

DATA MINING AND EXPERT SYSTEMS IN LAW ENFORCEMENT AGENCIES

Monica C. Holmes, Central Michigan University, holme1mc@cmich.edu
Diane D. Comstock-Davidson, Central Michigan University, comst1dd@cmich.edu
Roger L. Hayen, Central Michigan University, roger.hayen@cmich.edu

ABSTRACT

Criminal databases and artificial intelligence have proven to be effective in solving crime on a local basis. However, what is needed is an expert system that links law enforcement agencies nationwide. Evidence and clues relating to crime are located across numerous data bases and have no connection to each other. This research examines five different Law Enforcement Information Systems (LEIS) to determine current features and characteristics of these systems. Some systems have a narrow or local scope of use, whereas others are more widely used. Some utilized expert systems that implement artificial intelligence techniques while others have a data mining orientation. Traffic violations are identified as an important component for any LEIS, because routine traffic stops often lead to apprehension for more violent crimes. However, traffic violations appear in only the local LEIS reviewed in this study. Clearly, tools, techniques, and data can be shared in the development future LEIS.

Keywords: Law enforcement, expert systems, data mining

INTRODUCTION

Criminal databases and artificial intelligence have proven to be effective in solving crime on a local basis. However, what is needed is an expert system that links law enforcement agencies nationwide. Evidence and clues relating to crime are located across numerous data bases and have no connection to each other. In 2005, only 45.5 percent of violent crimes were cleared or solved and only 16.3 percent of property crimes were solved [13]. The estimated number of violent crimes was 1,390,695. This means that 632,766 violent crimes were solved and that an estimated 757,928 violent crimes (54 percent) were NOT solved [7]. Clearly, this illustrates the need for an integrated crime fighting system. Artificial intelligence programs could be applied to the data to detect patterns, which could be used to link evidence to suspects [8]. Since September 11, 2001, in an effort to protect the citizens of the United States, data on crimes - large and small - have been gathered.

However, people feel that this information could prove harmful to innocent victims [10].

Artificial intelligence is “the ability of a machine to perform at the level of a human expert” [6]. An Expert System is “a program which has a wide base of knowledge in a restricted domain, and uses complex inferential reasoning to perform tasks which a human expert could do” [9]. Therefore, an expert system is artificial intelligence stored in a computer about one specific problem or a group of associated problems in order that many individuals may access this knowledge without being experts on the subject themselves.

This research examines existing systems, such as COPLINK, an integrated information and knowledge-based system funded by the National Institute of Justice and the National Science Foundation [5]; and SHERPA, an integrated system developed to investigate drug crimes [1]. This research effort looks at the available local and national systems to explore the benefits currently being provided and the opportunities for improvement. Finally, recommendations are provided for a comprehensive national expert system that can not only record crimes (past and present), but will also be able to ‘predict’ and guide criminologists in their task of keeping our streets safe.

LITERATURE REVIEW

Several Law Enforcement Information Systems (LEIS) use experts systems with the related artificial intelligence to solve crimes. Other LEIS apply data mining and related data management in the support of crime investigation. Five of these LEIS are COPLINK, SHERPA, MATRIX, ViCAP, and ViCLAS. COPLINK, an integrative information and knowledge management system developed at the University of Arizona, in collaboration with the Tucson and Phoenix police departments, is being used to share information about various crimes in order to link suspects and therefore solve more crimes [5]. In Wisconsin, a multi-agent information system called SHERPA was developed to help the Division of Narcotics Enforcement investigate drug crimes [1]. The FBI provides law enforcement agencies with ViCAP for investigating violent crimes

against people [7]. MATRIX is a collaborative state program that shares information among its partners for crime investigations [12]. ViCLAS, was developed in Canada to help the Royal Canadian Mounted Police solve crimes that could be ten years or older [2], [3]. The characteristics of each of these crime analysis systems are examined in more detail. First, those that apply expert system/artificial intelligence are considered, and second, those with a data mining orientation are assessed.

