

LB/DON/106/2016
IT 01/134

Using Data Mining Techniques to Analyze Crime patterns in Sri Lanka National Crime Data

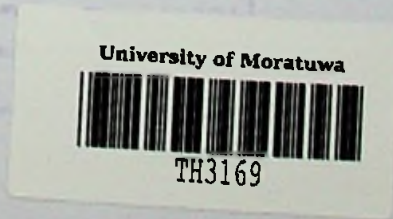
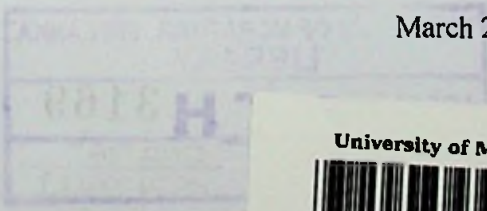
K.P.S.D. Kumarapathirana

139169A

LIBRARY
UNIVERSITY OF MORATUWA, SRI LANKA
MORATUWA

Dissertation submitted to the
Faculty of Information Technology, University of Moratuwa, Sri Lanka
for the partial fulfillment of the requirements of the
Degree of Master of Science in Information Technology.

March 2016



004 "16"
004 (013)

TH3169
+
DVD-ROM

(TH 3160 - TH 3180)

TH 3169

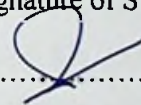
Declaration

We declare that this thesis is our own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Name of Student

K. P. S. D. Kumarapathirana

Signature of Student



.....

Date: 24.04.2016

Supervised by

Name of Supervisor

S. C. Premaratne

Signature of Supervisor

UOM Verified Signature

Date: 24/04/2016

Acknowledgements

I would like to express my gratitude to my supervisor, Mr. S. C. Premaratne, Senior Lecturer in University of Moratuwa, Sri Lanka whose expertise, understanding, and patience, added considerably to my research experience. I appreciate his vast knowledge and skill in many areas, and his assistance in writing reports.

I would like to thank the other lecturers of University of Moratuwa, Sri Lanka, especially, Prof. Asoka Karunananda for the knowledge and assistance they provided at all levels of the research project.

Moreover, a very special thank should go out to Mr. C. V. Millawithana, Criminologist, Crime Record Division, Department of Police, Sri Lanka and his staff members for assisting me in collecting data sets and giving me their valuable comments on the research goals and objectives.

I would also like to thank all the batch mates of the M.Sc. in IT degree program who gave their valuable feedbacks to improve the results of the research and my family for the support they provided me through my entire life and in particular. I must acknowledge my husband and best friend, Tharindu, without whose love, encouragement and editing assistance, I would not have finished this thesis.

Abstract

Crime is one of the dangerous factors for any country. Although crimes could occur everywhere, it is common that criminals work on crime opportunities they face in most familiar areas for them. The ultimate goal of crime analysis is to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions. Criminals and victims follow common life patterns and most of the time overlaps in those patterns indicate an increased likelihood of crime. Geographic and temporal features influence the where and when of those patterns.

Our proposed solution consists of four major modules namely; Hotspot Analysis Module, Offender Profiling Module, Victim Profiling Module and Predicting Suspects Module. Data related to fast crimes which is used for the analysis is collected from Department of Police, Sri Lanka. Hotspot analysis module identifies crime hotspots considering geographical data of past crime where victim profiling and suspect profiling modules identify the patterns or groups of victims who are most vulnerable and suspects who share same characteristics. First three modules are developed based on simple k-means clustering algorithm where the fourth module is based on simple k-means clustering and j48 algorithm to generate the classifier model which can be used to predict the cluster of suspects of a crime.

The results of this analysis can be used by law enforcers to find general and specific crime trends, patterns, and series in an ongoing, timely manner in order to take advantage of the abundance of information existing in law enforcement agencies, the criminal justice system, and public domain, to maximize the use of limited law enforcement resources, to have an objective means to access crime problems locally, regionally, nationally within and between law enforcement agencies, to be proactive in detecting and preventing crime, to meet the law enforcement needs of a changing society and to understand the criminal behaviors.

Table of Content

Declaration.....	i
Acknowledgements.....	ii
Abstract.....	iii
Table of Content	iv
List of Figures	viii
List of Tables	ix
Chapter 1.....	1
Introduction to Crime Record Analysis in National Crime Records in Sri Lanka.....	1
1.1 Prolegomena	1
1.2 Background and Motivation	1
1.3 Crime and Criminology	3
1.4 Data Mining Approach	4
1.5 Crime Analysis.....	4
1.6 Goal of the Research.....	6
1.7 Objectives	7
1.8 Acknowledgments.....	8
1.9 Overview of the Report.....	8
1.10 Summary	8
Chapter 2.....	10
State of the Art of Exploring Current Findings in Crime Record Analysis	10
2.1 Introduction.....	10
2.2 Framework 1	10
2.3 Framework 2	10
2.4 Framework 3	11
2.5 Framework 4.....	11
2.6 Framework 05	12
2.7 Framework 6	12

