

บทที่ 5

สรุปผลการวิจัย และข้อเสนอแนะ

5.1 สรุปผลการวิจัย

การนำเอสวีเอ็มซึ่งได้รับการออกแบบมาให้เป็นตัวจำแนกแบบสองกลุ่มไปใช้กับปัญหาแบบหลายกลุ่มต้องการขั้นตอนวิธีเพิ่มขึ้น การจำแนกแบบหลายกลุ่มโดยวิธีดีดีเอจีเป็นการจำแนกที่เร็วกว่าวิธีเดิมคือแมกซวิน แต่พบว่ายังไม่มีประสิทธิภาพสูงพอ วิธีการนี้จะทำให้การจำแนกมีความผิดพลาดสะสมโดยเฉลี่ยสูง และผลการจำแนกยังขึ้นกับลำดับของสมาชิกในลิสต์ด้วย งานวิจัยฉบับนี้ต้องการหาและวิเคราะห์ขั้นตอนวิธีจำแนกแบบหลายกลุ่ม เพื่อเพิ่มความถูกต้องโดยลดผลจากการเกิดความผิดพลาดสะสม และเพิ่มความเชื่อถือได้ของระบบโดยลดการขึ้นกับลำดับของสมาชิกในลิสต์

งานวิจัยนี้ได้เสนอวิธีการจำแนกแบบหลายกลุ่มที่เรียกว่าเอดีเอจี ด้วยการใช้โครงสร้างกราฟไม่วงแบบมีทิศทาง ที่ปรับโครงสร้างได้เมื่อกระบวนการจำแนกคืบหน้าไป โดยจะลดระดับความลึกของกราฟลง ซึ่งทำให้ความผิดพลาดสะสมลดลง นอกจากนี้ยังลดการขึ้นกับลำดับของสมาชิกในลิสต์ โดยทำให้การแจกแจงความน่าจะเป็นของการจำแนกมีการแจกแจงแบบยูนิฟอร์มมากขึ้น

ผลการทดลองกับชุดข้อมูล 5 ชุดข้อมูล และแปรค่าพารามิเตอร์ในแต่ละชุดข้อมูลอย่างน้อย 10 ค่า ทำการทดลองซ้ำ 120 20,000 50,000 ครั้งตามจำนวนกลุ่ม และเปรียบเทียบกันด้วยค่าเฉลี่ยของความถูกต้อง ระหว่างงานวิจัยนี้กับดีดีเอจี ผลที่ได้คือวิธีที่นำเสนอได้ผลดีกว่า 3 ชุดข้อมูล ไม่แตกต่างกัน 2 ชุดข้อมูล นอกจากนี้ยังพบว่าเมื่อจำนวนกลุ่มเพิ่มขึ้น จะสังเกตเห็นความแตกต่างของความถูกต้องของการจำแนกมากขึ้น หรืออีกนัยหนึ่งคือเอดีเอจีเหมาะกับปัญหาแบบหลายกลุ่มมากกว่าดีดีเอจี ในขณะที่เมื่อจำนวนกลุ่มน้อยทั้งสองวิธีไม่ได้แตกต่างกันมากนัก

5.2 ข้อเสนอแนะ

1. แนวทางหนึ่งนี้อาจเพิ่มความถูกต้องในการจำแนกโดยขั้นตอนวิธีเอดีเอจี คือการใช้ฮิวริสติกในการเลือกลำดับของสมาชิกในลิสต์ ฮิวริสติกดังกล่าวเป็นฟังก์ชันของมาริจนระหว่างกลุ่มสองกลุ่มในแต่ละตัวจำแนกแบบสองกลุ่ม ซึ่งมาริจนเป็นตัวบ่งชี้ถึงความแตกต่างระหว่างกลุ่มทั้งสองกลุ่ม แนวทางดังกล่าวสามารถทำได้โดยการกำหนดให้เลือกเปรียบเทียบระหว่างคู่ของกลุ่มที่มีความแตกต่างกันมากกว่าก่อนคู่ของกลุ่มที่มีความแตกต่างกันน้อยกว่าในโนนระดับบน เพื่อลดความเสี่ยงในการจำแนกผิดพลาดตั้งแต่ระดับบนลง ซึ่งเป็นการลดค่า p ในสูตร

