CRIME RISK PREDICTION IN URBAN AREAS USING UNSUPERVISED DOMAIN ADAPTATION TECHNIQUES

[1]D. Saikrishna, [2]S.Madhurya
[1]Assistant Professor, [2]MCA Student
Department Of MCA Student
Sree Chaitanya College of Engineering, Karimnagar

ABSTRACT
Accurate crime risk forecasting is essential for both city safety and inhabitants' quality of life. However, without tagged data, it is challenging to predict the likelihood of crime in a place. Many locations find it difficult to get high-quality tagged crime data because of maintenance costs and local regulations. For instance, the quantity of tagged data accessible for various cities may range significantly. It has shown possible to develop models for cities without labelled crime data by using what is known about crime prediction models in cities with an abundance of data. This prediction task is already difficult, though, and it is made even more difficult by the fact that crucial background information varies from place to place. In order to successfully estimate crime risk across cities, this study offers a feasible unsupervised domain adaptation model (UDAC) that accounts for the uneven environments. We start by identifying many similar source city grids for each target city grid in particular. To make sure that the contexts of the two cities are consistent, we construct additional contexts for the target city using the grids of the source cities. An unsupervised domain adaptation dense convolutional network was created to learn high-level representations for precise crime risk prediction as well as domain-invariant features for domain adaptation. The effectiveness of our approach is confirmed by extensive experiments using three real-world datasets.

Machine learning is a key component of the quickly growing field of data science. The training of many algorithms to provide predictions or classifications using statistical techniques has revealed the project's key discoveries. These kinds of insights should eventually drive application and business decision making, which in turn should affect crucial growth indicators.

Machine learning algorithms utilise this project data to build a model, which may subsequently make decisions or predictions without any special training. Machine learning algorithms are used by various datasets when traditional algorithm development is too difficult or impractical.

I. INTRODUCTION

For city people, crime is a constant threat and a significant contributor to their declining quality of life. This year, there have been 435 mass shootings in the United States, resulting in 1,648 injuries, 517 fatalities, significant property damage, and unimaginable grief [1]. Therefore, it is essential for people and society to be able to identify crime risk in order to prevent or reduce the chance of criminal acts. Fortunately, the availability of varied urban data in some places (e.g., Chicago) has given academics new chances to investigate crime-related issues [2, 3], [4], [5], [6], [7], [8], and [4], [5], [9], [10], [11]. 38 39 conclusions have been drawn from studies on the applicability of urban statistics to crime research. The possibility of stealing in highly populated places is one way that human migration may affect crime rates.

Some of the 44 cities do not make data available to the public for a variety of reasons, including the expensive cost of data collection and maintenance, the absence of clear laws, and rising privacy concerns. This is on top of the fact that they have different growth rates. Locals should thus have enough life experience to recognise danger when it arises. It is more challenging for tourists and other foreigners because not all residents have been there for a long time. For instance, a new paradigm known as transfer learning has arisen to address issues like population flow prediction and chain store site selection in places with limited data [12], [13]. This enables us to apply our knowledge from a more data-rich city (the source city) to a less data-rich city (the target city). Therefore, we use unsupervised transfer learning to examine crime risk prediction in cities without labelled crime data.

The problem is that without labelled crime data in the target city, a prediction model trained on sufficient labelled data from the source city would not be able to accurately forecast the risk of crime there. Inconsistencies in the available relevant context data between cities may be caused by various data gathering capabilities. For the sake of argument, let's say that NYC is the starting point and LA is the ending point. The city of New York has been collecting and making available multisource urban data, such as the distribution of points of interest (POIs) and taxi trip records, for a long time now, thanks to its extensive detection equipment deployment and long-running open data effort. A lot of urban data is also collected by certain cities. However, they withhold certain valuable and pertinent data, such taxi mobility statistics in areas like Los Angeles, from the public because of worries about 71 privacy problems or the hefty expenses of data collecting. As a result, cities using unlabeled data have worse crime risk prediction ability due to the contexts' inconsistency problem.

