

Review of Current Crime Prediction Techniques

Vikas Grover¹, Richard Adderley² and Max Bramer¹

University of Portsmouth, UK¹

A E Solutions²

{Vikas.Grover, Max.Bramer}@port.ac.uk

Rick.Adderley@A-ESolutions.com

Abstract

Police analysts are required to unravel the complexities in data to assist operational personnel in arresting offenders and directing crime prevention strategies. However, the volume of crime that is being committed and the awareness of modern criminals make this a daunting task. The ability to analyse this amount of data with its inherent complexities without using computational support puts a strain on human resources. This paper examines the current techniques that are used to predict crime and criminality. Over time, these techniques have been refined and have achieved limited success. They are concentrated into three categories: statistical methods, these mainly relate to the journey to crime, age of offending and offending behaviour; techniques using geographical information systems that identify crime hot spots, repeat victimisation, crime attractors and crime generators; a miscellaneous group which includes machine learning techniques to identify patterns in criminal behaviour and studies involving re-offending. The majority of current techniques involve the prediction of either a single offender's criminality or a single crime type's next offence. These results are of only limited use in practical policing. It is our contention that Knowledge Discovery in Databases should be used on all crime types together with offender data, as a whole, to predict crime and criminality within a small geographical area of a police force.

1. Introduction

In recent years police forces have been enhancing their traditional method of crime reporting with new technological advancements to increase their output by efficiently recording crimes to aid their investigation (Adderley and Musgrove 1999). Data is not just a record of crimes, it also contains valuable information that could be used to link crime scenes based on the modus operandi (MO) of the offender(s), suggest which offenders may be responsible for the crime and also identify those offenders who work in teams (offender networks) etc.

It is not an easy task for a Police analyst to manually unravel the inherent complexities within police data and this problem is compounded when the analysis is undertaken by a team, as each member is not in possession of all relevant facts and relevant information could be lost. For a long time, criminologists and statisticians have been applying their skills and knowledge trying to predict when and where the next set of crimes will occur, with varying degrees of success. The volume of crime and the greater awareness of modern criminals put a strain on the existing methods. Human reasoning fails when presented with millions of records. Therefore, there is clearly a requirement for a tool kit to assist in analyzing the data which will make the best use of limited resources. Knowledge Discovery in

Databases (KDD) techniques can be used to reveal knowledge which is beyond intuition.

The aim of this study is to examine the current techniques used in crime prediction.

2. Crime Facts

We have categorised crimes into two types; Major crimes: Murders, rapes etc. and Volume crimes: Burglary, vehicle crime, robbery, theft, damage etc. The cost of volume crimes, both in financial and emotional terms, makes their effective targeting essential.

2.1 Crime Recording Process

Whenever a crime is committed, a police officer visits the crime scene or the report is taken by telephone, which is known as the crime report. All UK police forces record their crime reports in a similar way but in different computer systems. The variables stored may be known in a variety of ways but comprise the following: - Time, day and date of the crime; Offence type (there are in excess of 800 different Home Office crime codes); Location of crime to include post code and Ordnance Survey grid references; Victim information; Modus operandi (MO) identifies how the crime has been committed.

Depending upon the crime recording system used by each individual force the data fields will be a mixture of free text and structured (and often) validated fields. The free text may not even contain key words or phrases and will contain non standard abbreviations, mis-spellings and, on occasion, contradictory information. The quantity and quality of information recorded varies considerably from case to case. It is often imprecise, and is almost certainly at times inaccurate. Unstructured and inconsistent data formats make it very complicated to automate the analytical processes.

2.2 Environmental Criminology

Understanding Environmental Criminology and the behaviour of offenders plays a significant role in understanding and predicting crime and criminality (Brantingham and Brantingham 1991, Rhodes and Conly 1991, Cohen and Felson 1979; Clarke and Felson 1993, Adderley and Musgrove 2003).

3. Current Crime and Offending Prediction Techniques

The 'holy grail' in policing is to be able to predict when and where the next crime or set of crimes will occur. This, of course, in a holistic sense is not currently possible. Many attempts have been made in the crime prediction arena, each of which has had limited success. Most of these attempts have been either concerned with the crime and criminality relating to either a single offender or a single crime type. The sections below discuss the current crime prediction techniques.

