Classification of Cybercrimes and Cyber-Incidents

Suvigya Hairiya[1], Ansh Sachdeva[2], A. Meena Priyadharsini[3]

[3]Assistant Professor
[1,2,3]Dept. of Computer Science and Engineering,
SRM Institute of Science and Technology, Chennai, India
[1]sh5895@srmist.edu.in, [3]as1972@srmist.edu.in, [3]meenapra@srmist.edu.in

Abstract

Cybercrimes very much are a product of today's modern technology environment. These offences occur across the cyberspace and have a risk associated with them of low to high intensity for our society and many industries, what in turn can pose a national as well as a global threat. Economy also becomes much pregnable if an instance like this occurs. Conventionally, emergency systems have been installed to counter and detect these incidents. But even these may sometimes become infeasible or insufficient for these incidents. But majority of incidents nowadays occur on a small or medium scale only. Hence, this project uses physical parameters of time and location to tag these crimes, which would aid to prevent them in future. Having these parameters in coordination with the algorithms such as Random Forest or Decision Trees aids to come to a conclusive result regarding the time-taken by each algorithm. Then, assess these results and devise a conclusive outcome from this detection.
In conclusion, it helps to improve the time-efficiency of the classification model which in turn will yield a secure conditions for the devices to work in.

Keywords: *Classification, Random Forest, Decision Trees, Data Cleansing, Data Wrangling.*

1. Introduction

Cyber-incidents are a mixture of discrete instances with new illegal acts. Cybercrime incidents occur as separate criminal offences and, according to the national crime statistics and surveys, the instances are increasing. In an FBI report, more than 260,000 complaints were registered of crimes over the internet in 2014, which is 1600% rise compared to the previous report. In accordance with a PwC report, there's been around 50% surge in information wrongdoings in 2014, which comes to be around 117,000 attacks per day.

Also, a hacking incident happened in Kudankulam, Tamil Nadu, India in the month of October, 2019 when the nuclear plant was hacked. Fortunately, this plant had two exclusive infrastructures, one operational and one technical. The hack was made in the organisational infrastructure. Had it been in the technical architecture, it would had been fatal. Similarly, cases of cyber malfunctions came up in nuclear plants like the Hatch power plant, near Baxley, Georgia, US, which without a proper emergency system could have proved to be catastrophic. Therefore, nuclear plants pose a grave danger to the security, confidentiality and secrecy of the nation.

It's very evident that these crimes pose serious danger to the world economy, its safety, and the overall functioning of society. In recent reports, it had been highlighted that these crimes aren't only increasing quantitatively but also becoming progressively destructive and affect a sizeable information range and vectors. Some reports also suggest that crimes are escalating not only in numbers but also in their gravity. There still isn't much information available of what these cyber-incidents can take the shape of. This also makes dealing with such evils a very tedious task. Hence, a way to classify, detect and counter these crimes is required accordingly.

Cyber-crimes have been categorised into two types: Type I, which identifies singular events by keystroke loggers, viruses and rootkits. Type II is not crime ware but repeated user's perspective crimes. Computer integrity offences and crimes assisted by computers are also two other broader types of classifications given by Wall. Also, these crimes have been classified on i) whether a system is used, ii) the target, iii) incidental.

2. Literature Survey

DL Schinder[1] states that these days awareness regarding the cybercrimes and threats have taken the topmost priority. Briefly stating, few risks are as universally comprehensive as cybersecurity where, moreover every industry sector around the world is hit or is under cyberattack.

Ross Anderson [2] provides a method of differentiating cybercrime from other crimes, and also further gives way to estimate cost of these crimes.

i) Cybercrime such as swindling and trickery, through electronic comm. networks and info. systems;
ii) Uploading inappropriate content over electronic multi-media (like hate speech, racism and abuse material;
iii) Crimes on electronic networks, like attacks on information systems, DOS and serious intentional hacking.

George Tsakilidis et al. [11] provides a method to classify cybercrimes into two broad classifications of level 1 and 2, based upon severity. Then these are divided into five types namely, type a, b, c, d and e based on seven parameters related to the crime, the victim, how the victim was targeted, what offences were made among other parameters.

The cybercrimes are basically detected upon the gathered or fetched data which indicates threat or have intentions to damage one's property. Various filtering tools and operations are operated to gather data which then are looked and processed further. Some data which is not serious at that moment is also kept under processing so that its future activity can be controlled. By making the data more comprehensible, the future of such activities can be better understood and handled.

Kai Lung et al. [4] suggests that there should be specific laws for the cyberattacks that are made so that the criminal activities can be deterred and the people who fear the threat and punishment tend not to get involved in these. Whenever such laws are crossed the system should be able to detect and deter the attacks which are to be caused. Convention of cybercrimes (COC) detects and acknowledges every attack or threat even DDOS. Enforcing COC will cause cybercrimes deterrence which will help the individual.

M Gercke [5] suggests that there should be stricter laws for cybercrimes as they now have physical, financial and social impacts. There must be consideration of digital evidences with proper fact-check and also, the jurisdiction of such crimes should be flexible. For this to happen, there must be international cooperation between nations and organisations to check on the offender at any remote location one is in.

