Crime Prediction

Prabhat Kanaujia
IIT Delhi
cs5160789@iitd.ac.in

Sunil Kumar
IIT Delhi
cs1160314@iitd.ac.in

1 Abstract

In modern times, crime is the biggest problem most countries are facing. It has become an inevitable part of the modern society and many big cities are plagued with this problem. According to the annual crime report, over half a million children and youth aged 10-24 years suffered non-fatal physical assault injuries which are related to stabbings and gunshots. Also, recorded crimes have increased from 2300 to 3000 for every 100,000 people during the period of 1980 to 2000. These figures continue to increase in the present-day societies as well.

In this project, we have worked on the problem of crime prediction, implementing a neural framework named DeepCrime (Huang et al., 2018), with certain modifications to th mentioned architecture and obtaining significant improvements over their results.
We have trained a model to predict the occurrence of a particular category of crime in a particular locality on a particular day, for 4 categories of crime and 77 localities, as detailed later.

After fine-tuning the hyper-parameters, extensive experimentation showed that our model's performance comes very close to the current state-of-the-art model, MiST (Huang et al., 2019).

2 Introduction

In this project, we have worked on the problem of crime prediction on a particular date, in a particular locality and of a particular kind, given the historic data of crime occurrences. Many existing frameworks for this task utilize demographic data. But this kind of data usually fails to capture the dynamic patterns of crime occurrence over a period of time because these demographic distributions are usually static and slow to change over time.
Hence, we have tried to address various factors important in the task of crime modelling and crime prediction.

1. **Temporal Dynamics**: to model the changing patterns of crime occurrences over time

2. **Category Inter-dependencies**: to model the causal relations between the occurrence of different types of crimes

3. **Correlation with surrounding events**: Certain events, such as a noise or a safety complaint can have correlation with crime occurrences. We have tried to capture this aspect of the crimes as well

3 Related Work

Related work in this field includes work pre-dating the DeepCrime (Huang et al., 2018) framework. It includes work in urban sensing applications, that can be translated to some extent to the field of crime prediction. As mentioned in (Huang et al., 2018), (Lian et al., 2017) studies the problem of restaurant survival prediction, based on geographical information and user-mobility.

Crime rate inference and crime hot-spot detection has been studied earlier by (Zhang et al., 2018) and (Wang et al., 2016). Other papers have studied the problem as well, using statistical as well as data mining approaches.

Our work is based on (Huang et al., 2018), which addresses crime prediction by focusing on modelling time-stamped data and also capturing various important inter-dependencies and correlations between the different categories crimes as well as between crimes and other surrounding events, such as public complaints on noise, safety, etc.

4 Methodology

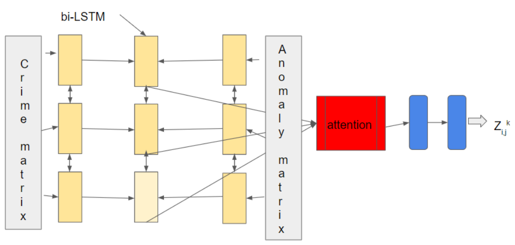


Figure 1: Model BI-LSTM

Model takes two inputs(one for each bi-LSTM) namely crime matrix and 311 matrix, more of which is explained later in the next section. Both are feed to Bi-LSTMs, whose outputs are combined(based on weights, which themselves are learned, to assign importance to these inputs) and passed to another Bi-LSTM layer.
The output from this Bi-LSTM is fed to an attention layer, which basically captures and assigns importance to the features of this output.
This is fed to a 2-layer MLP, which predicts the occurrence of crime class j, for region i at day T+1, given the historical data for the previous T days. The ground truth here is the actual crime occurence data for that day. The loss function is sigmoid cross-entropy with logits.

Loss, $L = z * -log(sigmoid(x)) + (1 - z) * -log(1 - sigmoid(x))$

where, x = logits, z = labels

5 Datasets

5.1 Description

Datasets used in the training and evaluation of our model were taken from the NYC OpenData collection, which is a collection of data collected in the city of New York over the recent years, spanning across different categories and departments.
For our work, we have taken the following datasets:

1. Crime Occurrence
2. Public 311 Complaints
3. Points of Interest: geographical location of important public places
4. Precincts: geographical divisions of the city into 77 areas

5.2 Collection

The data is openly available for download from the NYC OpenData website (https://opendata.cityofnewyork.us/data/). We were unable to download the most recent data for Public 311 complaints, for the year 2018, due to some issues with the downloaded file being corrupted. For 311 complaints, accessible data was for the years 2006-09.
Thus, our final model was trained on Crime records and 311 Public Complaints records for the year 2006. Precincts geographical divisions and Points of Interest were also downloaded from the same website.

5.3 Modification

After the data collection, a number of modifications were done to make the data relevant and usable for the neural network architecture.

