

5-21-2019

Towards Crime Prevention Using Big Data Analytics: A Literature Review with an Explorative Case Study

Andreas Vassilakos

Illinois Institute of Technology, avassilakos@hawk.iit.edu

Matthias Gottlieb

Technische Universität München, matthias.gottlieb@tum.de

Maurice Dawson

Illinois Institute of Technology, mdawson2@iit.edu

Follow this and additional works at: <https://aisel.aisnet.org/mwais2019>

Recommended Citation

Vassilakos, Andreas; Gottlieb, Matthias; and Dawson, Maurice, "Towards Crime Prevention Using Big Data Analytics: A Literature Review with an Explorative Case Study" (2019). *MWAIS 2019 Proceedings*. 24.

<https://aisel.aisnet.org/mwais2019/24>

This material is brought to you by the Midwest (MWAIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in MWAIS 2019 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Towards Crime Prevention Using Big Data Analytics: A Literature Review with an Explorative Case Study

Andreas Vassilakos

Illinois Institute of Technology
avassilakos@hawk.iit.edu

Matthias Gottlieb

Technical University Munich
matthias.gottlieb@tum.de

Maurice Dawson

Illinois Institute of Technology
mdawson2@iit.edu

ABSTRACT

Since the popularization of the big data concept, it has been implemented in various areas. Contemporary literature has proved the potential of using big data in crime prevention. In this research paper, we examine research on big data being used in Crime Prevention while implementing an author-centric to a concept-centric research approach. We also present the foundation for future research by analyzing data derived from the City of Chicago. We identified the neighborhoods in Chicago that are statistically more prone to crimes and used data of the last 10 years to make our observations. We additionally provide further discussion points for future research purpose.

Keywords

Crime, Prevention, Big Data Analytics, Case Study, Literature Review.

INTRODUCTION

The crime rate in some regions of metropolitan cities is high. A shift in these areas to an area with less crime rate is complicated. However, “big data” concepts can highlight best practices and enhance solutions to prevent crime even before it comes up. In this study, we applied a systematic literature review to identify what has been done for preventing crime. To conduct this study, we choose cities within the US due to data access reasons and the reliability of this study. The purpose of the paper is to analyze big data concepts and their application to prevent crime. Thus, we follow the research question: How can big data be applied to prevent crime? The remainder of this study is: First, in the section Related Work, we present an overview of the literature according to crime prevention. Second, in the section Method and Data Set, we explain our research approach and the explored data set. Third, we highlight our findings in section Result. Afterward, we discuss our results in the section Discussion. Finally, we conclude our research and present our thought for future research.

In our research, we gathered data provided by the City of Chicago. Data gathered was ranging from 2001 to 2018, but we focus this case study on the last 10 years. We use the results of our study to raise the research question and present our foundation for future research.

RELATED WORK

Different concepts are used to predict the crime rate. In this study, we highlight the commons and differences of these concepts. Ingilevich et al. 's (2018) research intended for providing insight into different statistical tools that can be implemented in the approximation of crime numbers. Linear regression, logistic regression, and gradient boosting (GB) where the predictive model's types that were used. Data of Saint-Petersburg were used so that the intended comparison can be made. Ingilevich et al. (2018) relied on the assumption that crimes are not randomly dispersed in regards to spatial factors (Ingilevich et al. 2018). Characteristics including GDP, population density, unemployment, and homeless population numbers were considered as the external factors responsible for the occurrence of crime in a location. These social factors were considered, as well as other spatial criminal patterns. All data used in this research was provided by the Ministry of Internal Affairs of the Russian Federation and was containing information about the Date a crime occurred, the Coordinates, and the Description. The data covered the timeframe of 1/1/2014 to 2/28/2017 (Ingilevich et al. 2018). Afterward, street crimes were prioritized and categorized to banditry, massacre, and robbery. A robbery was the most commonly occurring crime type with some 142,452 occurrences (Ingilevich et al. 2018). Then, through clustering, the sub-features of each type were examined. The clusters

included the number of population, police stations, schools, malls, churches, alcohol shops, numbers of buildings, and bars. It is explained that the feature selection technique is the solution of creating strong predictive models and Gradient Boosting was proved to be the most effective in presenting crime rate predictions (Ingilevich et al. 2018).

