Data mining for intelligence led policing

ir. RCP van der Veer
Sentient

H.T. Roos MSc
Amsterdam Police force

A. van der Zanden MSc
Amsterdam Police force

Singel 160, 1015 AH, Amsterdam, The Netherlands
+31 (0)20 5 300 325

rvdveer@sentient.nl

ABSTRACT

The benefit of data mining for police seems tremendous, yet only a few limited applications are documented. This paper starts with describing the implementation problems of police data mining and introduces a new approach that tries to overcome these problems in the form of a data mining system with associative memory as the main technique. This technique makes the system easier to use, allows uncomplicated data handling and supports many different data types. Consequently, data preparation becomes easier and results contain more information. A number of Dutch police forces have already been using this system for several years with over 30 users. Since the analytical process within the police is very knowledge-intensive, a high level of domain expertise is essential, which makes it harder to find a police data miner with sufficient domain knowledge plus technical skills in the area of databases, statistics and data mining. The police domain also has data quality issues and a very diverse information need. This is why the system design tries to reduce the need for technical skills as much as possible by working with one standard datawarehouse, techniques that can be configured automatically and active user guidance. The ease of use is also ensured by integrating many tools and techniques from statistics, business intelligence and data mining into one interactive environment that does not require the analytical process to be designed beforehand. Instead, the analysis is performed through step by step interaction. This paper discusses the benefit of police data mining, the design of the system, a number of practical applications, best practices and success stories. Experiments have shown a factor 20 efficiency gain, a factor 2 prediction accuracy increase, a 15% drop in crime rate, and 50% more suspect recognition.

General Terms

Algorithms, Management, Measurement, Design, Experimentation, Security, Human Factors.

Keywords

crime, police, data mining, prediction, GIS, hot spots, spatial, public safety, analysis

1. INTRODUCTION

In this digital era, police forces have access to a rapidly growing amount of data. Combined with the dynamic nature and complexity of criminal behavior, this sets the stage for successful data mining applications. Still, examples of consistently used police data mining implementations are scarce. In this paper we discuss the practical application of a standard police data mining system as used by a growing number of Dutch police forces. The system has been developed in a group effort of data mining software company Sentient and the Dutch police forces *Amsterdam-Amstelland, Midden en West Brabant* and *Brabant-Noord.* It consists of an integrated data mining tool called *DataDetective*, which is being developed and applied by Sentient since 1992, and an extensive datawarehouse containing data from various police systems and external sources, such as weather data, geographical data and socio-demographics. During the eight years of practical application, the data mining system has continuously been evaluated, improved and extended. The examples given in this paper illustrate how data mining now plays an important role at operational, tactical and strategic levels of decision making.

The main key to success has been a strong focus on simplicity. After a short training, selected police officers can quickly discover patterns and trends, make forecasts, find relationships and possible explanations, map criminal networks and identify possible suspects. Expert knowledge on statistics or data mining is not required. Additionally, a much larger audience for data mining results is reached through weekly reports containing statistics, prediction maps, crime clusters, trends and lists of suspects. These reports are automatically produced by the data mining system.

In Section 2 we start off by discussing why it is important for police to apply data mining, followed by a description of the shortcomings of traditional systems in section 3. Section 4 then describes the system we have built including how its use is organized. In section 5 we discuss various applications of the system, followed by a number of success stories in section 6. We finish with discussions in section 7 about the added value of data mining, the challenges for implementation and future work.

2. POLICE NEED FOR DATA MINING

The increasing adoption of the *Intelligence-Led Policing* model [1] puts analysis at the heart of operational, tactical and strategic decision-making. In this model, intelligence serves as a guide to operations, rather than the reverse. Therefore, it is now more important than ever to find out how data mining can help create better understanding and predictions.

Traditionally, police systems focus on small parts of the available data (e.g. year, month, type of crime) for a specific purpose (e.g. monitoring crime rates for strategy). Without data mining, the amount of data used in analysis is limited by the time that analysts have to go through it, step by step. It is simply unfeasible to analyze all the potentially useful data by hand. However, for many policing problems it is important to use as much data as possible, to be able to explain, understand, link and predict. The explanation of a phenomenon (e.g. the sudden increase of pickpocket activity) typically lies in small details, for example in the fact that in the recent period street festivals took place, with many potential pickpocket victims on the streets. This shows that by using more data, patterns offer more contextual information and help analysts reach the right conclusions. So, to understand crime, data is needed that goes beyond simple aspects of an incident or person, e.g. the type of neighborhood, the Modus Operandi (MO or manner of working), witness descriptions, stolen goods, vehicles involved and the background of the people involved (history, crime profile, socio-demographic profile). Linking crimes by similarity also benefits from rich data for the same reasons. Finding links can help to detect crime series, connect cases and solve them. This is where data mining comes in. Automated pattern recognition is necessary to turn the data overload into a manageable flow of information that matters.

Apart from handling the volume of data, data mining techniques also help to deal with the dynamic nature and complexity of criminal behavior. This can be seen apart from the data volume subject discussed above, simply because complex patterns can be present in just a few data elements. For example, the following question is very hard to answer using conventional techniques: where and when do crimes take place during the week looking at X- Y co-ordinates of the incidents, the time of day and the day of the week. A proven method to answer this question is to use clever clustering techniques (see 5.1).

It is a common misconception that data mining requires large data volumes in order to add value. On the contrary, it is our experience that when implementing data mining it is best to start with one, maybe two data sources to get acquainted with the possibilities and to manage expectations. Many organizations are surprised by how much information they can gain from data mining on just a small part of their data. Nevertheless, more data is always better for providing more depth and more context.

To conclude, in theory data mining allows the police to better understand and predict crime because many data sources can be analyzed and complex patterns can be found. In 2004, an extensive study by the program bureau of the Dutch Police (ABRIO) concluded that 'data mining enables more effective and goal driven decision making on strategic, tactical and operational levels' (internal report).

3. SHORTCOMINGS OF CRIME ANALYSIS SYSTEMS

Traditional crime analysis tool sets suffer from the following issues:

1. Based on selecting variables

Traditional analytical tools require the analyst to look at variables one by one. This way of working is not viable for rich data sets containing many variables.

2. Static results

Existing systems usually generate static reports that do not allow interaction. They cannot be used to find the explanations behind the numbers they present.

3. Based on simple patterns

When an analyst focuses on one or two variables, the traditional tools allow only the analysis of those individual variables. In rare cases, interaction between two chosen variables is analyzed, but all other combinations are not used which is why useful interactions between variables can be overlooked.