One obvious way to deal with inconsistent contexts is to train a model using just common context data from the source city and then fine-tune it to handle tasks in the target city. This way, you can ignore the inconsistent data particular to each city. The problem becomes much more severe when there is a lack of context data, since this might lead to the loss of valuable information for crime risk prediction. More data could lead to improved performance as a result of deep learning technology, which can extract more valuable information from networks. As a result, we want to develop a reliable model that can build source-city specific context data for the target city, solve the contexts' inconsistency issue, and use the source city's knowledge to forecast crime in the target city.

Here, we present UDAC, an unsupervised domain adaptation model, to forecast crime risk across cities while taking context inconsistency into account. By transferring knowledge from a source city with ample labeled data to a target city without such data, UDAC aims to tackle the problem of crime risk prediction in the target city. We take our cue from [15] and devise a system to solve the problem of inconsistent context by building potential city-specific context data for the target 98 city grids using context data from comparable source city grids. Later on, we provide a network that can successfully learn domain-invariant features for unsupervised domain adaptation while still learning effective features for crime risk prediction in the source city. The optimization procedure would take into account distribution discrepancy distance, domain categorization error, and crime risk prediction error all at once. Utilizing three real-world datasets from New York City, Chicago, and Los Angeles, comprehensive tests are carried out to confirm UDAC's efficacy.

Our work mostly contributes to the following areas.

1) We tackle the contexts' inconsistency problem across cities and take a promising start toward

unsupervised domain adaptation in crime prediction across cities.

2) Our proposed approach allows for the deep unsupervised domain adaptation technique to use information learnt from a source city with plentiful labeled data, making it effective for crime risk prediction in cities without labeled data. By building target-city-specific contexts and introducing a dense convolution network that learns effective features for accurate crime prediction and domain-invariant features for unsupervised domain adaptation, we may overcome the issue of inconsistent contexts across two cities. For crime prediction in the target city, the optimized network may be a viable option.

3) We use real-world datasets from three cities to demonstrate the efficacy of our proposed UDAC model via comprehensive trials. Our solution outperforms the state-of-the-art comparison methods, according to the experimental data.

II. LITERATURE SURVEY

Decision trees, random forests, and support vector machines are just a few examples of the machine learning algorithms that have shown promise in the field of crime risk prediction. Predictions of burglaries, robberies, and drug-related offenses are among the many crimes that have made use of these algorithms.

1. Adapting Domains Unsupervisedly for Cross-City Crime Risk Prediction

Applying adversarial training and feature alignment approaches to develop domain-invariant representations of crime data, this research provides an unsupervised domain adaptation method for crime risk prediction across cities. The authors review current methods for predicting crime risks and domain adaptation, while also highlighting the difficulties of adapting crime risk models across various cities. When tested experimentally on crime data from three separate cities, the suggested technique proved to be more accurate and resilient to domain changes than many baselines. In their last section, the writers address the strengths and weaknesses of their study, as well as future research directions.

2. Using urban open data, dynamically predicting the probability of road crime

This article suggests leveraging urban open data to make predictions about the likelihood of road crime using a machine learning method. Urban open data has the ability to provide valuable information for crime risk prediction in urban settings, according to the authors. They provide a machine learning pipeline that includes many data sources, such as traffic volume, weather, and crime statistics, and they examine current methods for predicting the likelihood of road crime. They also talk about how urban open data might be used for crime prediction. The authors demonstrate that their method is more accurate and efficient than many baseline models by comparing them side by side. At last, the writers wrap up by going over the highlights and lowlights of their work, along with possible directions for future studies. They draw attention to the relevance of dynamic risk prediction for road crime and the possibilities of urban open data for crime prediction in cities.

3. Metropolitan Area Theft Crime Risk Assessment

The authors emphasize the significance of stealing crime prediction in metropolitan environments and provide an outline of crime prediction in general. Various methods for crime prediction have been examined, such as classical statistical models and machine learning algorithms.

