

ไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะสำหรับการจำแนกประเภทอนุกรมเวลา



นางสาวพิชามณูชु อนันตเศรษฐ์

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย

The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)
are the thesis authors' files submitted through the University Graduate School.

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

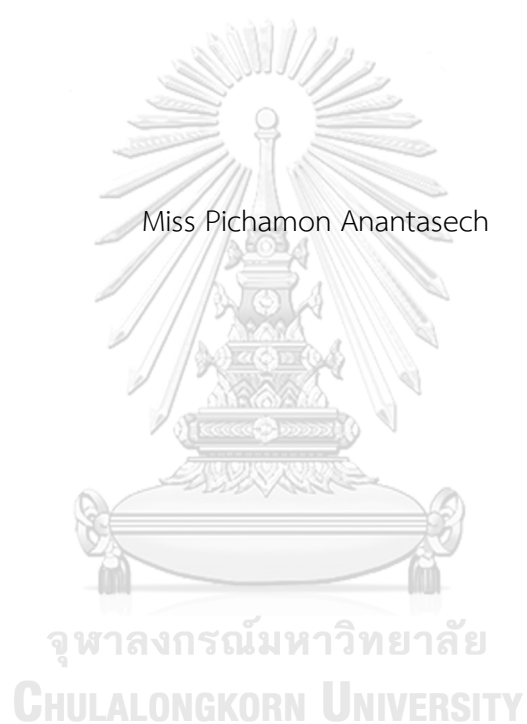
ปีการศึกษา 2560

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ENHANCED WEIGHTED-DYNAMIC TIME WARPING FOR TIME SERIES CLASSIFICATION



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Computer Science
Department of Computer Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2017
Copyright of Chulalongkorn University



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

หัวข้อวิทยานิพนธ์ ไตนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะ
สำหรับการจำแนกประเภทอนุกรมเวลา
โดย นางสาวพิชามณูช อันทศเรษฐ
สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก รองศาสตราจารย์ ดร.โชติรัตน์ รัตนามัทธนะ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้เป็นส่วน
หนึ่งของการศึกษาตามหลักสูตรปริญญาโท

.....คณบดีคณะวิศวกรรมศาสตร์
(รศ. ดร.สุพจน์ เตชวรสินสกุล)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(อาจารย์ ดร.ดวงดาว วิชาดากุล)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร.โชติรัตน์ รัตนามัทธนะ)

.....กรรมการภายนอกมหาวิทยาลัย
(ดร.เหมวรรณ ศิวรักษ์)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

พิชามณูชย์ อนันตเศรษฐ์ : ไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะสำหรับการ
จำแนกประเภทอนุกรมเวลา (ENHANCED WEIGHTED-DYNAMIC TIME WARPING FOR
TIME SERIES CLASSIFICATION) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: รศ. ดร.โชติรัตน์ รัตน
มัทธนะ, หน้า.

ไดนามิกไทม์วอร์ปิงเป็นมาตรวัดระยะห่างซึ่งเป็นที่นิยมและใช้กันอย่างแพร่หลายสำหรับ
การจำแนกประเภทข้อมูลอนุกรมเวลาเนื่องจากความยืดหยุ่นและความทนทานต่อข้อมูลใน
หลากหลายกรณีอันเป็นคุณสมบัติหลักของไดนามิกไทม์วอร์ปิง

อย่างไรก็ตามไดนามิกไทม์วอร์ปิงอาจนำไปสู่การยืดหดที่มากจนเกินไปส่งผลให้แนวการ
ปรับตรงของจุดหลายจุดบนอนุกรมเวลาหนึ่งสุดเพียงจุดเดียวบนอีกอนุกรมเวลาหนึ่งและอาจส่งผล
ให้เกิดการจำแนกประเภทผิดพลาดเมื่อนำไปใช้ในการจำแนกประเภทอนุกรมเวลา ซึ่งปัญหาดังกล่าว
ถือเป็นปัญหาสำคัญที่ทำให้มีงานวิจัยมากมายถูกนำเสนอออกมาเพื่อใช้แก้ปัญหา หนึ่งในนั้นคือการ
เสนอไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบต่างๆ ซึ่งสามารถแก้ปัญหาดังกล่าวได้เป็นอย่างดีแต่กลับ
ต้องแลกมาด้วยตัวแปรเสริมที่ยากต่อการหาค่าที่เหมาะสม

เพื่อขจัดปัญหาดังกล่าว วิทยานิพนธ์ฉบับนี้จึงได้เสนอมาตรวัดระยะห่างที่ยังคงคุณสมบัติที่
ดีของไดนามิกไทม์วอร์ปิงฉบับดั้งเดิมเอาไว้ซึ่งสามารถแก้ปัญหาแนวการปรับตรงที่ผิดพลาดของ
ไดนามิกไทม์วอร์ปิงแบบดั้งเดิมได้ โดยอาศัยหลักการของไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแต่ไม่
จำเป็นต้องหาค่าของตัวแปรเสริมเช่นไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบอื่นๆ ภายใต้ชื่อไดนามิก
ไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะ ที่นอกจากจะให้ประสิทธิภาพในการจำแนกประเภท
อนุกรมเวลาให้ดีขึ้นแล้ว ก็ยังคงไว้ด้วยความซับซ้อนของเวลาที่เท่ากันกับไดนามิกไทม์วอร์ปิงแบบ
ดั้งเดิมอีกด้วย

ภาควิชา วิศวกรรมคอมพิวเตอร์

ลายมือชื่อนิสิต

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์

ลายมือชื่อ อ.ที่ปรึกษาหลัก

ปีการศึกษา 2560

5970267021 : MAJOR COMPUTER SCIENCE

KEYWORDS: TIME SERIES / TIME SERIES CLASSIFICATION / DYNAMIC TIME WARPING /
WEIGHTED DYNAMIC TIME WARPING

PICHAMON ANANTASECH: ENHANCED WEIGHTED-DYNAMIC TIME WARPING FOR
TIME SERIES CLASSIFICATION. ADVISOR: ASSOC. PROF. CHOTIRAT
RATANAMAHAHATANA, Ph.D., pp.

Dynamic time warping (DTW) has been widely used as a distance measure for time series classification because its matching is elastic and robust in most cases. However, DTW may lead to over compression that could align too many consecutive points from one time series sequence to only one point on another and may be the cause of missclassification in time series classification. The over compression is an essential problem producing many research works that try to resolve the problem. One of those is a family of weighted dynamic time warping which can fix the problem very well but is difficult to find parameter's value.

To eliminate the problem, this thesis proposes a new distance measure called Enhanced Weighted Dynamic Time Warping which retains the advantages of original dynamic time warping and also fixes the too-many-to-one-point problem by using the idea of other weighted dynamic time warping approaches. Enhanced Weighted Dynamic Time warping does not only improve an accuracy for classification, but also maintain the original dynamic time warping's time complexity.

Department: Computer Engineering Student's Signature

Field of Study: Computer Science Advisor's Signature

Academic Year: 2017

กิตติกรรมประกาศ

วิทยานิพนธ์นี้สำเร็จลุล่วงได้ด้วยความกรุณาจาก รองศาสตราจารย์ ดร.โชติรัตน์ รัตนามัทธนะ อาจารย์ที่ปรึกษา ผู้คอยให้คำปรึกษาและแง่คิด ทั้งในด้านวิชาการรวมไปถึงด้านอื่นๆ ทั้งยังเป็นแรงผลักดัน ส่งเสริมความคิดเชิงบวก รวมไปถึงเป็นผู้ตรวจทานแก้ไขทำให้วิทยานิพนธ์ฉบับนี้ให้สำเร็จลุล่วงไปด้วยดี ขอขอบพระคุณเป็นอย่างสูงมา ณ ที่นี้

ขอขอบพระคุณอาจารย์ ดร.ดวงดาว วิชาดากุลและดร.เหมวรรณ ศิวรักษ์ ผู้ให้เกียรติเป็นประธานกรรมการสอบวิทยานิพนธ์และกรรมการสอบวิทยานิพนธ์ อีกทั้งยังเป็นผู้ชี้แนวทางในการปรับปรุงวิทยานิพนธ์ให้มีคุณภาพยิ่งขึ้น

ขอขอบคุณเพื่อนๆในแล็บฯที่ให้ความช่วยเหลือและส่งเสริมกันและกันเป็นอย่างดีตลอดระยะเวลาที่ผ่านมา

สุดท้ายนี้ขอขอบคุณสมาชิกในครอบครัว เพื่อนคนอื่นๆ เพลง VOICE ของศิลปิน Standing Egg รวมไปถึงแรงบันดาลใจอื่นๆที่คอยผลักดันอีกทั้งยังเป็นกำลังใจในการทำงานวิจัยนี้เสมอมากระทั่งวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยดี

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารบัญ

หน้า

บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
บทที่ 1 บทนำ	1
1.1 ที่มาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของงานวิจัย.....	2
1.3 ขอบเขตของงานวิจัย.....	2
1.4 ประโยชน์ที่ได้รับจากงานวิจัย.....	2
1.5 ขั้นตอนการดำเนินงาน	3
1.6 ผลงานวิจัยที่ได้ตีพิมพ์	4
บทที่ 2 ทฤษฎีและงานที่เกี่ยวข้อง.....	5
2.1 ข้อมูลอนุกรมเวลา (Time Series Data).....	5
2.2 การจำแนกประเภทแบบเพื่อนบ้านใกล้สุดลำดับที่ 1 (1-Nearest Neighbor Classification)	6
2.4 ไดนามิกไทม์วอร์ปิง (Dynamic Time Warping, DTW)	7
2.5 ฟังก์ชันซิกมอยด์ (Sigmoid Function)	8
2.6 ฟังก์ชันเอกซ์โพเนนเชียล (Exponential Function)	9
2.7 การกำหนดเงื่อนไขบังคับโดยรวม (Global Constraint)	10
2.8 เค-โพลด์ครอสวาเลดิชัน (k-fold cross validation).....	11
2.9 งานวิจัยที่เกี่ยวข้อง.....	12
บทที่ 3 ไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะ.....	15
3.1 การจัดเตรียมข้อมูล.....	15

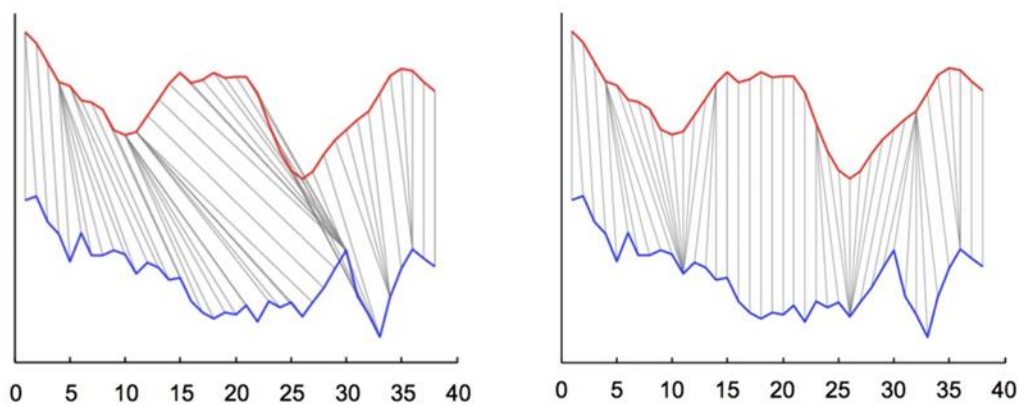
3.2 ไดนามิกใหม่วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะ	17
3.3 ความซับซ้อนของเวลา	18
3.4 การวัดผลมาตรวัดระยะห่าง	19
บทที่ 4 การทดลองและวิเคราะห์ผล	20
4.1 ชุดข้อมูลที่ใช้ในงานวิจัย	20
4.2 การทดสอบความแม่นยำของมาตรวัดระยะห่างเมื่อนำไปใช้ในการจำแนกประเภทของข้อมูล	21
4.3 การทดสอบความแม่นยำของมาตรวัดระยะห่างร่วมกับเงื่อนไขบังคับโดยรวม	33
บทที่ 5 สรุปผลงานวิจัยและข้อเสนอแนะ	35
5.1 สรุปผลงานวิจัย	35
5.2 ข้อเสนอแนะ	36
รายการอ้างอิง	38
ภาคผนวก.....	40
ภาคผนวก ก.....	41
ประวัติผู้เขียนวิทยานิพนธ์	64

บทที่ 1 บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ไดนามิกไทม์วอร์ปิง (Dynamic Time Warping) [1] เป็นหนึ่งในมาตรวัดที่ใช้กันอย่างแพร่หลาย โดยเฉพาะในการวัดระยะห่างระหว่างอนุกรมเวลา (time series) เพื่อใช้ในการจำแนกประเภท (classification) หรือการจัดกลุ่ม (clustering) ของข้อมูล เหตุผลที่ทำให้ไดนามิกไทม์วอร์ปิงได้รับความนิยมอย่างแพร่หลาย เป็นเพราะการเป็นมาตรวัดที่มีความยืดหยุ่น (elastic measurement) ที่ยังสามารถทำงานได้อย่างมีประสิทธิภาพ แม้มีการเลื่อนของข้อมูลในแกนของเวลา ด้วยคุณสมบัติดังกล่าวนี้ทำให้ไดนามิกไทม์วอร์ปิงได้รับความสนใจในการนำไปพัฒนาต่อยอดกันอย่างแพร่หลาย

จากงานวิจัยในปี 2011 [2] ได้นำเสนอแนวคิดที่ว่า การใช้ค่าถ่วงน้ำหนักส่งผลต่อความถูกต้องในการจำแนกประเภทหรือจัดกลุ่มของข้อมูลผ่านอัลกอริทึมไดนามิกไทม์วอร์ปิงแบบถ่วงน้ำหนัก จากรูปที่ 1.1 จะเห็นได้ว่าแนวการปรับตรง (alignment) ของไดนามิกไทม์วอร์ปิงในบางตำแหน่งไม่ได้นำไปสู่ตำแหน่งที่ควรจะเป็นบน อีกอนุกรมเวลาหนึ่ง ทั้งยังเกิดปัญหาการวอร์ปไปที่จุด ๆ เดียว (too-many-to-one-point problem) อีกด้วย



รูปที่ 1.1 ภาพแสดงแนวการปรับตรง (alignment) ของไดนามิกไทม์วอร์ปิง (ซ้าย) และแนวการปรับตรงของไดนามิกไทม์วอร์ปิงแบบถ่วงน้ำหนัก (ขวา)

(ที่มา : Y.-S. Jeong และคณะ [2])

อย่างไรก็ตามอัลกอริทึมที่ได้รับการพัฒนาต่อออกมาจากไดนามิกไทม์วอร์ปิงนี้เพื่อแลกมาด้วยความถูกต้อง (accuracy) หากไม่ใช้ระยะเวลาในการคำนวณรวมถึงการสอนพารามิเตอร์ต่าง ๆ ที่นานกว่า DTW ก็จะสามารถลดอัตราความผิดพลาด (error rate) ลงได้เพียงเล็กน้อย

เพื่อแก้ไขปัญหาดังกล่าว วิทยานิพนธ์นี้จึงนำเสนออัลกอริทึมใหม่ซึ่งเป็นการเพิ่มสมรรถนะให้แก่อัลกอริทึมแบบถ่วงน้ำหนัก โดยระยะเวลาที่ใช้ไม่มากไปกว่าไดนามิกไทม์วอร์ปิงแบบดั้งเดิม ทั้งยังสามารถลดอัตราความผิดพลาดได้ใกล้เคียงกับไดนามิกไทม์วอร์ปิงแบบถ่วงน้ำหนักด้วย

1.2 วัตถุประสงค์ของงานวิจัย

เพื่อเพิ่มสมรรถนะอัลกอริทึมไดนามิกไทม์วอร์ปิงแบบถ่วงน้ำหนักภายใต้ความซับซ้อนของเวลาที่เท่ากับไดนามิกไทม์วอร์ปิงแบบดั้งเดิมโดยไม่จำเป็นต้องอาศัยตัวแปรเสริมที่ปรากฏในอัลกอริทึมการถ่วงน้ำหนักแบบอื่น ๆ และยังคงให้ความถูกต้องที่ใกล้เคียงหรือสูงกว่าอัลกอริทึมที่พัฒนามาจากไดนามิกไทม์วอร์ปิงด้วยกัน

1.3 ขอบเขตของงานวิจัย

ข้อมูลอนุกรมเวลาที่น่าสนใจนำมาใช้ทดสอบผลของวิธีการที่นำเสนอมาจาก UCR Time Series Classification Archive [3] ในข้อมูลแต่ละชุดจะประกอบไปด้วยข้อมูลเรียนรู้ ข้อมูลทดสอบ คลาสของข้อมูลเรียนรู้แต่ละตัว คลาสของข้อมูลทดสอบแต่ละตัว จำนวนข้อมูลในแต่ละชุดข้อมูล ความยาวของข้อมูลอนุกรมเวลาแต่ละตัว และค่าความผิดพลาดในการคำนวณด้วยการใช้วิธีการจำแนกประเภทด้วยวิธีการจำแนกข้อมูลแบบเพื่อนบ้านใกล้สุดลำดับที่ 1 [4-6] โดยใช้การวัดระยะทางแบบไดนามิกไทม์วอร์ปิง ในกรณีที่ใช้ข้อมูลเรียนรู้ทุกข้อมูล ทดสอบกับข้อมูลทดสอบทุกข้อมูล

1.4 ประโยชน์ที่ได้รับจากงานวิจัย

การวัดผลของอัลกอริทึมในงานวิทยานิพนธ์นี้เปรียบเทียบจากผลการทดลองของการจำแนกประเภทของข้อมูลโดยวิธีเพื่อนบ้านที่ใกล้ที่สุดอันดับที่ 1 [4-6] ในเว็บไซต์ UCR Time Series Classification Archive [3] อันเป็นชุดข้อมูลที่ใช้โดยสากล ประกอบด้วยอนุกรมเวลาจำนวน 19 ชุด ซึ่งล้วนแล้วแต่เป็นข้อมูลที่ได้รับการแปลงค่าให้เป็นบรรทัดฐาน (normalization) [7] มาแล้วทั้งสิ้น

1.5 ขั้นตอนการดำเนินงาน

กิจกรรม	2017								2018					
	พ.ค.	มิ.ย.	ก.ค.	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.	พ.ค.	มิ.ย.
1 ทบทวนงานวิจัยที่เกี่ยวข้อง	✓	✓	✓	✓	✓	✓								
2 ทำการทดลองของงานวิจัยที่เกี่ยวข้องและศึกษาหาแนวทางในการปรับปรุงอัลกอริทึมไดนามิกโทมออร์บ์ปีงแบบถ่วงน้ำหนัก						✓	✓	✓						
3 ออกแบบมาตรวัดระยะห่าง								✓	✓	✓				
4 ทำการทดลองของงานวิจัยที่เกี่ยวข้องและศึกษาหาแนวทางในการปรับปรุงอัลกอริทึมไดนามิกโทมออร์บ์ปีงแบบถ่วงน้ำหนัก									✓	✓	✓	✓		
5 สรุป ตีพิมพ์งานวิจัย และเรียบเรียงวิทยานิพนธ์											✓	✓	✓	✓

1.6 ผลงานวิจัยที่ได้ตีพิมพ์

P. Anantasech, and C. A. Ratanamahatana, “Enhanced Weighted Dynamic Time Warping for time series classification,” in Third International Congress on Information and Communication Technology, ICICT 2018, at Brunel University London, London, United Kingdom, from 27 – 28 February 2018.



บทที่ 2 ทฤษฎีและงานที่เกี่ยวข้อง

2.1 ข้อมูลอนุกรมเวลา (Time Series Data)

ข้อมูลอนุกรมเวลา คือ ข้อมูลใดๆที่ได้รับการบันทึกค่าไว้ ณ จุดต่าง ๆ ของเวลา หรืออีกนัยหนึ่ง คือ ข้อมูลที่มีการเปลี่ยนแปลงไปตามจุดต่างๆของเวลา มีลำดับของการมาก่อนและหลัง ข้อมูลอนุกรมเวลาสามารถพบเห็นได้ทั่วไปในชีวิตประจำวัน ตัวอย่างที่เห็นได้ชัดเจน เช่น ข้อมูลราคาหุ้น (Stock price) ข้อมูลคลื่นไฟฟ้าหัวใจ (Electrocardiogram หรือ ECG) หรือข้อมูลความรุนแรงแผ่นดินไหวในปีต่างๆ (Earthquake magnitude) เป็นต้น ดังรูป 2.1 ซึ่งเป็นกราฟแสดงราคาหุ้นของบริษัท Google ในช่วงเวลาต่างๆของวันที่ 19 สิงหาคม จะเห็นได้ว่าราคาหุ้นมีการเปลี่ยนแปลงอยู่ตลอดเวลาที่เปลี่ยนไป



รูปที่ 2.1 ราคาหุ้นของบริษัท Google ในช่วงเวลาต่างๆของวันที่ 19 สิงหาคม

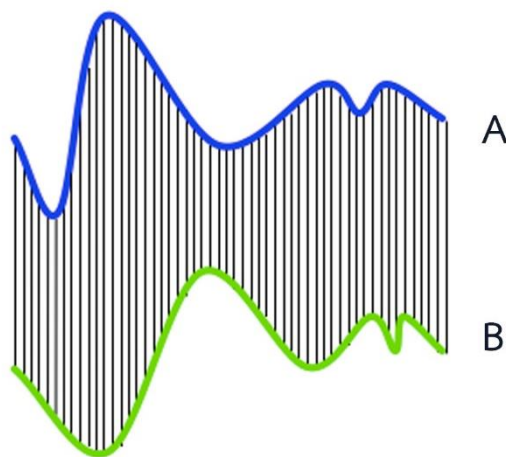
(ที่มา : <http://businessforecastblog.com/forecasting-googles-stock-price-goog-on-20-trading-day-horizons/>)

2.2 การจำแนกประเภทแบบเพื่อนบ้านใกล้สุดลำดับที่ 1 (1-Nearest Neighbor Classification)

การจำแนกประเภทแบบเพื่อนบ้านใกล้สุดลำดับที่ 1 [4-6] เป็นหนึ่งในวิธีการจำแนกประเภทของข้อมูลที่มีความนิยมมากที่สุดประเภทหนึ่ง โดยพิจารณาจากข้อมูลเรียนรู้ที่ใกล้ที่สุดกับข้อมูลสอบถามที่เราต้องการทราบประเภท การจำแนกข้อมูลประเภทนี้จำเป็นที่จะต้องรู้ประเภทของข้อมูลเรียนรู้มาก่อน จึงจะสามารถใช้การจำแนกประเภทนี้จำแนกประเภทข้อมูลสอบถาม (Query) ที่เราต้องการทราบประเภทได้ โดยอาศัยหลักการในการหาเพื่อนบ้านที่ใกล้ที่สุดคือหาระยะทางระหว่างข้อมูลเรียนรู้ทุกตัวกับข้อมูลสอบถามที่เราต้องการทราบประเภท จากนั้นเลือกข้อมูลเรียนรู้ที่มีระยะห่างกับข้อมูลสอบถามน้อยที่สุด เพื่อให้ได้ข้อมูลเรียนรู้ที่อยู่ใกล้กับข้อมูลสอบถามมากที่สุด และใช้ในการจำแนกประเภทข้อมูลของข้อมูลสอบถามนั้น โดยจะทำเช่นนี้กับข้อมูลสอบถามทุกตัวที่ต้องการจำแนกประเภทจนครบทั้งหมด

2.3 ระยะทางยุคลิด (Euclidean distance)

ระยะทางยุคลิด [8, 9] คือ การวัดระยะห่างระหว่างจุดสองจุดใด ๆ ในระนาบสองมิติโดยจะเป็นการ วัดระยะห่างระหว่างสองจุดใด ๆ ที่อยู่บนตำแหน่งเดียวกัน ดังแสดงในรูปที่ 2.2



รูปที่ 2.2 แสดงแนวการปรับตรงของระยะทางยุคลิดระหว่างอนุกรมเวลา 2 ชุด

(ที่มา : http://repmus.ircam.fr/_media/esling/dtwseuclid.jpg)

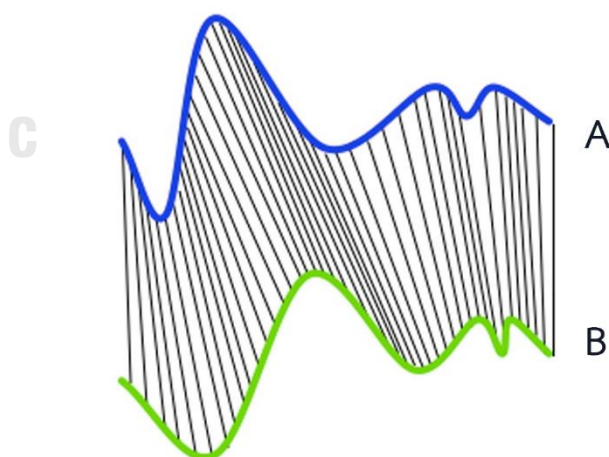
ให้อนุกรมเวลา $A = (a_1, a_2, \dots, a_n)$ และอนุกรมเวลา $B = (b_1, b_2, \dots, b_n)$ โดยทั้งสองอนุกรมมีความยาว n การหาระยะทางยุคลิดจะสามารถคำนวณได้ดังสมการ (2.1)

$$Euclidean(A, B) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (2.1)$$

ในงานวิทยานิพนธ์นี้จะใช้ระยะทางยุคลิดในการเปรียบเทียบผลการทดลอง แต่จะไม่นำมาพัฒนาต่อ เนื่องจากระยะทางยุคลิดไม่สามารถใช้หาระยะห่างของอนุกรมเวลา 2 ชุดที่มีความยาวไม่เท่ากันได้

2.4 ไดนามิกไทม์วอร์ปิง (Dynamic Time Warping, DTW)

ไดนามิกไทม์วอร์ปิง [1] เป็นวิธีการวัดระยะห่างระหว่างอนุกรมเวลาสองชุดที่ได้รับความนิยม อย่างแพร่หลายในการนำไปใช้เพื่อบอกถึงความต่างหรือคล้ายระหว่างอนุกรมเวลาสองชุดใด ๆ เนื่องจากคุณสมบัติของการเป็นมาตรวัดที่มีความยืดหยุ่นทำให้ทนทานต่อข้อมูลที่มีความคลาดเคลื่อนหรือการเลื่อนของข้อมูลในแนวแกนเวลาได้เป็นอย่างดี ดังแสดงในภาพที่ 2.3 ที่จะเห็นได้ว่าแม้ข้อมูลมีการเลื่อนไปในแนวแกนเวลา ไดนามิกไทม์วอร์ปิง ก็ยังสามารถจับคู่สองจุดใด ๆ บนอนุกรมเวลาได้อย่างเหมาะสม



