Algorithmic tools for data-oriented law enforcement
Cocx, T.K.

Citation
Cocx, T. K. (2009, December 2). *Algorithmic tools for data-oriented law enforcement*. Retrieved from https://hdl.handle.net/1887/14450

Chapter 4

Data Mining Approaches to Criminal Career Analysis

Narrative reports and criminal records are stored digitally across individual police departments, enabling the collection of this data to compile a nation-wide database of criminals and the crimes they committed. This collection of data over a number of years presents new possibilities of analyzing criminal activity through time. Augmenting the traditional, more socially oriented, approach of behavioral study of these criminals and traditional statistics, data mining methods like clustering and prediction enable police forces to get a clearer picture of criminal careers. This allows officers to recognize crucial spots in changing criminal behavior and deploy resources to prevent these careers from unfolding.

Four important factors play a role in the analysis of criminal careers: crime nature, frequency, duration and severity. We describe a tool that extracts these from the police database and creates digital profiles for all offenders. It compares all individuals on the basis of these profiles by a new distance measure and clusters them accordingly. This method yields a visualization of the criminal careers and enables the identification of classes of criminals. The proposed method allows for several user-defined parameters.

4.1   Introduction

The amount of data being produced in modern society is growing at an accelerating pace. New problems and possibilities constantly arise from this so-called data explosion. One of the areas where information plays an important role is that of law enforcement. Obviously, the amount of criminal data gives rise to many problems in areas like data storage, data warehousing, data analysis and privacy. Already, numerous technological efforts are underway to gain insights into this information and to extract knowledge from it.

This chapter discusses a new tool that attempts to gain insights into criminal careers:

the criminal activities that a single individual exhibits throughout his or her life. From the criminal record database described in Appendix B, our tool extracts four important factors (Section 4.2) in criminal careers and establishes a clear picture on the different existing types of criminal careers by clustering. All four factors need to be taken into account and their relative relations need to be established in order to reach a reliable descriptive image. To this end we propose a way of representing the criminal profile of an individual in a single year. We then employ a specifically designed distance measure to combine this profile with the number of crimes committed and the crime severity, and compare all possible couples of criminals. When this data has been compared throughout the available years we use the human centered visualization tool, described in Chapter 3, to represent the outcome to the police analysts. We discuss the difficulties in career-comparison and the specific distance measure we designed to cope with this kind of information.

In Section 4.3 we describe our approach to this problem. The main contribution of this chapter is in Section 4.4 and 4.5, where the criminal profiles are established and the distance measure is introduced. A prototype test case of this research was described earlier in [8].

4.2   Background

In Chapter 2, [10] we made an attempt at revealing what criminal characteristics were linked together through common co-occurrence. The research in this chapter aims to apply visualization of multi-dimensional data to criminal careers (rather than single crimes or crimes to demographic data) in order to constitute a representation of clusters or classes of these criminals.

Criminal careers have always been modeled through the observation of specific groups of criminals. A more individually oriented approach was suggested by Blumstein et al. [4]: little definitive knowledge had been developed that could be applied to prevent crime or to develop efficient policies for reacting to crime until the development of the criminal career paradigm. A criminal career is the characterization of a longitudinal sequence of crimes committed by an individual offender. Participation in criminal activities is obviously restricted to a subset of the population, but by focusing on the subset of citizens who do become offenders, they looked at the *frequency*, *seriousness* and *duration* of their careers. We also focus our criminal career information on the *nature* of such a career and employ data mining techniques to feed this information back to police analysts and criminologists.

4.3   Approach

Our "criminal career analyzer" (see Figure 4.1) is a multi-phase process that works on a criminal record database (in our case, the database described in Appendix B). From these files our tool extracts the four factors, mentioned in Section 4.2, and establishes the criminal profile per offender. We then compare all individuals in the database and calculate a

