Predicting crime classes based on spatial-temporal data

Noor Damstra
s2675188
Leiden University, Urban Computing
Niels Bohrweg 1, 2333 CA Leiden

Abstract

Crime occurs almost everywhere and at all times of the day. But it is very likely that there are locations and moments in time where crime rates are higher. By modelling spatial-temporal patterns crime rate and crime type could be predicted. Police officers in Chicago are stationed per beat, which is where they patrol. Modifying the stationing of police officers through the understanding of the patterns that steer the predictions could aid crime prevention. Furthermore, the knowledge of knowing when and where a certain type of crime might happen could be extremely useful information. This research found that although temporal and spatial trends are present, separately they aid little in the classification of crime type. While a combination of the two leads to a higher accuracy

Introduction

Almost all urban cities around the world are looking to prevent crime and lower crime rates. One way researchers in the world of machine learning are trying to aid this process is through the prediction of crime. An example is DevianceNet: Learning to Predict Deviance from a Large-Scale Geo-Tagged Dataset ((Park et al. 2022)). In their paper they aim to predict deviance, or crime, based on image data. Certain factors in the visual environment attribute to a heightened probability of deviance. Through these appearances they were able to classify 5 classes of deviance quite well. Being able to predict deviance based on the environmental factors could be used as crime prevention through city planning. Furthermore, if the knowledge from the model would become public then individuals could use it to avoid certain areas where crime is more likely. But there are more factors at play when it comes to crime and these two solutions are not the only way to reduce crime. A great way to be able to prevent criminal activity is to have police present at the right place and at the right time. The police department of Chicago uses a beat system. The city of Chicago consists of 277 beats. A beat is a small geographic area of the city where 8-9 police officers are stationed for at least a year (ChicagoPd). These same officers run shifts of patrols in their beat. So at any given time a certain number of officers are on patrol in their designated area. However, there might be a more effective way of distributing officers around the city. By capturing spatial-temporal trends in crime rates, it could be possible to station the right number of officers at a station and during patrol.

crime type	class	size
NON-CRIMINAL	0	3
OTHER OFFENSE LIQUOR LAW VIOLATION GAMBLING OBSCENITY PUBLIC INDECENCY PUBLIC PEACE VIOLATION OTHER NARCOTIC VIOLATION	1	15479
CONCEALED CARRY LICENSE VIOLATION PROSTITUTION INTERFERENCE WITH PUBLIC OFFICER WEAPONS VIOLATION NARCOTICS	2	13859
CRIMINAL DAMAGE THEFT DECEPTIVE PRACTICE BURGLARY INTIMIDATION STALKING CRIMINAL TRESPASS	3	107862
ARSON ROBBERY MOTOR VEHICLE THEFT ASSAULT KIDNAPPING	4	59324
CRIMINAL SEXUAL ASSAULT BATTERY SEX OFFENSE	5	43979
OFFENSE INVOLVING CHILDREN HUMAN TRAFFICKING HOMICIDE	6	2228

Table 1: Classification of crime types

Hossain et al. (2020) successfully build an algorithm that could classify two classes of crime, frequent and rare, through spatial-temporal modelling.

However their spatial model laid emphasis on districts. Districts are much larger than beats and it might be hard for police departments to become more effective at preventing crime based on the information from this model. Furthermore, splitting the data into two classes could lead to information loss. Having a way to distinguish more precisely between different types of crimes could also play a role in the distribution of officers. Specific types of crimes may have a higher priority due to the city's goals and or problems. Moreover, certain types of crime might be easier to prevent than others. This research aims to predict seven classes of crime data through spatial-temporal trends in the data and wishes to compare it to separate temporal and spatial models. Furthermore, it intends to reveal the capability of the algorithm to work with the data at hand versus it's capability when provided with extracted features.

Method

The Chicago 2023 crime data obtained from https://data.cityofchicago.org/Public-Safety/Crimes-2023/xguy-4ndq/about_data contains cases of 31 different types of crime recorded over the course of the entire year. These 31 different crime types have been sorted into 7 different classes. Where the class number corresponds to the severity of the crime. Crimes containing violence or direct danger to other individuals are ranked highest and theft or damage to property would rank intermediate, while crimes that caused minor disturbances would be ranked lowest. A list with the grouping of the crime types and their sizes can be found in table 1. The algorithm that will be used to model the different trends is Extreme Gradient Boosting (XGBoost). This algorithm is selected for it's ability to handle both categorical and numeric data and it's inherent capability of modelling complex relationships, which are often found in temporal and spatial data. In total 6 models are created, two for each trend: temporal, spatial, spatial-temporal. During hyper parameter optimization the performance of each model is assessed through 5-fold cross validation. The best model for each trend are investigated further by analysing their performance on the test set. More specifically their accuracy, precision and recall are calculated.