รูปที่ 2.3 ภาพแสดงแนวการปรับตรงของไดนามิกไทม์วอร์ปิงระหว่างอนุกรมเวลา 2 ชุด
(ที่มา : http://repmus.ircam.fr/_media/esling/dtwseuclid.jpg)

ให้อนุกรมเวลา $A = (a_1, a_2, \dots, a_n)$ และอนุกรมเวลา $B = (b_1, b_2, \dots, b_m)$ โดยอนุกรมทั้งสองมีความยาว n และ m ตามลำดับ ระยะทางไดนามิกโทมวอร์ปิงจะสามารถคำนวณได้ดังสมการ (2.2) และ (2.3)

$$DTW \text{ distance } (A, B) = \sqrt{\gamma(i, j)} \quad (2.2)$$

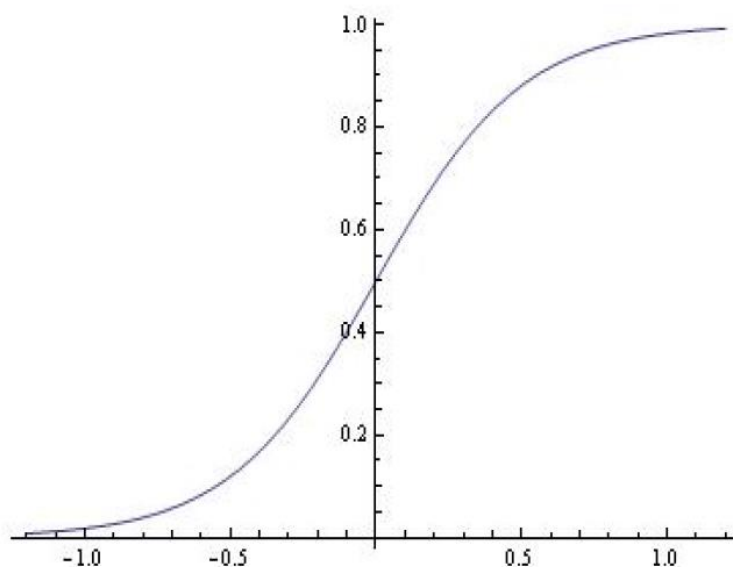
$$\gamma(i, j) = (a_i - b_j)^2 + \min \begin{cases} \gamma(i-1, j) \\ \gamma(i-1, j-1) \\ \gamma(i, j-1) \end{cases} \quad (2.3)$$

ในงานวิทยานิพนธ์นี้จะใช้ระยะทางไดนามิกโทมวอร์ปิงเป็นอัลกอริทึมพื้นฐานรวมถึงใช้เป็นพื้นฐานขั้นต่ำ (baseline) ในการเปรียบเทียบผลการทดลอง ด้วยไดนามิกโทมวอร์ปิงถือเป็นอัลกอริทึมที่ได้รับความนิยม อย่างแพร่หลาย รวมถึงเป็นแนวทางให้หลาย ๆ งานวิจัยได้นำไปใช้และศึกษาต่อ

2.5 ฟังก์ชันซิกมอยด์ (Sigmoid Function)

ฟังก์ชันซิกมอยด์ [10] คือ ฟังก์ชันคณิตศาสตร์ที่มีลักษณะของเส้นโค้งเป็นรูปตัวเอส (S-shape) หรือที่เรียกว่าเส้นโค้งซิกมอยด์ (Sigmoid curve) ดังแสดงในรูปที่ 2.4 โดยทั่วไปเมื่อก้าวถึงฟังก์ชันซิกมอยด์ จะหมายถึงฟังก์ชันพิเศษของโลจิสติกฟังก์ชันที่แสดงอยู่ในรูปของสมการ (2.4) ดังต่อไปนี้

$$S(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{e^x + 1} \quad (2.4)$$



รูปที่ 2.4 กราฟเส้นโค้งรูปตัวเอสของฟังก์ชันซิกมอยด์
(ที่มา: Nikolay Kyurkchiev และ Svetoslav Markov (10))

2.6 ฟังก์ชันเอกซ์โพเนนเชียล (Exponential Function)

ฟังก์ชันเอกซ์โพเนนเชียล [11] ในทางคณิตศาสตร์ คือ ฟังก์ชันที่อยู่ในรูปของสมการ (2.5) ดังแสดงต่อไปนี้

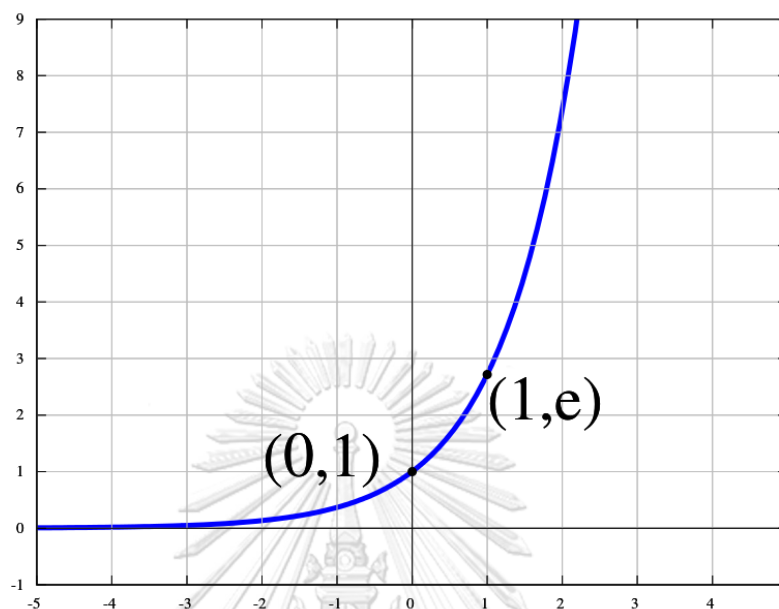
$$f(x) = b^x \quad (2.5)$$

ซึ่งจะเห็นได้ว่าตัวแปรที่รับค่าเข้ามาจะอยู่ในส่วนของเลขยกกำลัง เมื่อพิจารณาในมุมฟังก์ชันของตัวแปรจริง ฟังก์ชันเอกซ์โพเนนเชียล หมายถึง ฟังก์ชันที่มีลักษณะพิเศษอันเป็นหนึ่งเดียว คือ อัตราการเติบโตของฟังก์ชันที่เป็นสัดส่วนกับฟังก์ชัน ดังแสดงต่อไปนี้ในสมการ (2.6)

$$\frac{d}{dx}(b^x) = b^x \log_e(b) \quad (2.6)$$

จะเห็นได้ว่าค่าคงที่ที่เป็นสัดส่วนกับค่าของฟังก์ชันเอ็กโพเนนเชียล คือ ลอการิทึมธรรมชาติ (natural logarithm) ของ (2.5) ซึ่งหากเปลี่ยนค่าฐาน (base) ของฟังก์ชันเอกซ์โพเนนเชียลให้เป็นค่าคงที่ $e \approx 2.71828$ แล้ว ค่าคงที่ที่เป็นสัดส่วนกับค่าของฟังก์ชันเอ็กโพเนนเชียล จะมีค่าเท่ากับ 1 ซึ่งหมายถึง ค่าอนุพันธ์ของฟังก์ชันนี้มีค่าเท่ากับตัวมันเอง

$$\frac{d}{dx}(e^x) = e^x \log_e(e) \quad (2.7)$$



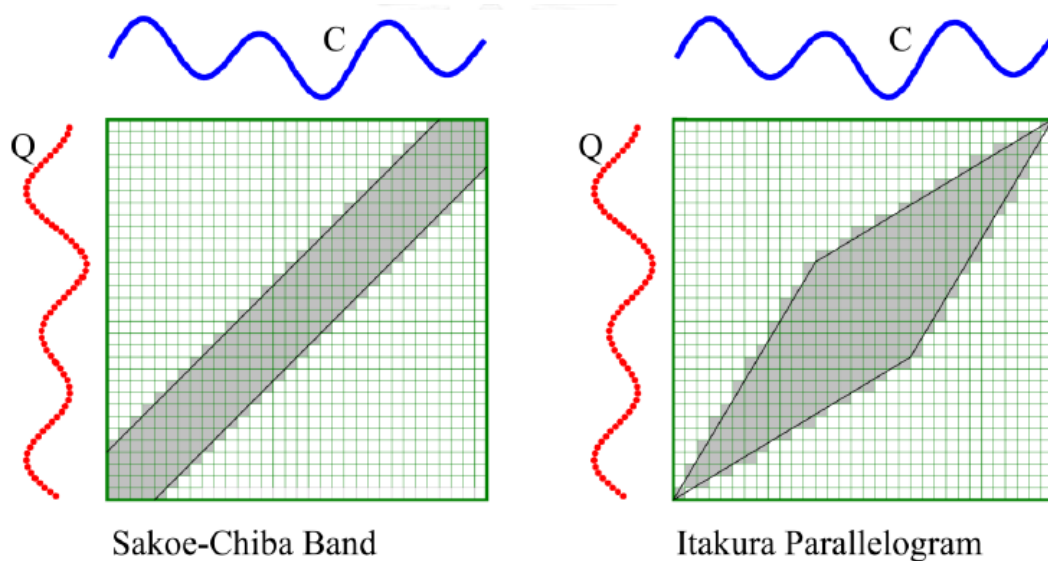
รูปที่ 2.5 กราฟเส้นโค้งของฟังก์ชันเอกซ์โพเนนเชียล

(ที่มา https://en.wikipedia.org/wiki/Exponential_function#/media/File:Exp.svg)

2.7 การกำหนดเงื่อนไขบังคับโดยรวม (Global Constraint)

การกำหนดเงื่อนไขบังคับโดยรวม [12-14] เป็นหนึ่งในวิธีการอันเป็นที่นิยมซึ่งมักจะใช้ควบคู่กันกับมาตรวัดระยะห่างไดนามิกโทมวอร์ปปีง เนื่องจากโดยธรรมชาติของไดนามิกโทมวอร์ปปีงจะอนุญาตให้จุดจุดหนึ่งบนอนุกรมเวลาสามารถจับคู่กับอีกจุดหนึ่งบนอีกอนุกรมเวลาได้โดยไม่จำกัดว่าจุดดังกล่าวจะต้องอยู่ใกล้หรืออยู่ติดกันกับจุดดังกล่าว ส่งผลให้ในบางกรณีจุดบนอนุกรมเวลาหนึ่งไปจับคู่กับจุดบนอีกอนุกรมเวลาหนึ่งซึ่งแตกต่างกันบนแกนเวลามากจนเกินไป เมื่อนำไปใช้ในการจำแนกประเภทจึงอาจทำให้เกิดความผิดพลาดตามมา การกำหนดเงื่อนไขบังคับโดยรวมจะเป็นการอนุญาตให้การจับคู่ระหว่างจุดสองจุดนั้นไม่ห่างกันมากจนเกินไปกว่าตามเงื่อนไขที่กำหนด

วิธีการกำหนดเงื่อนไขบังคับโดยรวมมีอยู่หลากหลาย โดยวิธีซึ่งเป็นที่นิยมได้แก่ การกำหนดเงื่อนไขโดยรวมแบบซาโก-ชิบะ [12] และการกำหนดเงื่อนไขโดยรวมแบบอิตาคูระ [13] โดยการกำหนดเงื่อนไขโดยรวมแบบซาโก-ชิบะจะกำหนดเป็นค่าระยะทางเป็นค่าคงที่ ส่วนการกำหนดเงื่อนไขโดยรวมแบบอิตาคูระจะเป็นการกำหนดโดยกำหนดเป็นฟังก์ชันระยะทางดังแสดงเป็นพื้นที่ในรูป 2.6



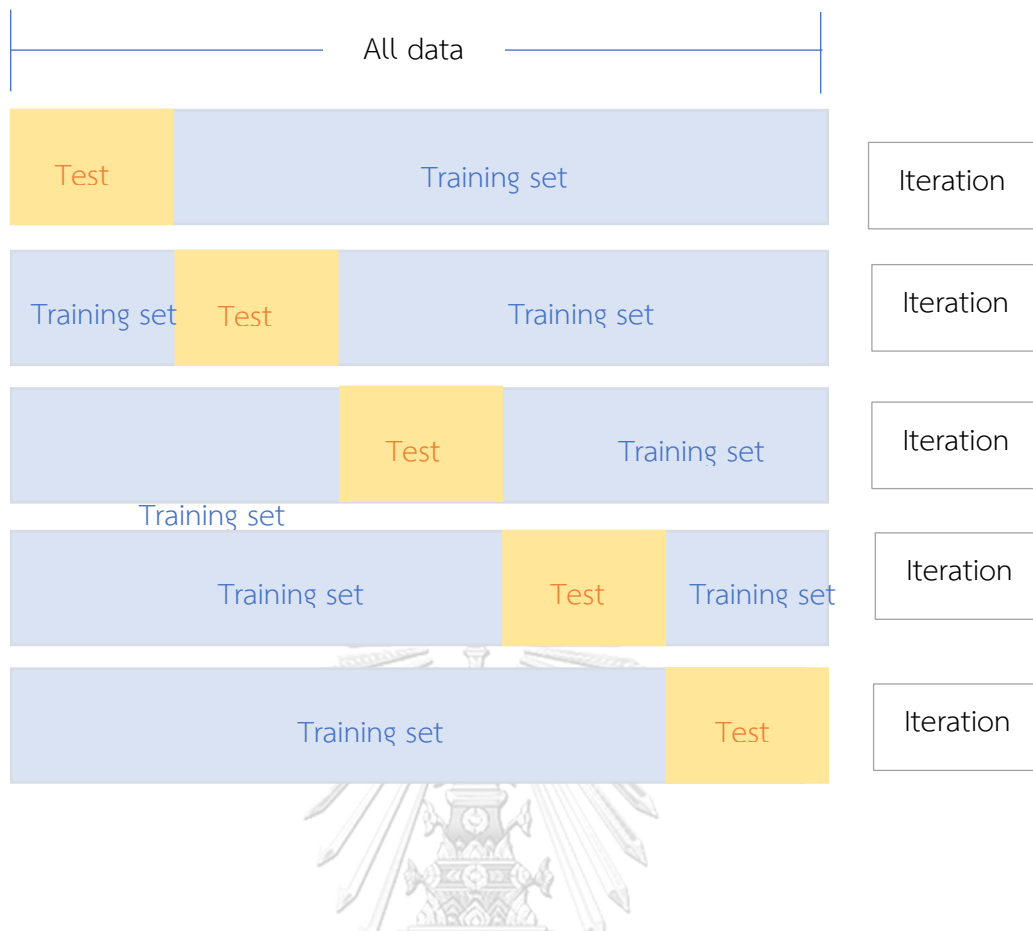
รูปที่ 2.6 แสดงขอบเขตของการกำหนดเงื่อนไขโดยรวมแบบซาโก-ชิบะ [12] และการกำหนดเงื่อนไขโดยรวมแบบอิตาคูระ [13] (ที่มา: Ratanamahatana และ Keogh [14])

นอกจากการกำหนดเงื่อนไขโดยรวมจะสามารถเพิ่มประสิทธิภาพในการจำแนกประเภทเมื่อนำไปใช้ร่วมกับมาตรวัดระยะทางไดนามิกโทมวอร์ปิงได้แล้วนั้น การกำหนดเงื่อนไขโดยรวมยังสามารถช่วยลดเวลาให้ในการคำนวณให้กับระยะทางไดนามิกโทมวอร์ปิงได้อีกด้วย

2.8 เค-โฟลด์ครอสวาเลชัน (k-fold cross validation)

เค-โฟลด์ครอสวาเลชัน [15] เป็นหนึ่งในวิธีการซึ่งเป็นที่นิยมในการวัดประสิทธิภาพของโมเดล ซึ่งจะสามารถทำได้โดยแบ่งข้อมูลที่มีออกเป็น k ส่วนเท่า ๆ กัน โดยเลือกใช้ข้อมูลจำนวน $k - 1$ ส่วนในการสอน (training) และใช้ข้อมูลส่วนที่เหลืออยู่ในการทดสอบ (testing) เป็นจำนวน k รอบ จากนั้นจึงหาค่าเฉลี่ยของความแม่นยำที่ได้จากทั้ง k รอบออกมา โดยข้อมูลทดสอบแต่ละชุดจะใช้เพียง 1 ครั้งเท่านั้นดังแสดงในรูป 2.7

นอกจากเค-โฟลด์ครอสวาเลชันจะสามารถใช้ทดสอบประสิทธิภาพของโมเดลได้แล้ว ยังเป็นที่นิยมในการใช้ปรับจูนตัวแปรต่าง ๆ ของโมเดล ซึ่งวิธีการดังกล่าวจะถูกใช้ในข้อมูลสอน ก่อนนำไปใช้จริง

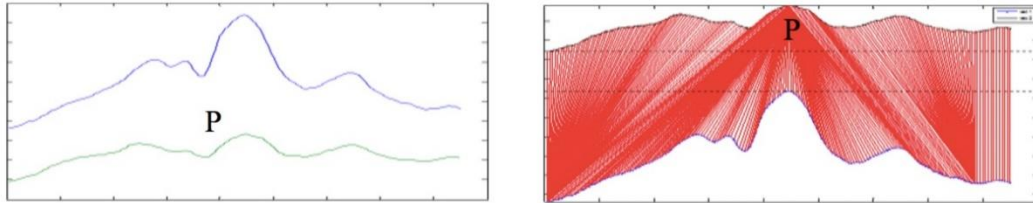


รูปที่ 2.7 แผนภาพแสดงการแบ่งข้อมูลสอนและข้อมูลทดสอบในการทำ 5-โฟลด์ครอส
วาลิเดชัน

2.9 งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่มุ่งหวังในการหามาตรวัดแบบใหม่หรือปรับปรุงจากมาตรวัดที่เป็นพื้นฐานขั้นต่ำให้ประสิทธิภาพในการจำแนกประเภทของข้อมูลดียิ่งขึ้น ถูกนำเสนอออกมาเรื่อย ๆ ในช่วงหลายปีที่ผ่านมา

ในปี 2001 E. Keogh และ M. Pazzani [16] ได้นำอัลกอริทึมของไดนามิกไทม์วอร์ปไปปรับปรุงภายใต้ชื่อ Derivative Dynamic Time Warping (DDTW) จากการนำอนุพันธ์อันดับที่หนึ่งเข้ามาแทนค่า แต่ละจุดในช่วงของการคำนวณระยะทางไดนามิกไทม์วอร์ป ซึ่งการทำเช่นนี้จะสามารถแก้ปัญหารวอร์ปไปที่จุด ๆ เดียวดังรูปที่ 2.8 ซึ่งเป็นปัญหาหลัก ของไดนามิกไทม์วอร์ปได้ แต่จะทำให้สูญเสียรูปร่างของอนุกรมเวลาและอาจ รวมถึงการสูญเสียคุณลักษณะ (feature) ของอนุกรมเวลาด้วย



รูปที่ 2.8 ภาพแสดงแนวการปรับตรงไปสู่จุดเดียวของไดนามิกไทม์วอร์ปิง

(ที่มา : X.Ying และ W.Bryan [17])

ในปี 2011 Y. -S. Jeong และคณะ [18] ได้นำอัลกอริทึมของไดนามิกไทม์วอร์ปิงไปปรับปรุงโดยได้ทำการทดลองและเสนอแนวทางให้เห็นถึงความสำคัญของค่าถ่วงน้ำหนักที่ส่งผลกระทบต่อความถูกต้องของการจัดกลุ่มและการจำแนกประเภทของอนุกรมเวลา สามารถคำนวณได้ดังสมการ (2.8) – (2.10)

$$WDTW(A, B) = \sqrt{\gamma^*(i, j)} \quad (2.8)$$

$$\gamma^*(i, j) = |w_{|i-j|}(a_i - b_j)|^2 + \min \begin{cases} \gamma^*(i-1, j) \\ \gamma^*(i-1, j-1) \\ \gamma^*(i, j-1) \end{cases} \quad (2.9)$$

$$w_k = \left[\frac{w_{max}}{1 + e^{-g(k-m_c)}} \right] \quad (2.10)$$

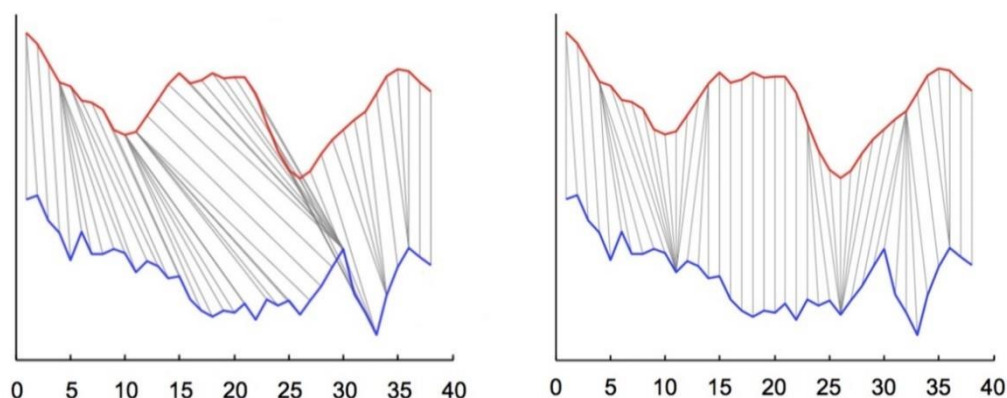
เมื่อ $k = |i - j|$, m_c คือ ตำแหน่งที่อยู่ตรงกลางของอนุกรมเวลา

w_{max} คือ ขอบเขตบนของค่าถ่วงน้ำหนัก

g คือ ตัวแปรเสริม (parameter) ที่เป็นค่าคงที่

จากรูปที่ 2.9 จะเห็นได้ว่าการเพิ่มค่าถ่วงน้ำหนักลงไปให้กับอัลกอริทึมไดนามิกไทม์วอร์ปิง จะทำให้แนวการปรับตรงของสองอนุกรมเปลี่ยน ซึ่งส่งผลกระทบต่อความถูกต้องเมื่อนำอัลกอริทึมดังกล่าวไปใช้ในการแบ่งประเภทของข้อมูลอนุกรมเวลา สังเกตได้จากแนวการปรับตรงของอนุกรมเวลาของทั้งสองอัลกอริทึม ซึ่งหากพิจารณาเรื่องตำแหน่งที่อนุกรมเวลาถูกวอร์ปไป จะเห็นได้ว่าแนวการปรับตรงของอัลกอริทึมไดนามิกไทม์ วอร์ปิงแบบถ่วงน้ำหนัก (ขวา) สามารถวอร์ปจุดใด ๆ บนเส้นของ

อนุกรมไปยังตำแหน่งที่เหมาะสมได้ดีกว่า อัลกอริทึมของไดนามิกไทม์วอร์ปิงแบบดั้งเดิม (ซ้าย) ทั้งยังสามารถแก้ปัญหาการวอร์ปไปยังจุด ๆ เดียวได้อีกด้วย



รูปที่ 2.9 แสดงแนวการปรับตรงของไดนามิกไทม์วอร์ปิง (ซ้าย) และแนวการปรับตรงของไดนามิกไทม์วอร์ปิงแบบถ่วงน้ำหนัก (ขวา)

(ที่มา : Y. -S. Jeong และคณะ (2))

อย่างไรก็ตาม การจะหาค่าที่เหมาะสมที่สุดให้กับตัวแปรเสริม g ในอัลกอริทึมดังกล่าว ยังไม่มีวิธีการที่แน่ชัด ซึ่งต่อมาในปี 2017 Y.Wan และคณะ [19] ได้นำแนวคิดดังกล่าวไปปรับปรุงต่อภายใต้ชื่อ Adaptive Cost Dynamic Time Warping (AC-DTW) สามารถคำนวณได้ตามสมการ (2.11) – (2.14) แต่ยังคงติดปัญหาเรื่องการหาค่าที่เหมาะสมที่สุดให้กับตัวแปรเสริมเช่นเคย อีกทั้งยังมีปัญหาเรื่องเวลาที่ใช้ที่ มากเกินกว่าไดนามิกไทม์วอร์ปิงแบบดั้งเดิมอย่างมหาศาลอีกด้วย

$$AC - DTW(A, B) = \min \begin{cases} c(b_{i-1, j}) \times d_{i, j} + AC - DTW(i - 1, j) \\ d_{i, j} + AC - DTW(i - 1, j - 1) \\ c(a_{i, j-1}) \times d_{i, j} + AC - DTW(i, j - 1) \end{cases} \quad (2.11)$$

$$d_{i, j} = (a_i - b_j)^2 \quad (2.12)$$

$$c(x) = g \cdot r \cdot x + 1 \quad (2.13)$$

$$r = \frac{\min(n, m)}{\max(n, m)} \quad (2.14)$$

บทที่ 3 ไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะ

จากงานวิจัยที่เกี่ยวข้อง [2, 17, 19] เป็นที่น่าสนใจว่าแม้แนวคิดเรื่องการถ่วงน้ำหนักให้กับไดนามิกไทม์วอร์ปิงสามารถเพิ่มประสิทธิภาพให้แก่ไดนามิกไทม์วอร์ปิงเมื่อนำไปใช้ในการจำแนกประเภทของข้อมูลอนุกรมเวลาได้ แต่วิธีการในการหาค่าถ่วงน้ำหนักที่เหมาะสม อีกทั้งยังใช้เวลาในการประมวลผลที่ใกล้เคียงกันกับไดนามิกไทม์วอร์ปิงแบบดั้งเดิมนั้นยังไม่สามารถทำได้ในงานวิจัยที่กล่าวมา วิทยานิพนธ์นี้จึงได้นำเสนอไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะที่ใช้เวลาในการประมวลผลเทียบเท่ากับไดนามิกไทม์วอร์ปิงแบบดั้งเดิม และให้ความแม่นยำเมื่อนำไปจำแนกประเภทของข้อมูลที่มากขึ้นในภาพรวมด้วย

ในบทที่ 3 นี้ จะนำเสนอการจัดเตรียมข้อมูล ไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะ ความซับซ้อนของเวลา รวมถึงการวัดผลมาตรฐานวัดระยะห่างที่สร้างขึ้น

3.1 การจัดเตรียมข้อมูล

ในขั้นตอนของการจัดเตรียมข้อมูลนั้นเป็นการเตรียมความพร้อมของข้อมูลก่อนที่จะนำไปใช้ในการจำแนกประเภทของข้อมูลเพื่อให้เกิดความเป็นบรรทัดฐาน เพื่อลดความคลาดเคลื่อนจากข้อผิดพลาดให้น้อยที่สุดเมื่อนำไปใช้ในการจำแนกประเภทของข้อมูล