4. Difficult extraction

It is typically hard to extract data from police source systems because of old and diverse database systems with data models based on transactions instead of analysis. There are several standard analytical applications working on small extractions, but when analysis needs to go a step further, the analyst faces a challenge in getting the right data from the systems. The fact that there are many different systems in the organization adds to the challenge and so does poor data quality. Typically, extraction, linking, correction and preparation need to be carried out for each analysis. To address this, a small number of Dutch police forces have implemented datawarehouses.

5. Diverse tool sets

Analysts have access to a range of tools, each with its own focus, look and feel and different method of reading data and creating results. There are tools for geographic visualization, statistical analysis, creating charts, defining queries, making reports, analyzing criminal networks, monitoring crime rates, etc. This requires analysts to learn all these tools and slows down the process because of the effort needed to transfer data between tools. There is no tight coupling to allow analysts to go from technique to technique.

6. Tool complexity

The available tools for statistical analysis require special training and sometimes an education in mathematics or statistics. Often, analysts do not have this background.

Do data mining tools solve these issues? The previous section argues that data mining offers tremendous added value for police. Indeed, data mining tools solve some of the problems mentioned above, by not requiring the selection of variables and the ability to find complex patterns, but they also make other problems worse by increasing tool diversity (yet another tool), tool complexity (datamining expertise required) and difficulties with extraction (techniques put new demands on data). The result is that there are many strong criteria for police data mining users to be effective: they need to be well-trained in IT, data mining, statistics, and have domain knowledge. In other words; they have to know police databases, how to extract databases, how to prepare data, how to use different analytical tools, to design a mining process, to select variables, to correct missing values, to pick the right techniques, to set the right parameters, and to know about psychology, criminals, society and police work. In addition, police data typically challenges the user because of quality issues.

We believe that the high demands for users of standard data mining tools are the main reasons why there seem to be just a few successful police mining applications ([2][3][4][5][6]) and most of these applications are either academic endeavors or small applied projects - not continuous activities. Furthermore, when these applications do appear to be ongoing activities they seem to be limited to a single police force.

4. SYSTEM OVERVIEW

The previous section argues that standard data mining tools are difficult to make continuously useful for police experts. With this in mind, our design philosophy for the data mining system has been to just require users to know their domain and to have analytical skills. No more knowledge and skills are required. The developed system brings various techniques from business intelligence, statistics, machine learning and GIS together in a comprehensive data mining infrastructure. This infrastructure includes a datawarehouse, a reporting module and a desktop tool to provide easy query definition, matching, data visualization, basic statistics, clustering, modeling for predicting, modeling for explaining, link analysis, geographic profiling and geographic visualization.

The following subsections discuss the key system elements, the contents of the database and the way in which the data mining system is used.

4.1 Key system elements

The shortcomings listed in the previous section were used to define requirements during the design of the data mining system. The resulting system has the following characteristics:

- **Ready database**

The requirement for expert IT and database knowledge is reduced by providing a single all-embracing database in which all the data has been extracted, linked, cleaned and augmented. The aim is that the database covers 99% of the information need. The remaining 1% needs dedicated data extraction and preparation. The extensive database works like a *single point of truth*: all analysts use the same standard data and definitions.

- **Automated data mining**

The requirement for expert statistical and data mining knowledge is reduced by automated selection and configuration of data mining techniques. Based on the task and the data, the tool chooses the right technique and optimizes parameters based on the data. Furthermore, the tool assists the user by watching for typical pitfalls, such as unreliable patterns and too many missing values. In some cases, a data mining expert would perhaps surpass the automated selection and configuration. This is the compromise that must be made to let data mining novices mine data. Still, when data mining experts use the tool, they save time and their quality of work is more consistent.

- **User friendly interface**

An intuitive graphical user interface is provided, with a task-based setup, instead of a technique-based design.

- **Interactive analysis**

The system works like an interactive analytical instrument in which every part of the results is clickable to 'zoom in'. In this way, the user can simply embark on an analytical journey without the need to first design the process, as is required in typical workflow-based data mining tools. To support this intuitive and ad hoc process, visualization is an important aspect of the user interface. These interactive possibilities support the discovery process and enhance the creativity and instinct of the analyst. In addition, they allow for interactive sessions with people requiring the information (e.g. someone in charge of an investigation). By working together at critical moments in the analytical process, the real question can be refined, new questions can be answered immediately and patterns can be selected based on their relevance.

- **Traceability**

Although the user is working interactively, it is important to keep track of the steps that were taken to reach a result, especially because such documentation can be required in court. Therefore, the system keeps track of the history of each result.

- **Data flexibility**

Associative memory [14][7] is used as the main technique for prediction, clustering and matching in the data mining system. This means that input data is matched to a representation of training data using a similarity principle, as in *Self Organizing Maps* [17] but with a much wider acceptance of different data types. Because of this, the mining process becomes easier and data usage richer:

1. Hardly any data preparation is required because the associative memory can handle a wide range of data types, such as symbolic data, cyclic ordinals (e.g. day of the week), lists, texts and categories with many values. As long as similarity between values can be calculated, a data type can be used. Furthermore, missing values do not need to be removed or guessed since similarity metrics are able to leave out the missing values in the calculation.
2. Because the technology can handle a wide range of data types, much more information is included in the mining process than would be feasible with other techniques because of their more strict data requirements. Working with non-standard data types when using other techniques is either impossible, introduces performance or training problems, or requires much work for data preparation.
3. The associative memory is able to explain how it reached a result by displaying relevant cases or persons from the memory - which is an intuitive way of explaining a decision to any user.
4. Building associative memory models converges well and fast, compared to for example back propagation neural networks [14].
5. Associative memories are robust against suboptimal parameters (*graceful degradation*) and therefore suitable to use in a situation where parameters are set automatically and the user is not an expert on optimizing the technique [14]

In other words: the user does not need to worry about fine tuning parameters, retraining, solving missing values, variable selection, and decoding variables into a usable form.

There is no free lunch, so there is a price to pay here. First: execution time of associative prediction models is not as fast as other techniques. In police practice, this does not pose a problem because the longer waiting times typically occur at the end of an analytical process, e.g. when there is time to wait for results from a batch prediction job. Furthermore, associative clustering is very fast. Second, other techniques produce more accurate models in some cases and in some cases the associative memory outperforms the rest. We found that the occasional differences in model quality do not outweigh the advantages of saving analyst time, the ability to use rich data and that non-experts are enabled to mine data.

