

**AUTOMATED RULE GENERATION  
FOR  
COMPLEX EVENT PROCESSING**

DILUM CHANDIMA VITHANA

(158254P)

Degree of Master of Science

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

February 2019

**AUTOMATED RULE GENERATION  
FOR  
COMPLEX EVENT PROCESSING**

DILUM CHANDIMA VITHANA

(158254P)

Thesis/Dissertation submitted in partial fulfillment of the requirements for the degree  
Master of Science

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

February 2019

## DECLARATION

I declare that this is my own work and this thesis/dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant the University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part in print, electronic or any other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

**Signature:**.....

**Date:** .....

**Name:** D.C Vithana

The above candidate has carried out research for the Masters/MPhil/Ph.D. thesis/dissertation under my supervision.

**Signature of the supervisor:**.....

**Date:** .....

**Name:** Dr. Surangika Ranathunga

## **Abstract**

The key concept of Complex Event Processing (CEP) is setting up accurate queries to pick up the important events. Since all CEP engines support SQL like queries, this rule setup requires some technical skills plus domain knowledge. The best solution to address this issue is to automate the query generation of CEP. Existing automated query generation methodologies are computationally expensive and are not fully automated processes. This study addresses the above two issues by proposing a shapelet based approach. This new approach is not computationally expensive, and it is a fully automated process with zero manual user intervention required. The proposed method uses the computationally efficient algorithm called Fast Shapelet Selection (FSS) algorithm. This FSS algorithm is used to extract the shapelets from data set. Then extracted shapelets are used to generate CEP queries. This proposed method can be used analyze to multivariant time series and this is more efficient than previously proposed shapelet based approaches.

## **Acknowledgements**

First and foremost, I would like to thank my supervisor, Dr. Surangika Ranathunga for the valuable guidance given throughout to successfully complete this research. I would also like to thank Department of Computer Science and Engineering, Faculty of Engineering, University of Moratuwa Sri Lanka for their support.

Finally, I must express my very profound gratitude to my family for providing continuous support and encouragement.

# TABLE OF CONTENTS

## Contents

|   |             |
|---|-------------|
| <b>DECLARATION</b> .....  | <b>I</b>    |
| <b>ABSTRACT</b> .....   | <b>II</b>   |
| <b>ACKNOWLEDGEMENTS</b> .....   | <b>III</b>  |
| <b>TABLE OF CONTENTS</b> .....  | <b>IV</b>   |
| <b>LIST OF FIGURES</b> .....  | <b>VI</b>   |
| <b>LIST OF TABLES</b> .....   | <b>VII</b>  |
| <b>LIST OF ABBREVIATIONS</b> .....  | <b>VIII</b> |
| <b>1 CHAPTER 1: INTRODUCTION</b> .....                                    | <b>1</b>    |
| 1.1 BACKGROUND.....   | 1           |
| 1.2 PROBLEM AND MOTIVATION .....  | 1           |
| 1.3 PROBLEM STATEMENT.....  | 2           |
| 1.4 RESEARCH OBJECTIVE .....  | 2           |
| 1.5 SCOPE OF THE RESEARCH .....   | 3           |
| 1.6 RESEARCH CONTRIBUTION .....   | 3           |
| 1.7 OUTLINE.....  | 3           |
| <b>2 CHAPTER 2: LITERATURE REVIEW</b> .....                               | <b>4</b>    |
| 2.1 COMPLEX EVENT PROCESSING .....  | 4           |
| 2.2 AUTOMATED COMPLEX EVENT GENERATION .....                              | 4           |
| 2.2.1 <i>Machine Learning Approaches</i> .....                            | 5           |
| 2.2.2 <i>Learning Event Detection rule with Hidden Markov Model</i> ..... | 9           |
| 2.2.3 <i>Recommendation Based Pattern Recognition</i> .....               | 11          |
| 2.2.4 <i>Shapelet Based Approach</i> .....                                | 13          |
| 2.3 SHAPELET AND SHAPELET EXTRACTION ALGORITHMS .....                     | 17          |
| 2.3.1 <i>Definition of the Shapelet</i> .....                             | 17          |
| 2.3.2 <i>Brute Force Algorithm</i> .....                                  | 18          |
| 2.3.3 <i>Shapelet Discovery Algorithms (FS)</i> .....                     | 18          |

