International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 - 5161

Available at http://inpressco.com/category/ijcet

General Article

Big Data to Combat Crime

Rishikesh C. Muchhala[Å*], Gaurang A. Shah[Å] and Prayag Shah[Å]

[Å]InformationTechnology Department, DJSCOE, Vile Parle (W), Mumbai -400056, India

Accepted 10 Oct 2014, Available online 20 Oct 2014, **Vol.4, No.5 (Oct 2014)**

Abstract

Big Data *analysis and the prediction of future outcomes based on the results of analysis obtained have found applications in business, finance, healthcare, etc. Similarly Big Data also found its way to* ***fight crime****. Various* ***Law Enforcement Agencies*** *(LEA) throughout the world have incorporated the use of Big Data technology to gather, store and retrieve large amounts of data in order to* ***draw out logical conclusions*** *by analyzing them. In this paper, we discuss the various dimensions explored by* ***crime analysts*** *to build systems which not only gather* ***structured and unstructured information*** *but also give/trigger solutions for the same in* ***real time****.*

Keywords: *Big Data, Crime, Law Enforcement, Analytics*

1. Introducing the Problem

A lot has been achieved, ever since the advent of Information Technology (IT), in terms of enhancing the operating methods within the organizations by using Technology. Soon enough, the same organizations began concentrating on the Information part of IT. With the rise in the number of computer users, smart phone users, internet and social media, the rate at which information was being generated overwhelmed everyone. This gave rise to a new way of storing and retrieving data. This new way had to be a step above the traditional relational database management systems (RDBMS). The reason was simple – we no longer generate structured form of data. We need a system which can store data of unstructured form. We needed a system which could handle huge **Volume** of data, which came in at a tremendous **Velocity,** and which had a lot of **Variety** in terms of data structures. These are the three V's which form the ethos of **Big Data** technology.

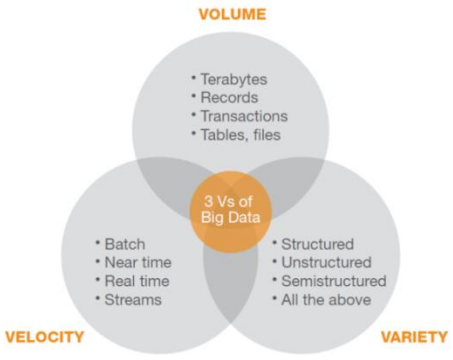

Fig 1: Three V's of Big Data

Information is no longer viewed as only the things that you type into your system and store. Information is everything that you do. Social networking websites like Facebook generate 500 terabytes of data a day. It really is a big task to manage such a huge volume of information. This information is in terms of texts, images, videos, location, etc. making it a very diverse form of information. And finally when we talk about velocity, there were 1155 million active Facebook users in the 2nd quarter of 2014. Which means Facebook has to not only store large variegated data, but it has to do it really fast. In fact, the experiments conducted at the CERN laboratories for the Large Hadron Collider (LHC) generate approximately 1 petabyte of data per second. And retrieval of this data is again equally as important. This meant Big Data would require large scale parallel processing capabilities.

Fig 2: Large Hadron Collider

Eventually, experts at statistics and analytics started using this technology to study the patterns evolving out of the information gathered. These patterns identified from the past helped them predict the future trends. Which meant what we do and what we choose in one aspect of our life can be used by analysts to predict what we may do and what we may choose in some other aspect of our life. This kind of predictive science witnessed increasing demand from extremely diverse disciplines. Organizations began

*Corresponding author: **Rishikesh C. Muchhala**

using data analyses to predict the favorable seasons, upcoming trends, target clients, probable locations, etc. One such discipline is **Criminology and Law Enforcement**.

2. Applying Big Data to Combat Crime

There are 4 basic steps involved:
1) Gathering information.
2) Analyzing information.
3) Predicting the crimes.
4) Preventing the crimes.

