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

Chapter 8

An Early Warning System for the Prediction of Criminal Careers

Dismantling networks of career criminals is one of the focus points of modern police forces. A key factor within this area of law enforcement is the accumulation of delinquents at the bottom of the criminal hierarchy. A deployed early warning system could benefit the cause by supplying an automated alarm after every apprehension, sounding when this perpetrator is likely to become a career criminal. Such a system can easily be built upon existing, strategic, analysis already performed at headquarters. We propose a tool that superimposes a two-dimensional extrapolation on a static visualization, that describes the movement in time of an offender through the criminal spectrum. Using this extrapolation, possible future attributes are calculated and the criminal is classified accordingly. If the predicted class falls within the danger category, the system could notify police officials. We outline the implementation of such a tool and highlight test results on a criminal record database. We also touch upon the issue of estimating the reliability of a single, individual criminal career prediction.

8.1   Introduction

The dismantlement of crime syndicates ranks seventh in the current list of top priorities of the FBI [21]. As the growth of crime syndicates starts at the bottom layer of the criminal hierarchy, which is most often shaded to law enforcement agencies, a tool that makes educated guesses about people who are most likely to enter such organizations can be a valuable asset. This chapter discusses a new tool that attempts to predict the continuation of individual criminal careers: the criminal activities that a single individual exhibits throughout his or her life. Analysis of a criminal record database (described in Appendix B) leads to a clustering of careers, that yields important information for the

formation of strategic policies [13]. Furthermore, the visualization can serve as a basis on which to track the movement in time of a certain perpetrator. A plotted line though the first few years of such a career could potentially be extended and a future class could be assigned to such an individual. Integration of this toolset into police software enables the automatic prediction of a criminal career each time a new entry for an offender is submitted. If the calculated career then falls within a preconfigured set of danger categories, an early warning will be sent to police officers in charge of for example organized crime, and action can be taken accordingly. In this chapter, we discuss the challenges in career prediction and the specific problems that arise with the implementation of software that transfers higher levels of knowledge discovery to prediction of individual cases.

8.2   Background

A new way of predicting the "movement in time" of items through predefined classes by analyzing their changing placement within a static, preconstructed two-dimensional visualization of other individuals was discussed in Chapter 7, [14]. It employs the visualization realized in previous steps within item analysis, rather than performing complex calculations on each attribute of each item. For this purpose a range of well-known mathematical extrapolation methods was adopted that were adapted to fit the need for two-dimensional extrapolation.

As a first step in this paradigm, the individual sequence to be extrapolated is selected and the sequence up to that moment is calculated for each time frame. Then, the distance between all time frames and all other sequences already in the visualization is calculated. Each time frame is then clustered in the original clustering, leaving all the existing elements in their original location. Note that this requires an iterative visualization and clustering method, like for example a randomized push-and pull-algorithm, rather than a Multi-Dimensional Scaling technique that calculates the entire visualization in a single pass. The resulting coordinates for each time frame can now be utilized by an extrapolation scheme.

The visual extrapolation paradigm has two distinguished advantages. On one hand, the results can immediately be visualized to the end-user, also enabling the user to derive how the results were reached in the first place, on the other hand, the computational complexity is very low, requiring only a few distances to be calculated and only a few elements to be plotted within an existing visualization. Note that the calculated x and y coordinates have no intended meaning; they serve only to give an idea where the item under consideration will be displayed relative to existing elements (initially also positioned on an arbitrary location).

The approach offers several different extrapolation schemes that are suitable for usage within a plane: second or third degree polynomial extrapolation, an x,y system with second or third degree polynomial extrapolation or spline extrapolation with straight line or polynomial continuation. Depending on the task domain the one with the best results should be selected. More information on visual extrapolation and the different extrapolation schemes can be found in Chapter 7.

8.3   Approach

The incorporation of a prediction tool into regular, police software comes with some time constraints; the early warning system should obviously not interfere with regular daily operations. Hence, the computational complexity of a prognosis tool should be minimal. Standard mathematical extrapolation methods that, for example, extrapolate every attribute separately, have difficulties complying with this demand. Next to that fact, a series of 0's will always be extrapolated to 0 by standard approaches. Some crimes, however, tend to have the property that they are far more likely to be committed after a few years of criminal activity, effectively creating a series of 0's that needs to be extrapolated by a number other than 0. This effectively renders standard extrapolation inapplicable here. Using the visualization as a depiction of domain knowledge, this problem can be dealt with effectively. We therefore resort to a more knowledge discovery oriented approach, specifically the two-dimensional extrapolation of visualization results mentioned above. As was shown in Chapter 7, the power of a clustering visualization resulting from career comparison can easily be used to reach accurate results without an abundance of computations. Using the temporal extrapolation method, only the *coordinates* of the careers in the visualization are used for the largest part of the algorithm. In Figure 8.1 we show the steps we take to come from the mentioned visualization to a decision on issuing a warning. Each step is described below.

