Artificial Intelligence and Law (2006) 14: 35–100
DOI 10.1007/s10506-006-9023-z

Decision support systems for police: Lessons from the application of data mining techniques to "soft" forensic evidence

GILES OATLEY[1], BRIAN EWART[2] and JOHN ZELEZNIKOW[3]
[1]*Centre for Adaptive Systems, University of Sunderland, Sunderland, UK*
[2]*Psychology Division, University of Sunderland, Sunderland, UK*
[3]*School of Information Systems, Victoria University, Melbourne, Australia*
(E-mail: giles.oatley@sunderland.ac.uk)

Abstract. The paper sets out the challenges facing the Police in respect of the detection and prevention of the volume crime of burglary. A discussion of data mining and decision support technologies that have the potential to address these issues is undertaken and illustrated with reference the authors' work with three Police Services. The focus is upon the use of "soft" forensic evidence which refers to modus operandi and the temporal and geographical features of the crime, rather than "hard" evidence such as DNA or fingerprint evidence. Three objectives underpin this paper. First, given the continuing expansion of forensic computing and its role in the emergent discipline of Crime Science, it is timely to present a review of existing methodologies and research. Second, it is important to extract some practical lessons concerning the application of computer science within this forensic domain. Finally, from the lessons to date, a set of conclusions will be advanced, including the need for multidisciplinary input to guide further developments in the design of such systems. The objectives are achieved by first considering the task performed by the intended systems users. The discussion proceeds by identifying the portions of these tasks for which automation would be both beneficial and feasible. The knowledge discovery from databases process is then described, starting with an examination of the data that police collect and the reasons for storing it. The discussion progresses to the development of crime matching and predictive knowledge which are operationalised in decision support software. The paper concludes by arguing that computer science technologies which can support criminal investigations are wide ranging and include geographical information systems displays, clustering and link analysis algorithms and the more complex use of data mining technology for profiling crimes or offenders and matching and predicting crimes. We also argue that knowledge from disciplines such as forensic psychology, criminology and statistics are essential to the efficient design of operationally valid systems.

Key words: data mining, decision support systems, matching, prediction

1. Introduction

1.1. SECTION OBJECTIVES

This section aims to describe some aspects of the forensic domain which illustrate the need for computer science methodologies and their ability to contribute to the investigative process. It also provides a brief summary of the succeeding sections comprising this paper.

1.2. THE INVESTIGATORS' TASK: DETECTING CRIME

In 2003/2004, approximately 5.9 million crimes were notified to 43 Police Services in England and Wales (Dodd et al. 2004). Given the examples used throughout this paper, it is also worth noting that property crimes such as theft and burglary account for 78% of recorded crimes and over 430,000 (14%) of these involve the burglary of a domestic dwelling.

Once notified, the Police must investigate. Sometimes this will constitute only the recording of the details of the offence from the victim. More often, the investigation will comprise a detailed examination of the crime scene, the collection of forensic material and recording of witness and victim statements. This level of process is not just reserved for serious crimes, but is routine practice even for relatively minor ones. It is apparent that regardless of the level of investigation, a single recorded crime will generate a considerable amount and diversity of information. The challenge is not only to store this information, but to use it to facilitate the investigative process. The features of one case at a particular point in time may have little value but as we shall see below, the ability to retrospectively interrogate such a comprehensive crime database is a powerful investigative tool.

The value of this technological challenge is best understood through considering the task of the investigator. In general terms, it is worth noting that the detection of crime in many Police Services will involve not just Police Officers, but civilian crime analysts and scene of crimes officers (SOCOs). In respect of crime detection, Canter (2000) describes the nature of forensic decision making. It is how the investigators draw inferences and conclusions from the data they have available in order to embark upon "the most appropriate way forward in order to identify and prosecute the culprit" (p. 24). However, there is an established literature on the problems of information processing and inference that may beset human decision making (see below). In conjunction with the point above about the volume of information, it should be evident that decision making with the forensic domain is problematic. This may be a contributory factor in the detection rate for a volume crime such as burglary is as low as 13% of recorded incidents (Dodd et al. 2004).

More specifically, Canter (2000) draws our attention to three key decision making challenges which confront the investigator. First, the amount of information available. When a crime happens it is recorded and a variety of information is available to the police officer. For instance, for burglary from a dwelling house (BDH), information (even for a single crime) will include forensic evidence (e.g. fingerprints, shoeprints), details of the victim, as well as when and where the crime occurred. If modus operandi data (e.g. property stolen, search strategy of the criminal) is included, it is clear that even for this single event, much information is available and it may vary in quality and validity.

Computer Science methodologies have the ability to select and display such information to the investigator. In this way, the "salience" Canter (2000) of crime features is revealed. These are aspects of the crime which have most potential to assist the investigation in a number of ways, including the identification of crime patterns, linking offences and the identification of suspects. Section 2 of this paper, which is summarized below, presents a discussion of role of computer science to this process.

Second, Canter (2000) describes the challenge of drawing conclusions from data and describes the inductive nature of the process. Here the investigator is again challenged by the quantity of information. Our cognitive capacity is limited and Newell and Simon (1972) characterise human inference as "bounded rationality". In order to cope with the complexity of decision making, a number of cognitive shortcuts or "heuristics" are evident in information processing. The problem is that these short cuts have the potential to introduce biases which may contaminate the process and ultimately the decision reached. For example, we exhibit a "confirmatory bias" (Oskamp 1965) in that we focus upon data which conforms to the initial ideas or hypotheses we have formed. In this way, we fail to *test* our ideas (e.g. about who is the suspect) by seeking evidence which disconfirms our notions. Instead, we seek information that substantiates these impressions.

We also find certain kinds of information easier than others to process. For example, case history information is frequently sought out and used as the foundation of our judgements (Kahneman and Tversky 1973) because we always believe it is important. So, aspects of a specific burglary will seem especially relevant if there is an apparently distinctive feature of the modus operandi. For example, if two burglaries at different addresses both involve entry from the roof space of an adjoining property, they may be readily attributed to the same offender. However, before this attribution is useful in linking the crimes to a suspect the overall frequency, or base rate, of this feature across all offenders and burglaries needs to be established. This is simply not possible for any Police Office to determine cognitively. Even when this base rate information is available, people find it difficult to incorporate it into their decision making.

However, Gigerenzer (1991) demonstrates that such biases and errors in processing and inference are not inevitable. Indeed, if information is presented appropriately, people incorporate a range of information types very successfully. The problem is that if it is not presented in a form which facilitates interpretation and "hypothesis" testing (e.g. about who may be the culprit) then the essence of the inductive process is compromised. Furthermore, one proactive strategy for detecting the criminal is to identify where they are next most likely to offend. Again, in conjunction with criminological concepts such as "near repeats" (Johnson and Bowers 2004) and psychological concepts such as "criminal range" (Barker 2000), visualising crime patterns allows this process to be executed. Section 3 describes computer science technologies which integrate disparate sets of information and enables individual case data to be set within a wider statistical, temporal or geographical context.

Finally, a fundamental part of crime investigation is the generation of suspects and this is known to be fraught with operational and ethical problems (Townsley and Pease 2002). The ability to interrogate a crime data base with the objective of linking crimes is an important prerequisite for suspect generation. As noted above, the importance of taking account base rate frequencies in specific behaviours is important. Section 4 presents examples of the successful application of computer science methods to identify and match suspects to unsolved burglaries.

1.3. THE POLICING TASK: PREVENTING CRIME

The statutory requirement under the Crime and Disorder Act (1998) for UK police and local partnerships to undertake crime and disorder audits and produce strategies based on these audits, has provided a powerful stimulus to the mapping and analysis of crime data. Ratcliffe and McCullagh (2001) make the point that the "recent shift within British policing towards a more decentralised, proactive style gives preventative policing a higher priority, and this puts the analytical focus onto analysts and intelligence officers at the police divisional level who are now expected to be the hub of the local intelligence gathering effort – for high volume crime, this has left an analytical void". Hirschfield's (2001) account of crime prevention targeting identified eight interest groups likely to want to scrutinise crime and disorder problems, ranging from the police through community safety co-ordinators, to residents and business managers. Different forms of crime data analysis and mapping are needed to meet the varying requirements of these groups, in fact the wide ranging needs of police officers alone has led to a proliferation of very varied crime analysis and general decision support software.

It is clear that the Police require software that assists them in:

- Targeting resources for prevention more efficiently,
- Providing an empirical foundation for inter-agency burglary reduction initiatives,
- Identifying important data to be collected at an event, leading to efficiency gains in personal time; and,
- Providing information on designing systems which incorporate hard (forensic) and soft (crime scene information) data, and police intelligence information.

While the detection of crime has been the focus of the arguments so far, it must be noted that the technologies and approaches under discussed also apply to crime prevention.

1.4. EVALUATION A PROVISO

The authors believe that because of the interest and pace of developments in forensic decision support applications, there is a need to review the existing state of the domain. Furthermore, the lessons learned have high ecological validity in that they are derived from "real world" work with Police Services. However, because of the speed of such developments, substantive evaluative research on the performance of systems is not yet available. Nonetheless, important components of the systems, such as the crime matching algorithms and the application of statistical methodologies to reveal psychological dimensions of offending have been tested and subjected to peer review. Furthermore, the psychological and criminological concepts employed in the design of the systems are empirically robust phenomena. To communicate the lessons learned sooner rather than later, we believe will benefit the future work.

1.5. SUMMARY OF SECTIONS

Section 2 of this paper, "Working with real world crime databases", presents examples of typical police data, and the types of problems likely to be encountered. The necessity for data preprocessing and data transformation is discussed and lessons for two projects are discussed.

Section 3 concerns "Visualizing crime data: From pins in maps to clustering and beyond" and describes a range of useful visualisation and data exploration techniques. The data mining technologies of association rules and classification rules are discussed in the context of exploratory techniques. Clustering, series analyses and association techniques are discussed.

Section 4 focuses upon “Data mining: From crime patterns to crime matching and profiling” and describes technologies that belong to statistics (including multi-dimensional scaling, binary logistic regression, empirical Bayes) and artificial intelligence (neural networks, case-based reasoning and feature selection, logic programming). The thrust is that the ability to link crimes is important to the Police in order to identify potential suspects. Pease (2001) asserts that “location is almost never a sufficient basis for, and seldom a necessary element in, prevention or detection”, and that non-spatial variables can, and should be, used to generate patterns of concentration. To date, little has been achieved in the ability of “soft” forensic evidence (e.g. the burglar’s modus operandi) to provide the basis of crime matching. However the authors’ work (Ewart et al. 2005), described in this section, investigates crime matching based upon the combinations of locational data with behavioural data. Additionally, a variety of data mining techniques are explored (classification and association rules, neural network clustering, survival analysis and Bayesian belief nets, case-based reasoning, ontologies and logic programming) to support police in detecting the perpetrators of burglary from dwelling houses, a volume crime with a low detection rate.

Section 5 entitled “Developing predictive systems: From predictive clustering to Bayesian models”, addresses the problems of prediction. It draws from statistics, artificial intelligence and the application of criminological and psychological concepts. The practical lesson from the West Midlands data is motivated by the observation that using officially reported burglaries, most houses are victimised only once and most repeat victims are “hit” twice only in a 12 month period. Defining high-risk properties by waiting for the second burglary has obvious limitations, and if police wait until a second victimisation, it appears from official statistics that it would be too late – the likelihood of a third burglary is small. The first issue described in this section is survival analysis (Ewart and Oatley 2003) used to explore if modus operandi distinguishes houses burgled once only from those suffering a revictimisation. This is important to the implementation strategy of both preventative and detection resources. The second method described is a Bayesian belief network (Oatley and Ewart 2002), which predicts the likelihood of burglary as a combination of varying kinds of forensic information.

Section 6 presents “Overall conclusions and Future Directions” and draws together and presents a discussion of the operational lessons relevant to future work based on the description of the main approaches and findings in this research project.

1.6. SECTION ONE CONCLUSION

Section 1 describes the conditions that challenge the criminal investigator and concludes that computer science has an important role to play. Therefore, three objectives underpin this paper. First, given the continuing expansion of forensic computing and its role in the emergent discipline of Crime Science, it is timely to present a review of existing methodologies and research. Second, it is important to extract some practical lessons concerning the application of computer science within this forensic domain. Finally, from the lessons to date, a set of conclusions will be advanced, including the need for multidisciplinary input to guide further developments in the design of such systems.

2. Working with Real World Crime Databases

2.1. SECTION OBJECTIVES

The section considers data from Cleveland Police and West Midlands Police (WMP). The WMP initial work was based around a project funded by the Home Office's Crime Reduction Programme. The primary objective was to assist the policing of the volume crime of Burglary from Dwelling Houses (BDH) using decision support systems, and involved both the University of Sunderland's Division of Psychology and its Centre for Adaptive Systems.

Examples of real world crime data sets illustrate the issues which must be addressed when working with real world crime databases, including the problems of diverse information sources and the lack of standards. The section begins with a description of knowledge discovery from databases, and the concepts of data decomposition, aggregation and transformation. These concepts are motivated with real world examples.

2.2. KNOWLEDGE DISCOVERY FROM DATABASES

Knowledge discovery from databases (KDD) is the non-trivial extraction of implicit, previously unknown and potentially useful information from data (Fayyad et al. 1996; Fayyad and Stolorz 1997). The KDD process begins with analysis of data stored in a database or data warehouse and ends with production of new knowledge. Fayyad et al. 1996) describe knowledge discovery as a process with five distinct stages: data selection, data pre-processing, data transformation, data mining and interpretation.

The first phase of any KDD process involves the selection of a sample of data from a database of records. Decisions must first be made regarding the nature of the problem of interest in order to assess its suitability for the KDD process. This phase is equivalent to sampling in statistical circles and involves selecting which records to include and which to omit. There are two distinct considerations; how to select records and how to select variables.

Data pre-processing involves preparing the sample data for further phases of the KDD process. This requires attention to two main factors; missing values and erroneous data.

Data may need to be transformed in order to discover useful knowledge. Transformation can involve changing the categories of values a variable may have. It can take one of three basic forms: (a) the decomposition of the data set into smaller parts where each part will be the subject of an independent data mining exercise, (b) the aggregation of variables and/or values to form a simpler, more general data set, (c) transforming values of variables in some way.

Following the acts of terrorism of September 11 2001, there has been an international focus upon the avoidance, detection and prosecution of acts of terrorism. Intelligence authorities have been very keen to embrace the use of data mining to help prevent acts of terrorism occurring.

The problem with data mining in this and other criminal domains is that information is often stored as free text and in unrelated data-sets. For example prior to September 11 2001, very few people would have taken much notice of students who wished to learn to fly airplanes but showed no interest in landing the planes. This issue only became important once it had become clear that the perpetrators of the acts of terrorism of September 11 2001 had no interest in landing planes. Similarly, US police authorities took some time to catch the Washington snipers (John Allen Muhammad and John Lee Malvo) in October 2002, because the necessary data was stored in diverse databases. The pair had lived in Washington State[1] and Lousiana, had changed their names and did not have a local place of residence.[2] Unfortunately, it was difficult to link the relevant databases.

Stranieri and Zeleznikow (2000) note that the lack of Knowledge Discovery from Database applications in Law is due to the fact that most legal cases are stored in free text rather than databases. To place legal cases in databases requires massive data preprocessing, data cleaning and data transformation efforts.

The need for data mining from heterogeneous data sets has been recognized by numerous intelligence agencies. For example, DARPA (The Defense Advanced Research Projects Agency of the United States Department of Defence) has funded the *Evidence Extraction and Link Discovery* (EELD) program. EELD is developing technologies and tools for the automated discovery, extraction and linking of sparse evidence contained in large amounts of classified and unclassified data sources. EELD is expected to provide detection capabilities to extract relevant data and relationships about people, organisations, and activities from message traffic and open source data. It will link items relating potential terrorist groups or scenarios, and learn patterns of different groups or scenarios to identify new organisations or emerging threats.

Police recording systems in the UK are not standardised, and although there are now conventions for address referencing (see for example, British Standard 7666) police forces do not regularly adhere to these standards. Most police data will be scattered over distributed information sources. To find relevant information, a police officer needs to know which data sources offer specific sets of data and how to access them, as well as understanding each individual source's query language and user interface. They must then manually integrate the retrieved data. There are several examples of important software for data aggregation, based on *pooling the various diverse information sources* available, that can be brought to bear on a problem. An example tool is *COPLINK* (Chen et al. 2003a, b, 2004), created in 1998 at the Artificial Intelligence Laboratory at University of Arizona at Tuscon. COPLINK is marketed as a "one-stop access to point for data to alleviate police officers information and cognitive overload". The COPLINK system consists of two major components: COPLINK CONNECT is designed to allow diverse police departments to share data seamlessly through an easy-to-use interface that integrates different data sources including legacy record management systems and COPLINK DETECT which uncovers various types of criminal associations that exist in police databases.

In a similar fashion *FLINTS* – "Forensic Led Intelligence System" (Leary 2001, 2002, 2003) integrates diverse data sources, including both "hard" (DNA, finger-prints, shoe-prints) and "soft" (behavioural) forensic data, and other software packages exist that also provide this fusion of the various sources of police data, for instance the popular *I2* (2004).

Several problems emerge when planning to interrogate a police recorded crime database, no matter how it is stored or aggregated. Primarily a crime database is designed to *store and track information* relating to a crime and its investigation. There may be multiple entries over time concerning the same offence, which is a significant contaminant in terms of revictimisation analyses. Certain locations may be used as "dumping sites" for records that the system is unable to geocode; large amounts of data may not have an *x* and *y* co-ordinate; or crime incidents could be referenced to the midpoint of streets when there is inadequate information to pin them to individual properties. Some entries may end up as "no crime" and need to be removed. There are the vagaries of input behaviour and practice which need to be discovered as they may be read by statistical packages as different cases. This is a common occurence, when the same address is recorded in slightly different ways. Pease (1998) and Ratcliffe and McCullagh (2001) provide other examples. Similarly, when considering burglary from dwelling houses, problems arise when incidents occur in houses under multiple occupation, flats and high-rise buildings.

This section illustrates these issues of decomposition, aggregation and transformation, based around data used by three UK police forces, and illustrates the lack of any adherence to a standard. The section also describes the requirements of cleaning and transformation of data for further exploration.

2.3. EXAMPLE DATA

The authors whilst working with Cleveland constabulary (Oatley et al. 2002) were supplied with 10,021 records, involving a 3-month time slice of data from a crimes database. A sample of the data can be seen in Table I (all presented data is anonymised by mixing field values between records), and a "repeat victim" can be seen, as the person reference number 10 (field: "REF") occurs three times. While the dataset contained the field "REP_-VIC", representing whether the person has been a victim of a repeat crime, this field was frequently inaccurately calculated. A new field was added so that if a person appeared more than once in the data set they were considered a repeat victim (938 records), else they were not considered a repeat victim (9083 records).

It should also be noted that the "ethnicity" for the person (reference number 10) was noted in only two of these records, with the value "999" indicating a missing value. The data had many other obvious inaccuracies.

The West Midlands Police burglary dataset contained over 70,000 records, which represented approximately three and a half years of data. The following tables contain details of the crime (Table II – 11,382 data points), stolen property (Table III – 59,216 data points), the victim (Table IV – 10,678 data points), and the offender (Table V – 1782 data points). The data figures stated are those at the time of development of the discussed models, January 2002.