EXPERT SYSTEM BASED

Some LEIS use expert systems that implement artificial intelligence technologies to solve crimes. Two of those are ViCAP sponsored by the FBI and COPLINK from the University of Arizona. The features of each of these are considered in more detail.

ViCAP

ViCAP or the Violent Crime Apprehension Program is a national program developed and sponsored by the FBI's National Center for the Analysis of Violent Crime (NCAVC). The NCAVC is organized into three components: Behavioral Analysis Unit (BAU), Child Abduction Serial Murder Investigative Resources Center (CASMIRC), and Violent Criminal Apprehension Program (ViCAP). Cases in the ViCAP database include (1) homicides or attempted homicides, (solved or unsolved), especially if they involved an abduction, appear random or sexually oriented or appear to be part of a series; (2) missing persons, where foul play may be possible and the victim is still missing; (3) unidentified deceased bodies, if a homicide is suspected; and (4) sexual assault cases [7].

The FBI provides the software to state and local law enforcement agencies at no charge. Any case meeting the ViCAP criteria can be entered into the ViCAP system by police investigators for comparison to the database and any possible matching with unsolved cases. Once a case is entered into ViCAP, it is compared continuously against all other entries on the basis of certain aspects of the crime. This process is to detect unique aspects or traits of homicide and patterns of modus operandi (MOs), which assist ViCAP personnel in identifying crimes that may have been committed by the same individual. Law enforcement agencies are notified if a pattern is found and they then can pursue this lead in their investigation [7].

COPLINK

COPLINK, deployed in 2001, is an integrated information and knowledge management system created to capture, analyze and share law enforcement-related information. Developed at the University of Arizona's Artificial Intelligence Lab in collaboration with the Tucson and Phoenix police departments, it is funded by the National Institute of Justice and the NSF [5]. COPLINK is made up of two components. COPLINK Connect allows police departments to share data with each other through a user-friendly interface. COPLINK Detect provides patterns of various crimes entered from police databases. There are four types of searches available using COPLINK Connect. The search types are by person, by vehicle, by incident and by location. These types of searches are the standard types that police officers perform [5].

COPLINK Detect shares the same record information as Connect. However Detect uses a new set of intelligence tools for its users. Detailed criminal case reports and the significant terms associated with each case are the base of the Detect module. The reports store both structured data (i.e. name, address, date, etc.) and unstructured data (narratives about the case written by officers). Another module, COPLINK Collaboration is in the process of being developed that will enable the sharing of crime information among team members. Also, these features of data searching and detection give COPLINK a data mining capability. That is, artificial intelligence techniques are applied in searching the crime data base [5]. This potential of COPLINK is explored in more detail in the next section.

DATA MINING ORIENTED

Researchers at the University of Arizona propose that data mining is a powerful tool that can enable criminal investigators, especially those who lack training as data analysts, to search large databases quickly and thoroughly. Since computers can process thousands of calculations in seconds, this would save a lot of time [4]. Data mining is the process of identifying patterns by searching large volumes of data using techniques such as classification and prediction, cluster analysis, and association analysis [14]. Other, newer techniques include entity extraction, clustering, association rule mining, sequential pattern mining, and deviation detection. Entity extraction can automatically identify persons, vehicles or addresses from narrative police reports or images. Overall, Figure 1 shows these relationships between analytical capabilities and crime types. String Comparator compares text

fields in pairs of database records to measure the similarities between the records, detecting invalid information such as name, address and Social Security number in a criminal's record. Social Network Analysis describes the roles and interactions of criminals in relation to the flow of information and goods. This 'network' description can be the key to solving crimes in relation to gangs and crime 'bosses' [4].

COPLINK adopted a modified version of the entity extractor system, which uses a three-part process to identify names, locations and organizations in a document. First, noun phrases are identified, using linguistic rules. Next, a score is calculated based on pattern matching. Lastly, it uses a 'feedforward/backpropagation' neural network to predict the most likely entity type for each pattern [4].