According to Emile Sahliyeh [7], these days, websites which are being hosted or served online have become a commonplace for cyberattack which comprises of terrorist threats as well. The terrorists' organisation hosts websites which mostly controls the planning of attacks and other criminal activities.

A lot of analytical work is done to gather the information from these terrorist organisations to track them and put them down, keeping them in range will help in assessing their strategies and confidential data. While accessing their data the person can get an insight on what this organisation future activity will lead to and what type of damage it can produce.

N. von Marees et al. [10] reports that cyberbullying is on an increasing trend in schools, colleges and even workplaces. Newer Info-Comm technologies have led to a better reporting of bullying but still a long way is to covered for an even more effective mechanism for the reporting of cyber-bullying in schools. M Gercke [8] explains online identity theft poses great danger for users. Malicious users can interact with users to gain information, interacting with them, and using it to commit a crime.

E. Blanzieri et al. [6] has provided a study that much of the cases of cyberbullying and cyber-laundering can happen through spams. Spam is unwanted email that was sent by the unknown intentionally with no content, moreover being consent. There is a chance that these spams can mostly be treated as online frauds or money laundering.

These spams can be evaluated on mails which are meant to be or mails which are fake, different kinds of algorithms can be operated to achieve accuracy for detecting and categorizing these mails.

Chi Shiang Cho et al. [9] explains that the use of nuclear power plants is essential, and the new NPP type plants have their architecture on internet. So, it becomes very important to discuss the cyber-physical aspects of the nuclear plants. There are different architectures for organisational structure and technical structures. Because any attack or even a slight cyber malfunction in a nuclear power plant can turn out to be catastrophic. Hence, the infrastructure must be divided in mutually exclusive segments and a physical emergency system should also be in place always.

3. Existing System

The existing system classifies the cybercrimes and cyber-incidents into five types. The crimes of relatively lesser severity are called “level one crimes” whereas of higher severity are “level two”. The system classifies them into five basic categories or types, namely Type-A, Type-B, Type-C, Type-D and Type-E. These are done on the following seven factors which are incident, identified offence, offender, access violation, target, victim and harm. This grouping is done so that the action taken to counter these crimes could be improved with specific steps for each crime.

Despite this classification, the detection is not full-proof. This method is independent of any unique criterion, hence types A, C and D which comprises of the object of legal protection as its focus can sometimes overlap with type B, which focuses on the ways of committing the crime. These overlapping in its result and the lack of a unique algorithm for classification is the main drawback of the existing model.

4. Proposed System

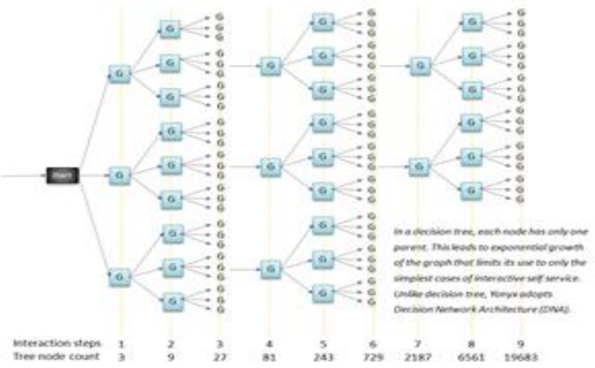
Start
G
In a decision tree, each node has only one parent. This leads to exponential growth of the graph that limits its use to only the simplest cases of interactive self service. Unlike decision tree, Yonyx adopts Decision Network Architecture (DNA).
Interaction steps 1 2 3 4 5 6 7 8 9
Tree node count 3 9 27 81 243 729 2187 6561 19683

FIGURE 1: Diagrammatic representation of a decision tree

The number of cyber offences have increased not only in numbers but also in the intensity. Most ways to counter or detect these offences rely on heavy computational power or expensive means. But most offences occur in small or medium scale, which become very infeasible to work upon. Hence, most such cases go unreported or are not followed-up. Therefore, in the proposed system, the physical details of the cybercrime or cyber incident such as the time or the location of the crime, which will help in classifying the cyber offences of different types.

First, the model fetches the different details of the committed cybercrimes from a crime database with physical parameters of time and location, whether it happened on the device or from a remote site. These details will be compiled dataset, which can be computed using different classification algorithms.

5. Implementation

Pre-processing:

This is done to transform the data into process-able form so as to perform operations by different algorithms on it. This is done to clean the data and make it more comprehensible.

Also, in this stage the different graphs for our dataset such as yearly, monthly and daily arrests among others are formed.

Decision Trees:

It is a decision-support algorithm in which each node acts as a step-test which can eventually lead to the conclusion.

In this model, all the parameters of date and location forms the nodes of the decision tree. These parameters are analysed and classified into a type of crime.

Random Forest:

In this algorithm, multiple decision trees are implemented and the best solution in the training set result is taken as the final solution to the data.
The project takes the above two algorithms as they are the most efficient when it comes to the classification of linearly categorical dataset. Also, one can understand the difference in time taken by the two algorithms in reaching a conclusive result for each of the given cases is very different despite not being explicitly displayed by the model.

