Problems of Classification in Investigative Psychology

Paul J. Taylor, Craig Bennell, and Brent Snook

Department of Psychology, University of Liverpool,
Eleanor Rathbone Building , L69 7ZA Liverpool,
United Kingdom, pjtaylor@liv.ac.uk

Abstract. Problems of classification in the field of Investigative Psychology are defined and examples of each problem class are introduced. The problems addressed are behavioural differentiation, discrimination among alternatives, and prioritisation of investigative options. Contemporary solutions to these problems are presented that use smallest space analysis, receiver operating characteristic analysis, and probability functions.

1 Introduction

Investigative Psychology (IP) is a growing discipline that studies the complex interaction between offender, victim and environment with the purpose of developing models of behaviour that can be used to provide actuarial support to police inquiries. Research in this area is based on the assumption that psychologically important information about an offender may be acquired by analysing and interpreting the patterns of behaviour that emerge during their criminal activity. This premise means that many questions in IP are essentially problems of classifying the differences in offender behaviour and offender information, as well as the correspondence between these two domains. Table 1 presents a framework for conceptualising the central problems of classification in IP, and gives examples of these problems to introduce the reader to the types of data available for analyses. Problems differ in form according to whether the data are individually assigned a score from a meaningful common measurement range (Ordering), or transformed to allow partitioning into a number of defining classes (Grouping). Problems differ in reference depending on whether data is classified through an independent theory-driven interpretation of content (Unsupervised) or with respect to some measured external criterion (Supervised). These distinctions combine to form three kinds of classification problems, unsupervised differentiation of subgroups, supervised discrimination among alternatives, and a more general supervised/unsupervised prioritisation of investigative options. Boundaries among these three kinds of problems are not mutually exclusive, and experience has shown that categories identified through differentiation are often the stimulus for hypotheses about effective discriminators, and findings from both these problems are precursors to prioritisation.

<table>
<tr><td colspan="2"></td><td colspan="2">Subject of Classification</td></tr>
<tr><td colspan="2"></td><td>Unsupervised</td><td>Supervised</td></tr>
<tr><td rowspan="8">Form of Classification</td><td rowspan="4">Grouping</td><td>Type I: Differentiation</td><td>Type II: Discrimination</td></tr>
<tr><td>Patterns of offender behaviour</td><td>Linking crimes</td></tr>
<tr><td>Types of offence</td><td>Statement validity</td></tr>
<tr><td>Types of interviewing</td><td>Threat credibility assessment</td></tr>
<tr><td rowspan="4">Ordering</td><td colspan="2">Type III: Prioritisation</td></tr>
<tr><td>Enquiry (resource) management</td><td>Geographical profiling</td></tr>
<tr><td>Filtering calls to service</td><td>Risk assessment</td></tr>
<tr><td>Targets for war against terrorism</td><td>Suspect prioritisation</td></tr>
</table>

Table 1. The three kinds of classification problems in Investigative Psychology.

The remainder of this paper presents examples of how current researchers are attempting to derive adequate solutions to each kind of classification problem. As is typical in IP, the solutions offered draw on disparate methods, but common to each is an emphasis on making minimal assumptions of the data. Such intrinsic analyses [7] are essential to IP because the data, collected from police investigations, are often incomplete, ambiguous, and inconsistent across cases, thus making the assumptions of many conventional methods misleading and unrealistic.

2 Behavioural Differentiation

IP was originally motivated by the problem of classifying the variety of ways in which an offender interacts with a victim during an offence. Research has sought to effectively differentiate a number of subgroups of highly co-occurring offence behaviours, where the number of groups is hypothesised from theoretical explanations of criminal motivation. Any support for such grouping substantiates the proposal that offenders bring different interpersonal styles to criminal activity. Aside from theoretical interest in individual differences, an effective classification of offence behaviour is valuable to investigative research because it is the necessary first step in linking variation in crimes to differences in the people who commit them. The challenge, then, is to take information regarding the occurrence (or not) of a set of B behaviours (e.g., hitting, threatening) over N offences and rearrange the resulting dichotomous matrix B x N so that classes of behaviours can be partitioned into psychologically meaningful groups.