การแปลงข้อมูลอนุกรมเวลาให้เป็นบรรทัดฐานถือเป็นสิ่งจำเป็นของการจัดเตรียมข้อมูลอย่างหนึ่ง เนื่องจากข้อมูลที่เก็บมาอาจมีความผิดเพี้ยนในมาตราส่วน (scale) เมื่อนำไปใช้กับมาตรวัดระยะห่างเพื่อทำไปจำแนกประเภทของข้อมูลจึงอาจส่งผลกระทบต่อความแม่นยำในการจำแนก ซึ่งการทำให้เป็นบรรทัดฐานนั้นสามารถทำได้โดยการปรับมาตราส่วนและแอมพลิจูด (amplitude) ของข้อมูลให้อยู่ในระดับเดียวกัน โดยงานวิจัยนี้เลือกใช้คะแนน Z (Z -normalization) [7] ในการแปลงอนุกรมเวลาให้เป็นบรรทัดฐานดังแสดงในรูป 3.1 และ 3.2

วิธีการแปลงอนุกรมเวลาให้เป็นบรรทัดฐานสามารถทำได้โดยใช้คะแนน Z เข้าไปแทนจุดของข้อมูล

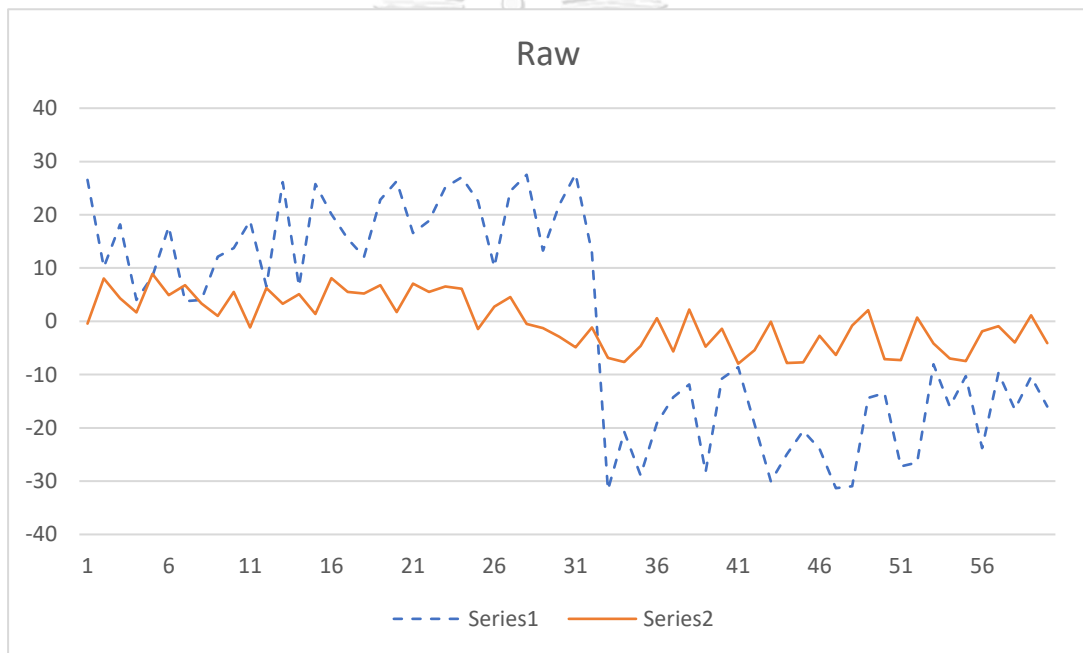
กำหนดให้อนุกรมเวลา $A = (a_1, a_2, \dots, a_n)$ เป็นอนุกรมเวลาที่ต้องการทำให้เป็นบรรทัดฐาน และให้อนุกรม $A' = (a'_1, a'_2, \dots, a'_n)$ เป็นอนุกรมเวลาที่ทำให้เป็นบรรทัดฐานแล้ว a'_i จะสามารถหาได้ดังสมการ (3.1)

$$a'_i = \frac{a_i - \bar{a}}{SD} \quad (3.1)$$

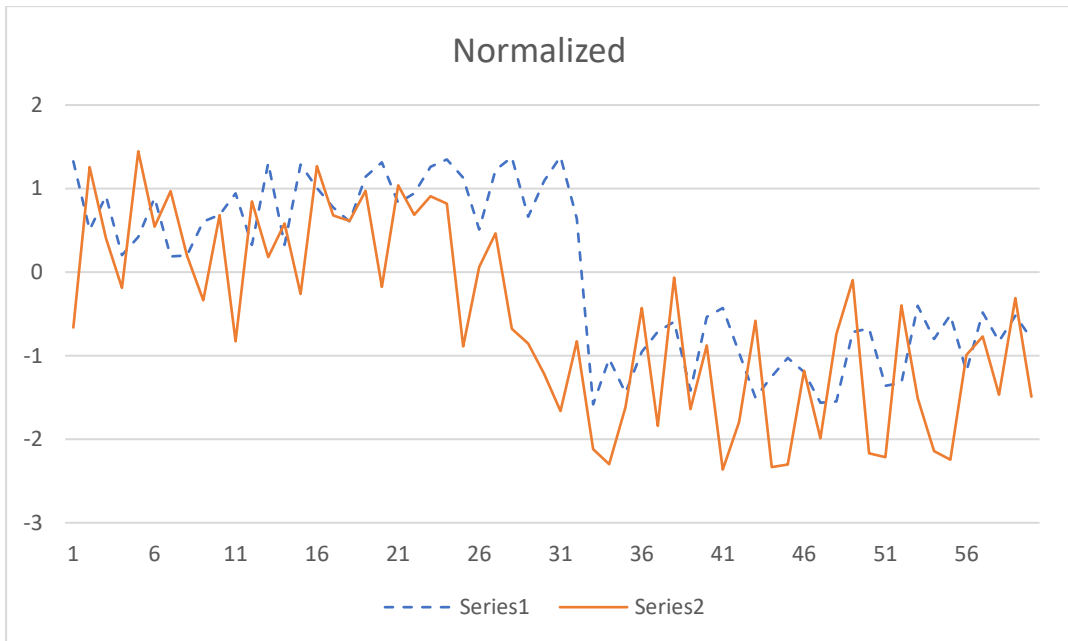
โดยที่ค่า \bar{a} คือค่าเฉลี่ยเลขคณิตของอนุกรมเวลา A และค่า SD คือส่วนเบี่ยงเบนมาตรฐานของค่าอนุกรมเวลา A ซึ่งสามารถคำนวณได้ดังสมการ (3.2) และ (3.3)

$$\bar{a} = \frac{1}{n} \sum_{i=1}^n a_i \quad (3.2)$$

$$SD = \sum_{i=1}^n \sqrt{\frac{(a_i - \bar{a})^2}{n}} \quad (3.3)$$



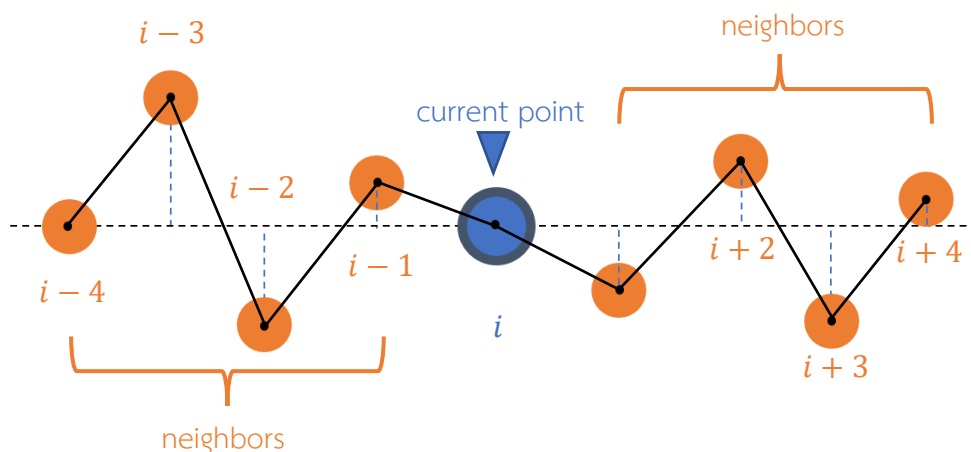
รูปที่ 3.1 แสดงข้อมูลอนุกรมเวลาที่ไม่ถูกทำให้เป็นบรรทัดฐาน



รูปที่ 3.2 แสดงข้อมูลอนุกรมเวลาในรูปที่ 3.1 เมื่อถูกทำให้เป็นบรรทัดฐานแล้ว

3.2 ไดนามิกไทม์วอร์ปปีงถ่วงน้ำหนักแบบเพิ่มสมรรถนะ

จากงานวิจัยที่เกี่ยวข้องซึ่งยังติดปัญหาในเรื่องของค่าถ่วงน้ำหนักที่จะถูกนำมาใช้กับไดนามิกไทม์วอร์ปปีง วิทยานิพนธ์นี้ได้นำเสนอฟังก์ชันซิกมอยด์เข้ามาช่วยในเรื่องการถ่วงน้ำหนักเพื่อทำให้แนวการปรับตรงเป็นไปอย่างถูกต้องมากยิ่งขึ้น เริ่มจากนำค่าของจุดที่สนใจไปลบกับเพื่อนบ้านแต่ละจุดจนครบทั้ง 8 จุดเพื่อนำไปหาค่าเฉลี่ยระหว่างจุดที่สนใจและจุดเพื่อนบ้าน ดังแสดงในรูปที่ 3.3 โดยจำนวนเพื่อนบ้านทั้ง 8 จุดนั้นได้มาจากการทดลองว่าเป็นจำนวนจุดที่เหมาะสมโดยทั่วไป อีกทั้งยังไม่ใช้ระยะเวลาในการประมวลผลที่นานจนเกินไป เพื่อให้ได้ระยะห่างระหว่างจุดที่สนใจและจุดรอบๆ จากนั้นนำไปหาค่าเฉลี่ยดังสมการ (3.4)



รูปที่ 3.3 แสดงจุดข้อมูลที่สนใจและจุดรอบ ๆ จำนวน 8 จุดบนอนุกรมเวลา

$$y_i = \sum_{k=i-4}^{i+4} \left[\frac{(x_i - x_k)}{8} \right]^2 \quad (3.4)$$

เมื่อได้ค่าเฉลี่ยระหว่างจุดที่สนใจและเพื่อนบ้านมาแล้ว จึงนำค่าเฉลี่ยดังกล่าวไปใช้ในฟังก์ชันชกมอยด์ ซึ่งใช้เป็นค่าถ่วงน้ำหนักในลำดับถัดไปในฟังก์ชันวัตรยะทางไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะดังสมการ (3.5) (3.6) และ (3.7)

$$w_i = \frac{1}{1 + e^{-y_i}} \quad (3.5)$$

$$EWDTW(A, B) = \sqrt{\gamma^*(i, j)} \quad (3.6)$$

$$\gamma^*(i, j) = w_{|i-j|} (a_i - b_j)^2 + w_i + \min \begin{cases} \gamma^*(i-1, j) \\ \gamma^*(i-1, j-1) \\ \gamma^*(i, j-1) \end{cases} \quad (3.7)$$

3.3 ความซับซ้อนของเวลา

ในการคำนวณระยะทางโดยใช้มาตรวัตรยะทางเป็นไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะระหว่างอนุกรมเวลาสองอนุกรมที่มีความยาว n และ m ตามลำดับ ไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะจำเป็นต้องคำนวณค่าถ่วงน้ำหนักของแต่ละจุด แต่เนื่องจากค่าถ่วงน้ำหนักดังกล่าวเป็นเพียงการหาค่าเฉลี่ยของระยะทางระหว่างจุดที่สนใจและเพื่อนบ้าน เวลาที่ใช้ในการคำนวณจึงเป็นค่าคงที่ k เมื่อรวมเข้ากับการหาเส้นทางวอร์ป (warping path) ที่ดีที่สุดในเมทริกซ์ $n \times m$ ความซับซ้อนของเวลาที่ใช้สำหรับมาตรวัตรยะทางที่เป็นไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะจึงเป็น $O(mn)$ เมื่อเพิกเฉยต่อค่าคงที่

ดังนั้นความซับซ้อนของเวลาเมื่อใช้มาตรวัตรยะทางเป็นไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะจึงมีค่าเท่ากับไดนามิกไทม์วอร์ปิงแบบดั้งเดิม คือ $O(mn)$

3.4 การวัดผลมาตรวัดระยะห่าง

สำหรับการวัดผลมาตรวัดระยะห่างที่ได้สร้างขึ้น ในวิทยานิพนธ์นี้จะใช้มาตรวัดระยะห่างดังกล่าวในการจำแนกประเภทข้อมูลอนุกรมเวลาจำนวน 20 ชุดข้อมูล ซึ่งเป็นข้อมูลในเว็บไซต์ UCR Time series classification [3] โดยวิธีเพื่อนบ้านที่ใกล้ที่สุดอันดับที่ 1 (1-NN) [4-6] จากนั้นจะทำการวัดผลโดยใช้ความแม่นยำ (accuracy) ที่สามารถคำนวณได้จากสมการที่ (3.8) เปรียบเทียบกับมาตรวัดระยะห่างที่แพร่หลายเพื่อใช้ในการวัดผลความน่าเชื่อถือในการจำแนกประเภทข้อมูลของมาตรวัดระยะห่างโดยที่ผลการทดลองจะถูกแสดงในลำดับถัดไป

$$\text{ความแม่นยำ} = \frac{\text{จำนวนข้อมูลที่จำแนกได้ถูกต้อง}}{\text{จำนวนข้อมูลทั้งหมด}} \quad (3.8)$$



บทที่ 4 การทดลองและวิเคราะห์ผล

ในบทนี้จะกล่าวถึงการประเมินประสิทธิภาพของงานวิจัยที่ได้นำเสนอ โดยจะทำการเปรียบเทียบกับวิธีที่เป็นพื้นฐาน (baseline) อย่างมาตรฐานวัดระยะทางยุคคิดและมาตรฐานวัดระยะทางไดนามิกใหม่เวอร์บปีง รวมถึงทำการเปรียบเทียบกับวิธีที่ใช้หลักการคล้ายคลึงกันอย่างไดนามิกใหม่เวอร์บปีงแบบถ่วงน้ำหนักด้วย

4.1 ชุดข้อมูลที่ใช้ในงานวิจัย

ข้อมูลอนุกรมเวลาที่ใช้ในการทดลองในงานวิจัยนี้ ประกอบด้วยข้อมูลอนุกรมเวลาจากทั้งหมด 19 ชุดจาก UCR Timeseries Classification Archive [3] ซึ่งเป็นข้อมูลอนุกรมเวลาที่มีการเปิดเผยเพื่อใช้สำหรับงานวิจัยที่เกี่ยวข้องกับอนุกรมเวลาโดยข้อมูลดังกล่าวจะถูกแบ่งออกเป็นข้อมูลสอนและข้อมูลทดสอบซึ่งมีความยาวที่หลากหลายและแตกต่างกันออกไป รวมถึงจำนวนคลาส (class) ความยาวของอนุกรมเวลา จำนวนของข้อมูลทดสอบและข้อมูลสอน ดังในตาราง 4.1 ที่แสดงรายละเอียดของข้อมูลต่าง ๆ ที่ใช้ในงานวิจัยนี้

ตารางที่ 4.1 รายละเอียดของข้อมูลที่ใช้ในการทดลอง

ชื่อชุดข้อมูล	จำนวนคลาส	จำนวนข้อมูลสอน	จำนวนข้อมูลทดสอบ	ความยาวของอนุกรมเวลา
Synthetic control	6	300	300	60
Gun Point	2	50	150	150
CBF	3	30	900	128
FaceAll	14	560	1690	131
OSULeaf	6	200	242	427
SwedishLeaf	15	500	625	128
50words	50	450	455	270
Trace	4	100	100	275
Two Patterns	4	1000	4000	128
FaceFour	4	24	88	350
Lightning-2	2	60	61	637

Lightning-7	7	70	73	319
Adiac	37	390	391	176
FISH	7	175	175	463
Beef	5	30	30	470
Coffee	2	28	28	286
Olive Oil	4	30	30	570
Wafer	2	1000	6174	152
yoga	2	300	3000	426

4.2 การทดสอบความแม่นยำของมาตรวัดระยะห่างเมื่อนำไปใช้ในการจำแนกประเภทของข้อมูล

ในส่วนของการทดสอบความแม่นยำของมาตรวัดระยะห่างในวิทยานิพนธ์เล่มนี้ จะทำโดยวิธีการเปรียบเทียบความแม่นยำของมาตรวัดความเหมือนที่ได้นำเสนอขึ้นมาใหม่ (Enhanced Weighted Dynamic Time Warping) เทียบกับมาตรวัดระยะห่างอื่น ๆ ที่มีอยู่รวมไปถึงวิธีที่เป็นพื้นฐาน (baseline) เมื่อนำไปใช้ในการจำแนกประเภทของข้อมูล โดยจะใช้วิธีการจำแนกประเภทของข้อมูลเป็นแบบเพื่อนบ้านใกล้สุดลำดับที่ 1 [4-6] ในการจำแนกข้อมูลทดสอบ ซึ่งเป็นข้อมูลที่ไม่เคยเห็นมาก่อนในการทดสอบ

นอกเหนือจากการนำมาตรวัดระยะห่างที่ได้นำเสนอขึ้นมาใหม่ไปใช้ในการเปรียบเทียบกับวิธีการที่มีอยู่ก่อนแล้ว ในวิทยานิพนธ์นี้ยังได้ทำการปรับฟังก์ชันถ่วงน้ำหนักที่ใช้ และได้ทำการเปรียบเทียบกันเอง เพื่อจะแสดงให้เห็นถึงความเหมาะสมของการเลือกใช้ฟังก์ชันในข้อมูลประเภทต่าง ๆ ได้อย่างชัดเจนมากขึ้น

ตารางที่ 4.2 ความแม่นยำในการจำแนกประเภทของข้อมูลโดยวิธีการเพื่อนบ้านที่ใกล้ที่สุดลำดับที่ 1
 ควบคู่กับมาตรวัดระยะห่างที่เป็นระยะทางยุคลิด ระยะทางไดนามิกไทม์-
 วอร์ปิง และไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะ

ชื่อชุดข้อมูล	ความแม่นยำ (%)		
	ระยะทางยุคลิด	ไดนามิกไทม์วอร์ปิง	ไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะ
Synthetic control	88	99.3	98.7
Gun Point	91.3	90.7	94.7
CBF	85.2	99.7	99.8
FaceAll	71.4	80.8	80.6
OSULeaf	52.1	59.1	70.2
SwedishLeaf	78.9	79.2	89.6
50words	63.1	69	81.3
Trace	76	100	83
Two Patterns	91	100	100
FaceFour	78.4	83	89.8
Lightning-2	75.4	86.9	93.4
Lightning-7	57.5	72.6	75.3
Adiac	61.1	60.4	61.6
FISH	78.3	82.3	83.4
Beef	66.7	63.3	66.7
Coffee	100	100	100
Olive Oil	86.7	83.3	86.7
Wafer	99.5	99.6	99.4
yoga	83	83.5	85.8

จากผลการทดลองในตาราง 4.2 จะเห็นได้ว่าไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะให้ความแม่นยำในการจำแนกประเภทของข้อมูลสูงสุดในภาพรวม เปรียบเทียบกับไดนามิก

ไทม์วอร์ปิงแบบดั้งเดิมและระยะทางยูคลิด และยังคงให้ความแม่นยำที่สูงที่สุดในภาพรวมเมื่อนำไปเปรียบเทียบกับไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักในกรณีที่ดีที่สุดดังตาราง 4.3

ตารางที่ 4.3 ความแม่นยำในการจำแนกประเภทของข้อมูลโดยวิธีการเพื่อนบ้านที่ใกล้ที่สุดลำดับที่ 1 ควบคู่กับมาตรวัดระยะทางไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักในกรณีที่ดีที่สุดและไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะ

ชื่อชุดข้อมูล	ความแม่นยำ (%)	
	ไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะ	ไดนามิกไทม์วอร์ปิงถ่วงน้ำหนัก
Synthetic control	98.7	99.8
Gun Point	94.7	99.6
CBF	99.8	99.8
FaceAll	80.6	74.3
OSULeaf	70.2	62.8
SwedishLeaf	89.6	86.2
50words	81.3	80.6
Trace	83	100
Two Patterns	100	100
FaceFour	89.8	86.4
Lightning-2	93.4	90
Lightning-7	75.3	80
Adiac	61.6	63.6
FISH	83.4	87.4
Beef	66.7	40
Coffee	100	86.7
Olive Oil	86.7	81.2
Wafer	99.4	99.8
Yoga	85.8	83.5

จากผลการทดลองที่แสดงในตารางที่ 4.3 จะเห็นได้ว่า กระทั่งในกรณีที่ไดนามิกไทม์วอร์ปิงถ่วงน้ำหนัก [2] สามารถเรียนรู้ได้ค่าถ่วงน้ำหนักที่ดีที่สุดออกมา ความแม่นยำในการจำแนกประเภทของข้อมูลโดยรวมก็ยังต่ำกว่าไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะที่วิทยานิพนธ์นี้ได้นำเสนอ

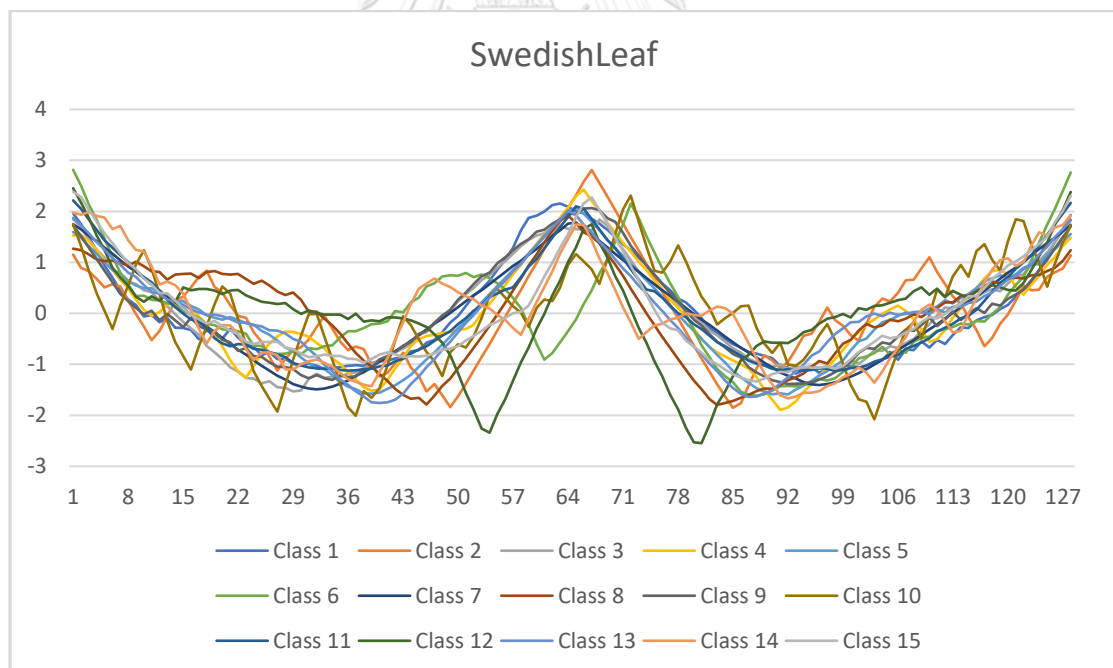
ตารางที่ 4.4 ความแม่นยำในการจำแนกประเภทของข้อมูลโดยวิธีการเพื่อนบ้านที่ใกล้ที่สุดลำดับที่ 1 ควบคู่กับมาตรวัดระยะห่างไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะซึ่งมีค่าถ่วงน้ำหนักเป็นฟังก์ชันซิกมอยด์และฟังก์ชันเอกซ์โพเนนเชียล

ชื่อชุดข้อมูล	ความแม่นยำ (%)	
	ไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะที่มีค่าถ่วงน้ำหนักเป็นฟังก์ชันซิกมอยด์	ไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะที่มีค่าถ่วงน้ำหนักเป็นฟังก์ชันเอกซ์โพเนนเชียล
Synthetic control	98.7	98.7
Gun Point	94.7	94.7
CBF	99.8	100
FaceAll	80.6	81.4
OSULeaf	70.2	70.2
SwedishLeaf	89.6	89.1
50words	81.3	81.3
Trace	83	83
Two Patterns	100	100
FaceFour	89.8	88.6
Lightning-2	93.4	93.4
Lightning-7	75.3	75.3
Adiac	61.6	61.1
FISH	83.4	83.4
Beef	66.7	66.7
Coffee	100	100
Olive Oil	86.7	86.7

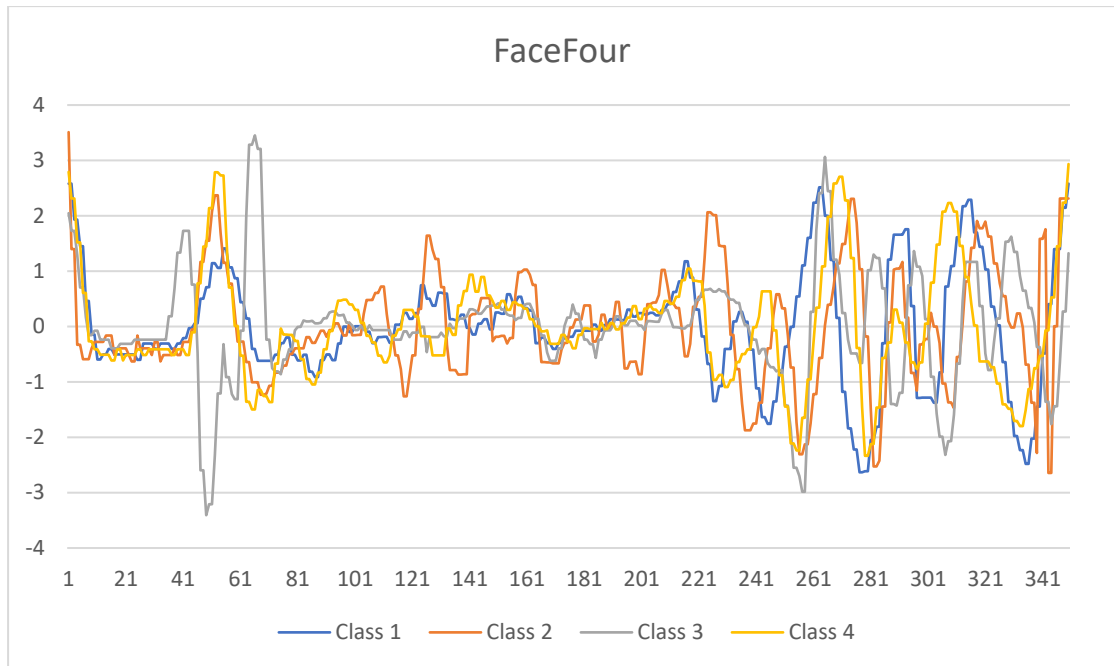
Wafer	99.4	99.4
yoga	85.8	85.8

จากตารางที่ 4.4 จะเห็นได้ว่าการปรับเปลี่ยนฟังก์ชันถ่วงน้ำหนักบนมาตรวัดระยะห่างไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะมีผลต่อความแม่นยำในการจำแนกประเภทของข้อมูลบางชนิด และจากผลการทดลองในตาราง สามารถแบ่งได้เป็น 2 กรณี คือ เมื่อฟังก์ชันถ่วงน้ำหนักเป็นซิกมอยด์ฟังก์ชันแล้วให้ความแม่นยำในการจำแนกประเภทของข้อมูลที่สูงกว่า กับกรณีที่เมื่อฟังก์ชันถ่วงน้ำหนักเป็นฟังก์ชันเอกซ์โพเนนเชียลแล้วให้ความแม่นยำในการจำแนกข้อมูลที่สูงกว่า