distance based upon this profile, crime severity and number of crimes. This information is summed over time and clustered using a human centered multi-dimensional visualization approach. Black boxed, our paradigm reads in the database and provides a visual career comparison report to the end user.
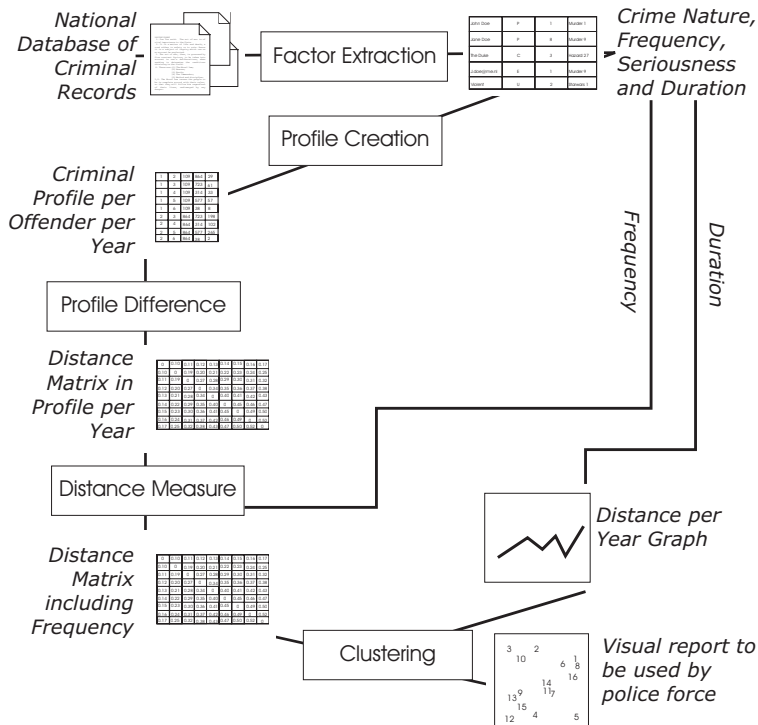

Figure 4.1: Analysis of Criminal Careers

First, our tool normalizes all careers to "start" on the same point in time, in order to better compare them. Second, we assign a profile to each individual. This is done for each year in the database. After this step, that is described in more detail in Section 4.4, we compare all possible pairs of offenders on their profiles for one year, while taking into account that seriousness is inherently linked to certain profiles. The resulting distance matrix is based upon only two of the four before mentioned criteria, crime nature and severity. We then employ a distance measure described in Section 4.5 to fairly incorporate the frequency of delinquent behavior into our matrix which is then summed over the available years. In the last step we visualize our distance matrix in a two-dimensional image using the method described in [5].

The database is an up-to-date source of information and therefore represents the current situation. This does, however, present us with the following problem: it also contains the criminal activities of people that started their career near the end of the database's range. Our approach suffers from a lack of information on these offenders and when

translated, it could be falsely assumed that their criminal activities terminated — however, they have just started. To cope with this problem we draw a semi-arbitrary line after two years, which means we dot not take criminals into account that have only been active for two years or less. In a later stage, it is presumably possible to predict in which direction the criminal careers of these offenders will unfold. More information about this prediction can be found in Section 4.7.

4.4   Obtaining Profile Differences

Interpreting the types of crimes in the database and compiling them into a criminal profile for individual offenders is done both by determining the percentage of crimes in one year that fall into each one of the categories and dealing with these crimes' seriousness. The categories are then ordered on severity. Note that while all categories are considered to be different when determining profile, their respective severity can be the same. We distinguish three types of crime seriousness: minor crimes, intermediate crimes and severe crimes. Each crime type falls within one of these categories. A typical profile of a person looks like the profile described in Table 4.1. Summing all the values in a row in the table will result in 1, representing 100% of the crimes across the categories.

Table 4.1: A typical Criminal Profile

Crime Type	*Traffic Crimes*	*Financial Crimes*	*...*	*Sex Crimes*
Severity	minor	minor	...	severe
Percentage	0	0.6	...	0.1

Now that we have established a way of representing the crime profile of an offender for a single year, it is possible to compare all individuals based upon this information and compile the profile distance matrix for each year. It is imperative that we take both the severity and nature of each profile into account.

