Crime Pattern Detection Using Data Mining

Shyam Varan Nath
Oracle Corporation
Shyam.Nath@Oracle.com
+1(954) 609 2402

Abstract

Data mining can be used to model crime detection problems. Crimes are a social nuisance and cost our society dearly in several ways. Any research that can help in solving crimes faster will pay for itself. About 10% of the criminals commit about 50% of the crimes. Here we look at use of clustering algorithm for a data mining approach to help detect the crimes patterns and speed up the process of solving crime. We will look at k-means clustering with some enhancements to aid in the process of identification of crime patterns. We applied these techniques to real crime data from a sheriff's office and validated our results. We also use semi-supervised learning technique here for knowledge discovery from the crime records and to help increase the predictive accuracy. We also developed a weighting scheme for attributes here to deal with limitations of various out of the box clustering tools and techniques. This easy to implement data mining framework works with the geo-spatial plot of crime and helps to improve the productivity of the detectives and other law enforcement officers. It can also be applied for counter terrorism for homeland security.

***Keywords**: Crime-patterns, clustering, data mining, k-means, law-enforcement, semi-supervised learning*

1. Introduction

Historically solving crimes has been the prerogative of the criminal justice and law enforcement specialists. With the increasing use of the computerized systems to track crimes, computer data analysts have started helping the law enforcement officers and detectives to speed up the process of solving crimes. Here we will take an inter-disciplinary approach between computer science and criminal justice to develop a data mining paradigm that can help solve crimes faster. More specifically, we will use clustering based models to help in identification of crime patterns[1].

We will discuss some terminology that is used in criminal justice and police departments and compare and contrast them relative to data mining systems. Suspect refers to the person that is believed to have committed the crime. The suspect may be identified or unidentified. The suspect is not a convict until proved guilty. The victim is the person who is the target of the crime. Most of the time the victim is identifiable and in most cases is the person reporting the crime. Additionally, the crime may have some witnesses. There are other words commonly used such as homicides that refer to manslaughter or killing someone. Within homicides there may be categories like infanticide, eldercide, killing intimates and killing law enforcement officers. For the purposes of our modeling, we will not need to get into the depths of criminal justice but will confine ourselves to the main kinds of crimes.

Cluster (of crime) has a special meaning and refers to a geographical group of crime, i.e. a lot of crimes in a given geographical region. Such clusters can be visually represented using a geo-spatial plot of the crime overlayed on the map of the police jurisdiction. The densely populated group of crime is used to visually locate the 'hot-spots' of crime. However, when we talk of clustering from a data-mining standpoint, we refer to similar kinds of crime in the given geography of interest. Such clusters are useful in identifying a crime pattern or a crime spree. Some well-known examples of crime patterns are the DC sniper, a serial-rapist or a serial killer. These crimes may involve single suspect or may be committed by a group of suspects. The below figure shows the plot of geo-spatial clusters of crime.

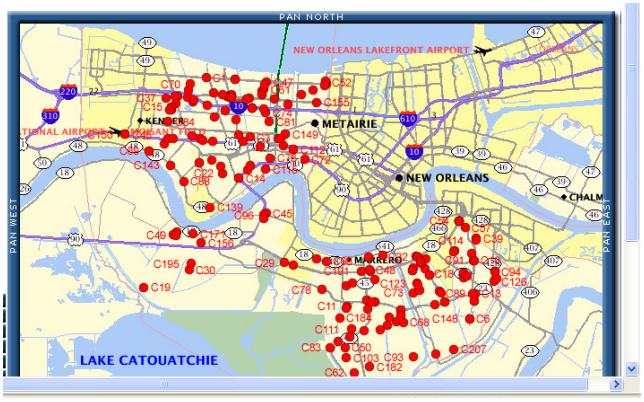

Fig 1 Geo-spatial plot of crimes, each red dot represents a crime incident.

