An Agent-Based Approach towards Crime Prediction:
Evaluating the Potential of Environmental Factors and Human Dynamics to Model Offender Behavior

Raquel Rosés Brüngger1
1 ETH Zurich, Weinbergstr. 56, 8092 Zurich, Switzerland
rroses@ethz.ch

Abstract. Crime reduction remains a critical challenge in modern societies. Thus, identifying regions in a city with an elevated risk for future crimes is of high interest to plan prevention strategies. Attempts have been made to apply various statistical and machine learning techniques to predict crime, while computational methods such as agent-based modeling (ABM) can account for the offender and its interactions with the environment in a new manner. In this thesis proposal, I describe an agent-based model to simulate crime, implementing offender mobility and offender decision-making informed by novel data and insights in human dynamics. The result will be a data-driven crime simulation for crime prediction, making it possible to study crime patterns and to reason about likely future high crime areas. This can then flow into a crime prediction tool for the police or into an experimental tool for policy makers and urban planners.

Keywords: Crime, Simulation, Agent-based modeling, Computational Social Science, Crime Prediction

PhD Advisor: Prof. Dr. Elgar Fleisch, ETH Zurich, Switzerland

PhD starting date: 15.03.2016 and estimated conclusion date: 15.12.2019.

1 Introduction

In recent years, the most useful information to prevent crime has been gained from past crime events and their attributes. In this sense, various actors engaging in crime prevention use crime analysis to identify patterns and hot spots in past crime data, in order to direct their prevention strategies to those areas [1]. By doing so, they attempt to target the most prolific areas, where the odds of success in preventing crime are higher. Today’s technology advancements, attempt to go one step further, by predicting the future development of crime [2]. This new approach has been advanced from different perspectives and by means of various techniques. Current crime prediction techniques include statistical methods, GIS methods or miscellaneous data mining techniques [3]. On one hand, the vast majority of crime predictions rely only on crime data and do not include environmental factors, while research in the field of Criminology has shown the importance of the environment for crimes to occur [4]. On the other hand, such
models still fail to include proxies for human activities, which influence the offenders’ position and therefore its decisions [5]. In that sense, crime is a complex phenomenon happening during the interaction between an offender and its environment [6], which can be very well represented in an Agent-Based Model (ABM). The potentials of ABM for simulating social phenomenon such as crime, raise from its ability to represent multiple autonomous decision-making entities called agents in its context [7]. This approach emphasizes the interactions of individuals opposed to an aggregated approach, leading to more realistic system behavior. ABM has been successfully used to describe social phenomena such as traffic jams, stock markets, voters, or workflows in organization [7]. The complexity of the phenomenon of crime added to the more recent advances in the field of computational social science has incremented the importance of newer computational approaches to solve problems in social science, giving way to the new field of computational criminology [8].

This interdisciplinary thesis will apply computational models to support crime prevention efforts, by building an ABM for crime prediction as an alternative to classical statistical learning techniques and by including new data sources to represent crime attractors and detractors as well as by mimicking human movement and decisions. The aim is to investigate the use of new data to build a model for predicting the development of crime. This model aims to be data-driven by implementing intelligent decision-making agents into a simulation. Upon successful completion, this model should mimic the real-world crime patterns and offer the opportunity to be implemented as a crime prediction tool and as an experimental test platform, to test crime prevention strategies, policies and the impact of changes in the urban environment.

2 Related Work

2.1 Environmental criminology and risk factors

To understand why and how crime can be predicted, criminology relies on a number of interconnected theories. In general, Environmental Crime Theories focus on the influence of the environment on crime and assume that relatively rational actors take a deliberate action aiming to maximize their return when committing crime [4]. The concept of a rational offender is derived from the Rational Choice Theory in Criminology [9]. This approach can be combined with Routine Activity Theory (RAT) [6], which identifies three elements that need to be present and co-occur in time and space for an offender to act: (1) a motivated offender, (2) a suitable target, and (3) a lack of guardianship. Not only does this theory define the conditions required for a crime to happen, but it also indicates that offenders engage in criminal activities along their usual paths to normal activities (e.g. along their path to work or leisure activities). Although this approach does rather define local offenders than other types, such as offenders traveling to a new area for crime purpose, this theory is generally accepted.

