ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

Investigating 3D-STDenseNet for Explainable Spatial Temporal Crime Forecasting

Brian Maguire[1] and Faisal Ghaffar[1] *

Innovation Exchange - IBM Ireland

Abstract. Crime is a well-known social problem faced worldwide. With the availability of large city datasets, the scientific community for predictive policing has switched its focus from people-centric to place-centric, focusing on heterogeneous data points related to a particular geographic region in predicting crimes. Such data-driven techniques identify micro-level regions known as *hotspots* with high crime intensity. In this paper, we adapt the state-of-the-art spatial-temporal prediction model *STDenseNetFus* to predict crime in geographic regions in the presence of external factors such as a region's demographics, seasonal events and weather. We demonstrate that STDenseNet maintains prediction performance compared to previous results [1] on the same dataset despite significantly reduced parameter count. We further extend STDenseNetFus architecture from two-dimensional to three-dimensional convolutions and show that it further improves the prediction results. Finally we investigate the use of the DeepShap model explanation method to provide insights into the important input features effecting the model forecasts.

1 Introduction

The analysis of historical crime events in urban areas in particular, the location and time of crime can play an important role in data analysis for intelligent policing. These spatial and temporal(ST) dimensions of city data can help in profiling regions for a particular type of crime.

ST data has a unique challenge of data sparsity in most domains, particularly crime. Investigating both spatial and temporal dimensions together has the effect of spreading available data across a potentially large number of individual cells/regions. This sparse data provides a weak signal for traditional temporal analysis techniques. In this work we address the above spatial and temporal challenges with Deep Neural Network techniques. Our contributions are as follows:

- We infuse heterogeneous external factors such census data, ethnicity stats and weather data with crime reports to investigate their use in crime forecasting.
- We compare STResNet with STDenseNetFus and demonstrate both perform almost equally.
- We extend STDenseNetFus to use 3D convolutions and observe a modest performance improvement.
- We investigate the use of DeepShap to provide insights into which features from the infused dataset contribute to predictions the most.

*This work has received funding from the European Union's Horizon 2020 research and innovation programme through the COPKIT project, under grant agreement No 786687

2 Related Work

Given the large quantities of crime report data available, Neural Networks (NN) based models make an obvious choice. Crime rate data can be displayed as a grid over a city, with the intensity of each grid representing the rate of crimes within that grid area. These crime rate grid maps can be thought of as low-resolution images, with each grid representing a pixel. Given Convolutional Neural Networks (CNNs) impressive performance when handling image tasks, there is a clear incentive to use CNNs to predict crime rates over a city. The CNN can learn features of how neighbouring grids interact with regards to crime in a similar fashion to learning spatial features within an image. It is expected that the influence that neighbouring grids have on each other is somewhat common across the city, allowing the shared weights of the CNN kernels to efficiently learn these interactions.

In [1] the authors introduce STResNet, a CNN based model which takes crime intensity maps as input and encodes temporal features as additional channels in these images. STResNet report impressive accuracy in predicting crime in Los Angeles (LA).

STDenseNetFus [2] builds on STResNet, also using three CNN sub-networks to model temporal features. In place of using ResNets, STDenseNetFus makes use of Densely Connected convolutions networks (DenseNets). These DenseNets use the same premise as ResNets, passing inputs through shortcut connections. DenseNets take this concept further, appending the output of all presiding layers to the input of the next layer. This makes the network more efficient, allowing the network to reuse features from previous layers. Where STResNet makes use of weather data, STDenseNetFus expands this concept to include several external contextual inputs Such as 1D weather data and 2D geographical context data such as points of interest in a region.

Both STResNet and STDenseNetFus model temporal data by adding multiple time steps as separate image channels, However, research into video classification has found that most of the temporal data is lost in the early layers of a CNN with this setup [3]. The researchers found that the later into a CNN the temporal data is merged the greater the results for video classification. Based on this premise, in [4] the authors develop a 3-dimensional convolutional network model for action classification in videos. 3D CNNs are similar in structure to 2D CNNs but in place of 2D feature kernels they use 3D kernels, moving through the video input with a 3D window of 3x3x3. These kernels maintain the temporal structure of the input video, keeping the order of the frames correct. Crime data can be interpreted as a video sequence, by treating each crime map per time period as a frame, these 3D CNN architectures show promise in predicting crime, similar to predicting the next frame of a video.

