Crime Prediction-Project Proposal

Team 30

Objectives:

Our aim is to build a tool for crime prediction that can predict the likelihood and category of crime in different geographical locations in the city of Chicago as well as identify the various factors that contribute to it.

How is it done today; what are the limits of current practice?

Historically the police have used their past experience and intuition to predict broad locations and times of crimes. While the police are now using more sophisticated technology like mining historical criminal records, individual criminal histories, modern equipment, the approach is same in its essence; it focuses on reactive measures to curb crime, not on its prevention.

In the past few years, the focus is shifting to preventive measures, mainly crime prediction by highlighting what factors lead to or facilitate a crime. Much work is happening in this field that we are using to inform our approach.

Almanie, Mirza and Lor (2015) use city crime history dataset that includes location, type of crime, date, time and proposes finding temporal and spatial crime hotspots using decision trees and Naive Bayesian classifiers in order to predict potential crime types. While Corman and Mocan(2005) similarly proposed prediction using spatial analysis using k-nearest neighbour approach. Although it suggests overcoming limitation decision trees as models but its limited in terms of source data and is purely based on historical crime records.

This approach also doesn't use other local level metrics specific to a small area like demographic data, vacant/abandoned buildings etc to get an integrated view of the factors that can affect crime rates and types of crime in the area.

Wallace and Schalliol (2015) measured the relationship between abandoned buildings and social disorder in a community. Their assumptions were based on the Broken Windows theory. Corman and Mocan (2005) investigates the impact of economic, demographic conditions and increased felony arrests on crime rates in New York City. It also suggests that the Broken Windows Theory has validity in the case of robbery and motor vehicle theft.

Hauk et al (2002) have used the concept space algorithm on crime data to detect abnormal activities. Once these activities are identified it may be possible to predict the next occurrence of such activity. Oatley et al (2004) have used algorithms to match and link burglary crimes together into a crime series for predicting where the next crime in that series will occur.

Wang et al (2012) perform automatic semantic analysis on Twitter Posts along with dimensionality reduction via latent Dirichlet allocation and prediction via linear modelling for forecasting hit-and-run crimes.

Dash et al (2018) fuse the spatiotemporal information about criminal records like schools, libraries, police stations, and 311 service calls using network analytic approach to identify observations that improve the quality of prediction. They use a regression-based approach by experimenting with polynomial, auto-regressive and support vector regression methods.

According to Johnson et al (1997), deprived regions are more affected by repeat victimization and in regions with high crime event rate, and in support Johnson et al. (1997) and Townsley et al. (2000) designated as hot-spots. Several researchers like Winkel (1991); Gill and Matthews, (1994); Ericsson, (1995) Johnson and Bowers, (2004) in their research work have interviewed offenders and established that they target the same property repeatedly suggests that burglary offences cluster in space and time.

The correlation of large-scale "Point-Of-Interest" data and "taxi flow data that serve as hyperlinks" in the city of Chicago, IL is explored by Wang and Kifer (2016) to be harnessed in crime rate estimation and reduce prediction error. Wang and Kifer 2016 gives us important insight to focus on seemingly trivial information to realize novel factors that might contribute to crime.

What's new in your approach; why it will be successful?

We aim to overcome the limitations of the above approach by incorporating other local level metrics including census data (education, unemployment levels, etc), Vacant/Abandoned buildings data, 311 requests in addition to historical crime records that contain the type, location, date and time of the crime.

As suggested by Wallace and Schalliol(2015), Corman and Mocan (2005) in their works, we believe that our model could be more accurate in its predictions of crime trends because it has a more integrated view of the local environment.