2. Crime Reporting Systems

The data for crime often presents an interesting dilemma. While some data is kept confidential, some becomes public information. Data about the prisoners can often be viewed in the county or sheriff's sites. However, data about crimes related to narcotics or juvenile cases is usually more restricted. Similarly, the information about the sex offenders is made public to warn others in the area, but the identity of the victim is often prevented. Thus as a data miner, the analyst has to deal with all these public versus private data issues so that data mining modeling process does not infringe on these legal boundaries.

Most sheriffs' office and police departments use electronic systems for crime reporting that have replaced the traditional paper-based crime reports. These crime reports have the following kinds of information categories namely - type of crime, date/time, location etc. Then there is information about the suspect (identified or unidentified), victim and the witness. Additionally, there is the narrative or description of the crime and Modus Operandi (MO) that is usually in the text form. The police officers or detectives use free text to record most of their observations that cannot be included in checkbox kind of pre-determined questions. While the first two categories of information are usually stored in the computer databases as numeric, character or date fields of table, the last one is often stored as free text.

The challenge in data mining crime data often comes from the free text field. While free text fields can give the newspaper columnist, a great story line, converting them into data mining attributes is not always an easy job. We will look at how to arrive at the significant attributes for the data mining models.

3. Data Mining and Crime Patterns

We will look at how to convert crime information into a data-mining problem [2], such that it can help the detectives in solving crimes faster. We have seen that in crime terminology a cluster is a group of crimes in a geographical region or a hot spot of crime. Whereas, in data mining terminology a cluster is group of similar data points – a possible crime pattern. Thus appropriate clusters or a subset of the cluster will have a one-to-one correspondence to crime patterns.

Thus clustering algorithms in data mining are equivalent to the task of identifying groups of records that are similar between themselves but different from the rest of the data. In our case some of these clusters will useful for identifying a crime spree committed by one or same group of suspects. Given this information, the next challenge is to find the variables providing the best clustering. These clusters will then be presented to the detectives to drill down using their domain expertise. The automated detection of crime patterns, allows the detectives to focus on crime sprees first and solving one of these crimes results in solving the whole "spree" or in some cases if the groups of incidents are suspected to be one spree, the complete evidence can be built from the different bits of information from each of the crime incidents. For instance, one crime site reveals that suspect has black hair, the next incident/witness reveals that suspect is middle aged and third one reveals there is tattoo on left arm, all together it will give a much more complete picture than any one of those alone. Without a suspected crime pattern, the detective is less likely to build the complete picture from bits of information from different crime incidents. Today most of it is manually done with the help of multiple spreadsheet reports that the detectives usually get from the computer data analysts and their own crime logs.

We choose to use clustering technique over any supervised technique such as classification, since crimes vary in nature widely and crime database often contains several unsolved crimes. Therefore, classification technique that will rely on the existing and known solved crimes, will not give good predictive quality for future crimes. Also nature of crimes change over time, such as Internet based cyber crimes or crimes using cell-phones were uncommon not too long ago. Thus, in order to be able to detect newer and unknown patterns in future, clustering techniques work better.

4. Clustering Techniques Used

We will look at some of our contributions to this area of study. We will show a simple clustering example here. Let us take an oversimplified case of crime record. A crime data analyst or detective will use a report based on this data sorted in different orders, usually the first sort will be on the most important characteristic based on the detective's experience.

Crime Type	Suspect Race	Suspect Sex	Suspect Age gr	Victim age gr	Weapon
Robbery	B	M	Middle	Elderly	Knife
Robbery	W	M	Young	Middle	Bat
Robbery	B	M	?	Elderly	Knife
Robbery	B	F	Middle	Young	Piston

Table 1 Simple Crime Example

We look at table 1 with a simple example of crime list. The type of crime is robbery and it will be the most important attribute. The rows 1 and 3 show a simple crime pattern where the suspect description matches and victim profile is also similar. The aim here is that we can use data mining to detect much more complex patterns since in real life there are many attributes or factors for crime and often

there is partial information available about the crime. In a general case it will not be easy for a computer data analyst or detective to identify these patterns by simple querying. Thus clustering technique using data mining comes in handy to deal with enormous amounts of data and dealing with noisy or missing data about the crime incidents.