2.8	Framework 7	12
2.9	Framework 8	13
2.10	Some Other Frameworks	13
2.11	Summary	14
Chapter 3	16
Technology adapted in Crime Record Analysis	16
3.1	Introduction.....	16
3.2	Methods related to predicting crimes.....	16
3.3	Methods to identify individuals at high risk of offending in the future.....	18
3.4	Methods used to identify likely suspects of a new crime.....	18
3.5	Methods used to identify crime victims.....	19
3.6	Summary	19
Chapter 4	20
A Novel Approach for Crime Record Analysis in Sri Lanka	20
4.1	Introduction.....	20
4.2	Proposed Model	20
4.3	System Overview	22
4.4	Data Collection	23
4.5	Data Integration	23
4.6	Data Analysis.....	24
4.7	Making Predictions about Potential Crimes.....	24
4.8	Predictive Analysis with Data Mining.....	25
4.9	Spatiotemporal Analysis	26
4.10	Users	26
4.11	Summary	26
Chapter 5	28
Analysis and Design of the Proposed Solution	28
5.1	Introduction.....	28

5.2	System Design	28
5.2.1	Hot spot analysis	29
5.2.2	Offender profiling	29
5.2.3	Victim profiling	31
5.2.4	Predicting suspects	31
5.3	Summary	32
Chapter 6.....		33
Implementation of the Solution		33
6.1	Introduction.....	33
6.2	Weka	33
6.3	Data Collection and Preprocessing	33
6.4	Hot Spot Analysis and Crime Mapping	35
6.5	Offender Profiling.....	39
6.6	Victim Profiling	41
6.7	Predicting Suspects	43
6.8	Summary	47
Chapter 7.....		48
Evaluation		48
7.1	Introduction.....	48
7.2	Evaluation of Hotspot Analysis Module.....	48
7.3	Evaluation of Suspect/Offender Profiling Module	49
7.4	Evaluation of Victim Profiling Module	51
7.5	Evaluation of Suspect Prediction Module.....	52
7.6	Summary of Evaluation	57
Chapter 8.....		58
Discussion.....		58
8.1	Introduction.....	58
8.2	Limitations	59

8.3	Further Developments.....	59
8.4	Summary.....	60
Chapter 9	61
Reference	61
Appendix A	63

List of Figures

Figure 2.1 Flow chart for Crime Analysis [11].....	15
Figure 4.1 Proposed Model.....	22
Figure 4.2 Summarized Model	23
Figure 5.1 System Design of the Proposed Solution	28
Figure 6.1 WEKA GUI.....	34
Figure 6.2 Sample of Preprocessed Dataset.....	35
Figure 6.3 Hot spot Analysis	36
Figure 6.4 Identified Hotspots	37
Figure 6.5 Hot Spot Analysis with k-Means Clustering	38
Figure 6.6 Identified Cluster Centroids of Hotspot Analysis Module	39
Figure 6.7 Sample Input to the Offender Profiling Module.....	40
Figure 6.8 Identified Cluster Centroids in Offender Profiling Module	41
Figure 6.9 Sample Dataset for the Victim Profiling Module	42
Figure 6.10 Identified Clusters and Their Centroids in the Victim Profiling Module	43
Figure 6.11 Summary of the new attribute 'cluster' in the dataset	44
Figure 6.12 Sample Dataset for the Suspect Prediction Module	45
Figure 6.13 Classification Model for the Suspect Prediction Module	46
Figure 6.14 Evaluation of the Model for the Suspect Prediction Module	47
Figure 7.1 Variation of Squared Errors within Clusters for Different number of clusters in Hotspot Analysis Module	49
Figure 7.2 Variation of Squared Errors within Clusters for Different number of clusters in Suspect Profiling Module.....	50
Figure 7.3 Variation of Squared Errors within Clusters for Different number of clusters in Victim Profiling Module.....	52
Figure 7.4 Variation of the Percentage of Correctly Classified Instances for Different number of suspect clusters in Suspect Prediction Module	53
Figure 7.5 Variation of the Kappa Statistic for Different number of suspect clusters in Suspect Prediction Module.....	54
Figure 7.6 Variation of the Mean Absolute Error for Different number of suspect clusters in Suspect Prediction Module	54
Figure 7.7 Variation of the Root Mean Squared Error for Different Number of Suspect Clusters in Suspect Prediction Module	55
Figure 7.8 Variation of the Relative Absolute Error for Different Number of Suspect Clusters in Suspect Prediction Module	55
Figure 7.9 Variation of the Root Relative Squared Errors for Different Number of Suspect Clusters in Suspect Prediction Module	56
Figure 7.10 Variation of the Percentage of Correctly Classified Instances with Different Confidence Factors for J48 Algorithm in Suspect Predicting Module	57

List of Tables

Table 5.1 The Meaning of Different Approach for the Crime	30
Table 5.2 The Meanings of Different Education Levels	30
Table 7. 1 Variation of Squared Errors within Clusters for Different number of clusters in Hotspot Analysis Module	48
Table 7.2 Variation of Squared Errors within Clusters for Different number of clusters in Suspect Profiling Module	50
Table 7.3 Variation of Squared Errors within Clusters for Different number of clusters in Victim Profiling Module	51
Table 7.4 Variation of percentage of correctly classified instances, Kappa statistic, mean absolute error, root mean squared error, relative absolute error and root relative squared error of the classifier model when different number of clusters for the suspect category is used in Suspect Prediction Module	53
Table 7.5 Variation of the Percentage of Correctly Classified Instances with Different Confidence Factors for J48 Algorithm in Suspect Predicting Module	56