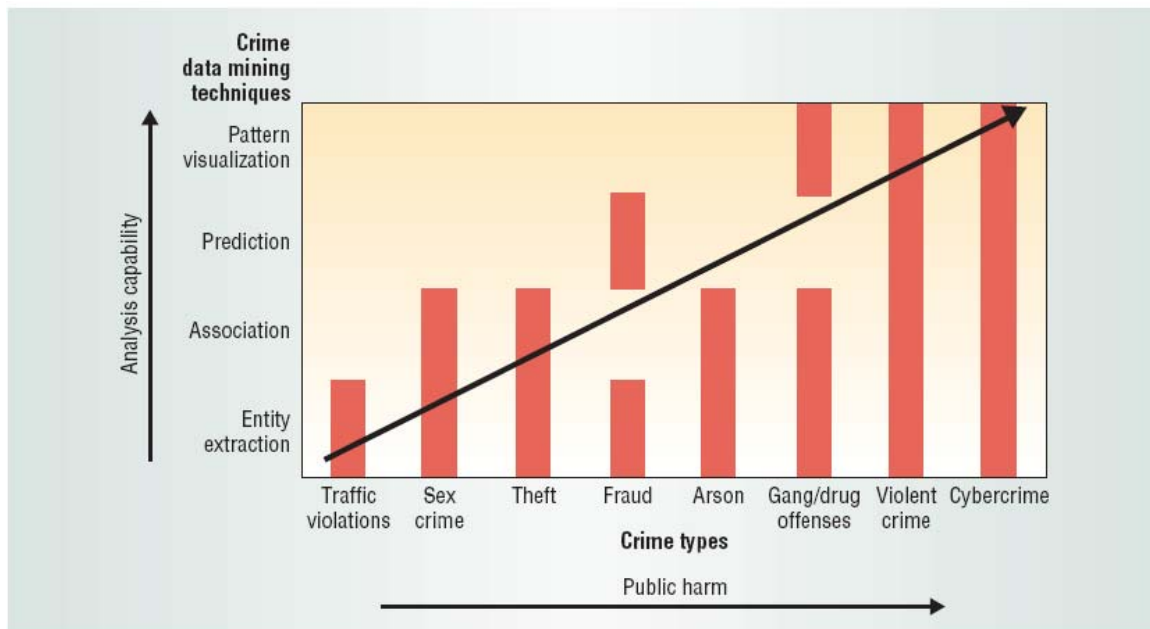


Figure 1: Crime data mining framework used by COPLINK [4]

SHERPA

The Wisconsin Division of Narcotics Enforcement has developed a distributed problem-solving multi-agent information system called SHERPA that is used to investigate crimes in the state of Wisconsin. SHERPA was designed to be a decision-support tool that can combine data from many sources to aid in the investigation of drug cases [1].

In designing SHERPA, the focus was on the data-analysis involved in drug investigations. The major problem can be sub-divided into sub-problems. Sub-problems might need different data to solve, since each sub-problem is independent of the others this is possible. The typical data sources needed are financial data, toll records, surveillance data, and other sources, such as checks of mail or trash [1]. In developing SHERPA, three problem-solving modules (agents) were used to analyze the data. After

analyzing the data, the system combines all solutions to identify a suspect. Figure 2 shows the architecture of SHERPA.

SHERPA contains 3 modules in total (Figure 2). The Language System handles the human-computer interaction used to input data and obtain output information. The Problem-Processing System compares past classification of information with current classification of information. The Meta-Level Knowledge System provides knowledge about knowledge, which means that it determines the importance of different knowledge sources when solving the problem. The primary goal of developing SHERPA was to increase the effectiveness and efficiency of the process of drug-crime data-analysis, providing more evidence and identifying more criminals.

