Connecting the Dots: Data Mining and Predictive Analytics in Law Enforcement and Intelligence Analysis

Colleen McCue, Ph.D., Program Manager, Crime Analysis Unit,and Colonel Andre' Parker, Chief of Police, Richmond Police Department, Richmond, Virginia

Law enforcement organizations are challenged by a staggering increase in data on a daily basis. In fact, it has been estimated that the amount of data in the world doubles every 20 months. Every transaction, every event, every blip of electricity has the potential to generate data. The events of September 11, 2001, have only served to increase the flood of data while underscoring the critical importance of timely and complete exploitation of law enforcement data resources. The changing face of the war on terrorism and the challenge to connect the dots faces all law enforcement professionals, including those in the local arena.

Data mining tools, which were once reserved for large federal agencies and research centers, are now available to enhance decision making and analysis in the state and local law enforcement arena. Used extensively in the business community, the newer data mining tools do not require huge IT budgets, specialized personnel, or advanced training in statistics. Rather, these products are highly intuitive, relatively easy-to-use, PC-based, and very accessible to the law enforcement community.

The Richmond, Virginia, Police Department is using data mining and predictive analytics for a variety of law enforcement and intelligence applications, including tactical crime analysis, risk and threat assessment, behavioral analysis of violent crime, and proactive deployment strategies.

Data Mining Overview

Data mining has been used successfully in the business world for years. For example, cell phone providers use "churn" models that predict the likelihood that certain customers will switch service providers, online retailers model buying habits in an effort to identify and suggest additional purchases, and grocery stores display ready-to-eat mashed potatoes in the meat section to remind you to buy side dishes to go with your pot roast. The same approach that lenders use to prequalify potential loan applicants can be used to assess the risk for escalation in a series of burglaries, while the models used to classify shopping patterns and purchasing decisions can be used to identify the motive in a homicide or predict the next incident in a crime series.

Data mining encompasses the process of discovering hidden patterns and relationships in large amounts of information. This also allows us to make accurate and reliable predictions of future events, based on the identification and characterization of these patterns and trends in historical data. Data mining helps solve a common problem: the more information, the more difficult and time-consuming it is to effectively analyze and draw meaning from the data. By using a clear

process and powerful analytic technologies, data mining quickly and thoroughly explores mountains of data, helping us identify the valuable or actionable nuggets of information.

Data mining can be used in law enforcement to discover new patterns or confirm suspected patterns or trends. One of the strengths of data mining, as opposed to more traditional statistical methods, is that it is not necessary to know exactly what you are looking for before you start. Data mining uses powerful analytic tools to quickly and thoroughly explore mountains of data and pull out the valuable, usable information. The primary use of data mining is to find something new in the data-to discover a new piece of information that no one knew previously. This is sometimes referred to as the bottom-up or data-driven approach because you start with the data and then build theories based on discovered patterns or trends. This approach tends to rely more on methods derived from computer science research, such as neural networks and genetic algorithms, as well as visualization techniques like link charts or graphs.

On the other hand, confirmation starts with an idea about a possible relationship (a hypothesis) and seeks to verify or refute the hypothesis based on the data. This application is strongly related to traditional statistical approaches, with an emphasis on hypothesis testing and model building. This is sometimes referred to as top-down or theory-driven data analysis because you start with a hypothesis and then check the data to determine whether it is consistent with the hypothesis.

The discovery and confirmation approaches are complementary and are often applied in alternating sequence. Discovery methods can be used to detect new patterns or relationships, while confirmation methods can be used to make sure that the new patterns identified are reliable. As you build the knowledge base by adding new information, new questions arise, and these new questions lead to new avenues of exploration and discovery.

Perhaps the most important required skill for data mining is domain knowledge, which simply means that you know your field and are in the position to evaluate the value or validity of the results. For example, we could find a strong relationship between icy roads and car wrecks, but it would have little value in policing because it is such an obvious relationship. On the other hand, identifying a relationship between cold weather and auto thefts associated with people preheating their vehicles could allow us to proactively deploy patrol units to residential neighborhoods and perhaps even deter auto thefts when the temperature dips below a certain level.

