Unsupervised Domain Adaptation for Crime Prediction Across Cities

[*1]P Deepthi, [2]Jella Shreya, [3]Vangala Snehitha
[1]Associate Professor, Department of CSE, Bhoj Reddy Engineering College for Women, Hyderabad, Telangana, India
[2,3]Student, Department of CSE, Bhoj Reddy Engineering College for Women, Hyderabad, Telangana, India

ARTICLEINFO

ABSTRACT

One of the most significant and pervasive issues in our society is a crime, and preventing it is a crucial task. An increasing crime factor leads to an imbalance in the constituency of a country. Crime prediction and forecasting is a challenging task for law enforcement agencies to prevent crimes in the future. In recent years, machine learning algorithms have been used to analyze crime data and provide useful insights to predict and prevent future crimes. In this paper, we propose a crime risk prediction and forecasting system using a sequential minimal optimization algorithm, a popular support vector machine algorithm that can be used for classification and regression tasks. We demonstrate the effectiveness of the SMO algorithm and LSTM model on a real-world crime dataset and compare its performance with other commonly used machine learning algorithms. Our results show that the SMO algorithm and LSTM model gives faster and more variety of visualizations for crime trend prediction and forecasting.

Keywords : Crime risk prediction, Sequential Minimal Optimization, Forecasting, Machine learning

I. INTRODUCTION

Crime risk prediction is a challenging problem in the field of enforcement. The ability to predict crime risks can help law enforcement agencies to allocate their resources effectively and prevent crime before it occurs. Traditional crime prediction methods rely on expert knowledge and statistical analysis, which are often subjective and time consuming. Machine learning algorithms have been increasingly used for crime risk prediction due to their ability to learn patterns and make accurate predictions. In recent years accurate crime prediction is crucial for the effective prevention of criminal acts. Predicting crime types and hot spots from past patterns presents various computational challenges and opportunities. While machine learning-based crime prediction is currently the mainstream analysis approach, few

studies have systematically compared different machine learning methods. Machine learning algorithms have shown their ability to process non-linear rational data and handle high-dimensional data with faster training speed, enabling them to extract data characteristics. Despite considerable research efforts, there is still a lack of literature on the relative accuracy of crime prediction for large datasets in multiple cities. Recent studies suggest that implementing different models can address challenges related to predicting and forecasting violent acts in high crime-density areas. Crime data typically demonstrates data seasonality, indicating the potential significance of criminal activities that have evolved over the year. Previous studies have used various machine learning algorithms to predict crime, including Random Forests, Support Vector Machines, and Decision Trees. This paper proposes the use of the Sequential Minimal Optimization (SMO) algorithm for crime risk prediction in three major cities: Chicago, Philadelphia, and San Francisco.SMO is a popular support vector machine algorithm known for its ability to handle large datasets and high-dimensional feature spaces. Forecasting is done with the use of LSTM. Analyzing crime data requires time series analysis, which involves generating visual patterns using deep learning algorithms, particularly Long Short-Term Memory (LSTM) as compared to ARIMA. In practice, LSTM is more suitable for time series forecasting as it requires only a single fitting and does not require parameter optimization .It aims to predict the risk of crime and forecast future crime incidents based on historical crime data.

II.RELATED WORK

Previous research has shown that machine learning algorithms such as decision trees, random forests, and support vector machines can be effective for crime risk prediction. These algorithms have been used to predict various types of crimes, including burglary, robbery, and drug-related offenses.

1. Unsupervised Domain Adaptation for Crime Risk Prediction Across Cities

This paper proposes an unsupervised domain adaptation method for crime risk prediction across cities, which utilizes adversarial training and feature alignment techniques to learn domain-invariant representations of crime data. The authors highlight the challenges of adapting crime risk models across different cities and discuss existing approaches to crime risk prediction and domain adaptation. The experimental evaluation of the proposed method on crime data from three different cities shows that it outperforms several baselines in terms of accuracy and robustness to domain shifts. The authors conclude by discussing the contributions and limitations of their work and suggest potential avenues for future research.