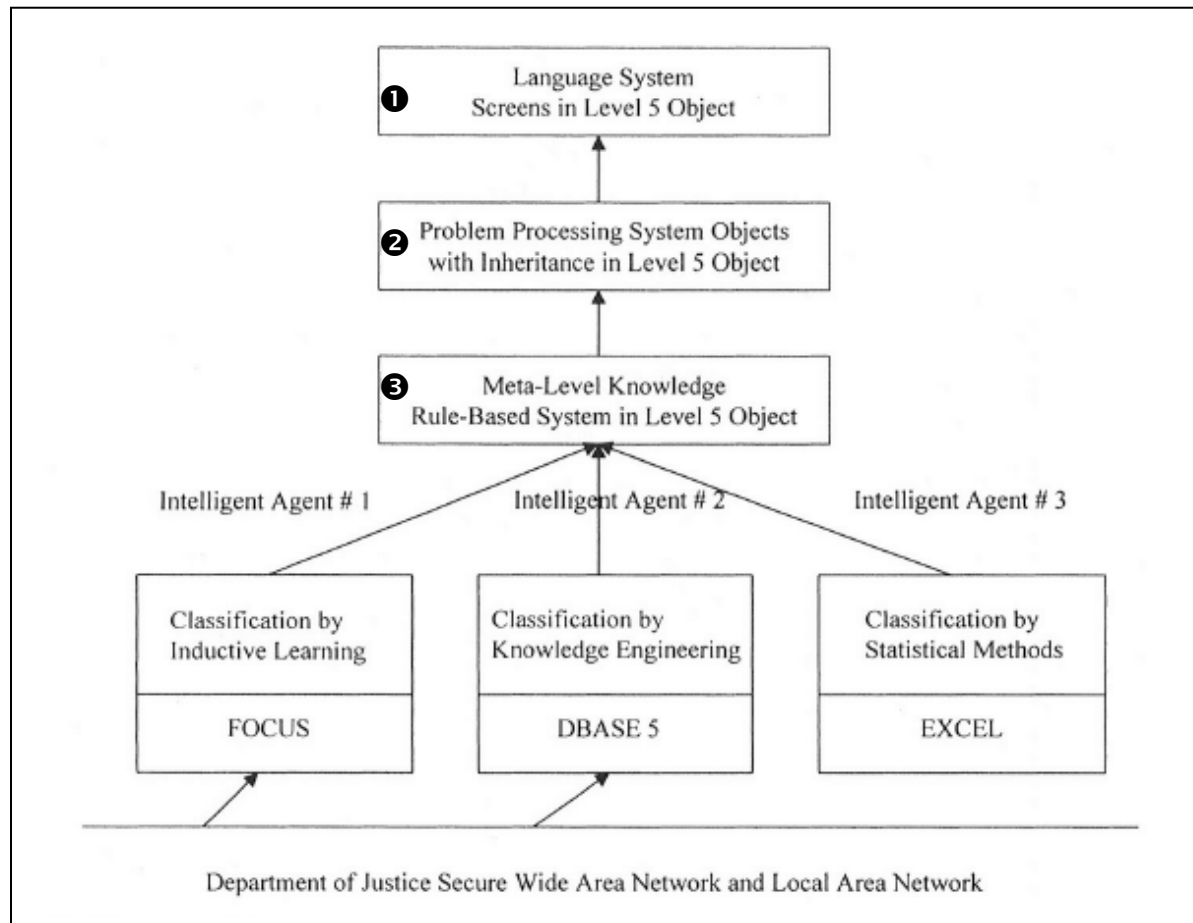


Figure 2. SHERPA Architecture [1]

MATRIX

In 2002, 16 states agreed to participate in a program called Matrix or Multistate Anti-Terrorism Information Exchange, an anti-crime database developed in the wake of 9/11. The system allows states to share vehicle, prison and criminal information with each other. The American Civil Liberties Union feels that Matrix could be abused by compiling information about people who have never been suspected of a crime. Matrix is run by a private company, Seisint, Inc. of Boca Raton, Florida on behalf of the participating state governments with partial funding from the federal government. Eleven of the original states are no longer participating in this multi-state project; perhaps because of the concerns that many have about privacy issues and how databases used by law enforcement agencies

should take measures to make sure that the data is not compromised in any way [12].