After outlining their methodology, the authors go into detail about how they choose features, preprocess data, and employ several machine learning models to make predictions.

The findings section showcases the experimental assessment of the suggested method using data collected from a Mexican city. The authors demonstrate that their method is more accurate and efficient than many baseline models by comparing them side by side.

Lastly, the authors address the work's limits and future research directions, stressing the need of broad and extensive datasets for better crime prediction in cities.

Predicting the Kind and Frequency of Crimes with the Use of Machine Learning Algorithms

This suggests a method for forecasting the frequency and kind of crimes in cities using machine learning.

In their introduction, the writers discuss crime prediction and its difficulties, including the absence of current and reliable data. Various methods for crime prediction have been examined, such as classical statistical models and machine learning algorithms.

The authors detail their method for predicting the types and occurrences of crimes using algorithms for machine learning. They provide an engineering and feature selection procedure to get useful characteristics from the given data. Also included are the models for predicting the incidence and nature of crimes, such as decision trees, random forests, and support vector machines.

Lastly, the authors address the work's shortcomings and future research directions, stressing the need of broad and extensive datasets for better crime prediction.

Fifth, a Smart Policing Method Based on Offense and Danger

This study presents a smart police strategy that uses machine learning to anticipate the kinds of crimes and the dangers they pose, in an effort to reduce crime rates. Also included in the article is research on where and how crimes have been detected using GIS and other data sources. A smart police strategy that is being proposed makes use of a machine learning pipeline that integrates several data sources, such as geographic, demographic, and crime statistics. The authors go into detail about the feature engineering process, the prediction models, the evaluation metrics, and the smart policing technique they propose. This technique employs a machine learning pipeline that gathers data from a variety of sources, including geographic, demographic, and crime statistics. Along with the assessment criteria, the writers go on the feature engineering process, prediction models, and more. Last but not least, they propose expanding their smart policing method to other areas and emphasize its potential for increasing enforcement efficiency and decreasing crime rates.

6. A Domain-Based Transfer Network for the Diagnosis of Cross-Domain Faults

By combining domain adaptation with deep learning approaches, this research introduces a novel method for defect diagnostics. The authors discuss the difficulties of cross-domain diagnosis and provide their solution, which use domain adversarial transfer learning to acquire sensor data representations that are domain invariant, therefore improving the accuracy of diagnoses. A domain adversarial transfer network for defect diagnosis is detailed in the methods section. This network uses an encoder-decoder design with a domain discriminator. The authors offer an approach that has potential uses outside of industrial systems, and the findings show that it works on two datasets from different areas.

III. SYSTEM ANALYSIS

EXISTING SYSTEM

Prior research on unsupervised domain adaptation and crime prediction is relevant to our approach. This section provides a concise overview of a few works that fall under these two broad groups.

In the last few decades, researchers have dabbled with the idea of predicting urban crime. It is crucial to identify pertinent external factors for the purpose of studying crime prediction. Information about the weather and temperatures, for example, may be useful to criminal investigations, according to Ranson's [16] analysis of meteorological data. 145

Using a variety of urban data, such as weather reports, point-of-interest distribution, and taxi trip records, Zhou et al. [17] performed a detailed investigation of crime. An important link between this data and criminal activity was discovered. It has become common to use feature analysis to build efficient models that provide accurate predictions. Over the last several decades, numerous spatio-temporal prediction models have been put forward to address a range of problems, including traffic prediction and inference, social event prediction, air quality prediction, logistics management optimization, and many more. These models capture both spatial and temporal dependencies. 1

Even without labelled data, these spatio-temporal prediction models made a pitiful showing in the prediction tests. Many approaches have been employed to enhance the prediction performance of crime data, including Twitter, Foursquare, demographic, and taxi trip data, as well as linear and count models, and machine learning models [9], [24], [25]. Using crime data, POI, and 311 public service complaint data, Huang et al. [4] suggested an attention-layer hierarchical recurrent neural network to learn the temporal significance of patterns and forecast future criminal occurrences. Using data from Twitter and points of interest (POIs), Yang et al. [3] trained several machine learning models (such as decision trees and random forests) to identify high-crime areas in New York City. 169