2.4. DATA PREPROCESSING AND TRANSFORMATION

The West Midlands Police data was "manually" inspected by a police officer, and various algorithms produced to clean or correct the data. An example of cleaning the data is eliminating records containing certain keywords [DUPLICATE | DUPLICATE RECORD | RECORDED IN ERROR | CREATED IN ERROR | WRITE OFF | WRITTEN OFF | NO CRIME | NO CRIMED]. These records are not errors, as we need to recall that the purpose of the database is to assist police officers in the routine tracking and auditing of crimes.

Also, although police routinely compile crime figures, these "descriptive" statistics are used primarily to explore the frequency of crimes types and

Table I. Example crime data. The “OCCUPATION” field is free text and is a rich source of information. The “RASCODE” (Police beat area) and “WARD” (Local Authority ward boundaries) contain geographic information. The “HOCLASS” is the Home Office crime category, the code referring to a specific type of crime, for instance, “HOCODE’s” 28, 29, 30, 31, are all types of burglary. “REF” is a unique reference for each person, “DATE_ADD” is the date added to database, “AGE” is the age of the person when added to database, “ETHNIC” is ethnicity (1. White 2. Black 3. Asian 4. Other), “REP_VIC” (1. No, 2. Yes)

REF	DATE_ADD	AGE	ETHNIC	OCCUPATION	HOCLASS	RASCODE	REP_VIC	WARD	SEX
1	31-Aug-99	51	1	HOUSEWIFE	30/1	M201A01	2	COATHAM	2
2	19-Jul-99	30	1	PROPRIETOR	49/10	L301P01	2	WESTBOURNE	1
3	4-Jul-99	40	1	PLATER	44	L204A01	2	ORMESBY	1
4	19-Jul-99	38	1	999	126	L206A02	2	GRANGETOWN	2
5	17-Aug-99	68	1	RETIRED	40	M112A01	2	GRANGE	2
6	17-Jul-99	34	1	HOUSEWIFE	28/1	S203A01	2	KADER	1
10	11-Aug-99	39	999	PIZZA CHEF	48/1	L201P01	2	SOUTH BANK	2
10	9-Sep-99	39	1	PIZZA CHEF	48/1	L201P01	2	THORNTREE	2
10	10-Sep-99	39	1	PIZZA CHEF	56/2	M204P04	999	BECKFIELD	2
11	20-Jul-99	35	4	UNEMPLOYED	58/3	M106P07	2	AYRESOME	1

Table II. Crimes table. The CRIME_NUMBER is the code consistent through all tables referring to a particular crime. CRIME_REF is an alternate unique crimes code. For the purposes of the work presented in this paper, OFFENCE is always "BURGLARY DWELLING", represented by the HOMC and HOOC (Home Office) _CODE's as 28:3. MO_DESC is a fixed format field which describes the M.O. There are 14 M.O. variables in this field, for instance ENTRY_POINT, DEGREE_OF_SEARCH. MO_NOTES is a free text field which represents an account of, and various notes associated with, the crime. Presently this is only used to search for number-plates using simple key word matching. BEAT_NUMBER_NUMERIC is not used for mapping as it reflects an arbitrary geographic region, however the GRID_REF's are used [Ordnance Survey 12-number grid system]

Feature	Data type	Example
CRIME_REF	INTEGER	400950
CRIME_NUMBER	CHAR(20)	20E1/4441/98
DATE_COMMITED	DATE	29/08/1998
LOCN_STREET	CHAR(255)	ST PETERS CAMPUS
LOCN_NUMBER	CHAR(25)	231,
LOCN_DISTRICT	CHAR(50)	SUNDERLAND
LOCATION_DESC	CHAR(255)	DETACHED - DWELLING
OFFENCE	CHAR(100)	BURGLARY DWELLING
MO_DESC	CHAR(255)	BURGLARY ENTRY:FRONT:DOOR:NOT KNOWN:WOOD: SMASH: ALL: UNTIDY:NOT KNOWN: AS ENTRY: DOOR: NOT KNOWN:NOT KNOWN:NOT KNOWN
MO_NOTES	BLOB SUB_TYPE TEXT	OFFENDER SMASHED GLASS IN FRONT DOOR AND OPENED SAME GAINING ENTRY. SEARCHED ALL ROOMS DISTURBED BY IP WHO WAS ASLEEP IN UPSTAIRS BEDROOM STATING "ALRIGHT MATE" WHEN IP DISCOVERED IN BED. OFFENDER RAN FROM HOUSE AND GOT INTO WHITE COLOURED VEHICLE NO MAKE

Table II. Continued

Feature	Data type	Example
BEAT_NUMBER_NUMERIC	INTEGER	17
GRID_REF_NORTHING	INTEGER	808080
GRID_REF_EASTING	INTEGER	909090
HOMC_CODE	INTEGER	28
HOOC_CODE	INTEGER	3

Table III. Property table. CAT_DESC is the description of the stolen property category. This is a nominal variable with 32 values. The CRIME_NUMBER is the code consistent through all tables referring to a particular crime

Feature	Data type	Example
CRIME_NUMBER	CHAR(20)	20E1/4441/98
CAT_DESC	CHAR(50)	BUILDING MATERIAL

Table IV. Victims table. Age is at date of the current crime. Sex is male, female or unknown

Feature	Data type	Example
CRIME_NUMBER	CHAR(20)	20E1/4441/98
SEX	CHAR(1)	F
AGE	INTEGER	37

Table V. Offender table. The URN is the "Unique Reference Number" for this offender. While the age will change throughout an offenders "career", the URN will not

Feature	Data type	Example
CRIME_NUMBER	CHAR(20)	20E1/4441/98
SEX	CHAR(1)	M
AGE	INTEGER	14
URN	INTEGER	19

calculate detection rates. The data was never intended to be subjected to "inferential" statistical techniques that are necessary for more substantive exploratory analyses. For example, for each burglary the crime scene information is recorded by a police officer on a questionnaire and entered as text in the original crime database – see the "MO_DESC" field in Table VI. This is unsuitable for use by commonly used packages such as SPSS, and it was

Table VI. Modus operandi features. The MO_DESC feature from Table II contains the above information. These were later converted in Boolean features indicating the presence or absence of a feature, for instance "LOCATION_OF_ENTRY_1Wall" {true or false}, "LOCATION_OF _ENTRY_2AdjoiningProperty" {true or false}, and so on. These predictor variables were used in the logistic regression described in Section 5.2

Grouping	Variables
Location of entry	1. Wall, 2. Adjoining Property, 3. Below, 4. Front, 5. Rear, 6. Side, 7. Roof, 8. Window, 9. Door, 10. Above
Entry methods and behaviour	1. Smash, 2. Cut, 3. Cutting equipment, 4. Duplicate Key, 5. Drill, 6. Force, 7. Remove Glass, 8. Ram, 9. Insecure door/window, 10. Climbed
Type of dwelling	1. Old, 2. Terrace, 3. Maisonette, 4. Bungalow, 5. Semi-detached, 6. Town House, 7. Flat
Search behaviour	1. Untidy Search, 2. Downstairs Only, 3. Many Rooms, 4. Upstairs Only, 5. Tidy Search, 6. Search All Rooms
Location of exit	1. Wall, 2. Adjoining Property, 3. Below, 4. Front, 5. Rear, 6. Side, 7. Roof, 8. Window, 9. Door, 10. Exit Same as Entry
Alarm/phone	1. Cut Phone, 2. Tamper with Alarm, 3. Alarm Activated
Bogus official crime	1. Social Services, 2. Bogus Official (type unknown), 3. Council, 4. DSS, 5. Home Help, 6. Gardener, 7. Other, 8. Water, 9. Police, 10. Distraction Bogus Crime

thus necessary to transform the data so that all burglaries have all crime scene variables in binary format indicating the presence of the factor. Fifty-six attributes are used in the (logistic regression) analyses (reported in Section 5.2) and are grouped as follows: Location of Entry (10 variables), Entry Methods and Behaviour (10 variables), Type of Dwelling (7 variables), Search Behaviour (6 variables), Location of Exit (10 variables), Alarm/Phone (3 variables) and Bogus Official Crime (10 variables). Table VI presents the groups and the variables.

An important aggregation required was due to the data being stored over four database tables, not permitting a simple analysis over all the data. All data relating to a particular case is best represented on one record within the database and ideally in numerical format. Particular examples of algorithms used in this reported work requiring "flat-files" are Trajan (2004) kohonen network (see Section 4.2), SPSS's logistic regression (see Section 5.2), and HUGIN (2005) Expectation Maximisation (EM) algorithm (see Section 5.3).

The Figures 1, 2 provide an idea as to how the data was transformed with respect to a specific range of experiments based upon the concept of "repeat victimisation". Polvi et al. (1991) provide a justification for a premise being

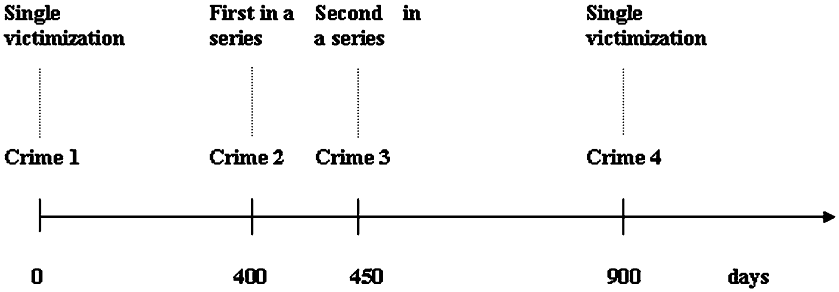


Figure 1. Time line for repeat victimization. A single premise is considered. Four crimes have occurred, base-lined at time 0 days for the first crime.

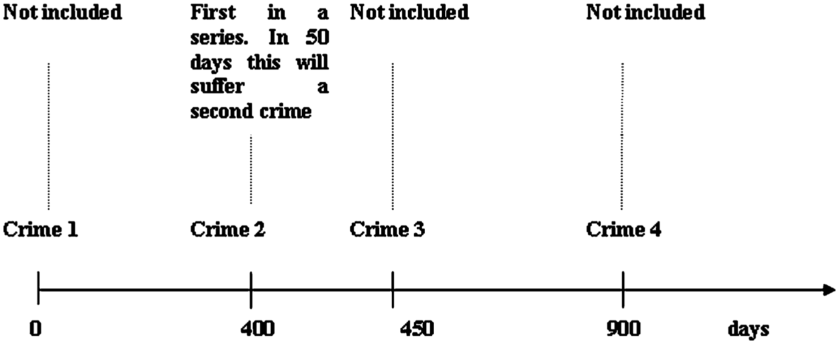
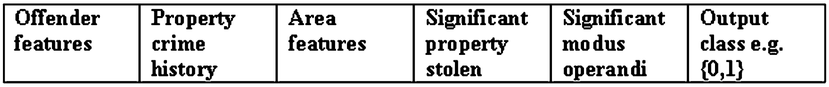


Figure 2. Time line for time till crime. The same data is used for a different experiment.

considered a repeat victim if it suffers a second victimisation within 365 days. Figure 1 presents data on a hypothetical single premise where crimes occur 400, 450 and 900 days after an initial crime. The first crime is *not* considered to be part of a series of repeat victimisation (the next crime is over 365 days later), however the second and third crimes are considered as part of a series of crimes. The dataset is then updated with an additional column representing this information, for instance "single victimisation" {0}, "first in a series" {1}, "second in a series" {2} and so on.

In Figure 2, the same data is used for a different purpose, determining how long it will be until the current premise suffers another victimisation. The results then constitute an output class. Depending on the particular algorithms used the data may also have to be significantly preprocessed. For instance, Figure 3 shows the additional features of "offender features", "property crime history", "area features", "significant property stolen", and "significant modus operandi", which are features used in the Bayesian network work (Section 5.3). Each one of these is a "calculated field" and did not exist in the original data. Some require significant calculation. For instance the "offender features" was computationally intense, including a kernel density estimate of each offender at the location of the premise in question. The specifics of these features are described fully in (Oatley and Ewart 2003).



Offender features	Property crime history	Area features	Significant property stolen	Significant modus operandi	Output class e.g. {0,1}

Figure 3. Calculated features in "algorithm" dataset. The flat-file format required by the HUGIN EM and other algorithms consists of many computationally intense features, each calculated for a particular premise, and then stored against the output class – in this case the "RepeatVictim?" network has a Boolean output, i.e. whether or not this premise suffers a further crime with 365 days.

In Figure 3, the flat-file format required by the HUGIN EM and other algorithms consists of many computationally intense features, each calculated for a particular premise, and then stored against the output class, taking either a Boolean value (i.e. whether or not this premise suffers a further crime with 365 days) or a "series" value (representing first in a series, second in a series etc.), or the time till crime value.

In a similar fashion the Cleveland crime data (see earlier Table I) was transformed into various "flat-file" formats for experimentation. From the repeat victims in the dataset, a new dataset was produced. If, for instance, a certain victim has three crimes recorded against them in the original dataset, then for the first two crimes the nature of the following crime will be known. This is included as an extra field (the output class), and the last record is discarded. Similarly the "next time" is known, and added in as an alternative output class. In this way the dataset was used in various neural network classifying experiments.

A further transformation was required with the Cleveland dataset, requiring significant preprocessing before presentation to the data mining algorithms. The presence of the free-text format "occupation" field was transformed using an amended version of the 1990 Standard Occupational Classification (SOC) (Elias 1995) into a very useful additional 11 class feature. Vague string matching techniques were used (Oatley et al. 2002) to compensate for misspellings, and a "look-up" table mapped the several hundred occupations to the derived feature "social codes".

2.5. SECTION TWO CONCLUSIONS

Computer science offers powerful tools to transform police databases into those amenable for the investigation of crimes and furthering crime prevention and detection. This section has detailed two such transformations that lay the ground for development of predictive systems.

However, police recording systems in the UK are not standardized, and the lack of a standard for data collected at a crime scene is a distinct problem for the development of generic techniques of data mining and decision support.

The PITO Corporate Data Model (PITO 2004) or Global Justice XML Data Model[3] crime scene descriptions are of no real use to the data miner, and inevitable adaptations are required. For instance the *Scene of Crime Information System* (SOCIS) project (Pastra et al. 2003) extracted from the PITO data model, semantic categories of interest to create their concept hierarchy.

Different analyses will require different data preprocessing (decomposition, aggregation and transformation), and this section has provided two examples of UK police force data sets and their preparation for predictive modeling. These experiences with crimes data are very common. For instance, Adderley and Musgrove (2001) document the significant recoding of textual data before presentation to their kohonen neural network. The transformations required for exploration and modeling the data will often be of differing complexity. Different police forces collect data of differing standards for the data mining task.

Once suitable algorithms for data mining from heterogeneous data sets have been developed, the efficiency of intelligence agencies and police in detecting possible crimes will be greatly enhanced. Data mining from crime databases can also be enhanced if police were made more aware of the benefits of collecting *more useful data*. They should be encouraged to ensure that all the important features of the data is recorded and that the data is clean. The authors were fortunate in that the West Midlands Police scenes of crime unit (SOCU) was keen to add extra variables of a more "subjective" nature into their reporting, and so the results by the authors have been used to inform improved data collection. For instance when describing crime scenes the importance of the data miner actually visiting the crime scenes and reviewing the data "picture" cannot be over emphasized. A simple addition of an "other" selection in a form category can prevent a rich crime scene being "squashed" into less appropriate fields, and whether significant property was visible but *not* stolen, are two simple but illustrative examples.

The education of data collectors as to the importance of capturing the data in assisting investigations is significant and culture dependent. In a proactive SOCU the intended users are actually collecting the data and they are able to see the benefit.

3. Visualizing Crime Data from Pins in Maps to Clustering and Beyond

3.1. SECTION OBJECTIVES

This section presents an overview of the visualisation tools and technologies used for the preliminary stages of data exploration. This is a wide ranging section because of the many uses the data will be put to. For instance whether the subsequent analysis has a crime perspective, an offender or victim perspective, we are always interested in the spread of criminal activity. Depending on the analysis certain technologies will be more or less appropriate – see for instance the review paper (Soomro et al. 2001), which provides a list of resources, including the categories of geographic/spatial analysis (23 entries), statistical analysis (12 entries), and crime analysis (14 entries).

Both this Section and the following Section 4 (Data Mining: From Crime Patterns to Crime Matching and Profiling) include technologies from data mining. The relevance to this section is in the sense of initial exploration of data through visualisation of complex relationships in the data, whereas Section 4's use of data mining is much more towards *deployment* of a useful technology.

According to Fayyad et al. (1996), data mining is a problem-solving methodology that finds a logical or mathematical description, eventually of a complex nature, of patterns and regularities in a set of data. Data mining techniques derive from several different sources, including artificial intelligence, inferential statistics and mathematical programming. This section includes means to match both single crimes and series of crimes, and the technologies include those from artificial intelligence and statistics, namely neural networks, case-based reasoning, logic programming, and Nave Bayes classification. Additional crime matching techniques include recent work comparing the use of spatio-temporal or behavioural or a combination of these features.

As Han and Kamber (2001) state, a data mining system has the potential to generate thousands of patterns or rules. Not all of the patterns are useful or interesting. Hence we need to define what is an interesting pattern and how can we generate all the interesting patterns and only the interesting patterns. A pattern is *interesting* if:

- The pattern is easily understood by humans;
- The pattern is valid (with some degree of certainty) on new or test data;
- The pattern is potentially useful; and,
- The pattern is novel.

A pattern is also interesting if it validates a hypothesis that the user wished to validate, or resemble a user's hunch. An interesting pattern represents knowledge. Several objective measures of pattern interestingness exist, based on the structure of discovered patterns and of the statistics underlying them. The concepts of support and confidence are examples of objective measures of pattern interestingness. In general, each interestingness measure is associated with a threshold, which may be controlled by the user.

Although objective measures help identify interesting patterns, they are insufficient unless combined with subjective measures that reflect the needs and interests of a particular measure. Subjective interestingness measures are based on user beliefs in the data. These measures find patterns interesting if they are unexpected (contradicting a user's belief) or offer strategic information on which the user can act.

It is often unrealistic and inefficient for data mining systems to generate all of the possible patterns. Instead, user-provided constraints and interestingness measures should be used to focus the search. Association rule mining (see later) is an example where the use of constraints and interestingness measures can ensure the completeness of mining.

According to Fayyad et al. (1997) KDD techniques, in general can be grouped into four categories:

- Classification – The aim of classification techniques is to group data into predefined categories. Examples include Bayesian classifiers, evolutionary computing, fuzzy logic, neural networks and rule induction.
- Clustering – The aim of clustering techniques is to group data into clusters of similar items. Research in data clustering comes from biology, machine learning, marketing, spatial databases and statistics. Clustering is an example of unsupervised learning. Unlike classification, clustering and unsupervised learning do not rely on predefined classes and class-labelled training examples.
- Series Analysis – A time-series database consists of a sequence of values or events changing with time. The values are typically measured at equal time intervals. Han and Kamber (2001) claim that there are four major components or movements that are used to characterise time-series data: (i) Long-term or trend movements, (ii) Cyclical movements or cyclic variations, (iii) Seasonal movements or seasonal variations, (iv) Irregular or random movements. By developing a systematic analysis of the movements of trend, cyclic, seasonal and irregular components, it is possible to make long-term or short-term predictions (forecasting the time series) with reasonable quality.
- Association – An association rule identifies a link between two or more attributes. An association rule is of the form $A1$ & $A2$ & ... & $Am \rightarrow P$ The association rule is interpreted as "database tuples that satisfy the conditions in the Ai are also likely to satisfy the conditions in P". Associated with each rule, is a confidence factor, that is how likely is the rule to be true and the support of the rule which states how many of the items in the data set are effected by this rule.

