LASSO: LINKAGE ANALYSIS OF SERIOUS SEXUAL OFFENCES

A DECISION SUPPORT SYSTEM FOR CRIME ANALYSTS AND INVESTIGATORS

By

DON CASEY, Sgt. Metropolitan Police Service, London
& London South Bank University

PHILLIP BURRELL , Professor of Intelligent Decision Systems,
London South Bank University

Abstract

One of the first and most important considerations when investigating a serious sexual offence is to see if the offence can be linked to others. If a link can be established then there is a very considerable dividend in terms of new evidence and lines of enquiry to be followed. It also raises what is already a serious incident to a higher level of significance with a corresponding increase in the resources allotted to investigation of the series of crimes. Computerised decision support systems which employ techniques from Artificial Intelligence are widely used in business and finance to assist practitioners in arriving at justifiable conclusions. In principle this is no different from the activities of a crime analyst or investigator in finding likely matches for a current crime in the overall set of crimes.

Aims of the Study

The aim of the study is to develop a computerised decision support system that can be used by crime analysts and investigators to suggest links between stranger rapes. It is intended that the characteristics of the crime under investigation can be entered into an easy to use computer interface and that the system will then be able to search its database of existing crimes and display a number of offences that have strong similarities.

The desirability of developing computer–based tools for linkage analysis has been recognised by the leading researcher into linking serious sexual offences:

> *"The ultimate goal is to create a computer-based screening system that will allow routine and systematic comparison of serious offences on a national basis, selecting cases on the basis of their behavioural similarity that are appropriate for more detailed attention by detectives or crime analyst"*
>
> Grubin (2000)

This viewpoint is acknowledged throughout the literature and it is recognised that the construction of a linkage tool is the necessary condition to progress this undertaking.

> "The development and test of theories and implementation of findings into computer-based, decision-support systems … has to be the proper basis for any professional derivation of inferences about offenders."
>
> Canter (2000)

It is interesting that here Canter widens the scope of computerised systems to include the possibility of inferring offender characteristics, the process known more widely as 'offender profiling' and the subject of a great deal of crime literature and Hollywood output.

Currently there are two computer systems that dominate the area of crime linkage: ViCAP – the Violent Crime Apprehension Program and ViCLAS, Violent Crime Linkage System. ViCAP is the creation of the FBI at Quantico and has been in existence in differing forms since 1985. ViCLAS was developed by the Royal Canadian Mounted Police in the early 1990's as an extension to the earlier system; The RCMP license ViCLAS, for a fee, and maintain control over it ; it is used in many EU jurisdictions : Belgium, Czech Republic, France, Germany, Ireland, Netherland and the United Kingdom. ViCLAS is also used in Australia, New Zealand, Switzerland and some states in the US.

Both these systems were developed primarily by practitioners, psychologists and criminologists and are essentially repositories of data which depend very much on the skill, training and experience of the user. The influence of Computer Scientists has been slight and there has been no involvement by researchers in A.I or Decision Support. As a result none of the advances that have been made in these areas are incorporated in either system and they remain essentially unchanged in the last 20 – 25 years.

There is a notable disparity between the amount of effort that police agencies invest in gathering and recording information that relates to these serious offences and the amount which has been expended on developing the computer systems onto which it is entered. ViCAP and ViCLAS are passive; the work spent in filling the database is not reciprocated by any corresponding functionality in the system. An effective crime linkage decision support system should and can assist the user in investigating the crime by using effective computer science technology to recommend answers to the questions : 'Which crimes are similar to this?' , ' How strong is the similarity?' and 'What are the factors that are most similar and most dissimilar between this set of crimes'.

Methodology

Fuzzy set theory (Zadeh 1965) is a well established approach in the field of Artificial Intelligence that can deal with imprecise or vague concepts such as 'young', 'old', 'tall', 'short' etc. These descriptions are defined as 'fuzzy sets', i.e. they are not specifications which have a yes or no answer. So a suspect described as 1.80m in height does not have to be either 'tall' or 'short' but can be accorded a degree of both qualities; in this case he could be 0.9 tall and 0.1 short. Or a person 35 years old could be 0.3 'young', 0.6 'middle-aged' and 0.1 'old'. This type of characterisation sits well with our own perceptions of what are known as linguistic variables in fuzzy set theory and give a richer picture of what we seek to describe. Most of all they allot meaningful numbers to the types of descriptions which we deal with in defining crimes.