Within this figure, only the boxed steps are taken every time the method is used to determine the development of a single career. The other steps are taken beforehand (the visualization and underlying clustering) or are predetermined by research (selection of extrapolation method, clustering reduction and reference points) or by domain experts (selection of danger categories).

As different steps of our approach take place in spaces with different dimensions, an overview is provided in Figure 8.2.

In this Figure, all steps within the high dimensional space are performed in the processes described in Chapters 4 and 5. These chapters end with the construction of a two-dimensional visualization (obviously a reduction to two-dimensional space) and a classification method that incorporates both the visualization and the raw high dimensional data from the database. The extrapolation methods in Chapter 7 built upon these purely two-dimensional results and combined them with the raw data to predict a future number of crimes. That information is then combined with the classification system in combined space to predict a future class.

8.3.1   Clustering Reduction and Extrapolation Selection

Although the speed of the temporal extrapolation scheme is high, the accuracy of the results can vary with the selection of a specific extrapolation method. It is important to determine the optimal option for extrapolation through field testing, which is explored in the experiments in Section 8.4.

Given the size of a typical criminal record database, ranging in the millions, a significant gain in speed could be realized by reducing the amount of offenders within the
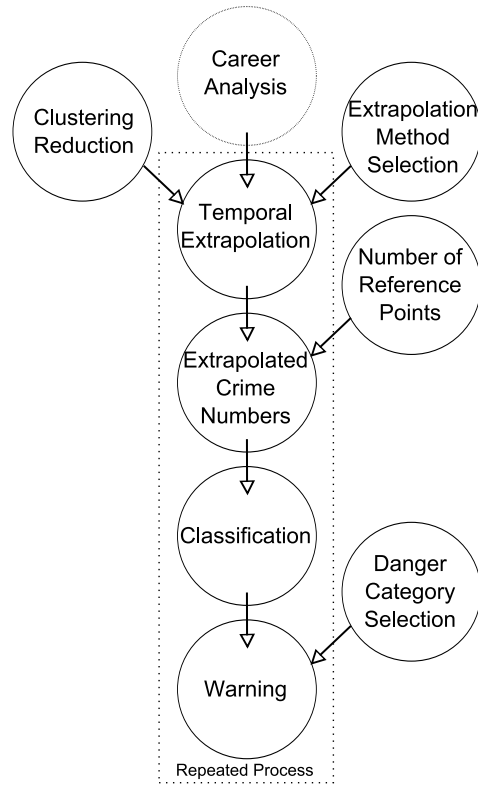
Figure 8.1: Approach for criminal career prediction

clustering. Naturally, care must be taken to realize this decrease without sacrificing the descriptiveness of the clustering itself, for example because certain clusters might lose enough "members" to cause a substantial reduction in "attraction" to newly entered individuals. Within our approach, the reduction is realized using a top down approach (as seen in Figure 8.3), that deletes items from a y-coordinate sorted list, keeping only every tenth individual.