- **Integration**

By integrating most of the tools and techniques into one tool, there is more consistency in usability, the user is no longer required to install and learn several tools and no longer needs to exchange results between tools by exporting and importing. The system features more than just techniques from data mining, ranging from simple data browsing to advanced OLAP analysis. For some types of results, popular standard tools have been linked in such a way that results can be exchanged automatically. Examples of these linked applications are Excel™, MapInfo™, Microsoft Word™, Cognos Reportnet™, Analyst's Notebook™, Weka and Google Maps™. Many analysts are already familiar with these tools.

- **Geo-spatial analysis**

The spatial aspect of crime is obviously important and therefore the data mining system is able to visualize results on maps and to use spatial aspects (e.g. co-ordinates, ground use, and census data) in models.

- **Automated routine work**

The data mining system features a reporting module to create a report for every district and for every priority crime type, containing the following elements:

1. Hot spot maps of the recent period.
2. Temporal hot spot maps to show what changed.
3. Prediction maps of the upcoming period.
4. A where/when analysis with description of the clusters found.
5. A predicted week-distribution of crime over the upcoming period: on what days and times is the highest crime rate to be expected?
6. Crime rate graphs with basic statistics, trends and key performance indicators.
7. Hot shot lists of the most frequent offenders with their social network status and photographs.
8. Maps showing residence and activity areas of offenders.

- **Best practice sharing**

The system allows users to store their best practices in the form of recipes: descriptions of problems with the steps taken to solve them. These best practices can be looked up and reused by all users.

4.2 Datawarehouse

The data mining system provides insight into thousands of variables from various police systems, census data, spatial information, weather, and lifestyle data. All this data is extracted, linked, cleaned, augmented and made available in an open datawarehouse by an automated module. This means that data does not have to be collected and cleaned before every analysis.

The datawarehouse integrates the following sources:

1. BPS/BVH/GIDS: the main transaction systems containing incidents, goods, persons, vehicles etc.
2. HKS: a system containing more details about offenders and a longer history of crimes
3. SHERPA: extensive geographic material, used to provide more information about the scene of the crime (type of area, surrounding infrastructure etc.) and the home addresses of persons
4. CBS: Dutch socio-demographic information on neighborhood level
5. Experian: extensive socio-demographic information on zip-code level
6. Sun/Moon: information about light conditions at a given date and time
7. Events: events taking place in certain areas (e.g. football matches, festivities, holidays)
8. KNMI: Dutch weather information at a given date and province

From these sources, information is gathered at the level of incidents and persons (victims, offenders, witnesses, others involved). Because the system is able to handle many columns and works with many different data types, relational data is also gathered, e.g. a list of crime types in an offender history.

4.3 Types of use

Three types of use of the data mining system can be distinguished: personal desktop analysis, group sessions and reports. The group sessions are special cases of the personal desktop analysis in which the analyst operates the system in the company of domain experts and stakeholders, to interactively look at a problem. In this way, new questions and theories can instantly be addressed and validated.

By distributing results of automated data mining reports, a large audience can benefit from the discovery of complex patterns without having to operate the data mining system themselves. If questions rise from the reports, readers can request an interactive data mining session with a trained user.

4.4 User organization

The Amsterdam police force started using the first version of the data mining system in 2001 and now has around 30 authorized and trained users. In order to match the available techniques to the responsibilities of the users, each user has access to a specific selection of functions in the tool.

In Amsterdam, data mining users are organized in teams of two for each district. These small teams are supported by a central team of analysts and one person responsible for managing the functionality of the system and communication with the user community. This user community consists of domain experts who are either police analysts, or researchers or detectives.

The amount of training required for these users varies from one to three days. New users are assessed to determine if they meet the requirements necessary to become a successful data mining user. This assessment is more focused on intellectual abilities plus analytical and communicative skills, rather than on technical knowledge and education. Simply put, the profile of the data mining user is: a clever person with analytical insight, good with numbers, experience with computer, good verbal and reporting skills, and knowledge of the police domain. Users are not required to have a background in statistics, data analysis or databases. It is a plus though, when they know the data model.

Importance of domain experts using data mining

Extensive police domain knowledge is essential for finding useful patterns and interpreting them. An example: a data mining expert from outside the police force was asked to analyze the activity pattern of drug-related crimes in and around a shopping centre. The expert found regularity but thought it wasn't interesting because there didn't seem to be a cause and effect. However, an experienced police officer noticed that this particular behavior corresponded with the time table of the so-called *methadone bus* that provides prescribed drug replacements to addicts. Furthermore, because the *methadone bus* is a helpful program, the solution should not be to stop the bus from visiting the area, but to find out the problem elements in the drug-related crimes at those times and locations. It turned out that a few addicts on the methadone program consistently misbehaved in the vicinity of the bus. The plan was made to confront these individuals, letting them know their prescriptions were on the line.

5. SYSTEM APPLICATIONS

The following subsections describe the main methods by which the data mining system is applied in practice.

5.1 Spatio-temporal clusters: where, when

Knowing when and where the risk of crime is highest allows proactive and effective deployment of resources. The traditional method for spatial risk analysis is to either list the neighborhoods with the highest recent activity or create a map that essentially does the same. This method ignores the fact that the crime rate in an area depends very much on the time of day and the day of the week. For example: some areas are crowded around rush hour, some areas are crowded when pubs close and some areas suffer from burglaries on Saturday evenings because many inhabitants are not at home. To find these patterns, one could create a giant cross table to combine neighborhood, time of day and day of the week, which would be hard to interpret. Also, more location detail is required for good tactical planning. Police presence in one street may not prevent problems a street further. Therefore it is important to use exact co-ordinates, since offenders do not take the boundaries of areas into account. When co-ordinates are used instead of neighborhood, traditional methods and visual inspection fall short.

One of the most popular applications of the data mining system is the so-called 'where-and-when-analysis' that clusters recent incidents and incidents from similar past seasons on co-ordinates, time of day and day of the week. Every cluster is reported with a typical profile of the incidents including MO and descriptions of the offenders. Together with the location and time, police employees have all the information to know where and when to go and what to look for.

The where-and-when-analysis employs an associative cluster technique that we based on the principle of Metric Multi Dimensional Scaling (MMDS) [8], projecting the high dimensional space onto two dimensions in a non-linear way. This process tries to preserve the distances between incidents as determined by the associative memory technique. Contrary to factor analysis, all variance is represented in the two-dimensional result space. This leads to some distortion, but that is not a problem since the purpose of the technique is to visualize and identify clusters. The advantage of the MMDS approach is that the results are easy to interpret for non-data mining experts, as it appears as a regular scatter plot, visualizing the clusters and the relations between them.