A variety of data mining techniques are described in this section, and later sections. These include classification and association rules, neural network clustering, survival analysis and Bayesian belief nets, case-based reasoning, ontologies and logic programming.

The technologies covered in this section are presented in the following order:

- Pie charts and histograms
- Geographical information systems, spatial data analysis and add-on algorithms
- Statistical and data mining algorithms
- Link and social network analysis
- Forensic psychology and criminology based approaches.

3.2. PIE CHARTS AND HISTOGRAMS

The use of pie charts as a visualisation aid can be demonstrated by our work with the Cleveland constabulary. Figure 4 presents some simple software which provide the ability to select certain criteria and see how the display changes. However even this display has a degree of sophistication, which clearly depends on the purpose in hand.

The project considered the issue of the repeat victimisation of individuals subject to varying crime types, and the various pie charts are ordered chronologically to give a summary of the kinds of crimes across multiple

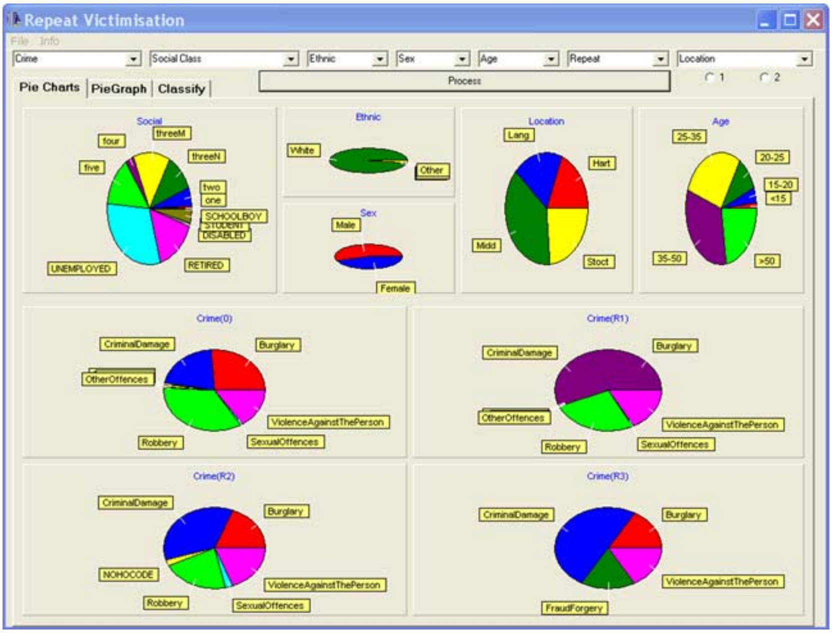
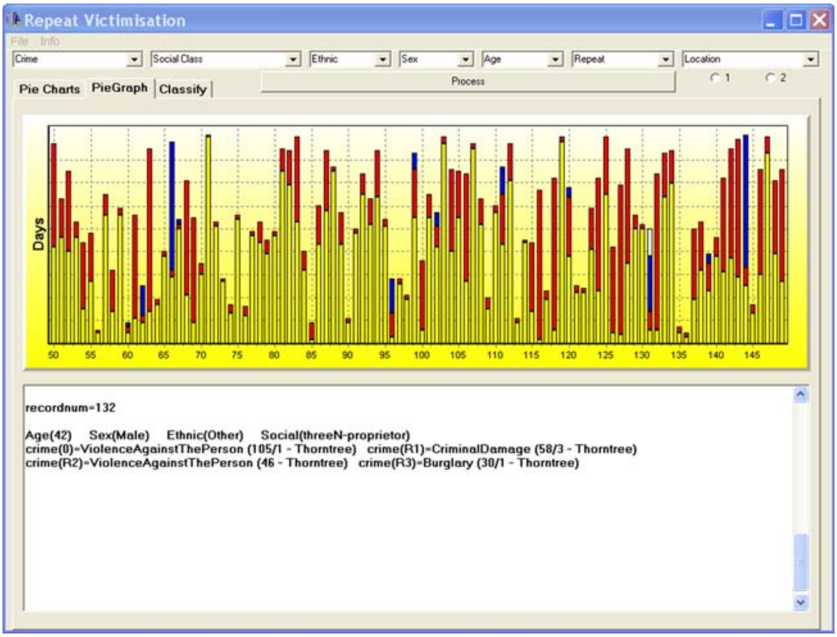


Figure 4. Pie chart presentation. Crime(0) is the first crime against a person, Crime(R1) is the first repeat crime against a person, Crime(R2) the second, and Crime(R3) the third repeat crimes against a person. The data is "unfiltered" and so shows the breakdown according to all features, e.g. the data points were mainly white, approximately half male, and the first repeat crime (CrimeR1) was evenly distributed between {Burglary | CriminalDamage | Robbery | ViolenceAgainstThePerson}.

victimisations. In this case the work is to study the victim, and so for instance the use of geographical information systems (GIS) would be less relevant. The data can be filtered according to the choices made in the drop-down boxes, including initial crime type, social class, ethnicity, sex, age, number of repeat victimisations, and location (the fields from Table I). These filters include all values for the attributes, for instance ‘‘ETHNIC’’ has the selections of {‘‘White’’, ‘‘Black’’, ‘‘Asian’’, ‘‘Other’’, ‘‘All’’}, and the social class is the 11-class categories mentioned earlier. For instance, in Figure 5, Crime(0) is the first crime against a person, Crime(R1) is the first repeat crime against a person, Crime(R2) the second, and Crime(R3) the third repeat crimes against a person. The data is ‘‘unfiltered’’ and so shows the breakdown according to all features, e.g. the data points were mainly white, approximately half male, and the first repeat crime (CrimeR1) was evenly distributed between {Burglary | CriminalDamage | Robbery | ViolenceAgainstThePerson}.

As stated, the Cleveland work took a person/victim centric view, and a more interesting visualisation can be seen in the use of the simple bar chart



recordnum=132

Age(42) Sex(Male) Ethnic(Other) Social(threeN-proprietor)
crime(0)=ViolenceAgainstThePerson (105/1 - Thorntree) crime(R1)=CriminalDamage (58/3 - Thorntree)
crime(R2)=ViolenceAgainstThePerson (46 - Thorntree) crime(R3)=Burglary (30/1 - Thorntree)

Figure 5. Bar chart presentation. Each bar represents an individual. The y-axis extends from 0 to 90 days. The start of the colored portions of the bar indicate the first crime, and successive borders between colored parts of the bar are further crimes. The individual (42-year-old male, ethnicity-not white, black or asian) that has been clicked on to examine is number 132, and is easily visible with a ‘‘white’’ section indicating a third repeat crime against them. The first crime was ‘‘ViolenceAgainstThePerson’’, Home Office code 105/1, in the Thorntree location, and was approximately 3 days into the study. They suffered a repeat crime 3 days later (CriminalDamage), another crime 20 days later (ViolenceAgainstThePerson), and a final crime 10 days later (Burglary).

feature shown in Figure 5, with a separate bar representing an individual. The *y*-axis is time, and ranges from 0 to 90 days. The lower shaded end of the bar indicates the point in this 90-day period when the first crime occurred against the person. The darker shaded part of the bar indicates when the next crime occurred, and so on. By clicking on any of these bars, the detail of the person is displayed in the text box below – for instance Figure 5 shows individual number 132 (42-year-old male, ethnicity – "other"), easily visible with a "white" section indicating a third repeat crime against him. The first crime was "ViolenceAgainstThePerson", Home Office code 105/1, in the Thorntree location, and was approximately 3 days into the study. He suffered a repeat crime 3 days later (CrimeR1 = CriminalDamage), another crime 20 days later (CrimeR2 = ViolenceAgainstThePerson), and a final crime 10 days later (CrimeR3 = Burglary).

The areas where each of these crimes occurred are also detailed. This kind of display can provide valuable information to the analyst, and while the earlier pie chart feature shows exactly the same data presented in a different format, both of these visualisations have a useful purpose.

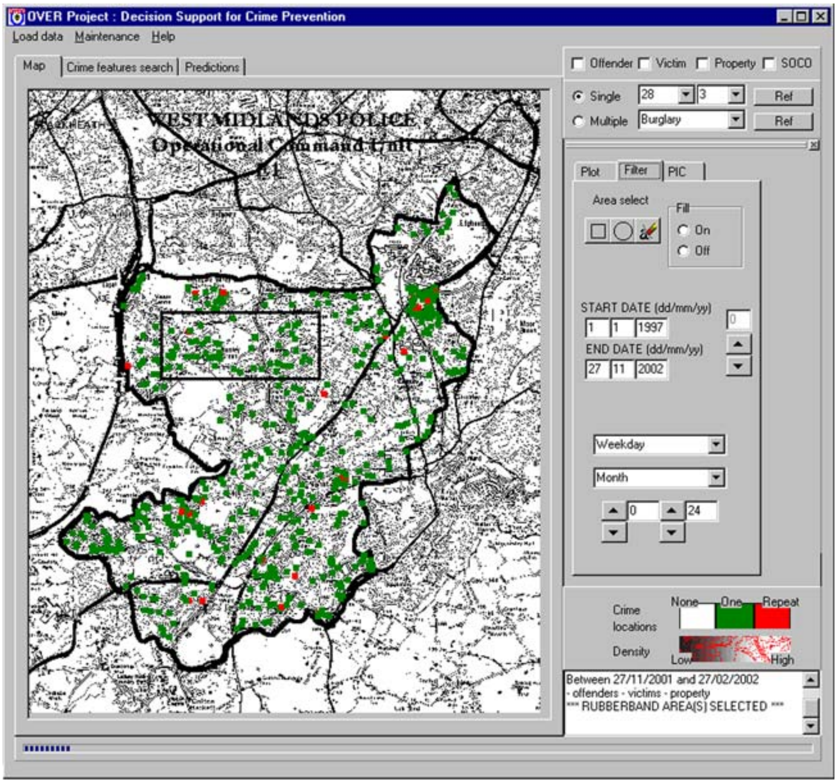
3.3. GEOGRAPHICAL INFORMATION SYSTEMS (GIS), SPATIAL DATA ANALYSIS AND ADD-ON ALGORITHMS

Crime mapping involves the manipulation and processing of spatially referenced crime data in order for it to be displayed visually in an output that is informative to the particular user. The use of GIS has been long established in police work, with many sophisticated GIS's on the market. For instance, a repository of crime analysis and mapping resources have been developed by *Community Oriented Policing Services* (COPS 2004), created as a result of the *Violent Crime Control and Law Enforcement Act* of 1994. The repository contains an introduction to crime mapping, and a review of a wide range of mapping software and geographic information systems, focusing on their functionality and ease of use by members of police departments.

The popularity of GIS is partly that they are easy to use, and that they provide an immediate visualisation of the data. However, its use is not without problems, for instance time is not easily represented, and "hotspots" can be misleading. This section will consider a variety of GIS approaches, and move onto metrics from spatial statistics and other algorithms that can be added-on to the basic GIS functionality.

3.3.1. *Crime mapping (pins in maps)*

Systems such as *MapInfo* and *Arcview* have robust and user-friendly graphical user interfaces. Crime initiatives may well require more than just this kind of data visualisation, however Bowers et al. (2001) implement a burglary-dwelling-house (BDH) initiative solely using GIS. Crime mapping for the visualisation and graphical exploration and presentation of data was an important part of the analysis of the West Midlands data. The software package that was developed (Oatley and Ewart 2003) can filter and display data on three levels of maps of increasing detail – the top level map is of the whole Operational Command Unit (see Figure 6), the second level maps represent quadrants (see Figure 7), in more detail, and, level three maps show street level detail. This is the classic presentation of crime data, as "pins in maps". It is then possible to select points of interest, investigating the details of the crimes. The benefits of writing bespoke map-plotting software over the use of a mainstream GIS are obviously programmability, although there are GIS products which provide a programmable interface that can be "embedded" in bespoke developed software.



□ Offender □ Victim □ Property □ SOCO

● Single 28 3

○ Multiple Burglary

○ On
○ Off

START DATE (dd/mm/yy) 1 1 1997

END DATE (dd/mm/yy) 27 11 2002

0

Weekday

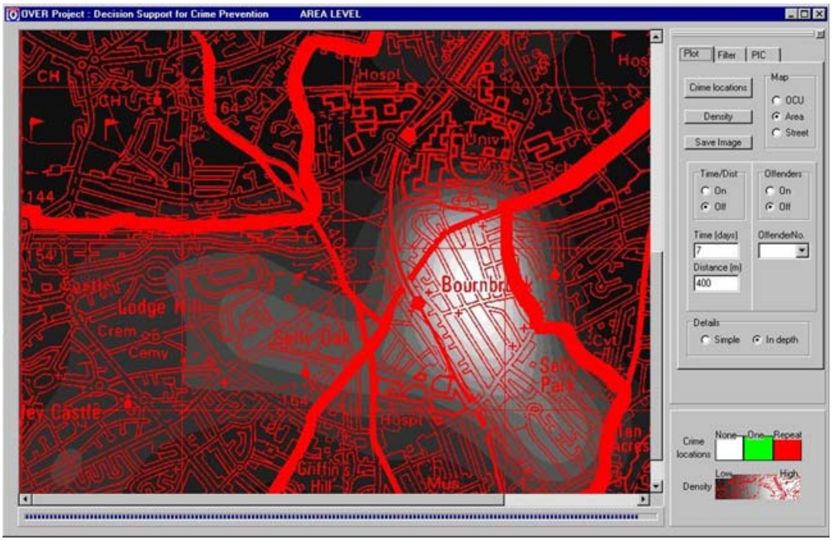
Month

0 24

Crime locations: None ⬜ One 🟩 Repeat 🟥

Between 27/11/2001 and 27/02/2002
- offenders - victims - property
*** RUBBERBAND AREA(S) SELECTED ***

Figure 6. Burglary data "GIS" plot. Data has been selected and loaded. The map is of the complete Operational Command Unit (OCU). The "Plot and Filter" form (docked) shows the main plotting functions, and a "rubber-band" area has been selected for investigation. The plotted squares either indicate single victimisations or repeat victimisations, as shown in the legend.



OVER Project : Decision Support for Crime Prevention AREA LEVEL

Map: ○ OCU ● Area ○ Street

Time/Dist: ○ On ● Off

Offenders: ○ On ● Off

Time (days): 7

Distance (m): 400

OffenderNo.:

Details: ○ Simple ● In depth

Crime locations: None □ One 🟩 Repeat 🟥

Density: Low High

Figure 7. Kernel based presentation. Any variable that is measured, such as the density of crime incidents, will be continuous over an area, being higher in some parts and lower in others – this gradient approach can provide a simple first view of data before exploring the explicit locations of the crimes.

3.3.2. *GIS and "add-on" algorithms*

There are many spatial data, or alternative statistical add-ons to GIS, different visualisation techniques, and Dykes and Mountain (2003) provide a survey of these geographic data plotting types and components. For instance *chloropeth maps* are shaded maps with colour varying according to the scale of the problem, *standard deviation ellipses* can be used to display "hotspots" created using clustering algorithms and *offence-offender residential location distance analysis* is self-explanatory. Williamson et al. (2001) provide a readable introduction to additional techniques of *block aggregation*, *kernel density estimation* (smoothing that result in crime "contour" maps), *Voronoi polygons diagrams* (maps indicating distances between offences) and *animation* (changes in crime distribution over time). Exploratory statistical techniques were implemented with the West Midlands data, including the ability to "toggle" between the explicit locations of crimes (see Figure 6) and a Gaussian (bell-shaped) *kernel-based approximation* for the probability density function – see (Speckt 1990; Bishop 1995; Patterson 1998; Hunter 2000) and Figure 7. In Figure 7, any variable that is measured, such as the density of crime incidents, will be continuous over an area, being higher in some parts and lower in others – this gradient approach can provide a simple first view of data before exploring the explicit locations of the crimes. The kernel had the added benefit that it can be weighted with the presence or absence of certain crime features, for instance a modus operandi or a property stolen feature, to bias the clustering in a more meaningful way. In this way the "hotspots" can be examined for dependency upon the modus operandi features, for instance the different kinds of property stolen in the burglary, or the search strategy of the offender.

Other examples of add-on technologies to GIS include Amsterdam police departments' analysis of criminal records using *Sentient's DataDetective* (Sentient 2004), where they combined the MapInfo GIS with decision trees and back propagation neural networks. These techniques are used to obtain simple profiles of perpetrators, to find organisational patterns, and predict criminal behaviour.

3.3.3. *GIS and "add-on" spatial statistics*

Among the more statistically oriented software, not specifically aimed at police officers, are the following, which are largely based around spatial statistics. *Spatial and Temporal Analysis of Crime* (STAC) developed by the Illinois Criminal Justice Information Authority, was one of the earliest crime mapping packages. Since STAC, there have been many other crime mapping packages that have started to be more widely used, including *CompStat*, *GAM* (Geographical Analysis Machines)[4] adaptations of MapInfo's *Vertical Mapper* facility, and *CrimeStat*[5] (Levine 2002). CrimeStat includes many spatial data statistical algorithms, and while the results (such as resulting hotspots) can be imported into many popular geographic information systems, it does not readily lend itself to use by police officers, and is aimed more towards the crimes analyst.

Many of the programs ask the user to lay down some a priori conditions concerning what constitutes a hotspot. For example, a cluster of crimes might be only defined as "hot" if there are at least 10 crimes within it. This means the user has to have certain knowledge to produce such maps.

The powerful Bayesian software *WinBUGS* (2004) is able to import spatial data (maps) in three common GIS formats (ArcInfo, Splus and EpiMap). This software is directed not even at the crimes analyst but at experts in the use of Bayesian statistics.

It should be noted that while some of these products are "badged" as crime data specific, they are more generally applicable to all types of spatial data, and do not incorporate any specific crime domain knowledge or theories.

Complete spatial randomness. Our further work with the West Midlands Police data has investigated the use of simple spatial statistical algorithms, including the check for *complete–spatial–randomness* (CSR) using *quadrat* methods. As the name suggests, CSR determines whether a set of spatial data is completely random and is the standard model against which a spatial point pattern is compared. The reasons for beginning an analysis with a test for CSR are that the rejection of CSR is a prerequisite for any serious attempt to model an observed pattern. CSR acts as a dividing hypothesis between regular and clustered patterns. Our experiments across offender data resulted in a CSR value ranging between 1.7 and 2.2 for all reasonable set quadrat sizes, indicating a high degree of clustering. In comparison the *k-function* on the same data indicates clustering at a low level but not at higher levels.

Other spatial algorithms currently included are *mean centre* and *centre of minimum distance* which although very simple can still provide useful information to the user. CSR and additional spatial routines can be seen as a complement to the kernel-based exploration of hotspots and the interpretation with the algorithm described in Section 3.3.2.

Spatio-temporal analysis of crimes. Of note in this subsection is the importance of the temporal representation for which (Ratcliffe and McCullagh 2001; Ratcliffe 2002) provide an interesting approach. The ability to use temporal data to plot changes over time is rarely used. Time stamping crime is difficult as the time of occurrence of a crime event, for instance burglary and motor vehicle theft, may not be known exactly. Therefore, when describing crime events, they must be viewed as singularities of variable length along the time line. This is represented by the following relationships in Ratcliffe's "aoristic" approach for two events X and Y as [XbeforeY | XequalsY | XmeetsY | XstartsY | XendsY | XoverlapsY | XduringY].

A number of different conceptual frameworks for temporal GIS have been suggested (Raafat et al. 1994; Peuquet and Duan 1995). However, a defining standard has yet to be agreed upon. One of the more comprehensive descriptions of the variety of possible temporal combinations comes from (Peuquet and Wentz 1994) and is applicable if the passage of time can be viewed as a fixed line and events are somehow fixed to this line.

