A SURVEY OF CRIME PREDICTION USING MACHINE LEARNING AND RECURRENT NEURAL NETWORKS

[1]Souglman Kabirou Kanlanfeyi, [2]Keshav Kishore
[1]Post Graduate Student, [2]Assistant Professor
[1]Departement of Computer Science and Engineering,
[1]AP Goyal Shimla University, Shimla (HP), India

Abstract: Since the evolution of AI techniques and their applications on in many domains such as banking, industries, administrations, agricultures... it has also been applied to on crime prediction or forecasting. Different studies showed that we can predict crime by using AI techniques such as machine learning classification algorithms to predict a type of crime by using time and location information. Since the deployment of different libraries designed for Deep Learning techniques like Tensorflow, Theanos and Keras, applications of DL on many fields has become very popular and so for crime prediction which leads to crime prediction by using DL which performs better than ML traditional algorithms. Because of the possible relationship between two or many occurrence of crime or in other words, considering that a crime occurrence can be related to previous ones, Recurrent Neural Networks (RNN) which is a DL architecture designed for time series problems has shown to perform better than not only ML but also than DNN algorithms. This paper reports different studies done in the domain of crime prediction using different AI techniques by considering different factors that can affect crime prediction and the limitation of some algorithms for predicting and their solutions such as RNN with LSTM.

Keywords: Crime Prediction, Machine Learning, Deep Learning, RNN, LSTM

I. Introduction

As it is often said: Prevention is better cure, this is the purpose of Deep Learning in the field of crime which is used to predict crime in order to prevent crime by finding hot spots and potential time of crimes occurrence. But what is Deep Learning? According to Wikipedia, Deep Learning is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms. Learning can be supervised, semi-supervised or unsupervised. As it is often said: Prevention is better cure, this is the purpose of Deep Learning in the field of crime which is used to predict crime in order to prevent crime by finding hot spots and potential time of crimes occurrence. But what is Deep Learning? According to Wikipedia, Deep Learning is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms. Learning can be supervised, semi-supervised or unsupervised.

1. Machine Learning

Machine Learning defined as the ability of a machine to learn for experience (data) without being explicitly programmed. It's a subset of AI which contains 3 main types of leanings which Supervised, Unsupervised and Reinforcement learnings designed for some specific problems. We will focus on Supervised Learning which consist on learning from data which is already labelled. In this type of learning, we have two techniques which are Regression which consists to predict a continuous value or in other words a numeric value and Classification which consist on classifying an input value into a given class.

Here is an illustration of Machine Learning structure:

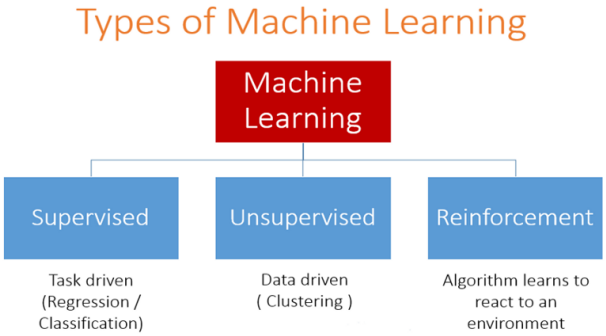

Fig 1.1 Types of Machine Learning

2. Deep Learning

DL is a subset of AI based on functional of human neural networks called Artificial Neural Networks. DL was adopted due to the fact that some ML techniques became to perform well on some complex problems and also due to the increase of data amount nowadays. As said previously, most DL architectures are based on ANN they are also designed for some specific task like CNN for Image based problems, RNN for time series… How does ANN work?

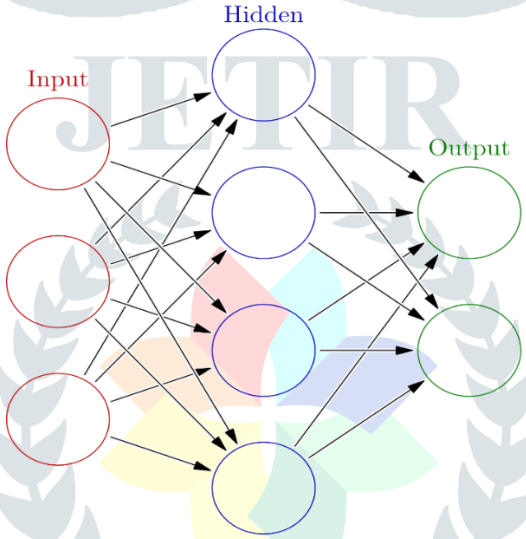

Fig 1.2 ANN architecture

We have three types of layers as shown on the figure Fig1.2 which are:

