An Early Warning System for the Prediction of Criminal Careers

Tim K. Cocx, Walter A. Kosters and Jeroen F.J. Laros

Leiden Institute of Advanced Computer Science, Leiden University, The Netherlands
tcocx@liacs.nl

Abstract. 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 2-dimensional extrapolation on a static clustering, 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 notifies police officials. We outline the implementation of such a tool and highlight test results on the Dutch National Criminal Record Database. Certain problematic situations, like time constraints, privacy concerns and reliability issues, are also discussed.

1 Introduction

One of the common factors in today's formation of policy is that of prediction. Usually, the ability to successfully estimate future behavior of individuals, has a fundamental role in the development of any corporate (communication) strategy. Prediction of later revenues might safeguard investments in the present. Large corporations invest heavily in this kind of activity to help focus attention on possible events, risks and business opportunities. Such work brings together all available past and current data, as a basis on which to develop reasonable expectations about the future. The increased availability of useful data has led to the development of more automated processes that enable the acquirement of such prognoses by constructing knowledge from different digital sources.

Current and future behavior are regularly described by employing a number of different algorithmic computer constructs. This branch of data mining, known as *predictive modeling*, provides predictions of future events and may be transparent and readable in for example *rule based systems* and opaque in others such as *neural networks*. Usually, *detection theory* [1] serves as a foundation whereupon these models are selected. They often employ a set of classifiers to determine the probability of a certain item belonging to a dataset, like, for example, the probability of a certain email belonging to the subset “spam”. These methods are well suited to the task of predicting certain unknown attributes of an individual by analyzing the available attributes, for example, estimating the groceries an individual will buy by analyzing demographic data. It might, however, also be of interest to predict shopping behavior based upon past buying behavior alone, thus predicting the future development of a certain sequence of recent events. Examples of this are the prediction of animal behavior when their habitats undergo severe changes, in accordance with already realized changed behavioral patterns, or the prediction of criminal careers based upon earlier felonies.

This paper 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. The national police annually extracts information from digital narrative reports stored throughout the individual departments and compiles this data into a large and reasonably clean database that contains all criminal records from the last decades. Analysis of this database leads to a clustering of careers, that yields important information for the formation of strategic policies [8]. Furthermore, the clustering 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 paper, 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.

2 Background

The number of data mining projects in the law enforcement area is slowly increasing. Both inside and outside of the academic world large scale projects are underway. In this section we discuss related work and provide an overview of our approach.

2.1 Related work

One of the larger academic projects, known as COPLINK, is a police-university collaboration in Arizona, where work has been done in the exploitation of data mining for cooperation purposes [4], the field of entity extraction from narrative reports [5], and social network analysis [6,16]. The FLINTS project and FinCEN [12] aim at revealing links between crimes and criminals and to reveal money laundering networks by comparing financial transactions. Also, Oatly et al. [14] linked burglary cases in the OVER project. Clustering techniques are widely used in the law enforcement arena as well, like for example by Adderly and Musgrove [2], who applied clustering techniques and Self Organizing Maps to model the behavior of sex-offenders, and by Cocx et al. [7, 8] who made an attempt at clustering criminal investigations to reveal what offenses were committed by the same group of criminals, revealed links between crimes and demographic data and the above mentioned automated analysis of criminal careers.

2.2 Criminal Careers

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. [3]: 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. Blumstein et al. [3] also look at one specific type of career, that of career criminals; offenders who make crime their life-time profession. A distinction can be made between minor and heavy career criminals, respectively those who are guilty of petty crimes, like theft or fencing, most often the result of addiction disorders, and those involved in more severe felonies, often through membership in a crime syndicate.

The dismantlement of crime syndicates ranks seventh in the current list of top priorities of the FBI [11]. 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.

Analysis The analysis of criminal careers is usually done through the extraction and aggregation of knowledge from a criminal record database, commonly containing an unexpectedly large portion of the population (6% in the Netherlands). It can best be accomplished by adapting a *multiset* approach. A multiset is a collection where each element can occur more than once. The set of all distinct elements in that multiset is called its *underlying set*. Figure 1 describes the relation between the two and shows how we employ it to represent a criminal's activities in a single year.

The multiset representation offers advantages, most notably the availability of standard approaches to compare multisets and calculate distances between them. Kosters and Laros [13] devised a distance function for multisets that generalizes well-known distance measures and allows for the incorporation of weights for an element, e.g., element x counts twice as much as element y. This metric contains a customizable function f that can be adapted to fit specific knowledge domains. A tailored f was developed that calculates the distance between two crime-multisets.