Work in the Artificial Intelligence community includes Aamodt's (1994) implementation of Allens' (1991) temporal system in the long-established Creek semantic network-based case-based reasoning system (Aamodt 1990, 1994, 1995). The author demonstrates that the interval-based approach is suitable for qualitative temporal reasoning, especially in their model-based reasoning system component, and that the representation is similar to the way that a human expert reasons in domains where qualitative changes of parameters over time are important. Case-based reasoning (CBR), as the name suggests, is reasoning with cases – instead of rules, where a case is a previously solved problem – this methodology is mentioned again in the later Section 4.4.

3.4. STATISTICAL AND DATA MINING ALGORITHMS

There are many statistical and data mining algorithms that can be applied to crime data investigation. This section starts with a review of some statistical technologies, and then proceeds with the commonly used data mining technologies of association rules and classification rules from decision trees. While association rules are exclusively exploratory, the latter two are also used in building classification models. The association rules experiments were carried out using the WEKA data mining toolkit (Witten and Frank 1999), and the remaining experiments used SPSS's CLEMENTINE data mining toolkit (Clementine 2004).

3.4.1. *Statistical techniques*

Previous statistical work related to police decision support systems includes various *clustering technologies* (including both those from the statistical and artificial intelligence literature) that can be usefully employed to explore the data, and cluster analysis and factor analysis (Everitt 1974; Everitt and Dunn 1991). They are methods of extracting clusters and factors (dimensions) from the data, and can provide valuable insight into the data that is available. These methods lend themselves to defining "hotspots". However where a line is drawn in order to define a "hotspot" is somewhat arbitrary. Instead of a boundary there is a *gradient* around these incidents, and an imaginary line has to be drawn to indicate the location at which the "hotspot" starts.

Multidimensional scaling (MDS) (Coxon 1982; Davies and Coxon 1982) can be considered to be an alternative to factor analysis and is a means of discovering meaningful underlying dimensions to explain observed similarities between the investigated objects. Green et al., (1976) are probably the first to use this method on criminal behaviour. A non-metric multidimensional scaling program was applied to simulated and actual modus operandi information on a sample of burglaries. In this instance, the technique is used for crime matching where the aim is to identify distinct clusters of related burglaries. The rationale is that each cluster is likely to have been committed by a particular individual or gang. New cases can be compared to this profile, thus assisting the generation of suspects. Merry and Harsent (2000) used smallest space analysis (SSA), a derivative of MDS, to examine burglars' modus operandi. The technique reveals distinct facets or clusters of actions which the authors interpret as representing underlying psychological dimensions of their criminal action. The authors present a typology of burglars using these dimensions.

However, when it comes to matching a "new" crime against prior offences in a database, SSA must be used with caution and has limitations. Limitations may be illustrated with reference to Merry (2000), who uses the SSA of a sample of burglaries as a template against which to evaluate a “new” crime. This template is based on the “Radex” hypothesis (Canter and Alison 2000) and generates two important features of burglars’ modus operandi. First, it illustrates spatially the relationship of actions, thus allowing behavioural themes to be identified. Second, the frequency of behaviour is represented in that the further they are from the center of the radex, the more infrequent they are across the sample of burglaries.

The problem for matching a new crime against this template is that it is acknowledged that while the presence of an action might confirm a link between crimes, the absence of a specific action does not prove that a link does not exist. Also, the features of a newly reported crime have no representation in dimensional space. Any linking with another “similar” crime is therefore done in the absence of spatial information, a critical feature of the Radex hypothesis. Nonetheless, SSA has shown its potential to reveal important commonalities in respect of offenders’ actions (Merry and Harsent 2000). It can be applied to offence groups as diverse as burglary, arson, armed robbery and fraud (Canter and Alison 2000). More recently, MDS has been used to investigate how burglars’ motivations and their psychological relationships with the property and the victim differs for houses which are repeatedly victimized in contrast to those which are not (Ewart and Oatley 2005). The “boost” hypothesis of revictimisation (Pease 1998) suggests there is something about the initial burglary which encourages the burglary to return. Exploring the “narrative” of burglary using MDS provides a more substantive psychological explanation of the motivations which may underpin this criminological concept.

However, in respect of offender profiling, Farrington and Lambert (2000) advocate a statistical approach where factor analysis plays an important role. They apply this technique to a range of crime data including demographic, modus operandi and temporal and spatial information. Their approach allows the extraction of important dimensions which may be contrasted for different crimes. Furthermore, they are able to test the extent to which the crime features are predicted by offender characteristics.

The studies above illustrate the range of tasks for which relatively familiar statistical approaches may be applied. In addition to their ability to reveal important features of crimes and criminals, they may be used to test their own crime matching and profiling performance, an essential but often absent feature of crime decision support systems.

3.4.2. *Association rules*

Link analysis, alternatively referred to as *affinity analysis* or *association*, refers to the data mining task of uncovering relationships among data. The best example of this type of application is to determine association rules

(Aggarwal et al. 1993; Aggarwal and Srikant 1994, 1995; Ahmed et al. 2003). Association rules are used to show the relationships between data items. These uncovered relationships are not inherent in the data, as is the case with functional dependencies, and they do not represent any sort of causality or correlation. Instead, association rules detect common usage of items.

The *Tertius* extension to the WEKA association rules by Deltour (2001) was used in our project. This was preferred for memory reasons over the *Apriori* algorithm, mainly because of the large number of Boolean features (over 130). Even when we used Tertius, our experiments were restricted to rules with only two literals.

Because so many different association rules can be derived from even a tiny dataset, interest is restricted to those that apply to a reasonably large number of instances and have a reasonably high accuracy on the instances they apply to. The *coverage* of an association rule is defined by how many of the items in the data set are effected by this rule divided by the data set and is often called its *support*. And its *accuracy* – often called *confidence* – is the number of instances that it predicts correctly, expressed as a proportion of all instances to which they apply.

Tertius presents the slightly different measures of *confirmation* value – the number of correctly classified instances divided by the total number of instances with these attribute values – and *frequency of counter instances* – the number of counter instances divided by the total number of data items. A rule is said to be better than another if it has a higher confirmation.

Table VII shows some example association rules from the West Midlands Police data. Rule number one reads as: *"If entry was not via the window then entry was via the roof or exit was not via the window"*. It can be seen that none of these rules have a very good confirmation, indicating the lack of simple relationships in this dataset.

3.4.3. *Decision trees (C5.0) classification rules*

Decision trees are a classification technique that uses nodes which involve testing a particular attribute. Usually the test at a node compares an attribute value with a constant. However, some trees compare two attributes with each other, or utilize some function of one or more attributes. Leaf nodes give a classification that applies to all instances that reach the leaf, or a set of classifications, or a probability distribution over all possible classifications. To classify an unknown instance, it is routed down the tree according to the values of the attributes tested in successive nodes, and when a leaf is reached the instance is classified according to the class assigned to the leaf.

It is easy to read a set of rules directly off a decision tree – one rule is generated for each leaf. The antecedent of the rule includes a condition for

Table VII. Selection of association rules. Cv – Confirmation value. Fci – Frequency of counter instances

	Rule	Cv	Fci
1	2via{_}WINDOW = 0 ==> 2via{_}ROOF = 1 or 10via{_}WINDOW = 0	0.541134	0.018750
2	[AUDIO, RADIO] = 0 and [PHOTOGRAPHIC] = 0 ==> [AUTO ACCESSORIES, FUEL] = 1 or [T/V, VIDEO] = 0	0.270994	0.162500
3	[ANTIQUES, PAINTINGS, CHINA, SILVERWARE] = 0 and [T/V, VIDEO] = 1 ==> [AUDIO, RADIO] = 1 or [PHOTOGRAPHIC] = 1	0.268692	0.162500
4	[CLOTHING] = 0 and [T/V, VIDEO] = 0 ==> [AUDIO, RADIO] = 0	0.283327	0.341772
5	[DOCUMENTS] = 0 and [PURSE, WALLET] = 0 ==> [CARDS, CHEQUES] = 0	0.246988	0.347826

every node on the path from the root to that leaf, and the consequent of the rule is the class assigned by the leaf.

The target class for the data set was determined simply on the basis of the proportion of offender crimes. Table VIII shows the number of crimes of each offender, and three roughly equivalent sized ranges were constructed: *class1_5* contains the offenders who have committed 1–5 offences (438 data points), *class6_15* have committed 6–15 offences (302 data points), and *class16_38* have committed 16–38 offences (380 data points).

Table VIII. Parameters retained after sensitivity analysis. VICSEX is the victim sex, ENTRY is ENTRY METHODS AND BEHAVIOUR, LOCATION is LOCATION OF ENTRY, SEARCH is SEARCH BEHAVIOUR, EXIT is LOCATION OF EXIT. The degree of sensitivity is the error rate resulting when each input parameter is missing. These errors can be compared with the baseline error. Parameters with the largest degree of sensitivity are ranked highest

	VICSEX	ENTRY	LOCATION	SEARCH	EXIT
Training					
Parameter ranking	2	3	1	5	4
Baseline error	0.6795877	0.5832178	0.7358213	0.5251378	0.5379434
Degree of sensitivity	1.772832	1.521433	1.919528	1.369921	1.403326
Verification					
Parameter ranking	2	3	1	4	5
Baseline error	0.7427634	0.644952	0.7825864	0.6121276	0.6034769
Degree of sensitivity	1.392062	1.208747	1.466697	1.147229	1.131016

In Figure 8, the average number of crimes per offender is approximately 15, with most offenders committing less than 5 offences. The total number of crimes is 1120. The classification rules from this dataset were produced using Quinlan's (1998) C5 algorithm, and numbered: 8 rules for *class1_5*, 3 rules for *class6_15*, and 8 rules for *class16_38*. An example rule can be seen, as the complete rule set (21 rules) is too numerous to present, and the decision tree was extremely hard to understand – a problem when using this for exploratory purposes. The numbers presented after the output class are coverage and accuracy, respectively.

The coverage of the rule is 27, i.e. this rule covers 27 data items, of which it correctly classifies 0.741 (accuracy). STOLEN13 = [DOMESTIC APPLIANCE] and STOLEN29 = [TV, VIDEO].

Dimensionality reduction. In order to construct a reasonably sized decision tree, or set of classification rules, various pruning methods can be used, or alternatively some form of dimensionality reduction can be used beforehand. The dataset was subjected to *Principal Component Analysis*, resulting in five components being extracted (from the input feature set numbering over 130). An example of the kind of representation for a component, or factor, can be seen in the shortened Equation 2. The first 32 features are types of stolen property and the following 105 features are modus operandi.

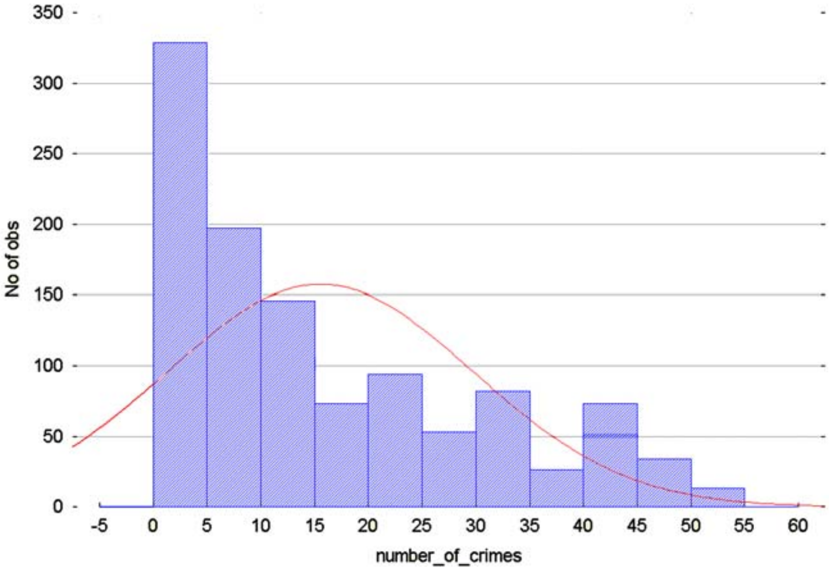


Figure 8. Number of crimes of each offender. The average number of crimes per offender is approximately 15, with most offenders committing less than five offences. The total number of crimes is 1120.

$$\begin{aligned}\text{Factor-1} =\, & 0.02286 \times \text{STOLEN1}\\ & +0.3363 \times \text{STOLEN2}\\ & +\cdots\\ & +0.07108 \times \text{STOLEN32}\\ & +0.01213 \times \text{1entry\_ABOVE}\\ & -0.05163 \times \text{1entry\_ADJOIN}\\ & +\cdots\\ & -0.3129 \times \text{12bogus\_OFFICIAL}\\ & -0.0713\end{aligned}$$

Decision trees (C5) – classification rules with dimensionality reduction. Using the previously determined components (factors: $F–1,$F–2,$F–3,$F–4, $F–5), a decision tree was induced and a set of five classification rules is derived. This is a much more manageable set of rules, however the draw back is obvious, as the factors are extremely opaque. While a decision tree or derived classification rules potentially offer high transparency, it is unfortunate that in high dimensional space the necessary preprocessing obscures these benefits.

It can be seen that the applied statistical and data mining approaches have several drawbacks when applied to this data. The following section describes technologies that are more commonly used in criminal investigations.

3.5. LINK AND SOCIAL NETWORK ANALYSIS

In law enforcement, intelligence analysts often refer to nodes and links in a criminal network as entities and relationships. The entities may be criminals, organisations, vehicles, weapons or bank accounts. The relationships between the entities specify how these entities are associated together. Law enforcement agencies and intelligence analysts frequently face the problems of identifying possible relationships among a specific group of entities in a criminal network. However, such tasks can be fairly time-consuming and labour-intensive without the help of tools to automate or assist with this process.

Since the creation of the United States Homeland Defence Agency, as a consequence of the Acts of Terrorism on September 11 2001, there has been a rapid increase in commercially available products that claim to perform link analysis. This term is meant in a more general sense than the previously mentioned association rules. Work in this area includes:

- The creation of links between transactional database records of individuals who have related financial transactions in order to identify money laundering networks (Goldberg and Wong 1998);
- The Link Discovery Tool which uses shortest path algorithms to discover association paths between two individuals that on the surface appear to be unrelated (Horn et al. 1997) and,
- The COPLINK system which is based on finding links between database elements of person's organisations, vehicles and locations (Hauck et al. 2002).

COPLINK contains additional analysis facilities. "COPLINK Detect" uses a technique called *concept space* (Chen and Lynch 1992; Hauck et al. 2002) to identify such associations from existing crime data automatically. A concept space is a network of terms and weighted associations within an underlying information space. While the basis of this algorithm is essentially TFIDF (Salton and Buckley 1988) with a weighting factor, it goes further by using hierarchical clustering to detect subgroups in the criminal networks, block modelling to reveal patterns of between-group interaction, and a further extension that identifies central members in a given subgroup. This is a good example of a useful add-on to assist analysis of such a large resource of data, as with many of the GIS and data integration systems. A drawback mentioned by the authors however was the effort required for the initial determination of concepts from the free-text case reports available from the Tuscon Police Department.

In a similar fashion, the concept of *offender networks* has been incorporated into FLINTS (Leary, 2001, 2002, 2003), which is very similar to link analysis, but is primarily motivated as a criminological theory. FLINTS was developed to support the detection of high volume crimes within the West Midlands, through querying databases of DNA, fingerprints, footwear and tool-marks. Analysis of the data reveals patterns, associations and links which would not have been detected had each evidence type been managed in separate systems. The latest version of FLINTS has the capability to deal with offender network analysis and assist in identifying groups of offenders. The new element of geographical profiling means officers can locate crime "hotspots" by either incident or crime type, displaying the information by area. Comparative and seasonal analysis maps show emerging trends and developing hotspots, which are presented in maps and animated formats for the user.

To date, the most popular general tool (in the UK) for an add-on to GIS visualisation is I2's *Analyst's Workstation* (I2 2004). This software has been specifically designed to help police forces analyse volume crime. Again aimed at data integration, this tool incorporates a data-warehouse and links to recognised GIS. However, the much touted "comprehensive range of analytical techniques" to "identify trends and patterns in volumes of data" is reduced to link and timeline analysis. The *PatternTracer* module is specifically aimed at locating calling patterns in telephone records and is very useful. This is an example of a very specific problem that has been solved using a "bespoke" algorithm.

The interested reader can find a survey of the importance of link analysis (and text analysis) and supporting software mentioned in (Mena 2003).

3.5.1. *Social network analysis*

The COPLINK algorithms of hierarchical clustering, block modelling and centrality measures of *degree*, *betweeness*, and *closeness* derive from the social sciences social network analysis literature. All of these measures are open to misinterpretation. It is very important to determine how the links or associations – the basis for the "network" – are created in the first place, as this will have a very significant impact on the interpretation of results. We show how adding in a geographical component (not present in other approaches) and then a temporal and frequency component can add to the interpretation of the network and its key players (Oatley and Ewart 2005).

A network was created from 342 offenders who committed 1121 crimes in the West Midlands burglary data. The network links are based upon whom the offender was arrested with for a particular crime and the geographical location of that offence. This represents a significant departure from previous methodologies in that links are on the basis of an established (albeit not proven in court) co-defendant relationship. Other approaches using mobile phone records or police intelligence employ such data to infer a criminal relationship. Here, there is little ambiguity. One advantage of this approach is that all police forces have arrest data on individuals and crime location information. They do not have to mount expensive surveillance operations or access phone records in order to apply the approach described here. One reservation, however, is that the links are on detected rather than unsolved crimes, which means the extent of the network and its range may be underestimated.

The measure of brokerage was used to explore the network of offenders. We focussed upon the point centrality concept termed *betweenness*, or the extent to which a particular point lies "between" the various other points in the graph. A point of relatively low degree (few links in or out from that point) may play an important "intermediary" role and so be very central to the network. The betweenness of a point measures the extent to which an agent can play the part of a "broker" or "gatekeeper" with a potential for control over others.

The experiments used the PAJEK software (Batagelj and Mrvar 2003) and resulted in many small sub-networks and two larger sub-networks, one of which is presented in Figure 9. The diagram is overlaid with the spatial density representation – the following comments use the orientations “northwest”, “southwest” etc in relation to the centre point of the nodes.

Offender #298 receives the highest “brokerage” value of 11, higher than #169 by merit of being a “gatekeeper” to a greater number of vertices. Offender #104 receives a very low value, because it provides no “new information” as #298 and #169 are already linked.

