Bilkent University

CS 490 - Introduction to Research

Introduction to Research in Computer Engineering and Science

Interim Report

Author
Selim Firat Yilmaz - 21502736
Advisor
Hamdi Dibeklioglu

April 8, 2019

Spatiotemporal Crime Prediction via Multitask Learning

Selim Firat Yilmaz
Bilkent University
Ankara/Turkey
firat.yilmaz(at)ug.bilkent.edu.tr

Hamdi Dibeklioglu
Bilkent University
Ankara/Turkey
last_name(at)cs.bilkent.edu.tr

1. Introduction

Criminology studies criminal actions and its patterns [9]. Although crime seems as a random event for many people due to its complexity, crime pattern theory argues that patterns exists in such complexities [9]. According to the crime pattern theory, every component of criminal events complexly related to its past trajectory, status, routines, and the environment [9]. Understanding the patterns of criminal events would lead us to understand crimes and their motives.

In our work, we aimed to exploit the existence of crime datasets from multiple cities. These datasets have different number of regions, and different number of crime types. Also, at least to our knowledge, there are no information exists from the dataset published that relates the classes between crime datasets of different cities.

2. Related Work

Crimes are a non-random events happening in place and time between an offender and a target (a living or non-living) [1]. Environmental criminology analyzes the relations between people's behaviors and external factors such as law, time, and geography and focuses on exploiting specific crime patterns within an environmental context [11]. Social criminology focuses on relations between criminal occurrences and socioeconomic, cultural, demographic indicators [11]. Both environmental and social criminology seem to have many useful features for machine learning tasks. Many integrible information such as census data of specific locations is not available for space-time tuples but usually is available in only regional or temporal or recent. Although the crime datasets has many occurrences, their spatiotemporal distribution is uneven, highly sparse and imbalanced for some crime types [8]. Hence, this section aims to answer following questions:

What are the used features?

How only spatial/temporal/recent information (such as surrounding buildings of today for specific region) are incorporated into space-time queries?

How prediction(gridding) regions are segmented?

Which algorithms/architectures of machine learning are used for crime prediction?

Which ways are used to deal with spatiotemporal sparsity of the problem?

What applications of crime prediction is used today?

Nyugen et al. [5] incorporates demographic, economic, educational and ethnic information into crime prediction. They concatenate the information vector and feeds it into support vector machines(SVM), Random Forest(RF), gradient boosting tree, and multilayer perceptron. Latter, they deal with imbalanced nature of the crime data by undersampling the majority crime classes.

Kang and Kang [4] incorporates demographic, housing, education, economic, weather and Google street view images. The categorizes these information into spatial, temporal, and environmental context groups. They propose a multi-modal deep neural network architecture by feeding them into 3 distinct sequence of hidden layers, concatenating their last layers and again 3 hidden layers. Prediction is done via softmax layer in output layer. They train whole network jointly.

Porzi et al. [6] proposes convolutional neural network based architecture for ranking safety levels of spatial regions. They feed scenes of urban crime location's scene to the network and predicts safety levels of the specific locations.

Chen et al. [2] and Gerber [3] integrates the information from Twitter. They collect the GPS-tagged data and calculates features per grid point. Chen et al. [2] applies lexicon based algorithm sentiment detection to collect polarity information per grid in addition to the weather information. Gerber [8] applies latent dirichlet analysis(LDA) to detect topical distribution of geo-tagged tweets. Latter, both [2] and [3] applies kernel density estimation.

Wang et al. [8] employs spatiotemporal ResNet (St-ResNet) architecture of [10] and emsembles that architecture with external features such as weather information feeded to fully connected.

3. Methodology

3.1. Problem Description

Crime prediction is the problem of predicting a future crimes location and types. Particularly, we divide the city into J number of distinct regions K number of time bins, and L number of crime types described in dataset section.

The challenging part of the problem is the need for capturing spatiotemporal relations between different geospatial regions and different times since we are using a datasets from distinct cities as New York city and Chicago city. Also, the crime types we want to predict is considerably different in datasets for which we capture relationships in between.

3.2. Evaluation

We see crime prediction as a discrete problem and so using precision, recall, and F1 metrics due to the sparsity of the datasets we use. Currently, since project is ongoing, only accuracy metrics exists but their effectivenes can still be observed regardless of the sparsity of the datasets.

Currently, accuracy is calculated as each crime type of each location-time tuples are distinct problems so that if there occur a crime in a location-time tuple and the model says 1 or there does not occur a crime and model says 0 for it is counted as correct classification. Otherwise, it is counted as false classification.

Experimentation protocol follows cross validation for time series. Firstly, the data are ordered in terms of time.

- Choose the part of (0, 25%] as training data, validate on the part of (25%, 35%], and test on the part of (35%, 50%].
- Choose the part of (0, 50%] as training data, validate on the part of (50%, 60%], and test on the part of (60%, 75%].
- Choose the part of (0, 75%] as training data, validate on the part of (75%, 85%], and test on the part of (85%, 100%].
- Then, average the accuracy scores for the experiments above.

Also, it is important to note that there exists some work[cite needed] who uses R^2 score to evaluate their models when predicting the rate of the crime.

3.3. Baseline – Meta Predictors

To deal with the comparison problem of the models via accuracy, we tested the datasets with the following meta predictors.

