A Survey On Cyber-Crime Prediction Techniques

Kaveh Shaerpour, Ali Dehghantanha, Ramlan Mahmod

Abstract

This paper is a review paper of journal and conference papers published in the field of crime prediction. Crime prediction is a growing field in the field of prediction. More and more law enforcement agencies are using or going to implement crime prediction software. In this paper firstly, data collection methods and findings of this paper are described and afterward the collected papers are described and their methods and findings are explained.

Keywords: Crime Prediction, Spatial-Temporal Event Prediction, Cyber Crime Prediction Models, Criminal Geographical Profiling

1. Introduction

Crime prediction is one the newly developed concept and technology for law enforcement agencies. This technology is so new that only a handful of law enforcement agencies use it around the world. Los Angeles Police Department (LAPD) was one of the first law enforcement agencies to use a crime prediction tool. Crime prediction tools’ main objective is to predict when or where a crime is going to happen in future. These applications use different algorithms to forecast the future. Some of the algorithms are tailor-made for the purpose of predicting crimes and some are algorithms which are developed from other forecasting algorithms such flood or earthquake prediction.

This paper is intended to review journal and conference papers published in the field of crime prediction to give the reader a sense of what is happening in this field around the world.

2. Data collection

This section provides an overview of all the available journal and conference papers about Crime Prediction.

2.1. Topics covered in the journals and proceedings

To find out the journal papers some databases were searched and the best and most relevant papers were identified. These databases were “Science Direct”, “Springer”, “Taylor and Francis”, “IGI Global” and “IEEE”.

In each database there were some journals and conferences which had relevant papers about crime prediction. The journals were “IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans”, “Procedia Computer Science”, “International Journal of Information Security”, “Computational Management Science”, “Journal of the American Statistical Association”, and “International Journal of Applied Evolutionary Computation (IJAEC)”.

Conference papers used for this paper were extracted from following conferences “IEEE 2nd International Conference on Software Engineering and Service Science (ICSESS)”, “Ninth International Conference on Machine Learning and Cybernetics”, “The Ninth International Conference on Web-Age Information Management”, “IEEE International Conference on Computer Systems and Applications”, “International Conference on Convergence Information Technology”, “International Conference on Computational and Information Sciences (ICCIS)”, “IEEE Systems and Information Engineering Design Symposium (SIEDS)” and “IEEE International Conference on Intelligence and Security Informatics (ISI)”.

International Journal of Advancements in Computing Technology (IJACT) Volume 5, Number 14, October 2013
Overall 61 papers were found and after reading their abstracts and some brainstorming sessions 12 of those were chosen as related. Once related papers were collected, their keywords were extracted to find out the most frequent keywords across the papers. Figure 1 shows the frequency of the keywords.

![Figure 1. Keyword Frequency in Papers](image)

After the keywords of papers were extracted, they were categorized into three categories:
- Modeling Cyber Crime
- Modeling Physical Crime
- Crime Prediction

As it is visible in Figure 2, Crime Prediction category has 6 papers in it, Modeling Physical Crime category includes 4 papers and 2 papers belong to modeling cyber crime category.

![Figure 2. Paper Categories](image)
3. Discussion and analysis

In this section the results of each paper in each category will be discussed and at the end of category an analysis will be given.

3.1. Crime prediction

As it was explained in section 2.1 there were 6 papers found about crime prediction. These 6 papers are categorized in two sections

- Theoretical Papers: In this section papers which introduced models or frameworks are explained. 5 papers belong to this sub-category.
- Practical Papers: In this section papers that developed a tool and explained its functionalities or showed its performance will be explained. One paper will be explained in this section.

This categorization is to divide the papers into even more detailed and more related categories, so it is more understandable for the readers to look up for more relevant discussions.

3.1.1. Theoretical papers

In this section all the models and frameworks suggested by the papers belonging to this category will be explained. These papers are more focused on group based crimes like criminal gangs, spatial-temporal crime prediction, serial crimes such as serial burglary.