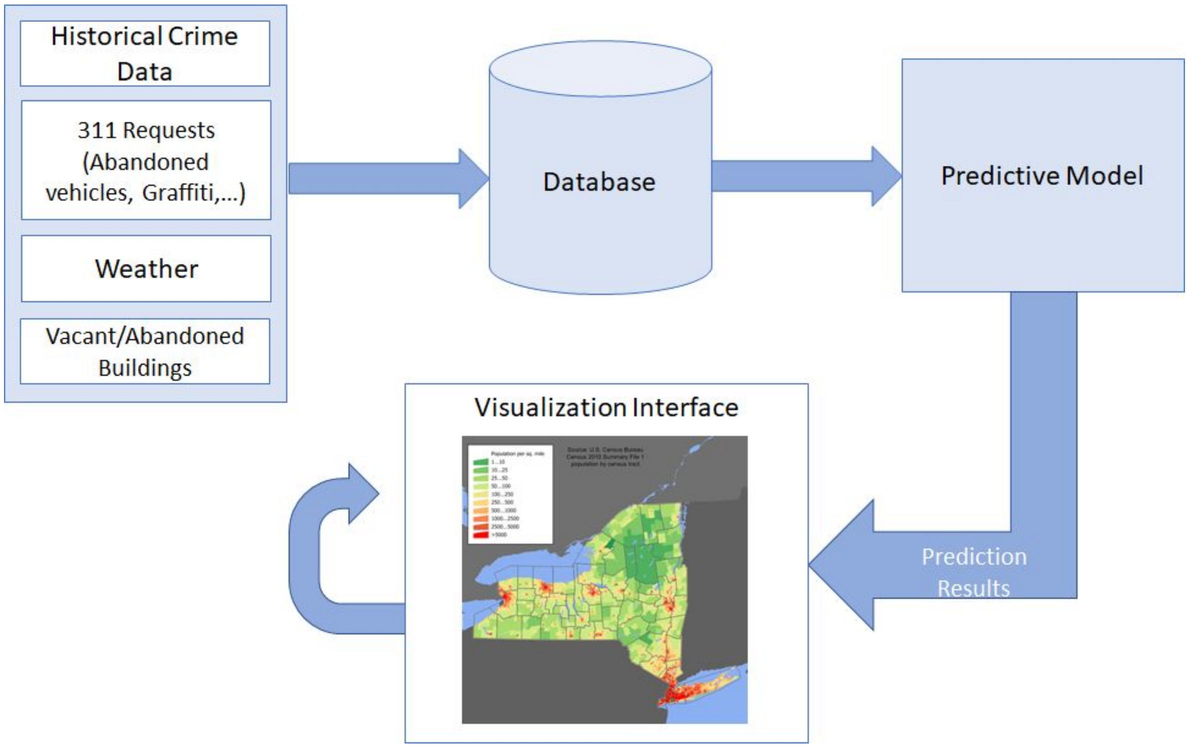

Fig 1: Overview of the proposed crime prediction tool

Who cares? Difference and Impact:

Law Enforcement can use this tool to reduce crime incidents by better planning and allocation of scarce police resources.

Risks and Payoffs:

Payoffs include improved prediction accuracy, which can lead to efficient patrolling and consequent reduction in crime.

The most important risk is that additional variables may not improve the accuracy of the predictions. We also risk bias in predictions based on demographic factors and historical data.

Cost:

Server time and developer time.

Plan of Action:

We plan to do the project in 6 weeks:

Week 1 (Oct 15 - Oct 21): Collect data including historical crime and other local level metrics from the Chicago Data Portal.

Week 2 (Oct 22 - Oct 28): Clean and merge the data

Week 3 (Oct 29 - Nov 4):

We will explore multiple approaches including KDE and Decision Trees to build the model and select the best approach that suits the requirement. Train a base model to predict the likelihood of each type of crime in a given area based only on historical crime data.
Week 4 (Nov 5 - Nov 11):
Design and build interactive visualizations and a web app for result exploration.
Week 5 (Nov 12 - Nov 18): Iteratively add other factors to base model to improve prediction accuracy.
Week 6 (Nov 19 - Nov 25): Find the most critical local features that contribute to crime.

Group Member wise work distribution:
Each participant contributed to the literature survey and are actively involved in data exploration, collection and cleaning.
Kuhu & Preeti - Investigating various visualization frameworks and techniques and building the front end.
Simha & Vatsal - Analysing different data modelling techniques. Building the backend for the web application.

References:

1. Almanie, T., Mirza, R. and Lor, E. (2015). Crime Prediction Based on Crime Types and Using Spatial and Temporal Criminal Hotspots. *International Journal of Data Mining & Knowledge Management Process*, 5(4), pp.01-19.
2. ToppiReddy, H., Saini, B. and Mahajan, G. (2018). Crime Prediction & Monitoring Framework Based on Spatial Analysis. *Procedia Computer Science*, 132, pp.696-705.
3. Wallace, D. and Schalliol, D. (2015). Testing the temporal nature of social disorder through abandoned buildings and interstitial spaces. *Social Science Research*, 54, pp.177-194.
4. Corman, H. and Mocan, N. (2005). Carrots, Sticks, and Broken Windows. *The Journal of Law and Economics*, 48(1), pp.235-266.
5. Hauk, R.V., Atabaksh, H., Ongvasith, P., Gupta, H., and Chen, H. (2002) Using Coplink to analyze criminal justice data, IEEE Computer, 35(3), pp. 30-37. Home Office (2000) Review of crime statistics: A discussion document.
6. Oatley, G.C., Zeleznikow, J., and Ewart, B.W., (2004), Matching and Predicting Crimes. In: Macintosh, A., Ellis, R. and Allen, T. (eds.), Applications and Innovations in Intelligent Systems XII. Proceedings of AI2004, The Twenty-fourth SGAI International Conference on Knowledge-Based Systems and Applications of Artificial Intelligence, Springer: 19-32. ISBN 1-85233-908-X
7. Dash, S. K., Safro, I., & Srinivasamurthy, R. S. (2018). Spatial-temporal prediction of crimes using network analytic approach. *arXiv preprint arXiv:1808.06241*.
8. Winkel, F.W. (1991) Police, Victims and Crime Prevention: Some Research-Based Recommendations and Victim-Oriented Interventions. British Journal of Criminology 31, 250-265.
9. Johnson, S.H. and Bowers, and Hirschfield, F.G. (1997) New Insights into the Spatial

and Temporal Distribution of Repeat Victimisation. British Journal of Criminology, 37(2), 224-241.

10. Trickett, A., Osborn D. R., and Ellingworth, D. (1995) Property crime victimization: The roles of individuals and area influences. International review of victimology 3, 273-95.
11. Johnson, S.H. and Bowers, K.J. (2004) The Burglary as Clue to the Future: The Beginnings of Prospective Hot-Spotting, European Journal of Criminology, Vol 1 (2): 237-255: 1477-3708.
12. 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).