This will reduce the amount of individuals with a factor 10, retaining a higher similarity with the original clustering than what would be the case if just the first tenth of the original database was used. The rationale behind this is that using this method, the same amount of individuals will be removed from every "height" in the image, preserving the shape of the image as much as possible. A strong argument against simple database removal is the fact that the database could be sorted in many ways, both implicitly and explicitly, without the user's knowledge. Therefore, removal of specific parts of the database could have unintended effects on the outcome, creating a "polluted" clustering. If necessary this process can be repeated to create even smaller clustering sizes. The effects of this reduction are analyzed and discussed in Section 8.4.
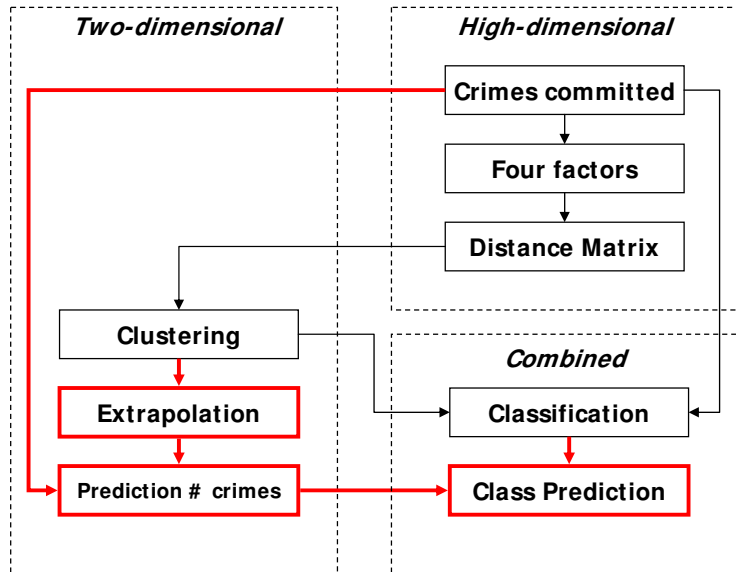
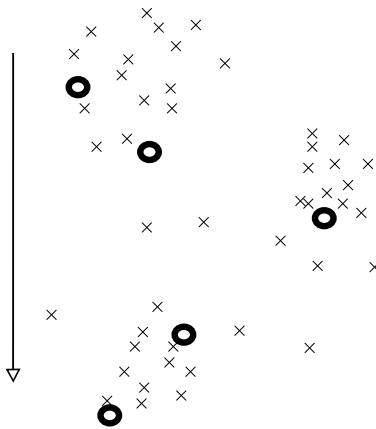Figure 8.2: Different prediction steps take place in different dimensions

Figure 8.3: Example for clustering reduction, shown in a visualization; only the circles are kept

8.3.2   Further Steps

A typical temporal extrapolation for an emerging minor career criminal is found in Figure 8.4. Here each cross represents a year of known criminal activity and the dotted line denotes the expected continuation of the sequence or criminal career.

Domain experts can easily scan such images and conclude what class of criminal ca-

reer this individual will probably belong to (minor career criminal in the case of the situation in Figure 8.4). If incorporation of the prediction is wanted, however, it is necessary to classify the offender under observation. Hence, we need to automatically calculate its attributes or the number of different crimes in his or her future. This can be accomplished by selecting a number of *reference points* close to the extrapolated line, averaging over their respective crime numbers to reach the expected crime data for the current offender. It was suggested in Chapter 7 that reference points closest to the last known year receive a higher weight in this process, following
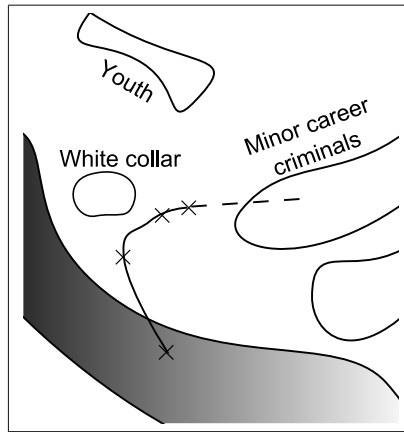

Figure 8.4: Example of temporal extrapolation for criminal careers

$$Attrib_j(new) = \frac{2}{r+1} \cdot \sum_{i=1}^{r} (r - i + 1) Attrib_j(i),$$

where r is the amount of reference points and j is one of the crime types. This process is illustrated in Figure 8.5.

Of course, the number of reference points to use is a matter of accuracy versus time complexity. Looking up a large number of reference points in a database can be very time consuming, but selecting a small amount can cause accuracy to drop greatly. Selection of the right number of reference points can therefore contribute to successful implementation of this tool and is discussed in Section 8.4.

Now that the possible future crime numbers are calculated, the individual can easily be classified in one of the categories. Domain experts can select which categories to monitor based upon their domain knowledge, the specific needs of their own district or the specific tasks of policing they are involved in. A warning can be issued on their computer every time a new individual falls within one of the selected danger categories.
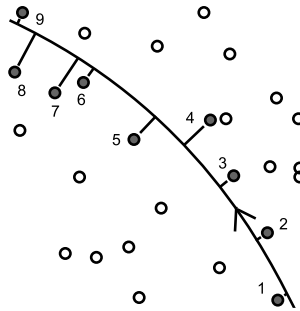
Figure 8.5: Selecting points with the shortest distance to the extrapolation line

8.4   Experimental Results

A number of experiments was performed to both reveal acceptable values for the needed parameters and test the validity of the approach as a whole. For this purpose the criminal record database described in Appendix B, was used to run the algorithm with a number of different settings. It contains approximately one million offenders and their respective crimes, of which 10% could be said to have finished their careers (no reported crimes for the last 10 years). Although this selection method is coarse, people can be incarcerated or they were simply not caught, it can still be used as a validation group.