We employ the method described in Table 4.2 to accomplish this. Between all couples of offenders we calculate the *Profile Distance* as follows. First we take the absolute difference between the respective percentages in all categories and sum them to gain the difference between these two offenders in crime nature. This number varies between 0, if the nature of their crimes was exactly the same for the considered year, and 2 where both persons did not commit any crime in the same category. An example of this can be seen in the top panel of Table 4.2, with 0.6 as difference.

One of the problems that arise in this approach is how it should deal with people with no offenses in the considered year. In [8], it was assumed that setting all the percentages in the crime nature table to 0 would provide a result that was true to reality. This does not seem to be the case; such an assumption would provide smaller distances between offenders and "innocent" people, than between different kinds of criminals for an arbitrary year. It would be desirable to assign a maximal distance in crime nature between

Table 4.2: Calculating the difference between Individual profiles

Crime Nature Difference

Ind. 1	0.1	0.0	0.0	0.0	0.5	0.4	0.0	0.0	
Ind. 2	0.0	0.2	0.1	0.0	0.4	0.3	0.0	0.0	Summed
Diff.	0.1	0.2	0.1	0.0	0.1	0.1	0.0	0.0	*0.6*

Crime Severity Difference

	Minor		Intermediate		Severe		
	#	Fac.	#	Fac.	#	Fac.	Summed
Ind. 1	0.1	1	0.5	2	0.4	3	
		0.1		*1.0*		*1.2*	*2.3*
Ind. 2	0.3	1	0.4	2	0.3	3	
		0.3		*0.8*		*0.9*	*2.0*
Diff.							*0.3*

Total Profile Difference: 0.9

innocence and guilt. Hence we propose to add to this table the *innocence column*. An individual without offenses in the year under consideration would have a 1 in this field, while an actual offender will always have committed 0% of his or her crimes of the innocent nature. This ensures that the algorithm will assign the desired maximal distance between these different types of persons.

Taking the severity of crimes into account is somewhat more complex. Not only the difference within each severity class needs to be calculated, but also the difference between classes. For example, an offender committing only minor crimes has a less similar career to a severe crime offender, than a perpetrator of the intermediate class would have to the same severe criminal. In compliance with this fact, our approach introduces a weighting factor for the three different categories. Minor offenses get a weight of 1, intermediate crimes get multiplied by 2 and the most extreme cases are multiplied by 3. Of course, these values are somewhat arbitrary, but the fact that they are linear establishes a reasonably representative and understandable algorithm. This multiplication is done for each individual and is summed over all severity classes before the difference between couples is evaluated. Obviously this will yield a greater distance between a small time criminal and a murderer ($|1 \cdot 1 - 3 \cdot 1| = 2$, which is also the maximum) than between an intermediate perpetrator and that same killer ($|2 \cdot 1 - 3 \cdot 1| = 1$). Naturally, two individuals with the same behavior, still exhibit a distance of 0. An example of this calculation can be found in the bottom panel of Table 4.2, with value 0.3.

Naturally, the innocence category in crime nature will be assigned a severity factor of

0, stating that a minor criminal (for example: theft) will be more similar to an innocent person than a murderer would be, when looking at crime severity.

Both these concepts will ultimately culminate into the following formula:

$$
PD_{xy} \quad = \quad \left(\sum_{i=1}^{9} \left| Perc_{ix} - Perc_{iy} \right| \right) +
$$

$$
\left(\left| \sum_{j=1}^{4} Fact_j \cdot Sev_{jx} - \sum_{k=1}^{4} Fact_k \cdot Sev_{ky} \right| \right)
$$

where the *profile distance per year* between person x and y is denoted as PD_{xy}, the percentage of person x in crime category a is described as $Perc_{ax}$, and Sev_{bx} points to the severity class b of person x, with attached multiplication factor $Fact_b$. This formula yields a maximum of 5 (the maxima of both concepts summed) and naturally a minimum of 0. In the example from Table 4.2 we would get $0.6 + 0.3 = 0.9$.

As an example of our approach we look at the following situation for an arbitrary year:

Table 4.3: Four people's behavior in one year

Person	*No Crime*	*Bicycle Theft*	*...*	*Murder*
Innocent Person	1.0	0.0	...	0.0
Bicycle Thief	0.0	1.0	...	0.0
Murderer	0.0	0.0	...	1.0
Combined	0.0	0.5	...	0.5

First we calculate the difference between all four offenders in both crime nature and crime severity. This yields the following table.