กรณีที่ 1 เมื่อฟังก์ชันถ่วงน้ำหนักเป็นซิกมอยด์ฟังก์ชันให้ความแม่นยำในการจำแนกข้อมูลที่สูงกว่า จากผลการทดลองที่ได้แสดงในตารางที่ 4.4 จะเห็นได้ว่า ชุดข้อมูลที่ตรงกับกรณีดังกล่าว ได้แก่ชุดข้อมูล Swedis Leaf, FaceFour และ Adiac ซึ่งตัวอย่างลักษณะของข้อมูลในแต่ละประเภทโดยคร่าวแสดงได้ดังรูปที่ 4.1 และ 4.2 โดยอนุกรมเวลาแต่ละเส้นจะแสดงให้เห็นถึงลักษณะของข้อมูลในแต่ละประเภท (class) ของชุดข้อมูลนั้นๆ

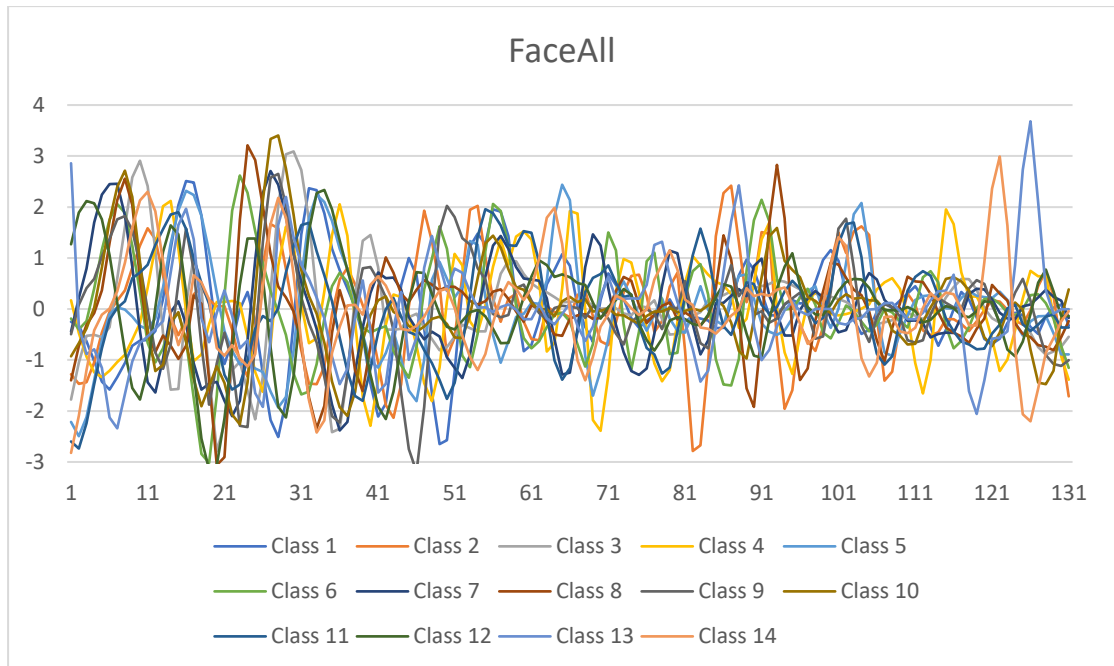


รูปที่ 4.1 แสดงลักษณะของข้อมูลแต่ละประเภทในชุดข้อมูล SwedishLeaf

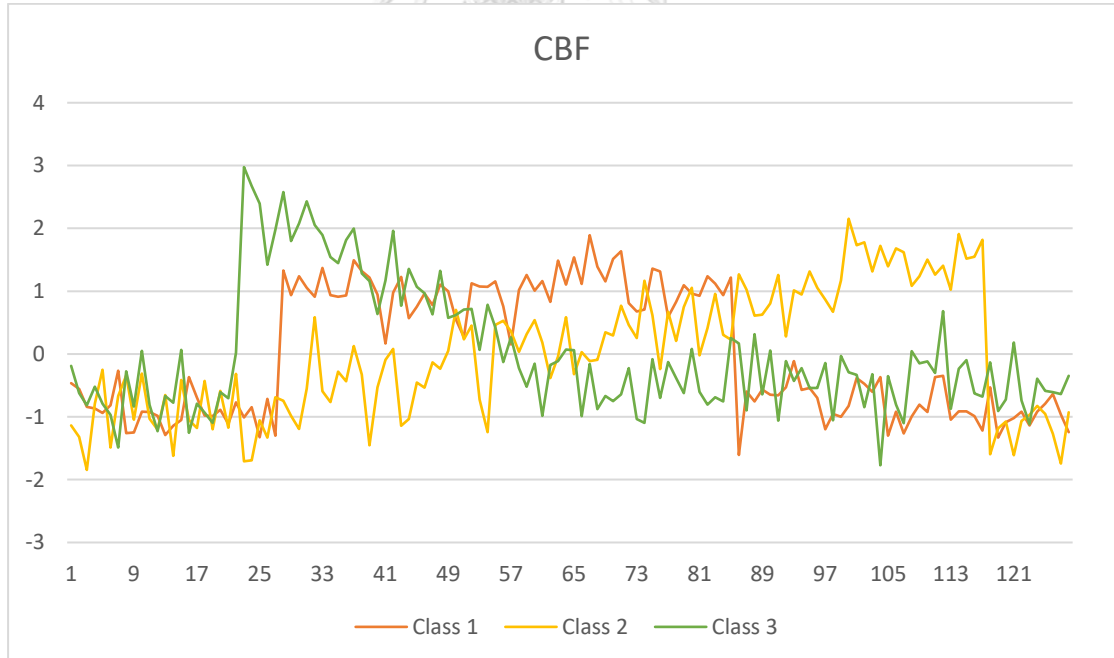


รูปที่ 4.2 แสดงลักษณะของข้อมูลแต่ละประเภทในชุดข้อมูล FaceFour

กรณีที่ 2 เมื่อฟังก์ชันถ่วงน้ำหนักเป็นเอกซ์โพเนนเชียลฟังก์ชันซึ่งให้ความแม่นยำในการจำแนกข้อมูลที่สูงกว่า จากผลการทดลองที่ได้แสดงในตาราง 4.4 จะเห็นได้ว่า ชุดข้อมูลที่ตรงกับกรณีดังกล่าว ได้แก่ชุดข้อมูล FaceAll และ CBF ซึ่งตัวอย่างลักษณะของข้อมูลในแต่ละประเภทโดยคร่าวแสดงได้ดังรูปที่ 4.3 และ 4.4 โดยอนุกรมเวลาแต่ละเส้นจะแสดงให้เห็นถึงลักษณะของข้อมูลในแต่ละประเภท (class) ของชุดข้อมูลนั้นๆ



รูปที่ 4.3 แสดงลักษณะของข้อมูลแต่ละประเภทในชุดข้อมูล FaceAll



รูปที่ 4.4 แสดงลักษณะของข้อมูลแต่ละประเภทในชุดข้อมูล CBF

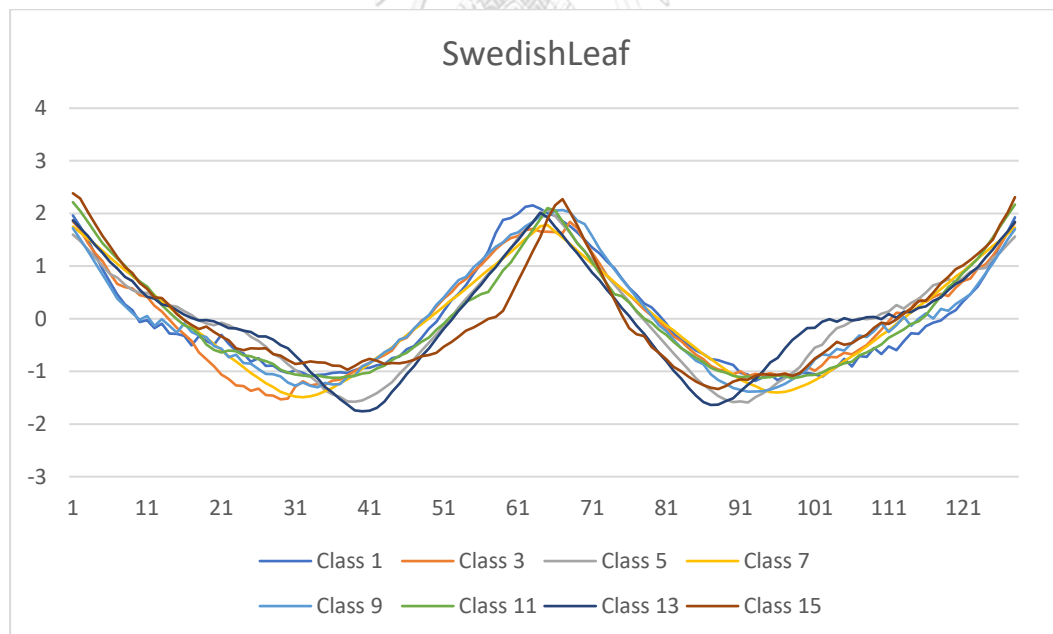
เมื่อเปรียบเทียบระหว่างกรณีที่ 1 และกรณีที่ 2 จะเห็นได้ว่ารูป 4.1 และ 4.2 ที่แม้จะมีความใกล้เคียงในการเรียงตัวของข้อมูลแต่ละคลาสอยู่มาก แต่เมื่อนำกราฟของสองข้อมูลดังกล่าวไป

เปรียบเทียบกับรูปที่ 4.3 และ 4.4 จะเห็นได้ว่าการแกว่งของข้อมูลที่น้อยกว่าสองรูปหลังที่ข้อมูลดีดค่าขึ้นลงสลับไปมาอยู่ตลอดอนุกรมเวลา

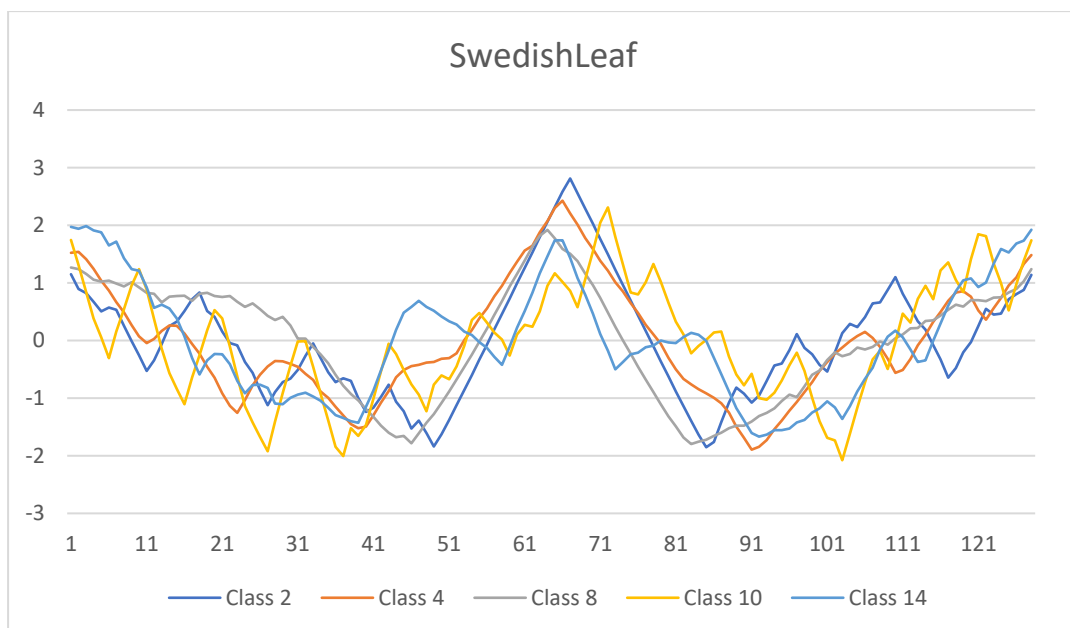
เมื่ออ้างอิงจากผลการทดลองในตารางที่ 4.4 พบว่าบนชุดข้อมูลที่มีความใกล้เคียงของข้อมูลระหว่างคลาส ฟังก์ชันถ่วงน้ำหนักที่เลือกใช้แล้วสามารถให้ความแม่นยำในการจำแนกประเภทของข้อมูลที่ดีกว่าจะเป็นฟังก์ชันซิกมอยด์ ขณะเดียวกันบนชุดข้อมูลที่มีข้อมูลมีการแกว่งตัวมากๆ ฟังก์ชันถ่วงน้ำหนักที่เป็นฟังก์ชันเอกซ์โพเนนเชียลจะให้ความแม่นยำในการจำแนกประเภทของข้อมูลที่สูงกว่า

เพื่อยืนยันข้อสรุปดังกล่าว วิทยานิพนธ์ฉบับนี้ได้ทำการทดลองเพิ่มเติมเพื่อให้ผลสรุปดังกล่าวเป็นที่กระจ่างยิ่งขึ้น โดยได้เลือกทำการทดลองบนชุดข้อมูลที่สอดคล้องกับข้อสันนิษฐานและการสรุปผลดังที่กล่าวมาดังต่อไปนี้

บนชุดข้อมูล SwedishLeaf ที่ข้อมูลแต่ละคลาสมีการเรียงตัวที่ใกล้เคียงกัน ข้อมูลที่มีหน้าตาใกล้เคียงกันจะแบ่งออกเป็น 2 ชุดเพื่อใช้ในการทดสอบความแม่นยำ ชุดข้อมูลที่ 1 ได้แก่ คลาส 1, 3, 5, 7, 9, 11, 13 และ 15 โดยลักษณะของข้อมูลแสดงดังรูปที่ 4.5 และชุดข้อมูลที่ 2 ได้แก่ คลาส 2, 4, 8, 10 และ 14 ลักษณะของข้อมูลแสดงดังรูปที่ 4.6



รูปที่ 4.5 แสดงลักษณะของข้อมูลที่ใกล้เคียงกันระหว่างประเภทบนชุดข้อมูล SwedishLeaf ชุดที่ 1



รูปที่ 4.6 แสดงลักษณะของข้อมูลที่ใกล้เคียงกันระหว่างประเภทบนชุดข้อมูล SwedishLeaf ชุดที่ 2

ผลการทดลองบนข้อมูล SwedishLeaf ที่ถูกแบ่งออกเป็น 2 ชุดตามความใกล้เคียงทางลักษณะของข้อมูล แสดงในตารางที่ 4.5

ตารางที่ 4.5 ความแม่นยำในการจำแนกประเภทของข้อมูลโดยวิธีการเพื่อนบ้านที่ใกล้ที่สุดลำดับที่ 1 ควบคู่กับมาตรวัดระยะห่างที่เป็นไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะซึ่งมีค่าถ่วงน้ำหนักเป็นฟังก์ชันซิกมอยด์และฟังก์ชันเอกซ์โพเนนเชียลบนข้อมูล SwedishLeaf ที่ถูกแบ่งตามลักษณะของข้อมูล

ชุดข้อมูล	ความแม่นยำ (%)	
	ฟังก์ชันถ่วงน้ำหนักเป็นฟังก์ชันซิกมอยด์	ฟังก์ชันถ่วงน้ำหนักเป็นฟังก์ชันเอกซ์โพเนนเชียล
ชุดข้อมูลที่ 1	86.1	85.2
ชุดข้อมูลที่ 2	92	92

จากผลการทดลองดังแสดงในตารางที่ 4.5 จะเห็นได้ว่า บนชุดข้อมูลที่มีการแกว่งของข้อมูลน้อยและมีลักษณะของข้อมูลระหว่างคลาสที่ใกล้เคียงกัน ไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่ม

สมรรถนะที่มีฟังก์ชันถ่วงน้ำหนักเป็นซิกมอยด์จะสามารถให้ความแม่นยำที่สูงกว่ากรณีที่ใช้ฟังก์ชันถ่วงน้ำหนักเป็นฟังก์ชันเอกซ์โพเนนเชียล

บนชุดข้อมูล FaceAll ที่มีการแกว่งของข้อมูลในหลายๆคลาสค่อนข้างมาก การทดลองจึงถูกแบ่งออกเป็น 2 ส่วนคือในส่วนแรก ใช้เพื่อทดสอบว่าข้อมูลประเภทใดที่ส่งผลให้ความแม่นยำในการจำแนกประเภทข้อมูลของฟังก์ชันซิกมอยด์และฟังก์ชันเอกซ์โพเนนเชียลแตกต่างกัน ส่วนที่สองเพื่อใช้ทดสอบและยืนยันว่าบนชุดข้อมูลดังกล่าว ฟังก์ชันเอกซ์โพเนนเชียลจะสามารถให้ความแม่นยำในการจำแนกประเภทของข้อมูลที่สูงกว่าฟังก์ชันซิกมอยด์จริง

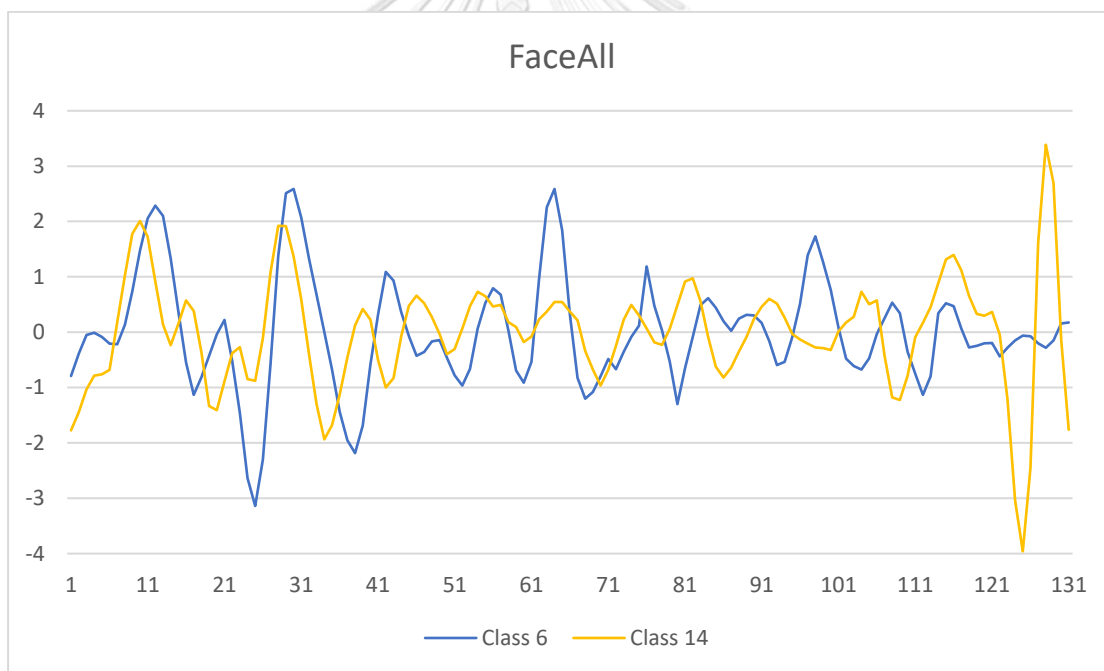
ตารางที่ 4.6 ความแม่นยำในการจำแนกประเภทของข้อมูลโดยวิธีการเพื่อนบ้านที่ใกล้ที่สุดลำดับที่ 1 ควบคู่กับมาตรวัดระยะห่างที่เป็นไดนามิกไทม์วอร์ปปีงถ่วงน้ำหนักแบบเพิ่มสมรรถนะซึ่งมีค่าถ่วงน้ำหนักเป็นฟังก์ชันซิกมอยด์และฟังก์ชันเอกซ์โพเนนเชียลบนข้อมูล FaceAll ที่ข้อมูลบางประเภทถูกตัดออกไป

ประเภทของข้อมูล (class) ที่ถูกตัดออก	ความแม่นยำ (%)	
	ฟังก์ชันถ่วงน้ำหนักเป็นฟังก์ชันซิกมอยด์	ฟังก์ชันถ่วงน้ำหนักเป็นฟังก์ชันเอกซ์โพเนนเชียล
1	98.1	99
2	98.1	99
3	98.1	99
4	98.1	99
5	98.1	99
6	99	100
7	98.1	99
8	98.1	99
9	98.1	99
10	98.1	99
11	98.1	99
12	98.1	99
13	98.1	99
14	99	99

จากตารางที่ 4.6 จะเห็นได้ว่าแทบจะในทุกๆชุดข้อมูลที่ถูกต้องจะให้ความแม่นยำที่เท่าๆกันของทั้งสองฟังก์ชัน ยกเว้นกรณีที่เมื่อชุดข้อมูลที่ 6 และชุดข้อมูลที่ 14 ถูกตัดออก

เมื่อชุดข้อมูลที่ 6 ถูกตัดออก จะเห็นได้ว่าฟังก์ชันเอกซ์โพเนนเชียลสามารถจำแนกประเภทของข้อมูลที่เหลืออยู่ได้ด้วยความแม่นยำ 100% ในขณะที่ฟังก์ชันซิกมอยด์สามารถจำแนกประเภทของข้อมูลได้ด้วยความแม่นยำ 99% เพิ่มจากเดิมที่สามารถจำแนกได้เมื่อมีชุดข้อมูลที่ 6 อยู่ 0.9%

ขณะเดียวกันเมื่อชุดข้อมูลที่ 14 ถูกตัดออก ทั้งฟังก์ชันซิกมอยด์และฟังก์ชันเอกซ์โพเนนเชียลจะสามารถจำแนกประเภทของข้อมูลได้ด้วยความแม่นยำที่เท่ากัน ซึ่งนำไปสู่คำถามต่อมาว่าบนชุดข้อมูลดังกล่าว ฟังก์ชันถ่วงน้ำหนักที่เป็นเอกซ์โพเนนเชียลสามารถทำงานได้ดีกว่าฟังก์ชันถ่วงน้ำหนักที่เป็นซิกมอยด์จริงหรือไม่ โดยลักษณะของข้อมูลในคลาสที่ 6 และ 14 จะถูกแสดงดังรูปที่ 4.7 โดยการทดลองในชุดที่ 2 จะใช้ข้อมูลทดสอบเฉพาะข้อมูลคลาสที่ 6 และ 14 ของชุดข้อมูล FaceAll ผลการทดลองดังตารางที่ 4.7



รูปที่ 4.7 แสดงลักษณะของข้อมูลในคลาสที่ 6 และ 14 ของชุดข้อมูล FaceAll

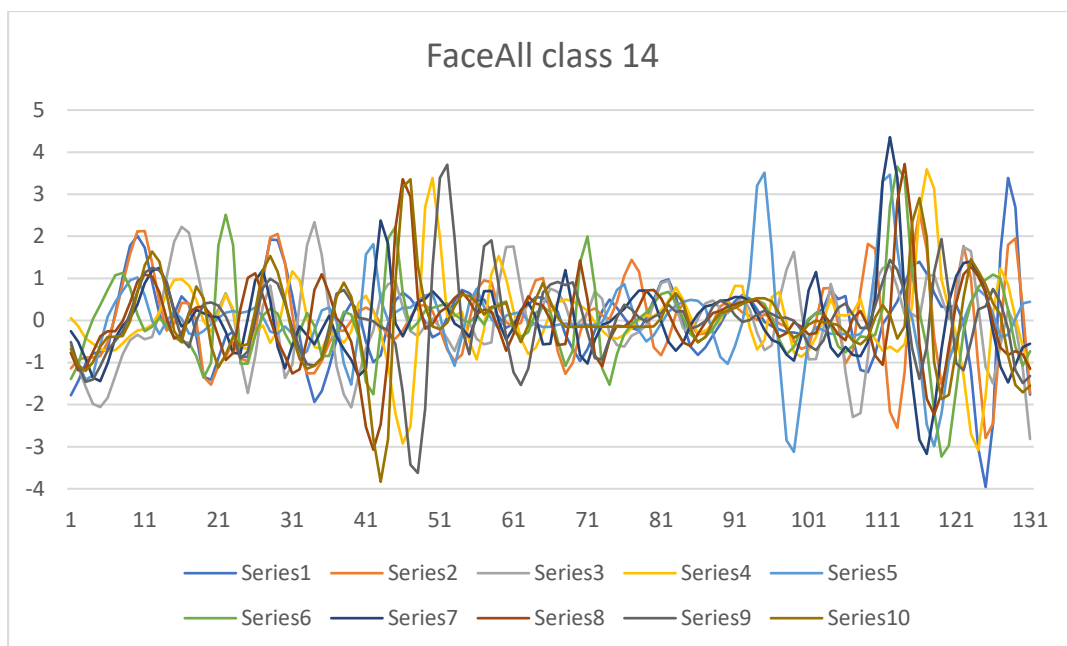
ตารางที่ 4.7 ความแม่นยำในการจำแนกประเภทของข้อมูลเฉพาะประเภทที่ 6 และ 14 โดยวิธีการเพื่อนบ้านที่ใกล้ที่สุดลำดับที่ 1 ควบคู่กับมาตรวัดระยะห่างที่เป็นไดนามิกไทม์วอร์ปปีงถ่วงน้ำหนักแบบเพิ่มสมรรถนะซึ่งมีค่าถ่วงน้ำหนักเป็นฟังก์ชันซิกมอยด์และฟังก์ชันเอกซ์โพเนนเชียล

บนข้อมูล FaceAll

ประเภทของข้อมูล	ความแม่นยำ (%)	
	ฟังก์ชันถ่วงน้ำหนักเป็นฟังก์ชันซิกมอยด์	ฟังก์ชันถ่วงน้ำหนักเป็นฟังก์ชันเอกซ์โพเนนเชียล
6	95.2	95.2
14	87.5	96.9

จากผลการทดลองที่แสดงในตารางที่ 4.6 จะเห็นได้ว่าเมื่อตัดข้อมูลประเภทที่ 6 ออกไดนามิกไทม์วอร์ปปีงถ่วงน้ำหนักแบบเพิ่มสมรรถนะที่มีฟังก์ชันเป็นซิกมอยด์และเอกซ์โพเนนเชียลมีความแม่นยำในการจำแนกประเภทของข้อมูลเพิ่มขึ้นฟังก์ชันละ 1% โดยประมาณ แต่เมื่อตัดข้อมูลประเภทที่ 14 ออกพบว่า ทั้งสองฟังก์ชันสามารถจำแนกประเภทของข้อมูลได้ด้วยความแม่นยำเท่ากันที่ 99%