One of the biggest challenges in using data mining and predictive analytics in law enforcement is that most, if not all, data encountered was never intended to be analyzed. Therefore, significant challenges associated with data form, content, reliability, and validity must be constantly evaluated and addressed.

One common example of this issue can be found in police calls for service or dispatch data. It is not uncommon for multiple callers to report the same incident of shots fired. To get an accurate representation of illegal firearms use in the community, however, these duplicate calls should be identified and deleted. The dispatch of multiple units to the same incident, on the other hand, might seem redundant; however, any analysis of patrol deployment and workload will require the retention of this duplication in order to get an accurate representation of resource allocation.

The data culling, management, and descriptive analysis features offered by commercially available data mining packages are extremely valuable to crime and intelligence analysts, frequently offering the first accurate view of some data previously unapproachable. Again, domain expertise is absolutely essential to ensure that the results are accurate, reliable, and of value to the organization.

Finally, data mining tools can enable analysts to merge and analyze data resources that do not traditionally coexist in the same environment. This type of value added analysis is a huge asset and allows the analyst to see and describe the big picture in related incidents or data resources, providing a more complete view of events or activities. One area that has been particularly helpful in our experience has been the combined analysis of weapon recovery data and violent crimes information.

At the Richmond Police Department we selected the SPSS Incorporated data mining workbench Clementine, but there are a number of other data mining tools available to law enforcement organizations. We have used Clementine in a variety of analytical applications, including the examples outlined below.

Deployment

The largest portion of most police budgets is devoted to personnel costs. With decreasing fiscal resources and expanded purview related to the war on terrorism police personnel are increasingly being asked to do more with less. Moreover, issues related to recruitment, retention, and military activation are further stretching these already limited personnel resources. The ability to effectively deploy patrol personnel when and where they are likely to be needed can help us better use this increasingly limited resource while enhancing public safety, the ultimate goal of policing. The challenge to police managers and command staff is to make smarter deployment decisions in an effort to ensure that their personnel resources are available when and where they are needed, while minimizing resource deployment when the demand is low.

A quick analysis of the problem reveals that time of day frequently is associated with police workload. For example, many more citizen complaints requiring police assistance are received at seven o'clock in the evening than at four o'clock in the morning. Further analysis of the data suggests that the nature of the calls also differs at different time of the day: officers might spend more time handling domestic issues in the evening, while alarm calls could account for a large percentage of calls during the midnight shift.

These different types of calls also require differing amounts of time to clear. Again, a domestic complaint would likely require a greater amount of time, perhaps even additional police units, while an alarm call might take only a few minutes to clear. Day of week also might come into play, particularly in communities with active nightclubs or bars.

We also have noted seasonal differences. For example, as the weather gets colder and people heat up their vehicles the number of vehicles stolen with keys increases. A similar trend can be anticipated in the warmer months as citizens leave their cars running in an effort to keep them cool. Changing drug patterns, trends, and prices also can have an impact on related crimes like

burglaries, armed robberies, and prostitution as users require increasing economic resources to supply their drug habit.

A thorough evaluation of all the variables likely to affect patrol personnel workload directly is complicated, requiring differential weighting of multiple factors under varying circumstances and conditions. This comprehensive analysis of citizen-initiated demand quickly became very complex, far exceeding the analytical capacity of the human brain or even traditional, computer-based methodologies.

Moreover, we have considered only the factors that we would expect to affect police calls for service. Frequently, it is the factors that we do not anticipate that can wreak havoc on even the most thoughtful deployment plans. Again, the discovery process associated with data mining affords police managers and command staff the opportunity to identify unusual or subtle patterns in very large datasets not readily apparent without the advanced methodologies incorporated in data mining and predictive analytics. These tools now give us the opportunity to analyze police calls for service data at a level previously unavailable to the law enforcement community and develop accurate and reliable models that significantly enhance patrol deployment decisions.

Risk-Based Deployment

Deployment of patrol units generally is linked to citizen-initiated calls or the anticipation of complaints. Like many other departments, the Richmond Police Department uses specialized tactical units that are proactively deployed to areas associated with specific challenges or issues, particularly illegal narcotics or violent crime.