Input layer: receives input data. An example of input can be the number of rooms, facilities, and distance to the downtown... in the case of price prediction of a house. The input layer passes the inputs to the first hidden layer.
Hidden layers: perform mathematical computations on our inputs. One of the challenges in creating neural networks is deciding the number of hidden layers, as well as the number of neurons for each layer.
Output layer: returns the output data. In our case, it gives us the prediction of the price of the house.
The process of prediction the price of the house is done as followed:

- Each connection between neurons is associated with a weight. This weight dictates the importance of the input value.
- The initial weights are set randomly.
- When predicting the price of a house, the number of rooms should be one of the heavier factors.
- Hence, the number of rooms' neuron connections will have a big weight.

3. Recurrent Neural Networks (RNN)

Basically, RNN is defined as a class of Deep Learning where output from previous step is used as input of the next step. Commonly used in Natural Language Processing and Speech Recognition, RNN follows a sequential pattern. In opposite to traditional Neural Network, RNN has a memory and input in the network are dependent to each other. For example, memory in RNN is used to save the previous word in order to predict the next one.

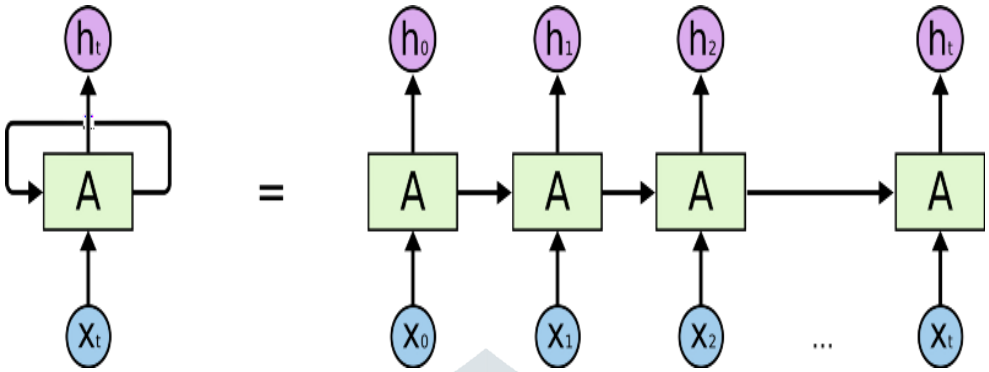

Fig 1.3 RNN Architecture

First, it takes the x(0) from the sequence of input and then it outputs h(0) which together with x(1) is the input for the next step.

4. LSTM (Long Short Term Memory).

Since RNN work by using previous step's information, they are also supposed to retained information for a long which is not evident in the case of classic RNN. This phenomena is called vanishing or exploding of gradient. To solve this problem another RNN architecture based on time series is used called LSTM (Long Short Term Memory). In order to retain information for long time, we have four types of NN with different memory blocks called cells. To insure manipulation of memories and retention of information we have three types of gates. Forget gate used to remove the information which is already used in the first cell state or which is no longer useful. Input gate used to add the information to the cell gate. Output gate used for extracting useful information from current cell state in order to present it as output.

II. Related Work

Ying-Lung Lin et al. [1]

In this paper, authors used machine learning algorithms such as Random Forest, Naïve Bayes and especially Deep Learning with Neural Networks to predict and assist crime prevention in Taiwan. To improve performance of model, the accumulate data to time scale and validate experimental results by visualizing potential crime hotspots on a map. For prediction, R language was used with h2O package and then the results were visualized on a map by QGIS. After prediction, the results showed that Deep Learning algorithm perform better than others.

Nishat Shama et al. [2]

Written by Nishat Shama, student of School of Engineering & Computer Science (Department of Computer Science & Engineering) in BRAC University, the purpose of this paper is to use machine learning classification algorithms to predict crime by considering time and location factors. The algorithms used are Decision Tree, Gaussian Naive Bayes, k-NN, Logistic Regression, Adaboost, Random Forest... Dataset used is provided by SF Opendata from SFPD Crime Incident Reporting System which contains information on crime incidents that occurred in San Francisco for the period of 1/1/2003 to 5/13/2015. Since some columns in the dataset were imbalanced, SMOTE, oversampling and undersampling methods such as Edited NN, Neighborhood Cleaning Rule were applied on the dataset in order to get an accuracy up to 81%.