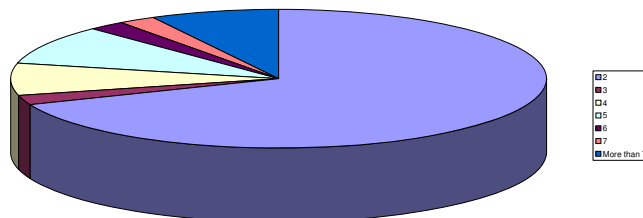Of these 10% the career length is distributed as is displayed in Figure 8.6.

Figure 8.6: Career length distribution within finished career subset

As a first step we clustered n criminals on their (in most cases) finished criminal careers, i.e., all the crimes they committed throughout their careers. In our first test, the number of reference points was set to $r = 100$.

A ten-fold cross validation was used to calculate the accuracy of a certain method: for each fold, one tenth of the population was used as a test group, where the other careers were used in the creation of the clustering. All the careers in the population were "cut off" after 3 or 4 years. For each of those careers, the future crime number was predicted and compared with the actual end-values of their careers. The accuracy for one career

Table 8.1: Time complexity and accuracy comparison between different methods and clustering sizes using $r = 100$ reference points

Clustering Size (n)	1,000,000 (all)	100,000	10,000	1,000
Second degree polynomial	961	98	10.4	1.0
	79.1%	77.9%	75.7%	61.3%
Third degree polynomial	965	101	10.5	1.1
	79.3%	78.2%	75.8%	62.0%
***x,y* system (second degree)**	971	104	11.3	1.9
	81.5%	81.1%	79.9%	66.6%
***x,y* system (third degree)**	973	105	11.7	2.1
	87.5%	87.3%	86.3%	74.4%
Spline (straight line)	982	113	**22.0**	13.4
	88.7%	88.2%	**87.3%**	74.3%
Spline (polynomial)	983	114	22.1	13.4
	79.6%	78.4%	76.9%	63.7%

prediction will then be described by the average similarity between all 50 predicted and actual crime numbers. The accuracy of the method using these settings is then described by the mean of all averages.

For all methods, a time factor was also calculated, This time factor represents how much time was consumed, using this method within this clustering size, relative to the fastest method-size combination (which is set to 1). On standard equipment this method needed an average of 60 ms.

The results are presented in Table 8.1, where the top box of every cell describes the time factor and the bottom box contains the calculated accuracy.

From the above presented results, the conclusion can be drawn that the *x,y* system with a third degree polynomial and the spline with straight line extrapolation largely outperform the other methods, especially when the amount of careers in the clustering decreases. The spline performs slightly better than the system methodology.

The decrease in clustering size appears to have a minor effect on accuracy, lowering it only marginally while reducing the clustering size with a factor 100. A steep drop, however, occurs when the size is lowered to 10,000 careers. Apparently, the quality of the clustering reaches a critical low, to make a reliable prediction.

Given the time demands put on this application, the best choice would be to overlay a straight line spline extrapolation on a 10,000 size clustering (bolded option in Table 8.1). The accuracy of this option can be considered high and provides a solid foundation for incorporation, while allowing for fast calculation (approximately 1.3 seconds).
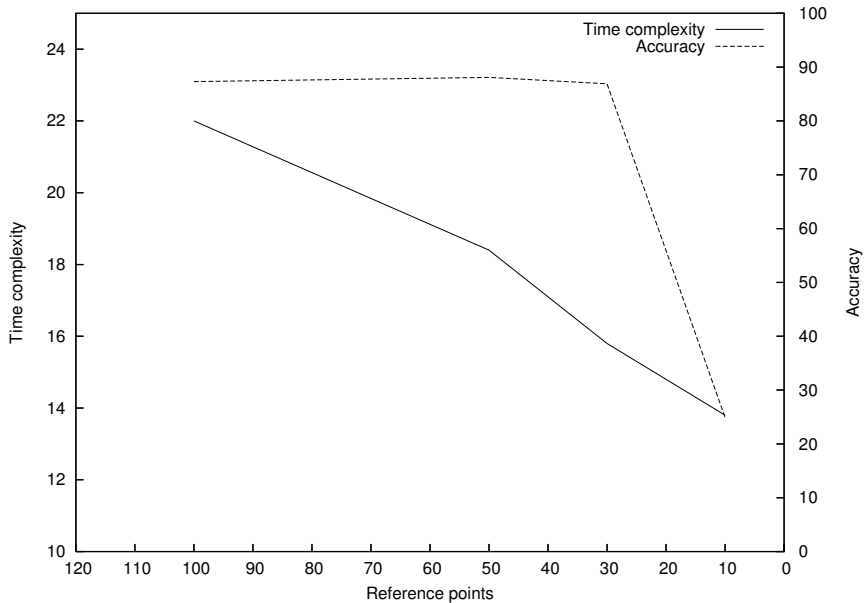
	100	50	30	10
Time complexity	22.0	18.4	15.8	13.8
Accuracy	87.3%	88.1%	86.9%	24.9%