One underlying assumption in the use of these tactical units is that they be placed in or near a location where it is likely that they will be needed so that they can respond rapidly when called. Through our involvement with Project Safe Neighborhoods, we have developed the concept of risk-based deployment.[1] Traditional approaches identify locations associated with an increased frequency of crime.

Using data mining and predictive analytics, however, we have developed models predicting areas at greater risk for violent crime. Specifically, we have evaluated robbery-related aggravated assaults and developed models of those armed robberies associated with an increased likelihood for an aggravated assault. Our findings have allowed us to identify subtle distinctions between areas associated with an increased number of armed robberies, as compared to those associated with an increased risk for escalation into an aggravated assault.

Moreover, the areas associated with increased risk also were somewhat smaller, facilitating a more concentrated deployment of the tactical units in the areas of greatest need. This information has been deployed in maps, which have been used to proactively deploy tactical units in the anticipation of an increased likelihood of violent crime.

Tactical Crime Analysis

Traditional tactical crime analysis generally involves creating a model that represents a crime or series of crimes, which can then be used to link cases, identify and apprehend suspects, and prevent future crimes. With tactical crime analysis, crimes or series of crimes can be characterized, linked, and even anticipated based on time of day, day of week, location, modus operandi, or many other variables.

Data mining tools can greatly enhance traditional methods by providing automated searches of extremely large datasets in search of subtle relationships, actionable patterns, and trends. These tools, when applied to tactical crime analysis, can be used to review large amounts of information and incorporate a vast array of variables, far beyond what a single analyst or analytical team or task force can accurately review. It also does this work in a timely fashion, which is critical to identifying and apprehending suspects before they can commit additional crimes.

Predictive analytics offers the added benefit of selectively weighting factors most likely to predict future events. In high profile cases or series, data mining tools also can be used to search large tip databases effectively; again, a task beyond the capacity of a single analyst or even a team of analysts or task force.

Behavioral Analysis of Violent Crime

Data mining also can be used to analyze and model violent crime. Behavior, even extremely violent or seemingly unusual criminal behavior, frequently can be modeled, anticipated, and even predicted, something that criminal investigative analysis or profiling exploits extremely well.

With data mining, we can discover relationships that might not be obvious or might be new to us. For example, using data mining we were able to discover a relationship between property crimes and stranger rapes.2

Through automated searches of large correctional databases we found that a prior property crime was a better predictor of a stranger rape than a prior sex offense. In fact, several high-profile sexually violent predators including Timothy Spencer, the first case where DNA evidence was used at trial, and Derrick Todd Lee, who was linked recently to a series of rape murders in Louisiana, had prior histories of burglary; and additional review of the sex offender literature supports this finding.

Similarly, characterizing normal can be very important in identifying suspicious or unusual behavior. In our experience, deviations from normal patterns of offending may indicate the potential for serious escalation. For example, most burglaries are economically motivated crimes. The successful burglar generally selects an unoccupied dwelling, which decreases the likelihood of detection and apprehension, and takes something that holds monetary value (such as a TV or electronics).

Upon further examination of the property crimes committed by some stranger rapists, we noted that many of their burglaries differed in that they seemed to preferentially target occupied dwellings and frequently stole property with little value, if they stole anything at all. Using this type of anomaly detection, we can identify unusual or suspicious incidents that are worthy of additional investigation and have been able to successfully identify cases associated with an increased risk for escalation based on subtle deviations from normal.

Therefore, another type of risk-based deployment includes situations where we believe that a certain type of incident or series has the potential for escalation based on our prior experience or the development of models. Ultimately, this gives us the opportunity to proactively deploy resources in an effort to prevent further escalation.

Another type of violent crime that the Richmond Police Department has controlled some success with is drug-related violence. Associated with increases in community violence, drug-related violence frequently presents challenges to investigators in that witnesses might be difficult to find, unreliable, or reluctant, and there often is no readily identifiable relationship between the victim and perpetrator. As such, drug-related violence has been associated with decreased solvability and can greatly reduce overall clearance rates.