1. Points of Interest(**POI**) contained geo-locations of around 24000 important, named locations around the city (e.g. Key residential properties, educational facilities, etc.). We extracted, from this list, only those POIs that were updated to the list in or before 2008 were considered, rest were discarded.
   This is because our model is trained on data from 2006 and to include the points existing in that year, we had to keep the threshold at 2008, as the last update before 2008 was in 1900s.

2. Crime records were extracted only for those categories which had a significant number of crime occurrences. These were:
   (a) **Robbery**: 43747
   (b) **Burglary**: 22785
   (c) **Felony Assault**: 22747
   (d) **Grand Larceny**: 16865

3. Similarly, 311 reports were also considered for only the following 4 categories, which were thought to be more correlated with crime occurrences:
   (a) **Blocked Driveway**: 51939
   (b) **Building/Use**: 32235
   (c) **Noise**: 43542
   (d) **Safety**: 1415

5.4 Input Matrices

The crime records and 311 complaints records are converted to matrices to be fed as inputs to the neural architecture.

1. C^{JXK}: for crime records
2. A^{JXL}: for 311 records

, where J is 77, i.e. the number of precincts
, K is 4, i.e. the number of crime categories considered
, L is 4, i.e. the number of 311 complaint categories considered

Since, we take the entire year's data as input, the final input vectors are 365-length arrays of above matrices, for crime records and 311 records.

6 Experimental setup

6.1 Data

The data for the year 2006 from NYC OpenData, along with the transformations mentioned above, has been used for training and evaluation.

6.2 Parameters

1. Optimizer: Adam optimiser
2. Learning Rate: 1e-4
3. Batch Size: 4
4. Context window: 10
5. MLP layers: 2
6. loss = sigmoid cross entropy with logits
7. Hidden Size: 64
8. Attention Size: 64

6.3 Training task

Given crime and 311 records for a period of T days, predict the occurrence/non-occurence of crime category j, in region i, at day T+1. This is done for all crime classes(total 4), for all regions(total 77).
Therefore, a total of 308 predictions for each step.

6.4 Evaluation

Metrics used for evaluation are:

1. Macro-F1 score
2. Micro-F1 score

The results have also been compared to those in (Huang et al., 2018) and (Huang et al., 2019), performing better than the former and slightly lower than the latter.

7 Results

	precision	recall	f1-score	support
robbery	0.62	0.83	0.71	2504
burglary	0.61	0.80	0.69	2399
felony	0.49	0.60	0.54	1607
grand larceny	0.71	0.97	0.82	3129
micro avg	0.63	0.83	0.71	9639
macro avg	0.60	0.80	0.69	9639
weighted avg	0.62	0.83	0.71	9639
samples avg	0.59	0.76	0.63	9639

Figure 2: predicted map of crime

With respect to the training task, mentioned in the experimental setup in the above section, these (figure 2) gave a score of less than 0.05 on running a t-test for best test score model to check if those results were true or just a statistical fluke, which showed that the predictions are statistically significant.

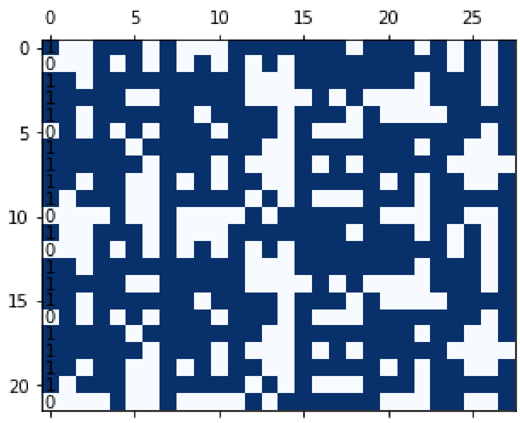


Figure 3: ground truth

So it's evident from F1 score graphs that Bi-LSTMs are more stable and consistent during the training phase, whereas LSTMs oscillate a lot during the training phase.

