Clustering of Psychological Personality Tests of Criminal Offenders

Markus Breitenbach[1], Tim Brennan[24], William Dieterich[34], and Gregory Z. Grudic[1]

[1] University of Colorado at Boulder, Boulder CO 80309, USA
[2] Institute for Cognitive Science, University of Colorado at Boulder, Boulder CO 80309, USA
[3] Graduate School of Social Work, University of Denver, Denver CO 80202, USA
[4] Northpointe Institute for Public Management Inc.,

Abstract. In Criminology research the question arises if certain types of delinquents can be identified from data, and while there are many cases that can not be clearly labeled, overlapping taxonomies have been proposed in [18], [20] and [21]. In a recent study Juvenile offenders (N = 1572) from three state systems were assessed on a battery of criminogenic risk and needs factors and their official criminal histories. Cluster analysis methods were applied. One problem we encountered is the large number of hybrid cases that have to belong to two or more classes. To eliminate these cases we propose a method that combines the results of Bagged K-Means and the consistency method[1], a semi-supervised learning technique. A manual interpretation of the results showed very interpretable patterns that were linked to existing criminologic research.

1 Introduction

Unsupervised clustering has been applied successfully in many applied disciplines to group cases on the basis of similarity across sets of domain specific features. A typical analytical sequence in the data mining process is to first identify clusters in the data, assess robustness, interpret them and later train a classifier to assign new cases to the respective clusters.

The present study applies several unsupervised clustering techniques to a highly disputed area in criminology i.e. the existence of criminal offender types. Many contemporary criminologist argue against the possibility of separate criminal types [22] while others strongly support their existence (see [20,23]). Relatively few studies in criminology have used Data Mining techniques to identify patterns from data and to examine the existence of criminal types. To date, the available studies have typically used inadequate cross verification techniques, small and inadequate samples and have produced inconsistent or incomplete findings, so that it is often difficult to reconcile the results across these studies. Often, the claims for the existence of "criminal types" have emerged from psychological or social theories that mostly lack empirical verification. Several attempts have been made to integrate findings from available classification studies These efforts have suggested some potential replications of certain offender types but have been limited by their failure to provide clear classification rules e.g. psychopathic offenders have emerged from a large clinical literature but there remains much dispute over how to identify them, and what specific social and psychological causal factors are critical, and whether or not this type exists among female offenders or among adolescents and whether there are "sub-types" of psychopaths. Thus, a current major challenge in criminology is to address whether reliable patterns or types of criminal offenders can be identified using data mining techniques and whether these may replicate substantive criminal profiles as described in the prior criminological literature.

In a recent study Juvenile offenders ($N = 1572$) from three U.S. state systems were assessed using a battery of criminogenic risk and needs factors as well as official criminal histories. Data mining techniques were applied with the goal to identify intrinsic patterns in this data set and assess whether these replicate any of the main patterns previously proposed in the criminological literature [23]. The present study aimed to identify patterns from this data and to demonstrated that they relate strongly to only certain of the theorized patterns from the prior criminological literature. The implications of these findings for Criminology are manifold. The findings firstly suggest that certain offender pattern can be reliably identified data using a variety of data mining unsupervised clustering techniques. Secondly, the findings strong challenge those criminological theorists who hold that there is only one general "global explanation" of criminality as opposed to multiple pathways with different explanatory models (see [24]).

The present paper describes some difficult analytical problems encountered in applied criminological research that stem from the kind of data produced in this field. A first major problem is that the data is noisy and often unreliable. Second, the empirical clusters are not clear cut so that cases range from strongly classified to poorly classified boundary cases with only weak cluster affiliations. Certain cases may best be seen as hybrids (close to cluster boundaries) or outliers. The distortion of clusters may also be a problem since it is well known many clustering algorithms assign a label to every point in the data, including outliers. Such "forcing" of membership - for both hybrids and outliers - may distort the quality and interpretation of the clustering results. Standard methods such as K-Means will assign cases to the closest cluster center no matter how "far away" from the cluster centers the points are. Some algorithms like EM-Clustering [2] output probabilities of class-membership, but nonetheless eliminating outliers in a unsupervised setting was a hard problem in this area of applied research. In this context we acknowledge that much work has been done to make clustering more robust against outliers, such as using clustering ensembles [3,4] or combining the results of different clustering methods [5], but we are not aware of a method to eliminate points in an aggressive way to obtain a more refined clustering solution, i.e. removing points that are not "close enough" to the cluster center.

