Q_1 \ll Q_1 = N$ |

ดังนั้นในการดำเนินการตามขั้นตอนวิธีการลำเลียงย้ายค่าเฉลี่ยแบบครอบคลุมนี้จะได้รับการวิเคราะห์หน่วยความจำที่ต้องใช้ในการประมวลผล (Space Complexity) คือ $O(nm)$ และส่วนที่สองเป็นการวิเคราะห์เวลาที่ต้องใช้ในการประมวลผล (Time Complexity) คือ $O\left(\sum_{i=1}^p Q_i(nm)\right)$ โดยที่ $Q_p \leq \dots \leq Q_1 \ll Q_1 = N$ (การที่ $\text{space} = 0$ คือ ไม่ได้ใช้พื้นที่ในการจัดเก็บ หรือ เก็บไว้ในตำแหน่งเดิม)

เมื่อเปรียบเทียบการวิเคราะห์หน่วยความจำที่ต้องใช้ในการประมวลผล (Space Complexity) และการวิเคราะห์เวลาที่ต้องใช้ในการประมวลผล (Time Complexity) ของขั้นตอนวิธีการย้ายค่าเฉลี่ยแบบเดิมกับขั้นตอนวิธีที่ได้ทำการพัฒนาขึ้นมา สรุปได้ดังตารางดังนี้

ตารางที่ 4 เปรียบเทียบการวิเคราะห์หน่วยความจำที่ต้องใช้ในการประมวลผล (Space Complexity) และการวิเคราะห์เวลาที่ต้องใช้ในการประมวลผล (Time Complexity)

| ขั้นตอนวิธี | Space Complexity | Time Complexity |
|--|------------------|--|
| 1. Mean Shift Algorithm (MS) | $O(nm)$ | $O(kn^2 m)$ |
| 2. Transport Mean Shift Algorithm (TMS) | $O(n^2)$ | $O(\sum_{i=1}^p Q_i nm)$ โดยที่ $Q_p \leq \dots \leq Q_p \ll Q_1 = N$ |
| 3. Generalized Transport Mean Shift Algorithm (GTMS) | $O(nm)$ | $O(\sum_{i=1}^p Q_i (nm))$ โดยที่ $Q_p \leq \dots \leq Q_p \ll Q_1 = N$ |

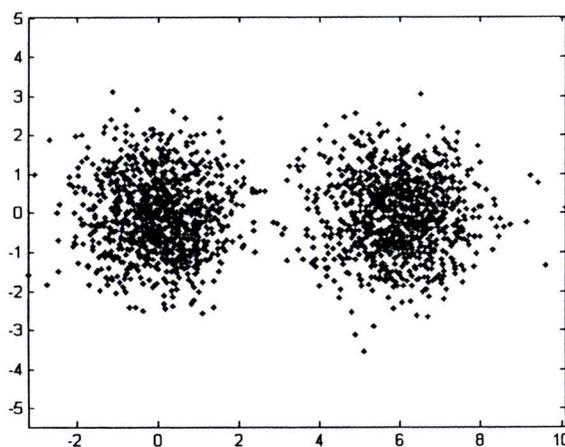
จะเห็นได้ว่า TMS ใช้เวลาในการประมวลผลลดลง แต่ใช้หน่วยความจำมากกว่า เมื่อเทียบกับ MS ส่วน GTMS นั้นใช้เวลาในการประมวลผลน้อยลงซึ่งเท่ากับ TMS และ ใช้หน่วยความจำที่ต้องใช้ในการประมวลผลเท่ากับ MS จะเห็นได้ว่าเมื่อเทียบกันทั้ง 3 ขั้นตอนวิธี GTMS เป็นขั้นตอนวิธีที่ให้ประสิทธิภาพที่ดีทั้งหน่วยความจำ และ เวลาที่ต้องใช้ในการประมวลผล

4. วิธีการดำเนินการวิจัย

ได้ดำเนินการนำวิธีการใหม่ที่น่าเสนอขึ้นมาประยุกต์ใช้กับการจัดกลุ่ม การแยกส่วนประกอบภาพทั่วไป และการแยกส่วนประกอบภาพทางการแพทย์ ซึ่งมีรายละเอียดดังต่อไปนี้