Figure 8.7: The relation between accuracy and time complexity when reducing the number of reference points

Potentially, an even greater gain in time can be reached by reducing the number of reference points, thus reducing calculation of the different averages. Figure 8.7 describes the effects on accuracy and time complexity of reference point reduction, using the optimal solution described above.

Again, the reduction in information does not (necessarily) lead to decrease in quality. Reducing the number of reference points to 30, slightly lowers the accuracy with only 0.4 percentage points. Furthermore, a reduction to 50 leads to an increase of 0.8 percentage points, probably because the selection method selects careers that are simply too far away from the extrapolated line to contribute positively to the calculation of crime numbers. A steep decline can be seen with the reduction of reference points below 30. Depending on the need for speed-up or the quality of the prediction any number between approximately 50 and 30 can be selected.

It may also be of interest to see the influence of the amount of years of criminal activity that are already known on the result. In the example above, either 3 or 4 years were selected. In Figure 8.8 we show how the accuracy depends on the availability of realized
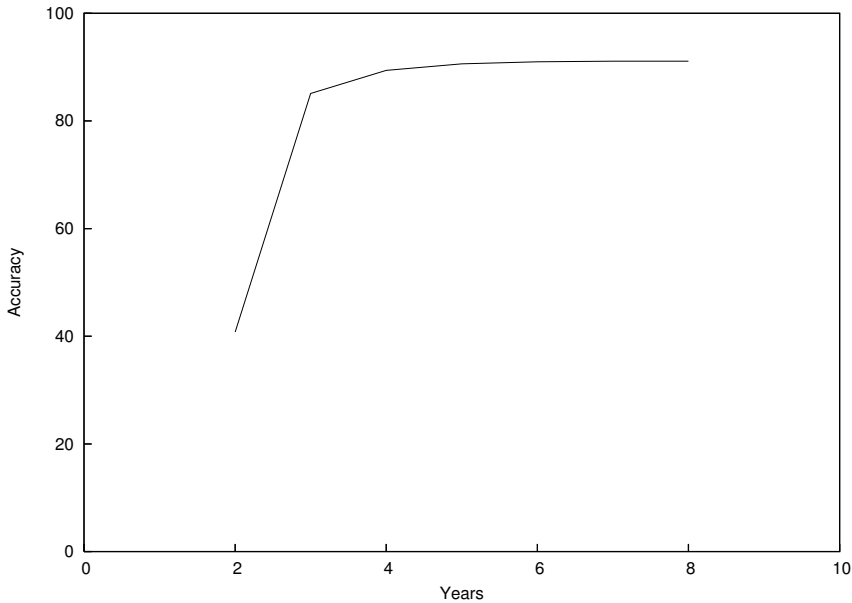
Figure 8.8: Accuracy as a function of known years

behavior. For this experiment we used the straight line spline extrapolation, a clustering of size 10,000 and 50 reference points. Only individuals with more than 10 years of activity in total were selected for testing the extrapolation accuracy in this experiment.

As can clearly be observed in the graph, a career of which less than 3 years of activity are already recorded can not be predicted accurately. However, the results cease to improve with the addition of more than 5 years of activity. As could be expected (2 points are best extrapolated by a straight line, which in most cases does not validly predict a career), prediction efforts should start only for criminals who are active for more than 2 years, in which case an accuracy of 88% can be reached within 1.2 seconds.