Thus, in this research we also demonstrate a methodology to identify well clustered cases. Specifically we combine a semi-supervised technique with an initial standard clustering solution. In this process we obtained highly replicated offender types with clear definitions of each of the reliable and core criminal patterns. These replicated clusters provide social and psychological profiles that bear a strong resemblance to several of the criminal types previously proposed by leading Criminologists [23,20]. However, the present findings firstly go beyond these prior typological proposals by grounding the type descriptions in clear empirical patterns. Secondly they provide explicit classification rules for offender classification that have been absent from this prior literature.

2 Method

We started with an initial solution obtained "manually" using standard K-means and Wards minimum variance method. These have been the preferred choice in numerous social and psychological studies to find hidden or latent typological structure in data [10,11].

However, despite its success standard K-means is vulnerable to data that do not conform to the minimum-variance assumption or expose a manifold structure, that is, regions (clusters) that may wind or straggle across a high-dimensional space. These initial K-means clusters are also vulnerable to remaining outliers or noise in the data. Thus, we proceeded with two additional methods designed to deal more effectively with these outlier and noise problems.

2.1 Bagged K-Means

Bagging has been used with success for many classification and regression tasks [9]. In the context of clustering, bagging generates multiple classification models from bootstrap replicates of the selected training set and then integrates these into one final aggregated model. By using only two-thirds of the training set to create each model, we aimed to achieve models that should be fairly uncorrelated so that the final aggregated model may be more robust to noise or any remaining outliers inherent in the training set.

In [6] a method combining Bagging and K-means clustering is introduced. In our analyses we used the K-means implementation in R [7]. We generated 1000 random bags from our initial sample of 1,572 cases with no outliers removed to obtain cluster solutions for each bag. The centers of these bags were then treated as data points and re-clustered with K-means. The final run of this K-means was first seeded with the centers from our initial solution, which was then tested against one obtained with randomly initialized centers. These resulted in the same solution, suggesting that initializing the centers in these ways did not unduly bias K-means convergence. The resulting stable labels were then used as our final centers for the total dataset and in the voting procedure outlined below.

2.2 Semi-Supervised Clustering

Zhou et.al. introduced the consistency method in [1], a semi-supervised learning technique. This method, given one labeled example per class, assigns all remaining unlabeled cases in accordance with the underlying intrinsic structure of the dataset. Thus, whereas K-means tends to favor (or impose) hyper-spherical clustering structure, the semi-supervised method is more sensitive to almost any arbitrary cluster structure or shape intrinsic to the data being analyzed as long as the point-clouds are connected. The method works by propagating labels from the labeled points to all other points over each iteration. However, the further the point is away from the labeled point, the fewer that label information is propagated. An unlabeled point is assigned to the class with the highest value of activation. This allows the method to follow the shape of arbitrarily shaped clusters as long as they are dense. This is illustrated in figure 1(a) and shows the label assignment for different steps of the propagation. The method has demonstrated good performance on high-dimensional domains such as various image classification tasks.

2.3 Obtaining a Refined Solution: Consensus Cases and Voting Procedure

To tackle the problem of hybrid case elimination we use a voting methodology to eliminate cases in which different algorithms produce a disagreement similar to [5] that combines hierarchical and partitioning clusterings.

In this paper we propose the following solution: First, we use Bagged K-Means [6] to get a stable estimate of our cluster centers in the presence of outliers and hybrid cases. To eliminate cases that are far away from the cluster centers, we will use the obtained centers in a semi-supervised setting with the consistency method [1] to obtain a second set of labels. The labels from the semi-supervised method are obtained with a completely different similarity measure than the K-Means labels. K-Means assigns labels by using the distance to the cluster center (Nearest Neighbor) and works best given clusters that are Gaussian. The semi-supervised consistency method assigns labels with respect to the underlying intrinsic structure of the data and follows the shape of the cluster. These two fundamentally different methods of label assignments are more likely to disagree the further away the point is from the cluster center. We eliminate cases in which the labels do not agree. Note that the consistency method has been demonstrated to work well on high-dimensional data such as images. On the other hand it has been demonstrated that assignments of labels using Nearest Neighbor in high dimensional spaces are often unusable [8].