Rapid development of a motive, however, can facilitate suspect identification and apprehension before a case grows cold. In some of our earliest work in this area, we were able to successfully model drug-related homicides using advanced statistical techniques and classify murders based on their motive.3

This preliminary work confirmed our ability to accurately and reliably model violent crime, and gave us some insight into the types of offenders likely to be involved in drug-related violence. Subsequent work in this area has been confined to information likely to emerge early in an investigation in an effort to assist in the identification of a motive and increase the pace of an investigation.

Risk and Threat Assessment

The ability to identify and characterize events or attributes associated with an increased threat level or risk gives agencies an analytical "crystal ball" to use for deployment, crime prevention, special operations, threat assessment, and forecasting. The Richmond Police Department has used data mining and predictive analytics to identify and characterize crime patterns and events associated with an increased risk of escalation associated with armed robberies and certain property crimes.

These tools represent a powerful tool for examining other potentially risky or threatening situations that law enforcement professionals encounter regularly including repeat domestic complaints, stalking, and other potentially volatile situations. Predictive analytics also can be used internally to examine policing issues such as the use of force and police pursuits in an effort to identify potential training needs or policy change.

Officer Safety

Data mining and predictive analytics can enhance officer safety as a direct result of the increased understanding of criminals, crime patterns, and trends. Understanding and characterizing how different factors might interact to create unsafe environments for police officers can result in operational changes that increase officer safety. For example, examination of victim risk factors revealed that drug dealers frequently use firearms for defensive purposes while violent offenders tend to use weapons for offensive purposes.4

Additional anecdotal information suggests that selection of specific weapons might be related to need; offenders carrying weapons for defensive purposes prefer reliable weapons that are easily concealed, while other more violent offenders may compromise reliability for a particularly menacing or popular weapon. Information like this can offer police professionals increased situational awareness as they approach and interact with different offender groups, ultimately resulting in enhanced officer safety.

Around-the-Clock Crime Analysis

As trauma surgeons have noted, violent crime frequently occurs at inconvenient times and during periods of low staffing.5 The same issue faces police managers. Many crimes are committed during evenings and weekends when analytical staffing is low. To wait for analytical personnel costs valuable time and can compromise the solvability of the case. One way to address this challenge has been to deploy the decision rules generated by data mining and predictive analytics directly to the sworn personnel. Using a Web-based tool, police managers, command staff, and other sworn personnel in the Richmond Police Department now are able to enter a small amount of relevant information and receive an analysis in a matter of seconds, regardless of the day or time. As need changes, additional models are prepared and deployed directly to the sworn personnel. Just as criminals know no boundaries, the possibilities for this type of tool are phenomenal, exploiting centralized or shared analytical capacity across geographical and jurisdictional boundaries.

Unlike business trends, where predictive variables such as demographic information are not likely to change, crime trends and patterns might change very rapidly and with some frequency in any community. Even seasonal changes or fluctuations in the availability or preference for illegal narcotics can be related to overall changes in crime patterns and trends.

Similarly, changes in groups of criminals, reorganization, or even arrests also can affect prevailing crime trends and patterns by removing key players or groups. In fact, rapid identification and apprehension of criminals will likely ensure that the models will change frequently, requiring ongoing analysis and reevaluation of the models. In fact, our experience with the risk-based deployment models developed for armed robbery-related aggravated assaults suggests that they should be refreshed approximately every six months.

Summary

Intelligent, timely and complete analysis of the thousands of incident reports, crime tips, and other pieces of information that law enforcement professionals confront everyday is critical to fighting crime. The massive volume of data law enforcement organizations work with on a daily basis requires a different approach to analysis. The Richmond Police Department has found data mining and predictive analytics to be the most effective approach to addressing the so-called volume challenge associated with this massive influx of information, and is pioneering the use of these tools in policing.6 The successful exploitation of data mining and predictive analytics in law enforcement and intelligence analysis truly represents a powerful new tool, as well as a significant paradigm shift for the police executive, raising the question, "Why just count crime, when you can predict it?"