Ashwini et al. [3]

In this paper written by Ashwini B. Bais, Prof. G. Rajesh Babu and Prof. Jayant Adhikari from Faculty of Computer Science and Engineering at TGPCET, Mohgao Nagpur, India, authors studied reasons, factors and relations between occurrence of different crimes and finding the most appropriate ways to control and avoid more crimes. In order to efficiently and accurately analyze the growing volumes of crime related data and due to the vast diversity and the complexity of crime patterns which made the analyzing and recording of crime data more difficult, a solution is present in this paper. The following objectives of this paper is:

- To implement web crawling to crawl news article
- To implement new approach for data preprocessing using NLP
- To implement document classification using SVM based classifier.

Yong Zhuang et al. [4]

Published in 2017 at the IEEE International Conference on Big Knowledge, the authors of this paper (Yong Zhuang, Matthew Almeida, Melissa Morabito and Wei Ding) studied how we can profit of large amount of data available nowadays to forecast hot spot of crimes by using spatial and temporal information available in datasets. The approaches are first to detect temporal patterns in time series by using RNN and then to build a Spatial-Temporal Network to forecast crime hot spots with embedding spatial information.

At the end of the study, we remark that a RNN architecture (LSTM) performs better than the other RNN architectures (RNN, GRU) and then other classic machine learning algorithms such as (Decision Tree, Gaussian Naïve Bayes, Random Forest, KNN, Logistic regression and Multi-layer perception). The performance metrics used are: Accuracy, F1 Score and Precision and Recall). Dataset used is the data of Call-For-Service provided by the Portland, Oregon Police Bureau (PPB) for a 5-year period from March 2012 through the end of December 2016.

Amit Gupta et al. [5]

Published by Amit Gupta, Ali Syed, Azeem Mohammad, Malka N.Halgamuge in the (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 7, 2016, this article describes our we can use data mining classification algorithms like BayesNet, NaiveBayes, J48, JRip, OneR and Decision Table to predict crime in Denver city in USA and to highlight trends of incidents that will in return help security agencies and police department to discover precautionary measures from prediction rates. The output of this work showed that the JRip classifier is more accurate with the prediction by using the confusion matrix to find the precision of each algorithm with an accuracy of 73.71%.

Trung T. Nguyen et al. [6]

In this paper, authors focused on prediction of crime by considering others factors than time and location ones which are: the demographic information of the area, educational background, economical and ethnic background of the people involved. This information was collected from different public sources provided. First, data collected was cleaned to deal with missing values then the data was reduced by doing a feature selection to eliminate features that are not important for prediction before applying data transformation, data discretization and then integrate new data to the data. After preprocessing the data, we obtained 4 classes where only two will be used for prediction. Predictions were done using machine learning classification algorithms such as Support Vector Machine (SVM), Random Forest, Gradient Boosting Machines, and Neural Networks which were finally compared to see which one performed well.

Panagiotis Stalidis et al. [7]

Panagiotis Stalidis, Theodoros Semertzidis and Petros Daras in 2018, the authors proposed an approach of examination different DL architectures for crime classification and prediction by presenting 3 fundamental DL architecture configurations for crime prediction based on encoding: a) the spatial and then the temporal patterns, b) the temporal and then the spatial patterns, c) temporal and spatial patterns in parallel which leads to a selection of the best configuration for a further investigation. Since the study is done on 5 different datasets, the models were compared on them by using 10 state-of-the-art algorithms. Finally, a proposition of a guide for designing DL models for crime hotspot prediction and classification was done.

Hyeon-Woo Kang et al. [8]

Since some previous studies for predicting crime have used data from multiple domains such as demographics, economics, and education. The authors of this paper noticed that the prediction models treat data from different domains equally which be a problem in crime occurrence prediction, such as difficulty in discovering highly nonlinear relationships, redundancies, and dependencies between multiple datasets so they also consider the environmental context.

The output of this research was obtained by the following three phases:

- collecting data from various types of data.
- analyzing the relationship between crime incidents and the collected data using a statistical approach in order to have a result that can be used more effectively in crime prediction.
- to accurately predict crime occurrences, authors employed a DNN using feature-level fusion with different weights to efficiently proportion the data in order to integrate spatial, temporal, and environmental context features.

The result of the prediction gives an accuracy of 84.25% which is such good to help police patrols to leverage predictive crime information to more effectively monitor crime hotspot areas and improve the overall effectiveness of police patrols. The authors of this paper are Hyeon-Woo Kang and Hang-Bong Kang.