Table 4.4: Crime Nature and Severity Distances

	Nature					*Severity*			
	I	*B*	*M*	*C*		*I*	*B*	*M*	*C*
I		2	2	2	I		1	3	2
B			2	1	B			2	1
M				1	M				1

We can now calculate the Profile Difference between the offenders. The result can be seen in Figure 4.2.
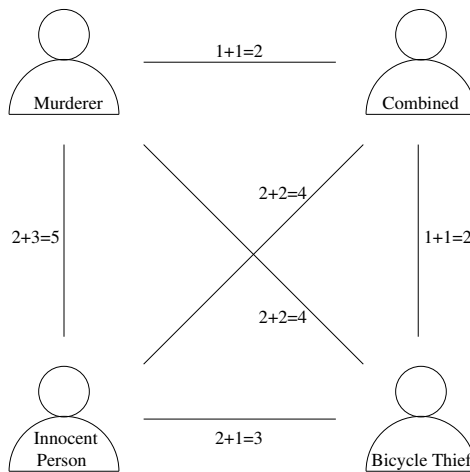
Figure 4.2: Example of Profile Difference between different criminals

We clearly observe from this figure that the innocent person from our list has a large distance from all offenders, varying from 3 to the bicycle thief to 5 to the murderer. As one would expect, the distances between respectively the thief and the thief-murderer, and the murderer and the thief-murderer are rather small.

The intermediate result we have calculated describes the difference in profile and crime seriousness between all individuals in the database. This distance matrix will be used as input for our main distance measure that will add crime frequency to the equation.

4.5   A New Distance Measure

The values the *PD* can get assigned are clearly bounded by 0 and 5. This is, however, not the case with the crime frequency of the number of crimes per year. This can vary from 0 to an undefined but possibly very large number of offenses in a single year. To get a controlled range of possible distances between individual careers it is desirable to discretize this range into smaller categories. Intuitively, beyond a certain number of crimes, the difference that one single crime makes is almost insignificant. Therefore we propose to divide the number of crimes into categories. These categories will divide the crime frequency space and assign a discrete value which will then be used as input for our distance measure. The categories we propose are described in Table 4.5. Figure 4.3 clearly shows that, using these categories, the significance of a difference in the number of committed crimes decreases with this number. Using this table the number of crimes value is bounded by 0 and 4, while the categories maintain an intuitive treatment of both one-time offenders and career criminals. Naturally, the difference between two individuals also ranges from 0 to 4.

A large part of the database described in Appendix B contains one-time offenders.

Table 4.5: Number of crimes categories

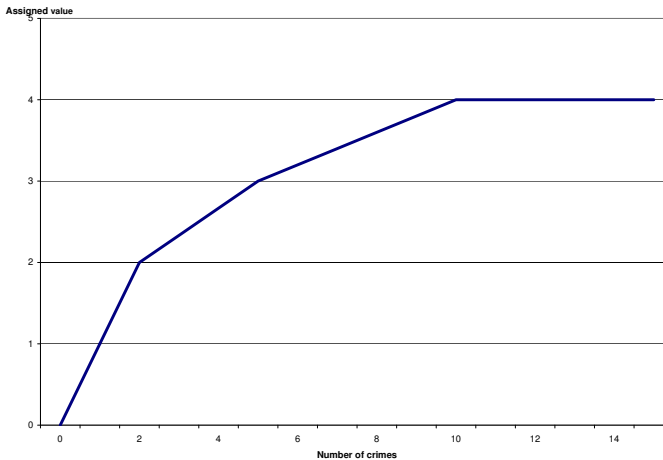| Category	Number of Crimes (#)	Assigned Value (FV)
0	0	0
1	1	1
2	2–5	$2 + \frac{\#-2}{3}$
3	5–10	$3 + \frac{\#-5}{5}$
4	>10	4

Figure 4.3: Relation between number of crimes and the assigned value (FV)