For predicting group related crime authors of [1] have provided three models to assess the threats from criminal gangs, insurgent groups and terror networks and then they combined the models to create greater models and at the end they measured the performance of each model.

Models that were explained in [1] are

- Pooled Model: It is a model in which all incidents, regardless of group, are used to fit the regression based upon the features of that location. This means that this model locates the incidents without considering the criminal group. The mathematical representation of this model is presented in formula (1).

\[
\pi(y_i = 1 | X) = \logit^{-1}(\alpha + \beta X_i).
\]  

(1)

- No-Pooling Model: Unlike pooled model, no-pooling model actually considers the group for which the threat assessment is done. This model is used for criminal preference models for specific groups. The mathematical representation of this model is shown in the formula (2).

\[
\pi(y_i = 1 | X) = \logit^{-1} \left( \alpha_{j|i} + \beta_{j|i} X_i \right).
\]  

(2)

- Multilevel Model: It is a 2 step model which uses several approximates. These approximate uses the Laplacian approximation to approximate the integrals in the likelihood function and minimize Green’s penalized quasi-likelihood function. More in depth explanations are provided in [1]. The mathematical equation of this model is quite difficult and it is in two formulas shown in formula (3a) and (3b).

\[
\pi_j \left( y_{ij} = 1 | X \right) = \mathcal{N} \left( \logit^{-1} \left[ \alpha_{j|i} + \beta_{j|i} X_i \right], \sigma^2_y \right)
\]

\[
\begin{pmatrix}
\alpha_j \\
\beta_j
\end{pmatrix}
\sim \mathcal{N} \left( \begin{pmatrix} \mu_{\alpha} \\ \mu_{\beta} \end{pmatrix}, \begin{pmatrix} \sigma^2_{\alpha} & \rho \sigma_{\alpha} \sigma_{\beta} \\ \rho \sigma_{\alpha} \sigma_{\beta} & \sigma^2_{\beta} \end{pmatrix} \right).
\]  

(3a)

(3b)
In this paper, the authors then combined these models to create more specific models to find out more detailed results for events based on groups and locations. After this step, the authors of [1] evaluated the combined models to find out the performance of each. This analysis was done with surveillance plot and sphere of influence analysis. The results of these tests show that the multilevel model provides more accurate results with much better performance.

For the topic of spatial-temporal crime prediction [2] discuss that many past researchers have either focused on spatial analysis or temporal analysis. Authors formulated the temporal indicators and then they formulated the spatial indicators based Generalized Linear Model (GLM) and they then create a new GLM by creating a new model of previous temporal and spatial formulas. The new formula or model is called GLM+.

The authors carried two types of test one simulated data set and one real life sample data set. The results show that in simulated data set test GLM+ shows 13% more accuracy compared to GLM and second test shows a 14% improvement over GLM. [2] shows the importance of combining both temporal and spatial indicators for predicting an accurate crime prediction but it does not show that whether these crimes were one time crimes or serial crimes.

For crime prediction for serial crimes [3] has proposed a serial crime prediction model based on Bayesian learning theory. In this paper, the authors have identified some factors related to the geographical distribution such as such as gender and age of the victims and characteristics of the crime site such as private or public where private includes private residences and apartments and public point towards to schools, subways and hospitals.

After considering the geographical factors two functions are introduced

- Distance decay function: It is to describe the geographical profile of each factor.
- Discrete distance decay function: To improve previous function they used the mixture of Gaussian distribution to approximate the discrete geographic distribution of the factors and weight each Gaussian distribution by the distance decay function.

By considering the factors and with two functions, now, they used Bayesian learning theory to create a model that learns from each related previous crimes that the serial criminal has done and based on those, the model will predict the next geographical site that the criminal may commit his next crime.

To test their models authors performed simulation in which they used a case study in which a serial killer had killed 9 women in Baiyin City in China in 1998. The results of the simulation showed that out of 3 predictions 2 were correct and only 1 was not.