Our cluster algorithm first employs the associative memory to calculate the distances between incidents. These distances are used to define gravitational forces between these incidents: the more incidents are alike, the stronger they want to move to another in a two-dimensional plane. Furthermore, a rotational force is introduced, as well as a force that drives away all incidents from the weighted center of the plane. These forces are then used to iterate through an optimization process in which the incidents that are alike move closer and create clouds in the plane (see figure 4). When movement drops below a threshold, the iteration stops and the resulting distances in two dimensions should resemble the multi-dimensional distances.

The results from where-and-when-analysis are of continuous value in the prevention and repression of priority crime types, which is why cluster overviews are generated automatically and included in the weekly standard reports. These overviews shows a hot spot map of the district with ellipses drawn where the clusters are, plus a list of clusters, each presented by a hot spot map, a description of time and day of the week plus a crime profile of what is typical for the cluster (e.g. type of offender, type of stolen property).

In Amsterdam, this analysis is also used to determine when and where to plan public search actions aimed at finding weapons on people in the streets. This is a co-operation between the city council and the Amsterdam police. Since the start of these data mining guided searches, weapon possession has dropped 27%.

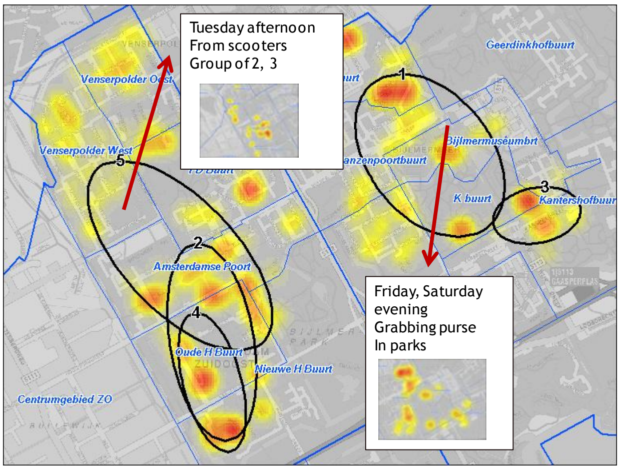

Figure1: Spatio-temporal clusters with profiles

5.2 Associative spatial prediction

Spatio-temporal clusters (see 5.1) find regularities in space and time that can be used as a general tactical plan for a specific period, e.g. a different suggested patrol route for every weekday during the month of November 2009. Associative spatial prediction is another crime prediction method, aimed at providing an optimal crime risk map for a specific short period; say November 1st 2009 during the 16:00 to 20:00 shift. Such a moment in time provides more context because there is a weather forecast for that day, in November the sun sets early and it is Halloween. Associative spatial prediction applies an associative memory to search for relevant situations in the past, after which these situations are superimposed to create a detailed predictive hot spot map, showing the spatial risk distribution of crime for the given future situation.

This approach is based on the theories of *repeat victimization* [9][10], *routine activity* [11][12] and *prospective hotspotting* [13]. Associative spatial prediction takes this work a step further by taking more into account than just location: recency, trends, seasonal influence, weather, time of day, day of the week, and events or holidays. Experiments have shown that for burglaries the hottest spots in the predicted maps contain 50% of all future incidents, whereas traditional methods (a hot spot map of the recent period) contain only 25% in the hottest spots.

Evaluation

Associative spatial prediction has been evaluated by specifying several random combinations of time periods for training and testing. For each combination, the prediction model was trained on one time period and tested on the other. The performance of the model was measured by calculating the error between the predicted number of crimes and the actual (future) number of crimes for each cell in a 30x30 grid covering the total area to be predicted. The total model error is the average error over all cells, averaged over the various test sets. The table below shows the model results compared to what is considered the standard method for geographic anticipation; creating a hotspot map of the crimes in the recent period (*Kernel density* – see 5.4):

Table 1: performance results of associative spatial prediction

Region	Technique	Total model mean square error
City centre of Tilburg	Kernel density	0.11
City centre of Tilburg	Associative	0.052
North Tilburg	Kernel density	0.086
North Tilburg	Associative	0.044

The table shows that the associative model outperforms the standard for these situations with about half the error. Further work is underway in looking at more situations and applying a performance measure that takes into account how the technique increases effectiveness of police work.

5.3 Analyze trends or behavior

This subsection discusses how the data mining system is used to describe and explain trends or behavior.

5.3.1 Explain trends

Police forces respond to changes in crime rates. Often it is necessary to understand the reason behind a trend to know whether the trend can be explained by a known phenomenon or whether it needs special attention. Also, finding a probable cause allows the cause itself to be addressed.

The data mining system provides trend explanations by comparing the recent period with the period before, performing chi square and Student T-tests for all the available data columns and then sorting on the index. The result is a *profile analysis*; an enumeration of the most significant differences between the periods, to offer clues for an explanation. The system takes this a step further by providing the option to build a decision tree using the same tests in order to find *combinations* of factors as trend explanations. For example: in the end of last year, more burglaries took place in the early evening in the down town area and on Saturday mornings in the western suburbs.

The same method can be used to determine the success of counter measures. For example: the effect of installing surveillance cameras in a public area was analyzed by comparing the period after the installation with the period before. The results showed that the destruction of public property was reduced, but reports of street robberies in parking lots (out of the cameras' sight) increased.

The example below shows how it can be explained why last December had more violent crimes than November. The tree starts out by looking at all violent crimes in November and December, where December is 60% of the total. The tree algorithm finds that street robberies show the most difference: 80% of all violent crime in December was a street robbery, which was 60% in November. Next, it turns out that robberies in Public transport have relatively increased, especially between 17:00 and 18:00. The latter pattern covers 40% of 75% of 80% is 24% of all violent crimes in December.