Although many techniques are available for analysing the structure of a B x N matrix, the currently favoured method in IP is non-metric multidimensional scaling, and in particular Smallest Space Analysis (SSA-I). The output of SSA-I represents the rank order of correlations among behavioural variables as distances between their representative points in a geometric configuration, such that the further apart two behaviours on the plot the less likely it is that they were both used by the same offender. Since the SSA-I configuration represents the interrelationships among behaviours in a single solution space and without reference to any arbitrary dimension or cluster [7], the dominant groups of criminal actions may be identified through a more relaxed criterion; coherent regions of behaviours with substantively similar meanings.. The utility of this regional approach is made evident by the growing body of publications which, despite the potential high degree of variation between offenders and offences, have been remarkably consistent in finding a three-fold circular radex pattern to the interrelationships among crime scene behaviours, both in serious (e.g., homicide and arson) and mass (e.g., burglary and juvenile delinquency) crimes [2].

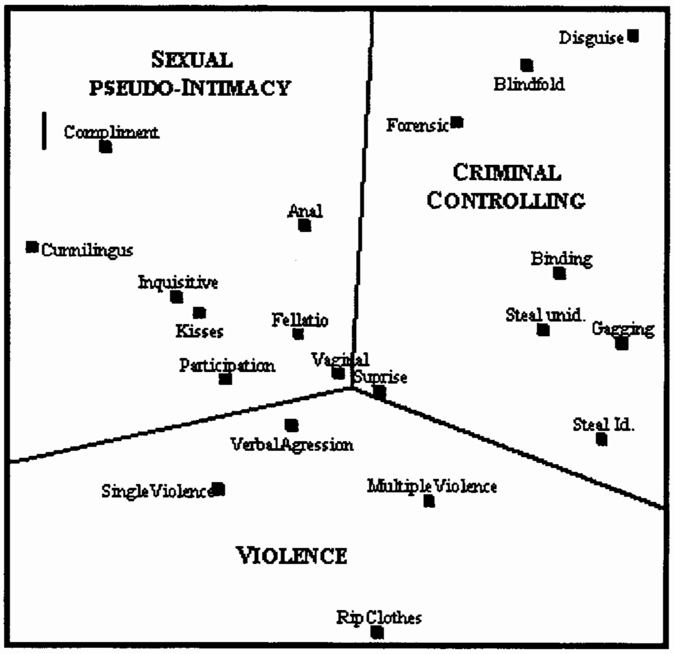

Fig. 1. SSA-I of 20 behaviours in 188 rape offences with regional interpretations. Coefficient of alienation = 0.15 in 15 iterations.

As an example of this approach to differentiating criminal behaviour, Figure 1 shows a 2-dimensional SSA-I configuration of 20 behaviours committed in 188 rape offences. The labels associated with each point correspond to one of the 20 behaviours defined in a content dictionary available from the authors. By drawing on previous research and theory to interpret the configuration, support was found for a regional classification of offenders behaviour into three themes: Sexual-Pseudo-Intimacy, Criminal-Controlling, and Violence. Evidence for these three interaction themes replicates the predominant