Their respective criminal careers 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 dependency of the *PD* on the crime frequency must become apparent in our distance measure. To this end we will multiply the profile oriented part of our distance measure with the frequency value and divide it by 4, to regain our original range between 0 and 5. Hence, the distance measure that calculates the *career difference per year* for offenders *x* and *y* ($CDPY_{xy}$) is as follows:

$$CDPY_{xy} = \frac{\frac{1}{4}(PD_{xy} \cdot FVD_{xy}) + FVD_{xy}}{9}$$

$$= FVD_{xy} \cdot \left(\frac{PD_{xy}}{36} + \frac{1}{9}\right)$$

Dividing by 9, as is done above, will result in a distance between 0 and 1 for all pairs of offenders, which is standard for a distance matrix. Usage of this distance measure will result in a distance matrix that contains all distances per year for all couples of criminals. The final preparation step will consist of calculation of all these matrices for all available years and incorporating the duration of the careers. Now that we have defined a measure $CDPY^i_{xy}$ to compare two persons x and y in a particular year i, we can combine this into a measure to compare full "careers". The simplest option is to use

$$\text{dist1}(x, y) = \frac{1}{years} \sum_{i=1}^{years} CDPY^i_{xy},$$

where *years* is the total number of years. This option is further explored in [8]. The idea can be easily generalized to

$$\text{dist2}(x, y) = \sum_{i=1}^{years} w(i)\, CDPY^i_{xy},$$

where the weights $w(i)$ must satisfy $\sum_{i=1}^{years} w(i) = 1$. For instance, the first year might have a low weight. Of course, dist1 is retrieved when choosing $w(i) = 1/years$ for all i.

Two problems arise from this situation. The first is that careers may be the same, but "stretched". Two criminals can commit the same crimes, in the same order, but in a different time period. This may or may not be viewed as a different career. On the one hand it is only the order of the crimes that counts, on the other hand the time in between also matters. The second is that if a criminal does not commit a crime in a given year, the distance to others (for that year) will be in some sense different from that between two active criminals.

In order to address these remarks, we propose a *stretch factor* δ_S, with $\delta_S \geq 0$. This factor can be set or changed by the career analyst if he or she feels that the provided result does not fully represent reality. If $\delta_S = 0$ we retrieve the original measure, if δ_S is high many years can be contracted into a single one. To formally define our measure, we look at two careers, where for simplicity we use capitals A, B, C, $\dots$ to indicate different crime types, and use multiset notation to specify the crimes in each year (so, e.g., $\{A, A, B\}$ refers to a year where two A-crimes and one B-crime were committed). Suppose we have the following 7 year careers x and y of two criminals:

$$x = (\{A, A, B\}, \{C\}, \{A\}, \{B\}, \{D\}, \varnothing, \varnothing),$$
$$y = (\{A, B\}, \varnothing, \{D\}, \varnothing, \{A\}, \{B, D\}, \varnothing).$$

The second career may be viewed as a somewhat stretched version of the first one, The differences are: x has two As in the first year, whereas y has only one; x has the singleton set $\{C\}$ in year 2, whereas y has $\{D\}$ in year 3; and year 6 from y is spread among year 4 and 5 of x. Evidently we are looking for some sort of *edit distance*.

Now we take $\delta_S \in \{0, 1, 2, \ldots\}$ and collapse at most δ_S years, for both careers, where we add $\varnothing$s to the right. Note that, when collapsing two sets, it is important to decide whether or not to use multisets. For instance, with $\delta_S = 2$, we might change y to

$$y' = (\{A, B\}, \{D\}, \varnothing, \{A, B, D\}, \varnothing, \varnothing, \varnothing).$$

If we have t years, there are

$$t + \binom{t-1}{2} + \ldots + \binom{t-1}{\delta_S}$$

ways to change a single career; by the way, many of them give the same result. For the above example, with $t = 7$ and $\delta_S = 2$, this gives 22 possibilities, and therefore $22 \times 22 = 484$ possible combinations of the two careers. We now take the smallest distance between two of these as the final distance. For the example this would be realized through

$$
\begin{aligned}
x' &= (\{A, A, B\}, \{C\}, \{A\}, \{B, D\}, \varnothing, \varnothing, \varnothing), \\
y'' &= (\{A, B\}, \{D\}, \{A\}, \{B, D\}, \varnothing, \varnothing, \varnothing).
\end{aligned}
$$

So finally the distance of x and y is determined by the distance between the singleton sets $\{C\}$ and $\{D\}$, and the way how the multiset $\{A, A, B\}$ compares to $\{A, B\}$. This last issue can be easily solved in the basic computation of PD, discussed in Section 4.4.