Temporal model

To determine if there might be some patterns in the data a plot of the occurrence of crime per class over the course of a week is created. This graph can be seen in figure 1. The figure shows that some temporal trends are definitely present and it seems that crime increases over the course of a day with a peak at midnight. It is hard to distinguish difference between the classes. To further inspect the pattern in each class the auto-correlation for each of them is calculated. A shift of half an hour is taken for each lag in this calculation of the auto-correlation. The visualizations of the auto-correlations are shown in figures 6, 7, 8, 9, 10and 11 in the Appendix. After inspection of the peaks of each class it's found that all classes have most of their peaks 24 and/or 48 lags apart. Which means that the spikes of crime rate occur at midday and midnight. It also means that there is little difference in temporal trends between classes.

During the analysis of just the temporal data, two models are built. One which contains the plain date and time data present in the dataset and one which also has the count per time of day as an extra feature. These features are synonymous to the probability that a certain crime belongs to it's class. The best parameters and results for both models are found in table 2 (temporal model 1 and temporal model 2). Model 2 performed equivalent to model 1 on the validation sets. The final evaluations were done on model 2.

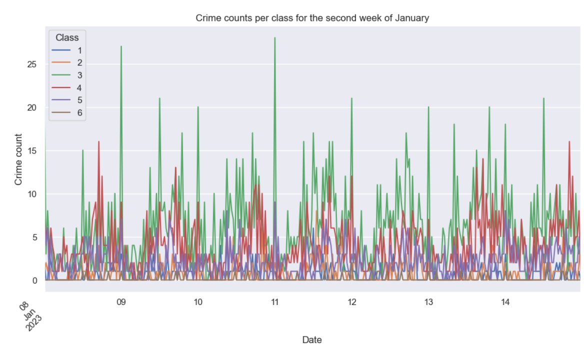

Figure 1: The crime rate per class from January 8th 00:00 to January 14th 23:59

Spatial model

In this section the spatial trends, specifically regarding beats, are explored. A shapefile regarding info on the beats was obtained from https://data.cityofchicago.org/Public-Safety/Boundaries-Police-Beats-current-/aerh-rz74. The shapefile was used to create visualizations about crime rates for each class types. These visualization are found in figure 2 . These visualizations made clear that there are differences in crime rate per beat per class but that there are also similarities. Beat 1834 were in the top 5 in crime count for classes x x and x.

Again two models were created one with just the location data present in the original data set and one with features that contains the crime counts per class per beat. The results of their hyperparameter optimization can be found in table 2 (spatial model 1 and spatial model 2). Model two outperformed model one and is evaluated on the test set.

Spatial-Temporal model

For the final model an extra feature regarding the daily pattern per beat is created for each class. Finally two spatial temporal models are created. One contains info present in the original dataset on both the location and the time and date of the crime. The other with the extra features, namely the counts of crime per class per beat per

hyperparameter	temporal model 1	temporal model 2	spatial model 1	spatial model 2	spatial-temporal model 1	spatial-temporal model 2
column sample per tree	0.8	0.8	0.6	0.6	0.6	0.6
learning rate	0.1	0.1	0.1	0.1	0.1	0.1
max. depth	5	5	3	3	5	3
subsample	0.7	0.7	0.5	0.7	0.7	0.7
accuracy	0.480	0.481	0.492	0.492	0.498	0.532

Table 2: The best hyperparameters for each model obtained through 5-fold cross-validation and their average accuracy