ToppiReddy et al. (2018) used R and libraries including RgoogleMaps and googleVis to present crime statistics visually and consequently predict crime occurrence through the use of Machine Learning. The result of ToppiReddy's research was the visualization of exact crime locations on google maps, the visualization of data based on the crime type, the visualization of crime occurrence frequency, and the graphical representations and bar charts of crime frequency. ToppiReddy et al. 's (2018) suggested framework implements Data Mining techniques and applies Rational Choice Theory, Routine Activity Theory, the K-Nearest Neighbour method m, and the Naïve Bayes theorem to conduct a predictive analysis (ToppiReddy et al. 2018).

Belesiotis et al. (2018) use the hypothesis that fear and perception of crime in a city affect the quality life of citizens. Crime hotspots are being examined in order to provide insight that could help security agencies and citizens. The research takes advantage of data-driven methodology and implements data mining from online sources. The experimentation process of the authors started by identifying valuable data sources. Then they retrieved the data and used it to produce statistics. Afterward, they trained prediction algorithms with data considered appropriate. Finally, they tested the model used to test its performance. Data used in this research was extracted from the Greater London geographic region after it was separated geographically to 4831 zones (Belesiotis et al. 2018). The researchers used Web APIs to mine data from 6 sources that include features related to demographics, points of interest, transportation, land use, and photos (Belesiotis et al. 2018). It was proved that crime prediction can be more successful thanks to data-driven research and can help both citizens be vigilant of specific areas and law enforcement agencies are more efficient (Belesiotis et al. 2018).

An application named "Are You Safe" utilizes crime statistics in local communities (Liu et al. 2015). They measure GPS with certain mathematical formulas to calculate the risk of crime occurrence in an area. They proposed the term "Individual's Local Community (ILC)" which is implemented to calculate the risk of an individual's risk in a given public space (Liu et al. 2015). Liu et al. (2015) are utilizing the concepts of "Personal Big Data" which is "a form of pervasive computing, consisting of a unified digital record of the tonality o an individual's experiences, captured multi-modally through digital sensors and stored permanently as a personal multimedia archive." Afterward, Liu utilized data pre-processing, personal computing, and mathematical models to demonstrate the research results on maps (Liu et al. 2015).

Cybercrime is defined as "criminal activity (such as fraud, theft, or distribution of child pornography) committed using a computer especially to illegally access, transmit, or manipulate data" (Merriam-Webster Dictionary n.d.). This type can take various forms including Spamming, Botnets, Denial of Service (DoS), Phishing, Malware, and Website Threats. The aftermath of a committed cybercrime created the urgency of strong infosec systems to be designed. Nefarious users adjusting their attacks to the existing cybersecurity approach pose a considerable threat against internet users. Because of the presence of Big Data in contemporary computer networks, threat agents can target enormous amounts of information and consequently damage companies and organizations to an intense degree. Due to the evolution of hardware and software capabilities, big data analytics can be implemented to prevent potential crimes (Mahmood et al. 2013).

METHOD AND DATA SET

We followed the research approach of Webster and Watson (2002) to come from an author-centric to a concept-centric approach. We searched the database Google Scholar and used the keywords crime, prediction, and prevention. To focus this study, we filter our findings by two steps and the years from 2014 to 2019. First, we only looked at the title. Second, we filter the findings from the first step by reading the (1) title, (2) abstract, and (3) keywords. We looked at contributions in detail that fits the filter in all parts. Afterward, a forward and backward search was applied to identify additional relevant literature.

Further, we cluster our findings to present our results in a concept-matrix. We derived from the matrix-specific cases and focused on one case. The case study on Chicago, IL we studied in detail. Therefore, we used the database X. We ended up with a data set of about 6.8 million observation of 22 variables: ID, Case Number, Date, Block, IUCR, Primary Type, Description, Location Description, Arrest, Domestic, Beat, District, Ward, Community Area, FBI Code, X Coordinate, Y Coordinate, Year, Updated on, Latitude, Longitude, and Location.

The primary type can be domestic violence, gambling, homicide, human trafficking, interference with public officer, intimidation, kidnapping, liquor law violation, motor vehicle theft, narcotics, non-criminal, non-criminal (subject specified), non - criminal, obscenity, offense involving children, other narcotic violation, other offense, prostitution, public indecency, public peace violation, ritualism, robbery, sex offense, stalking, theft, and weapons violation.

RESULT

We had a total of 1620 hits on Scholar Google. After filtering our findings, 5 relevant articles remained. In 4 from 5 articles the type of data were digital such as social media, none were non-digital such as books or mixed combination of these. 2 papers applied data from police departments, 1 from social media, and 1 from a Web API. 3 articles had precise measurements. While 2 had specific coordinates where the crime happens. 3 articles stick to urban data and 1 to country-wide based data. 1 paper presented the data size of the reported crimes. Table 1 concludes the findings in a concept-matrix.