ของค่าคาดหวังของความถูกต้องของเอดีเอซีตามทฤษฎีบทที่ 2 ลง ทำให้ค่าคาดหวังของความถูกต้องตามสูตร $\left(\frac{2N - 2^{\lceil \log_2 N \rceil}}{N}\right)(1-p)^{\lceil \log_2 N \rceil} + \left(\frac{2^{\lceil \log_2 N \rceil} - N}{N}\right)(1-p)^{\lceil \log_2 N \rceil - 1}$ สูงขึ้น

2. ข้อตกลงพื้นฐานของแบบจำลองกำหนดให้ค่าความผิดพลาดมีค่าคงที่เป็น p ในทุกโนดที่มีการเปรียบเทียบระหว่างกลุ่มที่ไม่ใช่กลุ่มที่ถูกต้อง ซึ่งในปัญหาจริงค่า p จะไม่เท่ากันในแต่ละตัว จำแนกแบบสองกลุ่ม ดังนั้นเพื่อสร้างโมเดลที่ใกล้เคียงความเป็นจริงมากขึ้น โมเดลควรจะรองรับการกำหนดค่า p ให้แตกต่างกันตามตัวจำแนกแบบ 2 กลุ่มได้ โดยกำหนดเป็นค่า p_u แล้วจึงคำนวณหาสูตรค่าคาดหวังของความถูกต้องตามแนวทางที่ได้เสนอไว้ในบทที่ 3
3. ในกรณีที่ไม่สามารถหาฟังก์ชันมาประมาณค่าคาดหวังของความถูกต้องได้ อาจใช้การจำลอง (Simulation) เพื่อหาค่าความถูกต้องจากโมเดลแทน การจำลองนี้จะช่วยตอบคำถามเบื้องต้นได้ว่าเอดีเอซีเหมาะกับปัญหาแบบใดและไม่เหมาะกับปัญหาแบบใด ในลักษณะที่เฉพาะเจาะจงลงไป เช่นปัญหาแบบที่กลุ่มแต่ละกลุ่มมีโอกาสเกิดขึ้นไม่เท่ากัน ดังในปัญหาการรู้จำตัวอักษรที่อักษรประเภทสละอย่าง เอ อี ไอ โอ หรือ ยู มีโอกาสเกิดขึ้นมากกว่าตัวอักษรประเภทอื่น นอกจากนี้การจำลองยังอาจช่วยให้ค้นพบความรู้ในการเลือกลำดับของสมาชิกในลิสต์ เพื่อให้เหมาะสมกับปัญหาแต่ละแบบได้ อันจะช่วยเพิ่มความถูกต้องในการจำแนก
4. แนวทางการเปรียบเทียบการขึ้นกับลำดับของสมาชิกในลิสต์สามารถทำได้โดยพิจารณาว่ามีรูปแบบของลิสต์ที่เป็นไปได้มากที่สุดที่รูปแบบที่สามารถทำให้ผลการจำแนกแตกต่างกันได้ โดยในเบื้องต้นเราทราบว่าในลิสต์แบบดีดีเอซี การสลับสมาชิกในลิสต์แบบย้อนลำดับ เช่น สลับ 1234 เป็น 4321 จะไม่ทำให้ผลการจำแนกแตกต่างจากเดิม ซึ่งทำให้ลิสต์ที่สามารถให้ผลการจำแนกที่แตกต่างกันเหลือเพียงครึ่งหนึ่งของจำนวนรูปแบบลิสต์ที่เป็นไปได้ทั้งหมด แต่สำหรับในแบบเอดีเอซีพบว่าการสลับลำดับในลิสต์ที่ไม่ทำให้ผลการจำแนกเปลี่ยนอาจมีมากกว่า 2 แบบ และจำนวนวิธีการสลับนี้ขึ้นกับจำนวนกลุ่ม เช่น ในปัญหาแบบ 4 กลุ่ม ลิสต์แต่ละแบบใน 8 แบบต่อไปนี้ 1234 4231 1324 4321 2143 3142 2413 3412 จะให้ผลการจำแนกที่เหมือนกันเสมอ การคำนวณนี้สามารถใช้เป็นแนวทางในการหาโครงสร้างที่สามารถลดการขึ้นอยู่กัลำดับของลิสต์ลงได้อีก ซึ่งจะทำให้ผลการจำแนกมีความเชื่อถือได้สูงขึ้นกว่าวิธีเอดีเอซี