An integrated model was suggested by Yi et al. [26] that extracts spatio-temporal characteristics and improves future crime prediction performance using a clustered continuous conditional random field (CCRF) technique. In addition to the stacked denoising autoencoder for learning pairwise interactions between spatial areas and long short-term memory (LSTM) units for learning nonlinear relationships between input and output, they also used the CCRF approach previously discussed [10]. To forecast crime prior knowledge at the road level, Zhou et al. [27] suggested a hierarchical framework that updates prediction results by introducing recurring crime characteristics. The architecture begins by establishing a pattern using spatio-temporal data. They went even further into the dynamics of crime by looking at it through the lens of impact propagation, and they came up with a model to forecast the likelihood of future road-level crimes using zero-inflated negative binomial regression [28].

Disadvantages

- Currently, there isn't a methodology that takes advantage of the Auxiliary Features' Construction method.
- Additionally, the system doesn't have a model that can effectively build context data specific to the target city, addressing the issue of context inconsistency. With this model, the target city can then use the knowledge it has learned from the source city to predict crime.

PROPOSED SYSTEM

To solve the problem of context inconsistency, the system suggests an unsupervised domain adaptation model (UDAC) for predicting crime risk across cities. By learning from a source city with plenty of labeled data, UDAC hopes to address the problem of crime risk prediction in a target city that doesn't have any such data. In order to solve the problem of inconsistent context, we build a method97 that takes inspiration from [15] to build potential city-specific context data for the target city grids using context data from comparable source city grids. Later on, we provide a network that can successfully learn domain-invariant features for unsupervised domain adaptation while still learning effective features for crime risk prediction in the source city. All three metrics—distribution discrepancy distance, domain classification error, and crime risk prediction error—would be optimized simultaneously. Utilizing three real-world datasets from New York City, Chicago, and Los Angeles, comprehensive tests are carried out to confirm UDAC's efficacy.

1) We tackle the contexts' inconsistency problem across cities and take a promising start toward unsupervised domain adaptation in crime prediction across cities.

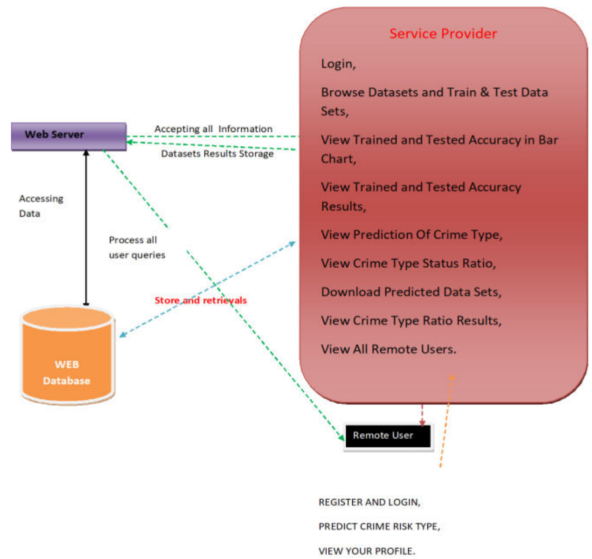
2) Our proposed approach allows for the deep unsupervised domain adaptation technique to use information learnt from a source city with plentiful labeled data, making it effective for crime risk prediction in cities without labeled data. As a solution to the problem of inconsistent city contexts, we build target city-specific contexts first, and then introduce a dense convolutional network that can learn domain-invariant features for unsupervised domain adaptation and effective features for accurate crime prediction. For crime prediction in the target city, the optimized network may be a viable option.124 total

3) We use real-world datasets from three cities to demonstrate the efficacy of our proposed UDAC model via comprehensive trials. Our solution outperforms the state-of-the-art comparison methods, according to the experimental data.