It can be quite clearly seen that the brokerage model accurately reflects offender #298 connecting the centre and northwest areas (#80, #53, #395, #348) with the southwest areas (#407). Also, the total offences that can be uniquely reached through #298 number 46, in comparison to #169 numbering 10. This would agree with #298 receiving a higher brokerage value than #169. What is not reflected however, is that we would wish #169 to receive a higher value as it is gatekeeper to #410 with eight crimes in the southeast area, a unique offender by right of area of operation. Similarly #169 is gatekeeper to a much wider geographic range of offenders: #104 in the east; #353 and #415 in the southwest area; and, #410 in the southeast. Offender #298’s connections only lie within the centre and slightly to the northwest. We would wish then that weightings were taken account of in the

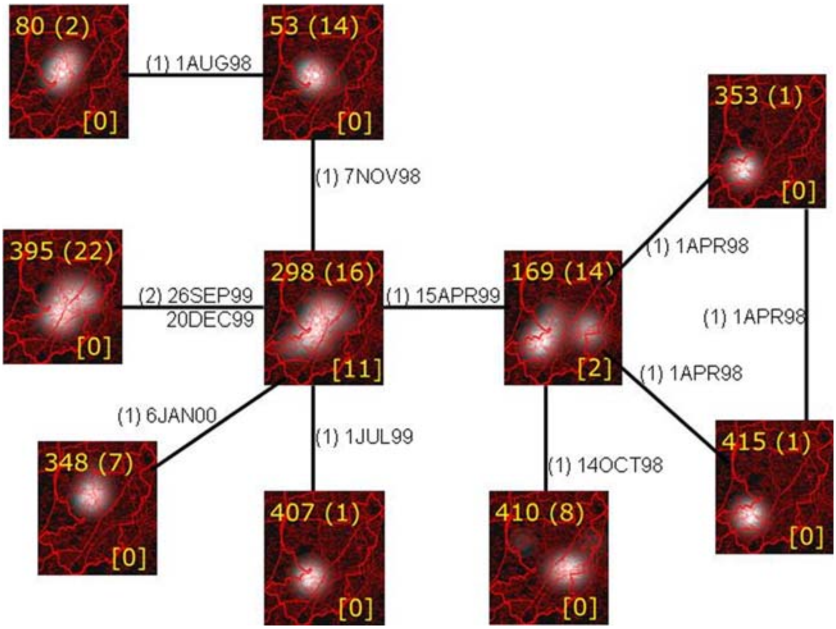


Figure 9. Brokerage, spatial, temporal and frequency analysis. The top-left figures of the nodes are unique offender identifier, with total number of crimes in brackets. The bottom-right figure is the brokerage value. The graphic displays the smoothed spatial location. Links between nodes are labeled with the number of crimes committed together by both parties and a list of dates of those crimes.

brokerage calculation (reflecting number of crimes committed between offenders), and also that the notion of criminal range (Merry 2000) was given suitable emphasis.

Examining the temporal and frequency of the links reveals for instance that the link between offenders #80 and #53 is a result of a single crime jointly committed on the 1st August 1998, and the link between offenders #395 and #298 is a result of two crimes jointly committed on the 29th September 1999 and the 20th December 1999. By examining this same network data we can understand offender behaviour in a more complete way. For instance this view of the data now reduces the brokerage values in emphasis because it can be seen that each link is mostly due to just one crime. The original interpretation of the brokerage network is also lessened by the fact that the subnetworks of { #353,#415,#169} and {#298,#169,#104} are due to *single crimes*, as the dates (and crime references – which are not shown) are the same.

The brokerage metric is very useful, however it is clear that in the case of burglary it needs to be balanced with spatial data. Consideration of the temporal and frequency analysis of the crimes constituting the links provides a better understanding of the nature of the links, and may indicate important links that are not considered important by the brokerage metric.

3.6. FORENSIC PSYCHOLOGY AND CRIMINOLOGY

This section includes work specifically driven by concepts and theories from forensic psychology and criminology. Ribaux and Margot (1999) describe the potential of computer science techniques as a means of representing the reasoning and inference processes of experienced investigators. Their approach produces sophisticated retrieval and crime matching algorithms for burglary. They intend to use their approach as a means of evaluating other data analytic techniques. A database of about 10,000 crimes containing "hard" forensic information (such as shoeprints), relatively "soft" modus operandi information (such as methods of entry) and time/space data is used to explore the ability of five different inference based structures as a means of linking crimes. Although no formal evaluative data is presented, the authors claim their computer prototype has shown "...great promise and has resulted in several operational successes".[6]

Testing their inference models with a real world data set reveals some valuable lessons (Ribaux and Margot 2003). One issue is the representation of the certainty or uncertainty of a link. Assigning a crime to a series means that one is adding the information of the new case to the other cases, but this in itself may give rise to anomalies as well as establishing inconsistencies within the crime series. They also note the need to explore how different pieces of data may be combined and how the data contributes to the analysis process. While "hard" forensic evidence provides the most rigorous evidence of a link, it is often absent at a crime scene while modus operandi information is always, to some degree, present.

A criticism of the work of (Ribaux and Margot 2003) is that while they formally define the many link types, and explain their definition of group membership and crime series membership, it is clear that the adaptation stage of their case-based reasoning (CBR) cycle is left to the user, requiring considerable understanding of the assumptions underlying the similarity metric. This is difficult to obtain as the assumptions have been automatically derived, and presumably different choices of links and thresholds would result in a completely different case-base and similarity metric. A better use of CBR could be seen as the simpler to understand approach described in Section 4.4 (*k-NN* variants based on selection of features from the original data), and better methods for determination of links as seen in the COPLINK methods (earlier Section 3.5) or techniques described in Section 4.2.3 concerning logic programming.

Modelling of the offender search area has been the subject of (behavioural) research by Brantingham and Brantingham (1981) who analysed the *geometry of crime* and conceptualised a criminal search area, a geographical area modified by the spatial distribution of potential offenders and potential targets, the awareness spaces of potential offenders, and the exchange of information between potential offenders. Rossmo (1993, 1995) has adapted *location theory*, particularly travel behaviour modelling, to serial offenders. He outlines a mathematical approach to identifying the home base location of a serial offender, given the distribution of the incidents. The mathematics represent a formulation of the Brantingham search area model, in which the search behaviour of an offender is seen as following a distance decay function with decreasing activity near an offenders home base. Rossmo's later work (1997) presents an alternative approach to geographic profiling. The Canter DRAGNET model (Canter et al. 2000) is different from the Rossmo model in that it suggests a *search strategy* by the police for a serial offender rather than a particular location aiming to indicate how narrow an area the police should concentrate on in order to optimise finding an offender. However the empirical model is strictly pragmatic, and is much less rigorous than Rossmo's method, and uses the negative exponential function without considering other distance decay models. Such models will produce a better fit to the data in certain circumstances (Crimestat Manual 2004).

The criminological statistical devices of *"prevalence, incidence and concentration"* were employed for the description of West Midlands Police geocoded burglary data. A useful paper on this issue is (Trickett et al. 1992). Their concept of "vulnerability" is the same as the concept of "concentration".

Some implications for policing using the definitions below may be found in (Pease 1998).

Incidence is the *number of crimes per household* who are available to be victimised in any given area.[7] Prevalence is the *proportion* of available households which *are* victimised in any given area. Concentration is the *number* of victimisations per "victim" household in any given area. For any given area, incidence can be expressed arithmetically as the product of prevalence and concentration. A high crime rate in an area may be because many people in that area become victims usually only once (high prevalence) *or* that same high crime rate may reflect that a few people suffer many victimisations (high concentration).

When interpreting changes the incidence rate is the first point of decision information. Where incidence is low, the number of victims and revictimisations will be low and so the area is perhaps a low priority area. *"High Prevalence, Low Concentration"* suggests that crime is spread across the area, whereas *"High Prevalence, High Concentration"* suggests that crime is spread across the area and there are a relatively high number of people who are suffering multiple victimisation. *"Low Prevalence, High Concentration"* suggests that crime is concentrated to a relatively few number of houses within an area. *"Low Prevalence, Low Concentration"* suggests that it must also be a low incidence area – so police should target the high incidence area and explore prevalence and concentration.

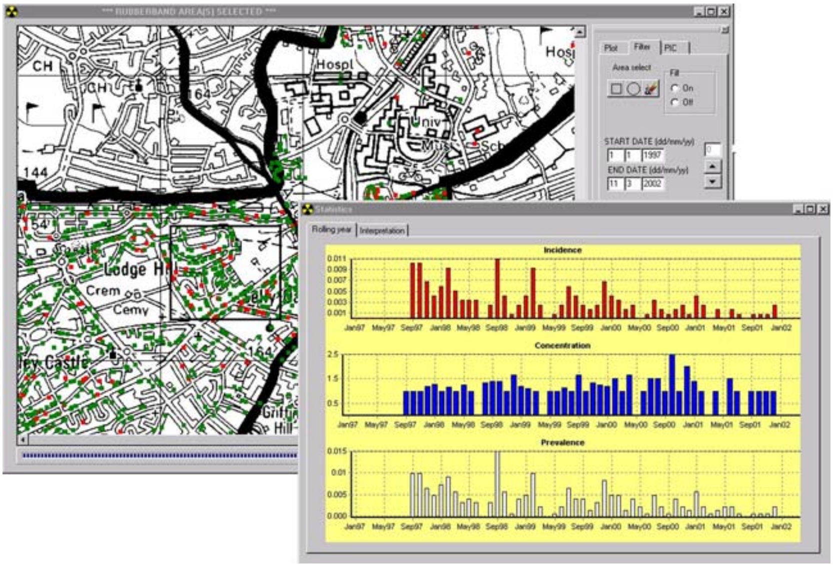
Figure 10 shows the presentation of this information and it is clear that the user is required to understand a great deal about the meaning of these features. Much more work needs to be performed to "automate" this concept as decision support software.

3.7. SECTION THREE CONCLUSIONS

The diverse approaches mentioned provide a way of visualizing data. This is important because of the otherwise overwhelming amount of data. Recall the investigators task (Section 1.2) and the volume of data recorded for even a single crime.

Typical crime data will have either a geographical nature, and/or contain a large number of features, and/or involve networks of offenders. Types of technologies relevant to these different aspects are:

- *Geocoded data.* GIS, hotspot analysis, spatial statistics (e.g. CSR: complete spatial randomness) and prevalence/incidence/concentration (PIC) metrics.
- *Large feature sets.* Dimensionality reduction using multidimensional scaling and principal components analysis.
- *Offender networks.* Link analysis and social network analysis technologies.



START DATE (dd/mm/yy) 1 1 1997

END DATE (dd/mm/yy) 11 3 2002

Figure 10. PIC statistics for a specific area, over 3 years. The prevalence, incidence and concentration are shown for the currently selected rubber-banded area (black square on map). The values for these factors are shown per month across a 3 year period (the time period of the selected data).

Any motivational assumptions or testing of offending/criminal range models in offender networks is conditional upon having substantive knowledge about the temporal characteristics of the network and the criminal activities of the key players. Conceptual issues such as how a network develops should be explored in terms of their spatio-temporal and behavioural characteristics. The effects of various disruption strategies can be modeled and evaluated. As yet, no one has explored characteristics that delineate a passive network from one which is currently engaged in criminal activity. Such research would allow for better targeting of policing resources as well as the prediction of when a passive network will become active.

The data mining technologies of association rules and decision trees/classification rules did not produce operationally useful knowledge. The association rules we derived were of poor quality, either indicating no significant relationships or relationships that are too complex to determine using this method. The classification rules (with or without dimensionality reduction) offered little insight in this instance. Similarly, the specification of the output class (the three roughly equivalent partitions of offender crimes) clearly biases the results, presupposing the knowledge in the data that is being mined. In fact once we have completed the task of data mining, we need to ask ourselves the question *"what do the results of the data-mining task* *signify?"* If the results are to indicate something more than mere data dredging, then we must provide some justification for our evidence. The evaluation of all computer software, let alone knowledge discovery from databases, is a significant issue.

The plots using pie charts and histograms provided a clear overview of the data that was previously inaccessible. Each technology aims to provide greater insight into the data to the user or decision maker. In fact it is the user who brings context and domain knowledge to the interpretation of the mined "patterns", and who must be taken into account when constructing police decision support systems. Interesting future work could investigate the best way to support the combination of CSR, PIC and kernel density algorithms, taking the technologies away from the domain of the crimes analyst, and developing methods that can be embedded in software useable by police officers.

4. Data Mining from Crime Patterns to Crime Matching and Profiling

4.1. SECTION OBJECTIVES

While the previous section provided a description of the kinds of preliminary investigations of data, which reveal interesting patterns and knowledge, this section builds upon this description by presenting techniques that support crime matching and profiling.

4.2. MATCHING SINGLE CRIMES WITH CLUSTERING TECHNOLOGIES

A neural network receives its name from the fact that it resembles a nervous system in the brain. It consists of many self-adjusting processing elements co-operating in a densely interconnected network. Each processing element generates a single output signal which is transmitted to the other processing elements. The output signal of a processing element depends on the inputs to the processing element: each input is gated by a weighting factor that determines the amount of influence that the input will have on the output. The strength of the weighting factors is adjusted autonomously by the processing element as data is processed.

In unsupervised learning, the system receives only the input, and no information on the expected output. The system learns to produce the pattern to which it has been exposed. There are many types of unsupervised networks including Self Organizing Maps (SOMs, or Kohonen networks), Grossberg nets, In-star, Out-star, Bi-directional Associational Maps and Hopfield networks.

SOMs are the most commonly used neural networks for crime investigation. They are generally used for the clustering of "crime and disorder". Wilson et al. (2002) and Adderley and Musgrove (2001) use this technology to link of records of crimes of serious sexual attacks. A number of clusters were used to form profiles of the offenders. However, before presentation to the Kohonen network, significant hand preprocessing/relabelling of text was required. In a similar fashion to their work with "bogus official" reports (Adderley and Musgrove 1999) also required significant recoding. Corcoran et al. (2001, 2003) also use SOMs, for hotspot prediction, with "explanation" of the clusters using rule abduction.

SOMs were also employed for offender predictions among the West Midlands burglary data. A genetic algorithm (the standard Holland algorithm with elitism and roulette selection) was utilised for parameter selection (Hunter 2000; Hunter et al. 2000) using the TRAJAN neural network package. The retained parameters can be seen in Table VIII.

Considering that over 120 features were input it is interesting that entry and exit point, entry method, victim sex, and the search strategy are the most discriminating variables – at face value these would be expected to be important. What may appear to be surprising is that there are no significant property stolen types listed. In a similar vein an analysis of Sunderland West crimes data by the authors (Ewart et al. 1997) showed that there was little evidence for the theory that offenders wait for certain items of property to be replaced by insurance companies and then strike again. This widely accepted "myth" is clearly not supported by these studies, and indicates the need for caution when utilizing findings from studies that may not be valid when being transferred to new domains.

Each identified cluster contained offenders with related crime behaviour, and the model was operationalised in a decision support system in two ways. First, given an unknown crime the cluster most closely matching the behaviour is identified and all offenders belonging to that cluster are marked as suspects. Second, as several offenders occurred in multiple clusters (the items stolen across their crimes varied significantly), given a particular offender find the crimes that they could have committed. To perform this task, the features of each cluster that the offender is found in are retrieved and matched against all unknown crimes. In the eventuality of a match, all the other offenders on that cluster are also listed as possibilities.

(Yokota and Watanabe (2002) use a probabilistic retrieval algorithm designed to interrogate modus operandi data held on a Japanese National Police Agency database. This method is described in Section 4.3 as it forms the motivation to recent work by the authors. In contrast to most of the studies mentioned above Yokota and Watanabe provide evaluative information on their models. For the offenders who actually committed the target crimes, their target ranks indicate the system's retrieval accuracy. On completion of their search, 20.2% of the offenders achieved a target rank score of 1. The median rank score was 29, which the authors describe as a success considering the 12,000 plus offenders in the data base.

4.3. MATCHING SINGLE CRIMES WITH SPATIO-TEMPORAL AND BEHAVIOURAL DATA

Modus operandi (MO), temporal and geographic information were used to evaluate the ability of three algorithms to match a target crime to the actual offender (Ewart et al. 2005). The motivation for this approach being the assertion by Pease (2001) that "location is almost never a sufficient basis for, and seldom a necessary element in, prevention or detection", and that non-spatial variables can, and should be, used to generate patterns of concentration.

The detected burglaries ($n = 966$) were each attributed to specific offenders ($n = 306$). The first approach (RCPA) uses only MO information, the second (RPAL) only temporal and geographic data and a third algorithm (COMBIN) is a combination of the two.

The RPAL algorithm represents spatio-temporal information and is the product of *recency*, *prolificness* and *actual location* data on each crime. *Recency* is the distance in time between any previous crime and the considered crime, while *prolificness* represents the amount of previous crimes committed by an offender before the considered crime. *Actual location* is the Euclidean distance between any previous crime and the considered crime. Each one of these metrics contained parameters that were assigned values drawn from empirical findings within the literatures of forensic psychology and criminology. However these values were also subject to optimisation. For instance, *recency* is calculated by looking for offender activity within 28, 27, 26 ... 1 days of the current crime in the local area. The more recent the last crime the more highly it is weighted. If the *recency* feature is an exact match in time, it will receive an arbitrary value of 56. This feature was later optimised, and is known as PARAM_RECENT_CLOSEST in the following Equation 2. The scaling factor known as PARAM_RECENT_SCALE receives the value of 28, and the inverse of this value is multiplied against the distance. If there is no previous crime, instead of receiving a value of zero, an additional feature of PARAM_RECENT_TINY is used to scale the distance in the appropriate direction, and which receives the value of 0.0001.

Similar equations for *prolificness* and *actual location* containing parameters that were optimised can be found in (Ewart et al. 2005).

$$\text{Recency} = \begin{cases} \dfrac{\text{PARAM\_RECENT\_SCALE}}{t_{\text{crime}} - t_{\text{last\_offence\_before\_crime}}} & t_{\text{crime}} \neq t_{\text{last\_offence\_before\_crime}} \\ \text{PARAM\_RECENT\_CLOSEST} & \text{Crimeoccurringtoday} \\ \text{PARAM\_RECENT\_TINY} & \text{No last crime} \end{cases}$$

where:PARAM_RECENT_SCALE = 28; PARAM_RECENT_CLOSEST = 56; PARAM_RECENT_TINY = 0.0001.

Yokota and Watanabe (2002) random choice probability algorithm (RCPA) is implemented by means of the empirical or naive Bayes approach (Carlin and Louis 2000) which incorporates modus operandi and information about property stolen. Naive Bayesian classifiers assume the effect of an attribute value on a given class is independent of the other attributes. This assumption is made to simplify computations – hence the use of the word naive. The method (see Equation 3) takes a sequence x_i of observations such that the distribution of the ith observation x_i depends on a parameter θ_i (see Equations 4 and 5), typically in such a way that $p(x_i|\theta_i)$ has the same functional form for all i.

The parameters θ_i are supposed to be a random sample from some (unknown) distribution, and it is this unknown distribution that plays the role of a prior distribution and so accounts for the use of the name of Bayes. There is a contrast with the "normal" use of the term Bayes, where the prior distribution represents our prior beliefs and so by definition it cannot be unknown. Further, the prior distribution in empirical Bayes methods is usually given a frequency interpretation, by contrast with the situation arising in true Bayesian methods

$$\frac{n_1}{N}\prod \frac{x_1^Z (n_1 - x_1)^{1-Z}}{n_1}. \tag{3}$$

Shrinkage estimator (x tests out of n):

$$\theta = \frac{x+1}{n+2} \tag{4}$$

$$\theta = \frac{x+\lambda}{n+2\lambda}. \tag{5}$$

The general approach involves creating a matrix containing all of the crimes for all offenders, with the presence or absence of the MO or property stolen features represented as Boolean attributes. Interpretation of this matrix enables new crimes to be matched against all the offender data. The more crimes an individual has committed in an area, the more likely they are responsible for the unknown crime.

Similarly, combining the algorithms RPAL and RCPA with their different properties within this forensic domain is an extemporized process. The lessons of this work should be used to guide the formal evaluation of such processes in the future. The COMBIN model results considered in this paper are from the best classifying RCPA and RPAL models combined together.