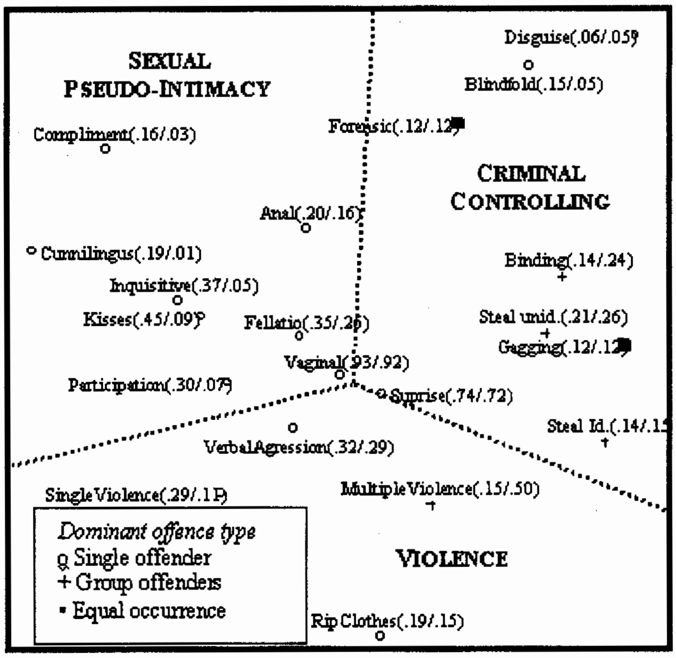

Fig. 2. SSA-I of rape offences separated into 112 single offender and 76 group rape offences. Numbers in parentheses denote proportion of occurrence for single offences / gang offences.

findings of previous studies [2], lending support to the proposal that similar behavioural themes may differentiate offender behaviour across many crimes.

The full potential of the SSA-I model emerges when the relationship between behaviours and external aspects of the data are explored. For example, of the 188 rape offences used in the SSA-I shown in Figure 1, 112 were committed by single offenders while the remaining 76 were gang rapes. Given the potential for differences between single offender and gang behaviour, Figure 2 shows the identical SSA-I plot where each behaviour is labelled according to whether it occurred relatively more often within single or group offences. As is shown in Figure 2, all of the behaviours that occurred more frequently in gang offences are situated towards the right side of the SSA-I space. This provides some initial evidence to suggest that group offences are predominantly controlled attacks on individuals that are not obviously motivated by sexual intent but often do involve extreme use of violence.

3 Discrimination

A second type of classification problem in IP requires a decision-maker to discriminate between two alternatives. Such situations exemplify a highly structured classification problem in which two possible outcomes are correct (hits and correct rejections) and occur when decisions correspond with reality, while two other outcomes are incorrect (false alarms and misses) and occur when decisions do not correspond with reality. Making a correct decision is difficult because evidence is typically ambiguous and can often arise for both alternatives [8]. One goal of IP, therefore, is to identify methods that improve the decision-makers ability to identify unambiguous discriminators and set appropriate decision thresholds that define how much evidence needs to be present before particular decisions are made.

One such task involves determining whether two crimes were committed by the same offender based only on information from the crime scene [4]. Previous approaches to the problem have had limited success because they do not easily allow decision-makers to evaluate the discriminatory power of various forms of behavioural evidence, or to assess how different decision thresholds affect discrimination accuracy. More recently, receiver operating characteristic (ROC) analysis has been proposed as a potential method for linking crimes [1]. Given information about a set of behavioural evidence over a number of crimes, the ROC approach begins by deriving the probability that any crime pair is linked for each item of evidence. Multiple decision thresholds are then set along this continuum of probabilities for each item, whereby any crime pair receiving a probability score above the specified threshold are classified as linked. One can then calculate the conditional probabilities of hits (pH) misses (pM), correct rejections (pCR) and false alarms (pFA) from their respective frequencies (e.g., pH=hits/(hits+misses)) and plot these probabilities on a graph as a function of the different decision thresholds [8]. The height of the resulting ROC curve relates to the ambiguity of the evidence used, such that the proportion of area lying beneath the curve (A) can be used as a measure of overall discrimination accuracy (perfect discrimination = 1.0, chance discrimination = 0.5). Threshold-specific accuracy measures can then be obtained by examining single points along the ROC curve.

In order to provide an empirical example of the ROC approach, behavioural information from 133 solved burglaries committed by 29 offenders was collected from a UK police force. The level of similarity between every combination of two crimes was calculated as a function of a variety of behavioural variables, including across crime distances, entry behaviours, internal behaviour, and property stolen. These 'discriminator' measures were then used in logistic regression models to calculate the probability that each crime pair was linked, and ROC curves were constructed from these probabilities using the previously described procedure. As the ROC curves in Figure 3 indicate, across crime distances are the most effective feature for classifying burglaries as linked vs. unlinked (A=0.84), though effectiveness depends largely on the threshold adopted. The practical significance of such a finding is clear when one considers how many more hits (or how many less false alarms) will be made at a particular decision threshold depending upon the evidence used for the task [8]. For example, a police force may decide, based on an evaluation of their available resources, that an appropriate decision threshold is one that prevents them from exceeding pFA=0.30. At pFA=0.30 an investigator would correctly identify 34 more linked crime pairs for every 100 crime pairs if across-crime distances were drawn on instead of entry behaviours (i.e., the difference between pH=0.82 and pH=0.48).