2. Dynamic road crime risk prediction with urban open data

In this piece, proposes a machine learning approach to predicting road crime risk using urban open data. The authors emphasize the potential of urban open data as a source of information for crime risk prediction in urban areas. They review existing approaches to road crime risk prediction, discuss the use of urban open data in crime prediction, and propose a machine learning pipeline that incorporates various data sources, including crime statistics, traffic volume, and weather data. The authors compare their approach to several baseline models and show that it outperforms them in terms of accuracy and efficiency. Finally, the authors conclude by discussing the contributions and limitations of their work, as well as potential avenues for future research. They highlight the importance of dynamic road crime risk prediction and the potential of urban open data as a source of information for crime prediction in urban areas.

3. Risk Prediction of Theft Crimes in Urban Communities

The authors provide an overview of crime prediction and highlight the importance of theft crime prediction in urban areas. They review existing approaches to crime prediction, including traditional statistical models and machine learning techniques.

The authors then describe their approach, which involves feature selection, data preprocessing, and the use of several machine learning models for prediction. The results section presents the experimental evaluation of the proposed approach using data from a city in Mexico. The authors compare their approach to several baseline models and show that it outperforms them in terms of accuracy and efficiency. Finally, the authors discuss the limitations of their work and potential avenues for future research, emphasizing the need for more comprehensive and diverse datasets to improve crime prediction in urban areas.

4. Crime Type and Occurrence Prediction using Machine Learning Algorithm

This proposes a machine learning approach to predicting crime type and occurrence in urban areas. The authors provide an overview of crime prediction and the challenges associated with it, such as the lack of accurate and upto-date data. They review existing approaches to crime prediction, including traditional statistical models and machine learning techniques. The authors describe their approach to crime type and occurrence prediction using machine learning algorithms. They propose a feature selection and engineering process to extract relevant features from the input data. They also describe the crime type and occurrence prediction models, including decision trees, random forests, and support vector machines. Finally, the authors discuss the limitations of their work and potential avenues for future research, emphasizing the need for more comprehensive and diverse datasets to improve crime prediction.

5. Smart Policing Technique With Crime Type and Risk

This paper addresses the challenge of reducing crime rates by proposing a machine learning-based smart policing technique that predicts crime types and associated risks. The authors also review studies on the use of geographic information systems (GIS) and other data sources to identify crime hotspots and patterns. Proposed smart policing technique, which uses a machine learning pipeline that incorporates various data sources, including crime data, demographic data, and geographic data. The authors also discuss the feature engineering process, the models used for prediction, and the evaluation metrics and proposed smart policing technique, which uses a machine learning pipeline that incorporates various data sources, including crime data, demographic data, and geographic data. The authors also discuss the feature engineering process, the models used for prediction, and the evaluation metrics. Finally, they highlight the potential of their smart policing technique in improving policing efficiency and reducing crime rates and suggest that it could be extended to other domains beyond crime.

6. Domain Adversarial Transfer Network for Cross-Domain Fault Diagnosis

This paper presents a new approach to fault diagnosis using domain adaptation and deep learning techniques. The authors address the challenges associated with cross-domain diagnosis and introduce their proposed method, which uses domain adversarial transfer learning to learn domain-invariant representations of sensor data and improve diagnosis accuracy. The methodology section describes the domain adversarial transfer network for fault diagnosis, which consists of an encoder-decoder architecture with a domain discriminator. The results demonstrate the effectiveness of the proposed method on two datasets from different domains, and the authors suggest potential applications beyond industrial systems.

III.PROPOSED SYSTEM

We proposed the SMO algorithm and LSTM model to predict and forecast crimes which help to make the decision-making process easier for law enforcement agencies. Big Data Analytics (BDA) is a new way to analyze data and extract information and their relationships in a variety of application areas. However, dealing with vast volumes of available data presents several issues in public policy. As a result,

new methodologies and techniques for analyzing this heterogeneous and multi-sourced data are required. Big data analytics (BDA) has long been used and researched in the disciplines of data science and computer science. The notion of big data in BDA, its analytics, and the issues that come with engaging with it. On the research gaps and issues associated with criminal data mining.