We used k-means clustering technique here, as it is one of the most widely used data mining clustering technique. Next, the most important part was to prepare the data for this analysis. The real crime data was obtained from a Sherriff's office, under non-disclosure agreements from the crime reporting system. The operational data was converted into denormalised data using the extraction and transformation. Then, some checks were run to look at the quality of data such as missing data, outliers and multiple abbreviations for same word such as blank, unknown, or unk all meant the same for missing age of the person. If these are not coded as one value, clustering will create these as multiple groups for same logical value. The next task was to identify the significant attributes for the clustering. This process involved talking to domain experts such as the crime detectives, the crime data analysts and iteratively running the attribute importance algorithm to arrive at the set of attributes for the clustering the given crime types. We refer to this as the semi-supervised or expert-based paradigm of problem solving. Based on the nature of crime the different attributes become important such as the age group of victim is important for homicide, for burglary the same may not be as important since the burglar may not care about the age of the owner of the house.

To take care of the different attributes for different crimes types, we introduced the concept of weighing the attributes. This allows placing different weights on different attributes dynamically based on the crime types being clustered. This also allows us to weigh the categorical attributes unlike just the numerical attributes that can be easily scaled for weighting them. Using the integral weights, the categorical attributes can be replicated as redundant columns to increase the effective weight of that variable or feature. We have not seen the use of weights for clustering elsewhere in the literature review, as upon normalization all attributes assume equal importance in clustering algorithm. However, we have introduced this weighting technique here in light of our semi-supervised or expert based methodology. Based on our weighted clustering attributes, we cluster the dataset for crime patterns and then present the results to the detective or the domain expert along with the statistics of the important attributes.

The detective looks at the clusters, smallest clusters first and then gives the expert recommendations. This iterative process helps to determine the significant attributes and the weights for different crime types. Based on this information from the domain expert, namely the detective, future crime patterns can be detected. First the future or unsolved crimes can be clustered based on the significant attributes and the result is given to detectives for inspection. Since, this clustering exercise, groups hundreds of crimes into some small groups or related crimes, it makes the job of the detective much easier to locate the crime patterns.

The other approach is to use a small set of new crime data and score it against the existing clusters using tracers or known crime incidents injected into the new data set and then compare the new clusters relative to the tracers. This process of using tracers is analogous to use of radioactive tracers to locate something that is otherwise hard to find.

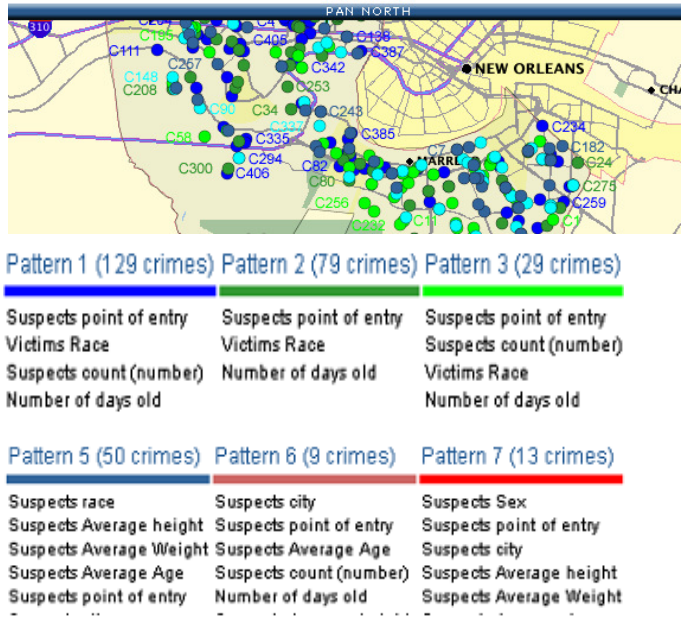

Figure 2 Plot of crime clusters with legend for significant attributes for that crime pattern

5. Results of Crime Pattern Analysis

The proposed system is used along with the geo spatial plot. The crime analyst may choose a time range and one or more types of crime from certain geography and display the result graphically. From this set, the user may select either the entire set or a region of interest. The resulting set of data becomes the input source for the data mining processing. These records are clustered based on the predetermined attributes and the weights. The resulting, clusters have the possible crime patterns. These resulting clusters are plotted on the geo-spatial plot.

