ADTA 5900 Module 2: Industry/Scholarly Review and References

Crime Patterns and Predictive Modelling in Chicago City

Literature Review

Table of Contents

Introduction

Crime data analytics has become one of the most important practices that help contemporary police organizations to learn from the past crime patterns and, therefore, prevent them in the future (Ourania, Alina, Araujo, & Leitner, 2020). Using statistical analysis, geographic information systems (GIS), and machine learning algorithms, analysts can highlight crime patterns and even make predictions about the likelihood of crime in certain spatial and temporal dimensions. This approach – often referred to as predictive policing – is a shift from the traditional reactive policing to data-driven prevention with the potential of increasing the effectiveness of resource allocation and improving safety (Mugari & Emeka, 2021). However, the development of big data analytics in policing has also met with a significant controversy regarding its effectiveness and fairness. In this review, we review current developments in crime data analytics research with an emphasis on new developments in predictive modelling and spatial-temporal analysis, the current state of practice in law enforcement, existing problems (prejudice and transparency issues) and new possibilities for change. The analysis includes academic articles published in the last ten years that can help to understand the ways of crime prediction and their real consequences.

Advancements in Crime Data Analytics

Recent literature reveals, new tools for analyzing and forecasting crime have been developed and tested, and Spatio-temporal modeling has emerged as central - analysts have realized that crime is not randomly occurring but rather occurs in the cluster in space and time (i.e., ‘hot spots’) based on other social and environmental factors (Yingjie & Ning, 2023). (Butt et al., 2020) performed a systematic review of the literature on crime hotspot detection and prediction techniques and observed that the current state of the art incorporates multiple strategies. The methods generally employed are clustering algorithms for defining high-crime regions, time series analysis for the study of trends over time, and increasingly, deep learning models for the analysis of complex patterns in the data. For instance, different types of neural networks such as CNNs and LSTMs have been used in crime analysis and have been found to be better than some of the conventional methods in the prediction of crime rates or crime locations. This suggests moving forward from the rudimentary regression or hotspot mapping to the advanced AI-based forecasting.

Another important advancement is the use of multiple types of data sources. Whereas traditional crime prediction models were based primarily on historical crime incident data, newer approaches enhance this with other factors such as demographic information, land use or even social media post content to enhance the predictive capability (Mohler et al., 2015). (Matthew, 2014) demonstrated that combined analysis of geotagged Twitter data and crime records provided a modest enhancement in the prediction of specific crimes in Chicago (for instance, thefts, assaults, and stalking) by providing indicators of normal activity in specific areas that are likely to be associated with crime.

The increasing use of such 'big data' (including socio-economic indicators and real-time feeds) is in line with broader smart city initiatives, in crime analytics. Multi-scale prediction model: For instance, (Yingjie & Ning, 2023) review crime prediction for short-term, medium-term and long-term horizons, and micro-level (street), meso-level (district) and macro-level (city) spatial scales. Analysts can refine predictions, from thinking next week's burglary hotspots in a neighborhood to forecasting yearly crime trends citywide, by tailoring models to right temporal and geographic scale.

Applications in Predictive Policing

With these analytical improvements, predictive policing applications have been developed and used in practice by police departments globally. The most frequent practice is the use of place-based predictive policing where algorithms are used to pinpoint the locations most likely to experience crime in order to deploy police presence in the areas. (Mugari & Emeka, 2021)Dozens of cities have tried or adopted systems like PredPol (now Geolitica), which are based on near-repeat and hotspot theories, or Risk Terrain Modeling, which includes environmental risk contributors. These tools are the real application of crime data analytics: the models outputs areas of predicted crime on a map, for example translated into daily operational decisions such as where to deploy officers. A few studies have been conducted to assess their effectiveness. As per Mohler et al. (2015) a randomized controlled field trial was conducted in Los Angeles and an algorithmic predictive model (an epidemic inspired point process) was used to inform patrols which resulted in a 7.4% reduction in crime volume while outperforming traditional hotspot policing based on human analysis. Not all implementations show such clear benefits, however. To some jurisdictions predictive policing is credited with crime declines, but in others it results in no significant change at all, which suggests that context and execution are key. Meijer and Wessels (2019) argue that although the tool is widely adopted, there is a gap between the perceived potential of predictive policing and empirical reality: some studies show small improvements in crime, other show no effect, and in general there is hardly evidence for the massive benefits promised. Moreover, a recent systematic review by Lee et al. (2024) revealed that only 6 out of 161 studies on data-driven predictive policing were strong in supporting evidence. In fact, Meijer and Wessels argue that the current momentum of predictive policing is based more on ‘compelling stories and examples’ than on evidence, and they demand more independent assessments to create real evidence.