3 Experiments

3.1 Dataset Description

We used the LA Crime reports dataset, which includes the location, date and time of incident, the type of crime and the weapon used. The dataset covers from 2010 to May 2019 and includes approximately 1.9 million events. Additional contextual data was added, for spatial context, demographics data was retrieved from the US census bureau. This data was collected in 2010 and includes population, average income, median age, average size of family and other demographics breakdowns. Temporal context was added as weather and event data. Weather data was retrieved from the National Oceanic and Atmospheric Administration for LA over the period of 2010 to May 2019 and a list of US holiday dates was compiled for the same period as temporal events.

3.2 2D DenseNet Model Design

The model design is based on the STDenseNetFus [2] model which has been altered to facilitate crime predictions in place of network demand predictions. The model consists of 3 separate parts, crime density DenseNets, a geographical context DenseNet and a temporal context fully connected layers sub network.

Crime Density DenseNet - Made up of 3 separate DenseNet blocks. Each sub network models a different time range, Daily, weekly and monthly time steps. For each day the model takes in a density map where each grid, or pixels value corresponds to the number of crimes that occurred in that grid over that time period. The output of the crime density DenseNet blocks are fused with a weighted fusion layer that can give weight to one time span over another. **Geographical context DenseNet** - Demographics data are input to a separate DenseNet sub network, the blue box in fig 1. **Temporal Context** - The temporal context data, holidays and weather data is fed into a fully connected network which increases the dimensions to match those of the grid dimensions. Finally the outputs of all sub networks are concatenated together and passed through a final 1x1 convolution layer. The final output represents the expected total crime rates of the next day.

3.3 3D Model Design

The 3D model makes use of 3-dimensional DenseNet blocks, which in turn follow the same design as traditional DenseNets but make use of 3D convolutional layers. As the separate time ranges are not aggregated together as they are in traditional CNNs, there is only a single DenseNet sub network used for all time ranges.

4 Results and Discussions

4.1 Evaluation

To evaluate the forecasts generated by the models we chose the Predictive Accuracy Index(PAI) and Predictive Efficiency Index(PEI). The PAI and PEI have

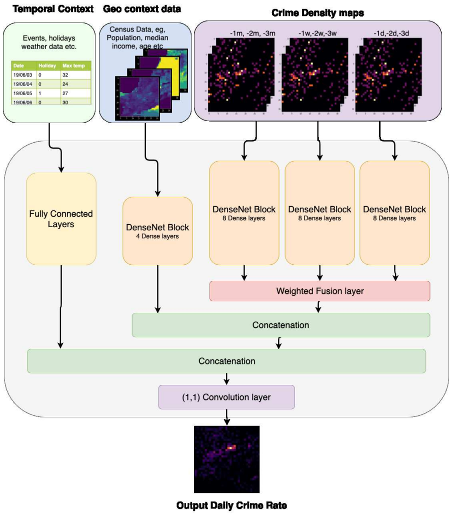

Fig. 1: STDenseNetFus

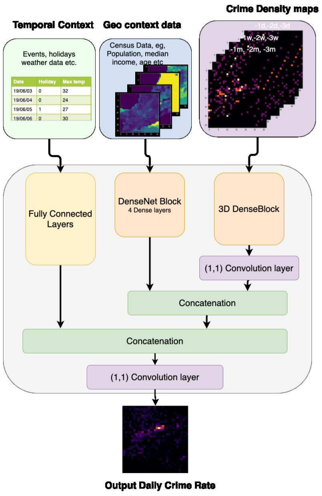

Fig. 2: STDenseNetFus3D

Table 1: Forecast results from 2016-2017

Model	rmse	f1	hit rate	PEI	Params
Naive	1.18	0.75	75.4	73.5	0
STResNet	0.89	0.75	96.1	82.7	1,794,117
STDenseNetFus	0.88	0.75	95.2	82.6	887,740
STDenseNetFus No Geo	0.88	0.77	95.9	82.5	794,565
STDenseNetFus3D	**0.86**	**0.79**	**96.3**	**83.1**	924,394
STDenseNetFus3D No geo	0.87	0.78	96.2	83.1	830,851

been used previously to determine the effectiveness of forecasts when it comes to choosing patrol areas for law enforcement agencies(LEA), for example in a 2017 competition held by the national institute of justice into predictive policing [1].

For all given PEI scores in table 1, the maximum area allowed to be predicted is set to 1% of the the total area, higher areas are assumed too large for useful patrol planning.