4.1 การเปรียบเทียบเวลาการทำงานของแต่ละขั้นตอนวิธีเมื่อมีจำนวนข้อมูลเพิ่มขึ้น

ข้อมูลที่ใช้เป็นข้อมูลที่จำลองขึ้นเพื่อทำการเปรียบเทียบเวลาเมื่อข้อมูลมีจำนวนที่แตกต่างกัน โดยข้อมูลชุดนี้จะมีลักษณะอย่างอื่นเหมือนกันทั้งหมดกล่าวคือ เป็นข้อมูล 2 มิติ มี 2 กลุ่มที่แยกกันอย่างชัดเจน



ภาพที่ 29 ข้อมูลทดลองที่สุ่มแบบการกระจายตัวปกติจำนวน 2000 ข้อมูล

เกิดจากการสุ่มแบบการกระจายตัวปกติทั้งสองกลุ่มข้อมูล แต่ละชุดจะมีความต่างกันที่จำนวนข้อมูลเท่านั้น โดยใช้จำนวนข้อมูลแตกต่างกันตั้งแต่ 500, 1000, 1500, 2000, 3000, ..., 10000 ตัวอย่างของข้อมูลที่มีจำนวน 2000 ข้อมูลแสดงในภาพที่ 29 จากนั้นทำการทดลองขั้นตอนวิธีการย้ายค่าเฉลี่ยแบบเดิม และแบบที่นำเสนอขึ้น เพื่อวัดเวลาในการประมวลผลโดยทำการทดลองจำนวน 10 รอบแล้วหาค่าเฉลี่ย

4.2 การประยุกต์ใช้กับการจัดกลุ่มข้อมูล

ข้อมูลที่ใช้เป็นข้อมูลจริงสามข้อมูล โดยข้อมูลแรก คือ ข้อมูลของดอก iris 3 ประเภทคือ Setosa, Versicolour และ Virginica แสดงดังภาพที่ 30 มีประเภทละ 50 ข้อมูลรวมเป็น 150 ข้อมูล ลักษณะของข้อมูลจะมีประเภทหนึ่งที่ยกออกจากอีก 2 ประเภทอย่างชัดเจน ข้อมูลมี 4 มิติ คือความยาวกลีบเลี้ยง ความกว้างกลีบเลี้ยง ความยาวกลีบดอก และความกว้างกลีบดอก (เซนติเมตร) ตามลำดับ จากนั้นทำการทดลองขั้นตอนวิธีการย้ายค่าเฉลี่ยแบบเดิม และแบบที่นำเสนอขึ้น เพื่อวัดเวลาในการประมวลผลโดยทำการทดลองจำนวน 10 รอบแล้วหาค่าเฉลี่ย



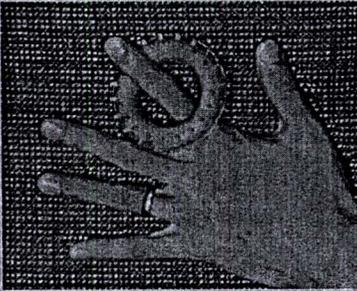
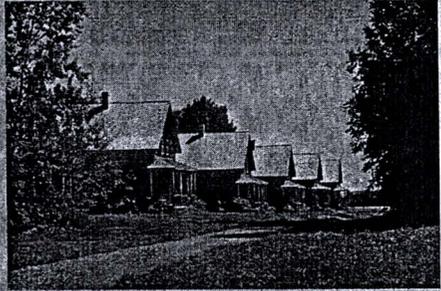
ภาพที่ 30 Iris Setosa, Iris Versicolour และ Iris Virginica ตามลำดับ

ข้อมูลที่สอง คือ ข้อมูลโรคของถั่วเหลือง (Soybean disease dataset) ซึ่งประกอบด้วยข้อมูลทั้งหมด 47 ข้อมูล มี 35 มิติ ซึ่งเป็นข้อมูลที่ถั่วเหลืองมีโรคประกอบด้วย 4 โรค คือ diaporthe stem rot จำนวน 10 ข้อมูล, charcoal rot จำนวน 10 ข้อมูล, rhizoctonia root rot จำนวน 10 ข้อมูล และ phytophthora rot จำนวน 17 ข้อมูล ในงานวิจัยนี้ผู้วิจัยได้ใช้ข้อมูลทั้ง 35 มิติ ในการจัดกลุ่มข้อมูลหลังจากใช้ขั้นตอนวิธีการย้ายค่าเฉลี่ย และแบบที่นำเสนอขึ้น เพื่อวัดเวลาในการประมวลผลโดยทำการทดลองจำนวน 10 รอบแล้วหาค่าเฉลี่ย