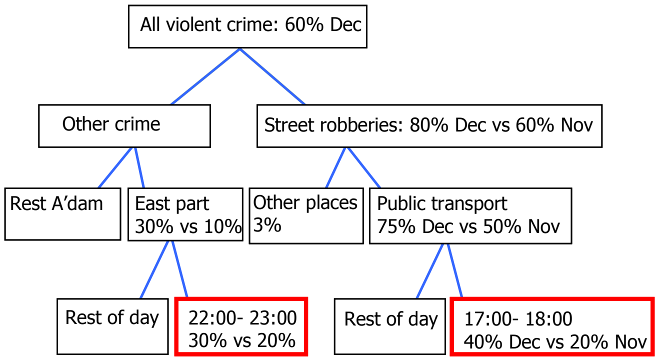

Figure 2: Decision tree explaining a trend

5.3.2 Find contextual trends

An alternative use of comparing periods is to find trends in context; not the trends in crime rates going up or down, but trends in the aspects of crime. For example: there is no strong increase of total number of burglaries but there is a sudden increase of flat screen theft from apartments, or a trend where a specific type of tool is used. These trends can be used to point out activity by a specific offender or offender group, or a more general phenomenon, such as an increase in motor cycle theft because spring has just begun.

5.3.3 Explain or describe behavior

The same approach can be used to explain behavior, by comparing the activities and/or the features of a person or group against a reference group. For example: compare violent young criminals in a specific neighborhood against all young criminals in that neighborhood. What makes these violent persons different? Are they from specific parts of town, are there patterns in their careers? What is their typical social background? Such patterns can be found with a decision tree analysis. By gaining insight in related aspects of such behavior, preventative measures can be designed. Another application is creating a description of the *signature* (typical way of working) of a person or group.

5.3.4 Explain spatial relations

Detecting relations between spatial aspects (e.g. type of neighborhood) and behavior (e.g. crime rates) is traditionally done by looking at areas: the crime rate for each area is used as a variable that needs to be explained by the aspects of the area (average income etc.). A new method of doing this was created, by using local crime density as the variable to be explained. In this way, the level of detail can be much higher, say on address level, without the requirement of having numerous incidents on that level. First, a density analysis is done (see paragraph 5.4), calculating the density of the crime for each address. Then, the techniques for explanation are applied. This allows for better and detailed explanations since crime rates and spatial aspects do not need to be averaged over large areas.

5.3.5 The use of text

Using textual information can be very helpful for gaining insight. For example: the system was asked to analyze the Ramadan period (Islamic month of fasting) and it found that this period typically has a strong increase of destruction of property. Before we had the chance to think about possible relations between fasting and this pattern, the system also showed that the words *firecracker* and *fireworks* occurred more often in incidents during Ramadan. Now, Ramadan takes place during the ninth month of the Islamic calendar, causing it to sometimes take place in December. This explains the pattern of fireworks and thus the destruction of property resulting from incidents during the days before New Year's Eve, which was confirmed by the resulting decision tree.

5.4 Kernel density estimation - hot spot maps

Knowing where crimes take place is crucial information for the police and is best shown on a map for optimal interpretation. The basic way of visualization is to plot single incidents as dots. If dots are too close to each other, they can be combined into larger dots. This method seems fairly obvious. However, such dot maps can be hard to interpret, especially where the concentration of dots is high. Hot spot maps provide a solution by interpolating incidents for each cell of a detailed grid on the map, resulting in a color-coded *hot spot* map. The data mining system uses kernel density estimation [11] for the interpolation on three detail levels, each for a different zoom range on the map. Thus, the hot spot map shows more detail when the user zooms in. An extension of this technique was implemented by allowing the map to incorporate parts of streets in addition to exact addresses, since a large percentage of incidents have been registered with just a street name.

5.5 Temporal hot spots

The data mining system allows temporal hot spots to be created for visualizing spatial trends in time. This is done by creating a kernel density grid for the recent period and one for the period before, which are then subtracted, resulting in a density map with positive (red) areas with increased crime rates and negative (blue) areas with decreased crime rates.

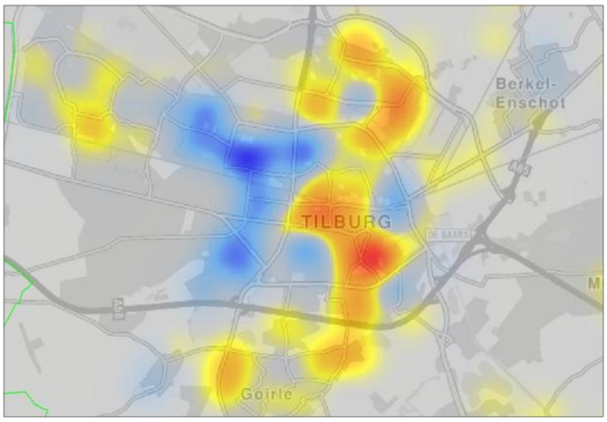

Figure 3: Temporal hot spot analysis of burglaries in Tilburg

5.6 Cluster series of crimes

A typical police analyst's task is to link crimes in order to either solve a case or to find interesting series that point to the activity of a single offender or a group. This linking process is supported by the data mining system by allowing crimes to be clustered on location, MO, time, day of the week, weather, and suspect descriptions.

Associative clustering (see paragraph 5.1) is used to create the result, showing clouds of incidents on a two-dimensional chart. The closer incidents are on the chart, the more similar they are. This chart can be used to detect series and to match a specific unsolved case to similar cases simply by looking it up on the chart. Especially the *solved* similar cases are interesting. This associative approach is similar to the application of Self Organizing Maps (or Kohonen maps [14]) for clustering crimes as discussed in [2]. The associative approach however is easier to use, has less data requirements and faster execution times.

5.7 Cluster sub problems

Spatio-temporal clustering, as discussed in paragraph 5.1, detects clusters in space and time, typically for one specific type of crime. By adding contextual information, the data mining system is able to detect clusters of similar incidents that together define a *sub problem*. In this way it is possible to cluster a wider range of incidents (e.g. all incidents) to find out what structural problems exist in order to address them individually through prevention and/or repression. Finding the right preventative measures is supported by a deeper analysis of the cluster to find its probable causes. Repression is done by using the tactical information just as with the spatio-temporal clusters: location, time, day, and clues to look for.

An example of a situation where this approach was successful is the analysis of sub problems on Queen's day. Queen's day in the Netherlands is a national street festival with its own typical safety challenges. In order to prepare for this day, the Amsterdam police ran a sub problem clustering on Queen's day data from recent years. The results showed the typical Amsterdam everyday problems plus typical Queen's day problems, such as pickpocket crimes near the Rembrandt Square in the early evening, inside bars and restaurants, with mostly tourist victims. This pattern, along with many other sub problems, allowed the police to define very specific instructions for officers to address these problems on the upcoming Queen's day.