This model only predicts the next crime site of a criminal and this crime site is not as accurate but highlights a small area for the law enforcement officers to search.

[4] is the next paper which discusses the crime prediction modelling for serial criminals. In this paper authors have chosen a different approach and they used psychology in order to find out a behavioral regularity for each criminal. In their discussion, the authors of [4] discuss that based on criminal psychology; criminals try to commit their crimes somewhere far from their homes or their workplace.

Another factor that they use is that when serial criminals commit an offense, they intend to follow a linear path to commit their next crimes. Lastly, serial criminals try to find a hiding zone where they hide in.

By using previously mentioned psychological facts they came up with seven models for different parts of the overall prediction process.

They modelled 4 models to predict the residence of the offender for which they used techniques such as circle fitting, linear fitting, probability theory. Each model could be used for different criminals.

A model for predicting the time of the next crime was suggested which uses linear fitting because the shape of the crime timeline is a straight line. According to this straight line they predict the next crime.

To predict the location of the next crime the authors used set theory and. And lastly they created a model to generate a geographical profile. To test their models, the authors used two case studies of past and they matched the results with the actual evidences and actual crime. The previous papers all developed a model for crime prediction but in [5] the authors want to show how other algorithms of prediction can be used in crime prediction models.

In [5] the authors show that highly clustered event sequences in some type of crimes such as burglary can be treated same as similar clustering patterns in seismology. In this research firstly, self-exciting point processes are reviewed in the context of seismology and then by using residential
burglary data that Los Angeles Police Department provided to the researchers, they illustrate the implementation of self-exciting point process models in the context of urban crime. The data that LAPD provided is about 5376 reported residential burglaries in an 18 km by 18 km region of the San Fernando Valley in Los Angeles. The tests show that 63% of the results are exact repeats of the actual crimes. This research actually shows that other types of prediction algorithms such as earthquake prediction algorithm can be used to predict similar crimes with similar attributes.

### 3.1.2. Practical papers

In previous section all the papers that were theoretical and develop models were discussed. In this section one paper is explained. In this paper a tool is developed.

In [6] authors developed a crime forecasting tool for law enforcement agencies. This tool is aimed to be integrated into WebCAT (Web-based Crime Analysis Toolkit) system which is used across U.S. state of Virginia.

In order to develop this tool, the developers had to choose three main components of the program which are algorithm, software technologies and data.

After testing different types of algorithms, the developers chose a general linear model (GLM) for their algorithm which is a linear DCM. The application was fully integrated into WebCAT and it is used across the state of Virginia. This program gives many advantages to the law enforcement agencies because by knowing where a crime is going to happen they can allocate their resources more efficiently.

### 3.1.3. Analysis

As it was explained in previous section many researchers are working on crime prediction models and tools. Although all of these models and tools are brilliant there are some problems that need to be addressed.

As it was explained before in section 3.1.1 the models that were developed are mostly only work with some very specific requirements. Some of the works are only used for group related crimes and they do not predict other types of crime. Some of the models are only working with serial crimes and they only work if there is a pattern and one person commits multiple criminal activities but this would not work on one time crimes. These models do not provide vast choices of crime to predict for law enforcement agencies. This would definitely make a huge chaos in police departments since each department has to have many different applications to predict different crimes.

Another issue that would be raised is the usage of other types of prediction for crime prediction models. Although this might work on very specific types of crimes but it is not a universal crime prediction model. Another problem with using other types of prediction for crime prediction is that criminal activities are very different nature with natural occurring events such as earthquake and there are a lot of factors that make crime prediction different from other types of prediction. One main difference is that crime prediction deals with humans and human behavior is very unexpected and each human being is different from others. Other factors that may raise issues are location, age, race and culture of the people. In some places there might be very little crime but if some criminals move there then this may change.

But all together all these models are very promising that in future there might be a universal crime prediction model or algorithm.