เมื่อพิจารณาร่วมกับผลการทดลองที่แสดงในตารางที่ 4.7 จะเห็นได้ว่า เมื่อต้องจำแนกข้อมูลประเภทที่ 6 ความแม่นยำในการจำแนกประเภททั้งของฟังก์ชันซิกมอยด์และฟังก์ชันเอกซ์โพเนนเชียลมีค่าเท่ากันที่ 95.2% นั่นหมายถึงความสามารถในการจำแนกข้อมูลประเภทที่ 6 ของทั้งฟังก์ชันซิกมอยด์และฟังก์ชันเอกซ์โพเนนเชียลมีค่าเท่ากัน ขณะที่การจำแนกข้อมูลประเภทที่ 14 ฟังก์ชันเอกซ์โพเนนเชียลสามารถทำได้ดีกว่าถึง 9.4% นั่นหมายถึงฟังก์ชันเอกซ์โพเนนเชียลจะสามารถให้ความแม่นยำในการจำแนกประเภทของข้อมูลที่สูงกว่าในข้อมูลที่มีลักษณะเดียวกันกับข้อมูลประเภทที่ 14 ของชุดข้อมูล FaceAll ดังตัวอย่างแสดงในรูปที่ 4.8



รูปที่ 4.8 แสดงลักษณะของข้อมูลในคลาสที่ 14 ของชุดข้อมูล FaceAll

จากการทดลองเบื้องต้นนำไปสู่การสรุปผลของการเลือกใช้ฟังก์ชันถ่วงน้ำหนักที่เหมาะสมกับประเภทของข้อมูล คือ หากข้อมูลมีการแกว่งตัวมากให้ใช้ฟังก์ชันถ่วงน้ำหนักเป็นฟังก์ชันเอกซ์โพเนนเชียล และในทางตรงกันข้ามหากข้อมูลมีการแกว่งตัวน้อยและมีลักษณะคล้ายคลึงกันในคนละประเภทของข้อมูลให้ใช้ฟังก์ชันถ่วงน้ำหนักเป็นฟังก์ชันซิกมอยด์

4.3 การทดสอบความแม่นยำของมาตรวัดระยะห่างร่วมกับเงื่อนไขบังคับโดยรวม

นอกเหนือจากการทดลองที่กล่าวมา ในระยะเบื้องต้นวิทยานิพนธ์ฉบับนี้ได้ทำการทดสอบความสามารถในของมาตรวัดระยะห่างไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะร่วมกับการกำหนดเงื่อนไขบังคับโดยรวมบนชุดข้อมูล SwedishLeaf และ Lightning-2 ซึ่งได้ทำการทดลองบน training data เพื่อหาเงื่อนไขบังคับโดยรวมที่เหมาะสมบนชุดข้อมูล Swedish Leaf และ Lightning-2 โดยวิธีการ 5-โพลด์ครอสวาไลเดชันก่อนจะนำเงื่อนไขบังคับโดยรวมที่ได้จากการสอนไปใช้ในการทดสอบบนข้อมูลสอนและข้อมูลทดสอบที่ทาง UCR Timeseries Classification Archive (3) กำหนดมาให้ ผลการทดลองแสดงดังตารางที่ 4.8 และ 4.9

ตารางที่ 4.8 แสดงจำนวนหน้าต่างสำหรับการกำหนดเงื่อนไขโดยรวมและความแม่นยำบนมาตรวัดระยะห่างที่เป็นไดนามิกไทม์วอร์ปิงและไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะกับชุด

ข้อมูล SwedishLeaf

	ความแม่นยำ (%) เมื่อไม่กำหนดเงื่อนไขโดยรวม	ความแม่นยำ (%) [จำนวนหน้าต่างของตัวแปรสำหรับการกำหนดเงื่อนไขโดยรวม%]
ไดนามิกไทม์วอร์ปิง	79.2	85.6 [2]
ไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะ	89.6	89.8 [6]

ตารางที่ 4.9 แสดงจำนวนหน้าต่างสำหรับการกำหนดเงื่อนไขโดยรวมและความแม่นยำบนมาตรวัดระยะห่างที่เป็นไดนามิกไทม์วอร์ปิงและไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะกับชุด

ข้อมูล Lighting-2

	ความแม่นยำ (%) เมื่อไม่กำหนดเงื่อนไขโดยรวม	ความแม่นยำ (%) [จำนวนหน้าต่างของตัวแปรสำหรับการกำหนดเงื่อนไขโดยรวม%]
ไดนามิกไทม์วอร์ปิง	86.9	83.6 [4]
ไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะ	93.4	91.8 [9]

จากตารางที่ 4.8 และ 4.9 จะเห็นได้ว่าไดนามิกไทม์วอร์ปิงแบบถ่วงน้ำหนักสามารถใช้ร่วมกับการกำหนดเงื่อนไขโดยรวมได้เช่นเดียวกับไดนามิกไทม์วอร์ปิงแบบดั้งเดิม ซึ่งหากเปรียบเทียบกับกรจำแนกประเภทตามปกติโดยไม่กำหนดเงื่อนไขโดยรวม

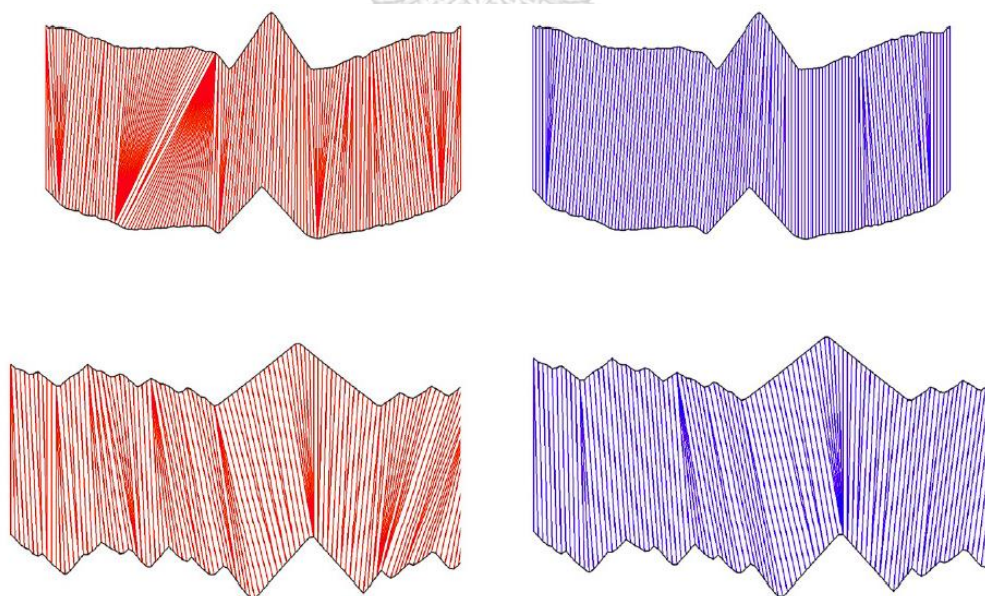
บทที่ 5 สรุปผลงานวิจัยและข้อเสนอแนะ

5.1 สรุปผลงานวิจัย

การจำแนกประเภทของอนุกรมเวลาโดยใช้วิธีการเพื่อนบ้านที่ใกล้ที่สุดลำดับที่ 1 เป็นหนึ่งในวิธีการที่ทรงประสิทธิภาพโดยเฉพาะในยามที่ใช้ควบคู่กันกับมาตรวัดระยะห่างที่เป็นไดนามิกไทม์วอร์ปิงซึ่งมีความสามารถในการรับมือต่อข้อมูลที่มีการเคลื่อนไหวในแนวแกนเวลาทำให้ส่งผลต่อความแม่นยำในการจำแนกประเภทของข้อมูลในทางที่ดีขึ้นอย่างมาก

อย่างไรก็ตามความสามารถดังกล่าวของไดนามิกไทม์วอร์ปิงยังมีข้อเสียอยู่บนข้อมูลอนุกรมเวลาบางประเภท ทำให้เกิดการจับคู่ (matching) ที่เป็นไปอย่างที่ไม่ควรจะเป็นและนำไปสู่ปัญหาที่หลาย ๆ จุดบนอนุกรมเวลาจับคู่กับจุดเพียงจุดเดียวบนอีกอนุกรมเวลาหนึ่งเท่านั้น ส่งผลให้เกิดความผิดพลาดเมื่อนำไปใช้ในการจำแนกประเภท

งานวิจัยนี้จึงได้นำเสนอมาตรวัดระยะห่างที่ยังคงคุณสมบัติอันเป็นข้อดีของไดนามิกไทม์วอร์ปิงแบบดั้งเดิม ทั้งยังสามารถรับมือต่อข้อมูลบางประเภทที่ทำให้เกิดปัญหาการที่จุดหลาย ๆ จุดบนอนุกรมเวลาหนึ่งจับคู่กับจุดเพียงจุดเดียวบนอีกอนุกรมเวลาซึ่งเกิดในไดนามิกไทม์วอร์ปิงแบบดั้งเดิมได้ ดังแนวปรับตรงของไดนามิกไทม์วอร์ปิงและไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะที่แสดงในรูปที่ 5.1



รูปที่ 5.1 แสดงแนวการปรับตรงระหว่างอนุกรมเวลาที่อยู่ในประเภทเดียวกันโดยมีมาตรวัดระยะห่างเป็นไดนามิกไทม์วอร์ปิงแบบดั้งเดิม (สีแดง) และไดนามิกไทม์วอร์ปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะ (สีน้ำเงิน)

ในส่วนของความแม่นยำของมาตรวัดระยะห่างที่นำเสนอเมื่อนำไปใช้ร่วมกับวิธีการเพื่อนบ้านที่ใกล้ที่สุดลำดับที่ 1 จากผลการทดลองในตารางที่ 4.2 ในบทที่ 4 จะเห็นได้ว่าไดนามิกโทมวอร์ปปีงถ่วงน้ำหนักแบบเพิ่มสมรรถนะสามารถให้ความแม่นยำที่ดีกว่าระยะทางยูคลิดและไดนามิกโทมวอร์ปปีงแบบดั้งเดิม อีกทั้งบนบางข้อมูลอนุกรมเวลาที่ระยะทางยูคลิดสามารถให้ความแม่นยำได้สูงกว่าไดนามิกโทมวอร์ปปีงแบบดั้งเดิม [1] มาตรวัดระยะห่างที่วิหยนิพนธ์นี้ได้นำเสนอก็สามารถให้ความแม่นยำที่สูงกว่าทั้งระยะทางยูคลิด [8, 9] และไดนามิกโทมวอร์ปปีง [1] อีกด้วย

นอกจากนี้บนอนุกรมเวลาบางประเภท ฟังก์ชันถ่วงน้ำหนักของไดนามิกโทมวอร์ปปีงถ่วงน้ำหนักแบบเพิ่มสมรรถนะยังสามารถปรับเปลี่ยนได้ตามความเหมาะสม ดังอธิบายในบทที่ 4 เพื่อให้ได้ประสิทธิภาพที่ดียิ่งขึ้น ซึ่งเมื่อพิจารณาโดยภาพรวม ไดนามิกโทมวอร์ปปีงถ่วงน้ำหนักแบบเพิ่มสมรรถนะสามารถให้ความแม่นยำเมื่อนำไปใช้ร่วมกับวิธีการเพื่อนบ้านที่ใกล้ที่สุดลำดับที่ 1 ได้เป็นอย่างดีในความสัมพันธ์ของเวลาที่เท่ากันกับไดนามิกโทมวอร์ปปีงด้วย

5.2 ข้อเสนอแนะ

ข้อเสนอแนะต่อไปนี้เป็นแนวทางในการวิจัยที่จะนำเสนอเพื่อพัฒนามาตรวัดระยะห่างไดนามิกโทมวอร์ปปีงถ่วงน้ำหนักแบบเพิ่มสมรรถนะให้ทรงประสิทธิภาพมากยิ่งขึ้น โดยแบ่งประเด็นที่ต้องการพัฒนาออกเป็น 2 ประเด็นหลัก

ประเด็นที่หนึ่ง คือ การเลือกใช้ฟังก์ชันอื่น ๆ เป็นฟังก์ชันถ่วงน้ำหนัก จากผลการทดลองที่ได้แสดงในบทที่ 4 ซึ่งกล่าวถึงการเลือกใช้ฟังก์ชันซิกมอยด์และฟังก์ชันเอกซ์โพเนนเชียลบนข้อมูลต่าง ๆ จะเห็นได้ว่าการเลือกฟังก์ชันสำหรับมาตรวัดระยะห่างไดนามิกโทมวอร์ปปีงถ่วงน้ำหนักแบบเพิ่มสมรรถนะส่งผลต่อความแม่นยำเมื่อนำไปใช้จำแนกประเภทของข้อมูล โดยในบทที่ 4 ได้ทำการอธิบายไว้ว่าบนข้อมูลที่มีการแกว่งตัวสูงให้ใช้ฟังก์ชันเอกซ์โพเนนเชียลเป็นฟังก์ชันถ่วงน้ำหนัก แต่การจะบอกว่าข้อมูลใดแกว่งตัวมากน้อยแค่ไหนนั้น มาจากการสังเกตข้อมูลต่าง ๆ โดยภาพรวมซึ่งอาจทำให้เกิดความลำเอียงในจะบอกว่าข้อมูลดังกล่าวแกว่งตัวในระดับที่เหมาะสมกับฟังก์ชันใดมากกว่ากัน ดังนั้นในการจะบอกว่าข้อมูลใดที่มีการแกว่งตัวของข้อมูลมากหรือน้อย ควรที่จะต้องมีการวัดบางอย่างในการระบุว่าข้อมูลดังกล่าวมีการแกว่งตัวที่มากพอในการจะเลือกใช้ฟังก์ชันเอกซ์โพเนนเชียลหรือฟังก์ชันซิกมอยด์ โดยอาจเป็นค่าขีดแบ่ง (threshold) ซึ่งได้มาจากการคำนวณในแต่ละชุดข้อมูล

ประเด็นที่สอง การหาค่าเฉลี่ยรอบ ๆ จุดที่สนใจ เนื่องจากตามมาตรวัดระยะห่างแบบที่ได้นำเสนอไปนั้น เสนอให้ใช้เพื่อนบ้านรอบ ๆ จุดที่สนใจจำนวน 8 จุดด้วยเหตุผลที่ว่าเป็นตัวเลขที่เหมาะสมกับข้อมูลโดยทั่วไป ทั้งยังไม่ใช้ระยะเวลาในการประมวลผลที่มากจนเกินไปนัก แต่เมื่อนำไปใช้จริงกับข้อมูลอื่น ๆ ที่อาจมีความจำเพาะและทำให้การใช้จำนวนจุด 8 จุดรอบ ๆ จุดที่สนใจ

ไม่ใช่จำนวนจุดที่เหมาะสมในการหาค่าเฉลี่ยเพื่อนำไปใช้ในฟังก์ชัน ดังนั้น การปรับให้จำนวนจุดเพื่อนบ้านดังกล่าวเป็นตัวแปรเสริมถือเป็นการเพิ่มความยืดหยุ่นในการทำงานสำหรับไดนามิกไทม์วอร์ปปีง ถ่วงน้ำหนักแบบเพิ่มสมรรถนะให้มากยิ่งขึ้นไปตามความเหมาะสมของแต่ละข้อมูล โดยเปลี่ยนให้จำนวนจุด 8 จุดรอบ ๆ จุดที่สนใจเป็นค่ากำหนดเริ่มต้นในกรณีที่ใช้ไม่ต้องการหาจำนวนจุดของเพื่อนบ้านรอบ ๆ จุดที่สนใจแทน



รายการอ้างอิง

1. D.J. Berndt and J. Clifford. Using Dynamic Time Warping to Find Patterns in Time Series. Proceedings of the Workshop on Knowledge Discovery in Databases. 1994:pp. 359-70.
2. Y. -S Jeong, M. K. Jeong , O.A. Omitaomu. Weighted dynamic time warping for time series classification. Pattern Recognition. 2011;44(9):2231-40.
3. Y. Chen, E. Keogh, B. Hu, N. Begum, A. Bagnall, A. Mueen and G. Batista. The UCR Time Series Classification Archive [Internet]. 2015. Available from: www.cs.ucr.edu/~eamonn/time_series_data/.
4. K. Fukunaga, P.M. Narendra . A Branch and Bound Algorithm for Computing k-Nearest Neighbors. IEEE Transactions on Computers. 1975;C-24(7):750-3.
5. J.M. Keller , M.R. Gray, J.A. Givens. A fuzzy K-nearest neighbor algorithm. IEEE Transactions on Systems, Man, and Cybernetics. 1985;SMC-15(4):580-5.
6. K. Ueno , X. Xi , E. Keogh , D. -J Lee . Anytime Classification Using the Nearest Neighbor Algorithm with Applications to Stream Mining. Proceedings of the Sixth International Conference on Data Mining. 1193365: IEEE Computer Society; 2006. p. 623-32.
7. A. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. Pattern Recognition. 2005;38:2270-85.
8. A. Meijster J.B.T.M. Roerdink, and W. H. Hesselink. A General Algorithm for Computing Distance Transforms in Linear Time. Mathematical Morphology and its Applications to Image and Signal Processing. 2000;18:331-40.
9. R.A. Lotufo and F.A. Zampiroli, editor Fast multidimensional parallel Euclidean distance transform based on mathematical morphology. Computer Graphics and Image Processing; 2001: Proceedings of XIV Brazilian Symposium on Computer Graphics and Image Processing.
10. N. Kyurkchiev and S. Markov. Sigmoid functions: Some approximation, and modelling aspects 2015.

11. M. Hazewinkel Exponential function: Springer Science+Business Media B.V. / Kluwer Academic Publisher; 1994.
12. H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. IEEE Transactions on acoustics, speech, and signal processing. 1978;ASSP-26(1).
13. F. Itakura. Minimum prediction residual principle applied to speech recognition. IEEE Transactions on acoustics, speech, and signal processing. 1975;ASSP-23(1).
14. C. Ratanamahatana and E. Keogh. Making Time-Series Classification More Accurate Using Learned Constraints. SDM. 2004.
15. H. Srivastava. What Is K-Fold Cross Validation? : Magoosh Data Science Blog; 2017 [Available from: <https://magoosh.com/data-science/k-fold-cross-validation/>].
16. E. J. Keogh and M. J. Pazzani. Derivative Dynamic Time Warping. Proceedings of the 2001 SIAM International Conference on Data Mining. 2001.
17. X. Ying and W. Bryan. Adaptive Feature Based Dynamic Time Warping. IJCSNS International Journal of Computer Science and Network Security. 2010;10:264-73.
18. Y. Wan, X. -L Chen, Y. Shi. Adaptive cost dynamic time warping distance in time series analysis for classification. Journal of Computational and Applied Mathematics. 2017;319:514-20.



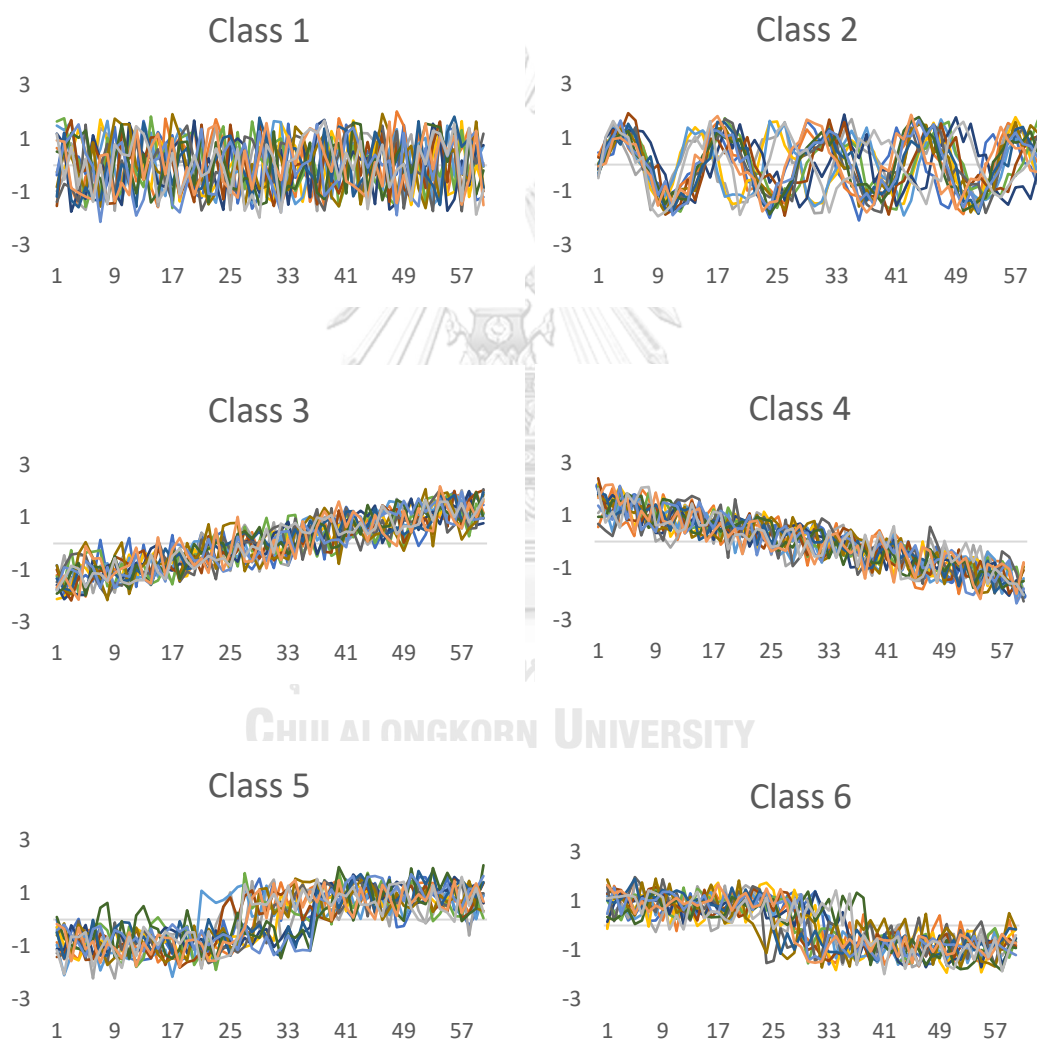
ภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภาคผนวก ก

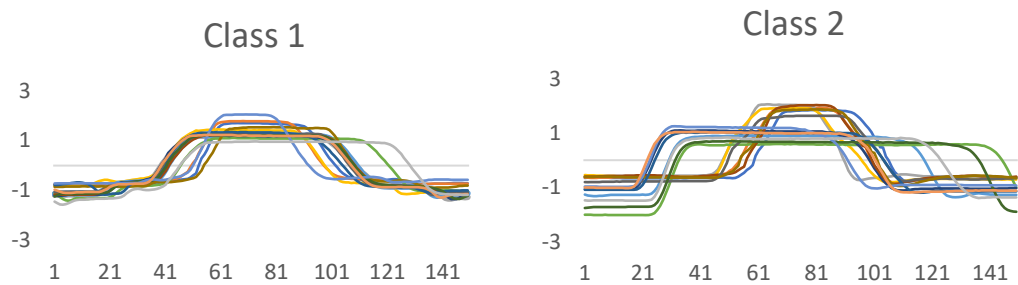
ข้อมูลทั้งหมด 19 ชุดที่ใช้ในการทดลองซึ่งได้กล่าวถึงในบทที่ 4 จะถูกแสดงในรูปแบบของกราฟ โดยในแต่ละกราฟจะประกอบด้วยตัวอย่างข้อมูลแต่ละประเภทของข้อมูลแต่ละชุด โดยจะแสดงเฉพาะข้อมูลเรียนรู้ที่นำมาใช้ในการทดลองเท่านั้น

1. ชุดข้อมูล Synthetic control มีข้อมูลทั้งหมด 6 ประเภทดังแสดงในรูปที่ ก.1



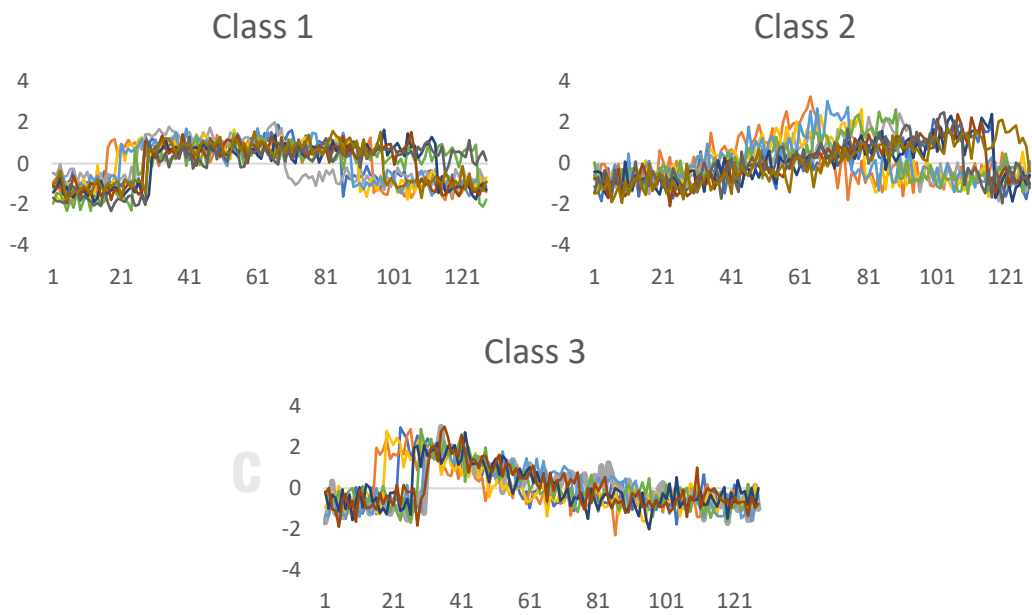
รูปที่ ก.1 กราฟของข้อมูลแต่ละประเภทในชุดข้อมูล Synthetic control