Another example is the application of sub problem clustering for a specific shopping area in the South East of Amsterdam. It took place in an interactive group session where experts and stakeholders were present: neighborhood police officer, neighborhood watch, shop owners and drug expert. The data mining system was operated by an analyst visualizing the problems, so questions could be addressed immediately.

Experiments with extensive socio-demographic data in problem clusters have demonstrated that causes of problems can be explained better by combining tactical data with household data. For example: clusters were found containing theft of flat screen televisions from low-income households, who have a tendency of showing off material possessions. This resulted in the theory that this group makes itself vulnerable by positioning the televisions so they can clearly be seen from the outside, making them more attractive to burglars. Such a theory also helps to determine other geographic areas that suffer from the same risks.

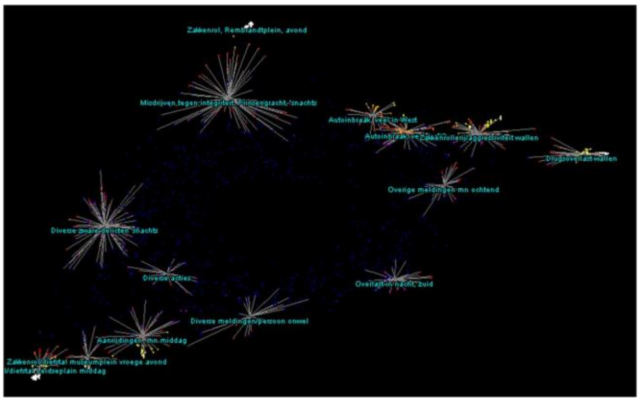

Figure 4: Clustering of sub problems on Queen's day

5.8 Geographic profiling

Geographic profiling is a method that uses locations of crimes to find the most likely areas of the offender's home address [15]. This is done by applying theories about the activity radius of offenders combined with theories about offenders usually not operating in the vicinity of their own home address. By combining this activity theory for every crime location, the data mining system creates a probability map that can be used to narrow down a selection process while solving a case. The parameters are calculated automatically based on the data.

5.9 Link analysis

Connecting crimes and criminals is important when solving cases and when looking into the networks of criminals to see what groups there are, who knows who, and what their roles seem to be. The data mining system provides a link analysis tool to visualize the structure and content of a collection of interconnected entities and to calculate measurements based on theories from *Social Network Analysis*, tailored to the criminal domain [16]. These measurements help in finding roles (e.g. who maintains the most connections). The data mining system is able to generate a network and to visualize it using the same dimensional scaling algorithm as in clustering (See 5.1). This way, the suspects and known offenders in the network are organized so that their closeness on the screen resembles closeness in links. This saves analysts hours of combining data. Networks can be exported to Analysts's Notebook.

5.10 Find possibly connected suspects or incidents

One of the standard methods for solving a case is the search for similar cases in the past to see if they contain useful information. Especially when those similar cases have a known offender, it can be interesting to consider that offender as a possible suspect of the new case. Performing such a search is a difficult task because there is hardly ever an exact match. For example: suppose the case to be solved is a violent street robbery. The same offender may have been a suspect for a very similar incident a few weeks ago, but the similarity may be complex: the past crime has been entered as a pickpocket incident and not as a street robbery; the pickpocket location was near the robbery location, but not in the same street; witness descriptions state blond hair for the pickpocket, but dark blond for the robber. There may be many similarities making the pickpocket case interesting, but it is typically not found using conventional query methods.

The police data mining system features associative techniques that find similar cases based on a search case. This is used as part of the algorithms for predictive modeling and clustering, and it can be used as a technique for flexible search as well. Such a search takes a case as input and creates an output of similar cases, ordered by similarity. These similar cases can then be used to find the related suspects. If one of those suspects has more than one similar case, that suspect becomes more interesting. This entire process has been automated in the system, also allowing multiple cases (for example a series of crimes) to be used as the input for a search. For example: a series of 10 crimes is used in an associative search that finds 33 similar cases in the past of which 4 have one and the same suspect and these 4 cases are similar to 8 out of the 10 crimes in the search series.

There are many applications for associative searching, e.g. looking for suspect photographs based on a witness description, or looking for incidents that are most similar to the entire criminal history of an offender in order to find other crimes that may be committed by this offender.

6. DATA MINING SUCCESS STORIES

The following subsections present success stories of applying the data mining system in practice.

6.1 Robbery

In the beginning of 2009, the district of Tilburg in the Netherlands suffered from a serious wave of robberies at gas stations, restaurants etc. Analysts used the data mining system to visualize the locations of these robberies and applied associative spatial prediction to determine when and where police actions (e.g. roadblocks) would be optimal. They applied link analysis on past robberies to determine which suspects were important to keep an eye out for. In addition, the system produced top X lists of offenders, ordered by their past robbery activity. These selected people were brought in if they had any outstanding fines. Police officers visited the rest and their photographs were used in briefings. Within two weeks, the wave of robberies was stopped.

6.2 Car burglaries

In 2006, a trend analysis in Amsterdam indicated an increase of motor vehicle theft in District 2 for the month of May. A profile analysis of this trend showed that the increase could be explained by thefts from private garages. More officers than normal were deployed and their actions were supported by spatio-temporal cluster analysis plus the top 10 list of car burglars. The result was that a repeat offender was caught in the act within one hour after the data mining information was provided and crime went down 90% in the first week.

6.3 Burglary

The following scenario illustrates how data mining techniques are combined in solving a problem. The analysis actually was performed by police to better understand the underlying problems.

1. The standard report indicates a sudden increase of burglaries in October, which is atypical for the time of year.
2. A time series chart is created to visualize the trend and validate the indication.
3. By clicking on October in the chart, an explanation of the trend is requested in the form of a decision tree to show what combinations of factors have changed (see 5.3). The tree shows that there are more break-ins at the back of houses than before, especially in the early evening. It also shows that the burglaries occurred more often after sunset. This leads to the theory that it is getting dark earlier and therefore there is more opportunity in the early evening especially in back yards where there is hardly any street lighting. This happens especially in a specific neighborhood. The pattern can be used purely to explain the trend, but it can also be used to start a program in that neighborhood to encourage placing motion-triggered lights in back yards.
4. The decision tree also shows a strong geographic difference. This leads to a temporal hot spot analysis (see 5.5) that shows a number of hot spots where crime increased.