It might be possible that offenders are already classified in their respective end-classes after examination of the data that is already known. If this is the case in a large portion of the database for the amount of reference years used within our approach, the necessity of extrapolation and the expressiveness of its accuracy would greatly decrease. However, Figure 8.9 shows that only a reasonably small percentage of offenders have reached their respective end-classes after the amount of reference years we used to reach the results described above.

Combined with the results displayed in Figure 8.8, the added value by using our prediction method over the simple assumption that criminal careers do not change after a set number of reference years is displayed in Figure 8.10.

Within this figure, the top portion of the graph denotes the gain that is reached using our method specified per amount of reference years used. It is clear that using two-
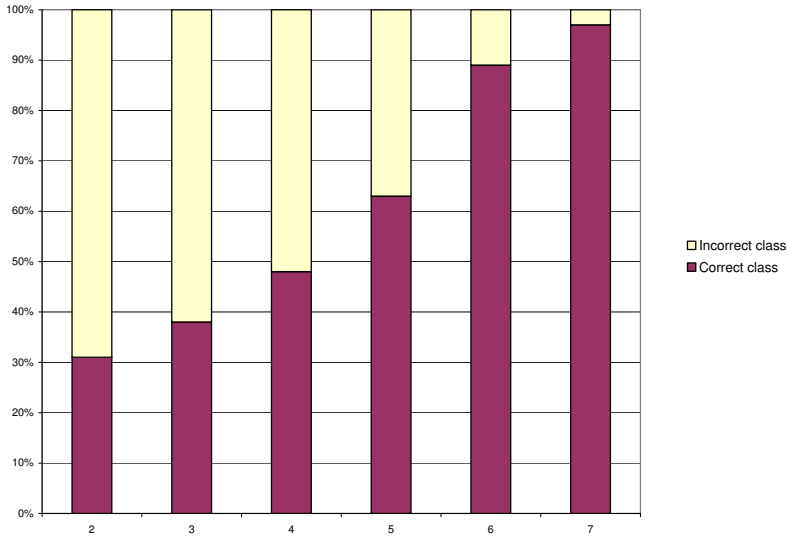
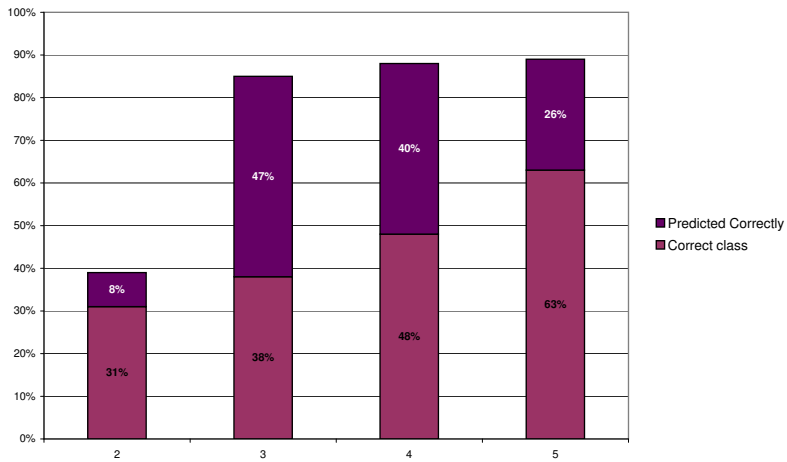Figure 8.9: Accuracy without prediction after known years

Figure 8.10: The additional accuracy reached through prediction

dimensional extrapolation, the prediction of careers can be greatly enhanced within the [2...5] reference year domain, especially when using 3 or 4 reference years. When 6 years of activity are known the added value of our approach approaches zero and after 7 years of activity the prediction yielded by the extrapolation becomes less accurate than

the actual class the offender under consideration was already placed in.

Summarizing all of the above results, a two years ahead advantage in the prediction of criminal careers can be reached with two-dimensional straight line spline extrapolation, using 3 or 4 reference years and between 30 and 50 reference points.

8.4.1 Accuracy of a Single Prediction

Obviously, the overall accuracy of 89% that can be reached is not representative for the extrapolation of an individual sequence of events, meaning that an individual career can be predicted with an 89% certainty, but that results may vary greatly, depending on the type of offences, the amount of crimes or the shape of the extrapolation line. Therefore, further insights into the accuracy of a single prediction are necessary when investigating a single criminal career. In such a case, a percentage denoting the chance that this specific prediction is correct is as a measure or reliability that law enforcers can use.