The optimisation was carried out using the C++ genetic algorithm library *GAlib* (Wall 1996). In order to represent the 105 modus operandi and 32 property stolen features, the *GAlib* genome *"GA1DBinaryStringGenome"* was chosen. Each bit in this string would represent the presence or absence of the feature. There were two main objective functions (OF's) used to determine the fitness of a solution. In both cases the rank order for each crime whose offender was known was compared against each of the algorithms (RCPA or RPAL). For each algorithm and for each offender the rank per crime is calculated. OF1 used the metric of the number of times an offender occurs in the top 30 for his/her own crime. The value 30 was based upon the value of 29 determined by Yokota and Watanabe (2002) and represents a reasonable number of offenders to search through. OF2's metric was to add the rank position onto the total. For example, consider the example dataset in Table IX, where offender number 11 has committed two crimes, and the algorithm has (very accurately) placed this offender rank 2 and rank 1 for these crimes. The OF values can be seen to be very different, with OF1 favouring the higher ranks, and OF2 biased towards optimising the whole dataset.

The final results that are considered in this paper, from the COMBIN model, are derived from the best classifying RCPA and RPAL models combined together, with the parameters optimised using OF1.

Yokota and Watanabe (2002) employ a similar tactic in that they chose the random choice probability procedure method after comparing its ease of mathematical applicability and performance against two other methods of probability calculation (the constant and the kernel). Tables X, XI show some results. The RPAL and COMBIN each achieve a perfect match for 24% of the crimes. For prolific offenders, matching using MO information alone is better than temporal and geographic data, although the best performance is achieved when in combination.

Future work will optimize the RCPA selection of modus operandi parameters.

4.4. MATCHING SINGLE CRIMES WITH SIMILARITY-BASED RETRIEVAL

Recent work in matching single crimes with similarity-based retrieval, includes a retrieve-and-propose CBR system (Oatley 2004). CBR problem solving involves: generating a case base; retrieving the most similar case(s) to

Table IX. Comparison of objective functions. Reference is the Unique crime reference, Number is the Unique offender number, Rank position is the rank position of the actual offender by the algorithm

Reference	Number	Rank position	Objective function	
			OF1	OF2
Ref 1	1	54		54
Ref 2	11	2	1	2
Ref 3	11	1	1	1
Ref 4	78	34		34
Ref 5	43	60		60
Ref 6	3	6	1	6
Total			3	157

Table X. Observed cumulative percentages. Values of the crimes given a ranking of 10 or less, 30 or less and 50 or less. The observed minus expected cumulative percentages (adjusted for the different rank ranges produced by the models) are given in parenthesis. * The RPAL model did not generate the exact ranks of 10 an 30 and so the observed percentages are based on the nearest alternatives of 8 and 27

Model	RCPA	RPAL*	COMBIN
Rank of 10 or less	25% (22%)	52% (50%)	59% (57%)
Rank of 30 or less	51% (42%)	52% (46%)	77% (68%)
Rank of 50 or less	62% (47%)	53% (43%)	94% (79%)

Table XI. Mean retrieval ranks. These ranks are generated by the RCPA, RPAL and COMBIN models for each of the three offender groups (SDs are in parenthesis)

Model	RCPA	RPAL	COMBIN
Group 1 (committed 1–5 crimes)	195.96 (88.42)	406.26 (158.17)	35.07 (24.86)
Group 2 (committed 6–10 crimes)	35.10 (9.28)	181.45 (117.25)	12.62 (8.69)
Group 3 (committed over 10 crimes)	13.73 (7.92)	84.02 (84.41)	7.53 (7.57)

the query case from the case base; calculating the similarity between the retrieved case and the query case; reusing or adapting the information and knowledge in the retrieved case to solve the query case; revising the proposed solution, according to the implemented result; and, retaining the parts of this experience likely to be useful for future problem solving (either as a new case or an update to the domain knowledge).

The CBR system used three different similarity metrics – nearest-neighbour metric, Tversky's contrast model, and the cosine and modified-cosine matching functions. The nearest neighbour metric is widely used in CBR systems (Weiss and Indurkhya 1998; Witten and Frank 1999; Wilson 2001) and computes an algebraic average similarity (Gebhardt et al. 1997) by totaling the weighted similarities for each attribute pair, and dividing by the total of the attribute weights.

Typically, a similarity between the new case n and a source case P_k is the weighted sum of the similarity of each attribute pair ~$(\overset{n}{a_i},\overset{pk}{a_i})$, divided by the sum of the attribute weightings, where n is the number of attributes, w_i is a weight of the ith attribute (Gupta and Montazemi 1997, p. 602):

$$\mathrm{NN} = \frac{\sum_{i=1}^{n} w_i \sim \left(a_i^n, a_i^{Pk}\right)}{\sum_{i=1}^{n} w_i}. \tag{6}$$

The weighting here is *global*, in the sense of being applicable across all cases. The nearest neighbour scheme is simple and effective, however there are a number of important problems with this method. First, it is costly for large case-bases because the entire case-base has to be scanned (Roiger and Geatz 2003). However, performance can be improved by using data mining algorithms to cluster the cases into a hierarchical structure (Weiss and Indurkhya 1998). Second, it performs poorly with noisy data, because the solution is determined locally, often from only the single nearest neighbour, without any averaging to help eliminate noise (Witten and Frank 1999). In addition, the nearest-neighbour metric is particularly sensitive to irrelevant attributes, as all attributes contribute equally to the similarity formula (Roiger and Geatz 2003).

One solution is to weight the attributes according to their relevance to solving the problem (Chklovski 2003), the less relevant attributes having a lower weighting and thus a lesser influence. Still, one of the main problems in assessing similarity is determining which attributes are the most relevant (Aggarwal 2003). Weights may be determined by a domain expert, or alternatively data mining algorithms can weight attributes based on their problem-solving performance (Aha 1992; Oatley et al. 1998).

Another solution is to adopt the *k-nearest neighbour* (K-NN) method (Roiger and Geatz 2003), where a number of nearest neighbours are retrieved and either the majority class, or the distance–weighted average if the class is numeric, is used for the solution (Witten and Frank 1999). However, this inevitably increases computation time, further adding to the computational cost of performing this metric.

The cosine rule (CR) has been used extensively in information retrieval systems for matching a query with the documents in a database (Hastie et al. 2003), and includes the benefit of *local* weighting for attributes, where the weight is dependent upon each particular case.

Tversky's contrast (TC) model (Tversky 1997) extends the nearest neighbour algorithm and cosine rule approaches (geometric representations of similarity) as they are unsuitable for certain types of data (Vargas et al. 1998), and do not mimic aspects of human similarity judgements, raising questions about their utility for determining similarity (Chklovski 2003). The method includes a contrast between attributes that match and those that do not. Application of this contrast has been limited in CBR because similarity between attributes is assumed to be binary, and it does not incorporate attribute weights. However, some authors (Vargas et al. 1998; Chklovski 2003) have developed methods to apply Tversky's model which use attribute weighting, and can produce interval-scaled similarity values.

To overcome the limitations of the nearest neighbour approach, Tversky's model and the cosine rule, Gupta and Montazemi (1997) proposed a modification to the cosine similarity metric (MCR), which not only includes local weighting, but also computes a contrast.

The case base for our experiments consisted of modus operandi, property stolen, time and location features – for each crime where the offender was known (1140 cases). In this way we propose an offender for an unsolved crime.

Each of the similarity metrics (K-NN, TC, CR, MCR) were coded into a CBR system, and tested against every case in the case base. The rank position of each offender was determined against their own crime (ideally they should be retrieved with rank position 1), and the average rank across all offenders was the measure by which to compare the similarity performance – this is admittedly a very simple metric, which does not consider the bias introduced by offenders committing varying numbers of crimes.

As the data consisted solely of globally weighted Boolean features, and each case was measured across the same features in query and case base (no contrast), an addition to these metrics was the selection of features before presentation to the similarity metrics. The four feature selection algorithms were: Best-First, Forward Selection, Decision Table and Genetic Search. These were taken from the WEKA data mining tool, as the CBR system was written in Java.

The results were that CR and MCR produced very similar similarity ratings. However, MCR discriminates better with ranking – average mean and median rank is always higher for all feature sets. Surprisingly KNN and TC produced the same average similarity and rank thus perform equally well, and they always (across all feature sets) produced higher average similarity ratings and rankings than CR and MCR.

It is expected that the modified methods (especially MCR) would be more useful when the data representation is more complex. The features selected are under review also, as is the most appropriate way to weight the features.

4.5. MATCHING SERIES OF CRIMES WITH LOGIC PROGRAMMING

A logic programming language (Prolog) was used to generate flexible matching facilities. The data (see earlier Tables II–VI) were converted into Prolog facts, for instance the example of Offender data given in Table V would be represented as the fact: offender ("20E1/4441/98", m, 14, 19).

The property stolen and modus operandi data are converted into ontologies to permit generalisation of matching, for instance "cash" and "cheques" can be generalized to "money". The ontologies are imported into the SWI-Prolog (Wielemaker 2003) environment as RDFS (Wielemaker et al. 2003). The benefit of using such an ontology (which is essentially a set of predicates) is that it is easy to view and adapt in an editor such as (Protege 2004), and to engage a domain expert in its construction (Noy et al. 2000).

The logic programming paradigm gives an extremely flexible way of querying and examining the data. Consider the way that a *series* of crimes can be determined. To be a member of a series each crime in the list has to be *linked* with another by some `location_link`,`property_stolen_link`, `property_description_link`, `time_link` and `gang_link`.

An example of the former is:

```
location_link(CRIMEREF1,CRIMEREF2,X_DISTANCE,Y_DISTANCE):-
crime(CRIMEREF1,..., X_COORDINATE1, Y_COORDINATE1),
crime(CRIMEREF2,..., X_COORDINATE2, Y_COORDINATE2),
not(CRIMEREF1 = CRIMEREF2),
abs(X_COORDINATE1 -- X_COORDINATE2) <X_DISTANCE,
abs(Y_COORDINATE1 -- Y_COORDINATE2) <Y_DISTANCE.
```

The remaining predicates are equally simple except the latter,`gang_link`, which is determined as follows.

```
same_crime(OFFENDER_ID1,OFFENDER_ID2):-
offender(CRIMEREF,_,_,OFFENDER_ID1),
offender(CRIMEREF,_,_, OFFENDER_ID2),
not(OFFENDER_ID1 = OFFENDER_ID2).
strong_offender_friends(OFFENDER_ID1,ListN):-
findall(OFFENDER_ID2,
same_crime(OFFENDER_ID1,OFFENDER_ID2), List),
remove_duplicates(List,ListN).
weak_offender_friends(OFFENDER_ID,ListN):-
strong_offender_friends(OFFENDER_ID,List1),
full_list(List1,List2),
flatten(List2,List3),
remove_duplicates(List3,ListN).
```

```
full_list([],[]). full_list([H|T],[ListN|T2]):-
strong_offender_friends(H,ListN),
full_list(T,T2).
gang_link(OFFENDER_ID1, OFFENDER_ID2):-
weak_offender_friends(OFFENDER_ID1,List),
% broader search than:
% strong_offender_friends(OFFENDER_ID1,List)
member(OFFENDER_ID2,List).
```

It is simple to determine a series, given the set of all the links – all that remains are that the links are ordered chronologically. This then provides the data miner the possibility of examining the features of crimes in the series, also of finding out all crimes carried out by a gang, and determining an appropriate metric for crime matching. As well as being able to relax the "friendship" (strong_offender_friends or weak_offender_friends), it is also possible to relax the criteria forproperty_stolen_link (i.e. move up one step in the ontology), relax the geographical search area, and relax the time period between crimes. Of course all of these "relaxing" criteria are not equivalent, and it might be pertinent on one occasion to relax the geographic catchment area, however on another occasion relax the property stolen criteria, or gang-membership.

This work is in its early stages, and it is clear that development of this approach for exploratory purposes will need to closely tie in with development with the graphical user interface.

Mooney et al. (2004) describe the benefits in using such an inductive logic programming approach to identify complex relational patterns that might indicate terrorism threats in large amounts of relational data. They have tested their approach on "nuclear smuggling" and "contract killing" data to great success, although optimizing the performance is an issue that is still to be addressed with large datasets.

A final point to mention is the non-monotonic nature of the Prolog negation, which provides the possibilities of a certain kind of default reasoning (Krause and Clarke 1993).

4.6. SECTION FOUR CONCLUSIONS

The ability to match crimes is an essential part of the investigative process. It is unlikely that a crime investigator will be able to consistently make links between crimes in anything other than exceptional cases. The sheer volume of crimes such as burglary, motor vehicle theft and street robbery (the so-called "volume crimes") means that the majority of significant crime patterns will remain undetected in the original crime recording database.

There are a wide range of techniques that make it possible to discover useful information to assist in crime matching, not only of single crimes, but also of series of crimes. These certainly are a diverse set, and the examples in this section are certainly not the only techniques, although they are illustrative of the potential in this area.

One interesting difference between these technologies lies in the amount of explicit encoding of domain knowledge. The artificial intelligence techniques of self organising maps, CBR (with feature selection for dimensionality reduction), and empirical Bayes (using genetic algorithms for dimensionality reduction) required no domain knowledge, in contrast to the domain knowledge intensive technology drawn from the same discipline, namely the logic programming approach, which was crafted explicitly for this problem domain.

The forensic psychology approach also required explicit encoding of domain knowledge, and was contrasted with the artificial intelligence approaches. In fact the combination of both approaches fared best. This particular set of experiments also demonstrated that the widely accepted belief that offenders wait for certain items of property to be replaced by insurance companies and then strike again is not supported by these studies, and indicates the need for caution when utilizing findings from studies that may not be valid when being transferred to new domains.

We conclude by noting that (i) the performance of these models relies completely on the quality of the data – not only the accuracy of collection, but of the actual features collected, and so features should be collected which will assist the development of crime matching algorithms, leading to efficiency gains in personal time, and, (ii) all of these methods could potentially be embedded in decision support software.

5. Developing Predictive Systems from Predictive Clustering to Bayesian Models

5.1. SECTION OBJECTIVES

There exists very little literature about predictive models for police decision support. Earlier work by the authors with the Cleveland data set included the use of neural networks for the prediction of repeat victimisation (Oatley et al. 2002), however this work was more of a demonstration system, being one of several components of a DTI-funded SSDM project.[8] This research predicted the occurrence of the next crime, and the expected time interval for this victimisation. The best neural network models trained on this data for predicting the next crime achieved an average performance, and it was impossible to find a neural network model for predicting the time interval without either obvious over-training, or extremely poor performance and it is unfortunate that in this earlier study there was no time to use a dataset where the time intervals were not left as single point values – previous studies had shown the significance of time intervals (Polvi et al. 1991; Ewart et al. 1997).

Until recently, very few computer systems have attempted to make decisions using evidence sources. Tillers and Schum (1988) and Schum (1994) discussed Wigmore's pioneering approach (Wigmore 1913) to proof in litigation. Wigmore's method of diagramming the body of evidence in a case is the central method behind (Walton 2002) treatise on legal argumentation and evidence. Schum and Tillers (1991) examines marshalling legal evidence. Kadane and Schum (1996) used probability and Wigmore's diagrams of evidence to analyse the trial of the American anarchists Sacco and Vanzetti. Various techniques have been used to construct criminal investigation decision support systems. Statistics has been used to analyse evidence (Schum 1994; Aitken 1995). Areas investigated include DNA testing, fingerprints, footwear and ballistics. This section presents a statistical and artificial intelligence method for crime prediction and the ways that evidence sources can be used. The section also illustrates the way that developed analyses can be embedded in decision support software for use by non-experts.

5.2. SURVIVAL/FAILURE TIME ANALYSIS

Survival/failure time analysis, developed primarily in the health sciences (Kleinbaum 1996), deals with the issue of censored observations, which arise whenever the dependent variable of interest represents the time to a terminal event, and the duration of the study is limited in time. For example, it is possible to study the "survival" of victims of a given crime, given the current point in time. This technique was used with the West Midlands Police data (Ewart and Oatley 2003). However, what is of interest here is the survival time within a group of households known to be at high risk of a burglary (referred to below as the Twelve Month Repeats group). The criminological literature consistently tells us (Pease 1998) that it is those who have suffered a prior burglary who are at greatest risk of being further burgled. The problem is that not all burglary victims will be revictimised so we wanted to explore if modus operandi data could be analysed to distinguish (albeit retrospectively) those who were not revisited from those who were revictimised. Binary logistic regression with the "*Single victimisations*" and "*First in a series*",[9] as the dependent groups were executed with the set of crime scene variables appearing in Table XII. Fourteen

Table XII. Probability densities and hazard rates of revictimization. Repeat properties are the No. of Properties Suffering a Repeat During the Interval. These are during successive 56 day intervals for properties comprising the Twelve Month Repeats group

Interval start time (in days)	No. of properties entering the interval	Repeat properties	Probability density	Hazard rate
0	606	196	0.0058	0.0069
56	410	87	0.0026	0.0042
112	323	87	0.0026	0.0056
168	236	63	0.0019	0.0055
224	173	66	0.0019	0.0084
280	107	73	0.0022	0.0185
336	34	34	0.0010	0.0357

discriminating modus operandi features were found through experiments with binary logistic regression. These are discussed in (Ewart and Oatley 2003). Examples are "ENTRY METHODS AND BEHAVIOUR = FORCED", "ENTRY METHODS AND BEHAVIOUR = CLIMBED", "SEARCH BEHAVIOUR = SEARCH ALL ROOMS".

The use of force, searching behaviour, type of property, place of entry, place of exit, alarm activation and use of a bogus official method of entry are discriminating features. Comparing non-repeats with "quick" repeats (i.e. within 365 days), searching behaviour, type of property, entry method and a bogus official strategy are discriminating features.

So, we are able to identify the features of a "first" burglary which are predictive of a revictimisation. We now wanted to explore if we could predict the time scale of that second burglary. Survival analysis was employed to examine if a burglar's behaviour could be used to identify which properties within this high risk group were likely to be burgled "sooner rather than later". Examining the proportions of the group that are revictimized within specified time intervals reveals the temporal pattern of repeat burglaries. This is represented at each time period by the *hazard rate* and the *probability density*. The former is of the rate of being revictimized at the midpoint of a time interval, while the latter is the *probability* of being revictimized by the midpoint. As mentioned, Polvi et al. (1991) found the risk of a repeat was highest within a month (28 days) of the first crime. To facilitate comparison, the time intervals here are taken in increments of 56 days. Cox regression examines the association of modus operandi variables and the timing of revictimisations. The period (in days) to the second victimisation is the "survival time" while the crime scene variables are the covariates. Separate analyses are conducted for each group of variables. Nineteen properties were revictimised on the same day (i.e. within hours of the first burglary) and so have a survival time of zero days. These would normally be dropped by SPSS as they have non-positive time values. Their inclusion was achieved by giving each a time value of 1 day.

The highest probability of revictimisation occurs within the first time interval. The probability of not surviving (i.e. being revictimized) to the mid point of the 0–56 day interval is 0.0058. The two subsequent intervals each have a probability density of .0026. The lowest probability of repeat burglaries is found in the longest interval beginning 336 days from the first crime.

In Table XIII, there are successive 56 day intervals for properties comprising the Twelve Month Repeats group. The rate of revictimisation is greatest within the longest time intervals. The hazard rate (0.0069) for the shortest time interval is the fourth highest across the seven intervals. The interval beginning at 56 days after the first burglary has the lowest hazard rate (0.0042) (see Table XII).

Table XIII presents the significant crime scene covariates. Ramming and removing glass to gain entry are strongly associated with early revictimisation. A search of all rooms is a marginally significant indicator of relatively early revictimisation. In contrast, exit by a window is significantly associated with a longer period between the first and subsequent burglary.