5. The largest hot spot is chosen by zooming in on the map and asking the system to create an explanation of this trend. This shows a very specific time of day, specific streets and some witness descriptions; useful information to support police officers on patrol in that area.
6. To determine who the police officers should look for, the hot spot selection is used to perform *associative searching* (see 5.10). This matches the incidents in the hot spot with the entire known crime history. The best matches and most recent incidents are selected to see who the connected suspects were, in order to provide their photographs to the police officers.
7. Next, link analysis is used to find the people that are directly or indirectly connected to the selected probable suspects, to also add their photographs.

7. DISCUSSION

In this section we will discuss added value for police work, the challenges involved and future work.

7.1 Added value

Section two discusses the various reasons why data mining is important for police analysis. The organizational impact of the discussed data mining system is that a large number of users are now able to gain more insight and predict criminal behavior. These users are responsible for providing information to the organization, so in other words, data mining has made the organization more intelligent. The growing number of users of the data mining system is a clear sign of the acceptance of data mining as an important tool for the police. But how can the added value be measured?

Five methods to assess the added value of data mining can be distinguished: model accuracy, experiments in practice, mimic practice, analyst efficiency and qualitative comparison.

7.1.1 Measure model accuracy

Measuring model accuracy can be done by basing the model on one part of the data and testing it on the rest. The resulting accuracy can then be translated to the predicted increase in police effectiveness using a simple model of police practice. For example: we calculated that the top 5% hotspot areas from associative spatial prediction eventually contain 50% of all crimes in the predicted period, whereas the top 5% areas from traditional tactical information contain 25%. We assume the police only have the capacity to be present in those 5% areas and that police presence reduces the risk of an incident taking place by 50%. This way we can calculate the projected reduction of crime. For the spatial prediction, 50% of all crime is reduced by 50%, so total crime is reduced by 25%, were the traditional approach reduces 50% of 25%, which is a reduction of the total crime rate by 12.5%. This method of projecting effectiveness typically suffers from many assumptions but it can be useful to illustrate the impact of these predictive models.

7.1.2 Run experiments in police practice

The proof of the pudding is in the eating and therefore the best way to measure effectiveness is to experiment in practice. However, because public safety is generally considered not something to experiment with, pure tests are very difficult to arrange. Furthermore, crime and environment are so dynamic that an exclusive explanation of success cannot be based on one experiment alone. Crime rates may change during an experiment because of many reasons other than data mining being used or not. Therefore, multiple experiments are required for a good measurement. Another challenge in measuring added value is how to measure safety, security and fear. Crime rates are useful but cover just a part of the whole picture.

Recently, the police force of Midden en West Brabant addressed a region-wide wave of crime. One district used data mining for providing information on these crimes, the others did not. After two months, the data mining district reduced crime by more than 15% whereas the other districts showed constant crime levels. There are more encouraging results like this, but it is no solid proof. Because of the difficulties with measuring this type of added value, we also use the other methods in this section to convince police organizations.

7.1.3 Mimic police practice in laboratory settings

An alternative to experimenting in practice is to mimic practice. We performed a field test with the Rotterdam and Haarlem Police forces in which volunteers were asked to sit in a waiting room, where a laptop was 'stolen' by an actor. The volunteers witnessed this act and were asked to provide descriptions that were then used to search for photos to show for identification. This was done using two systems: the standard police search system and an associative search tool, similar to the one used in the current data mining system. The result was that 50% more witnesses recognized photos of criminals selected by the associative system than from the standard police search system. The disadvantage of such tests is that they are expensive.

7.1.4 Measure analysts' efficiency

An alternative to measuring increase of effectiveness is to measure the gain in *efficiency* as a result of data mining. Measuring analysts' efficiency is more simple than measuring police effectiveness and the work situation is easier to control than criminals and their environment. Research in the region of Midden en West Brabant showed that analysts generally reduced their analysis time by a factor 20 when using the data mining system. The added value of that comes down to cost reduction or increased effectiveness because analysts can get more work done. Furthermore, response times are much shorter and analysis is done quick enough to support group sessions in which experts and stakeholders discuss and analyze a subject. In all police organizations we have seen, analysts always have much more to do than they have time for, which means that when the data mining system is used, the organization is better informed.

7.1.5 Qualitative comparison of results

In our experience, qualitative assessment of data mining advantages is often convincing to decision makers. Once analysts and their chiefs experience what data mining can offer them, they typically do not need quantitative proof that it will make the organization more effective. An analogy is the implementation of geographic information systems in police forces. Once officers start working with maps, the added value seems clear and there is no need for conducting experiments in which this new situation is compared with situations without maps. Research by Abrio [internal report] and Midden en West Brabant has shown that data mining analysis is more productive and the results better satisfy the information needs because of their level of detail and explanation.

7.1.6 Cost and return on investment

The costs for data mining can be divided in implementation and running costs. Implemenation requires the availability of a datawarehouse in which data sources are extracted, combined, cleaned and extended by calculating variables (e.g. domestic violence yes/no). Such a datawarehouse provides a *single point of truth* for the organization, and its merits go far beyond data mining. Therefore, police forces in the Netherlands see the availability of a datawarehouse as a general *must* and there are many datawarehouse initiatives. We have developed a datawarehouse based on the dominant standard police systems in the Netherlands, thereby reducing datawarehouse implementation costs. Remaining costs are for database server hardware and software, installation and configuration.

Other data mining implementation costs are: application server hardware, software license and training. Running costs are software license, retraining, functional management and operating cost of batch modules that update the datawarehouse and create standard reports.

We have reduced these costs greatly by making the data mining system easy to learn and easy to use. Training is done in two to three days and no expert users need to be hired.

It is possible to argue a positive return on investment, based on only the benefit of analyst efficiency (7.1.4). The efficiency gain is a factor 20, so theoretically the capacity in analyst personnel can be reduced strongly while maintaining the same level of information delivery. This would create a cost reduction that would more than make up for the costs of implementing and using data mining. However, police organizations do not intend to reduce analyst capacity because the information need strongly exceeds the information production. In other words: in a simplified model, using data mining can be compared to hiring more analysts for a fraction of the costs. This makes the business case for data mining, based on analyst efficiency alone, leaving all other benefits out of the equation.

7.2 Challenges

The implementation and use of the data mining system pose a number of challenges.

7.2.1 User skill management

Even though much attention has been given to making the system user friendly, it is important to maintain the user's level of skill. Applying data mining is an inherently complex task, even if all choices regarding parameters and techniques are done automatically. Maintaining the skill level is managed by pro-active user assistance, sharing best practices, stimulation of user communities and regular training sessions.