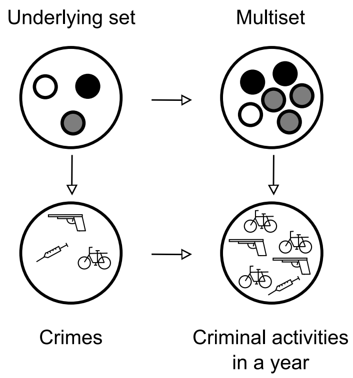


Fig. 1. A multiset representation of a criminal profile in a single year

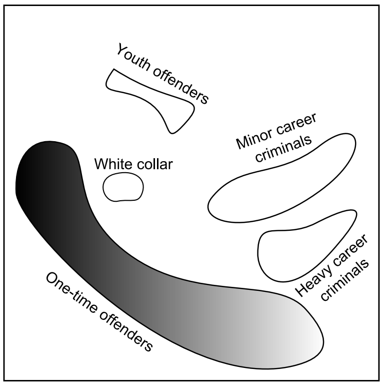


Fig. 2. Impression of a clustering of criminal careers

Instead of a strict number-wise comparison between years (comparing the first year of criminal a with the first year of criminal b, the second year of a with the second year of b, etc.) a novel *alignment* of the mentioned multisets was proposed. This method strives for an optimal automated matching of years, using the distance measure described above, dealing penalties for every mutation needed, which enables a police analyst to better cope with situations like captivity, forced inactivity or unpenalized behavior. Figure 3 shows the intuition behind such a method.

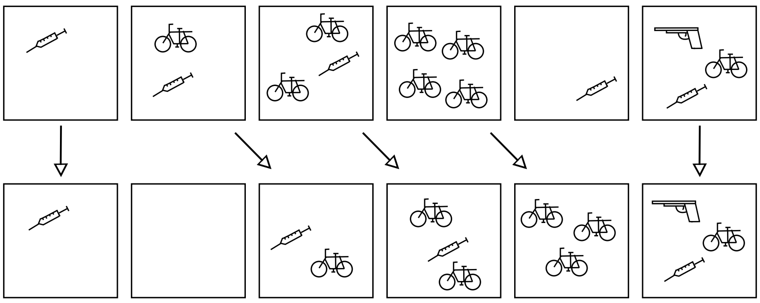


Fig. 3. Two criminal careers who's similarity is revealed by alignment

Based upon the distance between two aligned criminal careers, a distance matrix was constructed, containing the distances between all couples of criminal careers. This matrix was then visualized into a 2-dimensional clustering using some kind of *Multi-Dimensional Scaling* [10]. An impression of the results can be seen in Figure 2.

The ultimate goal, predicting if certain offenders are likely to become (heavy) career criminals, was not realized or elaborated upon. Further investigation of this possibility led to the need of a good 2-dimensional visual *extrapolation* system.

2.3 2-Dimensional Extrapolation

A new way of predicting the "movement in time" of items through predefined classes by analyzing their changing placement within a static, preconstructed 2-dimensional clustering of other individuals was discussed in [9]. 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 2-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 unit. Then, the distance between all time units and all other sequences already in the clustering is calculated. Each time-unit is then clustered in the original clustering, leaving all the existing elements in their original location. Note that this requires an iterative 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 unit 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 clustering. Note that the calculated x and y coordinates have no actual 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 [9].

3 Approach

The incorporation of a prediction tool into regular, police software comes with some serious 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

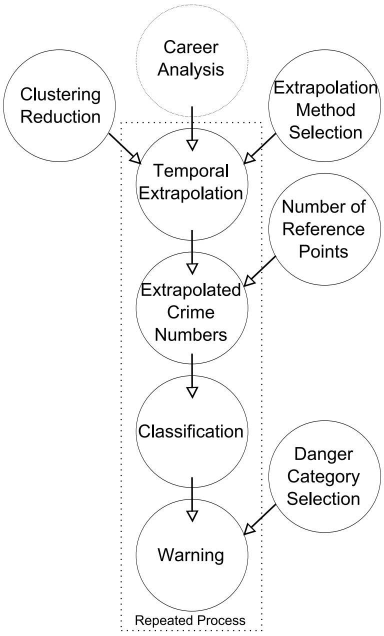


Fig. 4. Approach for criminal career prediction

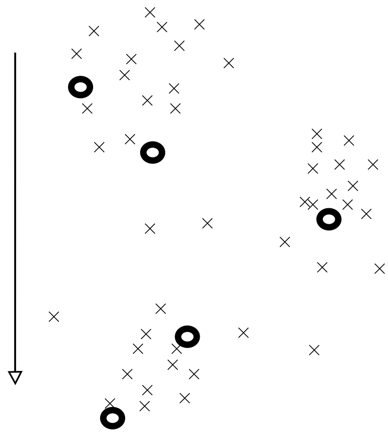


Fig. 5. Example for clustering reduction; only the circles are kept

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 methods useless. Using the clustering 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 2-dimensional extrapolation of clustering results mentioned above. According to [9] 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 clustering are used for the largest part of the algorithm. In Figure 4 we show the steps we take to come from the mentioned clustering 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 clustering) or are predetermined by research (selection of extrapolation method, clustering reduction and reference points) or by domain experts (selection of danger categories).