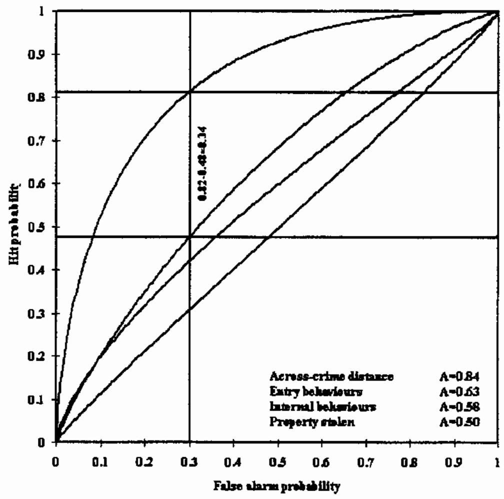

Fig. 3. Using ROC curves to classify burglaries as linked vs. unlinked.

4 Prioritisation

The third classificatory goal of IP is to identify how police investigations can make better use of resources. This goal is accomplished through some form of ordering or weighting of options, where options with a higher prioritisation are considered more likely to yield investigative success. Prioritisation is an intriguing problem that retains individual observations as the object of classification (i.e., each observation is a separate class) and instead requires modelling of the inter-relationships among these observations. This task is manageable in many investigative problems because observations must only be assigned to positions along a single measurement scale; typically the probability of detecting the offender.

One prescient example of this probabilistic modelling comes from research developing profiling systems that aim to classify geographic areas according to how likely they are to contain the offender's residence. The solution to area prioritisation begins with the assumption that crime site locations are an observed subset of an offender's overall spatial activity, and that the 'activity space' within these offences contains the offender's residence. Based on data of the x and y co-ordinates of offence locations, this activity space is classified by superimposing a grid (defined by the minimum and maximum x and y co-ordinates), and then applying a probability distance function around each of the crime locations to assign cells of the grid a probability value. Given that the accuracy of grid classification rests on the effectiveness of the prescribed function, research aimed at improving geographical profiling systems has focused on comparing the accuracy of different functions.

One popular system known as Dragnet [6] uses a function that assumes that the probability of an offender residing in a particular area decreases with increasing distance from an offence. Prioritisation is carried out by using one function from a family of negative exponential decay functions, defined as:

$$f(d_{ij}) = a \cdot e^{-c \cdot d_{ij}} \quad (1)$$

where $f(d_{ij})$ is the likelihood that the offender will reside at a particular location, d_{ij} is the distance from the centre of the grid cell i to an offence j, /emphe is a coefficient chosen from a basic understanding of decay functions to indicate of the maximum likelihood of finding a home, and c is an exponent, either arbitrarily chosen or predetermined using data from solved cases, that determines the steepness of the function [5]. A second alternative system called Criminal Geographic Targeting [3] makes the additional assumption that there is an area around each offence where the offender is less likely to live (a buffer zone). This initial increase in probability is modelled using a spline function that starts at zero likelihood of locating the offender's home and increases to a peak likelihood (defined by user), at which point the function follows the negative exponential function given in equation 1. This additional linear part of the function is defined as:

$$f(d_{ij}) = g + b \cdot d_{ij} \quad (2)$$

where g represents the probability of an offence site also being the offender's residence (typically zero), and b is the positive constant that defines the slop of the linear function, and hence how likelihood increases with distance away from the crime site [5]. The left panel of Figure 4 shows the probability curves resulting from these two approaches. In applying these functions around each crime site location, a degree of probability overlap will occur that can be summed to produce an overall probability for each grid cell. The resulting probability surface provides investigators with a method to order the classified areas within the offender's activity space in terms of the likelihood that the area contains the offender's residence.