The result is that a common description of a crime such as: *"A very violent attack on a middle-aged woman by a young man"* can be represented by a number of co-ordinates so that the degree of violence, middle-age and youth can be compared with other crimes. Consequently crimes and criminals can be described in highly descriptive terms and procedures to discover what the most significant differentiating features are, using mathematically and logically sound methods, can be undertaken. We have been fortunate in being successful in obtaining data on 545 serious sexual offences from the Serious Crimes Analysis Section of the U.K National Police Improvement Agency. We have excluded those offences that do not relate to serial rapes, by which we mean a set of rapes committed by a single individual, resulting in a much narrower dataset (n =110, development set n =83, test = 27) .

As a starting point we have adopted the dimensions identified as significant in the research undertaken by Grubin et al (2000) in linking serious sexual offences through behaviour: Sex, Control and Escape. Sex comprises 13 variables relating to the sexual assault; Control has 20 variables that define the controlling actions that the offender uses to subdue the victim and Escape (11 variables) includes those actions that the attacker adopts to ensure that he leaves a minimum of evidence at the scene , e.g. binding and blindfolding.

Results

The consequences of assigning a single set of numbers to a crime are far reaching. A great number of techniques can be employed to represent similarity between crimes and also to look for clusters of crimes. The fuzzy c-means algorithm (Bezdek 1981) looks for clusters in data and allows the user to specify the number of input dimensions and output. Table 1 shows the results where the three input dimensions of 'Sex','Control' and 'Escape have been input and three clusters specified. The double lines indicate series boundaries, i.e. crimes committed by the same offender, so it can be seen that three of the four crimes in the first series belong entirely in cluster 'C' while both crimes in the last series have very high memberships of cluster 'B'

Overall 88% of crimes were assigned to a cluster with > 80% degree of membership and 15 of the 28 series were assigned to a single cluster at 80% membership or more.
This demonstrates a far greater degree of consistency within series than the Grubin study which is the only comparable research in this area.

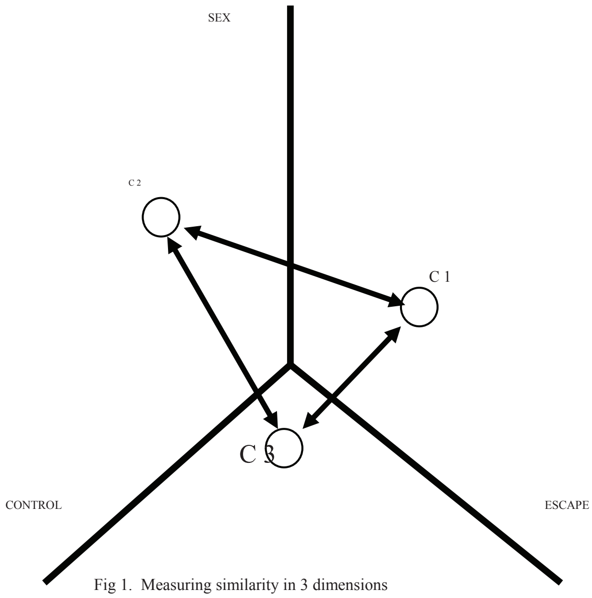

Fig 1. Measuring similarity in 3 dimensions

In Table 2 we've described the average distance between crimes as the value 'close' and then measured the degree of closeness between each crime in the dataset. The first three series are shown comprising fourteen crimes of length five, three and six offences; degrees of closeness greater than 0.6 are in bold.

There is a strong degree of closeness (> 0.6) between four crimes in the series 1 to 5 and all of the crimes in the second series, 6 to 8. The associations in the third series are less successful, but useful associations do exist. For instance crime 12 only has only two strong links but they are both with crime in the same series, 11 and 14. The result is to develop a structured search strategy for analysts and investigators from one crime to those other crimes that are likely be linked to it.