3.1 Clustering Reduction and Extrapolation Selection

Although the speed of the temporal extrapolation scheme is very high, the accuracy of the results can vary with the selection of a specific extrapolation method. It is obviously important to determine the optimal option for extrapolation through field testing, which is explored in the experiments in Section 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 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 5), 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.

3.2 Further Steps

A typical temporal extrapolation for an emerging minor career criminal is found in Figure 6. Here each cross represents a year of known criminal activity and the dotted line denotes the expected continuation of the sequence or criminal career.

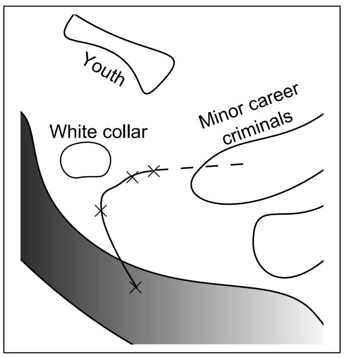


Fig. 6. Example of temporal extrapolation for criminal careers

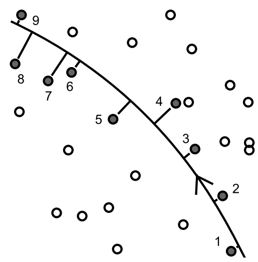


Fig. 7. Selecting points with the shortest distance to the extrapolation line

Domain experts can easily scan such images and conclude what class of criminal career this individual will probably belong to (minor career criminal in Figure 6). 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 [9] that reference points closest to the last known year receive a higher weight in this process, following

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

where r is the amount of reference points and j is one of the crime types. This process is illustrated in Figure 7.

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 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 own computer every time a new individual falls within one of the selected danger categories.

4 Experiments

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 Dutch National Criminal Record Database was used to run the algorithm with a number of different settings. This database is available for scientific research at CBS (Statistics Netherlands) in an anonymized version. It contains approximately one million offenders and their respective crimes (approximately 50 types), 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.

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: One tenth of the population was used as a test group, where the other careers were in the clustering. All the careers in the population were “cut off”

Table 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%

after an evenly distributed 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 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.

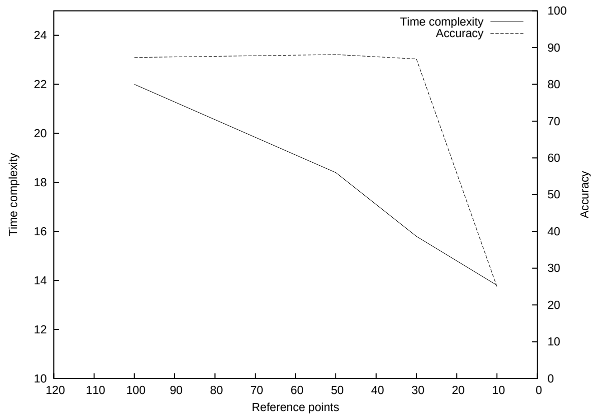
For all methods, a time factor was also calculated, This time factor represents how much time was consumed, using this method within this clustering size, relative to the fastest method-size combination (which is set to 1). On standard equipment this method needed an average of 60 ms.

The results are presented in Table 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 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%

Fig. 8. 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 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 9, we show how the accuracy depends on the availability of realized 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

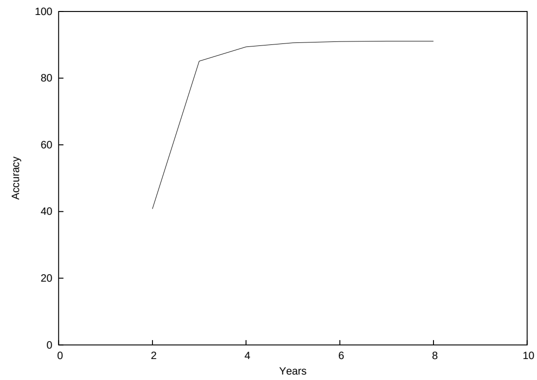


Fig. 9. Accuracy as a function of known years

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.

5 Discussion

In this paper 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 dramatically. We superimposed such an extrapolation on an existing clustering 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 Dutch National Criminal Record Database can serve as a solid basis for extrapolation. If this extrapolation is accomplished by a spline extrapolation with straight line continuation (see [9]) and 50 reference points, accuracy can reach 88%.

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 units 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.

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 sub-careers could be performed, that could reveal sub-careers 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.