ข้อมูลที่สาม คือ ข้อมูลของการบริจาคเลือดที่เมือง Hsin Chu ประเทศไต้หวัน เป็นข้อมูลที่ถูกนำมาใช้ กับ RFM Model ซึ่งเป็นวิธีหนึ่งในการช่วยการวิเคราะห์ข้อมูลทางการตลาด การบันทึกข้อมูลการบริจาคเลือดนี้ได้มาจากข้อมูลที่เกิดจากการที่รถบริจาคเลือดไปรับบริจาคที่มหาวิทยาลัยแห่งหนึ่งในเมือง Hsin Chu ประเทศไต้หวัน เพื่อรับบริจาคเลือดทุก ๆ สามเดือน โดยข้อมูลนี้มีจำนวน 748 ผู้บริจาค แต่ละข้อมูลประกอบไปด้วย 4 มิติ คือ จำนวนเดือนหลังจากที่บริจาคครั้งล่าสุด, จำนวนครั้งที่บริจาคทั้งหมด, จำนวนเลือดที่บริจาคทั้งหมดที่บริจาค (cc) และ จำนวนเดือนหลังจากที่บริจาคครั้งแรก จุดประสงค์เพื่อที่จะทำนายคนที่จะมาบริจาคในเดือนมีนาคม 2007 ผู้วิจัยได้นำข้อมูลนี้มาจัดกลุ่มหลังจากใช้ขั้นตอนวิธีการย้ายค่าเฉลี่ยและแบบที่นำเสนอขึ้น

เพื่อวัดเวลาในการประมวลผลโดยทำการทดลองจำนวน 10 รอบแล้วหาค่าเฉลี่ย เพื่อนำไปวัดค่าความถูกต้องกับกลุ่มของข้อมูลของผู้บริจาคในเดือนนั้นจริง

ตารางที่ 5 รายละเอียดของภาพที่ใช้ในการทดลองการประยุกต์ใช้กับการแยกส่วนประกอบภาพทั่วไป

| | |
|--|---|
|  <p>Cameraman (ระดับสีเทา) ขนาด 50x50 pixel</p> |  <p>Lena (ภาพสี) ขนาด 50x50 pixel</p> |
|  <p>Hand (ภาพสี) ขนาด 62x50 pixel</p> |  <p>Smallbowelloop (ระดับสีเทา) ขนาด 65x50 pixel</p> |
|  <p>Hawk (ภาพสี) ขนาด 75x50 pixel</p> |  <p>Houses (ภาพสี) ขนาด 75x50 pixel</p> |