We show the results in the figure below. The different clusters or the crime patterns are color-coded. For each group, the legend provides the total number of crimes incidents included in the group along with the significant attributes that characterize the group. This information is

useful for the detective to look at when inspecting the predicted crime clusters.

We validated our results for the detected crime patterns by looking the court dispositions on these crime incidents as to whether the charges on the suspects were accepted or rejected. So to recap the starting point is the crime incident data (some of these crimes already had the court dispositions/ rulings available in the system), which the measured in terms of the significant attributes or features or crime variables such as the demographics of the crime, the suspect, the victim etc. No information related to the court ruling was used in the clustering process.

Subsequently, we cluster the crimes based on our weighing technique, to come up with crime groups (clusters in data mining terminology), which contain the possible crime patterns of crime sprees. The geo-spatial plot of these crime patterns along with the significant attributes to quantify these groups is presented to the detectives who now have a much easier task to identify the crime sprees than from the list of hundreds of crime incidents in unrelated orders or some predetermined sort order. In our case, we looked at the crime patterns, as shown in same colors below and looked at the court dispositions to verify that some of the data mining clusters or patterns were indeed crime spree by the same culprit(s).

6. Conclusions and Future Direction

We looked at the use of data mining for identifying crime patterns crime pattern using the clustering techniques. Our contribution here was to formulate crime pattern detection as machine learning task and to thereby use data mining to support police detectives in solving crimes. We identified the significant attributes; using expert based semi-supervised learning method and developed the scheme for weighting the significant attributes. Our modeling technique was able to identify the crime patterns from a large number of crimes making the job for crime detectives easier.

Some of the limitations of our study includes that crime pattern analysis can only help the detective, not replace them. Also data mining is sensitive to quality of input data that may be inaccurate, have missing information, be data entry error prone etc. Also mapping real data to data mining attributes is not always an easy task and often requires skilled data miner and crime data analyst with good domain knowledge. They need to work closely with a detective in the initial phases.

As a future extension of this study we will create models for predicting the crime hot-spots [3] that will help in the deployment of police at most likely places of crime for any given window of time, to allow most effective utilization of police resources. We also plan to look into developing social link networks to link criminals, suspects, gangs and study their interrelationships. Additionally the ability to search suspect description in regional, FBI databases [4], to traffic violation databases from different states etc. to aid the crime pattern detection or more specifically counter terrorsim measures will also add value to this crime detection paradigm.

7. References

[1] Hsinchun Chen, Wingyan Chung, Yi Qin, Michael Chau, Jennifer Jie Xu, Gang Wang, Rong Zheng, Homa Atabakhsh, “Crime Data Mining: An Overview and Case Studies”, AI Lab, University of Arizona, proceedings National Conference on Digital Government Research, 2003, available at: http://ai.bpa.arizona.edu/

[2] Hsinchun Chen, Wingyan Chung, Yi Qin, Michael Chau, Jennifer Jie Xu, Gang Wang, Rong Zheng, Homa Atabakhsh, “Crime Data Mining: A General Framework and Some Examples”, IEEE Computer Society April 2004.

[3] C McCue, “Using Data Mining to Predict and Prevent Violent Crimes”, available at: http://www.spss.com/dirvideo/richmond.htm?source=dmpage&zone=rtsidebar

[4] Whitepaper, “Oracle’s Integration Hub For Justice And Public Safety”, Oracle Corp. 2004, available at: http://www.oracle.com/industries/government/IntegrationHub_Justice.pdf