5.1 Reliability and Privacy

Naturally, approaches such as this one come with some reliability and privacy concerns. Most data mining applications concern quests for truth or the discovery of general knowledge, rather than the translation of such newly discovered patterns or descriptions to individual cases. This approach is an exception to this situation, which raises some questions about the statistical validity of the results when applied in real life situations, especially areas as important as that of law enforcement.

This approach is best compared to the process of discovering suspicious financial transactions like for example in [12]. These methods also enforce criteria from a knowledge discovery support system on individual financial transactions in order to find outliers that might indicate for example money laundry activities. Naturally, not all transactions that yield warnings are actually fraudulent, but as long as the expected chance a warning is actually correct is reasonably high, the usage of such an decision support system is warranted as long as each warning is reviewed by experts before action is undertaken. The same should apply to the usage of our approach. Still it remains an important warning to police officers working with the software that for every 88 careers the system predicts correctly, there are 12 careers predicted incorrectly with varying margins.

Another important concern is that of privacy. Constant electronic monitoring of one's activities can be seen as an invasion on one's personal privacy as stressed in, for example, [15]. However, in contrast with, among others, the above mentioned financial monitoring systems, that monitor *all* transactions done by *everybody*, only the people who have been active criminals for more than three years are under surveillance and even then they are only electronically processed every time they commit a new crime. Therefore, the approach as described in this paper poses a minimal threat to the privacy experience of the regular population.

Acknowledgment

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

References

1. H. Abdi. Signal detection theory. In N.J. Salkind, editor, *Encyclopedia of Measurement and Statistics*. Thousand Oaks (CA): Sage, 2007.
2. R. Adderley and P. B. Musgrove. Data mining case study: Modeling the behavior of offenders who commit serious sexual assaults. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'01)*, pages 215–220, New York, 2001.
3. A. Blumstein, J. Cohen, J. A. Roth, and C. A. Visher. *Criminal Careers and "Career Criminals"*. The National Academies Press, 1986.
4. M. Chau, H. Atabakhsh, D. Zeng, and H. Chen. Building an infrastructure for law enforcement information sharing and collaboration: Design issues and challenges. In *Proceedings of The National Conference on Digital Government Research*, 2001.
5. M. Chau, J. Xu, and H. Chen. Extracting meaningful entities from police narrative reports. In *Proceedings of The National Conference on Digital Government Research*, pages 1–5, 2002.
6. H. Chen, H. Atabakhsh, T. Petersen, J. Schroeder, T. Buetow, L. Chaboya, C. O'Toole, M. Chau, T. Cushna, D. Casey, and Z. Huang. COPLINK: Visualization for crime analysis. In *Proceedings of The National Conference on Digital Government Research*, pages 1–6, 2003.
7. T.K. Cocx and W.A. Kosters. A distance measure for determining similarity between criminal investigations. In *Advances in Data Mining, Proceedings of the Industrial Conference on Data Mining 2006 (ICDM2006)*, volume 4065 of *LNAI*, pages 511–525. Springer, 2006.
8. T.K. Cocx, W.A. Kosters, and J.F.J. Laros. Enhancing the automated analysis of criminal careers. In *SIAM Workshop on Link Analysis, Counterterrorism, and Security 2008 (LACTS2008)*, 2008.
9. T.K. Cocx, W.A. Kosters, and J.F.J. Laros. Temporal extrapolation within a static clustering. In *Foundations of Intelligent Systems, Proceedings of ISMIS 2008*, volume 4994 of *LNAI*, pages 189–195. Springer, 2008.
10. M.L. Davison. *Multidimensional Scaling*. John Wiley and Sons, New York, 1983.
11. The FBI strategic plan, 2004–2009, `http://www.fbi.gov/`.
12. H.G. Goldberg and R.W.H. Wong. Restructuring transactional data for link analysis in the FinCEN AI system. In *Papers from the AAAI Fall Symposium*, pages 38–46, 1998.
13. W. A. Kosters and J. F. J. Laros. Metrics for mining multisets. In *Proceedings of the Twenty-seventh SGAI International Conference on Artificial Intelligence SGAI2007*, pages 293–303, 2007.
14. G.C. Oatley, J. Zeleznikow, and B.W. Ewart. Matching and predicting crimes. In *Proceedings of the Twenty-fourth SGAI International Conference on Knowledge Based Systems and Applications of Artificial Intelligence (SGAI2004)*, pages 19–32, 2004.
15. B. Schermer. *Software Agents, Surveillance, and the Right to Privacy: A Legislative Framework for Agent-enabled Surveillance*. PhD thesis, Leiden University, 2007.
16. Y. Xiang, M. Chau, H. Atabakhsh, and H. Chen. Visualizing criminal relationships: Comparison of a hyperbolic tree and a hierarchical list. *Decision Support Systems*, 41(1):69–83, 2005.