We will learn in the following section that different technologies and systems use different kinds of information. And the devices/interfaces used for gathering this information also vary from application to application. The analysis of this information is done using histograms, pie charts, Venn diagrams, bar graphs, etc. The tool or the software takes care of synthesizing and representing the information. It takes considerable human expertise though in the remaining two steps. The user of the system has to draw out logical conclusions from these representations and determine the probability of the crime to occur. The results of Big Data analysis and the real world scenario must together be considered to arrive at a prediction. And the final step is to do everything possible within the legal and feasible boundaries to prevent the crime. This kind of system is called Intelligence Led Policing (ILP). Let's see the existing technologies available:

Fig 3: Intelligence Led Policing

2.1 DNA samples from the crime scenes

DNA of the criminals from the crime scenes can be obtained in the most unusual of the ways. DNA is generally obtained either from the fluids that human secrete or from fragments of skin. Which means fluid such blood, spit, saliva can be used by the forensic experts to study the DNA. These fluids are obtained from rims of glasses, spoons, knives, ends of cigarette etc. Fragments of skin can be obtained from underneath the fingernails of the victim or from the clothes, cushions, etc. Hair samples may also be preserved by the forensic team. In England and Wales, anyone arrested on suspicion of a recordable offence must submit a DNA sample. In Scotland, the law requires the DNA profiles of most people who are acquitted be removed from the database. In Sweden, only the DNA profiles of criminals who have spent more than two years in prison are stored. In Norway and Germany, court orders are required, and are only available, respectively, for serious offenders and for those convicted of certain offences and who are likely to reoffend. Forty-nine states in the USA store DNA profiles of violent offenders. The United States maintains the largest DNA database in the world, with the Combined DNA Index System (CODIS) holding over 9 million records as of 2011. All these records are stored using Big Data. Using these samples the investigators can conclude whether the DNA sample is that of an innocent bystander or of a person who already has a criminal record.

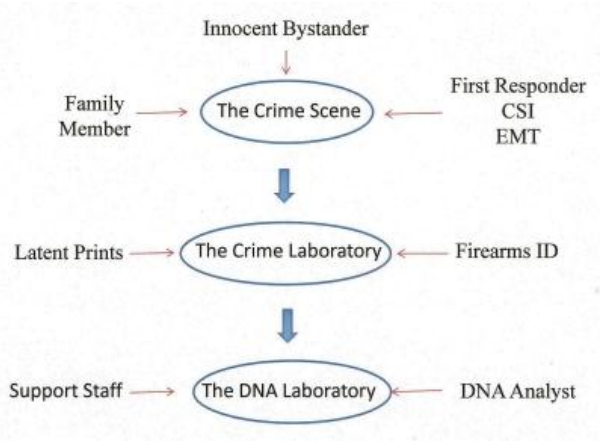

Fig 4: DNA evaluation

2.2 Gunfire Sensors/ Locators

The United States have successfully implemented this technology. These gunfire sensors can be used to cover small public places like parks, parking lots, museums, etc. or they can also be used to cover small townships, high profile complexes, military operating bases, etc.

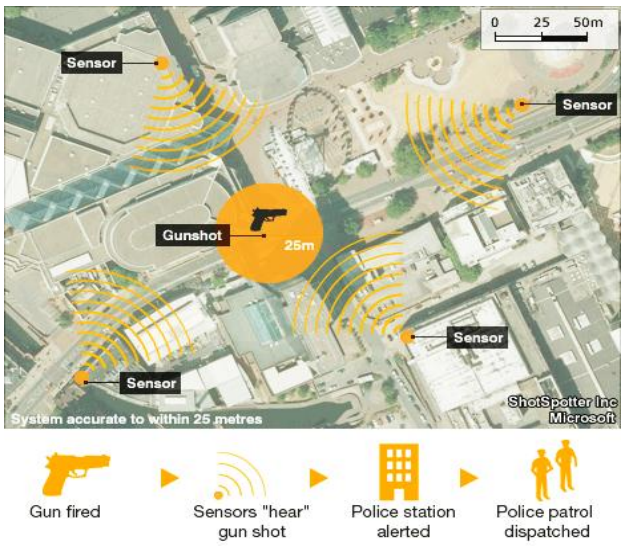

Fig 5: Gunfire Sensors

These sensors can not only detect a gunshot but they can also give the geographical location of incident. Gunfire acoustics is studied and the sensors detect either the sound of the gunfire or the trajectory of the gun shot. A typical gunshot produces 120 to 160 decibels of sound. Also, if

