International Journal of Engineering & Technology, 7 (4.32) (2018) 84-86

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Crime Prediction in the Era of Big

Hicham Ait El Bour [1]*, Mohamed Amine Talhaoui[1], Mohamed Azzouazi[1], Reda Moulouki[1]

[1] LTIM Laboratory, Faculty of science Ben Msik, Hassan II University Casablanca, Morocco
**Corresponding author E-mail: bourhicham@yahoo.fr*

Abstract

Big data has become in recent years the fashion for all companies and academic entities. The appearance of this field of research and discussion on a global scale is due to the development of society where each individual alone produces a large size of data of several types including videos, text, audio, logs, database etc. that are structured or unstructured that even this individual alone cannot be found there. If we cite social media, for example, we find unimaginable data.
The fight against crime remains a priority for all societies and deploys enormous human, financial and technological resources to minimize this scourge.
If we can know where, who and when with sufficient time lag, we can reach a goal of "zero crimes". Hence the need for a crime prediction system.

Keywords: *Big data; prediction; crime; machine learning.*

1. Introduction

1.1. Big data

Big data technologies offer technological potential for very intelligent systems, which in other cases help or replace human beings in very advanced analyses.
A technological evolution (hardware and software) was necessary to accompany the evolution of the size of the data, which caused the appearance of Big data [9].
To define the big data the scientific community used the 3 V (Volume, Variable, Velocity). Other speaks of the 5 V (Volume, Variable, Velocity, Veracity, Value): volume for large data size, variable for data diversity, velocity for rapid processing capability, veracity relative to the reliability of the processed data and the ability to derive values from the raw data [1] [10].

1.2 Prediction

Prediction is a way to see the future. It is an objective, on which man has always fought to see his future. Prediction has always been linked to miracles and oracles in the Greeks.
Now, big data technologies have allowed this concept, to take another detour to help the company and society in general to predict results and correct the present in order to have the expected results, and this thanks to these new applications and thanks to this new data extraction and analysis capability.
Prediction is a concept that is based on learning, testing and evaluation. An intelligent system must train on data and then be tested on data from other situations not encountered in the draft shield[11]. Subsequent evaluation can adjust the results.
These learning methods are based on supervised or unsupervised classification algorithms that depend on the nature of the problem.

There are several learning algorithms in the literature. We can note for example regression algorithms, k-means, naïve-bais and neuron networks [2].

1.3 Crime prediction.

Today's city that wants to be a smart city invest a lot of resources to find solutions to simplify the collective life on several areas for that many research are focusing on crime prediction as a system to decrease the crime rate in the city.
Three components are covered to achieve this objective. These are the actions before, during and after the crime. Before the crimes, the officers of the national security plan patrols. After the crimes of the teams of inspectors, forensic police conduct investigations to find the culprit. During the crime, teams intervene quickly to minimize the damage.
In this work, we will begin by explaining our problem. In the next section, we will discuss work in this area. We will then propose a model as a solution to the problem. We will conclude with a conclusion and some perspectives.

2. Challenge

Methods and techniques for prediction are numerous based on mathematical models and possibilities offered by technological development including data mining, big data analytics, machine learning and deep learning.

In this set of solutions, how to establish a predictive model f (equation 1) that will be the basis of a ubiquitous system that allows crimes to be predicted at a given point in the city with maximum precision (less error possible) and based on available data.

$$y = f_T(p, s) \quad (1)$$

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Where T : type of crime, p a point of the city and s a slice of the day.

3. Related Works.

In the literature one speaks of four classes of methods of peak prediction to his see the classical statistical techniques (regression, data mining, times series analysis, seasonality adjustment), simple method (heuristic method), complex application : methods that require a large data size plus sophisticated processing tools, and near repeat and finally adapted methods as heat map [12].
Researchers at the University of Virginia in America have begun work considered statistical techniques on crime prediction based on several types of data, namely historical crime data, twitter data and Foursquare GPS coordinates [3].
In this work, they used an approach based on density estimation by kernel with different parameters topics, GPS Coordinates, Time, Activity,...etc. [4].
They had to cut Chicago's surface at 200 meter points sometimes and sometimes depending on the zip code and linked each to a (local) venue category among a dozen categories already established [5].
To calculate the probabilities of the risk of a type of crime, they chose to use the logistic function on a polynomial sum of parameters (12, 14 or n parameters) derived, in cases, from the topics of the twitter messages, on the other hand, from the categories of the premises and on the other hand, from the activities of the individuals modelled by the routes travelled and periods of activities with always the first parameter calculated by the density estimate [3] [4] [5] [6].