Advantages

- An efficient model that may build potential source-city-specific context data for the target city in order to solve the issue of context inconsistency, and, using the information it has gained from the source city, forecast crime in the target city.
- The Dense Convolutional Network with Unsupervised Domain Adaptation system was used in this study.

IV. SYSTEM ARCHITECTURE



V. SYSTEM IMPLEMENTATION

Modules

Service Provider

The Service Provider must provide a valid user name and password to log in to this module. Following a successful login, one may do a number of tasks, including Look Through Datasets and Test & Training Data Sets, View the results of trained and tested accuracy, view the status ratio of crime type, download predicted data sets, view the results of crime type ratio, and view all remote users. You can also view the accuracy in a bar chart.

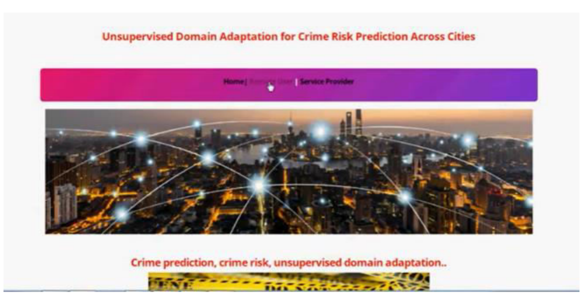
View and Authorize Users

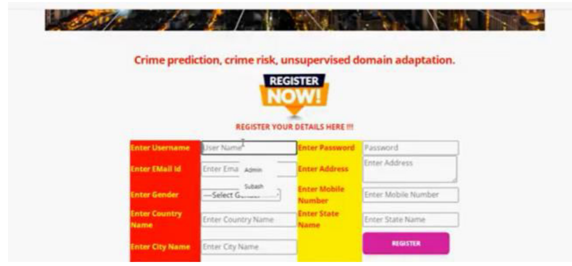
The administrator may see a list of all enrolled users in this module. The administrator may see user information here, including name, email address, and address, and they can also approve people.

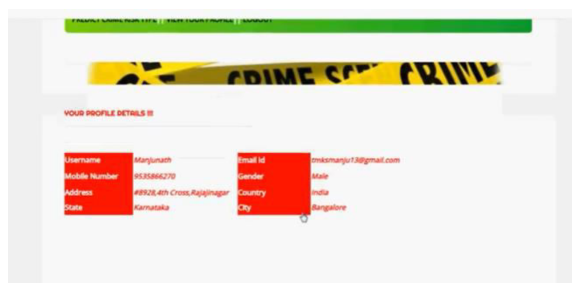
Remote User

There are n numbers of users present in this module. Prior to beginning any actions, the user must register. The user's information is saved in the database when they register. Upon successful registration, he must use his permitted user name and password to log in. Upon successful login, the user may do several tasks such as registering and logging in, predicting the kind of crime, and seeing their profile.

VI. RESULTS








Unsupervised Domain Adaptation for Crime Risk Prediction Across Cities


Unsupervised Domain Adaptation for Crime Risk Prediction Across Cities

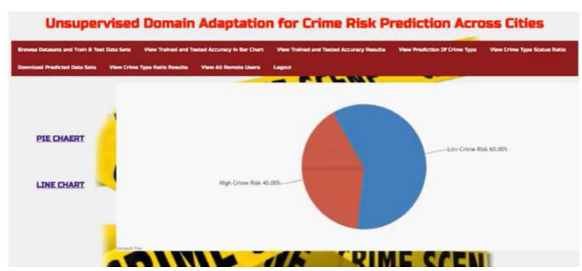

Unsupervised Domain Adaptation for Crime Risk Prediction Across Cities


Unsupervised Domain Adaptation for Crime Risk Prediction Across Cities


Unsupervised Domain Adaptation for Crime Risk Prediction Across Cities


Unsupervised Domain Adaptation for Crime Risk Prediction Across Cities
PIE CHAERT
LINE CHART