Historical crime data shows that the opportunities (e.g. vulnerable targets) and motivated offenders are not randomly distributed in time and space, moreover individual criminal behaviors result into certain recognizable patterns in crime data, as described by Crime Pattern Theory (CPT) [10]. Consequently, crime occurs in the context of
complex interactions between an offender and its environment, whereas characteristics of time and space can contribute to explain the appearance of this phenomenon. In detail, there are specific environmental characteristics of places with higher crime rates. This research stream started in ecological criminology, looking into persistent high crime neighborhoods and explaining the characteristics of such deviant places [11]. More recently, places with high crime density have been described in literature as criminogenic places, which can be categorized into crime attractors, crime generators and crime detractors [12, 13]. Crime attractors are places, which attract criminals, because there are known opportunities in those areas. Consequently, the probability of a crime happening in those places is higher (e.g. nightlife district) compared to other places. In crime-generating places, crime emerges at times, where large number of people are attracted to those places for reasons other than to offend (e.g. massive sports events), supporting the importance of including proxies for human dynamic in crime prediction models. In those situations, crime occurs as a side effect, due to a high volume of people leading to more opportunities. In turn, crime-detracting places are those with a deterrent effect on criminals, mostly places with extensive guardianship (e.g. police stations) or few opportunities. Not only do points of interest in proximity influence crime, but previous victimization of places and individuals has also been identified as a risk factor, referred to as near-repeat victimization [14]. Empirical research across several cities has shown relative stability of crimes in time and space, approximately 5% of places account for 50% of recurrent crimes [15, 16]. Thus, crime occurs in the context of complex interactions between a large number of elements present in the environment and the decision of the offender, resulting in emerging crime patterns. Therefore, I argue about the importance of including various environmental characteristics in crime prediction models.

Criminology has a large body of empirical research focusing on micro level risk factors related to crime, derived from the above-mentioned theories. For the purpose of this thesis, I focus on including environmental risk factors linked to criminogenic or deviant places into the crime prediction model. Based on an empirical review I identify the following factors being relevant: (1) road-network and traffic flow [17, 18]; (2) public transportation stops [19]; (3) land use characteristics, points of daily activity and near major personal attractors (e.g. home, shopping center, restaurants, work etc.) [18, 20], as well as commercial vs residential area delimitations [21]; (4) weather [22]; (5) vegetation in urban areas [23]; (6) police stations, fire stations, and hospitals [24]. A number of those environmental factors have already been used in crime prediction models (e.g. road-network [25], points of daily activity [5, 26], etc.), while no model, especially no ABM contains all of these factors in combination. As a result, crime prediction models including those factors among others could increase the models performance.

2.2 Crime prediction methods and challenges

Police forces have a long history analyzing past crime data to recognize patterns, e.g. statistical methods or hot spot mapping [27], because (as mentioned before) previous victimization has been identified as a good predictor of future risk of crime. More advanced models for crime predictions have been developed in other research fields, such
as of data mining and machine learning, to include additional data sources for improving predictive power: (1) [28] developed the spatio-temporal generalized additive model (ST-GAM) and modeled crime using demographic data, past crime and points of interest with spatiotemporal features; (2) 29 [29] applied Kernel Density Estimation to predict crime using crime data and twitter-inferred topics; (3) similarly 30 [30] compared logistic regression, support vector machines, neural networks and decision trees using demographic data and mobile data to predict crime; (4) 26 [26] applied several types of regression to demographic data, points of interest (POI) and taxi flow.