It is important to recognize that crime data analytics is not solely about predicting where crime will occur – some initiatives are based on people (e.g. identifying people most likely to reoffend or become victims). However, this review focuses on spatial analytics, person-based predictive systems (such as offender risk scores or potential shooters, the so-called 'heat lists') have also been discussed, stimulating similar levels of optimism and concern in the literature (Mugari & Emeka, 2021). In total, the current situation with crime analytics in practice can be characterized as somewhat progressive: the police have started using these tools and incorporating them into their plans, but the studies of the effectiveness of these efforts have been rather ambiguous and more work is needed. Nevertheless, the literature reviewed agrees that data-assisted policing holds the potential to improve the effectiveness of the policing institution in the achievement of its goals if properly applied (Mugari & Emeka, 2021). The problems that have inhibited the effectiveness of these tools up to this point must be solved.

Key Issues and Challenges

Despite the promise of crime data analytics however, scholars argue that critical challenges must be addressed for reliability and public trust to improve.

Bias and Fairness: One of the most controversial topics is the possibility of the analytical models to reinforce or even increase the existing biases in crime data (Alikhademi et al., 2022). It is possible that historical crime data contains information about the law enforcement activities rather than the crime rates – for example, Lum and Isaac (2016) show that more intense police presence results in more arrests and reported incidents which in turn lead to the cycle (Alikhademi et al., 2022). As a result, if such data is used in the model without any correction, the police will keep on visiting the same areas which are already over policed. Lum and Isaac argue that the use of 'bias free' algorithmic tools on biased data just provides a way of implementing biased police practices, while appearing to be objective (Alikhademi et al., 2022). This concern is also raised frequently in the literature: researchers argue that there is a risk of disparate impact, where people of color may be disproportionately stopped because of the built-in bias in the data (Alikhademi et al., 2022). Therefore, the problem of fairness becomes one of the main issues. However, not all the evidence available indicates that predictive policing is necessarily a racist approach – Brantingham et al. (2018) in the case of Los Angeles did not find important variations in the frequency of arrests by race between areas applying the predictive policing approach and control areas and also failed to establish overall higher arrest rates (Brantingham et al., 2018). This means that bias is not an inevitable result of using these tools; nevertheless, the issue of bias and the need for disclosure continue to be critical in order to gain the community's trust.

Evaluation and Validation: There is no framework agreed upon to evaluate crime prediction models until now, which results in the inconsistent reporting of the performance. The studies use accuracy, PAI, F1-score, etc. as metrics and train/test validation approaches (Kounadi et al., 2020). It is impossible to compare the results or know which method is the 'best' one to use due to the heterogeneity of the methods employed. Kounadi et al. (2020) point out that many publications do not provide sufficient information about the experimental process or employ unified performance criteria (Kounadi et al., 2020). As such, the field may be overclaiming successes and underrecognizing possible failures. This challenge is related to the need for more systematic empirical research: more randomized trials or controlled studies would help to provide more convincing evidence concerning what crime analytics can and cannot achieve (Meijer & Wessels, 2019).

Opportunities for improvement and Future Direction

The literature has for the most part outlined how to improve crime data analytics so that the advantages are not coupled with the present risks. First, an area of improvement is in the technological aspect: as researchers advise, applying new AI approaches to solve existing problems. For example, (Yingjie & Ning, 2023) proposed using transformer-based learning for better handling of sparse and complex data and integrating reinforcement learning for the dynamic modification of the policing strategies based on the model outputs (Yingjie & Ning, 2023). These approaches could enhance the precision and adaptivity of the predictions and, therefore, move away from the use of static hotspot lists to more dynamic systems. Likewise, the application of explainable AI is also essential – the creation of models that not only predict crime but also explain their predictions (e.g., explaining that an area is coded as high risk based on recent thefts combined with high foot traffic) will increase the credibility and support the decision-making process (Yingjie & Ning, 2023).

Standardizing evaluation practices is another clear opportunity as well. The field would also benefit from agreed upon benchmarks, data sharing, and perhaps open competitions (similar to the recent NIJ Crime Forecasting Challenge) that would accelerate innovation while rigorously comparing approaches (Butt et al., 2020). For instance, this entails making available open datasets and results so that independent researchers can validate and improve upon agency models. Such transparency can meet scientific and public accountability needs.

Crucially, several scholars have suggested that fairness should be a part of the next generation of crime analytics. The field of algorithmic fairness is still developing; for instance, there are current efforts to develop methods of ‘debiasing’ the training data or to set equity constraints on model outputs (Alikhademi et al., 2022) and these should be extended to predictive policing. In addition, social scientists, legal experts, and community stakeholders should be involved in the design and deployment of these tools to balance ethical considerations with technical performance. In practice, this might be bias audits of crime prediction software, using alternative data that can help counteract policing bias (e.g., community reports), or feedback loops where officers and analysts review algorithmic suggestions instead of following them blindly.