Ying-Lung Lin et al. [9]

The main purposed of this paper is to predict crime using geographic features of grid. This study incorporates the concept of a criminal environment in grid-based crime prediction modeling, and establishes a range of spatial-temporal features based on 84 types of geographic information by applying the Google Places API to theft data for Taoyuan City, Taiwan. Among all algorithms used, Deep Neural Networks performed well with an accuracy of a highest accuracy of 81.45% on average accuracy. We also remark an improvement of F1 score about 7% compared to our design's baseline 11-month moving average. This paper was written by *Ying-Lung Lin, Meng-Feng Yen and Liang-Chih Yu.*

Luiz G.A. Alves et al. [10]

Focusing on the correlation between crime and urban metrics, the authors of this paper (Luiz G.A. Alves, Haroldo V. Ribeiro, Francisco A. Rodrigues proposed a solution by not only considering common factor for crime prediction like time and location. Since there is non-Gaussian distributions and multicollinearity in urban indicators, it is common to find controversial conclusions about the influence of some urban indicators on crime then the proposed to use ML ensemble learning technique (Random Forest) to handle this problem by choosing the number of homicides at a city level in Brazil as crime indicator. At the end of the study, the reach an accuracy of 97% and then quantify the importance of urban metrics in crime prediction. The results shows that unemployment and illiteracy are the most factors that impact homicides in Brazilian cities.

Tianxiang Gao et al. [11]

In this paper published in 2016, Tianxiang Gao explained the optimization of Hyperparameters into three parts:

- Definition of Hyperparameters
- Exploration the space of Hyperparameters
- Evaluation of Hyperparameters.

Nikhil Ketkar et al. [12]

This paper published by Nikhil Ketkar in 2017 is an introduction to Keras library which is built on top of Tensorflow and Theanos which are also libraries designed for AI domains such as ML and DL. In this paper, the author explained the different aspect of Keras library and the different types of layers and networks that can be built with it with many examples. It also shows us how to build different DL architectures such as LSTM and CNN.

Shen Ting Ang Weichen Wang et al. [13]

In this report published in 2015, the authors (Shen Ting Ang Weichen Wang and Silvia Chyou) focused on the classification of an occurrence of crime based on time and location by using data mining techniques such as Naïve Bayes, Logistic Regression and Random Forest. The authors also provided some visualization with multivariate analysis. By considering all features such as District, Day, Hour, Month and Grid, Random Forest classifier were able to reach an accuracy of 0.24863 and a log-loss of 2.49745 performing better than the other classifiers.

Anish Krishnan et al. [14]

In this paper written by Anish Krishnan, Aditya Sarguru and Shantha Sheela in 2018, the authors focused on a prediction of crime by using Deep Learning Architectures such as RNN especially by focusing on the LSTM which is used to sole problem of vanishing and exploding of gradient in RNN. Keras framework built in top of Tensorflow and Theanos were used to build an efficient classifier for crime prediction. Result obtained is then compared with result obtained with an already existing model and it shows that the built model performs well.

From this survey, we can noticed the common usage of these technologies:

Most languages used	Python, R
Most Machine Learning Algorithms used	Support Vector Machine, Random Forest, Naïve Bayes, Logistic Regression, Decision Tree
Most Accurate Machine Learning Algorithms	Support Vector Machine, Random Forest, Naïve Bayes, Decision Tree…
Most Deep Learning Architectures used	Classic Artificial Neural Networks, Recurrent Neural Networks (LSTM, GRU)
Most frameworks used	TensorFlow, Keras

Tab.2.1 Summary of different technologies and algorithms used

III. Research Gaps

Based on this survey, we have identified these gaps on the papers which can be consider in the future research topics:

- Most of research focused on using ML techniques only in some complex problems where DL techniques are known to perform better
- Some topics used in the implementation parts designed software like Weka which are designed for both ML and Data Mining but can remark that the user doesn't have hands on different parameters.
- There is a lack of network optimization by using different callbacks and by tuning the Hyperparameters in order to find the optimal combination of Hyperparameters in some topics which implemented DL architectures.
- Few topics on crime prediction nowadays use RNN but it has been shown that this architecture performs very in these kinds of problems.

