

ห้องสมุดงานวิจัย สำนักงานคณะกรรมการวิจัยแห่งชาติ



E42167

การจัดกล่มลำดับข้อของบัญญัติอนุกรมเวลาแบบक्रमसोपाเป็นความหมาย

นายวิชาญ เบ็ญรนาททรงกูจ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2553

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

600256911

การจัดกลุ่มลำดับย่อยของข้อมูลอนุกรมเวลาแบบกระแสอย่างมีความหมาย

ห้องสมุดงานวิจัย สำนักงานคณะกรรมการวิจัยแห่งชาติ



นายวิชญ์ เนียรนาทตระกูล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2553

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย



MEANINGFUL SUBSEQUENCE CLUSTERING FOR TIME SERIES DATA STREAM

Mr. Vit Niennattrakul

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Computer Engineering

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2010

Copyright of Chulalongkorn University

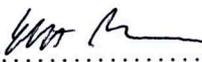
Thesis Title MEANINGFUL SUBSEQUENCE CLUSTERING FOR TIME
SERIES DATA STREAM
By Mr. Vit Niennattrakul
Field of Study Computer Engineering
Thesis Advisor Assistant Professor Chotirat Ann Ratanamahatana, Ph.D.

Accepted by the Faculty of Engineering, Chulalongkorn University in Partial Fulfillment of
the Requirements for the Doctoral Degree

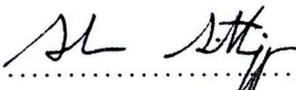

..... Dean of the Faculty of Engineering
(Associate Professor Boonsom Lerdhirunwong, Dr.Ing.)

THESIS COMMITTEE


..... Chairman
(Professor Boonserm Kijisirikul, Ph.D.)


..... Thesis Advisor
(Assistant Professor Chotirat Ann Ratanamahatana, Ph.D.)


..... Examiner
(Professor Prabhas Chongstitvattana, Ph.D.)


..... Examiner
(Assistant Professor Sukree Sinthupinyo, Ph.D.)


..... External Examiner
(Assistant Professor Charnyote Pluempitiwiriyaewej, Ph.D.)

วิษญ์ เนียรนาทตระกูล: การจัดกลุ่มลำดับย่อยของข้อมูลอนุกรมเวลาแบบกระแสอย่างมีความหมาย. (MEANINGFUL SUBSEQUENCE CLUSTERING FOR TIME SERIES DATA STREAM) อ.ที่ปรึกษาวิทยานิพนธ์หลัก : ผู้ช่วยศาสตราจารย์ ดร. โชติรัตน์ รัตนามัทธนะ, 192 หน้า.

E 42167

การจัดกลุ่มลำดับย่อยสำหรับข้อมูลอนุกรมเวลาแบบกระแสเป็นหนึ่งในปัญหาที่ท้าทายมากที่สุดของการทำเหมืองข้อมูลอนุกรมเวลาตั้งแต่การจัดกลุ่มลำดับย่อยได้ถูกแสดงให้เห็นว่าการจัดกลุ่มจะให้คำตอบที่ไร้ความหมายในเชิงการทดลองและทฤษฎี การจัดกลุ่มลำดับย่อยของข้อมูลอนุกรมเวลาที่ถูกใช้ในหลายร้อยงานวิจัยนั้นจะให้คลื่นไซน์เป็นตัวแทนกลุ่มเสมอ ถ้าให้ข้อมูลอนุกรมเวลาหนึ่ง ๆ การจัดกลุ่มลำดับย่อยของข้อมูลอนุกรมเวลาควรคืนค่าตัวแทนกลุ่มที่เป็นลักษณะของทุกลำดับย่อยในข้อมูลอนุกรมเวลา สาเหตุที่ทำให้เกิดความไร้ความหมายถูกระบุไว้มาจากสองสาเหตุได้แก่ การใช้ระยะทางยุคลิดเป็นตัววัดระยะทางที่ไม่เหมาะสมและการใช้การเฉลี่ยค่าตามแอมพลิจูดเป็นฟังก์ชันการเฉลี่ยที่ไม่เหมาะสม เพื่อที่จะได้มาซึ่งคำตอบของการจัดกลุ่มที่มีความหมาย ในวิทยานิพนธ์นี้การจัดกลุ่มลำดับย่อยของข้อมูลอนุกรมเวลาตามรูปได้ถูกเสนอโดยใช้ระยะทางไดนามิกโทมวอร์ปปีงและการเฉลี่ยค่าตามรูปแทนระยะทางยุคลิดและการเฉลี่ยค่าตามแอมพลิจูดตามลำดับ ดังนั้นการจัดกลุ่มลำดับย่อยของข้อมูลอนุกรมเวลาตามรูปจะคืนผลลัพธ์ที่มีความหมายที่มากกว่าการจัดกลุ่มลำดับย่อยของข้อมูลอนุกรมเวลาแบบเดิมแต่อย่างไรก็ตามการจัดกลุ่มลำดับย่อยของข้อมูลอนุกรมเวลาตามรูปไม่สามารถประยุกต์ใช้กับข้อมูลแบบกระแสได้ เนื่องจากการจัดกลุ่มลำดับย่อยของข้อมูลอนุกรมเวลาตามรูปใช้เวลาในการประมวลผลนานโดยคำนวณลำดับย่อยที่ผ่านมาทั้งหมดเมื่อมีจุดข้อมูลใหม่เข้ามา การจัดกลุ่มลำดับย่อยของข้อมูลอนุกรมเวลาแบบกระแสตามรูปจึงถูกเสนอให้รองรับกรณีข้อมูลแบบกระแสโดยคำนวณบนชุดข้อมูลขนาดเล็กของลำดับย่อยที่เก็บไว้แทนที่จะคำนวณจากลำดับย่อยทั้งหมดซึ่งชุดข้อมูลของลำดับย่อยที่เก็บไว้ถูกปรับปรุงสำหรับทุกๆจุดข้อมูลเพื่อรักษาจำนวนลำดับย่อยในชุดข้อมูลไม่ให้เกินกว่าจำนวนมากที่สุดที่อนุญาต ดังนั้นการจัดกลุ่มลำดับย่อยของข้อมูลอนุกรมเวลาแบบกระแสตามรูปจึงเร็วกว่าการจัดกลุ่มลำดับย่อยของข้อมูลอนุกรมเวลาตามรูปอย่างมาก