### 3.2. Modeling cyber crime

In this section two papers are reviewed. These papers are about cybercrime prediction unlike the next section’s papers which are about physical crime prediction. Cybercrime prediction models deal with those types of crimes that happen in cyber environment and they include internet related crimes or any other computerized crimes.

One of the most common and most devastating attacks that can happen to any internet user is the DoS attack. In [7], the authors try to develop a model to predict DoS attack’s distribution discrete probability.

Firstly, clustering problems are described and then a generic algorithm is utilized to implement the optimization of clustering methods. Then, based on the relation between network traffic data and the
amount of DoS attack data the authors propose a clustering method based on the genetic optimization algorithm to implement the classification of DoS attack data. This model builds the prediction sub-model of DoS attack. Then with the Bayesian method, the calculation of the output probability corresponding to each sub-model is deduced and then the distribution of the amount of DoS attack in some range in the future is obtained.

Another area that Internet users face many problems is worms which are malicious applications that replicate and spread themselves from one computer network to another. In [8] a detection, prediction and reaction system is proposed.

The proposed system allows the system to detect, predict, react using grouping traffic characteristics. According to the proposed system, traffic Factors generated by respective worms using k-means algorithms are grouped into N groups so that a great deal of information may be effectively understood and a worm generated afterward is involved with characteristics of relevant group using cosine similarity for prediction and reaction.

In addition there are some papers which discuss the issues that happen after prediction of digital crime and when an investigation is going on. These issues may arise in terms of privacy [9-18], or when a forensic investigation is happening [19-30], or when a crime happens on a mobile device [31-34]. One of the most used mediums to perform cybercrime is using malware applications which are discussed in [35, 36].

### 3.2.1. Analysis

As it is shown based on the findings of this paper a very small amount of research has been conducted about cybercrime prediction modeling. This might be due to the fact that the Internet is a huge and complex network and predicting one’s actions would be very difficult. But based on the two reviewed papers in this paper, it is quite obvious that researchers have conducted research in the areas in the cybercrime prediction that have a huge impact on the daily life of the Internet users.

### 3.3. Physical crime modeling

In this section, physical crime modeling papers are discussed. Physical crime models are models developed to simplify the process of predicting physical crimes. Physical crimes include every crime beside those which happen in the digital world. These crimes are burglary, murder, bank robbery and the list goes on. There are four papers in this category which are discussed in the following.

One of the models used in physical crime prediction is geographical profiling. As in [37], the authors have developed a model based on geographical profiling. To develop their prediction algorithm the authors have used Rosmo model which has four elements in it

- Anchor point: It is an area where you can pick up the track of the offender. It can be the offender’s home, working place, or some other location of importance to the offender.
- Buffer zone: a border area act as a barrier separating an area designated for special phenomena that the offender would not offend in this area.
- Euclidean distance: is the ordinary distance between two points.
- Manhattan distances: \[d_{1(p,q)} = ||p - q|| = \sum_{i=0}^{n} |p_i - q_i|\]

By using these factors, a model is developed which creates a grid of a city, in which the crime has happened, with calculating Manhattan distance and by calculating the anchor point and buffer point an area will appear on the map of the city which can be the area for the next crime. To test this model, the Yorkshire Ripper is used and the result is that 2 out of 3 results are correct.

Another aspect of physical crime prediction is hotspot prediction. In hotspot prediction rather than predicting where the actual crime will be happening, the areas that will be having more crimes were predicted, thus the name hotspot prediction. One of the researchers on hotspot prediction is done by Keivan Kianmehr and Reda Alhajj. In their research [38] they introduce Support Vector Machine (SVM) approach. They used one-class SVM to predict hotspot crime locations when a predefined level of crime rate and percentage for selecting portion of that are given. They also used k-means clustering algorithm to boost the performance of their model. In their experiment it is proven that the model can work quite fast and give accurate predictions.
One of the important sources of information for law enforcements in this millennium is social media such as Twitter. In these websites users share their views and information for other people. Some of the crime prediction models use different social medias to gather information about someone or a particular topic. In [39], the authors have merged two models of crime prediction and they try to gather information from Twitter.