The process is illustrated in figure 1(b) with a toy example consisting of three Gaussians and a couple of hybrid cases placed in between. The labeling resulting from K-Means and the consistency method differ. The final voting solution consists of less hybrid cases (marked in blue in the bottom right figure).

Using the method outlined above results in roughly half the cases of our data being eliminated. The stability of these central core cases - as retained in the consensus model - is shown by the almost identical matching of these core cases between the consensus model and the bagged K-means solution ($\kappa = .992, \eta = .994$) and also to the original K-means ($\kappa = 0.949, \eta = 0.947$).

3 Results

The core clusters obtained with this method were interpreted and relationships with types already identified in the criminology literature examined.

The clusters identified were Internalizing Youth A[20,13,16], Socialized Delinquents [12,14,15], Versatile Offenders[20], Normal Accidental Delinquents[18], Internalizing Youth B[20], Low-control Versatile Offenders[20,21] and Normative Delinquency [19]. All the clusters relate to types that have been previously identified various studies in the Criminology literature, but were never identified at the same time in one data set using clustering.

External validation requires finding significant differences between clusters on external (but relevant) variables that were not used in cluster development. By comparing the means and bootstrapped 95 percent confidence intervals of four external variables across the seven clusters from the core consensus solution we identified those variables. The external variables include three criminal history variables (total adjudications, age-at-first adjudication and total violent felony adjudications) and one demographic variable (age-at-assessment). These plots show a number of significant differences in expected directions. For example, clusters 4 and 7, which both match the low risk profile of Moffitt's AL type [18] have significantly later age-at-first adjudication compared to the higher risk cluster 6 that matches Moffitt's high risk LCP and Lykken's [20] Secondary Psychopath. This latter cluster has the earliest age-at-first arrest and significantly higher total adjudications - which is consistent with Moffitt's descriptions.

Finally, while our results indicate that boundary conditions of clusters are obviously unreliable and fuzzy, the central tendencies or core membership appear quite stable. This suggests that these high density regions contain sufficient taxonomic structure to support reliable identification of type membership for a substantial proportion of juvenile offenders.

Using the method in Section 2 we were able to remove most of the hybrid cases. In fact, the case removal was overly aggressive and removed roughly half the data set. However, the remaining cases were very interpretable on manual inspection. Our analyses also show that cluster boundaries are relatively unstable. Kappa's from 0.55 to 0.70, although indicating general overlap, also imply that boundaries between clusters may be imposed differently, and cases close to boundaries may be unreliably classified across adjacent clusters. Many of these cases may be regarded as hybrids with many co-occurring risk or needs and multiple causal influences. Lykken [20] recognized this by stating that many offenders will have mixed etiologies and will be borderline or hybrid cases (p. 21).

The presence of hybrids and outliers appears unavoidable given the multivariate complexity of delinquent behavior, the probabilistic nature of most risk factors and multiplicity of causal factors. Additionally, our findings on boundary conditions and non-classifiable cases must remain provisional since refinements to our measurement space may reduce boundary problems. Specifically, it is known that the presence of noise and non-discriminating variables can blur category boundaries [10]. Further research may clarify the discriminating power of all classification variables (features) and gradually converge on a reduced space of only the most powerful features.

4 Conclusion

In this paper we report on our experiences with finding clusters in the Youth COMPAS data set which contains 32 scale scores used for criminogenic assessment.

Cluster analysis methods (Ward's method, standard k-means, bagged k-means and a semi-supervised pattern learning technique) were applied to the data. Cross-method verification and external validity were examined. Core or exemplar cases were identified by means of a voting (consensus) procedure. Seven recurrent clusters emerged across replications.

The clusters that were found using unsupervised learning techniques partially replicate several criminal types that have been proposed in previous criminological research. However, the present analyses provide more complete empirical descriptions than in most previous studies and allow. Additionally, the presence of certain sub-types among these major types is suggested by the present analysis. This is the first study in which most of the well replicated patterns were identified purely from the data. We stress that many prior studies provided only partial theoretical or clinical descriptions, omit operational type-identification procedures or provide very limited feature sets.