7.2.2 Data quality

The higher the data quality, the better the results from data mining are. Managing the quality of police data is a notoriously difficult problem because of the diversity of people who enter the data, the varying circumstances in which this is done and because of difficulties in applying definitions when registering facts. This has not stopped data mining applications to be successful but it is obvious that results will be better as data quality increases.

It is interesting to note that data quality can be improved by applying data mining because of two effects: 1) all data is used (demonstrating to the organization how important it is to register all data correctly), and 2) some patterns found by data mining illustrate data problems. For example: an increase in specific MO may not be caused by an increase of that type of crime but rather by new instructions to the people that entered the data.

Two principles guide the handling of data quality problems by the data mining system:

1. Deal with potential data problems as early as possible while creating the datawarehouse by combining data sources and by using software to correct data errors, for example by simply removing a piece of information when rules detect it is erroneous.
2. Document data problems and make this documentation easily accessible from within the data mining system.

7.2.3 Managing expectations

It is our experience that in some cases the expectations of the impact of data mining are too high. This requires clear communication of what can be expected. Data mining will not solve all data problems. Data mining will not make analysts dispensable. Data mining will not be able to interpret textual information perfectly. Data mining will not easily succeed in bringing together all the available data because of privacy laws and the amount of effort it takes to create reusable extractions.

7.3 Future work

Although the data mining system has been operational for years, new ideas still come up to improve the system and to apply it in new ways.

7.3.1 More prediction models

Currently, prediction models are used to predict where crimes are likely to take place and to explain trends or behavior. Many other applications of predictive models are being studied:

1. Criminal career: who have the highest risk of becoming a repeat offender?
2. Criminal profiling: predict probable motive, offender age, gender, etc., based on a crime or series of crimes.
3. Weapon possession: what is the risk of a person carrying a weapon, based on personal profile and history?
4. Domestic violence: what is the chance of a reported domestic violence situation getting out of hand, to support the decision to pay extra attention to the case?
5. Predict crime rates based on infrastructure, buildings and socio-demographics in a new neighborhood, allowing what-if experiments.
6. What are the risks for a specific type of crime for an area based on the properties of that area and the situation (time, day, weather)? This alternative approach to associative spatial prediction (see 5.2) is useful when there are too few examples to make the spatial prediction work.

7.3.2 Dashboard

The standard reports generated by the data mining system are currently distributed in document form, created by transforming XML into HTML. There are plans to offer this information in the form of a clickable dashboard that starts with an overview of crime rates and allows clicking on rates to zoom in on areas and types of crime, showing more details such as maps and trends. In this way, the wealth of available report information could be navigated more effectively.

7.3.3 Text

Text is an important source of information within police forces because, for example, statements have more detail than the information recorded in the structured data. The implemented data mining system allows textual information to be included and used for selection and for pattern analysis. However, a system to regularly collect textual information from the source systems still needs to be built. A different way of working with text is to extract entities (license plates, phone numbers, addresses and names) to use them as a method for linking people and cases. When such entity extraction systems are implemented within police forces, we will add the entity information to the data mining system.

7.4 To conclude

In this paper we showed that data mining is important for police and that it has been made a continuous activity, performed by police employees without extensive expertise in databases or data mining. This has been realized by developing a data mining system in co-operation with police, bringing together interaction, visualization, tool integration, automated algorithms, a large and diverse datawarehouse and ease of use. Especially the latter is a key success element because a combination of factors make data mining for police relatively difficult: analytical diversity, strong need for domain expertise, data quality problems and difficult data extraction. The system has proven to provide more depth in analysis and in some cases an efficiency gain of a factor 20, resulting in more analytical capacity, shorter response times and the ability to analyze interactively during group sessions. Furthermore, crime rates dropped in field tests. After eight years of use, the system is still being extended constantly based on new ideas, which shows great promise for the future of crime fighting.

8. REFERENCES

[1] Ratcliffe, J.H. 2003. Intelligence-led Policing, Australian Institute of Criminology, Canberra, Australia, 248.

[2] Adderley, R. & P.B. Musgrove, P.B. 2001. Data Mining Case Study: Modeling the Behavior of Offenders Who Commit Serious Sexual Assaults, Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining.

[3] McCue, C. 2007. Data mining and predictive analysis: Intelligence gathering and crime analysis, Butterworth-Heinemann.

[4] Mena, J. 2003. Investigative Data Mining for Security and Criminal Detection, Elsevier Science (USA).

[5] Brown D.E. 1998. The Regional Crime Analysis Program (RECAP): A framework for mining data to catch criminals, University of Virginia.

[6] Bruin, J.S. de, Cocx, T.K.., Kosters, W.A., Laros, J.F.J., Kok, J.N. 2006. Data Mining Approaches to Criminal Career Analysis, Proceedings of ICDM '06.

[7] Uyl, M.J. den. 1986. Representing Magnitude by Memory Resonance, Proceedings of the 6th Annual Conference of the Cognitive Science Society, pp. 63-71.

[8] Borg, I. & Groenen, P. 2005. Modern Multidimensional Scaling: theory and applications (2nd ed.), SpringerlVerlag New York.

[9] Farrell, G. 2005. Progress and prospects in the prevention of repeat victimization.

[10] Eck, J.E., Chainey, S., Cameron, J.G., Letiner, M. & Wilson, R.E. 2005. Mapping crime: Understanding hot spots, Technical report 209393, National Institute of Justice.

[11] Chainey, S. & Ratcliffe, J. 2005. GIS and Crime Mapping, Mastering GIS: Technology, applications and management, West Sussex, England: Wiley.

[12] Clarke, R. V. & M. Felson, Eds. 2004. Routine Activity and Rational Choice (Advances in Criminological Theory), Transaction Publishers, New Brunswick (U.S.A).

[13] Bowers, K.J., Johnson, S.D. & Pease, K. 2004. Prospective hot-spotting: The future of crime mapping?, Br. J. Criminol 44(5), pp. 641-658.

[14] Kohonen, T. 1984. Self-organisation and associative memory, Springer series in information sciences, Vol8. Springer Verlag, New York, USA.

[15] Rossmo,D.K. 2000. Geographic profiling, Boca Raton, Fla, CRC Press.

[16] Freeman, L. 2004. The development of social network analysis, Vancouver, CA: Empirical Press.

[17] Kohonen, T. 2001. Self-Organizing maps, Springer series in Information Sciences, Vol. 30, Springer, Berlin, Heidelberg, New York