IV. Proposed solutions

In order to face these issues, the propose solution will be to consider different factors that can impact crime occurrence except time and location which are the common one used. These factor can be related to social factor of the victim or the criminal such as illiteracy, unemployment, sex…Except the different factors, we should also focus on using DL architectures designed for sequence or time series problems which are RNN and to improve networks quality, it's preferable to used different callbacks such as EarlyStop, LearningRateScheduler, ModelCheckPoint… Hyperparameters Tuning should also be a great advantage in the training process by using different techniques like Grid Search CV or Random Search CV. To resume, the proposed solutions will be:

- Consider many factors that can affect crime occurrence except time and location
- Consider using DL architecture especially RNN techniques such as LSTM which are designed to retain information for a long time
- Perform Hyperparameters tuning
- Use callbacks for our different networks during the training process
-

V. Conclusion

In this paper, we did a survey of the application of AI in crime prediction nowadays by using different techniques such as ML and DL. We also identify the architectures well known to perform well in these kind of problems and the common technologies and algorithms used for implementing these solutions.

RFERENCES

[1] Ying-Lung Lin, Liang-Chih Yu, Tenge-Yang Chen: Using Machine Learning to Assist Crime Prevention. 6th IIAI International Congress on Advanced Applied Informatics, 2017.
[3] Ashwini B. Bais, Rajesh Babu, Jayant Adhikari: An Intelligent Analysis of Crime through Newspaper Articles – Clustering and Classification. International Journal on Recent and Innovation Trends in Computing and Communication Volume: 5 Issue: 2, 2017.
[2] Nishat Shama and Dr. Mahbub Alam Majumdar: A Machine Learning Approach to Predict Crime Using Time and Location Data, 2015.
[4] Yong Zhuang, Matthew Almeida, Melissa Morabito, and Wei Ding: Crime Hot Spot Forecasting: A Recurrent Model with Spatial and Temporal Information. IEEE International Conference on Big Knowledge, 2017.
[5] Amit Gupta, Ali Syed, Azeem Mohammad, Malka N.Halgamuge: A Comparative Study of Classification Algorithms using Data Mining: Crime and Accidents in Denver City the USA. (IJACSA) International Journal of Advanced Computer Science and Applications, 2016.
[6] Trung T. Nguyen, Amartya Hatua, and Andrew H. Sung: Building a Learning Machine Classifier with Inadequate Data for Crime Prediction. Journal of Advances in Information Technology Vol. 8, No. 2, May 2017.
[7] Panagiotis Stalidis, Theodoros Semertzidis and Petros Daras: Examining Deep Learning Architectures for Crime Classification and Prediction. IEEE, 2018.
[8] Hyeon-Woo Kang, Hang-Bong Kang: Prediction of crime occurrence from multimodal data using deep learning. PLoS ONE 12(4):https://doi.org/10.1371/journal.pone.0176244, 2017.
[9]Ying-Lung Lin, Meng-Feng Yen and Liang-Chih Yu: Grid-Based Crime Prediction Using Geographical Features. International Journal of Geo-Information, 2018.
[10] Luiz G.A. Alves, Haroldo V. Ribeiro, Francisco and A. Rodrigues: Crime prediction through urban metrics and statistical learning. ELSEVIER, 2018.
[11]Tianxiang Gao: Hyper-parameter Optimization for Deep Learning, 2016.
[12] Nikhil Ketkar: Deep Learning with Python, DOI 10.1007/978-1-4842-2766-4_7, 2017.
[13] Shen Ting Ang, Weichen Wang, Silvia Chyou: San Francisco Crime Classification. Fall CSE 255 Assignment 2 Report, 2015.
[14] Anish Krishnan1, Aditya Sarguru, A. C. Shantha Sheela: Predicitve Analysis of Crime Data Using Deep Learning. International Journal of Pure and Applied Mathematics Volume 118 No. 20 2018.
Introduction to Recurrent Neural Network
https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/
Ensemble Methods in Machine Learning: What are They and Why Use Them?
https://towardsdatascience.com/ensemble-methods-in-machine-learning-what-are-they-and-why-use-them-68ec3f9fef5f
Understanding LSTM Networks
http://colah.github.io/posts/2015-08-Understanding-LSTMs/
Overview of Recurrent Neural Networks And Their Applications
https://www.analyticsindiamag.com/overview-of-recurrent-neural-networks-and-their-applications/
Deep Learning | Introduction to Long Short Term Memory
https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/
Definition of crime:
https://www.en.oxforddictionaries.com/definition/crime
Definition of Machine Learning:
https://medium.com/@ageitgey/machine-learning-is-fun-80ea3ec3c471