ViCLAS

The Canadian government took three years to develop their own crime database – ViCLAS or Violent Crime Linkage Analysis System. Unveiled in 1991, this database stores information on homicides, sexual assaults, missing persons and human remains, and identifies information that links solved and unsolved crimes. A crime analysis questionnaire of 263 questions is entered into a central database. In 1996 this database held over 15,000 entries. Because serial killers strike over and over again, storing information about unsolved homicides and then comparing to newly committed murders increases the chance of finding these repeat offenders [2], [3], [11]. As of May 1997, there were over 20,000 cases entered into the database. Over

3,200 linkages have been made. Of those linkages, 80% were confirmed and 3% were potential links.

ViCLAS is being used in a number of other countries, including Belgium, Austria, Australia, Holland and the United Kingdom. Also the states of Tennessee and Indiana are using it. The Royal Canadian Mounted Police (RCMP) has been giving it away for free. The negatives of the system have been identified as: (1) investigator resistance – to filling out 263 questions per violent case is a road block; (2) key fact evidence is the practice of holding back key evidence to help with the investigation, and investigators are hesitant to put this 'key evidence' into ViCLAS; and (3) the Information Privacy Acts and Privacy Legislation could slow down the use of this system to help fight violent crime [11].

ANALYSIS

Several LEIS utilize expert systems with artificial intelligence and data mining technologies and techniques to assist in crime investigations. The utilization of these systems by law enforcement agencies is elective. That is, they choose whether they will provide data and use that data for crime investigation. Of course, this greatly limits the available amount of information that can be accessed

and analyzed. Table 1 summarizes the deployment of these technologies in LEIS. Of the systems reviewed, different combinations of these technologies were found. Further, these systems focus on different crime types. COPLINK and MATRIX both includes a vehicle component, which appears as a significant element in locating criminals. That is, criminals are frequently located as a result of a traffic violation. Then, this leads to their further identification of being wanted for other crimes. There is no national level LEIS available to link all crimes, including traffic violations, which are quite often the key to apprehending much more dangerous criminals.

All of these LEIS utilize historic and current information for analysis. However, data mining provides the ability for the prediction of future behaviors. So far, based on the LEIS examined here, there are no systems that have implemented data mining technologies and techniques to predict future behaviors. On the other hand, it may be that law enforcement agencies are so overwhelmed in solving current crimes that future behaviors are beyond their capabilities.

Table 1: Law Enforcement Information Systems Classification

System	System Type	Crime Type	Geographic Coverage	Geographic Use Areas
ViCAP	ES/AI	homicide, abduction, missing person, sexual assault	wide	US provided to state and local agencies
COPLINK	ES/AI, DM	person, incident, vehicle, location	narrow	Arizona major metropolitan areas
SHERPA	ES/AI, DM	drug	narrow	State of Wisconsin
MATRIX	DM	person, vehicle	narrow	5 states
ViCLAS	DM	homicide, missing person, sexual assault	wide	Canada, Belgium, Austria, Australia, Holland, United Kingdom, 2 US states

As shown in Table 1, there is a mix of both wide and narrow geographic coverage. Narrow coverage indicates only state or local deployment of the LEIS. Clearly, ViCLAS has the broadest geographic area of

use. This would also indicate this system meets many of the requirements found in different countries. However, ViCLAS and MATRIX have significant concerns with privacy acts and legislation

that threaten to slow down their expansion and growth in solving crimes.

CONCLUSIONS AND RECOMMENDATIONS

This research has looked at several crime databases. Some based on expert systems that implement artificial intelligence techniques. Others employ data mining tools and techniques. The United States government has developed ViCAP, which stores only violent crimes and is available to all state and local law enforcement agencies, but it is not mandatory to use. Not being a mandatory participation system for state and local law enforcement appears as a major weakness of the program, along with the fact that this program stores information about 'violent' crimes and therefore leaves a wide gap by not including burglary, traffic violations, and other 'minor' crimes in the database. Many violent criminals are caught in routine traffic stops, if the crime was in the same state as where they were stopped.