Author(s) (Year)	Type of Data	Data based on	Measurements	Researched Region	Data Size N
Ingilevich, V., and Ivanov, S. (2018)	Digital	Social Media	Date of crime, Coordinates, Description	Urban	N/A
ToppiReddy, H.K.R., Saini, B., and Mahajan, G. (2018)	Digital	U.K. Police Department	Coordinates, Address, Type of Crime Prediction Probability	Country	N/A
Belesiotis, A., Papadakis, G., and Skoutas, D. (2018)	Digital	Web APIs	Type of Crime, Description, Number	Urban	921,250 Crimes
Liu, Y., Marthur, A., and Magno, C. (2015)	Digital	City of Erie, PA Police Department	N/A	Urban	N/A
Mahmood, T., and Afzal, U. (2013)	N/A	N/A	N/A	N/A	N/A

Table 1. Concept Matrix of the Literature Review

We identified the top five cities with the highest crime rate and the top five cities lowest crime rate within the US. One significant distinction is the number of inhabitants. Surprisingly we found Monroe Township, NJ with about 45,000 inhabitants nearly located to New York, NY as one of the biggest cities under the top 100 in the safest regions in the US.

We looked at Chicago in detail as one of the five cities with a high crime rate in the US. We identified the areas: West Englewood, Englewood, Riverdale, Auburn Gresham, Fuller Park, Gage Park, Chicago Lawn, West Garfield Park, Grand Crossing, and Chatham as the areas in Chicago with the highest density on crime. Figure 1 highlights these ten hotspots of crime in comparison to the density of crime in Chicago, IL over time.