the bullet travels within the specified range of the sensors, the sensor can sense infrared movements (optics) of the bullet. Once a gunshot has been identified, the authorities are notified and in some cases, the nearby cameras are immediately aligned to capture the image. This is done using Geographic Information System (GIS). Now, the collection of so much of data in terms of sound, light, position, etc. needs Big Data technology. Such a system is a part of ILP (intelligence led policing).

2.3 Biometrics for antitheft

Biometrics can be used in a number of ways to prevent theft. Modern day researchers have proposed to use sensors in and around the driver's seat.

Fig 6: Sensors on a Gun

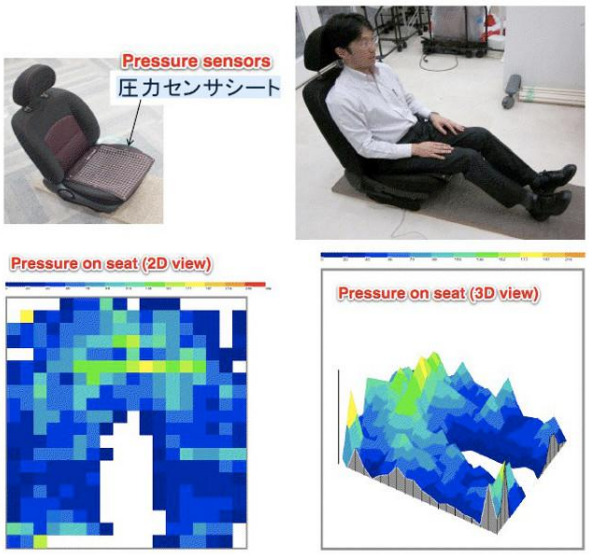

Fig 7: Car Anti theft sytem

The entire biometric information of the driver will be recorded and stored in the system. This information includes weight, body dynamics, retina scan, finger print detector at the steering wheel, etc. If the person who tries to drive the car does not have his/her biometric information pre-fed into the system, the system will immediately lock the doors of the car and may alert the police using Global Positioning System (GPS). The driver is locked inside unless a password is entered into the system by someone whose biometric information is identified by the system. Similar type of biometric system is proposed to be used for guns as well. The gun will not fire if it is not held by the owner of the gun. Such biometric systems in fact produce large amounts of information every time a request is triggered. This information is generated extremely rapidly and is not of a uniform datatype.

2.4 Image identification

Image identification is in itself a vast branch. Image identification includes face recognition, video imaging, Automated Number Plate Recognition (ANPR), etc. The law enforcement agencies which have a digitized collection of images of all the citizens can benefit a lot from such systems. Facial recognition features are used by social networking sites like Facebook. And over the years, these systems have become fairly accurate. Identifying faces from a moving background (like a motion picture) is a great boon as it helps in getting the identity of the identified face if there exists a comprehensive database. This implies that practically any video recording done of a crime can be used to identify the wrongdoer. Also, systems like ANPR can be extremely useful. It can help the police keep a track of all the cars going out and coming into their jurisdiction. Traditional systems only photographed the number plate and the car. But ANPR gives the entire identification of the owner of the car. Such systems help us correlate vehicle movements with the crimes. Big data makes it possible.