Of course, there are other practical considerations. For instance, some (perhaps even many) crimes will go unnoticed by police forces, giving an incorrect view. Furthermore, in the current approach there is a clear, but maybe unwanted separation between crimes committed in December and in January of the next year.

4.6   Visualization

The standard construction of our distance matrix enabled our tool to employ existing techniques to create a two-dimensional image. The intermediate Euclidean distances in this image correspond as good as possible to the distances in the original matrix. Earlier research [11] showed that automated clustering methods with human intervention might get promising results. The method that we incorporated in our tool was described in Chapter 4 and [5], where we built upon the push-and-pull architecture described in [20], and allows data analysts to correct small mistakes made by naive clustering algorithms that result in local optima. The user is assisted by coloring of the nodes that represent how much a single individual or a collection of individuals contribute to the total error made in the current clustering. The end result is a clickable two-dimensional image of all possible careers. When the user clicks on an individual he or she is presented with the profile information for that particular offender.

A standard K-means algorithm was used to automatically find clusters in the visualization. These clusters correspond to the large groups in the visualization and can be seen as an internal model of the visualization. If such a cluster is viewed as a representation of a certain class of criminal careers the internal clustering can be used to classify new individuals in the visualization by finding out in which cluster they reside. Note that an incremental visualization method is needed to correctly add such a new individual in the first place.

4.7 Prediction

Once a (large) group of criminals has been analyzed, it is possible to predict future behavior of new individuals. To this end we can readily use the tool from Chapter 3, [5]. Suppose that the given database has already been converted into a two-dimensional visualization and a clustering diagram has been created, as described in Section 4.6. We can and will assume that the original database consists of "full" careers. For a new partially filled career, e.g., consisting of only two consecutive years with crimes, we can compute its distance to all original ones taking into account only the number of years of the new career. The system now takes care of finding the right spot for this individual, where it reaches a local minimum for the distances in the two-dimensional space in comparison with the computed distances. The neighboring individuals will then provide some information on the possible future behavior of the newly added career; this information is of a statistical nature.

Of course, in practice it is not always clear when a career ends, not to speak of situations where a career ends or seems to end prematurely, e.g., because a criminal dies or is in prison for a substantial time period. More on the possibilities concerning prediction can be found in Chapter 5, where some initial steps are taken to facilitate extrapolation and in Chapter 7 and Chapter 8, where an actual attempt is made at the prediction of criminal careers based upon this data.

4.8 Experimental Results

We tested our tool on the actual database described in Appendix B, containing approximately one million offenders and the crimes they committed. Our tool analyzed the entire set of criminals and presented the user with a clear two-dimensional visualization of criminal careers as can be seen in Figure 4.5.

Three different clusters were automatically identified. The largest of them, seen in the bottom of Figure 4.5, was easily identified as the group of one-time offenders. The other two visible clusters were identified through manual sampling, investigating the four factors of the criminal careers of the sampled perpetrators in the cluster under observation. Combined with a default state of "unclassified", we can distinguish four different classes of criminal careers, proportioned as described in Figure 4.4, where the amounts are specified in thousands.