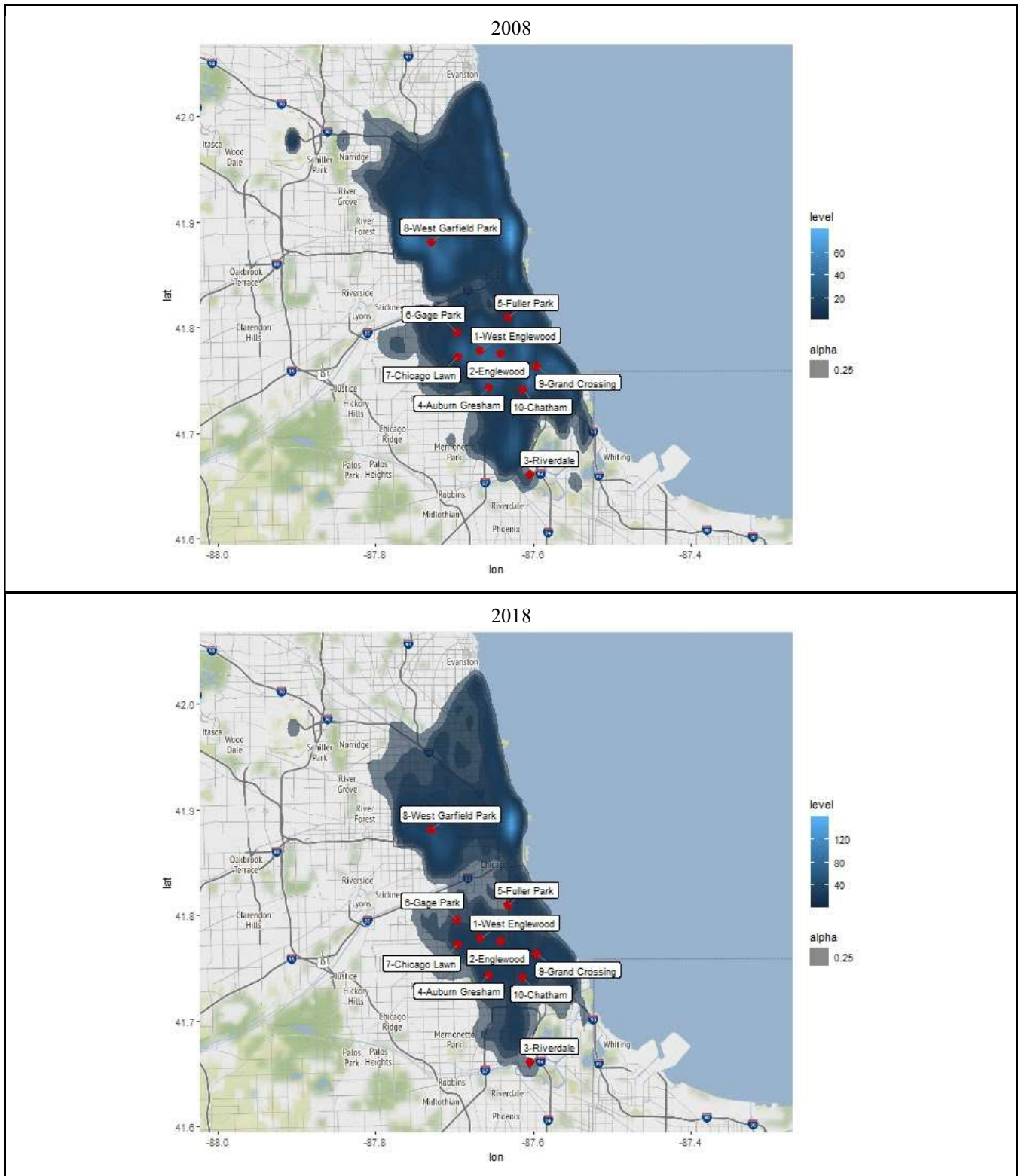


Figure 1. Crime Rate Density in Chicago, IL

The crime rate in Chicago from 2008 to 2018 decrease significantly. In 2008 the total number of crime observations was 427,053. The number went down to 264,028 in 2015. Since 2015 it is almost stable ending up by 265,650 in 2018. Figure 2 presents the number of observations per primary crime type by year from 2008 to 2018.