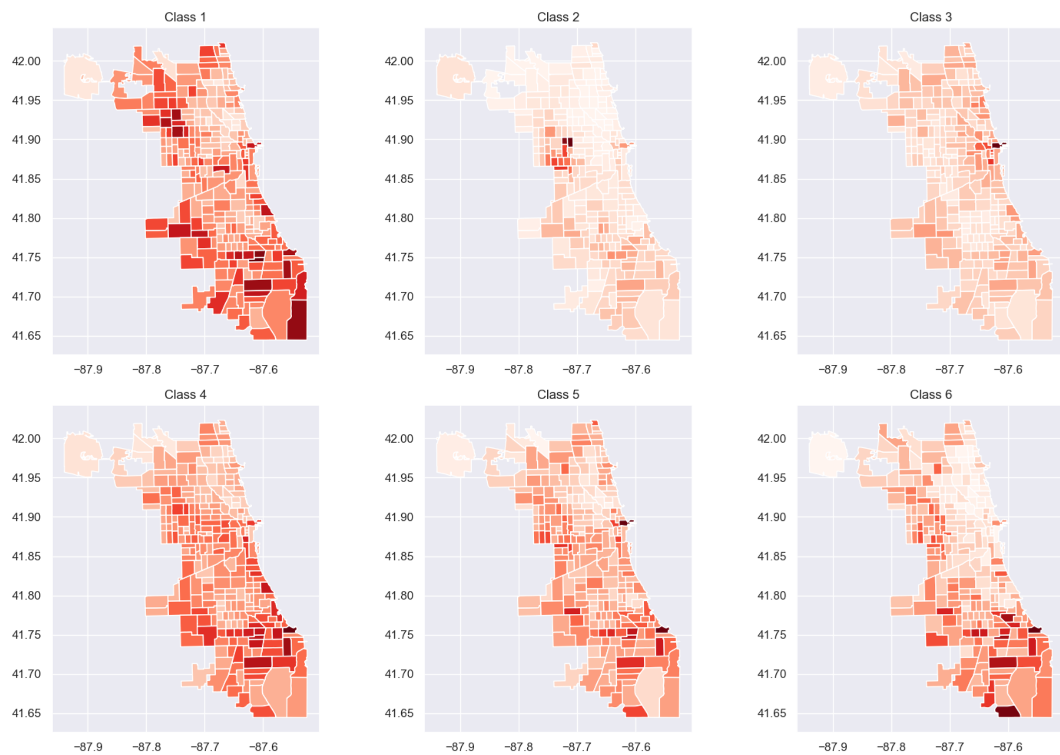

Figure 2: Crime count per beat for each class. Darker colors indicate the larger amounts of crime within that class but are not valued the same over the subplots.

time of day. The results of the hyperparameter optimization for these models can be found in table 2. The second model performed best and is evaluated on the test set.

Results

The temporal model with the crime counts for each half hour of the day for each class obtained an accuracy of 0.278 (table 3). The final spatial model which contained crime counts per beat for each class obtained an accuracy of 0.297 (3). The precision of both models (spatial and temporal) scored the highest in class 3 and their recall scored highest in class 3 as well(table 4 and table 5). Both also scored lowest in terms of precision in class 1 and 6. They both were not able to score above 0.1 percent on recall in class 1. The spatial-temporal model performs better in terms of accuracy, precision and recall. It's especially better at correctly classifying class 1. The most important feature for the spatial and temporal models was the location description (figure 3 and 5). This was also an important feature for the spatial-temporal model but the extracted pattern features were even more important.

Discussion

Correctly classifying crime types based on spatial-temporal data obtains more accurate results than using just spatial or temporal data. But this model does not yet yield outstanding results. The added features in the temporal and spatial model did not perform better than their basic counterparts temporal model. On the test sets the models with added featured performed much lower in terms of accuracy than on the validation sets. This is probably due to overfitting on the train data. The spatial-temporal model however did see a real increase in accuracy over the basic model.The temporal and spatial

	temporal model	**spatial model**	**spatial-temporal model**
accuracy	0.278	0.297	0.444

Table 3: Accuracy per model

	1	2	3	4	5	6
temporal model precision	0.0625	0.295	0.488	0.250	0.298	0.006
spatial model precision	0.049	0.366	0.491	0.243	0.243	0.006
spatial-temporal model precision	0.213	0.410	0.508	0.419	0.362	0.031

Table 4: Class precision for each model

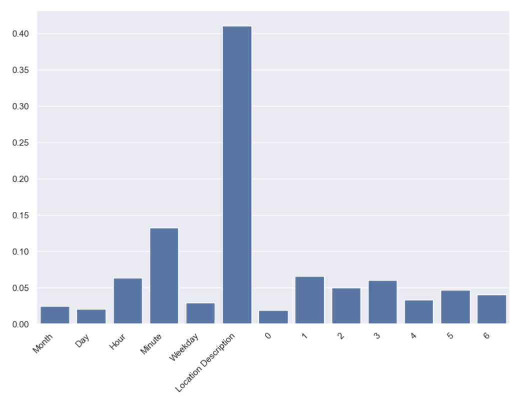

Figure 3: Feature importance of the final temporal model

model both performed well in terms of precision and recall on class 3. Which also happens to be the largest class. The bad performance in terms of recall on all other classes, except class 6, means that out of all instances of a certain class the model only classified as that class a small number of times. The slightly better precision means that when he classified something in a certain class it was correct a slightly higher percent of the time. The two models do perform similarly on both metrics and no conclusions about which one has a better performance can be made. The spatial-temporal model clearly performs better overall. This increase in performance was clearly obtained through the added features as they were the most important. From this it's possible to conclude that there definitely are temporal patterns in per beat.