ภาควิชา วิศวกรรมคอมพิวเตอร์.....

ลายมือชื่อนิสิต

สาขาวิชา วิศวกรรมคอมพิวเตอร์.....

ลายมือชื่ออ.ที่ปรึกษาวิทยานิพนธ์หลัก

ปีการศึกษา 2553

4971828021: MAJOR COMPUTER ENGINEERING

KEYWORDS: DATA MINING / SUBSEQUENCE CLUSTERING / TIME SERIES / DATA STREAM

VIT NIENNATTRAKUL : MEANINGFUL SUBSEQUENCE CLUSTERING FOR TIME SERIES DATA STREAM. ADVISOR : ASSISTANT PROFESSOR CHOTIRAT RATANAMAHAHATANA, PH.D., 192 pp.

E42167

Subsequence clustering for time series data streams is one of the most challenging issues of time series data mining since subsequence clustering has been proven both theoretically and empirically that it produces meaningless clustering results, where hundreds of research works that utilize Subsequence Time Series Clustering (STSC) as a preprocessing step and a subroutine are all affected. Given a time series sequence, subsequence clustering should return cluster representatives which represent characteristics of all subsequences in time series. Therefore, if cluster representatives are always sine waves regardless of inputs, clustering results are meaningless since they do not reflect characteristics of the subsequences. The causes of meaninglessness are identified in twofold, i.e., inappropriate uses of Euclidean distance as a distance measure and Amplitude Averaging as an averaging function. To achieve meaningful clustering results, in this thesis, Shape-based Subsequence Time Series Clustering (2STSC) is proposed to use Dynamic Time Warping (DTW) distance measure and Shape-based Averaging function. Therefore, 2STSC returns more meaningful results than those from STSC. However, 2STSC cannot directly apply to data streams since 2STSC consumes large computational complexity by considering all previous subsequences for every new incoming data point. Shape-based Streaming Subsequence Time Series Clustering (3STSC) is then proposed to handle the streaming case by calculating a clustering result on a small set of stored subsequences instead of calculating from all previous subsequences. The small set of stored subsequences is updated for every new incoming data point to maintain the number of stored subsequences not to exceed the maximum allowance. 3STSC, therefore, is much faster than 2STSC, while 3STSC returns small distortions of clustering results.

Department: Computer Engineering Student's Signature ... 
Field of Study: Computer Engineering Advisor's Signature ... 
Academic Year:2010.....

Acknowledgments

I would like to express my sincere gratitude to my thesis advisor, Dr. Chotirat Ann Ratanamahatana for her invaluable guidance and support during my graduate studies at Chulalongkorn University. She always encourages my progress, new ideas, as well as motivates my research work. She is always full of energy and untiringly available for giving me insightful advices. In addition, I have learned precious lessons through her past and present remarkable research work, as well as her exceptional presentation. I truly consider it a great privilege in having the opportunity to work with her as my graduate advisor.