Efforts to add various data sources for crime prediction have pushed research forward, for instance, they show the importance of including static and dynamic factors for improving predictive models. Yet, only few of the approaches compare the value of different data sources, and therefore only gain insight on the chosen environmental factor. Moreover, the above-mentioned methods fail to account for a realistic representation of the environment and the dynamic interactions between the individual actors and the environment, identified as an important factor in criminology. As such, predictive models based on statistical methods and GIS methods are limited in their predictive accuracy [31]. Data mining techniques can include environmental characteristics influencing crime in their models, but cannot account for interactions between the elements of the model on an individual level. In contrast, ABM has the potential of addressing the previously outlined challenges, as it allows representing dynamic individual agent behavior in its context or environment [32], including interactions between the elements of the model [33]. This makes ABM a promising approach for studying and understanding crime.

2.3 Previous ABM to Predict Crime

There is a new trend towards using ABM in crime prediction [34]. This relatively new simulation technique, allows for a more realistic representation of the (1) environment, (2) individual actors, i.e. agents, and (3) for individual behavior, in particular for (a) interactions between agents, and (b) interactions between the agents and the environment. Researchers from various disciplines have presented models to simulate crime by means of ABM. Such simulations aim at testing crime theories or relevant hypotheses drawn from theory [2, 3, 25], analyzing the performance of prevention strategies [35–37], and forecasting crime developments [25, 38, 39]. While the majority of such models do not include real data, research in this direction has shed light on the use of ABM to better understand crime, with the latest advances in computational criminology using this technique to predict crime.

On one hand, very few crime simulation models have attempted to crime integrating real data as characteristic of the environment: (1) [40] included the street network in a spatial model to test criminological theories such as RAT with basic offenders; (2) [25] simulate burglary in Beijing for prediction purpose, using street and subway network and a statistic-based human mobility pattern, suggesting to include more details about the environment to increase predictive accuracy; (3) [41] and [42] simulate burglary in Leeds to test the use of ABM for crime prediction, representing the environment (census data, household data, building data and street network) and detailed offender behavior only for known offenders, and using real crime data to calibrate the model. All
of those models include real data for the static environment, while their agents either are implemented manually or only represent known offenders, limiting their validity and scalability. On the other hand, offender mobility has only been modeled for known offenders caught by the police, with a known starting point for their activities (i.e. home address) and implementing empirical and theoretical knowledge to model their movement through a limited environment [37, 43–45]. To the best of our knowledge, no model has been build deriving agent mobility and decision whether to offend from data. Especially, informing offender mobility from human dynamic insights for unknown offenders and decision whether to offend or not at a certain position in space from environmental data. In addition, the models lack important static and dynamic environmental factors influencing crime such as crime attractors, generators and detractors. I therefore argue, that crime prediction models can be improved by: (1) including a larger set of static and dynamic environmental factors known to influence crime [25], (2) implementing offender mobility for a general set of offenders, derived from human dynamics insights and empirical findings on offender mobility, and (3) implementing a more realistic [46] and data-driven representation of the agents decision whether to offend.

3 Contribution and Research questions

In this interdisciplinary thesis, I aim at building an ABM-based crime prediction model using novel data sources and investigate how to increase its accuracy. This model will at first, include a number of offender agents moving randomly through the transportation network. At each time-step of the simulation, the agents will move from one point to another. The decision where to move will gradually be improved to simulate realistic offender movement, dependent on environmental characteristics such as POIs. At a later point, the offender will decide whether to offend along its path, a positive decision to offend will result in a mark recorded in the environment. The output in form of predicted crime points of each model will be compared to the real crime points for the same area using predictive accuracy index (PAI), a generally accepted accuracy measure for crime prediction models [47]. The following research questions will guide this thesis:

Overall goal: To what extent can ABM be applied to predict crime? The overall goal of this research is to study how an ABM can be design to reproduce existing crime patterns using environmental information and historical crime data. In a second step, I will examine how AMB can contribute to predicting future crime. The following research questions describe the purpose in more detail.