The United States should study Canada's ViCLAS to see how they could improve ViCAP. ViCLAS has been used successfully in identifying suspects in solving crimes. The system has worked so well that other countries (and 2 US states) are currently using this software. Hence, ViCAP could benefit by trying to incorporate the successful aspects of Canada's ViCLAS system. Minor crime data, such as traffic violations should be included in the database. This would provide more information for linking crimes nationwide. ViCAP use should be mandatory for state and local agencies, so that all agencies are both using and contributing to the system. The interface should be user-friendly, with fewer questions to be answered for each crime.

MATRIX, COPLINK, and SHERPA have limited areas of use. COPLINK is used only in Arizona. SHERPA was developed by the State of Wisconsin to be used in their drug enforcement program. MATRIX was developed in the wake of 9/11, when Homeland Security was looking at any opportunity to monitor and detect terrorists or future terrorists. This system has come under fire with concerns of invasion of privacy, especially since individuals could be included in this database without ever having been convicted of any crime. Most of the 16 states that originally signed up to participate and share data have since left this group. Nonetheless, these LEIS should be examined for features that might be included in those systems that are used more widely.

One of the biggest drawbacks of any of these existing systems is the large amount of questions that must be answered. A true expert system would be able to guide the individual interactively, by asking general

questions and based on those initial questions, more specific questions would be asked. Hundreds of questions to be answered for each crime would be very discouraging and would not encourage officers to enter all crimes.

In conclusion, a true 'expert system' for solving crime would be a valuable tool to solving crime in the United States. Even if only one more case were solved that would be a benefit for this country. But the work of enrolling all states in the program is tough in a country which values freedom. This includes the freedom of a state to participate in a national crime solving system. Clearly, LEIS can assist criminologists in their task of keeping our streets safe.

REFERENCES

1. Bhaskar, R., & Pendharkar, P. C. (1999). The Wisconsin division of narcotics enforcement uses multi-agent information systems to investigate drug crimes. *Interfaces*, 29(3), 77-87.
2. Burger, D. (1997). RCMP turns to high-tech in bid to get their man. *Computing Canada*, 23(8), 1-2.
3. *Canadian Business*. (1996). Looking for Mr Wrong: not all criminals leave fingerprints A new police tool looks for deeper clues (Violent Crime Linkage Analysis System), 69(1), 15-16.
4. Chen, H., Chung, W., Xu, J.J., Wang, G., Qin, Y. & Chau, M. (2004). Crime data mining: A general framework and some examples. *Computer*, 4, 50-55.
5. Chen, H., Zeng, D., Atabakhsh, H., Wyzga, W., & Schroeder, J. (2003). Coplink. Managing law enforcement data and knowledge. *Communications of the ACM*, 46(1), 28-34.
6. Dharmaraja, R. (2002). Is the debt-collection industry a suitable candidate for artificial intelligence. *Business Credit*, 104(1), 39-42.
7. Federal Bureau of Investigation. (2006). *Investigative Programs, Critical Incident Response Group*, National Center for the Analysis of Violent Crime, Retrieved December 1, 2006, from <http://www.fbi.gov/hq/isd/cirg/ncavc.htm>,
8. Kurlander, N. (2005). The crime database battle. *Security*, 42(10), 30-32.

9. Metaxiotis, J.P., & Psarras, J. (2003). Expert systems in business: Applications and future directions for the operations researcher. *Industrial Management & Data Systems*, 103(5/6), 361-368.
10. Munro, N. (2002). The ever-expanding network of local and federal databases. *Communications of the ACM*, 45(7), 17-19.
11. Royal Canadian Mounted Police. (2006). *Violent Crime Linkage Analysis System (ViCLAS)*. Retrieved December 1, 2006, from http://www.rcmp-grc.gc.ca/viclas/viclas_e.htm.
12. Swartz, N. (2004). Anti-crime database losing support. *Information Management Journal*, 38(3), 16-17.
13. U.S. Department of Justice. (2006). FBI releases its 2005 crime statistics. September 18, 2006.
14. Wikipedia. (2007). *Data Mining*. accessed Retrieved March 10, 2007, from http://en.wikipedia.org/w/index.php?title=Data_mining.