5.3. BAYESIAN NETWORKS

Bayesian methods provide formalism for reasoning about partial beliefs under conditions of uncertainty. In this formalism, propositions are given numerical values, signifying the degree of belief accorded to them. Bayesian classifiers (Section 4.3) are statistical classifiers that can predict class membership probabilities – such as the probability that a given

Table XIII. Significant crime scene variables. Variables which survival analyses reveals are significantly and marginally significantly associated with the time to revictimisation. These are for properties comprising the Twelve Month Repeats group. The mean time to revictimization is presented for those properties where the variable is present (code P) and for properties where the factor is absent (code A)

Grouping	Variables	Beta value	Significance	Mean time (days) to revictimization
Entry methods and behaviour	7. Remove Glass	−0.57	0.03	P = 73, A = 146
	8. Ram	−1.55	0.003	P = 26, A = 144
Search behaviour	6. Search All Rooms	−0.25	0.06	P = 131, A = 154
Location of exit	8. Window	0.35	0.05	P = 168, A = 133

sample belongs to a particular class. Han and Kamber (2001) claim that studies comparing classification algorithms have found that the nave Bayesian classifier is comparable in performance with decision tree and neural network classifiers.

In practice, however, dependencies can exist between variables. Bayesian belief networks, however, are graphical models, which unlike nave Bayesian classifiers allow the representation of dependencies among subsets of attributes. Bayesian belief networks can also be used for classification.

There are many excellent books on the use of probabilistic reasoning for the construction of intelligent systems. Pearl (1988) and Schum (1994) are two such examples.

Bayesian belief networks specify joint conditional probability distributions. They allow class conditional independencies to be defined between subsets of variables. Bayesian belief networks provide a graphical model of causal relationships on which learning can be performed. Bayesian belief networks are also known as belief networks, Bayesian networks and probabilistic networks.

A Bayesian belief network is defined by two components:

- A directed acyclic graph where each node represents a random variable and each arc represents a probabilistic dependence. Each variable is conditionally dependent of its non-descendents in the graph, given its parents. The variables may be discrete or continuous.
- A conditional probability table (CPT) for each variable, P(X|parents(X))

The network structure may be given in advance or inferred from the data. The network variables may be observable or hidden in all or some of the training samples.

If the network structure is known and the variables are observable, then training the network consists of computing the CPT entries; similar to the case of computing the probabilities involved in naive Bayesian classification.

When the network structure is given and some of the variables are hidden, then a method of gradient descent can be used to train the Bayesian belief network. The object is to learn the values for the CPT entries. See (Han and Kamber 2001) for details.

A Bayesian belief network (Pearl 1988; Jenson 1996; Cowell et al. 1999; Pearl 2000) was constructed as part of the West Midlands burglary initiative, as shown in Figure 11. "Offender Features" have been already mentioned in Section 4.3 as they were involved in other experiments. However as previously mentioned, they also included a kernel estimate, and were the most computationally intense feature to calculate.

"Property Crime History" was simply the number of previous crimes suffered by this premise (irrespective of 365 day "cut-offs"). "Area Status"

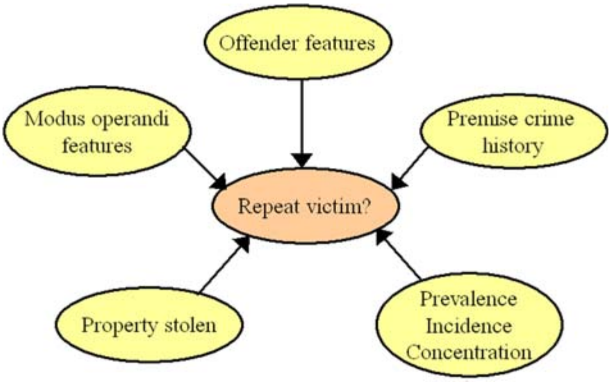


Figure 11. Burglary data Bayesian network.

are the aforementioned PIC features. "Significant Property Stolen at the First Crime" (45 categories) had been explored by the Police before our analysis, using a frequency count – with "AUDIO/RADIO" and "JEWELRY" being demonstrated to be the most significant variables. The remaining feature is "First Crime: M.O.", which was subjected to analysis through binary logistic regression analysis. The specific details of this approach can be found in (Oatley and Ewart 2003).

The results of the Bayesian network (for instance calibration) are not presented, indeed this approach has never been validated – this is because it contained many arbitrary decisions and was intended as a prototype "proof of principle" demonstration – although considerable thought was given to how this algorithm (and others mentioned throughout this paper) is incorporated into decision support software, for use by police officers.

This is shown in Figure 12, and also shows the interface to the Kohonen neural network for matching crimes against an offender list (see earlier Section 4.2). The user interface is simple enough for a Police officer to use. As the figure illustrates it is possible to predict whether or not a property (a single property, or within a selected area) will be re-victimised, and if so predict the time frame within which the re-victimisation will occur. Based on the developed offender profiles, it is possible to reveal which unsolved crimes can be attributed to a certain offender, and, for a given victimisation, provide a rank ordered list of possible offenders.

The end user need not know how to use the sophisticated technologies of Bayesian belief networks or neural networks to use this interface. Each selection is based on features that the end user would find useful over the available data. The algorithms can be run on a specific premise selected with the "mouse" ("clicked property"), or a "rubber-banded" area ("banded property") or all of the currently loaded data ("loaded data").

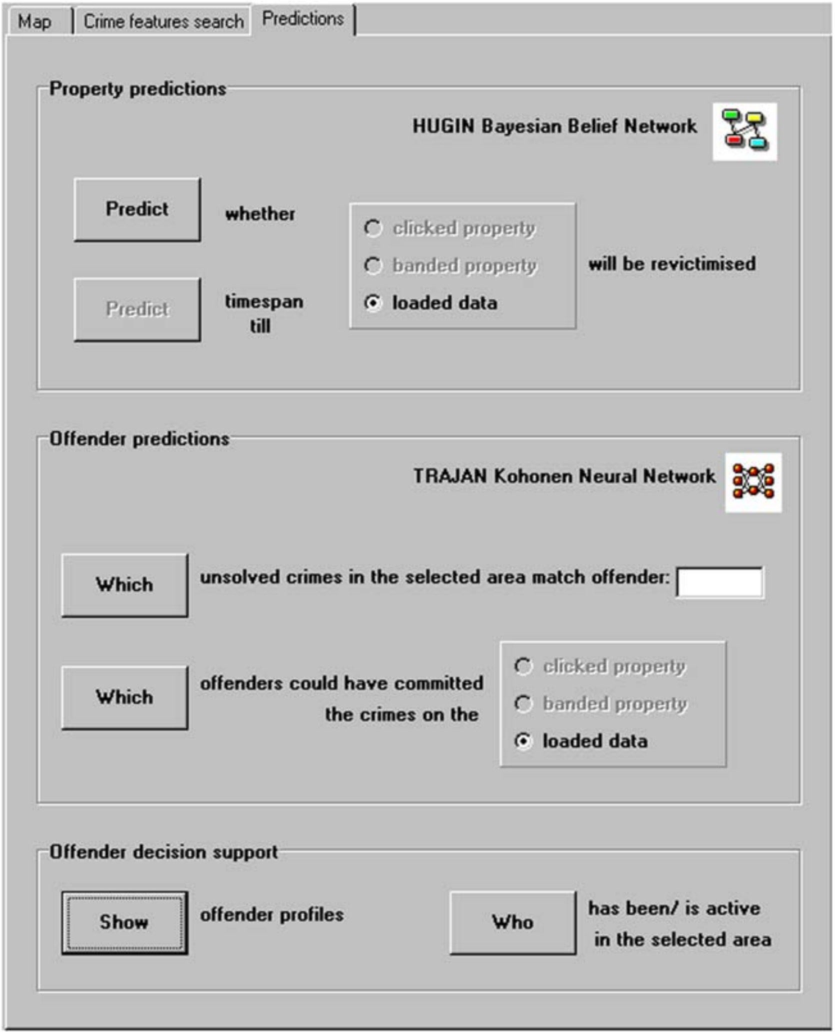


Figure 12. Available predictive models. The models include the Bayesian belief network. The end user need not know how to use the sophisticated technologies of Bayesian belief networks or neural networks to use this interface. Each selection is based on features that the end user would find useful over the available data. The algorithms can be run on a specific premise selected with the "mouse" ("clicked property"), or a "rubber-banded" area ("banded property") or all of the currently loaded data ("loaded data").

5.4. SECTION FIVE CONCLUSIONS

The ability to predict where crimes will occur has obvious benefits for resource allocation. Two different technologies have been presented that can be used for prediction, namely survival analysis from statistics and Bayesian networks from artificial intelligence. Both technologies required significant data transformation from the format of the original crime recording database, in fact the Bayesian approach utilised features that were extremely computationally expensive to calculate. While the Bayesian technique remains very much a demonstrator system, demonstrating how differing evidence types can be combined and brought to bear in a complicated bespoke predictive system, the presentation of this algorithm to its end-users as decision support software has received considerable attention, and illustrates how such algorithms can be operationalised.

The development of criminal decision support systems is greatly influenced by the intended user and by the concepts that guide and inform the design of the system. In the work presented in this section the focus has been upon crime detection and in particular to support police work, and so there has been no need to focus upon admissibility questions affecting proof. Similarly, while the provision of explanations is a central feature of any automated reasoner and especially KDD systems, neither system incorporated explanations. In fact the outputs of most KDD systems will be essentially statistical, and the only justification for the answer received was that it achieved a certain level of statistical significance. Such explanations are rarely acceptable to end-users, and so explanations that are distinct from the inferencing are vital in convincing a user of the value of using a KDD system.

However, despite the lack of explanations, and that the systems were costly in data processing and analysis, this has to be weighed against the benefits to a police officer of simply selecting data and receiving a spreadsheet of likely crimes – before the events have happened.

6. Overall Conclusions and Future Directions

In this paper we have surveyed the use of police decision support systems and described ones that have been constructed for various UK police forces primarily in the domain of Burglary from Dwelling Houses (BDH). The findings of this work demonstrate the value of crime scene information held by the police. When guided by an appropriate criminological and psychological frameworks, the benefits of more substantive statistical analyses to crime prevention and detection initiatives are clear.

Technologies have been described to interrogate the database of recorded crimes in order to explore the temporal and spatial characteristics of BDH across an entire operational command unit. The objectives were to allow police to test their beliefs about BDH trends and patterns against the empirical realities, thereby providing a more substantive empirical foundation for the development and implementation of preventative and detection strategies. Furthermore, the data described was used to develop diverse decision support tools. All of these sophisticated matching and predictive abilities are based on analysis of the data using techniques from statistics and artificial intelligence. However, it is important to recognize that the design of the visualisation, matching and predictive systems have been guided and underpinned by concepts from forensic psychology and criminology. Without recourse to such work, design decisions would be at best somewhat arbitrary and at worst, ill informed and actually compromise the validity and operational value of the system. However, it is important to rely only on those concepts which are empirically established, ecologically valid and reliable.

The paper commenced by considering what are the user requirements of those people active in criminal investigation and a discussion throughout of knowledge discovery from databases. The disparate elements that can be deployed in Police decision support systems have been illustrated. The work is necessarily broad, ranging from well established fields of geographical information systems, to the newer applications of classification and predictive technologies in this domain. While the simple visualisation of data will always be important, it is to the latter approaches that we look for significant operational advances. It is these developments which provide significant gains to the inference processes of the crime investigator.

While crimes classifiers have been presented, it is suggested that the symbolic approaches are preferred in this domain because of the transparency of the logic and method of inference. For instance the modified cosine rule matching function in the case-based reasoning experiments does not incorporate complex domain knowledge, which is costly to acquire and easy to get wrong, and is also very quick to implement over the original data. Some of the reviewed algorithms developed for specific purposes required much more bespoke data processing. While the authors' Bayesian network contained, at face value, most of the factors that could possibly be brought to bear on predicting future crimes, the processing required, especially for the offender features, was expensive. Future work needs to identify which are the most salient features for these and other methodologies.

The paper has concentrated upon the volume crime of burglary. In drawing any conclusions to other criminal domains, it is important to note that the number of attributes required to model Burglary from Dwelling Houses (BDH) is small and the domain is not very open-textured.[10] Other intelligence and policing domains[11] may be more difficult to model. However, the lessons from this experience are important to such developments.

The paper began by describing the challenges facing the crime investigator. It is not just a question of the quantity of information available, but the very nature of the cognitive processes which may operate. We conclude that computer science has a central role to play in utilizing and developing decision support technologies founded upon statistical, criminological and psychological frameworks. Furthermore, we have also provided examples where such technologies may be used to test specific theories about offending behaviour. The value of these multidisciplinary collaborations illustrates the importance of computer science to the emergent discipline of Crime Science.

Acknowledgements

The authors are grateful to staff at Bourneville Lane Police Station, West Midlands Police, for supplying the crime data and the Home Office for funding under the Crime Reduction Programme. Some experiments in this paper used the GAlib genetic algorithm package, written by Matthew Wall at the Massachusetts Institute of Technology.

Notes

[1] And hence their automobile did not have local numberplates.

[2] They lived in their vehicle.

[3] Global JXDM – http://www.it.ojp.gov/jxdm/ – intended as a data reference model for the exchange of information within the justice and public safety communities.

[4] Which use a search algorithm to find significant clusters of points.

[5] CrimeStat is a spatial statistics program for the analysis of crime incident locations, developed by Ned Levine & Associates under grants from the National Institute of Justice (grants 1997-IJ-CX-0040 and 1999-IJ-CX-0044).

[6] At p. 193.

[7] See page 6 onwards.

[8] Smart Software for Decision Makers, part of the Information Society Initiative – see: http://www.dti.gov.uk/iese/vol1summary.pdf.

[9] Refer back to the earlier Figure 1 for a graphical explanation of these terms.

[10] Open textured legal predicates contain questions that cannot be structured in the form of production rules or logical propositions and which require some legal knowledge on the part of the user in order to answer.

[11] And in particular detecting, preventing and prosecuting acts of terrorism.

References

Aamodt, A. (1990). Knowledge-intensive Case-based Reasoning and Sustained Learning: In Aiello, L. (Ed.), Proceedings of the 9th European Conference on Artificial Intelligence, ECAI-90, Stockholm, August, 6–10. Pitman Publishing London, 1–6.

Aamodt, A. (1994). Explanation Driven Case Based Reasoning. In Aamodt, A., Wess, S., Althoff, K. and Richter, M. (eds.), Topics in Case Based Reasoning. Springer Verlag: 1848 Berlin, 274–288.

Aamodt, A. (1995). Knowledge Acquisition and Learning by Experience – The Role of Case Specific Knowledge. In Kodratoff, and Tecuci (eds.), Machine Learning and Knowledge Acquisition, 197–245, Academic Press Ltd, ISBN 0-12-685120-4.

Adderley, R. and Musgrove, P. B. (1999). Data Mining at the West Midlands Police: A Study of Bogus Official Burglaries. BCS Special Group Expert Systems, ES99, 191–203. Springer-Verlag: London.

Adderley, R. and Musgrove P. B. (2001). Data Mining Case Study: Modeling the Behavior of Offenders who Commit Serious Sexual Assaults. In Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 26–29, 215–220. ACM: San Francisco, CA, USA.

Aggarwal, C. C. (2003). Towards Systematic Design of Distance Functions for Data Mining Applications. In Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Washington D.C, USA, 9–18. ACM Press: New York, USA.

Agrawal, R. and Srikant, R. (1994). Fast Algorithms for Mining Association Rules. In Proceedings of the 20th International Conference on Very Large Data Bases, 487–499. Santiago, Chile.

Agrawal, R. and Srikant, R. (1995). Mining Sequential Patterns. In Proceedings of the International Conference on Data Engineering (ICDE), Taipei, Taiwan.

Agrawal, R., Imielinski, T., and Swami, A. (1993). Mining Association Rules Between Sets of Items in Large Databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, D.C., May 26–28, ACM Press, 207–216.

Aha, D. (1992). Tolerating Noisy, Irrelevant and Novel Attributes in Instance-Based Learning Algorithms, International Journal of Man–Machine Studies 36(2): 267–287.

Ahmed, S., Cohen, F., and Peng, P. (2003). Strategies for Partitioning Data in Association Rule Mining. In Proceedings AI-2003, 23rd SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, Dec 2003, Springer: Cambridge, 127–140.

Aitken, C. (1995). Statistics and the Evaluation of Evidence for Forensic Scientists. John Wiley and Sons: Chichester, UK.

Allen, J. F. (1991). Temporal Reasoning and Planning: In Allen, J. F. (eds.Reasoning About Plans. Morgan Kaufmann San Mateo, 8–33.

Barker, M. (2000). The Criminal Range of Small-Town Burglars. In Canter D. and Alison L. (eds.), Profiling Property Crimes. Ashgate Publishing Co.

Batagelj, V. and Mrvar, A. (2003). Pajek – Analysis and Visualization of Large Networks. In Junger, M. and Mutzel, P. (eds.), Graph Drawing Software, 77–103. Springer (series Mathematics and Visualization): Berlin, ISBN 3-540-00881-0.

Bishop, C. (1995). Neural Networks for Pattern Recognition. Clarendon Press: Oxford.

Bowers, K., Newton, M. and Nutter, R. (2001). A GIS-linked Database for Monitoring Repeat Domestic Burglary: In Hirschfield, A. and Bowers, K. (eds.), Mapping and Analysing Crime Data – Lessons from Research and Practice. Taylor and Francis London, New York, 120–137.

Brantingham, P. L. and Brantingham, P. J. (1981). Notes on the Geometry of Crime: In Brantingham, P. J. and Brantingham, P. L. (eds.), Environmental Criminology. Waveland Press, Inc. Prospect Heights, IL, 27–54.

Canter, D. (2000). Offender Profiling and Criminal Differentiation, Legal and Criminological Psychology 5: 23–46.

Canter, D. and Alison, L. (2000). Profiling Property Crimes. In Canter, D. and Alison, L. (eds.), Profiling Property Crimes, 1–31. Ashgate Publishing Ltd.

Carlin, J. B. and Louis, T. A. (2000). Bayes and Empirical Bayes. Methods for Data Analysis. (nd ed.). Chapman and Hall: New York.

Canter, D., Coffey, T., Huntley, M. and Missen, C. (2000). Predicting Serial Killers' Home Base Using a Decision Support System, Journal of Quantitative Criminology 16(4): 457–478.

Chen, H. and Lynch, K. J. (1992). Automatic Construction of Networks of Concepts Characterising Document Databases. IEEE Transactions on Systems Sept/Oct, 885–902.

Chen, H., Zeng, D., Atabakhsh, H., Wyzga, W. and Schroeder, J. (2003a). COPLINK Managing Law Enforcement Data and Knowledge, Communications of the ACM 46(1): 28–34.

Chen, H., Schroeder, J., Hauck, R., Ridgeway, L., Atabakhsh, H., Gupta, H., Boarman, C., Rasmussen, K. and Clements, A. (2003b). COPLINK Connect: Information and Knowledge Management for Law Enforcement. Decision Support Systems (DSS), Special Issue "Digital Government: technologies and practices' 34(3): 271–285.

Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., and Chau, M. (2004). Crime Data Mining: A General Framework and Some Examples. IEEE Computer 37(4).

Chklovski, T. (2003). Learner: A System for Acquiring Commonsense Knowledge by Analogy. In The Proceedings of the International Conference on Knowledge Capture. Florida, USA, October 23–25 2003, 4–12. ACM Press: New York, USA.

Clementine SPSS Clementine home page. [Online]. Available from: http://www.spss.com/clementine/ [Accessed 2004].

