# Onto Clustering of Criminal Careers*

Jeroen S. de Bruin, Tim K. Cocx, Walter A. Kosters,
Jeroen F.J. Laros, and Joost N. Kok

LIACS, Leiden University, The Netherlands
tcocx@liacs.nl

**Abstract.** We analyze criminal careers through crime nature, frequency, duration and severity. We propose a tool that yields a visual clustering of these criminal careers, enabling the identification of classes of criminals.

## 1 Introduction

The Dutch national police annually extracts information from digital narrative reports stored throughout the individual departments. This data is compiled into a large and reasonably clean database that contains all criminal records from the last decade. This paper discusses a new tool that attempts to gain new insights into the concept of *criminal careers*, the criminal activities that a single individual exhibits, from this data.

The main contribution of this paper is in Section 4, where the criminal profiles are established and a distance measure is introduced.

## 2 Background

Background information on clustering techniques in the law enforcement arena can be found in [1, 4]. Our research aims to apply multi-dimensional clustering to criminal careers (rather than crimes or linking perpetrators) in order to constitute a visual representation of classes of these criminals. A theoretical background to criminal careers and the important factors can be found in [2].

## 3 Approach

We propose a *criminal career analyzer*, which is a multi-phase process visualized in Figure 1. Our tool normalizes all careers to "start" on the same point in time and assigns a profile to each offender. After this step, we compare all possible pairs of offenders on their profiles and profile severity. We then employ a specifically designed distance measure that incorporates crime frequency and the change over time, and finally cluster the result into a two-dimensional image using the method described in [3].

* This research is part of the DALE project (Data Assistance for Law Enforcement) as financed in the ToKeN program from the Netherlands Organization for Scientific Research (NWO) under grant number 634.000.430

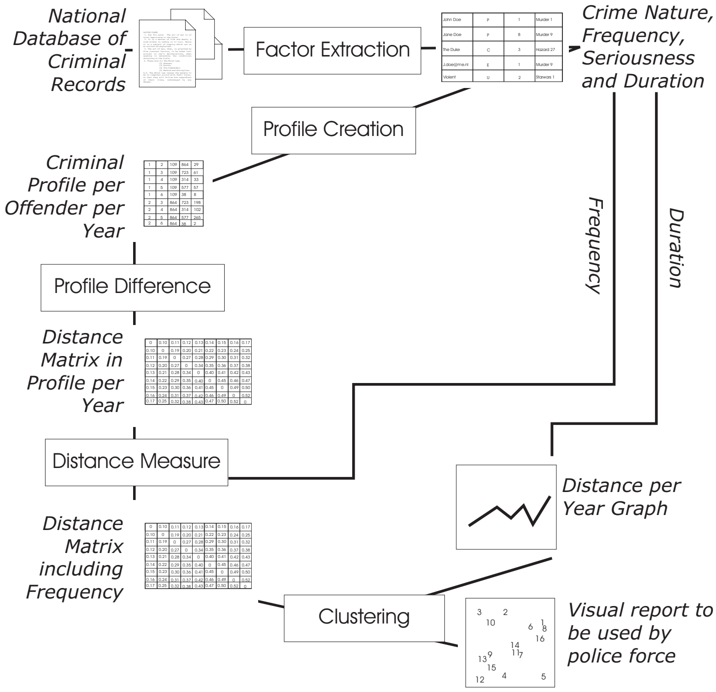


**Fig. 1.** Analysis of Criminal Careers

## 4 Method of Career Comparison

We distinguish eight different types of crime (vandalism, ..., sexual violence) divided into three severity classes (minor, intermediate, severe). Each offender's profile $x$ is described by a table containing the percentages $Perc_{ix}$ of crimes that fall within each category $i$, summing to 1 (if at least one crime is committed). We also assign a severity class $k$ to each crime type. Each severity class gets its own weighting factor $Fact_k$ (1,2,3, respectively). The total severity of crimes in class $k$ for person $x$ (the sum of the appropriate $Perc_{ix}$'s) are then described as $Sev_{kx}$.

After compilation of all individual profiles, we employ the method described in Formula 1 to create a *profile difference* matrix $PD$, where $PD_{xy}$ denotes the profile difference between persons $x$ and $y$:

$$PD_{xy} = \sum_{i=1}^{8} |Perc_{ix} - Perc_{iy}| + |\sum_{k=1}^{3} Fact_k \cdot Sev_{kx} - \sum_{k=1}^{3} Fact_k \cdot Sev_{ky}| \quad (1)$$

This intermediate distance matrix describes the profile difference per year for each possible pair of offenders. Its values all range between 0 and 4.