	A	B	C
c28	0.00	0.00	**1.00**
c29	0.00	0.00	**1.00**
c30	0.00	0.00	**1.00**
c31	0.00	0.39	0.60
c32	0.04	0.01	**0.95**
c33	0.00	0.00	**1.00**
c34	0.45	0.01	0.54
c35	0.00	0.00	**1.00**
c36	0.00	**1.00**	0.00
c37	0.00	**0.99**	0.01
c38	0.02	**0.98**	0.00
c39	0.03	**0.97**	0.01

table 1. crime membership 3 clusters

Conclusion

The need for a computerised decision support system to assist in linking serious crimes has been identified and the current systems in use shown to be inadequate. Established techniques from Artificial Intelligence, in particular fuzzy set theory, can be applied to crime linkage and have been shown to produce promising results. This could be further developed to set the area on a sound theoretical base and introduce the possibility of profiling offenders by discovering similar offender characteristics in linked crimes.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1.00	**0.87**	**0.75**	0.45	**0.80**	**0.69**	**0.80**	**0.60**	**0.76**	**0.81**	0.13	0.26	0.51	0.47
2	**0.87**	1.00	**0.76**	0.35	**0.79**	**0.73**	**0.86**	**0.64**	**0.79**	**0.78**	0.12	0.23	0.48	0.41
3	**0.75**	**0.76**	1.00	0.45	**0.64**	0.58	**0.76**	0.41	**0.97**	**0.71**	0.33	0.44	0.45	**0.59**
4	0.45	0.35	0.45	1.00	0.32	0.19	0.31	0.08	0.43	0.39	0.21	0.35	0.41	**0.60**
5	**0.80**	**0.79**	**0.64**	0.32	1.00	**0.87**	**0.85**	**0.68**	**0.67**	**0.89**	0.11	0.23	0.32	0.41
6	**0.69**	**0.73**	0.58	0.19	**0.87**	1.00	**0.82**	**0.69**	**0.61**	**0.80**	0.08	0.19	0.22	0.33
7	**0.80**	**0.86**	**0.76**	0.31	**0.85**	**0.82**	1.00	**0.61**	**0.79**	**0.85**	0.19	0.30	0.35	0.45
8	**0.60**	**0.64**	0.41	0.08	**0.68**	**0.69**	**0.61**	1.00	0.43	0.57	0.00	0.00	0.29	0.10
9	**0.76**	**0.79**	**0.97**	0.43	**0.67**	**0.61**	**0.79**	0.43	1.00	**0.74**	0.32	0.43	0.44	0.58
10	**0.81**	**0.78**	**0.71**	0.39	**0.89**	**0.80**	**0.85**	0.57	**0.74**	1.00	0.21	0.34	0.32	0.52
11	0.13	0.12	0.33	0.21	0.11	0.08	0.19	0.00	0.32	0.21	1.00	**0.84**	0.00	0.58
12	0.26	0.23	0.44	0.35	0.23	0.19	0.30	0.00	0.43	0.34	**0.84**	1.00	0.00	**0.73**
13	0.51	0.48	0.45	0.41	0.32	0.22	0.35	0.29	0.44	0.32	0.00	0.00	1.00	0.21
14	0.47	0.41	**0.59**	**0.60**	0.41	0.33	0.45	0.10	0.58	0.52	0.58	**0.73**	0.21	1.00

table 2. Degree of 'closeness' between crimes

References

Bezdek J.C (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York, Plenum.

Canter D (2000). Offender profiling and criminal differentiation. *Legal and Criminological Psychology* **5**: 23-46.

Canter D.V, Bennell C, et al. (2003). Differentiating Sex Offences. *Behavioural Sciences and Law* **21**.

Grubin D, Kelly P, et al. (2000). Linking Serious Sexual Assault through Behaviour *Home Office Research Study 215. London.*

Zadeh L (1965). Fuzzy Sets. *Information and Control*(8): 228-353.

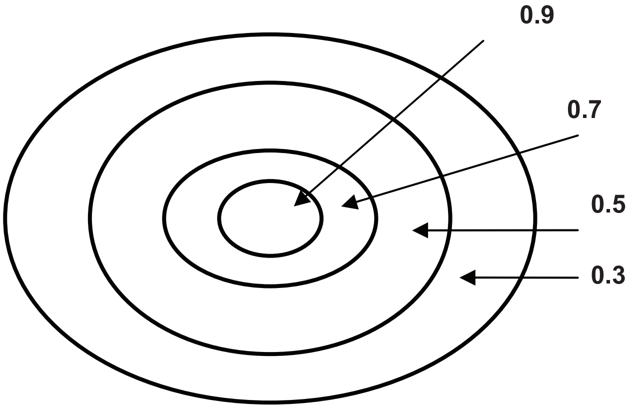

Fig 2 Closeness to the index crime