3.1 Statistical Methods

Canter (1994) found that scene of a crime is a key feature to the address or home base of an offender which has been confirmed by Rossmo (2000) and there is a distance decay pattern for crime trips (Brantingham and Brantingham 1984).

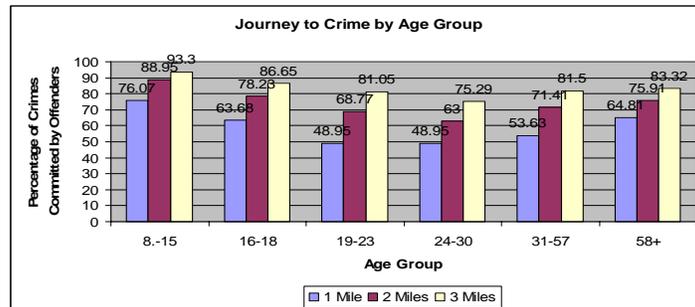


Figure 1: Distance travelled by age

Using West Midlands Police crime data, Figure 1, we have confirmed the statement mentioned in academic papers listed above. The left hand column of each age band clearly demonstrates that a large proportion of offenders commit their crime within one mile of their home address. The figures in the chart are accumulative.

If there is a lack of physical evidence, the behaviour of an offender has been used to suggest that the offender may be responsible for other undetected crime. In these instances the main emphasis is on the MO i.e. offender's 'way of working' (Davies 1992; Grubin et al 2001). The search for a common behavioural thread linking the offence of a single individual may be like searching for a needle in a haystack that is not there. Here data mining could help to find the links.

3.2 Miscellaneous Methods

Several academic ventures have made an abortive attempt to use artificial intelligence (AI) to identify volume crimes such as burglary (Lucas 1986 and Charles 1998). This has engendered suspicion within policing regarding the usefulness of AI techniques. There are only few examples of using data mining on crime data (see Adderley 2004; Brown 1998).

There are several other techniques that have been used by several researchers. Chau et al (2002) has used entity extraction to discover the patterns that identify person names, their addresses, vehicles and other characteristics. Some of the approaches, such as string comparator, social network analysis and deviation detection are described in Chen et al (2004) to use on crime data in understanding criminal behaviour. Hauk et al (2002) have used the concept space algorithm on crime data to detect abnormal activities. Once these activities are identified it may be possible to predict the next occurrence of such activity. Several algorithms have been used by Oatley et al (2004) to match and link burglary crimes together into a crime series. Having ascertained that a series is occurring it is possible to suggest, from that data, where the next crime in that series will occur. Repeat victimisation has also been used to assist the prediction (Ewart and Oatley 2003).

3.3 Geographical Information System Methods

(Brantingham & Brantingham 1995) propose that crime hot spots are developed in areas of the community that can be labelled as crime generators, such as entertainment areas and shopping malls. The techniques used by police forces to identify hot-spots are not always consistent. Crime problems in areas designated as

hot-spots may be momentary, and may disappear before resources are officially allocated to those areas. Other than being a crime generator, there are a variety of reasons why a particular geographical area is regarded as a hot-spot. For example, the crime rate could be caused by a prolific offender being released from prison or due to a particular community event occurring. These hot-spots can be used as good predictors of crime and criminality.

Residential burglary has been the focus of geographical research. Johnson and Bowers, (2004) suggests that burglary offences cluster in space and time in affluent areas, "a burglary event is a predictor of significantly elevated rates of burglary offences within 1-2 months and within a range of up to 300-400 metres of a burgled home" (Johnson and Bowers 2004).

4. Discussion

Current prediction techniques have had limited success in operational policing. Many researches have spent time analysing large amounts of police related data with a view to predicting either where the next crime or set of crimes will occur. There are two main areas where these prediction techniques have been concentrated: - An individual crime type and an offender's movements. There are two subsidiary areas where these techniques have been concentrated; Repeat victimisation and Hot-spot analysis.

In our view by limiting the research to a single crime type or offender or geographical area the ability to predict has limited value to operational policing. In order to effectively prevent crime and arrest offenders, it is necessary to effectively target the geographical area where crimes are occurring or will occur. This is ALL crime and not just an individual crime type. Therefore, we suggest that KDD could be and should be used on all crime types together with offender data, as a whole, to predict crime and criminality.