I am grateful to Prof. Eamonn Keogh who supported me when I was in the United States. I also express my thankfulness to my dissertation committee: Dr. Prabhas Chongstitvatana, Dr. Boonserm Kitsirikul, Dr. Sukree Sinthupinyo, and Dr. Charnyote Pluempitiwiriyawej. I am indebted to every teacher, especially, Dr. Proadpran, Dr. Atiwong, Dr. Athasit, Dr. Pizzanu, Dr. Vishnu, Dr. Somchai, Ajarn Mandhana for introducing me a rabbit hole of computer engineering for nine years since my undergraduate years. I would like to thank my lovely friends, Soung, Nart, Aim, Ji, Ping, Heng, Ton, Nui, Yong, Jen, Rote, Pick, Pong, Kwang, Guk, N'Pam, N'Bird, N'Bim, N'Pao, N' Au, N'Rong, N'Pop, N'Pun, Kook, Tohn, Poo, P'Lin, P'Komate, P'Woon, P'Petch, P'Jung, P'Yui, P'Nan, P'Noot, P'Ae, P'O, P'Woot, and P'Ko for their wonderful friendships, encouragement, and many valuable discussions. I also would like to thank Took, Moo, Nakorn, P'Tookta, P'Lek, P'Fad, P'Pang, P'Art, P'Dao, P'Hao, P'May, P'Pui, P'Tee-Guay, P'Ae, P'O, N'Parn, N'Neoy, N'Ping, N'Nanah, and N'Giffy for making my stay in the U.S. wonderful; without them, my visit would be colorless. Additionally, I am thankful to the administrative staffs for always being helpful during my whole time attending the Department of Computer Engineering.

I greatly appreciate the financial support from the Thailand Research Fund given through the Royal Golden Jubilee Ph.D. Program (PHD/0141/2549) and the Chulalongkorn University Graduate Scholarship to Commemorate the 72nd Anniversary of His Majesty King Bhumibol Adulyadej for giving me the invaluable opportunity and providing financial support during my precious a half decade of years studying in Ph.D. program and going abroad to the United States. I also greatly appreciate the research fund from the 90th Anniversary of Chulalongkorn University Fund (Ratchadaphiseksomphot Endowment Fund) to make my research significantly forward and my idea becomes reality.

Finally, with my utmost gratitude, this dissertation is dedicated to my beloved parents for shaping my life for what it is today, and to my sisters for always being there. Without their love, encouragement, understanding, and support, this research could not have been completed.

Contents

	Page
Abstract (Thai)	iv
Abstract (English)	v
Acknowledgments	vi
Contents	vii
List of Tables	x
List of Figures	xi
Chapter	
I Introduction	1
1.1 Objective of the Thesis	5
1.2 Scopes of the Thesis	5
1.3 Contributions of the Thesis	6
1.4 Research Methodology	6
II Meaninglessness of Subsequence Time Series Clustering	8
2.1 Background	8
2.1.1 Subsequence Time Series Clustering (STSC)	8
2.1.2 K -Hierarchical Clustering	10
2.1.3 K -Means Clustering	12
2.1.4 Euclidean Distance	13
2.1.5 Amplitude Averaging	14
2.1.6 Z -Normalization	15
2.2 Related Work	16
2.3 Experiments	19
2.3.1 First Experiment	20
2.3.2 Second Experiment	22
2.4 Causes of Meaninglessness	24
2.5 Conclusion	27
III Shape-based Averaging	28
3.1 Background	29
3.1.1 Dynamic Time Warping (DTW) Distance	29
3.1.2 Dynamic Time Warping (DTW) Averaging	30
3.2 Related Work	31

Chapter	Page
3.3 Shape-based Averaging	33
3.3.1 Cubic-Spline Dynamic Time Warping (CDTW) Averaging	33
3.3.2 Iterative Cubic-Spline Dynamic Time Warping (ICDTW) Averaging	35
3.4 Experimental Evaluation	37
3.5 Averaging Trivial-Matched Subsequences	39
3.6 Conclusion	39
IV 2STSC: Shape-based Subsequence Time Series Clustering	41
4.1 Related Work	42
4.2 Shape-based Subsequence Time Series Clustering (2STSC)	47
4.3 Experimental Evaluation	48
4.4 Conclusion	53
V Incremental Shape-based Averaging	54
5.1 Incremental Shape-based Averaging	54
5.2 Experimental Evaluation	56
5.2.1 First Experiment	56
5.2.2 Second Experiment	57
5.3 Conclusion	60
VI 3STSC: Shape-based Streaming Subsequence Time Series Clustering	61
6.1 Related Work	62
6.2 Shape-based Streaming Subsequence Time Series Clustering	63
6.3 Experimental Evaluation	65
6.3.1 First Experiment	65
6.3.2 Second Experiment	67
6.4 Conclusion	68
VII Conclusion	70
VIII Publications	72
References	83
Appendices	84