4.2 Results

Table 1 shows results comparing four models. **Naive** which takes the last known value for each grid as the current prediction. **STResNet**, defined in [1], implementation taken from GitHub [2], changes were made to predict over 24 hours rather than the papers original 1 hour predictions. **STDenseNetFus**, using 2D convolutions with geographical and spatial context input data and finally **STDenseNet-**

[1] https://nij.gov/funding/Pages/fy16-crime-forecasting-challenge-document.aspx

[2] https://github.com/lucktroy/DeepST/tree/master/scripts

Fus3D. For both STDenseNet models one is trained with Geographical context data in the form of Census data, for models labeled "No Geo" this data was not included. All models were trained on LA crime report data using June 2010 to December 2016 for training data and December 2016 to December 2017 for validation data. The city was divided into 32x32 uniform grids enclosing the area from Latitude, Longitude (33.79, -118.7) to (34.38, -118.11). One-step predictions were made for each day in the training set. Table 1 shows the average metrics over this period. The results show a small improvement in all metrics for STDenseNet3D model. This finding is consistent with work done in [3] which showed that keeping the temporal ordering of video clips further into a model produced better results for the task of action recognition. STDenseNetFus shows very similar results to STResNet, in particular in PEI forecasting efficiency. This is despite having 51% of the parameters, shown in table 1 under 'Params'. This is consistent with the initial findings of using DenseNets on image recognition tasks in [5] which found DenseNet models could achieve similar results with reduced parameter counts. Adding Geographical Context data appears to have no significant effect on model performance. This may be due to the fact that the census data is static and out of date, having been collected in 2010. A more recent and dynamic geographical dataset may produce better results.

4.3 Forecast Explanations

Forecasting bias is a considerable concern of any predictive policing strategy. Black box models, such as deep learning architectures, provide little insight into why one area is considered a hotspot over another. As a possible mitigation to this, we investigated the use of the DeepShap[3] model explainability technique described in [6]. In brief, DeepShap can provide a weight of how much each feature in the input was responsible for a given output. It does this by calculating the difference in activation values for the given input compared to a baseline input. This difference is then back propagated through the model to arrive at an estimation for the shapely values for each input feature. An LEA can use this information to determine which input datatype is causing an area to be a hotspot and make a judgment call as whether to use the forecast or not. To make use of DeepShap we reconfigure the model as a regression problem, with the crime intensity of a single grid set as the output for the overall model. The DeepShap method requires a set of baseline inputs which represents the default input to compare against, for these experiments we chose a baseline of zero for all crime intensities.

We generated explanations for a number of dates and verified the explanations produced expected results. As an example, on Grid 440 [4] on the 6th of February 2019, the total number of crimes (the target variable) in a region dominates the forecast weight with 83% and 16% positive and negative effects respectively. Temporal context values make small contributions to the forecast, such as the maximum temperature, having a 0.03% positive effect on the forecast. Addi-

[3] https://github.com/slundberg/shap

[4] block Latitude and Longitude (34.050825, -118.2441625), (34.032387, -118.2626)

tionally we viewed the contribution of individual input grids and viewed how the shapely weights are spatially distributed. As expected the total crime intensity of the target grid itself had the highest weight of 10.05% along with a number of key neighbouring grids, such as grid number 471 which had a 2.9% positive impact on the forecasted value. This information may provide LEAs insights into how neighbouring areas effect each others at particular times and circumstances.

5 Conclusion and Future Work

In these experiments we have found that the DenseNet based STDenseNet model produces comparable results to the ResNet based STResNet with significantly reduced parameter counts. We have found that the addition of geographical context data, in the form of census information, did not produce appreciable gains in the performance. Using a 3D DenseNet design did provide modest performance improvements in PEI scores, likely due to the ability to keep temporal information intact further into the model. Additionally we explored the DeepShap model explanation method as a possible remedy to opaque forecasts being used by LEAs. We confirmed from shapely weight values that census data had a minimal affect on forecasts. We plan to explore the impact of other geographic contexts such as point of interest in future work along with getting direct LEA feedback on Deepshap values.

References

[1] Bao Wang, Duo Zhang, Duanhao Zhang, P Jeffery Brantingham, and Andrea L Bertozzi. Deep learning for real time crime forecasting. *arXiv preprint arXiv:1707.03340*, 2017.

[2] Haytham Assem, Bora Caglayan, Teodora Sandra Buda, and Declan O'Sullivan. Stdennetfus: A new deep learning approach for network demand prediction. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 222–237. Springer, 2018.

[3] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1725–1732, 2014.

[4] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. Learning spatiotemporal features with 3d convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 4489–4497, 2015.

[5] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[6] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc., 2017.