For this purpose, we propose to use the shape of the extrapolation curve as an indication for the reliability of its prediction, assuming that a curve with a variance in its direction has a less reliable outcome than a curve that slowly changes towards its end-class. Within this effort, we will be using the modification of the derivative of the curve as an indication for directional changes. Since we are using the straight line spline extrapolation, only the interpolated part of the curve needs to be observed (the derivative does not change after the last known data point), providing us with a fixed domain to be investigated. This allows for a change detection based upon small changes in t, that, knowing its start- and end values, can be set at a static interval. A natural way to denote the modification of the derivatives is in degrees of the smallest angle between the old and the new tangent, which should yield the correct result as long as Δt is small enough and the curves do not change direction too suddenly.

Figure 8.11 describes the mathematical situation when calculating the distance between two tangents. Figure 8.12 shows a situation where the calculation is somewhat different.

The two lines in these figures represent the tangents, angle $\angle xy$ is the target angle to be calculated and $\angle x$ and $\angle y$ are the angles between the tangents and the horizontal assisting line. Using this situation,

$$\angle xy = \begin{cases} \tan^{-1} \max(|\alpha_x|, |\alpha_y|) - \tan^{-1} \min(|\alpha_x|, |\alpha_y|) & \alpha_x \alpha_y \geq 0 \\ \tan^{-1} |\alpha_x| + \tan^{-1} |\alpha_{y|} & \alpha_x \alpha_y < 0 \end{cases},$$

where α_x and α_y are the slopes of tangents x and y, respectively. Within our experiments we used three reference years and set Δt on 0.005 ($t \in [0, 3]$), having 600 intervals, each with its own $\angle xy_t$. The directional change of interpolation curve c, Δ_c can now be calculated as follows:

$$\Delta_c = \sum_{t=0}^{600} \angle xy_{\frac{t}{600}}$$

In order to evaluate the fitness of this function as an indicator for prediction reliability a Δ_c was added to every possible prediction. As can be seen in Figure 8.13, Δ_c is
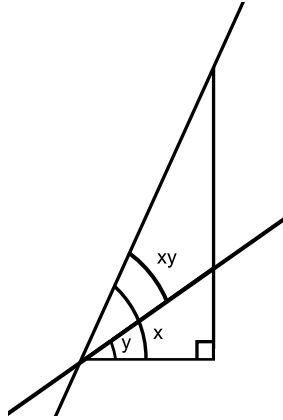
Figure 8.11: The mathematical situation when calculating the angle between two tangents
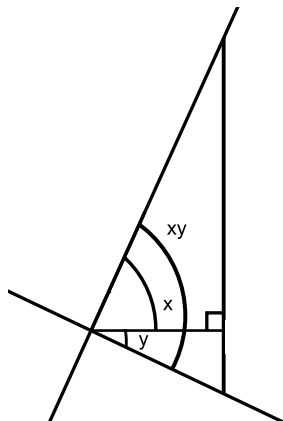

Figure 8.12: Another possible calculation of the angle between two tangents

approximately normally distributed with a mean around 90 and a standard deviation of about 50 degrees. Each occurrence is calculated for a 10 degree interval and is specified in thousands.

According to our previous results, the average of 89% accuracy should be reached at approximately 90 degrees curvature. A closer observation of these 10 degree intervals should reveal if there is any connection between Δ_c and the accuracy of the prediction. If we calculate the percentage of correctly predicted offender classes for each interval separately, and add a trend line, a relation could be established together with its average error. Figure 8.14 shows the result of this investigation.