	Page
Appendix A Datasets	85
Appendix B Complete Experimental Results of the First Experiment in Chapter II	89
Appendix C Complete Experimental Results of the Experiment in Chapter III . .	105
Appendix D Complete Experimental Results of the Experiment in Chapter IV . .	109
Appendix E Complete Experimental Results of the First Experiment in Chapter V	128
Appendix F Complete Experimental Results of the Second Experiment on Chapter V	133
Appendix G Complete Experimental Results of the First Experiment in Chapter VI	150
Appendix H Complete Experimental Results of the Second Experiment in Chapter VI	171
Biography	192

List of Tables

Table	Page
2.1 Pseudo code of Subsequence Time Series Clustering (STSC)	9
2.2 Agglomerative hierarchical clustering algorithm (AGNES)	10
2.3 Pseudo code of single linkage distance function	11
2.4 Pseudo code of complete linkage distance function	12
2.5 Pseudo code of average linkage distance function	12
2.6 Pseudo code of k -means clustering	13
2.7 Pseudo code of Amplitude Averaging function	15
3.1 Pseudo code of Dynamic Time Warping distance measure	30
3.2 Pseudo code of Dynamic Time Warping averaging function	31
3.3 Pseudo code of generating a warping path	32
3.4 Pseudo code of Shape-based Averaging scheme	33
3.5 Pseudo code of Cubic-Spline Dynamic Time Warping (CDTW) averaging function . . .	35
3.6 Pseudo code of Iterative Cubic-Spline Dynamic Time Warping (ICDTW) averaging function	36
3.7 SUMDIST of each averaging method	38
4.1 Pseudo code of Shape-based Subsequence Time Series Clustering (2STSC)	48
5.1 Pseudo code of Incremental Shape-based Averaging	55
5.2 Updating stored sequences in Incremental Shape-based Averaging	55
5.3 Averaging stored sequences in Incremental Shape-based Averaging	56
6.1 Pseudo code of Shape-based Streaming Subsequence Time Series Clustering (3STSC) .	64
6.2 Updating stored sequences in 3STSC	65
A.1 Details of the UCR classification/clustering datasets used in Chapters III and V	88

List of Figures

Figure	Page
1.1 Examples of time series data in real world.	2
1.2 Multivariate time series collected from SmartCane system. (Wu et al., 2008).	3
1.3 Cluster representatives generated from STSC	3
1.4 Trivial-matched subsequences of CBF sequence	4
2.1 Overview of Subsequence Time Series Clustering (STSC)	9
2.2 Example of Euclidean distance calculation.	14
2.3 Example of Amplitude Averaging calculation.	14
2.4 Example of z -normalization.	16
2.5 Examples of Cylinder-Bell-Funnel dataset	17
2.6 Some part of Cylinder-Bell-Funnel sequence	18
2.7 Cluster representatives generated from STSC	18
2.8 Datasets from TSDMA used in the experiments.	20
2.9 Cluster representatives generated from STSC of Buoy1 when $k = 3$ and $w = 64$	21
2.10 Cluster representatives generated from STSC of CBF when $k = 3$ and $w = 64$	21
2.11 Constructed sine waves generated from STSC of Buoy1 when $k = 3$ and $w = 64$	22
2.12 Constructed sine waves generated from STSC of CBF when $k = 3$ and $w = 64$	22
2.13 KLMMs of STSC using k -means clustering.	23
2.14 KLMMs of STSC using k -hierarchical clustering.	24
2.15 Trivial-matched subsequences of CBF sequence	25
2.16 Euclidean distance cannot capture similarity between trivial-matched subsequences	26
2.17 Amplitude Averaging produces a smoothed averaged result.	27
3.1 Comparison between two averaged results generated from Amplitude Averaging and Shape-based Averaging.	28
3.2 Alignment obtained from a DTW distance calculation.	30
3.3 Result generated from DTW Averaging	31
3.4 Comparison between DTW averaging and CDTW averaging functions	34
3.5 Averaged results before and after re-sampling in CDTW averaging function.	35
3.6 Examples of some classes in evaluated datasets.	37
3.7 Averaged results of CBF	38
3.8 Averaged results of ECG	38
3.9 Trivial-matched subsequences b) extracted from a) CBF sequence.	39
3.10 Averaged results generated from Amplitude Averaging.	40
3.11 Averaged results generated from Shape-based Averaging with CDTW function.	40

Figure	Page
3.12 Averaged results generated from Shape-based Averaging with ICDTW function.	40
4.1 Three sets of trivial-matched subsequences.	42
4.2 a) Euclidean cannot capture the similarity of trivial-matched subsequences, while b) DTW can.	43
4.3 a) Amplitude Averaging cannot construct meaningful representatives of trivial- matched subsequences, while b) Shape-based Averaging can.	44
4.4 a) STSC produces a meaningless clustering result, while b) 2STSC produces a meaningful clustering result.	44
4.5 Overview of 2STSC using DTW distance and Shape-based Averaging.	48
4.6 Datasets used to evaluate meaningfulness of STSC and 2STSC	49
4.7 SMMs of Buoy1 when the number of clusters (k) is 3 and the length of sliding window (w) is varied.	51
4.8 SMMs of CBF when the number of clusters (k) is 3 and the length of sliding win- dow (w) is varied.	51
4.9 SMMs of Buoy1 when the length of sliding window (w) is 64 and the number of clusters (k) is varied.	52
4.10 SMMs of CBF when the length of sliding window (w) is 64 and the number of clusters (k) is varied.	52
4.11 Cluster representatives generated from 2STSC of Buoy1 with complete linkage (left) and average linkage (right) when $k = 3$ and $w = 64$	53
4.12 Cluster representatives generated from 2STSC of CBF with complete linkage (left) and average linkage (right) when $k = 3$ and $w = 64$	53
5.1 Examples of some classes in evaluated datasets.	56
5.2 Computational time of Incremental Shape-based Averaging and Shape-based Av- eraging when a new incoming sequence arrives.	57
5.3 Difference of SUMDIST and speedup of Buoy1 when the number of stored se- quences to an original dataset is varied.	58
5.4 Difference of SUMDIST and speedup of CBF when the number of stored sequences to an original dataset is varied.	59
5.5 Averaged results of some classes of CBF from Incremental Shape-based Averaging. . .	59
5.6 Averaged results of some classes of ECG from Incremental Shape-based Averaging. . .	60
6.1 Overview of Shape-based Streaming Subsequence Time Series Clustering (3STSC). . .	63
6.2 Some datasets from TSDMA used in the experiment.	66
6.3 Computational time of 3STSC and 2STSC of Buoy1 when a new incoming se- quence arrives.	66

Figure	Page
6.4 Computational time of 3STSC and 2STSC of CBF when a new incoming sequence arrives.	66
6.5 Percentage difference of SMM and speedup of 3STSC of Buoy1 when $k = 3$, $w = 64$, and number of stored sequences are varied.	67
6.6 Percentage difference of SMM and speedup of 3STSC of CBF when $k = 3$, $w = 64$, and number of stored sequences are varied.	68
A.1 Datasets from TSDMA used in the experiments of Chapters II, IV, and VI.	86
A.2 Examples of some classes of the UCR classification/clustering datasets used in Chapters III and V.	87
B.1 Cluster representatives generated from STSC using k -means clustering when $k = 3$ and $w = 32$	90
B.2 Cluster representatives generated from STSC using k -means clustering when $k = 3$ and $w = 64$	90
B.3 Cluster representatives generated from STSC using k -means clustering when $k = 3$ and $w = 128$	91
B.4 Cluster representatives generated from STSC using k -means clustering when $k = 5$ and $w = 64$	91
B.5 Cluster representatives generated from STSC using k -means clustering when $k = 7$ and $w = 64$	92
B.6 Constructed sine waves generated from STSC using k -means clustering when $k = 3$ and $w = 32$	92
B.7 Constructed sine waves generated from STSC using k -means clustering when $k = 3$ and $w = 64$	93
B.8 Constructed sine waves generated from STSC using k -means clustering when $k = 3$ and $w = 128$	93
B.9 Constructed sine waves generated from STSC using k -means clustering when $k = 5$ and $w = 64$	94
B.10 Constructed sine waves generated from STSC using k -means clustering when $k = 7$ and $w = 64$	94
B.11 Cluster representatives generated from STSC using k -hierarchical clustering with complete linkage (left) and average linkage (right) inter-distance functions when $k = 3$ and $w = 32$	95
B.12 Cluster representatives generated from STSC using k -hierarchical clustering with complete linkage (left) and average linkage (right) inter-distance functions when $k = 3$ and $w = 64$	96

Figure	Page
B.13 Cluster representatives generated from STSC using k -hierarchical clustering with complete linkage (left) and average linkage (right) inter-distance functions when $k = 3$ and $w = 128$	97
B.14 Cluster representatives generated from STSC using k -hierarchical clustering with complete linkage (left) and average linkage (right) inter-distance functions when $k = 5$ and $w = 64$	98
B.15 Cluster representatives generated from STSC using k -hierarchical clustering with complete linkage (left) and average linkage (right) inter-distance functions when $k = 7$ and $w = 64$	99
B.16 Constructed sine waves generated from STSC using k -hierarchical clustering with complete linkage (left) and average linkage (right) inter-distance functions when $k = 3$ and $w = 32$	100
B.17 Constructed sine waves generated from STSC using k -hierarchical clustering with complete linkage (left) and average linkage (right) inter-distance functions when $k = 3$ and $w = 64$	101
B.18 Constructed sine waves generated from STSC using k -hierarchical clustering with complete linkage (left) and average linkage (right) inter-distance functions when $k = 3$ and $w = 128$	102
B.19 Constructed sine waves generated from STSC using k -hierarchical clustering with complete linkage (left) and average linkage (right) inter-distance functions when $k = 5$ and $w = 64$	103
B.20 Constructed sine waves generated from STSC using k -hierarchical clustering with complete linkage (left) and average linkage (right) inter-distance functions when $k = 7$ and $w = 64$	104
C.1 Averaged results generated from CDTW function of each dataset	106
C.2 Averaged results generated from ICDTW function of each dataset	107
C.3 Averaged results generated from NLAAF of each dataset.	108
D.1 SMMs of AEM2 when the number of clusters (k) is 3 and the length of sliding window (w) is varied.	110
D.2 SMMs of TOR96 when the number of clusters (k) is 3 and the length of sliding window (w) is varied.	110
D.3 SMMs of Buoy1 when the number of clusters (k) is 3 and the length of sliding window (w) is varied.	111
D.4 SMMs of CBF when the number of clusters (k) is 3 and the length of sliding window (w) is varied.	111

Figure	Page
D.5 SMMs of ERP when the number of clusters (k) is 3 and the length of sliding window (w) is varied.	112
D.6 SMMs of Field4 when the number of clusters (k) is 3 and the length of sliding window (w) is varied.	112
D.7 SMMs of Fortune5004 when the number of clusters (k) is 3 and the length of sliding window (w) is varied.	113
D.8 SMMs of MITDBX108 when the number of clusters (k) is 3 and the length of sliding window (w) is varied.	113
D.9 SMMs of AEM2 when the length of sliding window (w) is 64 and the number of clusters (k) is varied.	114
D.10SMMs of TOR96 when the length of sliding window (w) is 64 and the number of clusters (k) is varied.	114
D.11SMMs of Buoy1 when the length of sliding window (w) is 64 and the number of clusters (k) is varied.	115
D.12SMMs of CBF when the length of sliding window (w) is 64 and the number of clusters (k) is varied.	115
D.13SMMs of ERP when the length of sliding window (w) is 64 and the number of clusters (k) is varied.	116
D.14SMMs of Field4 when the length of sliding window (w) is 64 and the number of clusters (k) is varied.	116
D.15SMMs of Fortune5004 when the length of sliding window (w) is 64 and the number of clusters (k) is varied.	117
D.16SMMs of MITDBX108 when the length of sliding window (w) is 64 and the number of clusters (k) is varied.	117
D.17Cluster representatives generated from 2STSC with complete linkage (left) and average linkage (right) using CDTW function when $k = 3$ and $w = 32$	118
D.18Cluster representatives generated from 2STSC with complete linkage (left) and average linkage (right) using CDTW function when $k = 3$ and $w = 64$	119
D.19Cluster representatives generated from 2STSC with complete linkage (left) and average linkage (right) using CDTW function when $k = 3$ and $w = 128$	120
D.20Cluster representatives generated from 2STSC with complete linkage (left) and average linkage (right) using ICDTW function when $k = 5$ and $w = 64$	121
D.21Cluster representatives generated from 2STSC with complete linkage (left) and average linkage (right) using ICDTW function when $k = 7$ and $w = 64$	122

Figure	Page
D.22 Cluster representatives generated from 2STSC with complete linkage (left) and average linkage (right) using ICDTW function when $k = 3$ and $w = 32$	123
D.23 Cluster representatives generated from 2STSC with complete linkage (left) and average linkage (right) using ICDTW function when $k = 3$ and $w = 64$	124
D.24 Cluster representatives generated from 2STSC with complete linkage (left) and average linkage (right) using ICDTW function when $k = 3$ and $w = 128$	125
D.25 Cluster representatives generated from 2STSC with complete linkage (left) and average linkage (right) using ICDTW function when $k = 5$ and $w = 64$	126
D.26 Cluster representatives generated from 2STSC with complete linkage (left) and average linkage (right) using ICDTW function when $k = 7$ and $w = 64$	127
E.1 Computational time of Incremental Shape-based Averaging and Shape-based Averaging with CDTW function when a new incoming sequence arrives.	129
E.2 Computational time of Incremental Shape-based Averaging and Shape-based Averaging with CDTW function when a new incoming sequence arrives. (cont.)	130
E.3 Computational time of Incremental Shape-based Averaging and Shape-based Averaging with ICDTW function when a new incoming sequence arrives.	131
E.4 Computational time of than Shape-based Averaging around Incremental Shape-based Averaging and Shape-based Averaging with ICDTW function when a new incoming sequence arrives.	132
F.1 Difference of SUMDIST of Incremental Shape-based Averaging with CDTW when the number of stored sequences to an original dataset is varied.	134
F.2 Difference of SUMDIST of Incremental Shape-based Averaging with CDTW when the number of stored sequences to an original dataset is varied. (cont.)	135
F.3 Difference of SUMDIST of Incremental Shape-based Averaging with ICDTW when the number of stored sequences to an original dataset is varied.	136
F.4 Difference of SUMDIST of Incremental Shape-based Averaging with ICDTW when the number of stored sequences to an original dataset is varied. (cont.)	137
F.5 Speedup of Incremental Shape-based Averaging with CDTW when the number of stored sequences to an original dataset is varied.	138
F.6 Speedup of Incremental Shape-based Averaging with CDTW when the number of stored sequences to an original dataset is varied. (cont.)	139
F.7 Speedup of Incremental Shape-based Averaging with ICDTW when the number of stored sequences to an original dataset is varied.	140
F.8 Speedup of Incremental Shape-based Averaging with CDTW when the number of stored sequences to an original dataset is varied. (cont.)	141

Figure	Page
F.9 Averaged results of some classes from Incremental Shape-based Averaging with CDTW when $\alpha = 1$	142
F.10 Averaged results of some classes from Incremental Shape-based Averaging with CDTW when α is 25% of total number of each class.	143
F.11 Averaged results of some classes from Incremental Shape-based Averaging with CDTW when α is 50% of total number of each class.	144
F.12 Averaged results of some classes from Incremental Shape-based Averaging with CDTW when α is 100% of total number of each class.	145
F.13 Averaged results of some classes from Incremental Shape-based Averaging with ICDTW when $\alpha = 1$	146
F.14 Averaged results of some classes from Incremental Shape-based Averaging with ICDTW when α is 25% of total number of each class.	147
F.15 Averaged results of some classes from Incremental Shape-based Averaging with ICDTW when α is 50% of total number of each class.	148
F.16 Averaged results of some classes from Incremental Shape-based Averaging with ICDTW when α is 100% of total number of each class.	149
G.1 Computational time of 3STSC and 2STSC with CDTW function and complete linkage when a new incoming sequence arrives, where $k = 3$ and $w = 64$	151
G.2 Computational time of 3STSC and 2STSC with CDTW function and complete linkage when a new incoming sequence arrives, where $k = 3$ and $w = 32$	152
G.3 Computational time of 3STSC and 2STSC with CDTW function and complete linkage when a new incoming sequence arrives, where $k = 5$ and $w = 64$	153
G.4 Computational time of 3STSC and 2STSC with CDTW function and complete linkage when a new incoming sequence arrives, where $k = 7$ and $w = 64$	154
G.5 Computational time of 3STSC and 2STSC with CDTW function and complete linkage when a new incoming sequence arrives, where $k = 3$ and $w = 128$	155
G.6 Computational time of 3STSC and 2STSC with CDTW function and average linkage when a new incoming sequence arrives, where $k = 3$ and $w = 64$	156
G.7 Computational time of 3STSC and 2STSC with CDTW function and average linkage when a new incoming sequence arrives, where $k = 3$ and $w = 32$	157
G.8 Computational time of 3STSC and 2STSC with CDTW function and average linkage when a new incoming sequence arrives, where $k = 5$ and $w = 64$	158
G.9 Computational time of 3STSC and 2STSC with CDTW function and average linkage when a new incoming sequence arrives, where $k = 7$ and $w = 64$	159

Figure	Page
G.10 Computational time of 3STSC and 2STSC with CDTW function and average linkage when a new incoming sequence arrives, where $k = 3$ and $w = 128$	160
G.11 Computational time of 3STSC and 2STSC with CDTW function and complete linkage when a new incoming sequence arrives, where $k = 3$ and $w = 64$	161
G.12 Computational time of 3STSC and 2STSC with CDTW function and complete linkage when a new incoming sequence arrives, where $k = 3$ and $w = 32$	162
G.13 Computational time of 3STSC and 2STSC with CDTW function and complete linkage when a new incoming sequence arrives, where $k = 5$ and $w = 64$	163
G.14 Computational time of 3STSC and 2STSC with CDTW function and complete linkage when a new incoming sequence arrives, where $k = 7$ and $w = 64$	164
G.15 Computational time of 3STSC and 2STSC with CDTW function and complete linkage when a new incoming sequence arrives, where $k = 3$ and $w = 128$	165
G.16 Computational time of 3STSC and 2STSC with CDTW function and average linkage when a new incoming sequence arrives, where $k = 3$ and $w = 64$	166
G.17 Computational time of 3STSC and 2STSC with CDTW function and average linkage when a new incoming sequence arrives, where $k = 3$ and $w = 32$	167
G.18 Computational time of 3STSC and 2STSC with CDTW function and average linkage when a new incoming sequence arrives, where $k = 5$ and $w = 64$	168
G.19 Computational time of 3STSC and 2STSC with CDTW function and average linkage when a new incoming sequence arrives, where $k = 7$ and $w = 64$	169
G.20 Computational time of 3STSC and 2STSC with CDTW function and average linkage when a new incoming sequence arrives, where $k = 3$ and $w = 128$	170
H.1 Percentage difference of SMM and speedup of 3STSC with CDTW function and complete linkage when $k = 3$, $w = 64$, and number of stored sequences are varied. . . .	172
H.2 Percentage difference of SMM and speedup of 3STSC with CDTW function and complete linkage when $k = 3$, $w = 32$, and number of stored sequences are varied. . . .	173
H.3 Percentage difference of SMM and speedup of 3STSC with CDTW function and complete linkage when $k = 5$, $w = 64$, and number of stored sequences are varied. . . .	174
H.4 Percentage difference of SMM and speedup of 3STSC with CDTW function and complete linkage when $k = 7$, $w = 64$, and number of stored sequences are varied. . . .	175
H.5 Percentage difference of SMM and speedup of 3STSC with CDTW function and complete linkage when $k = 3$, $w = 128$, and number of stored sequences are varied. . . .	176
H.6 Percentage difference of SMM and speedup of 3STSC with CDTW function and average linkage when $k = 3$, $w = 64$, and number of stored sequences are varied. . . .	177

Figure	Page
H.7 Percentage difference of SMM and speedup of 3STSC with CDTW function and average linkage when $k = 3$, $w = 64$, and number of stored sequences are varied. . . .	178
H.8 Percentage difference of SMM and speedup of 3STSC with CDTW function and average linkage when $k = 5$, $w = 64$, and number of stored sequences are varied. . . .	179
H.9 Percentage difference of SMM and speedup of 3STSC with CDTW function and average linkage when $k = 7$, $w = 64$, and number of stored sequences are varied. . . .	180
H.10 Percentage difference of SMM and speedup of 3STSC with CDTW function and average linkage when $k = 3$, $w = 128$, and number of stored sequences are varied. . . .	181
H.11 Percentage difference of SMM and speedup of 3STSC with ICDTW function and complete linkage when $k = 3$, $w = 64$, and number of stored sequences are varied. . . .	182
H.12 Percentage difference of SMM and speedup of 3STSC with ICDTW function and complete linkage when $k = 3$, $w = 32$, and number of stored sequences are varied. . . .	183
H.13 Percentage difference of SMM and speedup of 3STSC with ICDTW function and complete linkage when $k = 5$, $w = 64$, and number of stored sequences are varied. . . .	184
H.14 Percentage difference of SMM and speedup of 3STSC with ICDTW function and complete linkage when $k = 7$, $w = 64$, and number of stored sequences are varied. . . .	185
H.15 Percentage difference of SMM and speedup of 3STSC with ICDTW function and complete linkage when $k = 3$, $w = 128$, and number of stored sequences are varied. . . .	186
H.16 Percentage difference of SMM and speedup of 3STSC with ICDTW function and average linkage when $k = 3$, $w = 64$, and number of stored sequences are varied. . . .	187
H.17 Percentage difference of SMM and speedup of 3STSC with ICDTW function and average linkage when $k = 3$, $w = 32$, and number of stored sequences are varied. . . .	188
H.18 Percentage difference of SMM and speedup of 3STSC with ICDTW function and average linkage when $k = 3$, $w = 64$, and number of stored sequences are varied. . . .	189
H.19 Percentage difference of SMM and speedup of 3STSC with ICDTW function and average linkage when $k = 7$, $w = 64$, and number of stored sequences are varied. . . .	190
H.20 Percentage difference of SMM and speedup of 3STSC with ICDTW function and average linkage when $k = 3$, $w = 128$, and number of stored sequences are varied. . . .	191