Method	Accuracy
Ones Only M.P.	0.691
Zeros Only M.P.	0.309
Random M.P.	0.496
Baseline – SVR	**0.703**

Table 1. Current esults for Chicago P.D.'s crime dataset for binned for 6 hours in time, and for 24x24 spatial bins. Only the data between from beginning of 2015 to the end of 2017 is used for now.

Ones Only Meta Predictor predicts 1 only, saying that there will be a crime, for each time-location tuple regardless of the given training data.

Zeros Only Meta Predictor predicts 0 only, saying that there will not be a crime, for each time-location tuple regardless of the given training data.

Random Meta Predictor predicts 0 or 1 uniformly at random for each time-location tuple regardless of the given training data.

3.4. Baseline – Support Vector Regression

Given a training data with input Grid ID and time id; converted to two different one-hot vectors and concatenated. Then, it is given to the SVR machine that is predicting N different outputs representing different crime types where N is 34 for Chicago crime dataset, and N is 7 for New York crime dataset.

3.5. Multitask Network Architecture

We propose a multi-task model consisting of deep convolutional neural networks. Firstly, we grid the datasets of Chicago and New York as same and number of bins for each of the input variables: latitude, longitude, and time. These bind indices are propagated to the embedding layers, firstly. Then they will be propagated to the convolutional LSTM network [7].

In this model, we aim to capture the relationship of neighborhood as well as temporal relevancy between the crime data of distinct cities.

4. Datasets

4.1. Chicago Crime Dataset

We employ Chicago Police Department's crime dataset to model crime prediction task. The dataset contains many information such as street name, but we only use crime type, location in terms of latitude and longitude, the time of occurrence of crime. There are 34 different types of crimes in the dataset: Arson, assault, battery, burglary, concealed carry license violation, sexual assault, criminal damage, criminal treepass, deceptive practice, gambling, homicide, human trafficking, interference with public officer, intimidation, kidnapping, liquor law violation, motor vehicle theft, narcotics, non-criminal, obscenity, offense involving children, other narcotic violation, other offense, prostitution, public indecency, public peace violation, ritualism, robbery, sex offense, stalking, theft, weapons violation. The location boundaries of crimes are limited with the city of Chicago. Also, the time of crimes' occurences are from 2002 to 2017.

4.2. New York Crime Dataset

Since we propose to model a multitask network, we chose second task as the predicting an classifying crimes in New York city. For this reason, we employ New York City Police Department's crime dataset to model multitask crime prediction task. The dataset contains many information such as street name, but we only use crime type, location in terms of latitude and longitude, the time of occurrence of crime. There are 7 different types of crimes in the dataset: the violent crimes of murder, rape, robbery, and aggravated assault; and the property crimes of burglary, larceny and motor vehicle theft. The location boundaries of crimes are limited with the New York city. Also, the time of crimes' occurrences are from 2000 to 2015.

5. Conclusion and Ongoing Work

We aimed to find patterns in criminal events throughout the study using multitask learning. Note that our work may also be represented as a transfer learning where we transfer the pre-trained initial layers in structured manner.

We believe that we can improve the current baseline models. Thus, The goals of remaining part for this study is as follows:

- Evaluating the models with F1, precision, and recall metrics
- Evaluating the models with different grid settings.
- Completing the implementation of multitask model and evaluate it on different settings.
- Evaluating all the models with all the data.
- Implement a Spatially gridded LSTM baseline.
- Fully describe the details of the network.
- Complete this report and make it consistent.

References

[1] P. J. Brantingham, P. L. Brantingham, et al. *Environmental criminology*. Sage Publications Beverly Hills, CA, 1981. 1

[2] X. Chen, Y. Cho, and S. Y. Jang. Crime prediction using twitter sentiment and weather. In *2015 Systems and Information Engineering Design Symposium*, pages 63–68. IEEE, 2015. 1

[3] M. S. Gerber. Predicting crime using twitter and kernel density estimation. *Decision Support Systems*, 61:115–125, 2014. 1

[4] H.-W. Kang and H.-B. Kang. Prediction of crime occurrence from multi-modal data using deep learning. *PloS one*, 12(4):e0176244, 2017. 1

[5] T. T. Nguyen, A. Hatua, and A. H. Sung. Building a learning machine classifier with inadequate data for crime prediction. *Journal of Advances in Information Technology Vol*, 8(2), 2017. 1

[6] L. Porzi, S. Rota Bulò, B. Lepri, and E. Ricci. Predicting and understanding urban perception with convolutional neural networks. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 139–148. ACM, 2015. 1

[7] X. SHI, Z. Chen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. WOO. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 802–810. Curran Associates, Inc., 2015. 2

[8] B. Wang, D. Zhang, D. Zhang, P. J. Brantingham, and A. L. Bertozzi. Deep learning for real time crime forecasting. *arXiv preprint arXiv:1707.03340*, 2017. 1

[9] R. Wortley and M. Townsley. *Environmental criminology and crime analysis*. Taylor & Francis, 2016. 1

[10] J. Zhang, Y. Zheng, and D. Qi. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017. 1

[11] X. Zhao and J. Tang. Crime in urban areas:: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 20(1):1–12, 2018. 1