One of the main issues that probably makes the classification very difficult is the manner in which the classes are created. It would be useful to learn what the priorities of the police department are and use their expertise to build a class system. Another way to potentially see an increase in performance is to represent the data in different

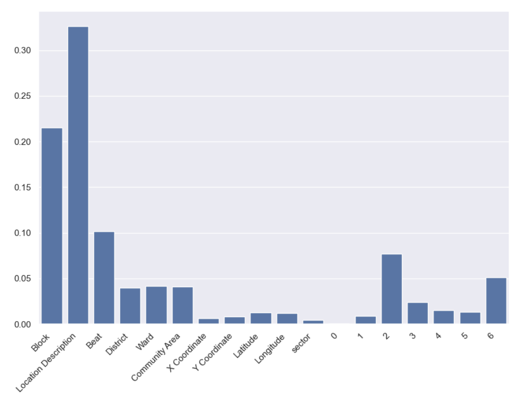

Figure 4: Feature importance of final spatial model

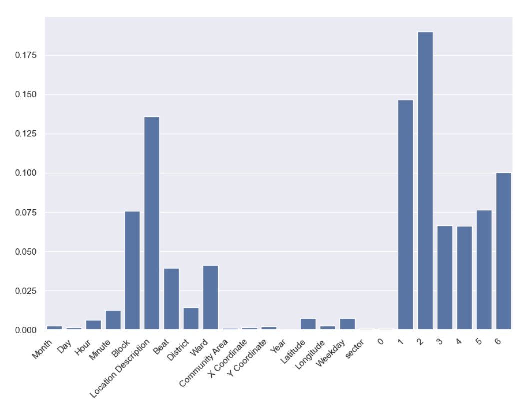

Figure 5: Feature importance of final spatial-temporal model

	1	2	3	4	5	6
temporal model recall	>0.001	0.047	0.530	0.044	0.146	0.251
spatial model recall	0.001	0.035	0.601	0.043	0.086	0.240
spatial-temporal model recall	0.016	0.128	0.784	0.190	0.215	0.249

Table 5: Class recall for each model

ways. More complex features could be extracted to boost the models effectiveness. Interactions, for example, could prove to be very informative. Crime events at a certain location could influence other locations, i.e. there might be correlations.
One other feature that was included in all models that did not give temporal or direct spatial information was the location description. This feature was an important feature in every single model. It would be useful to dive deeper into this variable and see if relevant information could be retrieved from it to aid in crime prevention. This variable also goes hand in hand with the research done by Park et al. (2022). But instead of an image of the area it is a very short description.
It is not yet possible to say whether spatial-temporal models can accurately classify Chicago crime. But through the above mentioned improvements it might be possible.

References

ChicagoPd. ???? Beat Officers | Chicago Police Department — home.chicagopolice.org. https://home.chicagopolice.org/community-policing-group/how-caps-works/beat-officers/. [Accessed 01-02-2024].

Hossain, S.; Abtahee, A.; Kashem, I.; Hoque, M. M.; and Sarker, I. H. 2020. *Crime Prediction Using Spatio-Temporal Data*, 277–289. Springer Singapore. ISBN 9789811566486.

Park, J.-H.; Park, Y.-J.; Lee, J.; and Jeon, H.-G. 2022. DevianceNet: Learning to Predict Deviance from a Large-Scale Geo-Tagged Dataset. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(11): 12043–12052.