|          |  |           |
|----------|--|-----------|
| 2.3.4    | <i>Shapelet Transformation Algorithm (ST)</i> .....  | 21        |
| 2.3.5    | <i>Learning Shapelet Algorithm (LS)</i> .....        | 22        |
| 2.3.6    | <i>Fast Shapelet Selection Algorithm (FSS)</i> ..... | 22        |
| <b>3</b> | <b>CHAPTER 3: METHODOLOGY</b> .....                  | <b>30</b> |
| 3.1      | HIGH LEVEL DESIGN .....                              | 30        |
| 3.1.1    | <i>Shaplet Learner Module</i> .....                  | 30        |
| 3.1.2    | <i>Query Generator Module</i> .....                  | 35        |
| <b>4</b> | <b>CHAPTER 4: RESULTS AND ANALYSIS</b> .....         | <b>36</b> |
| 4.1      | EVALUATION OF THE ALGORITHM .....                    | 36        |
| 4.1.1    | <i>Total Time for Shapelet Extraction</i> .....      | 36        |
| 4.1.2    | <i>Data Set</i> .....                                | 38        |
| 4.1.3    | <i>Evaluation</i> .....                              | 38        |
| 4.2      | RESULT COMPARISON .....                              | 41        |
| 4.3      | RESULT SUMMARY .....                                 | 44        |
| <b>5</b> | <b>CHAPTER 5: CONCLUSION AND FUTURE WORK</b> .....   | <b>46</b> |
| 5.1      | CONCLUSION .....                                     | 46        |
| 5.2      | FUTURE WORK .....                                    | 46        |
| <b>6</b> | <b>REFERENCES</b> .....                              | <b>48</b> |

## LIST OF FIGURES

|  |    |
|--|----|
| Figure 2.1 Sample Event Sequence [4] .....                                 | 10 |
| Figure 2.2 HMM for pattern ABC [4] .....                                   | 11 |
| Figure 2.3 Distance computation required between a pair of subseq[6] ..... | 19 |
| Figure 2.4 Sequence which contains subclasses .....                        | 23 |
| Figure 2.5 Identify IDP Point [19] .....                                   | 25 |
| Figure 2.6 Accuracy comparison [15] .....                                  | 27 |
| Figure 3.1 Steps of the Process .....                                      | 31 |
| Figure 4.1 Accuracy Evaluation .....                                       | 37 |
| Figure 4.2 Running Time Comparison .....                                   | 39 |
| Figure 4.3 ST Algorithm Running Time Comparison .....                      | 43 |
| Figure 4.4 Outlier of the EEG Data Set .....                               | 44 |

## LIST OF TABLES

|  |    |
|--|----|
| Table 4.1 Calculation Formula.....                             | 37 |
| Table 4.2 Occupancy Detection Data Set.....                    | 38 |
| Table 4.3 EEG Eye State Data Set .....                         | 38 |
| Table 4.4 FSS algorithm Running Time.....                      | 39 |
| Table 4.5 FSS - Occupancy Data Evaluation .....                | 40 |
| Table 4.6 FSS-EEG Data Evaluation .....                        | 41 |
| Table 4.7 Running Time (ms) Comparison between FSS and ST..... | 42 |
| Table 4.8 Accuracy Comparison .....                            | 43 |

## LIST OF ABBREVIATIONS

|         |  |
|---------|--|
| AEP     | Admissible Entropy Pruning                                     |
| BF      | Brute Force Algorithm  |
| CEP     | Complex Event Processing                                       |
| DBMS    | Database Management Systems                                    |
| EDL     | Event Description Language                                     |
| FS      | Fast Shapelet Discovery  |
| FSS     | Fast Shapelet Selection Algorithm                              |
| GRSF    | Generalized Random Shapelet Forest                             |
| HMM     | Hidden Markov Model (HMM)                                      |
| IDP     | Important Data Point   |
| LS      | Learning Shapelet Algorithm                                    |
| nHMM    | noise Hidden Markov Model                                      |
| PLR     | Piecewise Linear Representation (PLR)                          |
| PLR_IDP | Piecewise Linear Representation based on Important Data Points |
| RS      | Random Shapelet (RS)   |
| ST      | Shapelet Transformation Algorithm (ST)                         |
| SDEA    | Subsequence Distance Early Abandon (SDEA)                      |
| LFDP    | Local Farthest Deviation Points (LFDP)                         |