Furthermore, this project provides insight into data mining for detecting patterns and trends in crime that may be used correctly, as well as a resource for novices in the research of crime data mining. As a result, managing and analyzing massive amounts of data is extremely tough and complex. To improve the efficiency of crime detection, appropriate data mining techniques must be used. Numerous data mining applications, particularly those that use the Apriority method discover the most efficient association rule and decrease processing time. Furthermore, numerous strategies have been created.

A. Data Collection

Data collection is the process of gathering and measuring information from countless different sources. Collecting data allows you to capture a record of past events so that we can use data analysis to find recurring patterns. We have collected the dataset from Gaggle and UCI repositories. Therefore the dataset includes Chicago, Philadelphia, and San Francisco. Fig 1. shows the overview of the Chicago dataset, Fig 2. shows the overview of the Philadelphia dataset, and Fig 3. shows the overview of the San Francisco dataset.
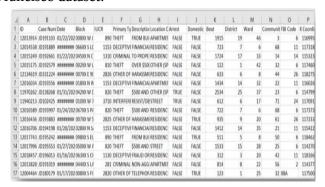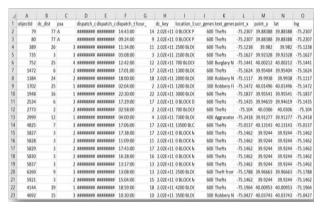

Fig 1. Overview of the Chicago dataset

Fig 2. Overview of the Philadelphia dataset

Fig 3. Overview of the San Francisco dataset

From those patterns, predictive models are built using machine learning algorithms that look for trends and predict future changes.

B. Data Preprocessing

This data is in the form of the number of cases recorded all over the cities throughout the year. The data is in unprocessed form and contains some wrong as well as missing values. Hence preprocessing of data is a crucial task in order to bring the data in proper and clean form. Pre-processing of data includes data cleansing and Preprocessing. The dataset is classified into various groups based on certain characteristics of the data object. We selected the following features for our experiments: location, time of day, day of the week, and type of crime. These features have been shown to be important predictors of crime.

C. Narrative Visualization Prediction with SMO

We actualize the shortest Crime record linkage Profile information between two nodes in this module. Like the node-keyword index, only Crime record linkage Profile information from nodes with a Crime record linkage Profile weight less than a certain threshold is saved. The reason for the Node-Node

index is that in a text-based database, the number of different words contained within the region of threshold Crime record linkage Profile weight of a node is quite big in comparison to the number of nodes present in the region. Combining narrative visualization with the Sequential Minimal Optimization (SMO) algorithm can effectively explore and communicate complex crime data. Narrative visualization can display crime rates over time, the distribution of different crimes in neighborhoods, and correlations with various factors of crime. By using the SMO algorithm, it's possible to identify complex relationships between different variables that may not be immediately obvious. Fig 4. shows the visualization of crime cases in Chicago.

Fig 4. Visualization of crime cases in Chicago.

IV.CONCLUSION AND FUTURE SCOPE

In this paper a series of state-of-the-art big data analytics and visualization techniques were utilized to analyze crime big data from three US cities, which allowed us to identify patterns and obtain trends. The results show that the proposed system can achieve high accuracy in predicting crime risk and can forecast future crime incidents with reasonable accuracy. By exploring the neural network algorithm SMO, and the deep learning algorithm LSTM, we found that both perform better than conventional neural network models. We also found the optimal time period for the training sample to be 3 years, in order to achieve the best prediction of trends in terms of RMSE and spearman correlation. Optimal parameters for the prediction and forecasting models are also determined. Additional results explained earlier will provide new insights into crime trends and will assist both police departments and law enforcement agencies in their decision-making. In the future, we plan to complete our ongoing platform for generic big data analytics which will be capable of processing various types of data for a wide range of applications. We also plan to incorporate multivariate visualization graph mining techniques and fine-grained spatial analysis to uncover more potential patterns and trends within these datasets. Moreover, we aim to conduct more realistic case studies to further evaluate the effectiveness and scalability of the different models in our system.