4. Proposed Model

As a solution we present an automatic learning model based on three types of data, namely social media data (SM DATA) for example twitter, history of observed crimes (CH DATA) and GPS data (GPS DATA).
This model, which uses a mathematical equation (equation 2), makes it possible to calculate the probability of a type T crime in a point p during a slice of the day s.

$$Prob\left(Label_x = T / f_1(x), f_2(x), f_3(x), f_4(x)\right) = F(f_1(x), f_2(x), f_3(x), f_4(x)) \quad (2)$$

$x = (p, s)$ / p point of the city and s: slice of the day

$f_1(x)$ Risk rate calculated from messages.

$f_2(x)$ Spatial density of crimes at p.

$f_3(x)$ Temporal density of crimes in the s-day range.

$f_4(x)$ Importance of used routes.

We can schematize this model as in figure fig.1 in the form of layers, which takes as input the three types of data SM DATA, CH DATA and GPS DATA. As this data is raw, a cleaning layer is required to eliminate voids, duplicates, errors and inconsistencies in this data so that it is ready for use by other programs.

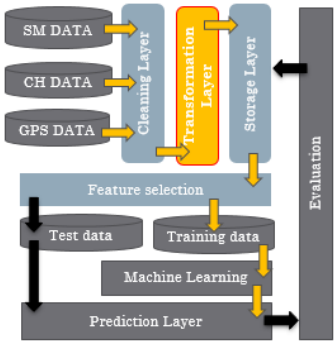

Fig.1: proposed model.

The transformation layer, which is the main contribution of this work, has the mission of consuming the data from the cleaning layer and transforming them into usable measurable values. This layer (Fig.2) contains several modules: topics for discussion, spatial density of crimes, temporal densities of crimes and importance of used routes, which are the main entries in our mathematical model.

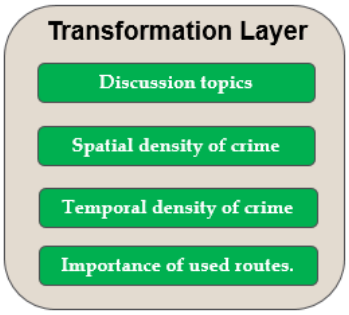

Fig.2: Transformation layer.

The routine activity allows the system to learn the activities around the crimes already committed and then predict the suspicious activity[7]. Therefore, we are led to model the activities of individuals. We can characterize an individual's activity by his time of inactivity, his subjects of discussion and his spatiotemporal movements. For discussion topics, a semantic analysis can be a LSA method (Latency Semantic Analysis) [8] which considered among the best methods in this sound. For time, we can divide the day into significant units (four 6-hour periods for example) and for space, we can divide it into points (200 meters for example) which will be classified according to categories that have meaning for human beings (Table 1).

Table 1: Rental category list.

Number	Location categories
1	Residence
2	Professional
3	Shop and service
4	Travel and transport
5	Food
6	Nightlife spot
7	Outdoors and recreation
8	Arts and entertainment
9	College and University
10	Event

We can from the social media messages have the GPS points source of sending the messages and the time of sending the messages.
Then an activity will then be a route taken with the time density in relation to the time points (slices) of the day.

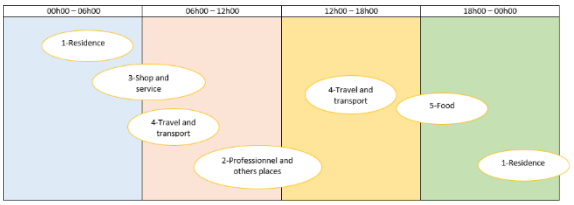

Fig.3: Example of a route.