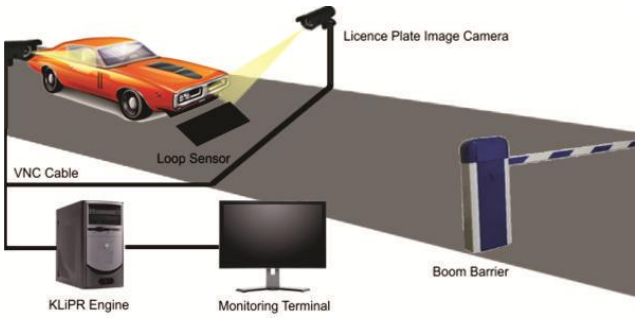

Fig 8: Automated Number Plate Recognition

2.5 Zoning applications

The zoning applications are the ones which truly incorporate the step of predicting. All the previous applications that we studied had gathering, analyzing and preventing methods. Zoning applications are successful if the Law Enforcing Agencies (LEA) have a strong GPS or any GIS system. The zoning applications have set parameters where it records particular incidences. A given zoning application may be used by the traffic police to store information about recorded accidents, traffic congestions, traffic law violations, etc. These incidences are all recorded along with the location on a regular basis. Using this information, the analysts may prepare

representations of what type of crimes occur at which part of the city on which day of the week on what time. Using this, it becomes easy for the government officials and sometimes also the policy makers to troubleshoot the civil issues if any. It may be noted that traffic jams are reported during school time. In this case, the school buses maybe asked to explore alternate routes. Or it may be observed that traffic laws are violated the most during the morning on the weekdays. When the systems come up with such statistical results, the authorities can take the required steps with much ease and with a sense of direction.

Similarly, we can have a system which keeps track of all the crimes location wise all throughout the year. Now it may be noticed that during the festivals that shopkeepers report more number of cases of shoplifting. Or it may be noticed that the incidents of juvenile driving under influence (DIU) increases in a particular area after midnight once the semester gets over. Using this data, the analysts can cogently draw out results and take pro-active measures. This is how we can predict the future crimes and help prevent potential damages.

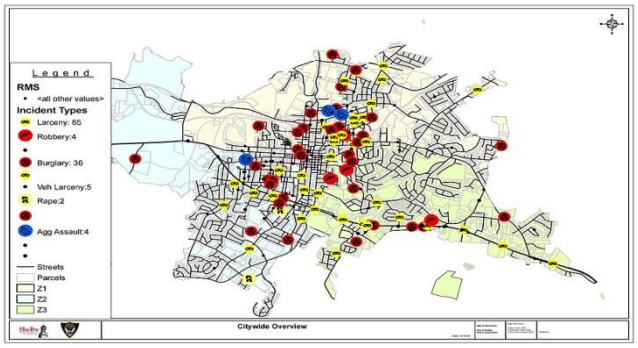

Fig 9: Crime Areas Highlighted

Conclusions

Advanced systems use advanced tools to predict crimes. They also make year on year or month on month comparison to determine whether the crime rate is going up or down. Big companies like IBM, Dell, Google, Microsoft invest billions of dollars in research and development of Big Data technologies. The developers have their own communities for technologies like Hadoop, NoSQL, Real time, machine learning, etc. So much has been achieved already in terms of developing and deploying technologies which help combat crimes. But there are challenges to face. There still are a lot of dimensions within the domain of information utilization which are waiting to be exploited which will help to record, identify, predict and prevent the crimes. Also, with improving technologies, the organizations will have to train themselves to adapt to newer ways of dealing with crime, because technology will help to drastically reduce the operating cost and to enhance the investigation.

References

https://gigaom.com/2012/08/22/facebook-is-collecting-your-data-500-terabytes-a-day/
http://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/
http://home.web.cern.ch/about/updates/2013/04/animation-shows-lhc-data-processing
http://www.kritikalsolutions.com/images/stories/KLiPR_Image.jpg
http://news.bbcimg.co.uk/media/images/50358000/gif/_50358939_birmingham_sensor_graphic.gif