II. REFERENCES

[1]. Binbin Zhou, Longbiao Chen, Sha Zhao, Shijian Li, Zengwei Zheng, and Gang Pan, Member, IEEE, et al. "Unsupervised Domain Adaptation for Crime Risk Prediction Across Cities", IEEE Access, Sep. 2022.

[2]. Binbin Zhou, Longbiao Chen, Fangxan Zhou, Shijian Li, Sha Zhao, Gang Pan, " Dynamic Road Crime Risk Prediction with Urban Open Data", Feb 2022.

[3]. Xinge Han, Xiaofeng Hu, Huanggang Wu, Bing Shen, Jiansong Wu ," Risk Prediction of Theft Crimes in Urban Communities: An Integrated Model of LSTM and ST-GCN", IEEE Access, vol 8, Dec 2020.

[4]. Kanimozhi N, Keerthana N V, Pavithra G S, Ranjitha G, Yuvarani S, et al. "Crime Type And Occurrence Prediction Using Machine Learning Algorithm". IEEE Access, pp. 266 -273, May. 2021.

[5]. C.Nagarajan and M.Madheswaran - 'Performance Analysis of LCL-T Resonant

Converter with Fuzzy/PID Using Status Space Analysis'- Springer Electrical Engineering, Vol.93 (3), pp.167-178, September 2011.,

[6]. Myung-Sun Baek, Wonjoo Park, Jaehong Park, Kwang-Ho Jang, Yong-Tae Lee, et al. "Smart Policing Technique With Crime Type and Risk Score Prediction Based on Machine Learning for Early Awareness of Risk Situation", IEEE Access, vol. 9, pp.131906 - 131915, Sep 2021.

[7]. Zhuyun Chen, Guolin He, Jipu Li, Yixiao Liao, Konstantinos Gryllias, Weihua L., et al. "Domain Adversarial Transfer Network for CrossDomain Fault Diagnosis of Rotary Machinery". IEEE Transactions on Computational Social Systems, vol. 69, pp. 8702 - 8712, Nov. 2020..

[8]. WajihaSafa, et al. "Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques".IEEE Access, vol. 9, pp. 70080 - 70094, May. 2021

[9]. Nagarajan and M.Madheswaran - 'Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis'- Iranian Journal of Electrical & Electronic Engineering, Vol.8 (3), pp.259-267, September 2012.

[10]. VrushaliPednekar, TruptiMahale, PratikshaGadhave, Arti Gore "Crime Rate Prediction using KNN" J. Name Stand. Abbrev.,Volume 6, Issue 1, Jan 2018.

[11]. M. Fang, L. Tang, X. Yang, Y. Chen, C. Li, and Q. Li, "FTPG: 758 A fine-grained traffic prediction method with graph attention network 759 using big trace data," IEEE Trans. Intell. Transp. Syst., vol. 23, no. 6, 760 pp. 5163–5175, Jun. 2021.

[12]. R.Srinivasan, G.Neelakrishnan, D.Vinoth and P.Iraianbu, "Design and Implementation of Novel Three Phase Multilevel Inverter for Smart Grid" International Journal of Multidisciplinary Educational Research, jan 2020, Volume 9, Issue 1(3) pp: 125-135

[13]. B. Liu, S. Yan, J. Li, Y. Li, J. Lang, and G. Qu, "A spatiotemporal 765 recurrent neural network for prediction of atmospheric PM2.5: A case 766 study Beijing," IEEE Trans. Computat. Social Syst., vol. 8, no. 3, 767 pp. 578–588, Jun. 2021

Cite this article as :