The first component of this model is the spatio-temporal generalized additive model (STGAM), which predicts the probability of criminal activities at any given location and time using feature-based approach. STGAM is explained more in [40] which represent it in a mathematical equation. This equation is shown in formula (4)

$$\logit \left( p \left( \text{inci}_{s_i,t_j} = 1 \right) \right) = \sum_{n=1}^{N} f_n \left( x_n, s_i, t_j \right) = K_{s_i,t_j-t_0}$$ (4)

In Equation 1, \( p(\text{inci}_{s_i,t_j} = 1) \) is the probability of at least one incident occurring in the spatial grid \( s_i \) at time \( t_j \); \( \logit(p) = \log\left( \frac{p}{1-p} \right) \) is a logit link function; \( N \) is the total number of features; \( x_n, s_i, t_j \) is the nth feature associated with location \( s_i \) and time \( t_j \); \( f_n \) is a smooth function of the nth feature to be estimated from data; and \( \kappa_{s_i,t_j-t_0} \) is a dummy variable indicating the length of the continuous zeros (no incident occurring) that precede the current observation at location \( s_i \) and time \( t_j \).

The second component involved textual analysis which analyzes Twitter posts. Extracting information from text and structuring textual data as numerical vectors are fundamental text mining tasks. A widely applied method to structure text is the vector space model, where documents are represented by term-document matrices. Weighting schemes such as term-frequency/inverse-document-frequency are often used in conjunction with vector space models. This type of model represents a collection of documents within a high-dimensional feature space.

To test the model the developer used incident data from Charlottesville, Virginia. They built two systems first one with the combination of both models and second one based only on STGAM. The results show that first system could accurately predict 60% of the actual crimes whereas second prototype only predicted 50% crimes.

### 3.3.1. Analysis

As it is explained in previous section, the physical crime modeling is located where the crime will happen. By comparing the two models explained above, the result of comparison is that the second model which is a combination of STGAM and text analysis is more effective. The first model just predicts the hotspot and it is only effective for serial criminals. This works because serial criminals have a pattern of committing their crime whereas combination of STGAM and text analysis of Twitter is more generic and could be used for broader crime prediction. Another model is the geographical profiling which may be a bit confusing because it works based on certain points and locations related to the criminal himself. If the criminal is unknown and if the criminal leaves false traces about himself then this program will be wrong. Based on the research and understanding the three models discussed above, the combination of STGAM and text analysis could be more useful to find the criminal himself, whereas the other two models could only provide where the next crime will be happening and these models are only useful in case of serial criminals with distinct patterns.

### 4. Summary

As explained in this paper, there have been some very creative researches on crime prediction. In this paper, crime prediction papers have been categorized into three different categories crime prediction papers, cybercrime modeling papers and physical crime modeling papers. Crime prediction papers are solely about crime prediction and its related aspects. Cybercrime modeling papers are about cybercrime prediction and physical crime modeling papers are about those papers that focus on developing models for finding the location of crimes.

As explained in this paper, all the models and solutions are working properly and the results of each model are satisfactory but the problem with most of these models is that each model only works on a targeted problem for example the model targets only serial criminals or gangs or one specific crime.
such as burglary but there is not any generic model in which law enforcement agencies could get results for most of the crimes with one application. Although crime prediction applications are helping law enforcement agencies but if there is one unified crime prediction application then the workload on officers would be less and more crimes could be prevented. This might be a problem because each crime type has its own attributes but this could be achieved in near future.

Another problem is that many of the algorithms used in the models are developed from other sources of prediction such as earthquake prediction. There is a lack of a tailor-made algorithm for crime prediction only. If this algorithm is developed the performance and the accuracy of the results of crime prediction applications would be much higher.

This paper only reviewed those papers which provide a crime prediction solution and their models or applications are completely related to crime prediction.

5. References