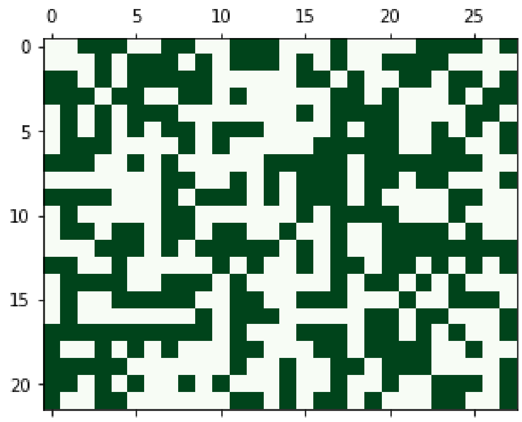


Figure 4: overlap

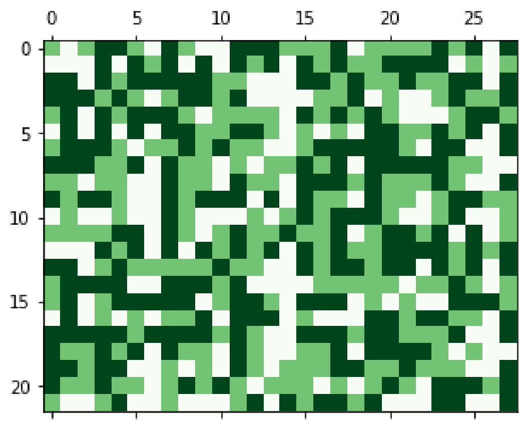


Figure 5: Results

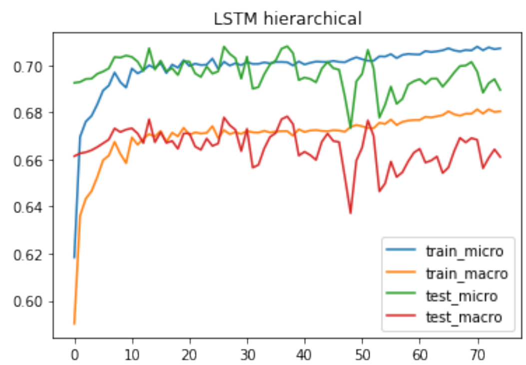


Figure 6: LSTM hierarchal with attention

8 Observations and Key Learnings

Some of the interesting observations and learning from this project can be listed down as:

1. In some of our experiments, we trained the model on concatenated data for 4 consecutive years. However, contrary to our expectations, this led to a decrease in F1 scores.
   A possible explanation can be that over such a long period there are a lot of trends of crime occurrence that keep changing over time. Trying to model all these trends at once require a higher-order dependence on time, whereas LSTMs and other recurrent networks are able to capture only linear-time variations based on our experimental setup.

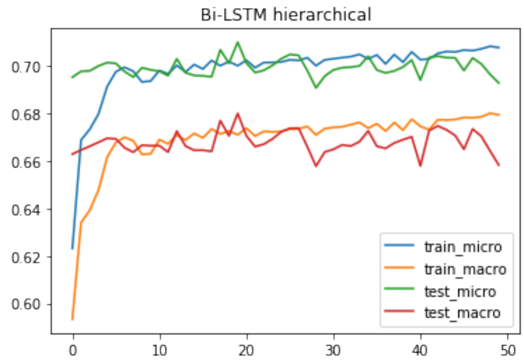
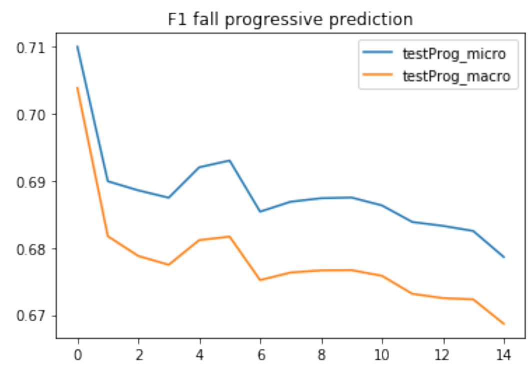


Figure 7: Bi-LSTM hierarchal with attention

2. We fed the weighted results from the Bi-LSTM trained on crime reports and that on 311 reports to another Bi-LSTM, whose output was used for further steps. Analysing these weights showed that:
   (a) Forward direction(in time) of crime occurrences carried more than double the importance of the data in backward direction(in time). This led us to believe that even an LSTM might work in this case. And it did, achieving a F1-score that was 0.001 less than that with Bi-LSTMs.
   (b) For 311 reports, forward and backward directions seemed to be of equal importance



3. As discussed earlier, when trying to predict crimes for a certain number of days ahead

of the T-day window, by recursively predicting for (T+i) day and bootstrapping, there is a visible decrease in performance.// This can be attributed to the fact that the uncertainty gets compounded over a series of predictions as we are using predicted crime for (T+j) day as well, to predict the crime occurence on (T+j+1) day and so on.

References

Chao Huang, Chuxu Zhang, Jiashu Zhao, Xian Wu, Dawei Yin, and Nitesh Chawla. 2019. Mist: A multiview and multimodal spatial-temporal learning framework for citywide abnormal event forecasting. In *The World Wide Web Conference*, WWW '19, pages 717–728, New York, NY, USA. ACM.

Chao Huang, Junbo Zhang, Yu Zheng, and Nitesh V. Chawla. 2018. Deepcrime: Attentive hierarchical recurrent networks for crime prediction. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, pages 1423–1432, New York, NY, USA. ACM.

Jianxun Lian, Fuzheng Zhang, Xing Xie, and Guangzhong Sun. 2017. Restaurant survival analysis with heterogeneous information. In *Proceedings of the 26th International Conference on World Wide Web Companion*, WWW '17 Companion, pages 993–1002, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

Hongjian Wang, Daniel Kifer, Corina Graif, and Zhenhui Li. 2016. Crime rate inference with big data. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 635–644, New York, NY, USA. ACM.

Chuxu Zhang, Chao Huang, Lu Yu, Xiangliang Zhang, and Nitesh Chawla. 2018. Camel: Content-aware and meta-path augmented metric learning for author identification. pages 709–718.