The crime frequency, or number of crimes, will be divided into categories (0, 1, 2–5, 5–10, >10 crimes per year) to make sure that the absolute difference shares the range 0–4 with the calculated profile difference, instead of the unbounded number of crimes per year offenders can commit. The *frequency value difference* will be denoted by $FVD_{xy}$.

Criminal careers of one-time offenders are obviously reasonably similar, although their single crimes may differ largely in category or severity class. However, when looking into the careers of career criminals there are only minor differences to be observed in crime frequency and therefore the descriptive value of profile becomes more important. Consequently, the dependence of the profile difference on the crime frequency must become apparent in our distance measure. This ultimately results into a proposal for the *crime difference per year* distance $CPDY_{xy}$ between persons $x$ and $y$:

$$CDPY_{xy} = \frac{\frac{1}{4} \cdot PD_{xy} \cdot FVD_{xy} + FVD_{xy}}{8} = FVD_{xy} \cdot \left( \frac{PD_{xy}}{32} + \frac{1}{8} \right) \qquad (2)$$

The factor $1/8$ guarantees that $0 \leq CDPY_{xy} \leq 1$.

We have now calculated the career difference distance matrix containing all career comparison information between individual offenders. The clustering method that we incorporated in our tool was described by Broekens et al. [3] and allows data analysts, using node-coloring, to correct small mistakes made by naive clustering algorithms that result in local optima.

## 5 Experimental Results

Figure 2 gives an impression of the output produced by our tool when analyzing the beforementioned database.

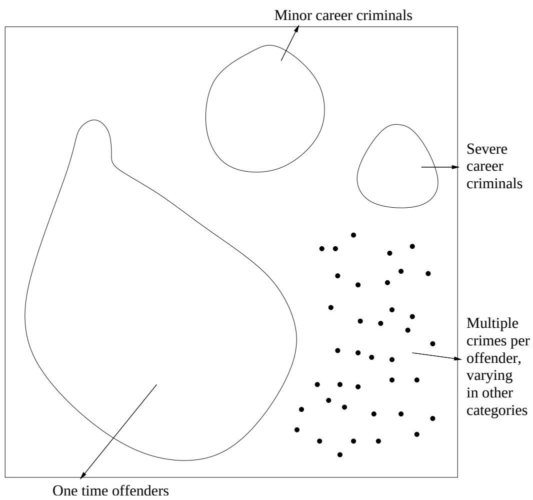


**Fig. 2.** Experimental results of tool usage

This image clearly shows what identification could easily be coupled to the appearing clusters after examination of its members. It appears to be describing reality very well. The large "cloud" in the left-middle of the image contains (most of the) one-time offenders. This seems to relate to the database very well since approximately 75% of the people it contains has only one felony or misdemeanour on his or her record. The other apparent clusters also represent clear subsets of offenders.

## 6 Conclusion and Future Directions

The tool we described compiled a criminal profile out of the four important factors describing a criminal career for each individual offender. We developed a specific distance measure to combine profile difference with crime frequency and the change of criminal behavior over time to create a visual two-dimensional clustering overview of criminal careers that is ready to be used by police experts.

The enormous "cloud" of one-time offenders gave a somewhat unclear situational sketch of our distance space. This problem, however, can not be easily addressed since a large part of the National Criminal Record Database simply consists of this type of offenders. Its existence shows, however, that our approach easily creates an identifiable cluster of this special type of criminal career, which is promising. One possible solution to this problem would be to simply not take these individuals into account when compiling our distance matrices.

The used method of clustering provided results that seem to represent reality well, and are clearly usable by police analysts, especially when the above is taken into account. The speed of the chosen approach, however was sub-optimal thus far. In the future, an approach like Progressive Multi Dimensional Scaling [5] could be more suited to the proposed task in a computative way, while maintaining the essence of career analysis. Future research will aim at solving both above mentioned concerns.

## References


1. R. Adderley and P. B. Musgrove. Data mining case study: Modeling the behavior of offenders who commit serious sexual assaults. In *KDD '01: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 215–220, New York, 2001.
2. A. Blumstein, J. Cohen, J. A. Roth, and C. A. Visher. *Criminal Careers and "Career Criminals"*. The National Academies Press, 1986.
3. J. Broekens, T. Cocx, and W.A. Kosters. Object-centered interactive multi-dimensional scaling: Let's ask the expert. To be published in the Proceedings of the 18th BeNeLux Conference on Artificial Intelligence (BNAIC 2006), 2006.
4. T.K. Cocx and W.A. Kosters. A distance measure for determining similarity between criminal investigations. In *Proceedings of the Industrial Conference on Data Mining 2006 (ICDM2006)*, LNCS. Springer, 2006.
5. M. Williams and T. Muzner. Steerable, progressive multidimensional scaling. In *IEEE Symposium on Information Visualization (INFOVIS'04)*, pages 57–64. IEEE, 2004.