2. ชุดข้อมูล Gun Point มีข้อมูลทั้งหมด 2 ประเภทดังแสดงในรูปที่ ก.2



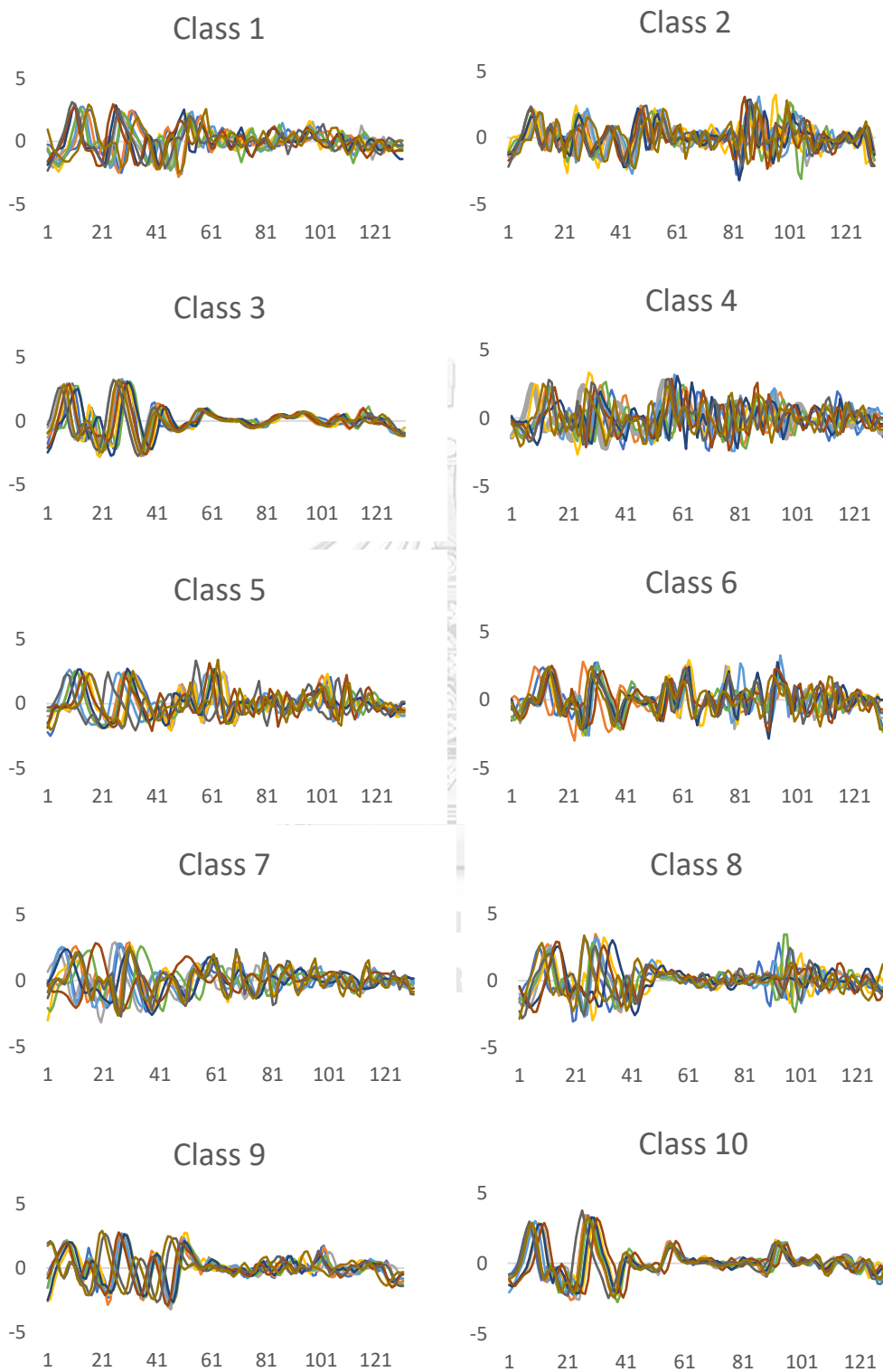
รูปที่ ก.2 กราฟของข้อมูลแต่ละประเภทในชุดข้อมูล Gun Point

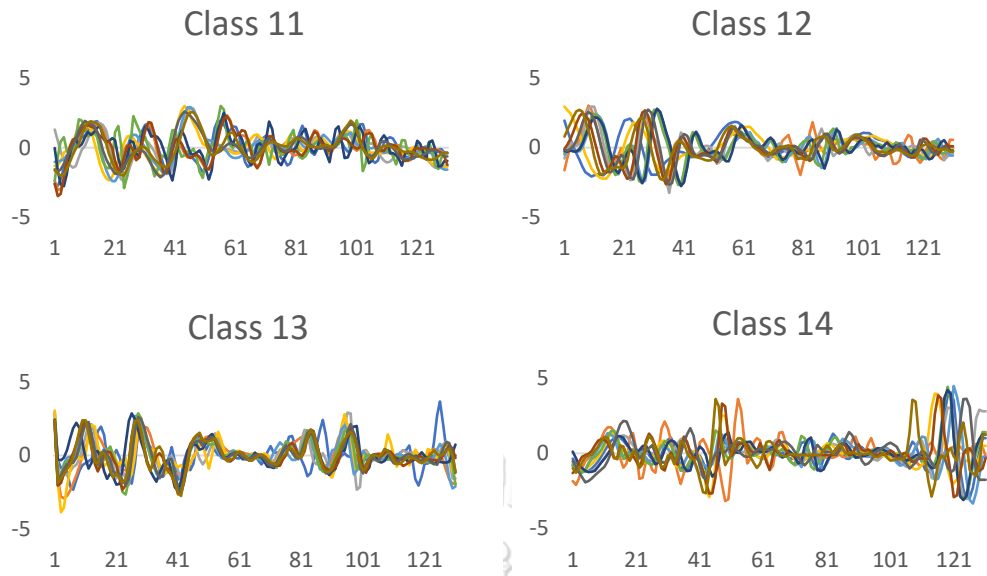
3. ชุดข้อมูล CBF มีข้อมูลทั้งหมด 3 ประเภทดังแสดงในรูปที่ ก.3



รูปที่ ก.3 กราฟของข้อมูลแต่ละประเภทในชุดข้อมูล CBF

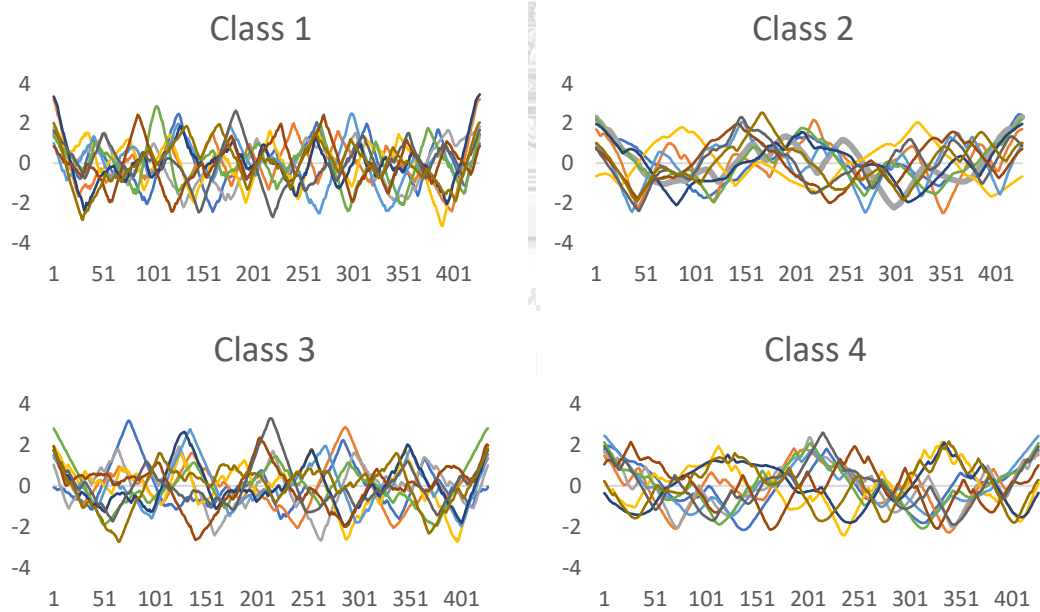
4. ชุดข้อมูล FaceAll มีข้อมูลทั้งหมด 14 ประเภทดังแสดงในรูปที่ ก.4

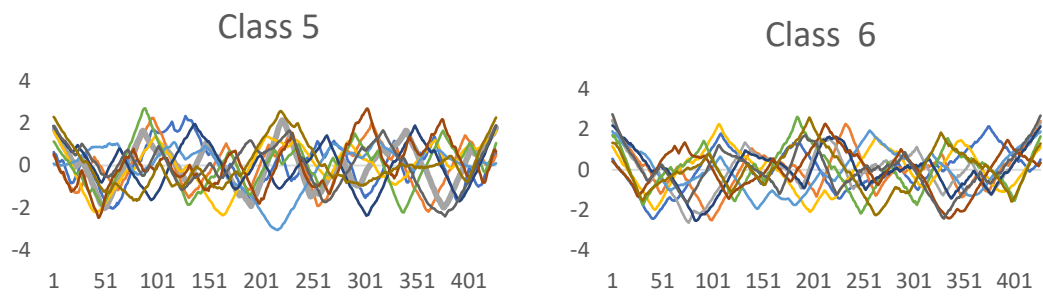




รูปที่ ก.4 กราฟของข้อมูลแต่ละประเภทในชุดข้อมูล FaceAll

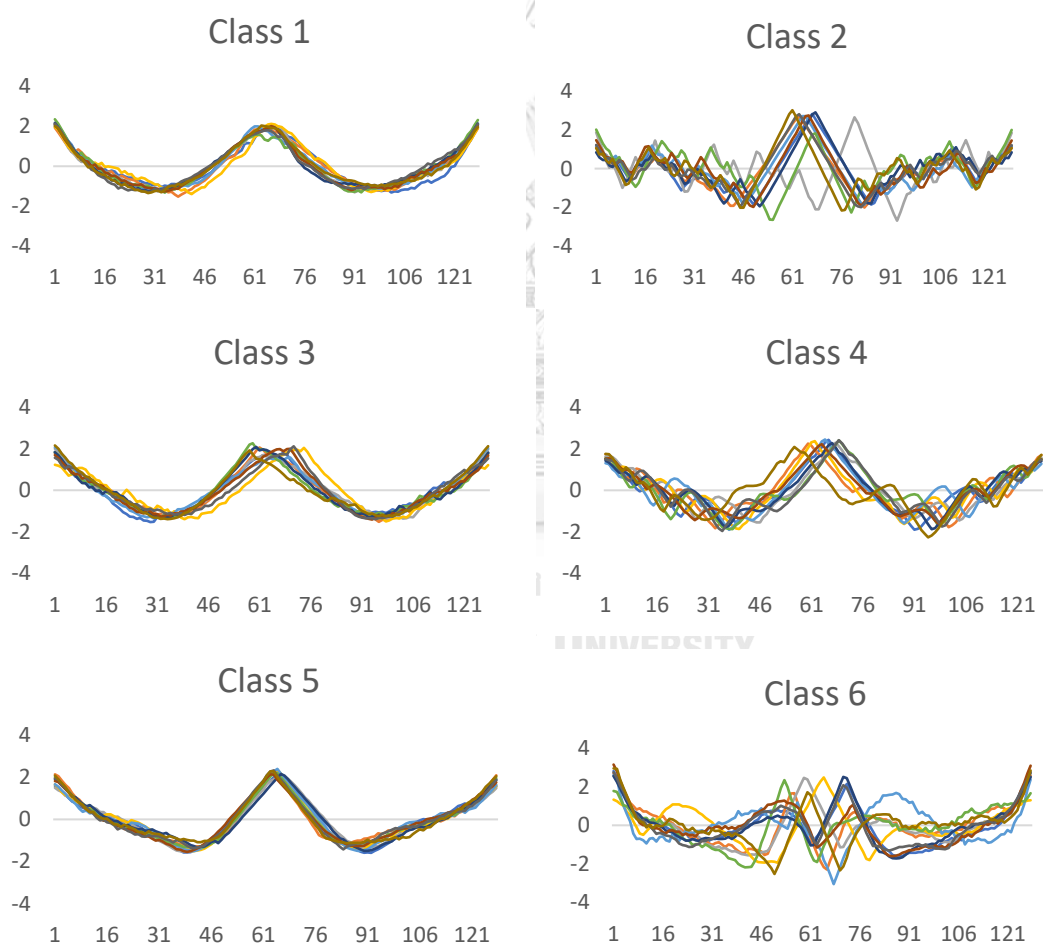
5. ชุดข้อมูล OSULeaf มีข้อมูลทั้งหมด 6 ประเภทดังแสดงในรูปที่ ก.5

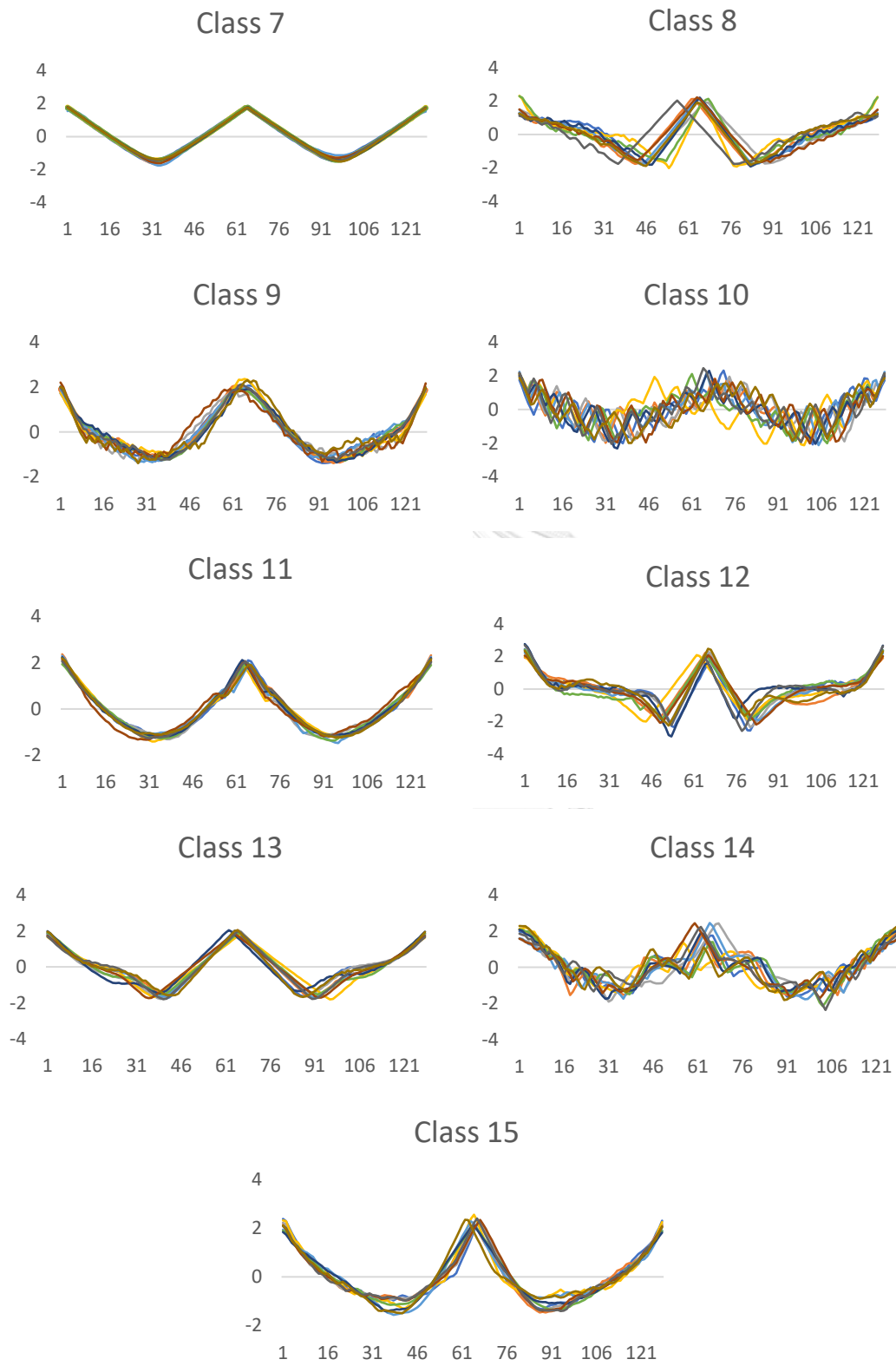




รูปที่ ก.5 กราฟของข้อมูลแต่ละประเภทในชุดข้อมูล OSULeaf

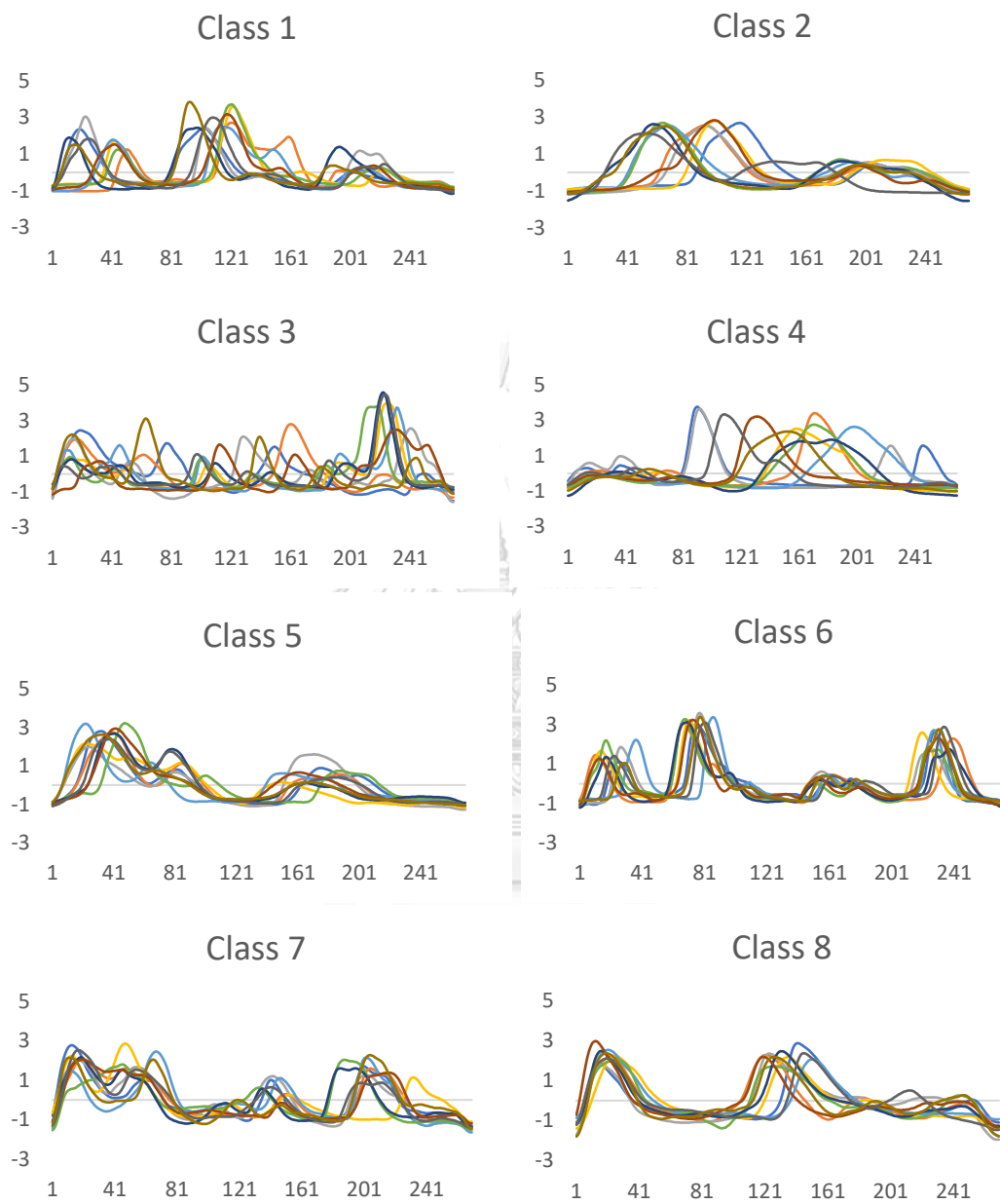
6. ชุดข้อมูล SwedishLeaf มีข้อมูลทั้งหมด 15 ประเภทดังแสดงในรูปที่ ก.6



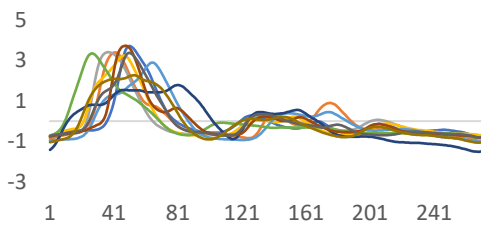


รูปที่ ก.6 กราฟของข้อมูลแต่ละประเภทในชุดข้อมูล SwedishLeaf

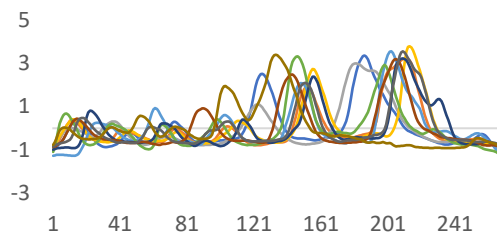
7. ชุดข้อมูล 50words มีข้อมูลทั้งหมด 50 ประเภทดังแสดงในรูปที่ ก.7