In the end, it is crucial to emphasize that future work should provide more rigorous tests of the methodologies employed. As various reviews have pointed out, there is a lack of independent field experiments and longitudinal studies to determine the real impact of crime data analytics on crime rates and policing outcomes (Meijer and Wessels, 2019). These studies should try to determine not only if crime decreases but also any negative impact on trust in the community or the police. The models should be validated empirically in various settings (other cities, crimes, and policing strategies) to enhance the knowledge base from promising theory to practical application. This evidence-based approach will allow the discussion to move from the question of whether predictive analytics is effective to the question of when and how it can be effective and how it can be used appropriately.

Conclusion

In summary, crime data analytics is thus seen as a new and evolving field within criminology and policing. Current best practice involves complex statistical tools such as the machine learning and space-time models that have been applied to the detection of patterns in crime and to some extent in the identification of locations and times of high crime risk. In predictive policing programs, these methods have been found to have the potential of assisting the police in their efforts to be more proactive and efficient (Mohler et al., 2015). However, as the existing literature suggests, we are still some ways from realizing that potential. The technology has, to some extent, outgrown the old patrol strategies, but research on validation and practical guidance is still developing (Meijer & Wessels, 2019). There are still issues of bias, understandability, and standardization that need to be addressed before crime analytics tools can be considered accurate, objective, and credible.

Encouragingly, the literature points to clear opportunities to improve these systems. They suggest that by using more sophisticated data science methods (and combining them with the knowledge of domain from criminology) and by paying more attention to the ethical issues the next generation of crime analytics can be more accurate and equitable. As law enforcement agencies continue to experiment with data driven approaches, researchers call for strong evidence-based framework – one that rigorously tests predictive tools and guides their use. If properly developed and supervised, crime data analytics can offer practical findings to improve the fight against crime and improve public safety (Yingjie & Ning, 2023). It is, therefore, important that the next few years see a continuation of the current academic work into practice so that the integration of big data with policing is beneficial for all concerned.

Reference

- Kounadi, O., Ristea, A., Araujo, A., Jr, & Leitner, M. (2020). A systematic review on spatial crime forecasting. Crime science, 9(1), 7. https://doi.org/10.1186/s40163-020-00116-7
- Mugari, Ishmael, and Emeka E. Obioha. (2021). Predictive Policing and Crime Control in The United States of America and Europe: Trends in a Decade of Research and the Future of Predictive Policing. Social Sciences 10: 234. https:// doi.org/10.3390/socsci10060234

- Du, Y., & Ding, N. (2023). A Systematic Review of Multi-Scale Spatio-Temporal Crime Prediction Methods. ISPRS International Journal of Geo-Information, 12(6), 209. https://doi.org/10.3390/ijgi12060209
- Butt, Umair Muneer, Letchmunan, Sukumar, Hassan, Fadratul Hafinaz, Ali, Mubashir, Baqir, Anees and Sherazi, Hafiz Husnain Raza ORCID: https://orcid.org/0000-0001-8152-4065 (2020) Spatiotemporal crime HotSpot detection and prediction: a systematic literature review. IEEE Access, 8. pp. 166553-166574. http://dx.doi.org/10.1109/access.2020.3022808
- Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L., & Brantingham, P. J. (2015). Randomized Controlled Field Trials of Predictive Policing. Journal of the American Statistical Association, 110(512), 1399–1411. https://doi.org/10.1080/01621459.2015.1077710
- Gerber, M. S. (2014). Predicting crime using Twitter and kernel density estimation. Decision Support Systems, 61, 115–125. DOI: 10.1016/j.dss.2014.02.003
- Meijer, A., & Wessels, M. (2019). Predictive Policing: Review of Benefits and Drawbacks. International Journal of Public Administration, 42(12), 1031–1039. https://doi.org/10.1080/01900692.2019.1575664
- Lee, Y., Bradford, B., & Posch, K. (2024). The Effectiveness of Big Data-Driven Predictive Policing: Systematic Review. Justice Evaluation Journal, 7(2), 127–160. https://doi.org/10.1080/24751979.2024.2371781
- Alikhademi, K., Drobina, E., Prioleau, D. et al. Correction to: A review of predictive policing from the perspective of fairness. Artif Intell Law 30, 19–20 (2022). https://doi.org/10.1007/s10506-021-09299-z

- Brantingham, P. J., Valasik, M., & Mohler, G. O. (2018). Does Predictive Policing Lead to Biased Arrests? Results From a Randomized Controlled Trial. Statistics and Public Policy, 5(1), 1–6. https://doi.org/10.1080/2330443X.2018.1438940
- Kristian Lum, William Isaac, To Predict and Serve?, Significance, Volume 13, Issue 5, October 2016, Pages 14–19, https://doi.org/10.1111/j.1740-9713.2016.00960.x
- OpenAI, ChatGPT 4 was used to brief some of the research papers to write this literature review.
- Google, NotebookLM used for reasoning task on the research papers to get ideas.