As a simple examination of the effectiveness of these functions, a measure known as error distance was calculated for a sample of 68 German serial murder series. Error distance is the distance between the area within the offender's activity space that is classified as most likely to contain the offender's residence and the area that actually contains the offenders residence. The

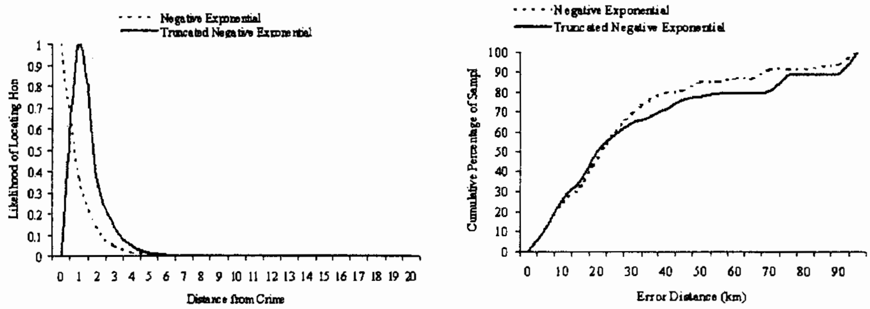

Fig. 4. Two probability distance functions and their relationship to error distance.

right panel of Figure 4 portrays the effectiveness of the functions by plotting the percentage of sample correctly located as a function of error distance. A steeper curve indicates that home locations were, on average, closer to the point of highest probability and that, consequently, the probability distance function was more efficient. Functions were compared by examining the difference in error distance across all 68 series. This difference was non-significant (Negative exponential: median = 22 km, standard deviation = 119; Truncated negative exponential: median = 18 km, standard deviation = 122 km; Wilcoxon Z = -0.283, *ns*).

5 Conclusions

In addition to answering a variety of theoretical questions, the value of IP rests on the extent to which solutions developed from research can be integrated into effective decision support tools. Before this is done, however, it would be useful for others within the field of classification research to provide solutions to the problems identified above through other techniques. Such comparisons across methodologies will allow research to maximise model effectiveness and so improve the accuracy of inferences drawn from criminal behaviour.

6 Acknowledgements

PT, CB and BS are Ph.D. candidates in the Department of Psychology at the University of Liverpool. The sections of this paper dealing with discrimination and prioritisation are based in part on dissertations in progress by CB and BS respectively, while PT should be contacted for information regarding behavioural differentiation. The authors thank Louise Porter for providing rape data and Christian Setzkorn for software development and LaTeX formatting. Parts of this research were supported by Overseas Research Student Scholarships awarded to CB and BS.

References

1.Craig Bennell. *Using ROC analysis to link serial crimes.* Liverpool - UK, 2001. Paper presented at the Merseyside Police conference on investigative strategies for domestic burglary.

2.David Canter. Offender profiling and criminal differentiation. *Legal and Criminological Psychology*, 5:23–46, 2000.

3.David Canter, Toby Coffey, Malcolm Huntley, and Chris Missen. Predicting serial killers' home base using a decision support system. *Journal of Quantitative Criminology*, 16:457–478, 2000.

4.Don Grubin, Peter Kelly, and Chris Brunsdon. *Linking Serious Sexual Assaults through Behaviour.* London: Home Office, 2001.

5.Ned Levine. *CrimeStat: A Spatial Statistics Program for the Analysis of Crime Indicdent Locations (version 1.1).* Washington, D.C.: National Institute of Justice, 2000.

6.Darcey Kim Rossmo. *Geographic Profiling.* Boca Raton: CRC Press, 2000.

7.Samuel Shye. Modern facet theory: Content design and measurement in behavioural research. *European Journal of Psychological Assessment*, 14:160–171, 1998.

8.John A. Swets, Robyn M. Dawes, and John Monahan. Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, 1:1–26, 2000.