



Universiteit
Leiden
The Netherlands

Algorithmic tools for data-oriented law enforcement

Cocx, T.K.

Citation

Cocx, T. K. (2009, December 2). *Algorithmic tools for data-oriented law enforcement*. Retrieved from <https://hdl.handle.net/1887/14450>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/14450>

Note: To cite this publication please use the final published version (if applicable).

Algorithmic Tools
for
Data-Oriented Law Enforcement

Tim Cocx



The work in this thesis has been carried out under the auspices of the research school IPA (Institute for Programming research and Algorithmics).



Netherlands Organisation for Scientific Research

This research is part of the DALE (Data Assistance for Law Enforcement) project as financed in the ToKeN program from the Netherlands Organization for Scientific Research (NWO) under grant number 634.000.430.

Cover design: Daphne Swiebel en Tim Cocx.

ISBN: 978-90-9024805-9

Algorithmic Tools
for
Data-Oriented Law Enforcement

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus prof. mr. P.F. van der Heijden,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 2 december 2009
klokke 11.15 uur

door

Tim Kristian Cocx
geboren te Amstelveen
in 1979

Promotiecommissie

Promotor:	prof. dr. J.N. Kok	
Co-promotor:	dr. W.A. Kusters	
Overige leden:	prof. dr. Th. Bäck	
	prof. dr. H. Blockeel	(Katholieke Universiteit Leuven)
	prof. dr. S. Haring	
	dr. M.-F. Moens	(Katholieke Universiteit Leuven)
	prof. dr. A.P.J.M. Siebes	(Universiteit Utrecht)

Contents

1	Introduction	1
1.1	Data Mining	2
1.2	Law Enforcement	4
1.2.1	Data on Criminal Activities	4
1.2.2	Related Work on Data Mining for Law Enforcement	5
1.2.3	Motivation	6
1.2.4	Overview	6
	Bibliography	8
I	Algorithmic Tools for Strategic Law Enforcement	11
2	Adapting and Visualizing Association Rule Mining Systems for Law Enforcement Purposes	13
2.1	Introduction	13
2.2	Background	14
2.3	Approach	15
2.3.1	Database Refit	15
2.3.2	Attribute Ban	16
2.3.3	Semantic Split	17
2.3.4	Detection	18
2.4	Trie Visualization	20
2.5	Experimental Results	21
2.6	Conclusion and Future Directions	22
3	Object-Centered Interactive Multi-Dimensional Scaling: Ask the Expert	25
3.1	Introduction	25
3.2	Expert Interaction Helps Heuristic Projection	27
3.3	Object-Centered Interactive MDS	28
3.4	Experimental Results	29
3.5	Conclusion and Future Directions	32

4	Data Mining Approaches to Criminal Career Analysis	33
4.1	Introduction	33
4.2	Background	34
4.3	Approach	34
4.4	Obtaining Profile Differences	36
4.5	A New Distance Measure	39
4.6	Visualization	42
4.7	Prediction	43
4.8	Experimental Results	43
4.9	Conclusion and Future Directions	44
5	Enhancing the Automated Analysis of Criminal Careers	47
5.1	Introduction	47
5.2	Background and Overview	48
5.3	A Distance Measure for Profiling	49
5.4	Alignment of Criminal Careers	50
5.5	Toward a New Visualization Method	52
5.6	Prediction of Unfolding Careers	53
5.7	Experimental Results	53
5.8	Conclusion and Future Directions	56
6	Detection of Common and Defining Subcareers	57
6.1	Introduction	57
6.2	Background	58
6.3	Approach	64
6.3.1	Changes to the Large Itemset Phase	64
6.3.2	Changes to the Transformation Phase	64
6.3.3	Finding Defining Subcareers	66
6.4	Experimental Results	68
6.5	Conclusion and Future Directions	71
	Bibliography for Part I	74
II	Algorithmic Tools for Tactical Law Enforcement	79
7	Temporal Extrapolation within a Static Visualization	81
7.1	Introduction	81
7.2	Background	82
7.2.1	Clustering	82
7.2.2	Extrapolation	83
7.3	Approach	87
7.3.1	Distance Matrix and Static Visualization	87
7.3.2	Distance Vector Time Frame and Sequence Clustering	88
7.3.3	Extrapolation	88

7.3.4	Final Steps	89
7.4	Experimental Results	90
7.5	Conclusion and Future Directions	90
8	An Early Warning System for the Prediction of Criminal Careers	93
8.1	Introduction	93
8.2	Background	94
8.3	Approach	95
8.3.1	Clustering Reduction and Extrapolation Selection	95
8.3.2	Further Steps	97
8.4	Experimental Results	99
8.4.1	Accuracy of a Single Prediction	104
8.5	Conclusion and Future Directions	107
9	A Distance Measure for Determining Similarity between Criminal Investi- gations	109
9.1	Introduction	109
9.2	Project Layout	110
9.3	System Architecture	110
9.4	Entity Extraction and Table Transformation	112
9.5	Multi-Dimensional Transformation	113
9.6	Distance Measure	115
9.7	Visualization	118
9.8	Experimental Results	119
9.9	Conclusion and Future Directions	121
10	Identifying Discriminating Age Groups for Online Predator Detection	123
10.1	Introduction	123
10.2	Background	124
10.2.1	Social Networking Sites	124
10.2.2	Online Identities	125
10.2.3	Online Sexual Predator	126
10.3	Predator Presence on Social Networking Sites	127
10.3.1	Amount of Under-Aged Friends: A First Glance	128
10.4	Approach	129
10.4.1	Genetic Algorithm	132
10.4.2	Greedy Group Selection	136
10.4.3	Thresholding	137
10.5	Experimental Results	138
10.6	Conclusion and Future Directions	140
A	Statistical Significance and Privacy Issues	143
A.1	Statistics	143
A.2	Privacy	145

B The HKS database	147
B.1 Structure	147
B.2 Ownership and Residence	148
B.3 Common Usage	149
B.4 Usage during Research	149
Bibliography for Part II	151
Acknowledgments	155
Nederlandse Samenvatting	157
Curriculum Vitae	161
Publication List	163