227; 21%

One-time offenders
Heavy career criminals
Minor career criminals
Unclassified

91; 8%

31; 3%

750; 68%

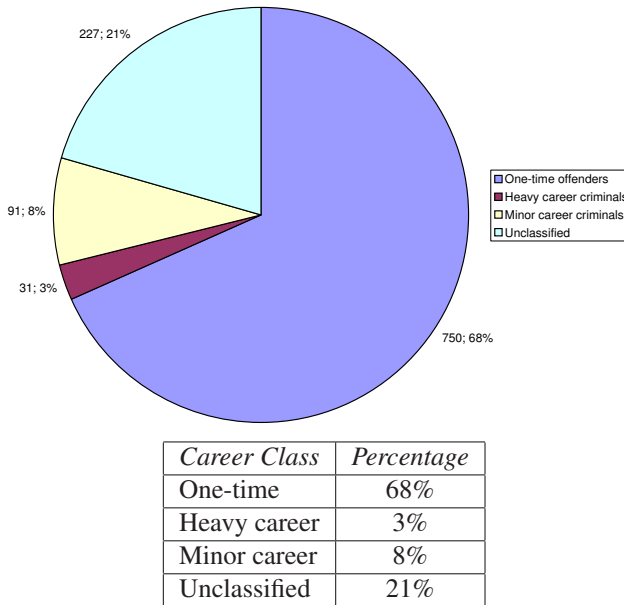| Career Class	Percentage
One-time	68%
Heavy career	3%
Minor career	8%
Unclassified	21%

Figure 4.4: Proportion of offenders in each class of criminal career

The confidential nature of the data used for criminal career analysis prevents us from disclosing more detailed experimental results reached in our research.

The image in Figure 4.5 gives an impression of the output produced by our tool when analyzing the before mentioned database. This image 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. There is however a reasonably large group of "un-clustered individuals". The grouping of these individual criminal careers might be influenced by the large group of one-timers. Getting more insights into the possible existence of subgroups in this non-cluster may prove even more interesting than the results currently provided by our approach. Future research will focus on getting this improvement realized (cf. Section 4.9).

4.9   Conclusion and Future Directions

In this chapter we demonstrated the applicability of data mining in the field of criminal career analysis. The tool we described compiled a criminal profile out of the four

Minor career criminals

Severe
career
criminals

Multiple
crimes per
offender,
varying
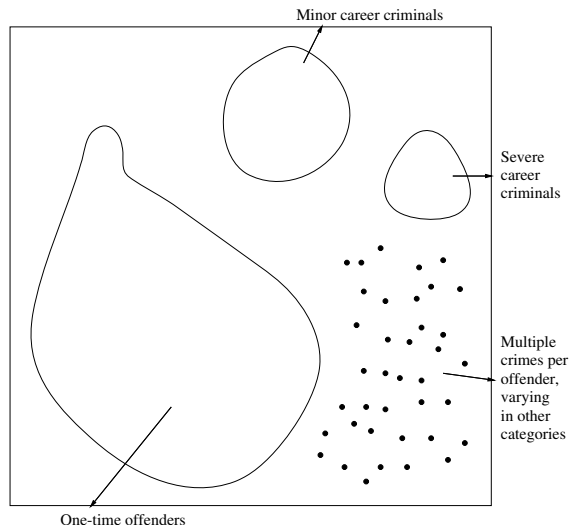in other
categories

One-time offenders

Figure 4.5: Visualization of Criminal Career clusters

important factors describing a criminal career for each individual offender: frequency, seriousness, duration and nature. These profiles were compared on similarity for all possible pairs of criminals using a new comparison method. We developed a specific distance measure to combine this profile difference with crime frequency and the change of criminal behavior over time to create a distance matrix that describes the amount of variation in criminal careers between all couples of perpetrators. This distance measure incorporates intuitive peculiarities about criminal careers. The end report consists of a two-dimensional visualization of the results and 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 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 visualization and clustering provided results that seem to represent reality well, and are clearly usable by police analysts, especially when the above is taken into account. However, the runtime of the chosen approach was not optimal yet. The visualization method was too intensive in a computational way, causing delays in the performance of the tool. In the future, an approach like Progressive Multi-Dimensional Scaling [35] 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 concerns mentioned above. After these issues have been properly addressed, research will mainly focus on the automatic comparison between the results provided by our tool and the results social studies reached on the

same subject, in the hope that "the best of both worlds" will reach even better analyzing possibilities for the police experts in the field. Incorporation of this tool in a data mining framework for automatic police analysis of their data sources is also a future topic of interest.

An inherit issue with these kind of approaches is that applicability in courts and daily police practice depends on statistical insightfulness and relevance, current law and that the fact that they are subject to privacy considerations. These matters are discussed in detail in Appendix A.