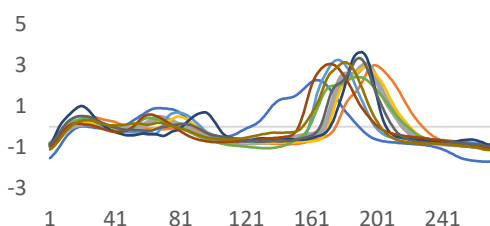
Class 9



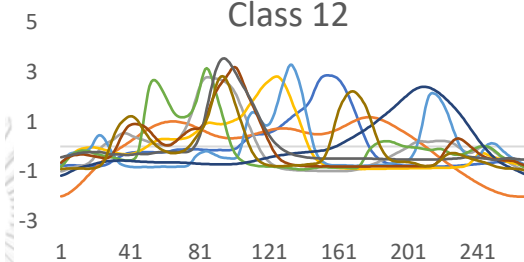
Class 10



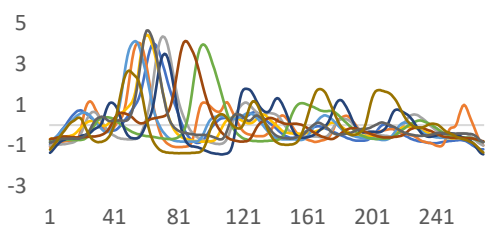
Class 11



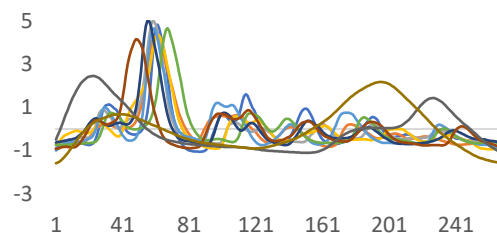
Class 12



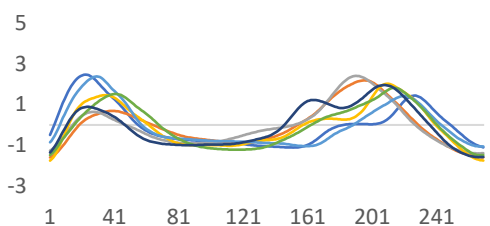
Class 13



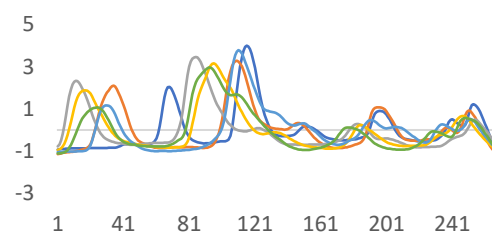
Class 14



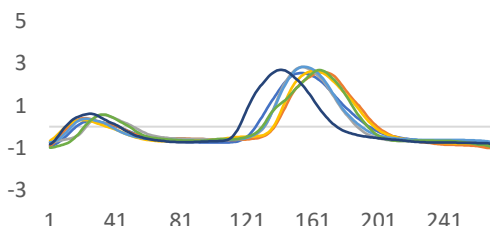
Class 15



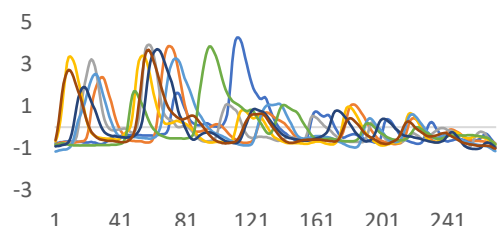
Class 16



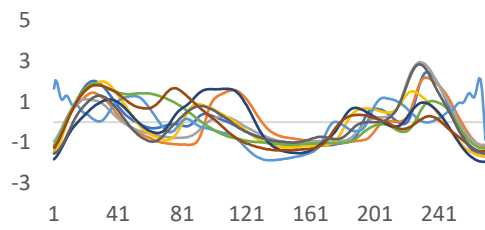
Class 17



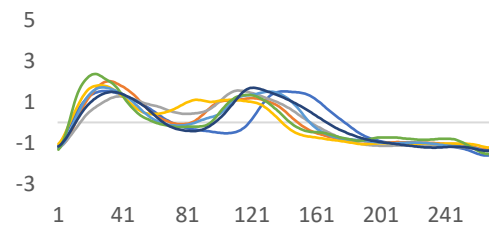
Class 18



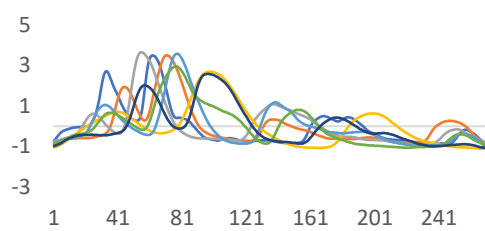
Class 19



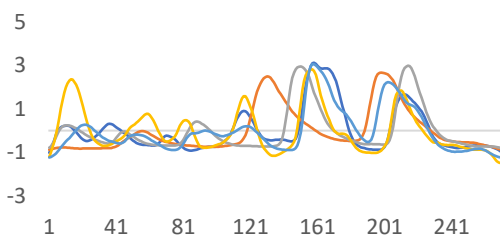
Class 20



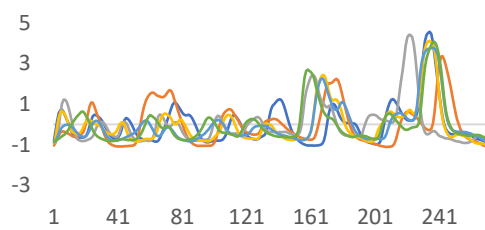
Class 21



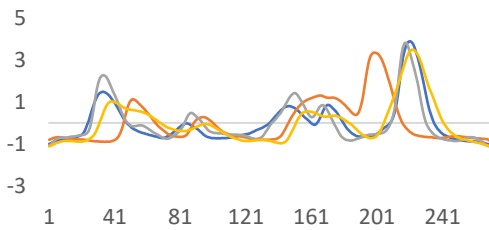
Class 22



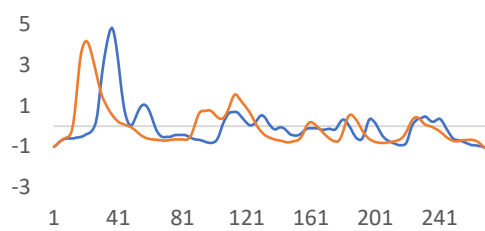
Class 23



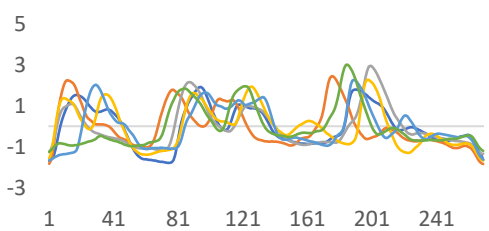
Class 24



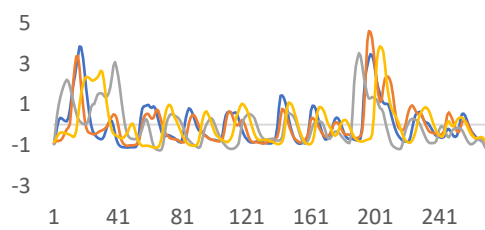
Class 25



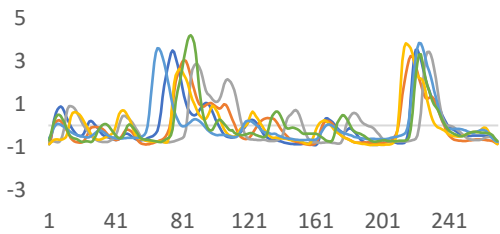
Class 26



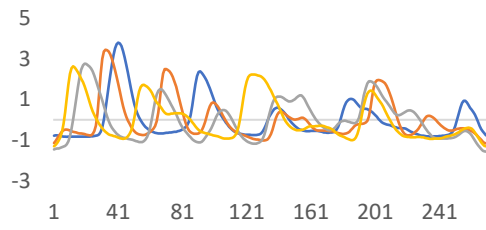
Class 27



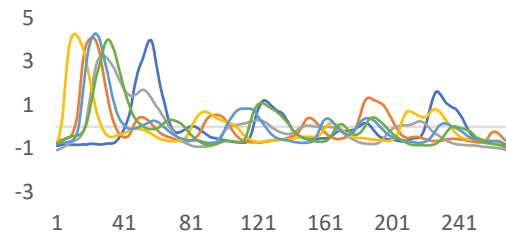
Class 28



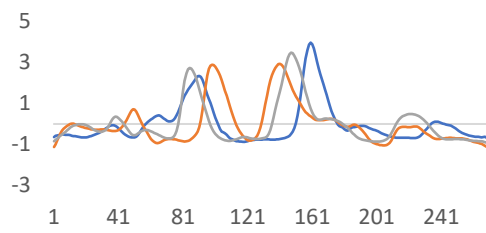
Class 29



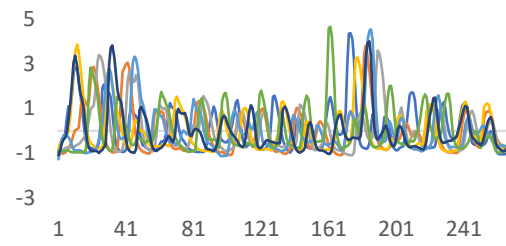
Class 30



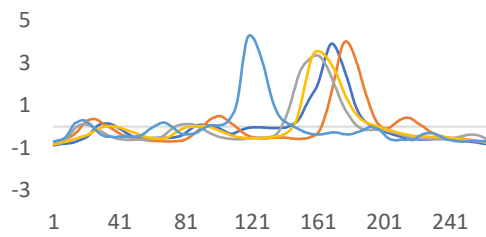
Class 31



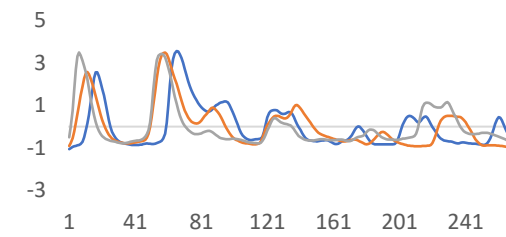
Class 32



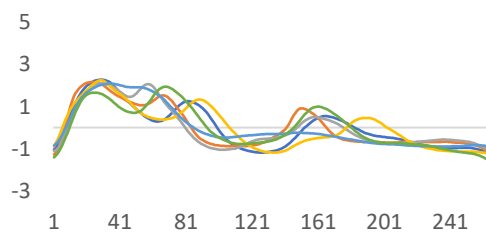
Class 33



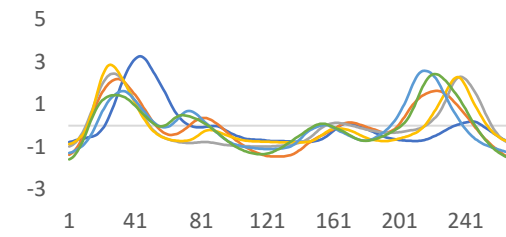
Class 34



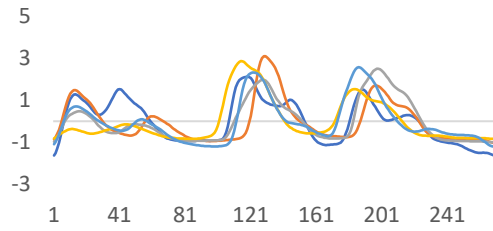
Class 35



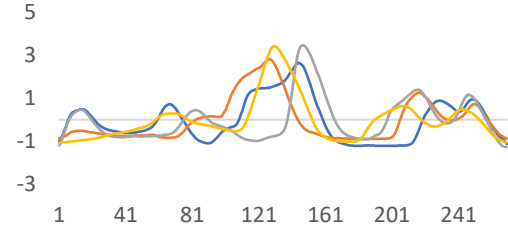
Class 36



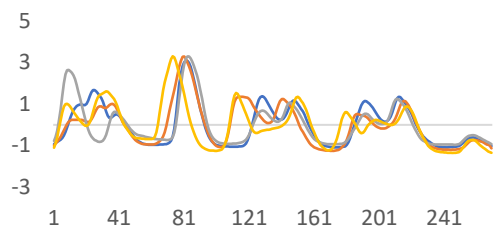
Class 37



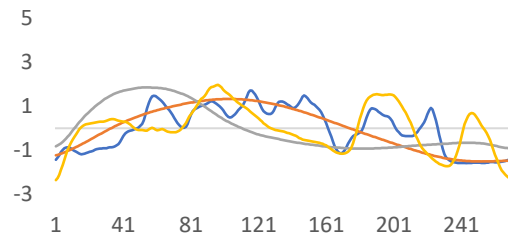
Class 38



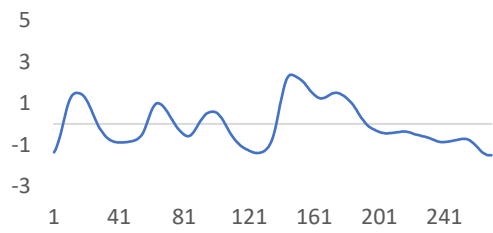
Class 39



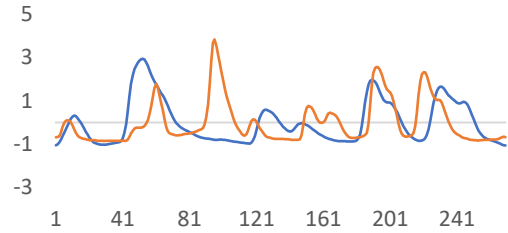
Class 40



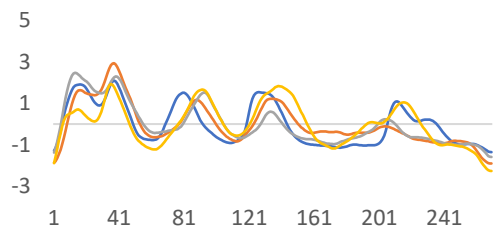
Class 41



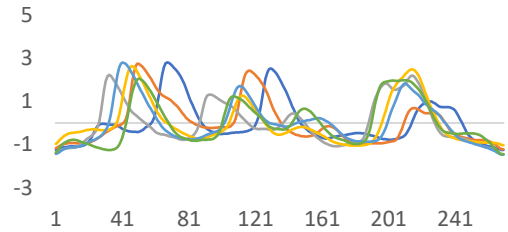
Class 42



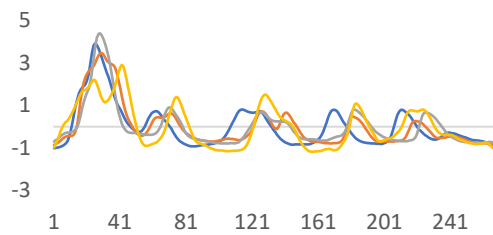
Class 43



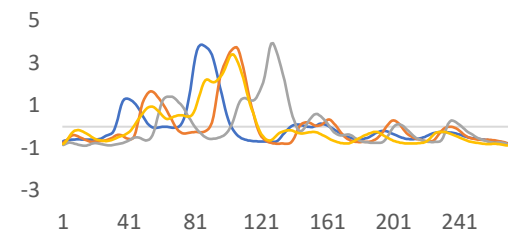
Class 44



Class 45



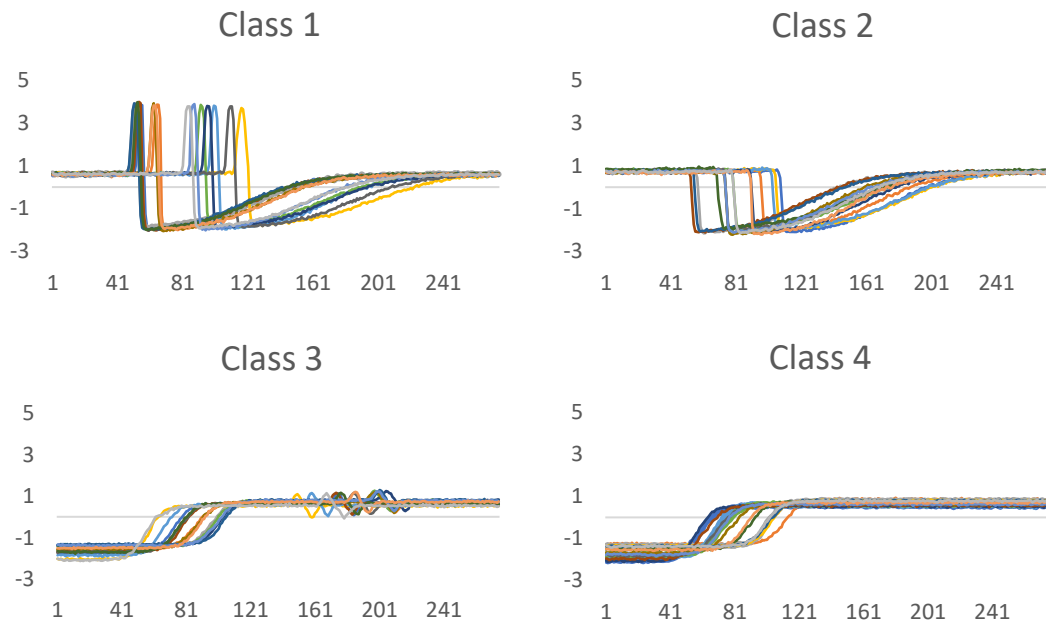
Class 46





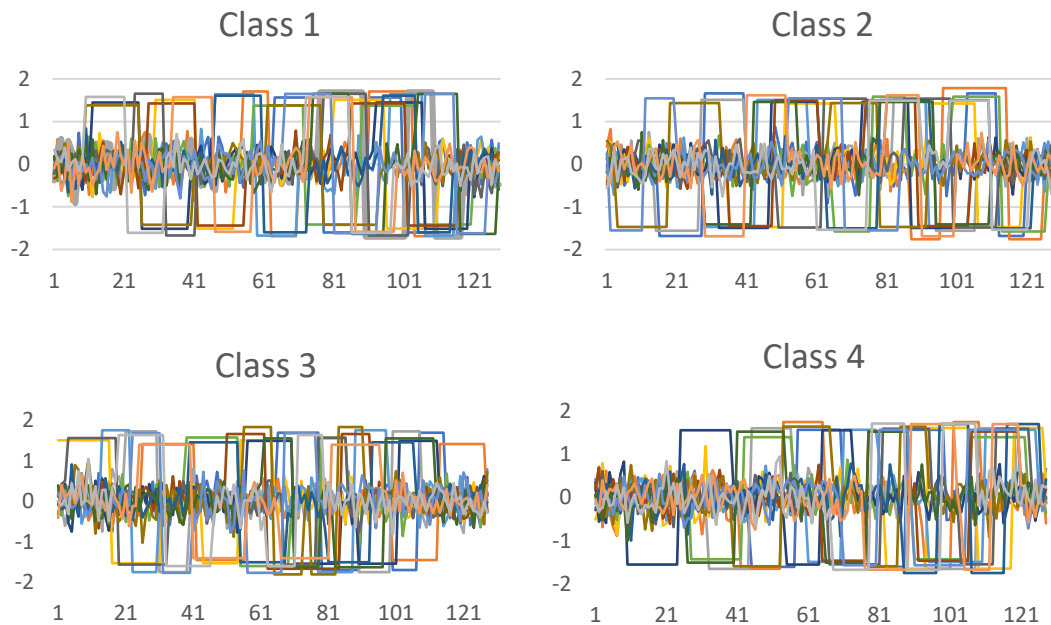
รูปที่ ก.7 กราฟของข้อมูลแต่ละประเภทในชุดข้อมูล 50words

8. ชุดข้อมูล Trace มีข้อมูลทั้งหมด 4 ประเภทดังแสดงในรูปที่ ก.8



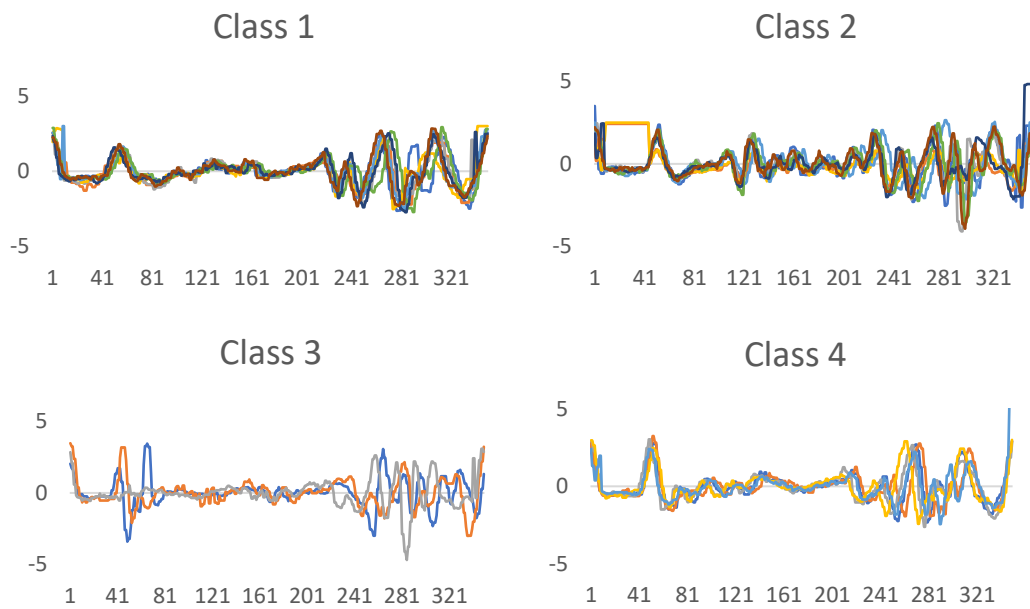
รูปที่ ก.8 กราฟของข้อมูลแต่ละประเภทในชุดข้อมูล Trace

9. ชุดข้อมูล Two Patterns มีข้อมูลทั้งหมด 4 ประเภทดังแสดงในรูปที่ ก.9



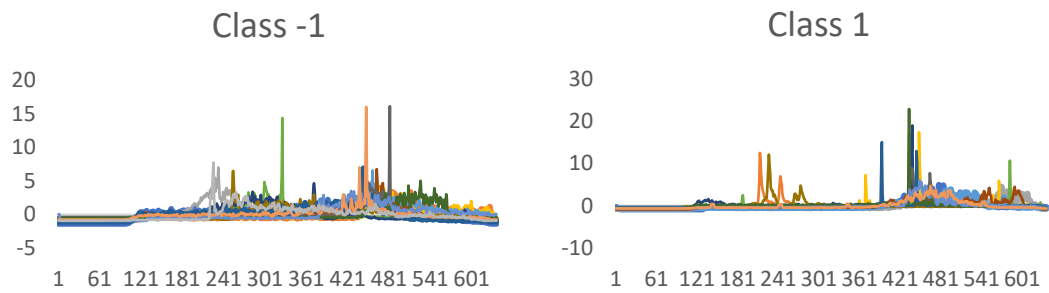
รูปที่ ก.9 กราฟของข้อมูลแต่ละประเภทในชุดข้อมูล Two Patterns

10. ชุดข้อมูล FaceFour มีข้อมูลทั้งหมด 4 ประเภทดังแสดงในรูปที่ ก.10



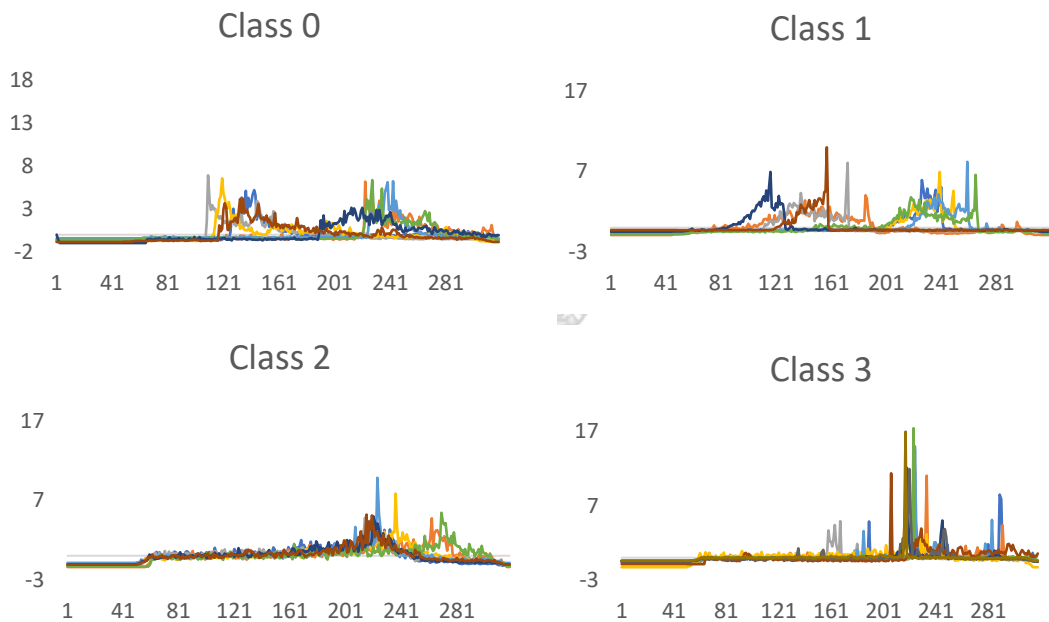
รูปที่ ก.10 กราฟของข้อมูลแต่ละประเภทในชุดข้อมูล FaceFour

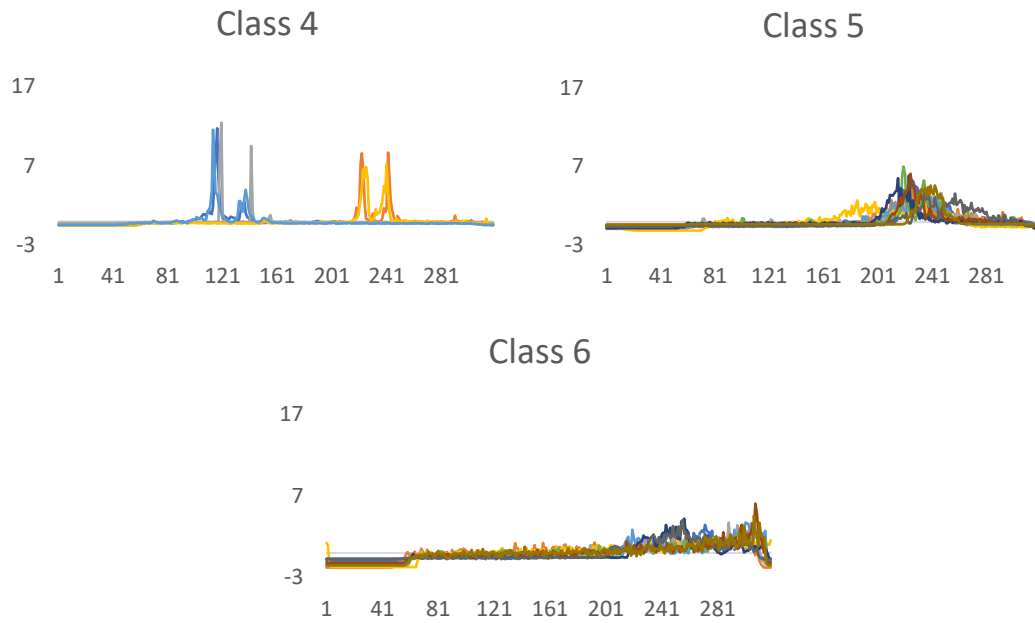
11. ชุดข้อมูล Lightning-2 มีข้อมูลทั้งหมด 2 ประเภทดังแสดงในรูปที่ ก.11



รูปที่ ก.11 กราฟของข้อมูลแต่ละประเภทในชุดข้อมูล Lightning-2

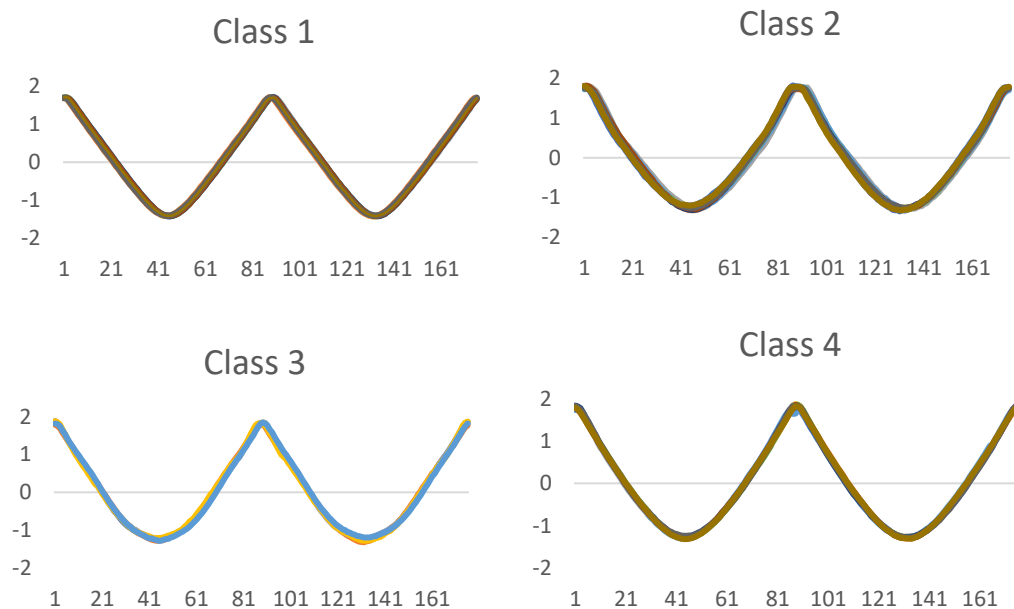
12. ชุดข้อมูล Lightning-7 มีข้อมูลทั้งหมด 7 ประเภทดังแสดงในรูปที่ ก.12



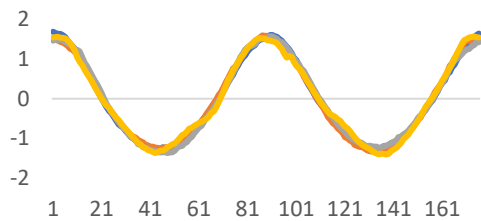


รูปที่ ก.12 กราฟของข้อมูลแต่ละประเภทในชุดข้อมูล Lighting-7

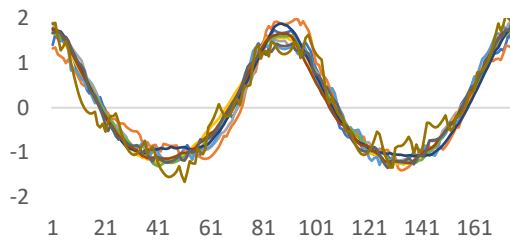
13. ชุดข้อมูล Adiac มีข้อมูลทั้งหมด 37 ประเภทดังแสดงในรูปที่ ก.13