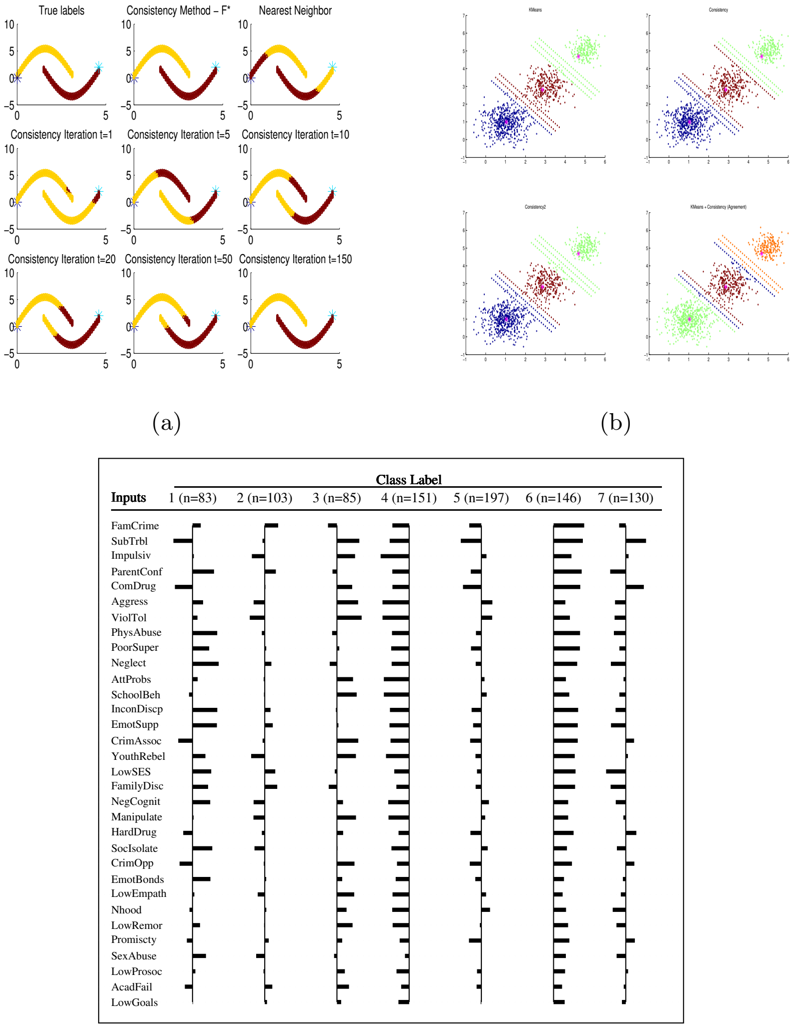
We introduced a novel way of hybrid-case elimination in an unsupervised setting and although we are still working on establishing a more theoretical foundation of the technique it has generally resulted in good results and very interpretable clusters. From the resulting clusters a classifier was build from the data in order to classify new cases.

It is noteworthy that the initial solution we obtained with an elaborate outlier removal process using Ward's linkage and regular K-Means was easily replicated using Bagged K-Means without outlier removal or other "manual" operations. In this instance Bagged K-Means appears to be very robust against noise and outliers.

References

1. Zhou, D., Bousquet, O., Lal, T., Weston, J., Schölkopf, B.: Learning with local and global consistency. In S. Thrun, L.S., Schölkopf, B., eds.: Advances in Neural Information Processing Systems 16, Cambridge, Mass., MIT Press (2004)
2. Dempster, A., Laird, N., Rubin, D.: Maximum-likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society **39** (1977)
3. Dimitriadou, E., Weingessel, A., Hornik, K.: Voting-merging: An ensemble method for clustering. In: Lecture Notes in Computer Science. Volume 2130., Springer Verlag (2001) 217