References

- Adderley, R. (2004) The Use of Data Mining Techniques in Operational Crime Fighting, *Intelligence and Security Informatics, Second Symposium on Intelligence and Security Informatics*. Springer, ISBN: 3-540-22125-5
- Adderley, R., and Musgrove, P.B., (1999) Data Mining at the West Midlands Police: A Study of Bogus Official Burglaries, *BCS Specialist Group on Expert Systems, ES99*, London, Springer – Verlag, pp191-203, 1999.
- Adderley, R., and Musgrove, P.B., (2003) Modus operandi modeling of group offending: a data mining case study, Accepted by: *The International Journal of Police Science and Management*, 2003.
- Brantingham, P., & Brantingham, P., (1984) *Patterns in crime*. New York: Macmillan.
- Brantingham, P., & Brantingham, P., (1991), Notes on the geometry of crime, in *Environmental Criminology*, USA: Wavelend Press Inc.
- Brantingham, P, & Brantingham, P., (1995) Criminality of place: Crime generators and crime attractors. *European Journal on Criminal Policy and Research* 3,3, special issue on Crime Environments and Situational Prevention, 5-21.

- Brown, D.E. (1998) The Regional Crime Analysis Program (RECAP): A Framework for Mining Data to Catch Criminals. in *IEEE International Conference: Systems Man and Cybernetics Society*.
- Canter, D.V. (1994) *Criminal Shadows* London: Harper Collins.
- Charles, J., (1998) AI and Law Enforcement, *IEEE Intelligent Systems*, pp77-80.
- Chau, M., Xu, J., and Chen, H (2002) Extracting Meaningful Entities from Police Narrative Reports. In: Proceedings of the National Conference for Digital Government Research (dg.o 2002), Los Angeles, California, USA.
- Chen, H., Chung, W., Xu, J. J, Qin, G. W. Y, and Chau, M (2004), Crime Data Mining: A General Framework and Some Examples. *IEEE Computer Society*. 50-56.
- Clarke, R.V., & Felson M. (1993), Introduction: Criminology, Routine activity, and rational choice in *Routine activity and rational choice: Advances in criminological theory, volume 5*. Clarke, R.V., Felson, M. (eds.) New Jersey, USA: Transaction Publishers.
- Cohen, L.E. and Felson, M., (1979), Social Change and Crime Rate Trends: A Routine Activity Approach. *American Sociological Review*, Vol 44, 588-608.
- Davies, A. (1992) Rapists Behaviour: A three aspect model as a basis for analysis and the identification of serial crime. *Forensic Science International*, 55, 173-194.
- 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 Strategies and Management, Vol.5 (2)*.
- Grubin, D., Kelly, P., and Brunson, C. (2001) Linking serious sexual assaults through behaviour. Home Office Research Study 215. ISBN 1-84082-560-X
- Hauk, R.V., Atabaksh, H., Ongvasith, P., Gupta, H., and Chen, H. (2002) Using Coplink to analyze criminal justice data, *IEEE Computer*, 35(3), pp. 30-37.
- Johnson, S.H. and Bowers, K.J. (2004) The Burglary as Clue to the Future: The Beginnings of Prospective Hot-Spotting, *European Journal of Criminology*, Vol 1 (2): 237-255: 1477-3708.
- Lucas, R. (1986) An Expert System to Detect Burglars using a Logic Language and a Relational Database, *5th British National Conference on Databases, Canterbury*.
- Oatley, G.C., Zeleznikow, J., and Ewart, B.W., (2004), Matching and Predicting Crimes. In: Macintosh, A., Ellis, R. and Allen, T. (eds.), Applications and Innovations in Intelligent Systems XII. Proceedings of AI2004, *The Twenty-fourth SGAI International Conference on Knowledge Based Systems and Applications of Artificial Intelligence*, Springer: 19-32. ISBN 1-85233-908-X
- Rhodes, W.M., Conly, C., (1991), *The criminal commute: A theoretical perspective in Environmental Criminology*, USA: Waveland Press Inc.
- Rossmo, D. K (2000) *Geographic profiling*: CRC Press. ISBN 0-8493-8129-0. pp 97-110.

An extended version of this paper is available as Grover, V., Adderley, R. and Bramer, M. (2006). A Review of Current Crime Prediction Techniques. University of Portsmouth, School of Computing, Technical Report.