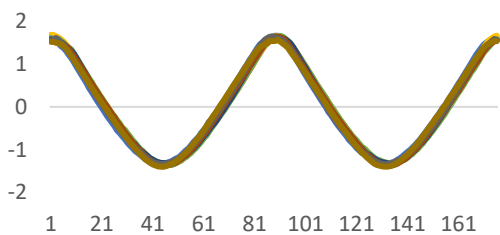
Class 5



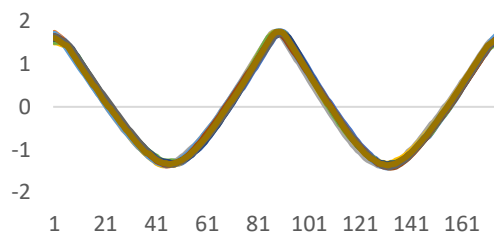
Class 6



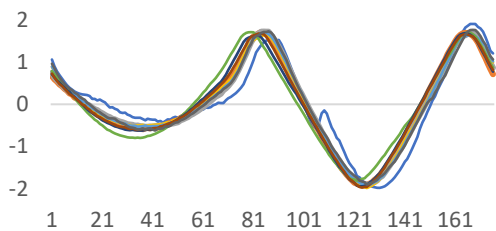
Class 7



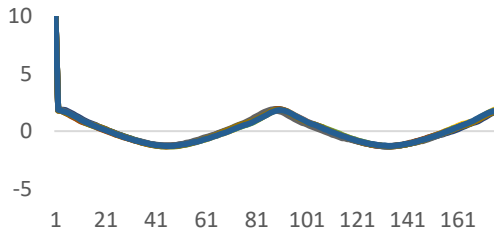
Class 8



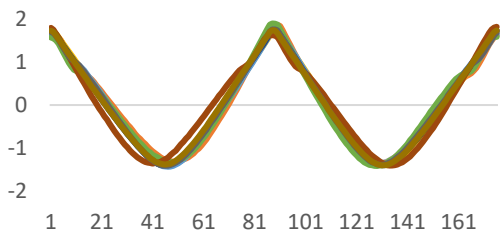
Class 9



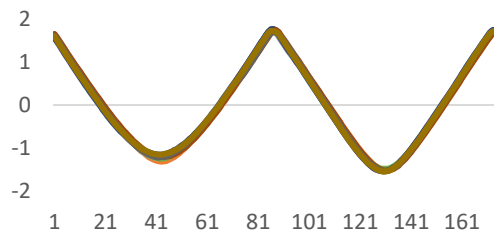
Class 10



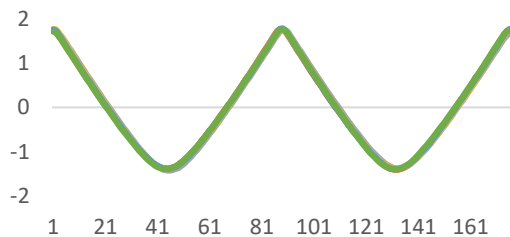
Class 11



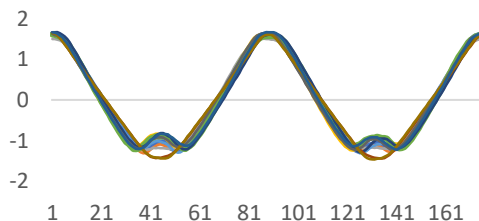
Class 12



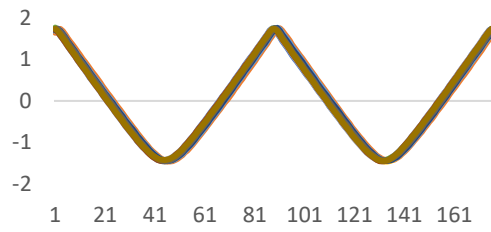
Class 13



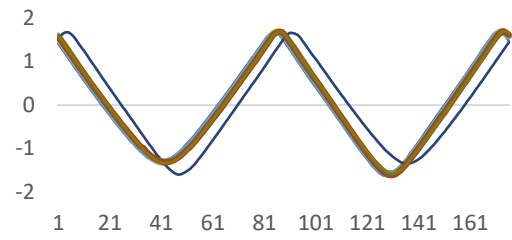
Class 14



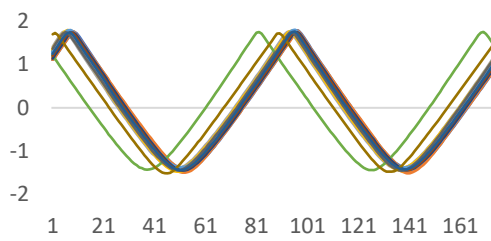
Class 15



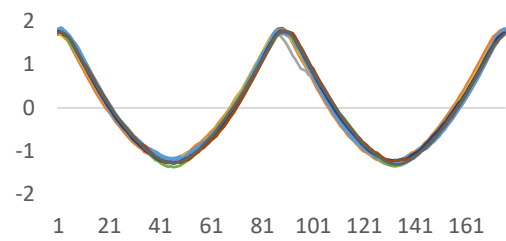
Class 16



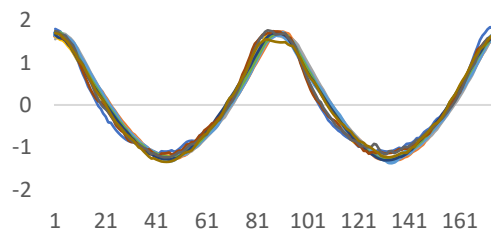
Class 17



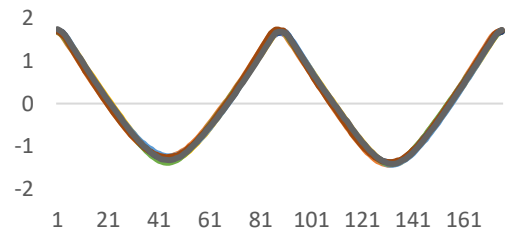
Class 18



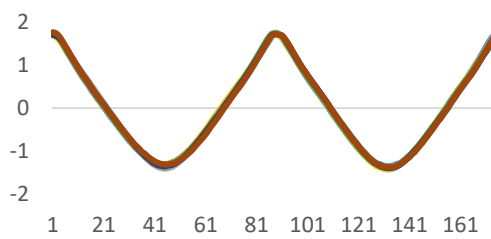
Class 19



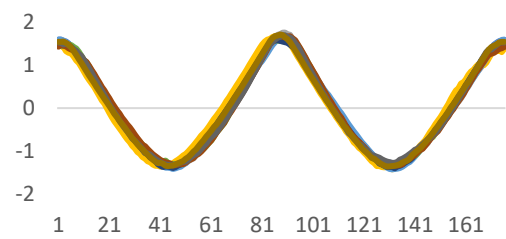
Class 20



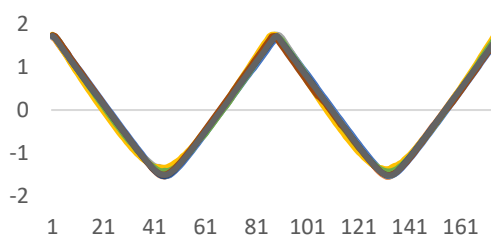
Class 21



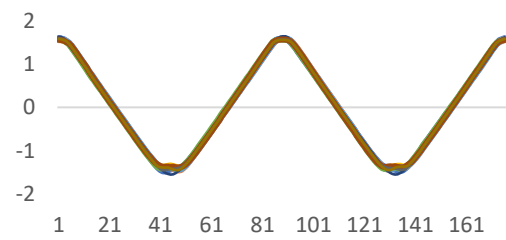
Class 22



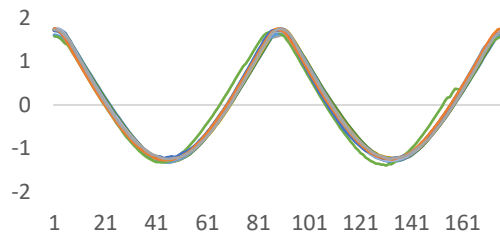
Class 23



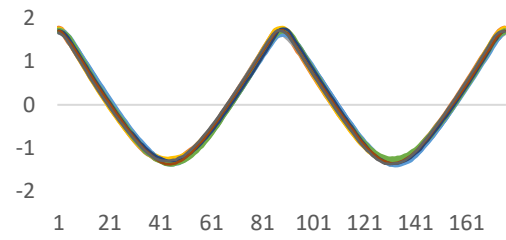
Class 24



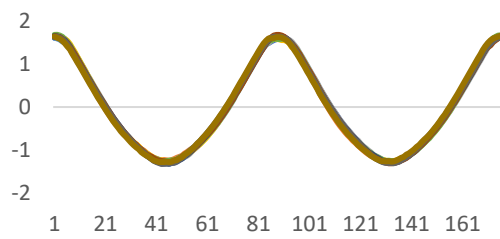
Class 25



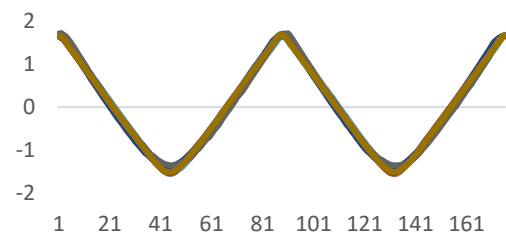
Class 26



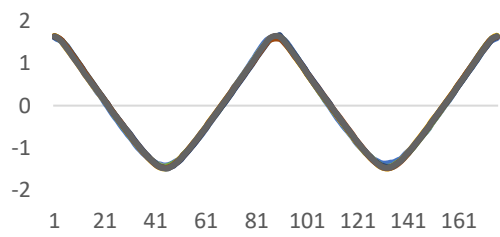
Class 27



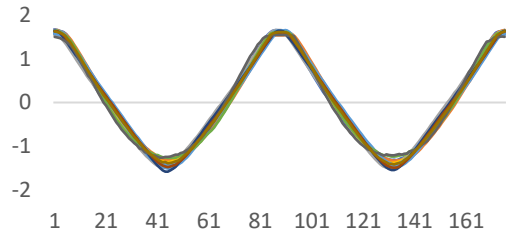
Class 28



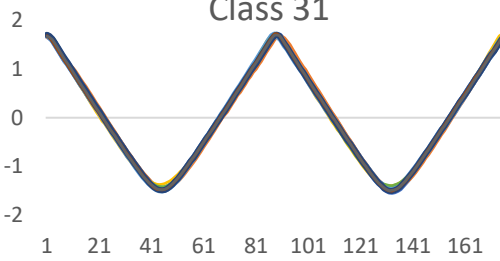
Class 29



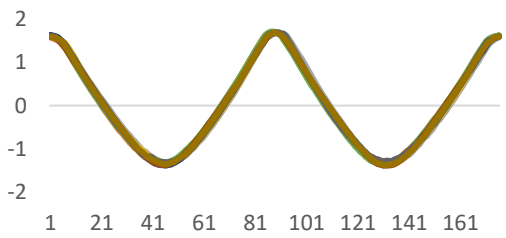
Class 30



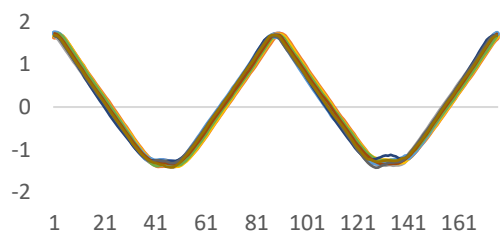
Class 31



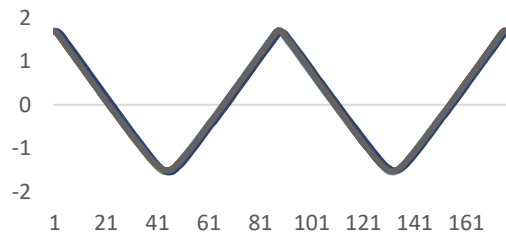
Class 32

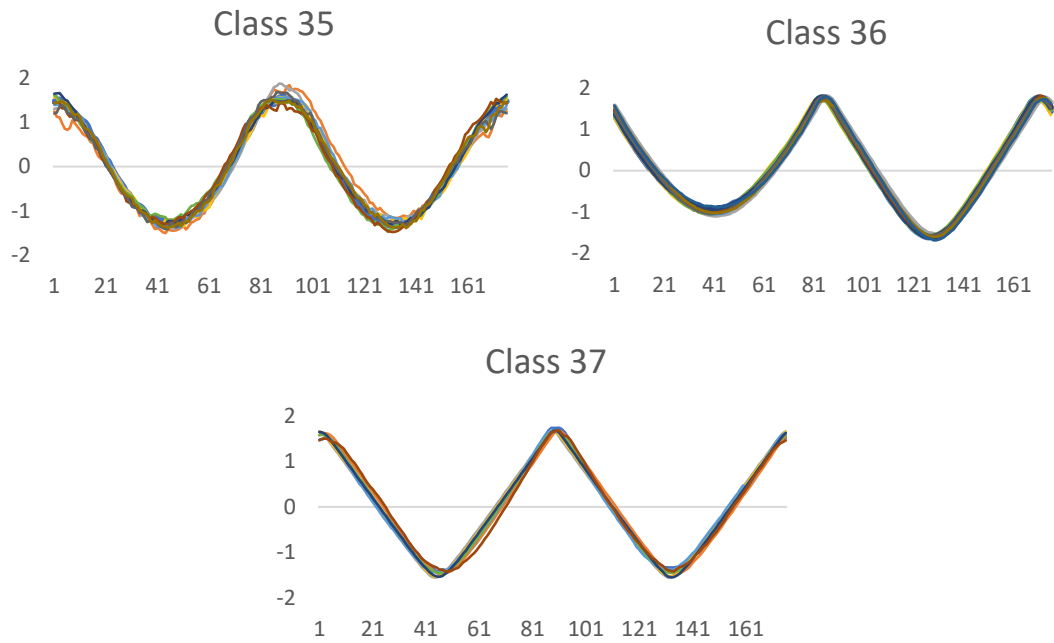


Class 33



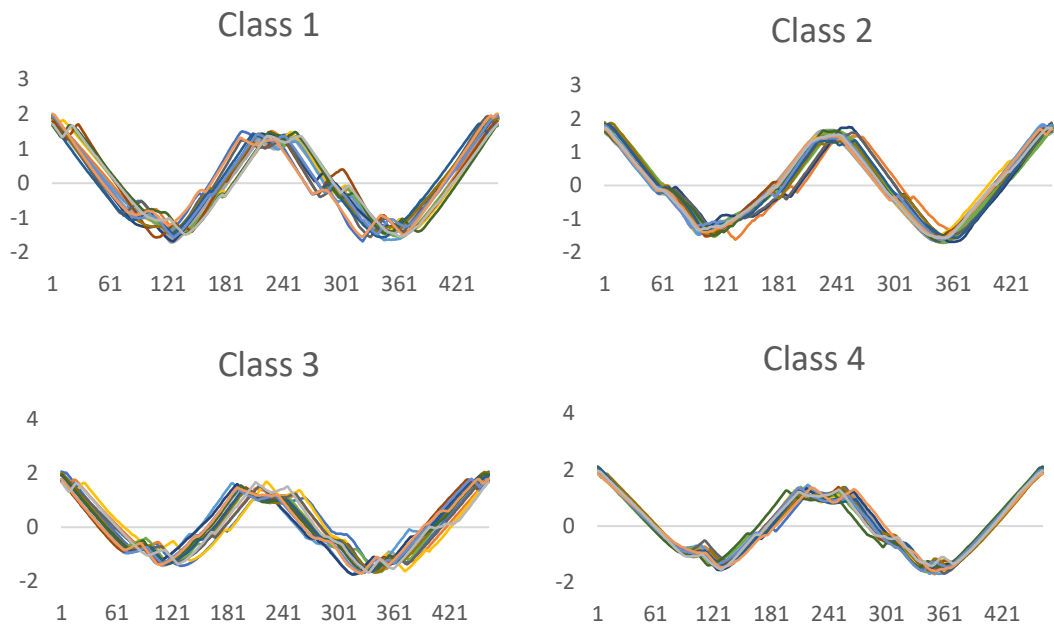
Class 34

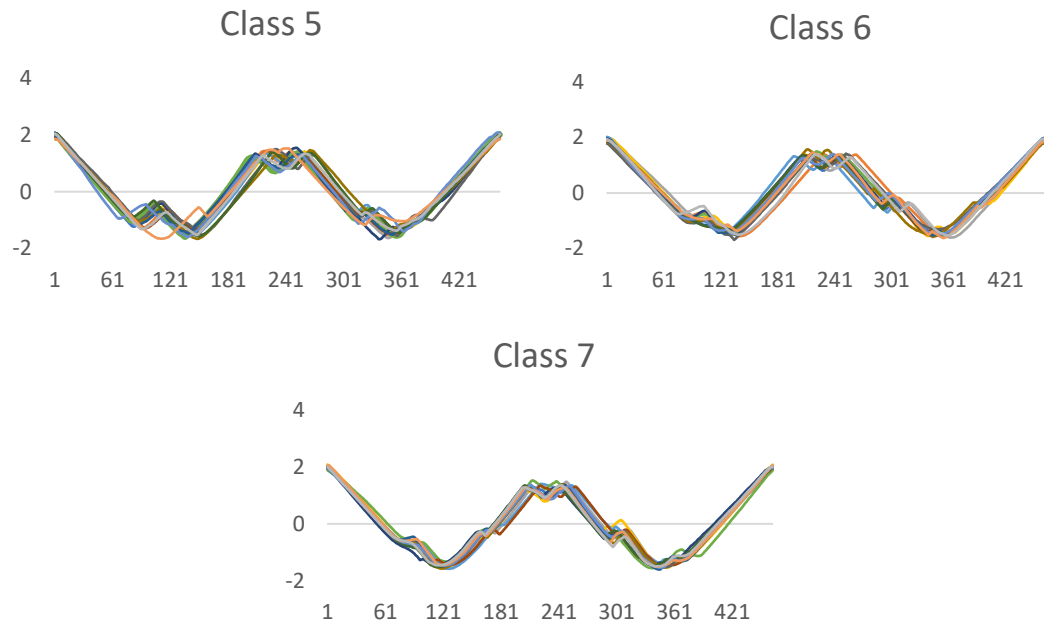




รูปที่ ก.13 กราฟของข้อมูลแต่ละประเภทในชุดข้อมูล Adiac

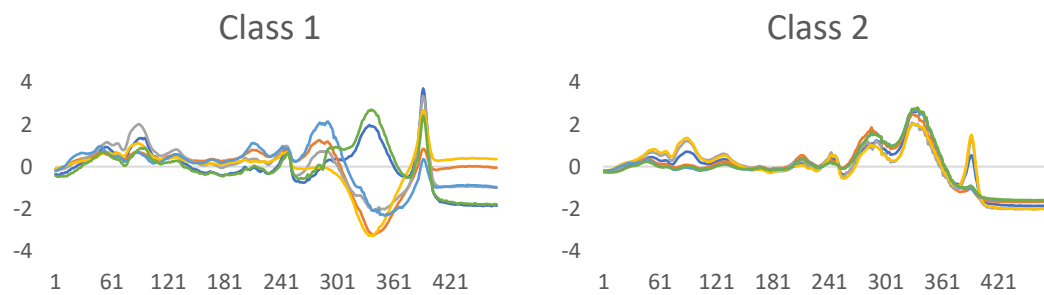
14. ชุดข้อมูล FISH มีข้อมูลทั้งหมด 7 ประเภทดังแสดงในรูปที่ ก.14

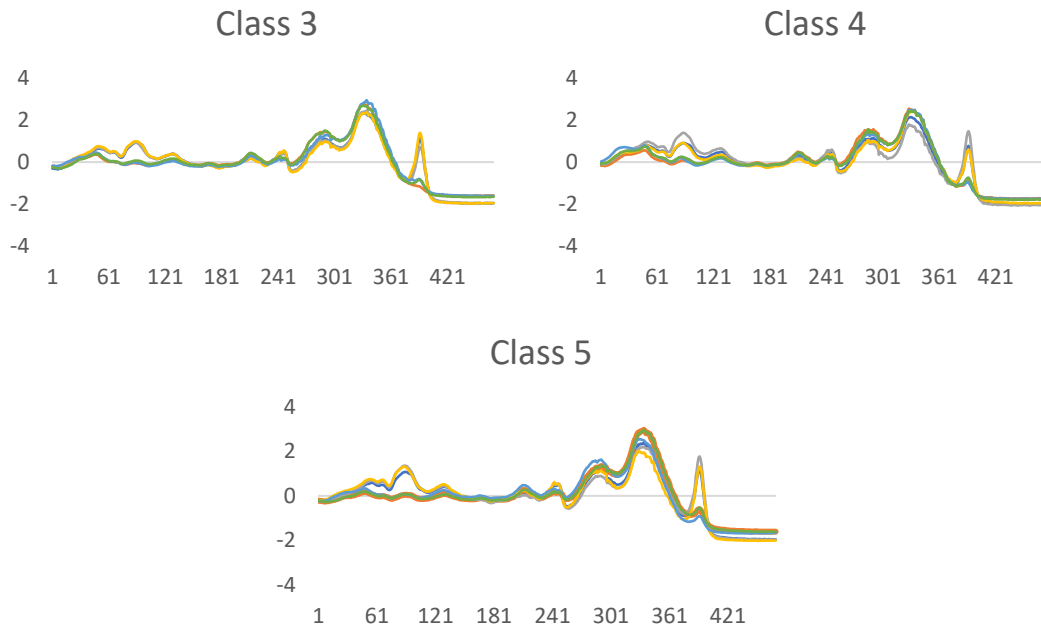




รูปที่ ก.14 กราฟของข้อมูลแต่ละประเภทในชุดข้อมูล FISH

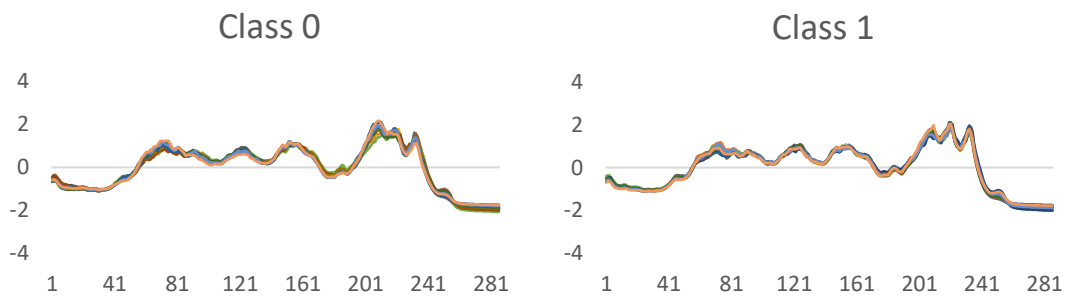
15. ชุดข้อมูล Beef มีข้อมูลทั้งหมด 5 ประเภทดังแสดงในรูปที่ ก.15





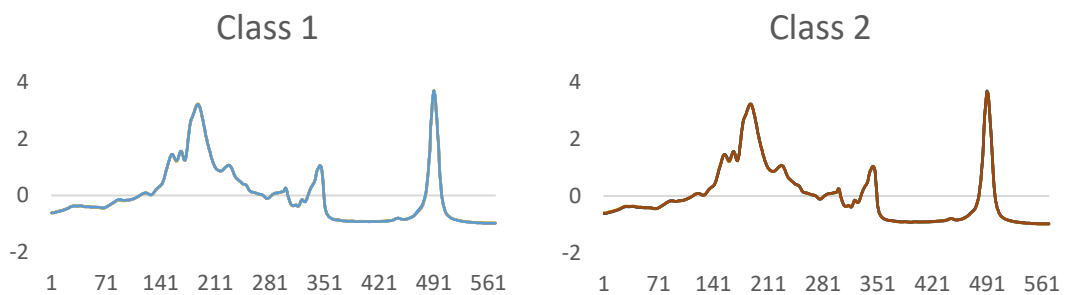
รูปที่ ก.15 กราฟของข้อมูลแต่ละประเภทในชุดข้อมูล Beef

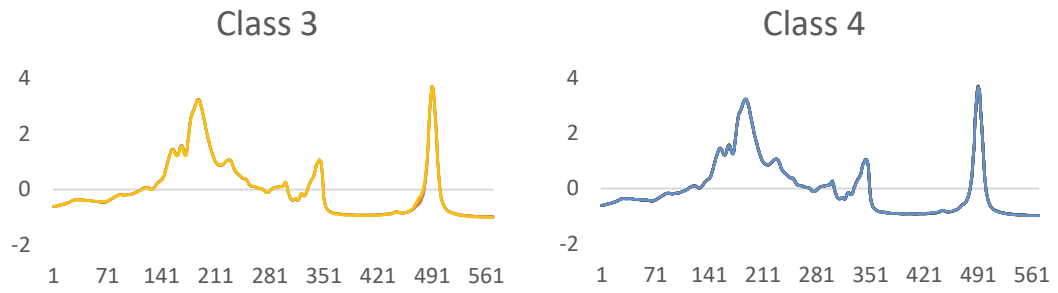
16. ชุดข้อมูล coffee มีข้อมูลทั้งหมด 2 ประเภทดังแสดงในรูปที่ ก.16



รูปที่ ก.16 กราฟของข้อมูลแต่ละประเภทในชุดข้อมูล coffee

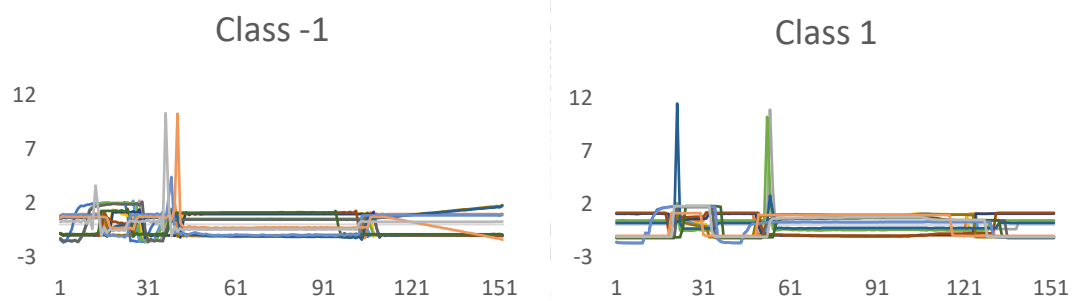
17. ชุดข้อมูล OliveOil มีข้อมูลทั้งหมด 2 ประเภทดังแสดงในรูปที่ ก.17





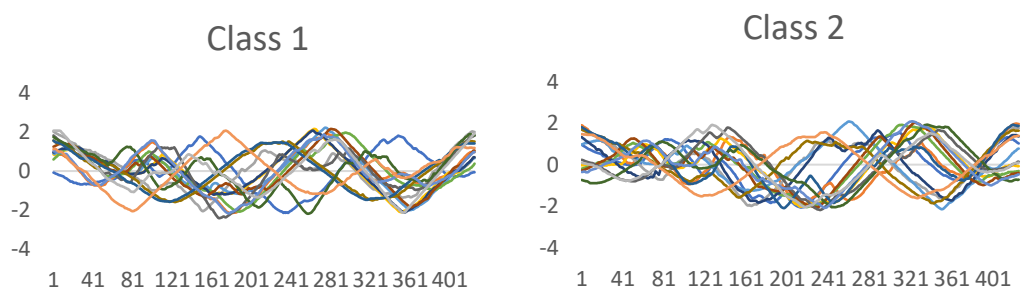
รูปที่ ก.17 กราฟของข้อมูลแต่ละประเภทในชุดข้อมูล OliveOil

18. ชุดข้อมูล wafer มีข้อมูลทั้งหมด 2 ประเภทดังแสดงในรูปที่ ก.18



รูปที่ ก.18 กราฟของข้อมูลแต่ละประเภทในชุดข้อมูล wafer

19. ชุดข้อมูล yoga มีข้อมูลทั้งหมด 2 ประเภทดังแสดงในรูปที่ ก.19



รูปที่ ก.19 กราฟของข้อมูลแต่ละประเภทในชุดข้อมูล yoga



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ประวัติผู้เขียนวิทยานิพนธ์

นางสาว พิชามณูช อนันตเศรษฐ์ เกิดวันที่ 27 กรกฎาคม พ.ศ. 2536 สำเร็จการศึกษา
ระดับมัธยมจากโรงเรียนสาธิตมหาวิทยาลัยเชียงใหม่ จากนั้นทำการศึกษาต่อที่คณะวิทยาศาสตร์
มหาวิทยาลัยมหิดล ในปีการศึกษา 2555 และสำเร็จการศึกษาปริญญาวิทยาศาสตรบัณฑิต
ภาควิชาคณิตศาสตร์ ในปีการศึกษา 2558 และเข้าศึกษาในหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิทยาศาสตร์คอมพิวเตอร์ที่ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์
จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2559



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY