

## DOCTORAL THESIS

### Shapelet Discovery for Time Series Analysis

LI, Guozhong

*Date of Award:*  
2021

[Link to publication](#)

#### **General rights**

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

# Abstract

Time series shapelets (or simply, shapelets) are discriminative subsequences that have been recently found effective for time series analysis, including classification and clustering. The quality of shapelets is crucial to both the accuracy and efficiency of time series analysis. However, the major of research has focused on building accurate models from some shapelet candidates rather than *discovering* high-quality shapelets. Discovery of high-quality shapelets is known to be computationally costly. Furthermore, shapelet discovery for time series analysis has a few challenges. Among the few existing work on shapelet discovery, they cannot be applied to multivariate time series classification (MTSC) since the shapelet candidates of MTSC may come from different variables of different lengths and thus cannot be directly compared. The current unsupervised shapelet discovery method cannot determine clustering quality on both univariate and multivariate time series. In this dissertation, first, we propose BSPCOVER to discover a set of high-quality shapelets for model building and then propose a matrix profile-based shapelet approach (MPS) to further improve the efficiency of shapelet discovery. Second, we propose ShapeNet for MTSC, a deep learning approach for discovering shapelets, and Autoencoder for Shapelets (AUTOSHAPe) for clustering, which takes the advantages of both shapelets and autoencoder for determining shapelets in an unsupervised manner. Experimental evaluation on well-known time series datasets (UCR and UEA archive) shows the superiority of our proposed approaches on both accuracy and efficiency. Last but not least, one of the strengths of shapelets is its interpretability. For all of our proposed

solutions, we illustrate the interpretability of shapelets with some case studies.

**Keywords:** Time Series Analysis, Shapelet Discovery, Classification, Clustering, Efficiency, Accuracy

# Table of Contents

<b>Declaration</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>Acknowledgements</b>	<b>iv</b>
<b>Table of Contents</b>	<b>v</b>
<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xiii</b>
<b>List of Algorithms</b>	<b>xv</b>
<b>Chapter 1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Efficient Shapelet Discovery for Univariate Time Series Classification - A Heuristic Approach . . . . .	3
1.3 Efficient Shapelet Discovery for Univariate Time Series Classification - A Matrix Profile Approach . . . . .	7
1.4 A Shapelet-Neural Network Approach for Multivariate Time Series Classification . . . . .	11
1.5 Shapelet Autoencoder for Time Series Clustering . . . . .	14
1.6 Dissertation Organization . . . . .	15
<b>Chapter 2 Preliminaries and Literature Review</b>	<b>17</b>
2.1 Preliminaries . . . . .	17
2.2 Univariate Time Series Classification . . . . .	19

2.2.1	Symbolic Representation-based Methods . . . . .	20
2.2.2	Matrix Profile-based methods . . . . .	20
2.2.3	Shapelet-based methods . . . . .	21
2.3	Multivariate Time Series Classification . . . . .	21
2.3.1	Model-based Methods . . . . .	21
2.3.2	Neural Network-based Methods . . . . .	22
2.4	Time Series Clustering . . . . .	22
2.4.1	Autoencoder-based Methods . . . . .	23
2.4.2	Shapelet-based Methods . . . . .	23

**Chapter 3 Efficient Shapelet Discovery for Univariate Time Series Classification - A Heuristic Approach** **25**

3.1	Preliminaries and Problem Statement . . . . .	25
3.2	Efficient SAX Subsequence Computation . . . . .	27
3.2.1	SAX Transformation . . . . .	27
3.2.2	Bloom Filter and Bitmap of SAX Words . . . . .	30
3.2.3	Similar SAX Word Pruning . . . . .	33
3.2.4	SAX Word Cover for Model Building . . . . .	38
3.2.5	$p$ -Cover Setting . . . . .	40
3.3	Experimental Results . . . . .	44
3.3.1	Environment . . . . .	44
3.3.2	Datasets and Parameters . . . . .	45
3.3.3	Baselines . . . . .	45
3.3.4	Experiments on Efficiency . . . . .	47
3.3.5	Experiments on Accuracy . . . . .	51
3.3.6	Experiments on Interpretability . . . . .	55
3.4	Chapter Summary . . . . .	56

<b>Chapter 4</b>	<b>Efficient Shapelet Discovery Univariate Time Series Classification - A Matrix Profile Approach</b>	<b>59</b>
4.1	Preliminaries . . . . .	59
4.1.1	Terminologies and Notations . . . . .	59
4.1.2	Three Issues of an MP Baseline Method . . . . .	60
4.2	MPS for Shapelets . . . . .	63
4.2.1	Shapelet Candidates Pool Generation . . . . .	64
4.2.2	Top- $k$ Shapelet Generation . . . . .	66
4.3	Experimental Results . . . . .	69
4.3.1	Experimental Settings . . . . .	70
4.3.2	Experiments on Efficiency . . . . .	70
4.3.3	Experiments on Accuracy . . . . .	73
4.3.4	Experiments on Interpretability . . . . .	76
4.4	Chapter Summary . . . . .	77
<b>Chapter 5</b>	<b>ShapeNet: A Shapelet-Neural Network Approach for Multivariate Time Series Classification</b>	<b>78</b>
5.1	Problem Statement . . . . .	78
5.2	Shapelet Neural Network Approach for MTS classification . . . . .	80
5.2.1	Multi-length-input Dilated Causal CNN (Mdc-CNN) . . . . .	80
5.2.2	Unsupervised Representation Learning . . . . .	81
5.2.3	Differentiation of the Loss Function . . . . .	84
5.2.4	Data Preprocessing . . . . .	86
5.2.5	Multivariate Shapelet Transformation . . . . .	89
5.3	Experimental Results . . . . .	93
5.3.1	Environment . . . . .	93
5.3.2	Datasets and Parameters . . . . .	93
5.3.3	Baselines . . . . .	94

5.3.4	Experiments on Convergence of Mdc-CNN . . . . .	95
5.3.5	Experiments on Accuracy . . . . .	95
5.3.6	Experiments on Interpretability . . . . .	99
5.4	Chapter Summary . . . . .	101
<b>Chapter 6</b>	<b>Shapelet Autoencoder for Time Series Clustering</b>	<b>103</b>
6.1	Preliminaries and Problem Statement . . . . .	103
6.2	Autoencoder for Shapelets (AUTOSHAPe) . . . . .	104
6.2.1	Shapelet Discovery . . . . .	104
6.2.2	Shapelet Adjustment . . . . .	107
6.3	Experimental Results . . . . .	110
6.3.1	Experimental Settings . . . . .	111
6.3.2	Baselines . . . . .	111
6.3.3	Experiments on Univariate Time Series . . . . .	112
6.3.4	Experiments on Multivariate Time Series . . . . .	119
6.4	Chapter Summary . . . . .	122
<b>Chapter 7</b>	<b>Conclusions</b>	<b>124</b>
7.1	Contributions . . . . .	124
7.2	Future Directions . . . . .	126
<b>Appendix A</b>	<b>Learning the Shapelets</b>	<b>128</b>
<b>Bibliography</b>		<b>131</b>
<b>List of Publications</b>		<b>142</b>
<b>Curriculum Vitae</b>		<b>144</b>

# List of Figures

1.1	The overview of <code>BSPCOVER</code> for shapelet discovery . . . . .	3
1.2	Concatenations of time series ( $T_A$ and $T_B$ ) of two classes (namely, $A$ and $B$ ) taken from the ArrowHead dataset of UCR Archive [17] . . . . .	8
1.3	Two shapelets of class A for ArrowHead in Figure 1.2 . . . . .	8
1.4	The matrix profiles of ( $T_A, T_B$ ) and ( $T_A, T_A$ ) of Figure 1.2 – $P_{AB}$ and $P_{AA}$ . . . . .	8
1.5	The difference between the matrix profile $P_{AB}$ and $P_{AA}$ of Figure 1.4 – $\text{diff}(P_{AB}, P_{AA})$ . . . . .	9
1.6	The overview of matrix profile for shapelets (MPS) for time series classification . . . . .	10
1.7	The overview of ShapeNet for multivariate time series classification . . . . .	11
1.8	The overview of <code>AUTO SHAPE</code> for time series clustering . . . . .	14
3.1	An example of PAA with $\omega = 10$ and SAX alphabet size with $ \Sigma  = 8$ from the Beef dataset for SAX sequence <b>DEFFF</b> . . . . .	26
3.2	The step function $\pi$ of $\omega$ . . . . .	29
3.3	An example of SAX transformation . . . . .	30
3.4	Pruning redundant and non-discriminative SAX words (of 3 classes) using bloom filters . . . . .	32
3.5	An example of (partial) constructing bitmaps of SAX words in Class 1 ( $C_1$ ) . . . . .	33



3.6	An example of finding SAX words for model building (The figure is best viewed in color) . . . . .	40
3.7	Efficiency vs $p$ (of $p$ -Cover) on four UCR datasets . . . . .	51
3.8	Accuracy of Random vs BSPCOVER in box plots on four UCR datasets	54
3.9	Accuracy vs $p$ (of $p$ -Cover) on four UCR datasets . . . . .	55
3.10	Accuracy vs iterations of BSPCOVER and ELIS on four UCR datasets	56
3.11	ECGFiveDays shapelets encodes hyperacute T wave . . . . .	57
3.12	A shapelet of ItalyPowerDemand highlighting the morning heating demand difference of summer and winter months . . . . .	58
4.1	An example of matrix profiles and the difference to show the first issue Discords as “shapelets” . . . . .	61
4.2	Efficiency vs Top- $k$ (shapelet number) of BSPCOVER, BASE, MPS methods on BeetleFly . . . . .	71
4.3	Efficiency vs Top- $k$ (shapelet number) of BSPCOVER, BASE, MPS methods on TwoLeadECG . . . . .	73
4.4	Critical difference diagram of the pairwise statistical comparison of 11 methods on the UCR Archive . . . . .	74
4.5	Accuracy by varying five kinds of shapelet’s numbers on four UCR datasets . . . . .	76
4.6	Different shapelets (green) from MPS and BSPCOVER of ItalyPowerDemand highlighting the morning heating demand difference of summer (red) and winter (blue) months . . . . .	77
5.1	An illustration of the best match location of a subsequence in a time series . . . . .	78
5.2	An illustration of a multivariate time series dataset — extracted from Basicmotions [42] . . . . .	79

5.3	An elaboration of the Multi-length-input dilated causal Convolutional Neural Network (Mdc-CNN) . . . . .	81
5.4	An example from Basicmotions of the violation of the second requirement of word2vec: subsequences that are far away but have a small distance between them . . . . .	82
5.5	A comparison between our cluster-wise triplet loss (multiple positives and multiple negatives, both with intra distances) and original triplet loss (one positive and multiple negatives without intra distance) on ArticulatoryWordRecognition [42] . . . . .	83
5.6	An illustration of the effect of training a model using the cluster-wise triplet loss function, positives are closer to each other and the anchor, negatives are closer to each other but farther from the anchor . . . .	86
5.7	An illustration of triplet sampling, black for Anchor, red for Positive samples, blue for Negative samples [best viewed in color] . . . . .	89
5.8	An illustration of transforming an mrs instance into the MST representation . . . . .	91
5.9	The convergences of the learning algorithm on some multivariate time series datasets . . . . .	96
5.10	Triplet sampling vs random sampling of final shapelets . . . . .	98
5.11	Utility-based vs random selection of final shapelets . . . . .	99
5.12	Multivariate time series classification accuracy by varying 6 shapelet numbers . . . . .	100
5.13	An example of multivariate shapelet transformation on Basicmotions [42] . . . . .	100
5.14	An example of multivariate shapelet transformation on Atrialfibrillation [42] . . . . .	102

6.1	Critical difference diagram of the pairwise statistical comparison of 15 methods on the UCR Archive . . . . .	115
6.2	NMI by varying 5 numbers of shapelet on four UCR datasets [17] . .	115
6.3	An example of shapelet transformation on ToeSegmentation1 [17] .	117
6.4	An example of shapelet transformation on ItalyPowerDemand [17] .	118
6.5	The original BirdChicken dataset [17], the red lines show the shapelet $S_1$ corresponding area . . . . .	118
6.6	An example of shapelet transformation on BirdChicken [17] . . . .	119
6.7	Critical difference diagram of the pairwise statistical comparison of 6 methods on the UEA Archive . . . . .	121
6.8	NMI by varying 5 numbers of shapelet on four UEA datasets [42] . .	121
6.9	An example of shapelet transformation on one mts dataset, called Epilepsy [42] . . . . .	122

# List of Tables

1.1 Shapelet discovery and model training time complexity of some shapelet-based methods [2] . . . . .	4
2.1 Summary of frequently used notations in this thesis . . . . .	19
3.1 Summary of frequently used notations in BSPCOVER . . . . .	27
3.2 Datasets and parameters on the tested UCR ARCHIVE . . . . .	46
3.3 Efficiency of BSPCOVER and related methods on UCR ARCHIVE (units is second; d and h denotes days and hours) . . . . .	49
3.4 Efficiency of BSPCOVER and ELIS on UCR ARCHIVE . . . . .	50
3.5 Accuracy of BSPCOVER and 10 related methods on UCR ARCHIVE . . . . .	53
4.1 Summary of frequently used notations in MPS . . . . .	60
4.2 The accuracy of the different top- $k$ Shapelets on MP baseline, 1NN-ED, and 1NN-DTW on four datasets from UCR Archive . . . . .	63
4.3 Efficiency of MPS and related methods on UCR Archive (second as unit) and the speedup . . . . .	72
4.4 Accuracy of MPS and related methods on UCR Archive . . . . .	75
5.1 Multivariate time series datasets information from UEA ARCHIVE . . . . .	94
5.2 Accuracy of ShapeNet and the related methods on UEA ARCHIVE . . . . .	97
6.1 Normalized mutual information (NMI) comparison on the UCR ARCHIVE . . . . .	114

6.2 Normalized mutual information (NMI) ablation comparisons on UCR  
ARCHIVE . . . . . 116

6.3 Normalized mutual information (NMI) comparison on the UEA  
ARCHIVE . . . . . 120

# List of Algorithms

3.1	SAX transformation for class $C$ . . . . .	31
3.2	Construction of bloom filters and the SAX words' bitmaps of class $C$ . . . . .	34
3.3	Computing the weights for non-similar SAX words . . . . .	37
3.4	ONESAXCOVER: Heuristic algorithm for the weighted bitmap cover problem . . . . .	41
3.5	PSAXCOVER: Determining $p$ -Cover SAX words . . . . .	42
3.6	Determining $p$ for BSPCOVER . . . . .	43
4.1	Generation of the pool for shapelet candidates . . . . .	66
4.2	Top- $k$ shapelet generation . . . . .	69
5.1	Shapelet candidate generation . . . . .	87
5.2	Selection of triplet (APN) . . . . .	88
5.3	Multivariate Shapelet Transformation . . . . .	92
6.1	Shapelet candidate generation . . . . .	105
6.2	Determining final shapelets . . . . .	110
A.1	Shapelet learning [1] . . . . .	130