In this figure, the vertical axis is the accuracy and the horizontal axis is Δ_c. As the
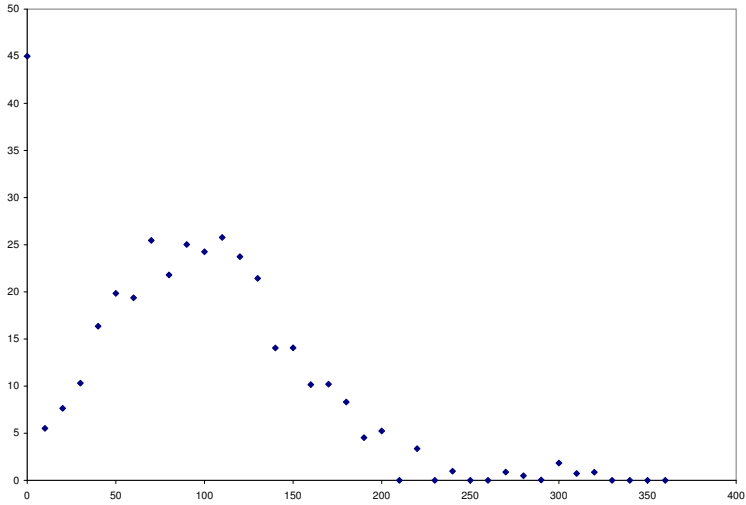
Figure 8.13: Scatterplot of complete population test of Δ_c
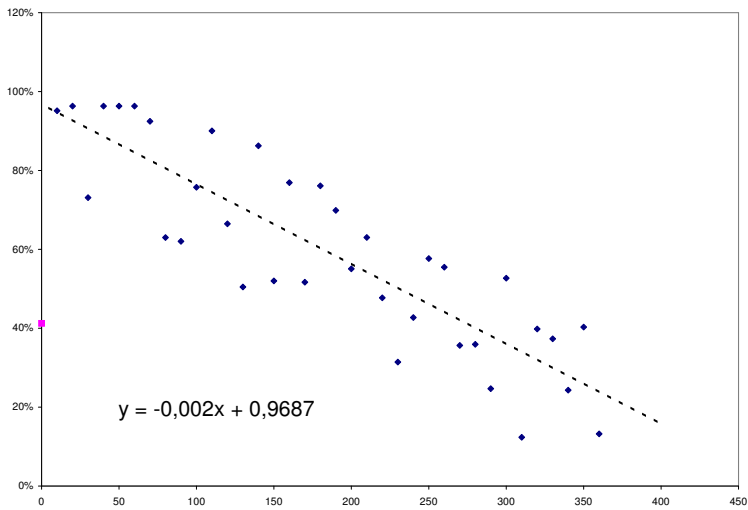

$$y = -0{,}002x + 0{,}9687$$

Figure 8.14: Scatterplot of complete population test of prediction accuracy with added trend line

trend line shows the accuracy of a single prediction is approximately linearly dependent on its Δ_c according to:

$$\text{Accuracy} = -0.002\Delta_c + 97,$$

where Accuracy is in percentages. The error associated with this trend line is 8 percentage points on average, resulting in an approximate 90% reliability of reliability prediction.

As a conclusion we can state that it is reasonably possible to associate a reliability with a single criminal career prediction using its curvature as the main indicator.

8.5   Conclusion and Future Directions

In this chapter we demonstrated the applicability of temporal extrapolation for the extrapolation of criminal careers. This method assumes that the visualization of a clustering inherently contains a certain truth value that can yield powerful results but reduces time complexity. We superimposed an extrapolation on an existing visualization of criminal careers and performed a series of tests that determined the speed and accuracy of the approach as well as the values of the necessary parameters. A clustering reduction was also performed to speed up calculation of a criminal career prediction.

It turns out that a clustering size of 10,000 criminal careers from the database can serve as a solid basis for extrapolation. If this extrapolation is accomplished by a spline extrapolation with straight line continuation and 50 reference points, accuracy can reach 88%. We also implemented and tested an approach that determines if an individual career is predicted reliably.

As time constraints are essential to successful implementation within actual police software services, it was important to reach significant gains in computational complexity. As an end-result, all necessary tasks to be repeated for every offender to be analyzed, can be completed in approximately 1 second.

Next to the fact that predictions can immediately be visualized to police end users because of their visual nature, offender's careers can be predicted with very high accuracy in a single second. These properties make the method very well suited for incorporation in background processes at police stations, allowing alerts to be sent to dedicated machines.

A weakness to the approach, that is native to extrapolation problems, is that the lack of enough information can cause very unreliable predictions, resulting in a minimum of 3 time frames of activity before anything valuable can come out of the extrapolation process. Unfortunately, data in the Netherlands is collected on a year-by-year basis, effectively establishing the demand that only third or higher year offenders can have their careers predicted.

The judicial constraints and privacy issues concerning the material in this chapter are discussed in Appendix A.

Future research will aim at reaching even higher accuracy values by improving the selection of reference items close to the extrapolation line. A triangular shape can, among others, be employed, that selects more reference points further away (more description of the future), or closer (more reliability) from the last known data point. Also, a search for common subcareers could be performed, that could reveal subcareers that "define"

certain classes. These subcareers, especially if they occur in the beginning of a career, could possibly improve both the speed and accuracy of future career prediction.