6. Result and Inferences

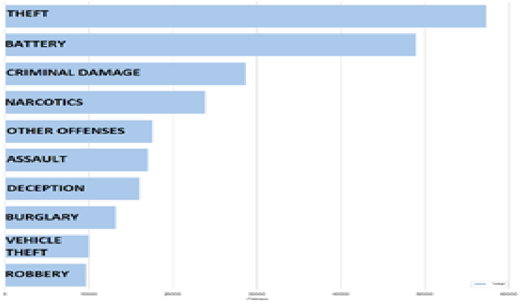

FIGURE 2: Frequency of the various crimes

The system takes the values of latitude, longitude, date and time of the commitment of the crime and outputs the type of crime which occurred on that location at that specific time. Also, it forms the different graphs such as the frequency of different crimes and the arrests made during a specific month, week and day.

This model also represents data through a correlation matrix upon several variables. Along with computing frequency charts for various crimes and displaying the monthly, weekly, daily offences and arrests.

The classification of the crime based on the above given parameters can be achieved using two different algorithms, the first being Decision Trees and the other one being Random Forest. Although, both the algorithms give fairly accurate results for the give dataset, decision trees are much faster. Hence, here the parameter for comparison between algorithms become the computation time within which the output can be generated.

Algorithm	**Time-Taken (s)**
Decision Trees	3.567
Random Forests	8.812

TABLE 1: This Table Gives the Average Time-Taken by the Two Algorithms on Various Inputs

As can be observed, Decision Trees are way more time-efficient than Random Forests, almost twice as efficient.

The model could also have used some other algorithm such as Support Vector Machines or KNN, but they are much more time-taking than either of the above algorithms for the given dataset. SVMs and KNN does not work very well for linearly categorised data. Here the tree-based algorithms can compute the result in seconds, whereas other algorithms could take up to hours or even days for the same result.

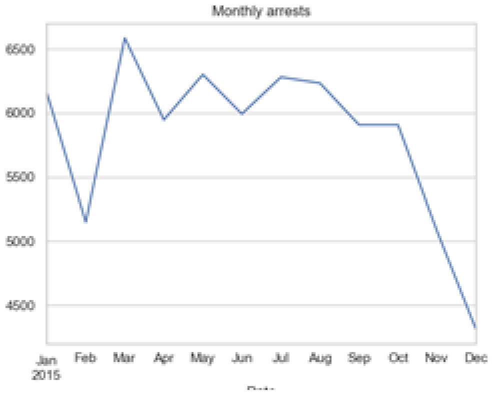

FIGURE 3: Graph showing monthly arrests in 2015

7. Conclusion

The paper gives an exclusive way of classifying various crimes depending on the physical factors such as time and date. It gives a solution to students to carry out an easy project making process using a cybercrime classifier. It uses the grouping of the dataset by either decision trees or random forest to build up a model to prepare over a preparation set in order to get the most exact outcomes. It is a modest and productive approach to group cybercrimes with the goal that the affected can identify the kind of occurrence and follow-up correspondingly. Additionally, it examines the information and creates various charts for the correct portrayal of the information.

References

[1] D. L. Shinder and M. Cross, "*Scene of the Cybercrime*", Burlington, MA, USA: Syngress, 2008.

[2] R. Anderson et al., "*Measuring the cost of cybercrime*", The Economics of Information Security and Privacy. Berlin, Germany: Springer, 2013

[3] D Hall, C. Hargreaves and D. Prince, "*Understanding Cyber Criminals and Measuring Their Future Activity*", Lancaster, U.K.: Lancaster Univ.2013.

[4] Kai-Lung Hi Seung, C. Xi, Hun Son Qiu-Hangkok and W. Qiu-Hong, "*Cybercrime Enforcement: A Comparative Study of US, UK, China, and Other European Countries*", 2014.

[5] M. Gercke, "*Understanding Cybercrime: Phenomena, Challenges and Legal Response*", ITU, Geneva, Switzerland, Sep. 2012

[6] E. Blanzieri and A. Bryl, "*A survey of learning-based techniques of email spam filtering*", Artif. Intell. Rev., vol. 29, no. 1, pp. 63–92, 2008

[7] Emile Sahliyeh, Linda Schamber, David McEntire, Herman L. Totten, Michael Monticino,"*Assessing perceived credibility of websites in a terrorism*", 2009.

[8] M. Gercke, "*Internet-related identity theft*", in Economic Crime Division, Directorate General of Human Rights and Legal Affairs. Strasbourg, France: Council Europe, 2007

[9] Chi-Shiang Cho, Wei-Ho Chung, and Sy-Yen Kuo, Fellow, IEEE, "*Cyber-physical Security and Dependability Analysis of Digital Control Systems in Nuclear Power Plants*", 2016

[10] N. von Marées and F. Petermann, "*Cyberbullying: An increasing challenge for schools*", School Psychol. Int., vol. 33, no. 5, pp. 467–476, 2012.

[11] George Tsakalidis, Graduate Student Member, IEEE, and Kostas Vergidis, "*A Systematic Approach Toward Description and Classification of Cybercrime Incidents*", 2017.