4. Topchy, A.P., Jain, A.K., Punch, W.F.: Combining multiple weak clusterings. In: Proceedings of the ICDM. (2003) 331–338
5. Lin, C.R., Chen, M.S.: Combining partitional and hierarchical algorithms for robust and efficient data clustering with cohesion self-merging. In: IEEE Transactions on Knowledge and Data Engineering. Volume 17. (2005) 145 – 159
6. Dolnicar, S., Leisch, F.: Getting more out of binary data: Segmenting markets by bagged clustering. Working Paper 71, SFB 'Adaptive Information Systems and Modeling in Economics and Management Science" (2000)
7. R Development Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2004) 3-900051-07-0.
8. Beyer, K., Goldstein, J., Ramakrishnan, R., Shaft, U.: When is 'nearest neighbor" meaningful? Lecture Notes in Computer Science **1540** (1999) 217–235
9. Breiman, L. (1996). Bagging predictors. *Machine Learning* 24(2): 123–140.
10. Milligan, G. W. (1996). Clustering validation: Results and implications for applied analyses. In Arabie, P., Hubert, L., and De Soete, G. (eds.), *Clustering and Classification*, World Scientific Press, River Edge, NJ, pp. 345–379.
11. Han, J., and Kamber, M. (2000). *Data Mining - Concepts and Techniques*, Morgan Kauffman, San Francisco.
12. Miller, W. (1958). Lower-class culture as a generating milieu of gang delinquency. *Journal Of Social Issues* 14: 5–19.
13. Miller, M., Kaloupek, D. G., Dillon, A. L., and Keane, T. M. (2004). Externalizing and internalizing subtypes of combat-related PTSD: A replication and extension using the PSY-5 scales. *Journal of Abnormal Psychology* 113(4): 636–645.
14. Jesness, C. F. (1988). The Jesness Inventory Classification System. *Criminal Justice and Behavior* 15(1): 78–91.
15. Warren, M. Q. (1971). Classification of offenders as an aid to efficient management and effective treatment. *Journal of Criminal Law, Criminology, and Police Science* 62: 239–258.
16. Raine, A., Moffitt, T. E., and Caspi, A. (2005). Neurocognitive impairments in boys on the life-course persistent antisocial path. *Journal of Abnormal Psychology* 114(1): 38–49.
17. Brennan, T., and Dieterich, W. (2003). *Youth COMPAS Psychometrics: Reliability and Validity*, Northpointe Institute for Public Management, Traverse City, MI.
18. Moffitt, T. E. (1993). Adolescence-limited and life-course persistent antisocial behavior: A developmental taxonomy. *Psychological Review* 100(4): 674–701.
19. Moffitt, T. E., Caspi, A., Rutter, M., and Silva, P. A. (2001). *Sex Differences in Antisocial Behaviour*, Cambridge University Press, Cambridge,Mass.
20. Lykken, D. (1995). *The Antisocial Personalities*, Lawrence Erlbaum, Hillsdale, N.J.
21. Mealey, L. (1995). The sociobiology of sociopathy: An integrated evolutionary model. *Behavioral and Brain Sciences* 18(3): 523–599.
22. Farrington, D.P. (2005) *Integrated Developmental and Life-Course Theories of Offending*, Transaction Publishers,London (UK).
23. Piquero A.R. and Moffitt T.E. (2005) Explaining the facts of crime: How developmental taxonomy replies to Farrington's Invitation *Chapter in Farrington D.P (Ed) Integrated Developmental and Life-Course Theories of Offending*, Transaction Publishers,London (UK).
24. Osgood D. W. (2005). Making sense of crime and the life course *Annals of AAPSS, November 2005*, 602:196-211.



(a)

(b)

(c)

Fig. 1. (a) **Consistency Method**: two labeled points per class (big stars) are used to label the remaining unlabeled points with respect to the underlying cluster structure. F^* denotes the convergence of the series. (b) **Toy example**: Three Gaussians with hybrid cases in between them. Combining the labels assigned by K-Means (top, left) and the Consistency Method (top, right; bottom, left) with two different σ results in the removal of most of the hybrid cases (blue dots; bottom, right) by requiring consensus between all models build. The K-Means centers have been marked in magenta. (c) **Resulting Cluster Means**: Mean Plots of External Criminal History Measures Across Classes from the Core Consensus Solution with Bootstrapped 95% Confidence Limits.