4.3 การประยุกต์ใช้กับการแยกส่วนประกอบภาพทั่วไป

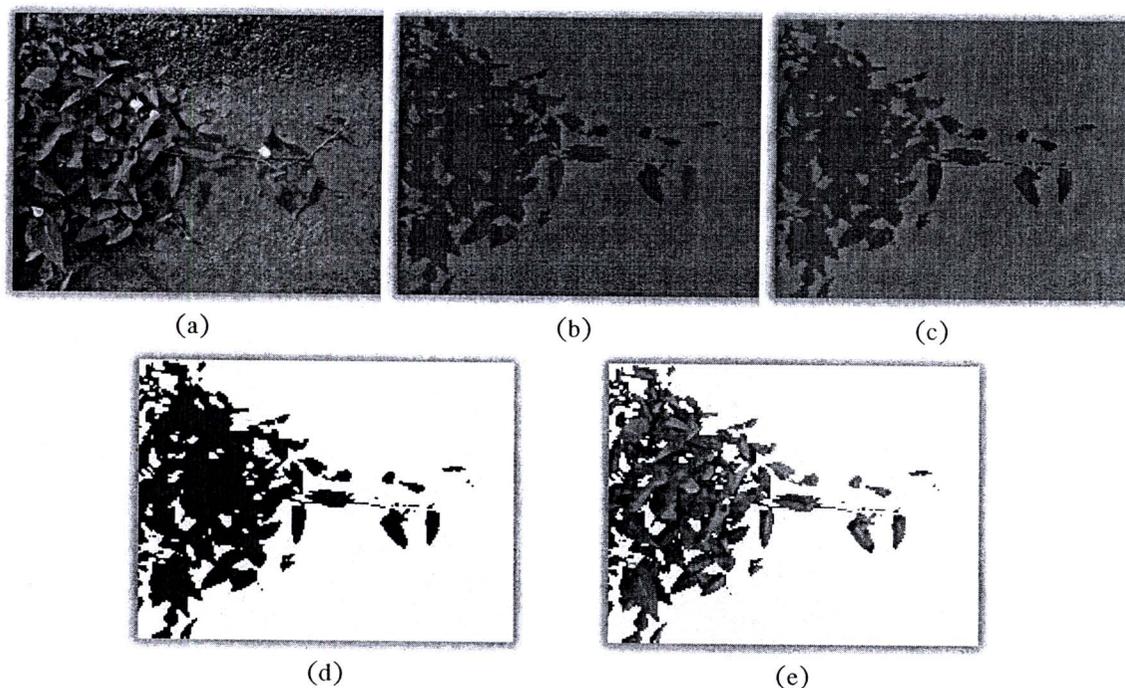
ผู้ทดลองได้ทำการทดลองแยกส่วนประกอบของภาพโดยใช้ภาพที่นิยมใช้ในการเปรียบเทียบผลของการแยกส่วนประกอบของภาพคือ Cameraman, Hand และ Lena และภาพทางการแพทย์คือภาพ Smallbowelloop ซึ่งเป็นภาพ CT scan ของการหดของลำไส้เล็ก รวมทั้งภาพ Hawk และ Houses จาก Berkeley Segmentation Dataset and Benchmark โดยได้แปลงภาพออกมาเป็นเวกเตอร์ขนาด 5 มิติ ประกอบด้วยตำแหน่งในแนวแกน x ตำแหน่งในแนวแกน y ค่าความเข้มของสีแดง ค่าความเข้มของสีเขียว และค่าความเข้มของสีน้ำเงิน โดยแทนด้วย (x, y, R, G, B) ตามลำดับ ในการเปรียบเทียบผลการทดลองของภาพทั้งหมดเนื่องจากขนาดของภาพไม่เท่ากัน บทความนี้ได้ปรับข้อมูลสองมิติแรกซึ่งเป็นตำแหน่งบนภาพในแนวนอนและแนวตั้งให้มีค่าอยู่ในช่วง $[1,100]$ เหมือนกันทุกภาพ รายละเอียดและคุณสมบัติของภาพแสดงในตารางที่ 5 นอกจากนี้ผู้วิจัยได้นำภาพทางการแพทย์คือภาพ B, D และ F ในตารางที่ 6 โดยได้แปลงภาพออกมาเป็นเวกเตอร์ขนาด 7 มิติตามหัวข้อที่ 4.4 ด้านล่าง

จากนั้นทำการทดลองขั้นตอนวิธีการย้ายค่าเฉลี่ยแบบเดิม และแบบที่นำเสนอขึ้น เพื่อเปรียบเทียบเวลา ค่าความคลาดเคลื่อน พร้อมทั้งวัดจำนวนเท่าของการเพิ่มความเร็ว และแสดงจำนวนข้อมูลที่เหลืออยู่ใน 5 รอบแรกของ TMS และ GTMS

และได้ทำการพิจารณาลักษณะของข้อมูลที่ใช้ในการทดลองโดยได้ใช้ทฤษฎีการวิเคราะห์องค์ประกอบหลัก (PCA) มาใช้ลดมิติข้อมูล เพื่อพิจารณาลักษณะการกระจายตัวความหนาแน่นของข้อมูล (probability density distribution) ว่าขั้นตอนวิธีที่ใช้นั้นเหมาะสมกับข้อมูลลักษณะแบบใด