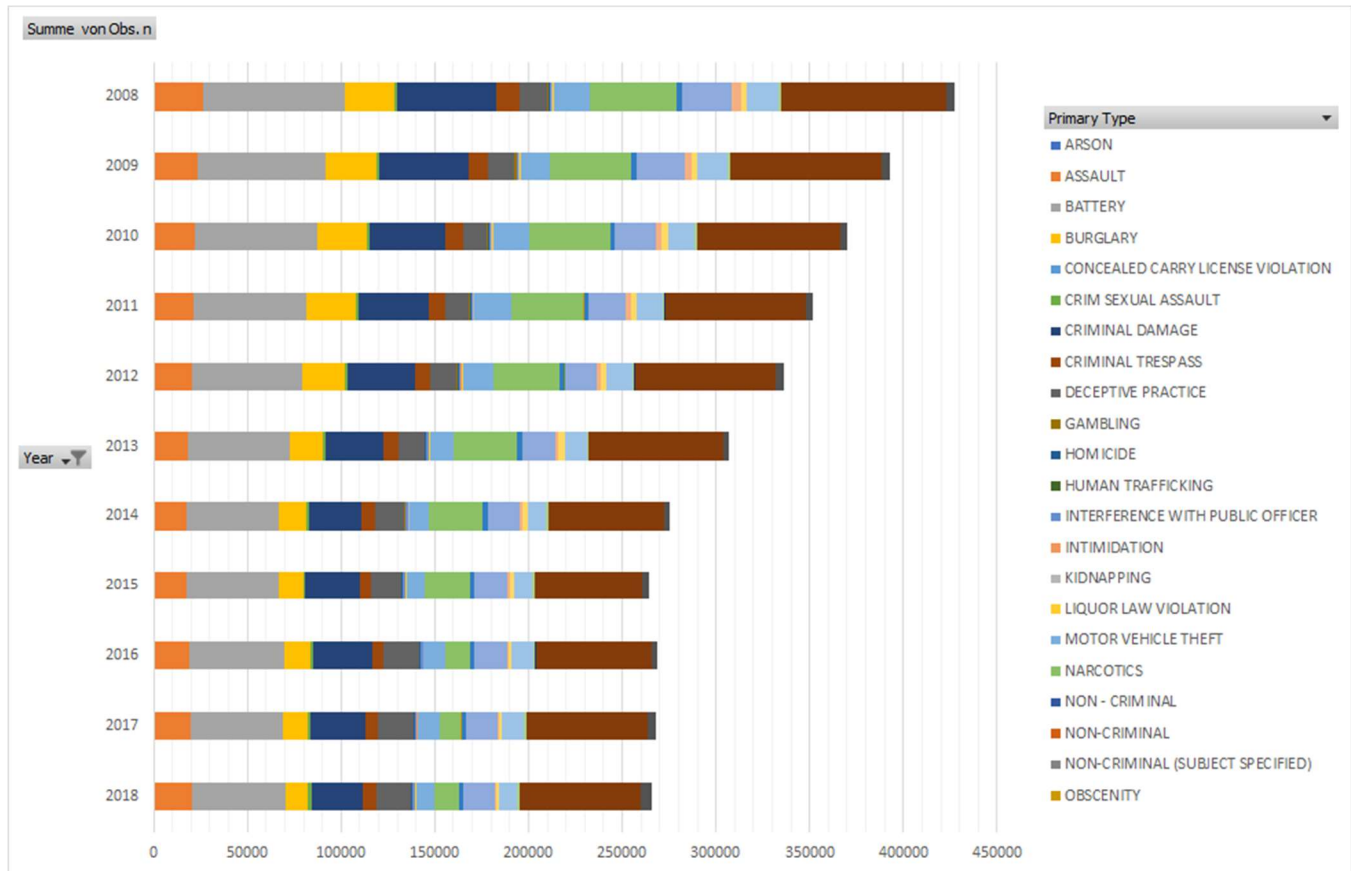


Figure 2. Primary Crime Type from 2008 to 2018 in Chicago, IL

DISCUSSION

The main observation was a decrease in the crime rate for the past ten years. In the future, we shall examine why does the crime rate decrease. Since 2015 the crime rate seems to be stable. However, a look at the locations illustrates that the density decreases in areas with a low crime rate. From the low crime rate in “good” areas by a stable number of crime observations follow that in “bad” areas the number of crime observations raises. In total, Chicago is observed to be getting safer. The stability of the crime rate raises the new question(s) requiring research: Is there a minimum of crime in Chicago? How can the crime rate be decreased even more? How can information technology be applied to lower the crime rate?

The question of what happens if future crime does no longer belong to the location, e.g., a hacker attack from a certain point on earth is raised. Is the location of crime where the hacker is located or where the crime happened creating a correlation to the possibility of higher crime rate? Does this shift from location-based crime to a non-location based crime affects the “good” areas?

CONCLUSION, LIMITATIONS AND FUTURE WORK

In this study, we identified in a case study at Chicago, IL a decrease in the crime rate. In addition, we highlight the 10 most dangerous areas in Chicago comparing on the crime rate to other neighborhoods. We elaborated five contributions to topic crime prevention. We identified a research gap which includes cybersecurity, and the measurement of the crime rate as well as a stabilizing process on the crime rate started in 2015 in Chicago.

The study is limited to the US. Thus, we might miss experiences from other nations such as those located within Europe with a lower crime rate.

We will develop a model from these findings to implement big data concepts to prevent crime and might predict crime. Research can implement this fundamental analysis on other metropolitan cities to identify different patterns of crime. Practitioners are able to identify hotspots of criminal activities and use appropriate strategies to decrease crime rate where it is observed.

ACKNOWLEDGMENTS

We thank Census, USA and the City of Chicago, IL for providing us with data and their access.

REFERENCES

1. Belesiotis, A., Papadakis, G., & Skoutas, D. (2018). Analyzing and Predicting Spatial Crime Distribution Using Crowdsourced and Open Data. *ACM Transactions on Spatial Algorithms and Systems*, 3(4), 1-31. doi:10.1145/3190345
2. Cybercrime. (n.d.). Retrieved from <https://www.merriam-webster.com/dictionary/cybercrime>
3. Ingilevich, V., and Ivanov, S. (2018) Crime rate prediction in the urban environment using social factors, *Procedia Computer Science* (136), 472-478.
4. Liu, Y., Marthur, A., & Magno, C. (2015). Crime Risk Evaluation in Individuals Local Community. *Proceedings of the 2015 Workshop on Mobile Big Data - Mobidata 15*. doi:10.1145/2757384.2757387
5. Mahmood, T., & Afzal, U. (2013). Security Analytics: Big Data Analytics for cybersecurity: A review of trends, techniques and tools. *2013 2nd National Conference on Information Assurance (NCIA)*. doi:10.1109/ncia.2013.6725337
6. ToppiReddy, H.K.R., Saini, B., and Mahajan, G. (2018) Crime Prediction & Monitoring Framework Based on Spatial Analysis, *Procedia Computer Science* (132), 696-705.
7. Webster, J., and Watson, R.T. (2002) Analyzing the Past to Prepare for the Future: Writing a Literature Review, *MIS Quarterly* (26, 2), xiii-xxiii.