Unsupervised Domain Adaptation for Crime Risk Prediction Across Cities
PIE CHAERT
LINE CHART


Unsupervised Domain Adaptation for Crime Risk Prediction Across Cities
PIE CHAERT
LINE CHART


Unsupervised Domain Adaptation for Crime Risk Prediction Across Cities
PIE CHAERT
LINE CHART


Unsupervised Domain Adaptation for Crime Risk Prediction Across Cities
PIE CHAERT
LINE CHART

VII. CONCLUSION

In order to overcome the prediction issues with unlabelled data, this research suggests an unsupervised domain adaption method. By sending information with labelled data from a source city to a destination city to assess crime risk, this method may address the issue of context inconsistency between cities. We identify many similar source city grids for each target grid and split the cities into multiple grids of the same size. Based on these pairings, we construct auxiliary characteristics for the target city to address the problem of context inconsistency between cities. In order to learn from the source city and apply it to the target city for future crime risk prediction, we then propose an unsupervised domain adaptation module built on dense-convolutional networks. Domain-invariant qualities are taught to facilitate information transfer. Through comprehensive experiments utilising real-world data from LA, NYC, and Chicago, we demonstrate the effectiveness of our approach. According to the experimental results, our proposed method outperforms existing cutting-edge comparison techniques.

In the future, we hope to improve our work from a number of perspectives. Our initial goal is to examine prediction performance in more extreme situations when the source and destination cities differ. Our second goal is to investigate a more thorough kind of unsupervised criminal risk prediction, which includes calculating the probability of crime along roads. Due to the severe data scarcity problem, this would be more challenging.

REFERENCES

[1] (2020). GunViolenceArchive. List of Mass Shootings in the United States in 2019. [Online]. Available: https://www.gunviolencearchive.org/reports/mass-shooting

[2] M. S. Gerber, “Predicting crime using Twitter and kernel density estimation,” Decision Support Syst., vol. 61, pp. 115–125, May 2014.

[3] D. Yang, T. Heaney, A. Tonon, L. Wang, and P. Cudré-Mauroux,“CrimeTelescope: Crime hotspot prediction based on urban and social media data fusion,” World Wide Web, vol. 21, pp. 1323–1347, Sep. 2018.

[4] C. Huang, J. Zhang, Y. Zheng, and N. V. Chawla, “DeepCrime: Attentive hierarchical recurrent networks for crime prediction,” in Proc. 27th ACM Int. Conf. Inf. Knowl. Manage., 2018, pp. 1423–1432.

[5] L. Xia et al., “Spatial–temporal sequential hypergraph network for crime prediction with dynamic multiplex relation learning,” in Proc. 30th Int.Joint Conf. Artif. Intell., Aug. 2021, pp. 1631–1637.

[6] P. Das, A. K. Das, J. Nayak, D. Pelusi, and W. Ding, “Incremental classifier in crime prediction using bi-objective particle swarm optimization,”Inf. Sci., vol. 562, pp. 279–303, Jul. 2021.

[7] H. Wang, D. Kifer, C. Graif, and Z. Li, “Crime rate inference with big data,” in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Aug. 2016, pp. 635–644.

[8] H. Wang, H. Yao, D. Kifer, C. Graif, and Z. Li, “Non-stationary model for crime rate inference using modern urban data,” IEEE Trans. Big Data, vol. 5, no. 2, pp. 180–194, Jun. 2019.

[9] L. Vomfell, W. K. Härdle, and S. Lessmann, “Improving crime count forecasts using Twitter and taxi data,” Decis. Support Syst., vol. 113,pp. 73–85, Sep. 2018.

[10] F. Yi, Z. Yu, F. Zhuang, and B. Guo, “Neural network based continuous conditional random field for fine-grained crime prediction,” in Proc. 28th Int. Joint Conf. Artif. Intell., Aug. 2019, pp. 10–16.