COPS. Community Oriented Policing Services home page. [Online]. Available from: http://www.cops.usdoj.gov [Accessed 2004].

Corcoran, J., Wilson, I. D., Lewis, O. M. and Ware, J. A. (2001). Data Clustering and Rule Abduction to Facilitate Crime Hot Spot Prediction, Lecturer Notes in Computer Science 2206: 807–822.

Corcoran, J., Wilson, I. D., and Ware J. A. (2003). Predicting the Geo-Temporal Variations of Crime and Disorder. International Journal of Forecasting 19(4): 623–634, Elsevier.

Cowell, R. G., Dawid, A. P., Lauritzen, S. L., and Spiegelhalter, D. J. (1999). Probabilistic Networks and Expert Systems. Springer-Verlag: New York.

Coxon, A. P. M. (1982). The Users' Guide to Multidimensional Scaling. Heinemann: London.

Crime and Disorder Act (1998). ISBN 0 10 543798 0.

Crimestat Manual (2004). Chapter 9 – Journey to Crime Estimation. [Online]. Available: http://www.icpsr.umich.edu/NACJD/crimestat.html#DOWNLOAD [2004, May 6].

Davies, P. M. and Coxon, A. P. M. (1982). Key Texts in Multidimensional Scaling. Heinemann: London.

Deltour, A., (2001). Tertius Technical Report CSTR-01-001 Department of Computer Science, University of Bristol, September 2001. [Online]. Available from: http://www.cs.bris.ac.uk/Publications/pub_info.jsp?id = 1000568 [Accessed 2004, May 20].

Dodd, T., Nicholas, S., Povey, D., and Walker, A. (2004). Home Office Statistical Bulletin, Crime in England and Wales 2003/2004. Research, Development and Statistics Directorate, Crown.

Dykes, J. A. and Mountain, D. M. (2003). Seeking Structure in Records of Spatio-Temporal Behaviour: Visualization Issues, Efforts and Applications, Computational Statistics and Data Analysis 43: 581–603.

Elias, P. (1995). Social Class and the Standard Occupational Classification. In Rose, D. (ed.), A Report on Phase 1 of the ESRC Review of Social Classifications. ESRC: Swindon.

Everitt, B. (1974). Cluster Analysis. Heinemann Educational books, Ltd: London.

Everitt, B. S. and Dunn, G. (1991). Applying Multivariate Data Analysis. Edward Arnold.

Ewart, B. W. and Oatley, G. C. (2003). Applying the Concept of Revictimization: Using Burglars' Behaviour to Predict Houses at Risk of Future Victimization. International Journal of Police Science and Management 5(2): 69–85.

Ewart B. W. and Oatley G. C. (2005) Dimensions of Burglary: A Disaggregated Approach. Paper presented to the 15th Conference of the European Association of Psychology and Law, June 30th–1st July, Vilnius, Lithunia.

Ewart, B. W., Inglis, P., Wilbert, M. N. and Hill, I. (1997). An Analysis of Time Intervals of Victimisation, Forensic Update 50: 4–9.

Ewart, B. W., Oatley, G. C. and Burn, K. (2005). Matching Crimes Using Burglars' Modus Operandi: A Test of Three Models, International Journal of Police Science and Management 7(3): 160–174.

Farrington, D. P. and Lambert, S. (2000) Statistical Approaches to Offender Profiling. In Canter D. and Alison L. (eds.), Profiling Property Crimes. Ashgate Publishing Co.

Fayyad, U. M. and Stolorz, P. (1997). Data Mining And KDD: Promises and Challenges, Future Generation Computer Systems 13: 99–115.

Fayyad, U. M, Piatetsky-Shapiro, G. and Smyth, P. (1996). The KDD Process for Extracting Useful Knowledge from Volumes of Data, *Communications ACM* 39(11): 27–41.

Gebhardt, F., et al. (1997). Reasoning with Complex Cases. Kluwer Academic Publishers: Boston, Massachusetts, USA.

Gigerenzer, G. (1991). How to Make Cognitive Illusion Disappear: Beyond Heuristics and Biases. In Stroebe, W. and Hewstone, M. (eds.), European Review of Social Psychology, Vol 2. John Wiley and Sons: London.

Goldberg, H. G. and Wong, R. W. H. (1998). Restructuring Transactional Data for Link Analysis in the FinCEN AI System. In Jensen, D. and Goldberg, H. (eds.), Artificial Intelligence and Link Analysis. Papers from the AAAI Fall Symposium. Orlando, Florida, Tech Report FS-98-01.

Green, E. J., Booth, C. E. and Biderman, M. D. (1976). Cluster Analysis of Burglary M/O's, Journal of Police Science and Administration 4: 382–388.

Gupta, K. M. and Montazemi, A. R. (1997). Empirical Evaluation of Retrieval in Case-Based Reasoning Systems Using Modified Cosine Matching Function, *Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans* 27(5): 601–612.

Han, J. and Kamber, M. (2001). Data Mining: Concepts and Techniques. Morgan Kaufmann: San Francisco, CA.

Hastie, T., Tibshirani, R. and Friedman, J. (2003). The Elements of Statistical Learning. Springer-Verlag: Canada.

Hauck, R., Atabakhsh, H., Onguasith, P., Gupta, H. and Chen, H. (2002). Using Coplink to Analyse Criminal-Justice Data, IEEE Computer 35: 30–37.

Hirschfield, A. (2001). Decision Support in Crime Prevention: Data Analysis, Policy Evaluation and GIS: In Hirschfield, A. and Bowers, K. (eds.), Mapping and Analysing Crime Data – Lessons from Research and Practice. Taylor and Francis London, New York, 237–269.

Horn, R. D, Birdwell, J. D. and Leedy, L. W. (1997). Link Discovery Tool. Proceedings of the Counterdrug Technology Assessment Center's ONDCP/CTAC International Symposium Chicago, IL. August 18–22, 1997.

HUGIN. Hugin Expert Home Page. [Online]. Available from: http://www.hugin.com/. [Accessed 2005].

Hunter, A. (2000). Feature Selection Using Probabilistic Neural Networks. In Neural Computing and Applications Vol. 9, 124–132. Springer-Verlag: London.

Hunter, A., Kennedy, L., Henry, J. and Ferguson, R. I. (2000). Applications of Neural Networks and Sensitivity Analysis to Improved Prediction of Trauma Survival, Computer Methods and Algorithms in Biomedicine 62: 11–19.

I2. Investigative Analysis Software home page. [Online]. Available from: http://www.i2.co.uk/Products/Analysts_Workstation/default.asp [Accessed 2004].

Jenson, F. V. (1996). An Introduction to Bayesian Networks. UCL Press.

Johnson, S. D. and Bowers, K. J. (2004). The Stability of Space-time Clusters of Burglary, The British Journal of Criminology 44(1): 55–65.

Kadane, J. B. and Schum, D. A. (1996). A Probabilistic Analysis of the Sacco and Vanzetti Evidence. John Wiley and Sons.

Kahneman, D. and Tversky, A. (1973). On the Psychology of Prediction, *Psychological Review* 80: 237–251.

Kleinbaum, D. G. (1996). Statistics in the Health Sciences: Survival Analysis. Springer-Verlag: New York.

Krause, P. and Clarke, D. (1993). Uncertain Reasoning: An Artificial Intelligence Approach. Intellect Books.

Leary, R. M. (2001). Evaluation of the Impact of the FLINTS Software System in West Midlands and Elsewhere. Home Office Policing & Reducing Crime Unit: Home Office, London.

Leary, R. M. (2002). The Role of the National Intelligence Model and ‘‘FLINTS’ in Improving Police Performance. http://www.homeoffice.gov.uk/docs2/.

Levine, N. (2002). CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 2.0). Ned Levine & Associates/National Institute of Justice: Houston, TX/ Washington, DC.

Leary, R. M. (2003). New Intelligence of the 21st Century: How Smart is it? Forensic Technology News November: 6, 2003.

Mena, J. (2003). Investigative Data Mining for Security and Criminal Detection. Butterworth–Heinemann, ISBN 0-7506-7613-2.

Merry, S. (2000).Crime Analysis: Principles for Analysing Everyday Serial Crime. In Canter, D. and Alison, L. (eds.), Profiling Property Crimes, 297–318. Ashgate Publishing Ltd.

Merry, S. and Harsent, L. (2000). Intruders, Pilferers, Raiders and Invaders: The Interpersonal Dimension of Burglary. In Canter, D. and Alison, L. (eds.), Profiling Property Crimes, 31–57. Ashgate Publishing Ltd.

Mooney, R. J., Melville, P., Rupert Tang, L.P., Shavlik, J., Dutra, I., Page, D., and Costa, V. S. (2004). Inductive Logic Programming for Link Discovery. In Kargupta, H., Joshi, A., Sivajumar, K., and Yesha, Y. (eds.), Data Mining: Next Generation Challenges and Future Directions. AAAI Press.

Newell, A. and Simon, H. A. (1972). Human Problem Solving. Prentice-Hall: Englewood Cliffs, NJ.

Noy, N. F., Grosso, W., and Musen, M. A. (2000). Knowledge-Acquisition Interfaces for Domain Experts: An Empirical Evaluation of Protege-2000. Twelfth International Conference on Software Engineering and Knowledge Engineering (SEKE2000), Chicago, IL.

Oatley, G. C. (2004). Case-based Reasoning (chapter). In Addison, D., and MacIntyre, J., (eds.), Intelligent Computing Techniques: A Review. Springer-Verlag, ISBN: 1-85233-585-8.

Oatley, G. C. and Ewart, B. W. (2002). Constructing a Bayesian Belief Network to determine the likelihood of burglary. In Proceedings of the Fifth International Conference on Forensic Statistics (ICFS5), Isola di San Servolo, Venice, Italy, August 30 – September 2, 2002.

Oatley, G. C. and Ewart, B. W. (2003). Crimes Analysis Software: ‘‘Pins in Maps’, Clustering and Bayes Net Prediction, Expert Systems with Applications 25(4): 569–588.

Oatley, G. C. and Ewart, B. W. (2005). The Meaning of Links. In Nelson, D., Stirk, S., Edwards, H., and McGarry, K. (eds.), Data Mining and Knowledge Discovery in Databases Workshop, 22nd British National Conference on Databases, Vol. 2, 68–76. Bncod 22, July 5–7, 2005, Proceedings (Lecture Notes in Computer Science), Springer-Verlag Berlin and Heidelberg GmbH & Co. K, Vol. 2, pp. 68–76.

Oatley, G. C., Tait, J., and MacIntyre, J. A. (1998). Case-Based Reasoning Tool for Vibration Analysis. In Milne, R., Macintosh, A., and Bramer, M. (eds.), Proceedings of the 18th Annual International Conference of the British Computer Specialist Group on Expert Systems (ES'98) – Applications and Innovations in Expert Systems VI, December 14–16, Springer, BCS Conference Series: Cambridge, 132–146.

Oatley, G. C., MacIntyre, J., Ewart, B. W. and Mugambi, E. (2002). SMART Software for Decision Makers KDD Experience, Knowledge Based Systems 15: 323–333.

Oskamp, S. (1965). Overconfidence in Case Study Judgements, Journal of Consulting Psychology 29: 261–265.

Pastra, K., Saggion, H., and Wilks, Y. (2003). Intelligent Indexing of Crime-Scene Photographs. IEEE Intelligent Systems, Special Issue on "Advances In Natural Language Processing' 18(1): 55–61.

Patterson, D. W. (1998). Artificial Neural Networks: Theory and Applications. Prentice Hall (Sd), ISBN: 0132953536.

Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers, Inc.

Pearl, J. (2000). Causality: Models, Reasoning, and Inference. Cambridge University Press: UK.

Pease, K. (1998). Repeat Victimisation: Taking Stock. Police Research Group, Crime Detection and Prevention Series, Paper 90. London, Crown Copyright.

Pease, K. (2001). What to do About it? Lets Turn off our minds and GIS: In Hirschfield, A. and Bowers, K. (eds.Mapping and Analysing Crime Data – Lessons from Research and Practice. Taylor and Francis London, New York, 225–237.

Peuquet, D. J. and Duan, N. (1995). An Event-based Spatiotemporal Data Model (ESTDM) for Temporal Analysis of Geographical Data, International Journal of Geographical Information Systems 9(1): 7–24.

Peuquet, D. J. and Wentz, E. (1994). An Approach for TimedBasedSpatial Analysis of Spatio-Temporal Data. Advances in GIS Research Proceedings Vol. 1, 489–504.

PITO (2004). PITO Core Data Model. [Online]. Available from: http://www.pito.org.uk/microsite/cordm/ [Accessed 2006, Jan 20].

Polvi, N., Looman, T., Humphries, C. and Pease, K. (1991). The Time Course of Repeat Burglary Victimization, British Journal of Criminology 31(4): 411–414.

Protg. Homepage of Protege. Available: http://www.protege.stanford.edu/ [2004, April 10].

Quinlan, R. (1998). C5.O: An Informal Tutorial. Rule Quest.

Raafat, H., Yang, Z. and Gauthier, D. (1994). Relational Spatial Topologies for Historical Geographical Informations, *International Journal of Geographical Information Systems* 8(8): 163–173.

Ratcliffe, J. H. (2002). Aoristic Signatures and the Spatio-temporal Analysis of High Volume Crime Patterns. Journal of Quantitative Criminology 18(1): 23–43.

Ratcliffe, J. H. A. and McCullagh, M. J. (2001). Crime, Repeat Victimisation and GIS: In Hirschfield, A. and Bowers, K. (eds.), Mapping and Analysing Crime Data – Lessons from Research and Practice. Taylor and Francis London, New York, 61–93.

Ribaux, O. and Margot, P. (1999). Inference Structures for Crime Analysis and Intelligence: The Example of Burglary Using Forensic Science Data, Forensic Science International 100: 193–210.

Ribaux, O. and Margot, P. (2003). Case Based Reasoning in Criminal Intelligence using Forensic Case Data, Science and Justice 43(3): 135–143.

Roiger, R. J. and Geatz, M. W. (2003). Data Mining, a Tutorial-Based Primer. Pearson Education, Inc: USA.

Rossmo, D. K. (1993). Multivariate Spatial Profiles as a Tool in Crime Investigation. In Block, C. R. and Dabdoub, M. (eds.), Proceedings of the Workshop on Crime Analysis through Computer Mapping. Illinois Criminal Justice Information Authority and Loyola University Sociology Department: Chicargo. Library of Congress HV7936.C88 W67.

Rossmo, D. K. (1995). Overview: Multivariate Spatial Profiles as a Tool in Crime Investigation: In Block, C. R., Dabdoub, M. and Fregly, S. (eds.), Crime Analysis through Computer Mapping. Police Executive Research Forum Washington, DC, 65–97.

Rossmo, D. K. (1997). Geographic Profiling. In Jackson, J. L. and Bekerian, D. A. (eds.), Offender Profiling – Theory, Research and Practice, 159–177. John Wiley and Sons.

Salton, G. and Buckley, C. (1988). Term Weighting Approaches in Automatic Text Retrieval, Information Processing and Management 24(5): 513–523.

Schum, D. A. (1994). The Evidential Foundation of Probabilistic Reasoning. John Wiley and Sons.

Schum, D. and Tillers, P. (1991). Marshalling Evidence for Adversary Litigation, 13 Cardozo Law Review 13: 657–704.

Sentient Sentient Informations Systems home page. [Online]. Available from: http://www.sentient.nl/ [Accessed 2004].

Soomro, T. R., Naqvi, M. R., and Zheng, K. (2001). GIS: A Weapon to Combat the Crime. In Proceedings of SCI 2001/ISAS 2001 (International Conference on Information Systems, Analysis and Synthesis), World Multiconference on Systemics, Cybernetics and Informatics, Information Systems Development: Part I.

Speckt, D. F. (1990). Probabilistic Neural Networks, Neural Networks 3(1): 109–118.

Stranieri, A. and Zeleznikow, J. (2000). Knowledge Discovery for Decision Support in Law, *ICIS* 2000: 635–639.

Tillers, P. and Schum, D. (1988). Charting New Territory in Judicial Proof: Beyond Wigmore, 9 Cardozo Law Review 907 (1988): 907–966.

Townsley, M. and Pease, K. (2002). How Efficiently can We Target Prolific Offenders?, International Journal of Police Science and Management 4(4): 323–331.

Trajan Homepage of Trajan Software [Online]. Available from: http://www.trajan-software.demon.co.uk [Accessed 2004, March 22].

Trickett, A., Osborn, D., Seymour, J. and Pease, K. (1992). What is Different About High Crime Areas?, British Journal of Criminology 32(1): 81–89.

Tversky, A. (1997). Features of Similarity, *Psychology Review* 84: 327–352.

Vargas, J. E. et al. (1998). Similarity-Based Reasoning about Diagnosis of Analogue Circuits. In Proceedings of the First International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems. Tennessee, USA, 1988, 83–86. ACM Press: New York, USA.

Wall, M. GAlib: A C+ + Library of Genetic Algorithm Components, version 2.4. Matthew Wall – Mechanical Engineering Department, Massachusetts Institute of Technology. http://www.lancet.mit.edu/ga, August 1996.

Walton, (2002). Legal Argumentation and Evidence. The Pennsylvania State University: University Park, Pennsylvania.

Weiss, S. M. and Indurkhya, N. (1998). Predictive Data Mining. Morgan Kaufmann Publishers: San Francisco, California, USA.

Wielemaker, J. (2003). An Overview of the SWI-Prolog Programming Environment: In Mesnard, F. and Serebenik, A. (eds.), Proceedings of the 13th International Workshop on Logic Programming Environments. Katholieke Universiteit Leuven Heverlee, Belgium, WLPE Report (CW371), 1–16.

Wielemaker, J., Schreiber, G., and Wielinga, B. (2003). Prolog-based Infrastructure for RDF: Performance and Scalability. In Fensel, D., Sycara, K., and Mylopoulos, J. (eds.), The Semantic Web – Proceedings ISWC'03, Sanibel Island, Florida, 644–658. Springer Verlag.

Wigmore, J. H. (1913). The Principles of Judicial Proof. Little Brown and Company: Boston, Massachussetts.

Williamson, D., McLafferty, S., McGuire, P., Ross, T., Mollenkopf, J., Goldsmith, V. and Quinn, S. (2001). Tools in the Spatial Analysis of Crime: In Hirschfield, A. and Bowers, K. (eds.), Mapping and Analysing Crime Data – Lessons from Research and Practice. Taylor and Francis London, New York, 187–203.

Wilson, D. C. (2001). Case-Base Maintenance: The Husbandry of Experience. Ph.D. thesis. Department of Computer Science, University of Indiana, USA.

Wilson, I. D., Corcoran, J., and Ware, J. A. (2002). Predicting the Geo-temporal Variations of Crime and Disorder. In Proceedings of the Sixth Annual International Crime Mapping Research Conference: Bridging the Gap Between Research and Practice, Denver, Colorado.

WinBugs. WinBUGS home page. [Online]. Available from: http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml [Accessed 2004].

Witten, I. H. and Frank, E. (1999). Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann, ISBN 1-55860-552-5.

Yokota, K. and Watanabe, S. (2002). Computer Based Retrieval of Suspects Using Similarity of Modus Operandi, International Journal of Police Science and Management 4(1): 5–15.

Zeleznikow, J. (2002). Designing Decision Support Systems for Crime Investigation. In Proceedings of the Fifth International Conference on Forensic Statistics (ICFS5), Isola di San Servolo, Venice, Italy, August 30 – September 2, 2002.