Appendix

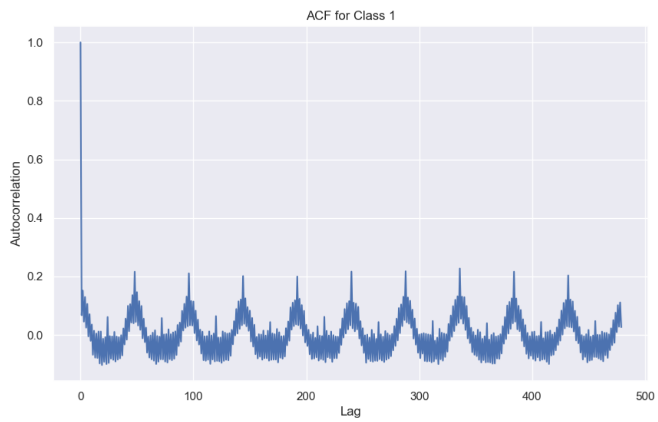

Figure 6

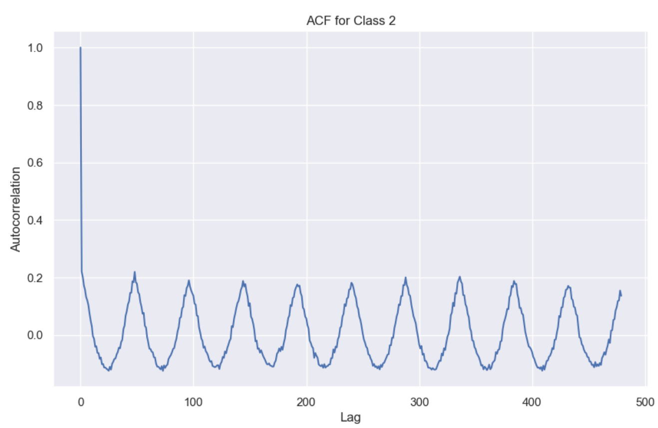

Figure 7

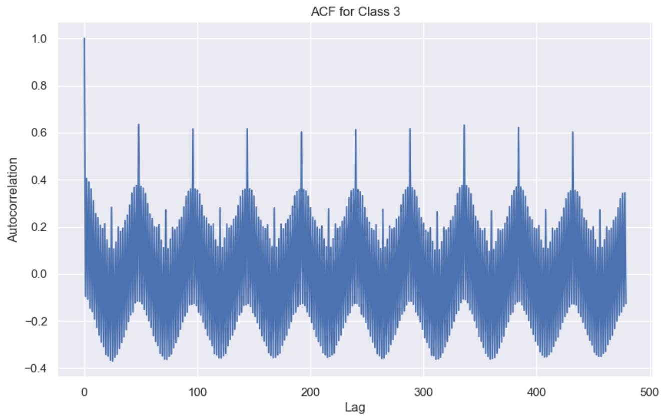

Figure 8

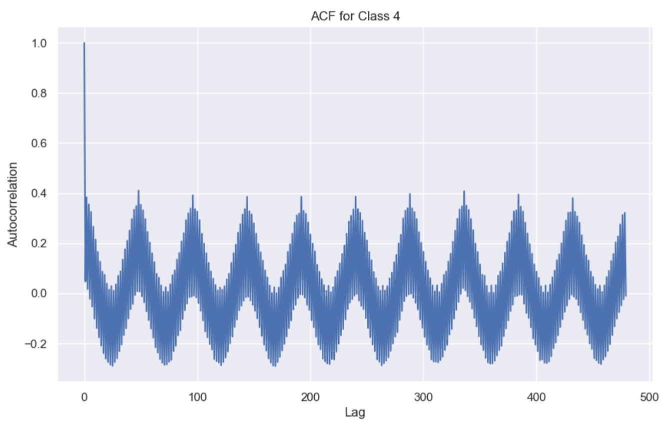

Figure 9

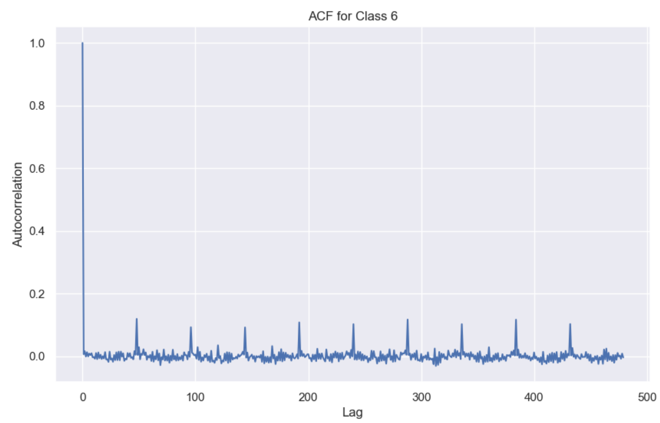

Figure 11

Figure 10