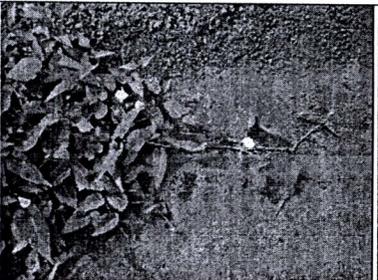
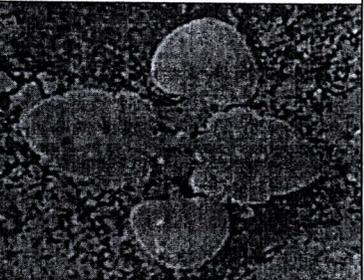
4.4 การประยุกต์ใช้ขั้นตอนวิธีการลำเลียงย้ายตามค่าเฉลี่ยแบบครอบคลุมกับการแยกส่วนประกอบภาพทางการแพทย์

นำขั้นตอนวิธี GTMS มาใช้ในงานแยกส่วนประกอบภาพระหว่างดินกับพืชสีเขียวจากแปลงพืชผลทางการแพทย์ตามบทความ [12] โดยได้แปลงภาพออกมาเป็นเวกเตอร์ขนาด 7 มิติ ประกอบด้วยตำแหน่งในแนวแกน x ตำแหน่งในแนวแกน y ค่าความเข้มของ G-R, ค่าความเข้มของ G-B, ค่าสีในระบบ HSV อีก 3 มิติ โดยแทนด้วย $(x, y, G-R, G-B, H, S, V)$ ตามลำดับ และได้ปรับข้อมูลสองมิติแรกซึ่งเป็นตำแหน่งบนภาพในแนวนอนและแนวตั้งให้มีค่าอยู่ในช่วง $[1,20]$ ในขั้นตอนของการแยกส่วนประกอบภาพหลังจากผ่านขั้นตอน GTMS มาแล้ว เนื่องจากข้อมูลต้องการแยกข้อมูลออกเป็น 2 ส่วนที่ต่างกันอย่างชัดเจนคือ ข้อมูลดินและข้อมูลพืช ดังนั้นผู้วิจัยจึงพัฒนาวิธีการแยกข้อมูลของ 2 ส่วนที่กล่าวมานี้โดยใช้การแปลงค่าสีในระบบ HSV จากตำแหน่งสุดท้ายที่ได้จากขั้นตอนวิธี GTMS เปลี่ยนไปเป็นค่าสีในระบบ RGB ทำให้ตำแหน่งที่อยู่ในกลุ่มเดียวกันจะถูกแทนด้วยสีเดียวกัน จากนั้นก็ทำการเปลี่ยนภาพให้เป็นภาพระดับสีเทา เมื่อถึงขั้นตอนนี้แล้วข้อมูลทั้ง 2 ส่วนจะมีสีที่แยกต่างกันอย่างชัดเจน ซึ่งสามารถใช้ค่าขีดแบ่ง (Threshold) ในการแยกข้อมูลออกจากกันได้อย่างชัดเจน ดังแสดงในภาพที่ 31 และรายละเอียดและคุณสมบัติของภาพแสดงในตารางที่ 6



ภาพที่ 31 (a) ภาพต้นฉบับ (b) ภาพที่ทำการแปลง HSV กลับมาเป็น RGB หลังจากใช้ GTMS (c) ภาพระดับสีเทาของภาพ (b) (d) ข้อมูลทั้ง 2 ส่วนหลังจากใช้ขีดแบ่ง (Threshold) ในการแยกข้อมูลออกจากกัน (e) ข้อมูลของพืช หลังจากรวมภาพ (d) กับภาพต้นฉบับ

ตารางที่ 6 คุณสมบัติของภาพพืชผลทางการเกษตร

| | | |
|---|---|--|
|  |  |  |
| A. ภาพสี ขนาด 200x150 pixel | B. ภาพสี ขนาด 200x150 pixel | C. ภาพสี ขนาด 200x150 pixel |
|  |  |  |
| D. ภาพสี ขนาด 200x150 pixel | E. ภาพสี ขนาด 225x300 pixel | F. ภาพสี ขนาด 196x175 pixel |