

บทที่ 5

บทสรุป

Conclusion

โครงการวิจัยนี้ได้นำเสนออัลกอริทึม Shape-based Streaming Subsequence Time Series Clustering (3STSC) ซึ่งสามารถให้ผลลัพธ์การจัดกลุ่มกระแสข้อมูลอนุกรมเวลาที่มีจุดข้อมูลใหม่เข้ามาเรื่อย ๆ ได้ในเวลาจริง โดยอัลกอริทึม 3STSC นี้เป็นการต่อยอดมาจากอัลกอริทึม Shape-based Subsequence Time Series Clustering (2STSC) ที่ได้มีการแสดงให้เห็นแล้วว่าสามารถให้ผลลัพธ์การจัดกลุ่มที่มีความหมายในเชิงรูปร่าง ตามมาตรวัด Shape-based Meaningfulness Measurement (SMM) ทั้งนี้อัลกอริทึม 2STSC ได้ใช้มาตรวัดระยะทางแบบไดนามิกไทม์วอร์ปิง (DTW) และฟังก์ชันการเฉลี่ยตามรูป ซึ่งเป็นสองส่วนสำคัญที่ทำให้อัลกอริทึมสามารถให้ผลการจัดกลุ่มที่มีความหมายได้อย่างเหมาะสม เนื่องจากคุณสมบัติของการวัดระยะแบบไดนามิกไทม์วอร์ปิง และฟังก์ชันการเฉลี่ยตามรูปนี้ สามารถรับมือกับลำดับย่อยติด ๆ กันที่เป็น Trivial Match ซึ่งมีความแตกต่างกันในเชิงเวลาเพียงเล็กน้อย อันเป็นผลให้ไดนามิกไทม์วอร์ปิงสามารถปรับแนว Warping Path ของคู่ของลำดับย่อยได้เหมาะสมที่สุด ก่อนที่จะเข้าสู่กระบวนการเฉลี่ยตามรูป ซึ่งเป็นการปรับปรุงอัลกอริทึมจากเดิม (Subsequence Time Series Clustering – STSC) ที่ใช้มาตรวัดระยะทางแบบยุคลิดและการเฉลี่ยตามแอมพลิจูด เพราะมาตรวัดระยะทางแบบยุคลิดไม่สามารถเห็นถึงความคล้ายกันระหว่างลำดับย่อยสองลำดับได้อย่างถูกต้อง ทั้งฟังก์ชันการเฉลี่ยตามแอมพลิจูดเองก็ไม่สามารถคงคุณลักษณะของลำดับย่อยที่ทำการเฉลี่ยได้อย่างเหมาะสม หรืออีกนัยหนึ่งคือ การใช้ระยะทางแบบยุคลิดสำหรับอัลกอริทึมการจัดกลุ่มนี้ สามารถนำไปสู่การจัดกลุ่มลำดับย่อยข้างเคียงที่มีความคล้ายกันสูง (Trivial Match) ได้ไม่ถูกต้อง และเมื่อนำลำดับย่อยเหล่านี้ไปทำการเฉลี่ย ก็จะทำให้เกิดผลข้างเคียงการปรับเรียบ (Smoothing) ของลำดับย่อยที่ไม่พึงปรารถนาได้

จากผลการวิจัยที่ผ่านมา ได้มีการพิสูจน์ให้เห็นแล้วว่าอัลกอริทึม STSC นั้นเป็นอัลกอริทึมที่ไร้ความหมาย ทั้งในเชิงทฤษฎีและในเชิงประจักษ์จากการทดลอง ซึ่งได้แสดงให้เห็นว่าอัลกอริทึม STSC นี้จะได้ตัวแทนกลุ่มของการจัดกลุ่มเป็นคลื่นไซน์เสมอ ไม่ว่าข้อมูลอินพุตจะเป็นอย่างไรก็ตาม ซึ่งเป็นตัวแทนกลุ่มที่ไม่ถูกต้องและไม่สามารถนำไปใช้ได้ ดังนั้นโครงการวิจัยนี้จึงมีจุดประสงค์หลักที่จะแก้ปัญหาที่เกิดขึ้นดังกล่าว โดยเริ่มจากการนำเสนออัลกอริทึม 2STSC ที่เป็นการจัดกลุ่มลำดับย่อยที่มีความหมาย จากนั้นได้นำเสนออัลกอริทึม 3STSC ซึ่งเป็นการต่อยอดเพื่อให้สามารถรับมือกับข้อมูลแบบกระแสได้ด้วย

โครงการวิจัยนี้สามารถต่อยอดเพิ่มเติม เพื่อเพิ่มประสิทธิภาพของกระบวนการทำเหมืองข้อมูลที่หลากหลาย กล่าวคือ ฟังก์ชันการเฉลี่ยตามรูป สามารถนำไปใช้ประโยชน์ในปัญหาการหาแม่แบบ (Template Matching) และการจำแนกประเภท (Classification) นอกจากนี้อัลกอริทึม 2STSC และ 3STSC สามารถนำไปใช้ในกระบวนการก่อนประมวลผล (Preprocessing) หรือขั้นรูทีนต่าง ๆ ของการทำเหมืองข้อมูลอย่างหลากหลายได้ เช่นการหากฎความสัมพันธ์ (Association Rules) การจำแนกประเภท (Classification) การค้นพบแบบรูป (Pattern Discovery) และการสร้างมโนทัศน์ (Visualization) เป็นต้น

ในการปรับปรุงอัลกอริทึมที่นำเสนอในงานวิจัยนี้ เราสามารถเพิ่มการปรับแนวของข้อมูลอนุกรมเวลา (Sequence Alignment) และการซีกตัวอย่าง (Re-sampling) เพื่อให้ให้อัลกอริทึมการเฉลี่ยตามรูป มีประสิทธิภาพดีขึ้น นอกจากนี้เรายังสามารถปรับกระบวนการเฉลี่ยให้สามารถหาค่าเฉลี่ยที่เหมาะสมที่สุด (Optimal) ได้ และถ้าหากจะปรับปรุงอัลกอริทึม 3STSC เพิ่มเติม ก็สามารถเลือกใช้ อัลกอริทึมการจัดกลุ่มอื่น ๆ ที่เร็วขึ้น และถ้าจัดพารามิเตอร์ที่ผู้ใช้ต้องเป็นคนกำหนดค่าออกไป ซึ่งได้แก่ จำนวนของกลุ่มข้อมูล และขนาดความยาวของหน้าต่าง (Sliding Window) เป็นต้น รวมทั้งอัลกอริทึมที่ทำการอัปเดตลำดับย่อยที่เก็บไว้ ก็ยังสามารถปรับปรุงให้ดีขึ้นได้ในแง่ของการลดความบิดเบือนของควมมีความหมายของผลลัพธ์การจัดกลุ่ม