You can write a route as a vector r=(c1,c2,…,c10) where 10 is the number of rental categories (table 1). Each ci is the number of times the ci category. The route on figure Fig.3 will be r=(2,1,1,2,1,1,0,0,0,0,0,0). Using the TF/IDF algorithm, we can calculate the importance vector i.e. the importance of each category ci in the route r [4].

This modelling of routine activity will allow us to calculate the spatial density and time density of the crimes as well as the degree of traffic around a point p and a slice of the day s.

In this solution, we will take advantage of existing machine learning techniques by exploiting the results of these transformations to allow the system to learn the behavior or activity of individuals and then predict suspicious activities.

Learning process: The execution of the learning program will be based on a 7-day window for example and predict the 8th [4]. Thereafter, an evaluation can be made between this prediction and the actual data. Then the window is moved from the 2nd day to the 8th to predict the 9th and so on until the prediction day (Fig.4).

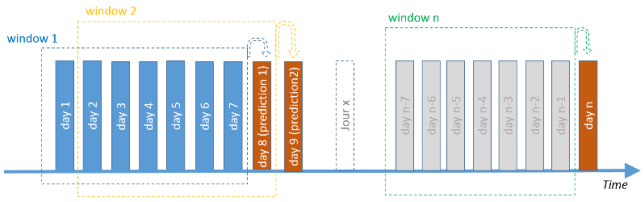

Fig.4: Learning windows

5. Conclusion and Perspectives.

We proposed in this work a 4-dimensional crime prediction model based on the activity of individuals that we could model it in mathematical forms.

In future works we intend to further explain our model and give it an experimental aspect for the validation of its results through data of the crimes of the city of Casablanca.

References.

[1] J. Chan, L. Bennett Moses. « Is Big Data challenging criminology? », Theoretical criminology, vol. 20, no 1, p. 21–39, 2016. doi : 10.1177/1362480615586614

[2] B. Cavadas, P. Branco, S. Pereira. « Crime Prediction Using Regression and Resources Optimization », in Progress in Artificial Intelligence, 2015, p. 513‑524. doi : 10.1007/978-3-319-23485-4_51

[3] M. S. Gerber. « Predicting crime using Twitter and kernel density estimation », Decision Support Systems, vol. 61, p. 115–125, 2014. doi : 10.1016/j.dss.2014.02.003

[4] M. A. Boni, M. S. Gerber. « Predicting crime with routine activity patterns inferred from social media », in 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2016, p. 001233-001238. doi : 10.1109/SMC.2016.7844410

[5] M. Al Boni, M. S. Gerber. « Area-Specific Crime Prediction Models », in Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on, 2016, p. 671–676. doi : 10.1109/ICMLA.2016.0118

[6] M. Al Boni, M. S. Gerber. « Automatic Optimization of Localized Kernel Density Estimation for Hotspot Policing », in Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on, 2016, p. 32–38. doi : 10.1109/ICMLA.2016.0015

[7] A. Newton, M. Felson. « Editorial: crime patterns in time and space: the dynamics of crime opportunities in urban areas », Crime Science, vol. 4, no 1, p. 11, juin 2015. doi : 10.1186/s40163-015-0025-6

[8] S. Bergamaschi, L. Po, S. Sorrentino. « Comparing Topic Models for a Movie Recommendation System », p. 12. doi : 10.5220/0004835601720183

[9] A. G. Ferguson, « Big Data and Predictive Reasonable Suspicion », 2014.

[10] L. B. Moses, J. Chan. « Using big data for legal and law enforcement decisions: Testing the new tools », UNSWLJ, vol. 37, p. 643, 2014.

[11] G. Dreyfus. Réseaux de neurones : méthodologie et applications ; [prévision, data mining, bio-ingénierie, reconnaissance de formes, robotique et commande de processus], 2. Éd. Paris : Eyrolles, 2004.

[12] W. L. Perry, B. McInnis, C. C. Price, S. C. Smith, J. S. Hollywood. Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations. Rand Corporation, 2013.