RQ1: How should the agent mobility be implemented? In an ABM, implementing realistic agent mobility is a critical challenge, as well as the biggest difference compared to current machine learning models implemented for crime prediction. I aim at starting with an agent randomly moving though the road network, and gradually increasing the complexity of movement rules. The rules will be derived from logic, crime theories and enriched using human mobility data and insights.
RQ2: How should the agent decision-making process be implemented? A data-driven ABM model for crime prediction can implement agents learning where to offend from past patterns in crime data. I aim to build a learning agent, informed by crime and environmental data, reproducing past crime patterns and predicting future crime.

RQ3: Which criminogenic places included as environmental factors attracting or deterring offenders increase accuracy of the prediction model? Including criminogenic places (places of interest identified in empirical research as crime attractors, generators and detractors, see chapter 2.1) in an extended model which influence agents mobility decisions, as opposed to random path selection, will reveal those places which influence offender movement and increase predictive accuracy of the model.

RQ4: Does including proxies for human activity increase accuracy of the prediction model? Proxies for human activity (such as population flow densities derived from location-based social networks or transportation counts) can be included as a risk factor of the environment attracting the offender agent, in addition to static criminogenic factors. I will study if this improves the agent mobility pattern and therefore increases predictive accuracy of the model. For instance, human activity can be implemented in the environment as a density function on the road network.

4 Methodology or Approach

4.1 Data

In line with literature in the field of criminology, a large number of data types can be used to build a simulation of crime, criminal activity and its environment. The most important data especially for testing the performance of the model is past crime data. This kind of data includes information about the location, time, type, and potentially attributes like short description, stolen goods, modus operandi, and instruments used. Crime data is publicly available for New York City and will be used for implementing the first models.

To create an accurate model to simulate crime it is also important to account for the environment in which crimes take place, and to include the risk factors identified in section 2.1. Therefore, open source data on the road network, public transportation network (from NYC open data platform) as well as data on POI (from Foursquare) will be included to model the environment in which the agents act. To model the POI, I will pay special attention to places, which either attract, generate or deter crime and to human dynamics data by including proxies for aggregated human activity (e.g. by means of location based social network data). This is in line with research using other approaches to predict crime [30], and to extract dynamic social events information (e.g. for concerts, large sports events, etc.) to account for sporadic dynamic risk factors.

4.2 Technique

ABM is described in literature as a “powerful simulation modeling technique” with various applications in social sciences [7]. This technique simulates social systems and complex social phenomena using autonomous decision-making entities called agents, which interact with their environment. Additionally, each agent is assigned a set of rules
guiding the decision-making process of the agent in its interaction with the environment or with other agents. In other words, agents take decisions on how to act based on what they perceive from their environment and according to their internal set of rules. An ABM is characterized by three basic elements: agents, the environment and interactions between the agent [33, 48]. This technique has been chosen because it allows to create a data-driven adaptive model of crime, and to project the development into the future for prediction purpose [49]. The potential of ABM lies in its ability to capture emergent behavior, include natural environment of a system and integrate geographic models, while it presents issues due to its sensitivity to initial conditions and difficulty to reproduce results [50]. Combining the advantages of ABM over other techniques and the potentials identified by other researcher to use ABM for crime prediction, this thesis attempts to investigate further applications of ABM to predict crime. The aim is to build a generalizable data-driven crime prediction model, offering the possibility to be implemented as a crime prediction tool for police forces or as an experimental platform for policy makers and urban planners to test eventual changes or even for police forces to test the effect prevention strategies.

The ABM for crime prediction will be implemented in Repast – a Java based platform and will use the above-mentioned data as initial simulation parameters [51]. The baseline model will be built using demographic and transportation network data as well as a random moving agent. Additional data will be included in batches (e.g. static environmental data, crime data, and dynamic data) to the extended model to be able to identify the contribution of each subset of data. The model will be tested on an unseen environment or time and cross validated with real crime data. Furthermore, the agents will be mobile individuals throughout the transportation network, modeled as learning agents deciding where to travel and whether to offend along their path. Human dynamics data will be implemented as a density function in a grid, and will contribute not only as a risk factor but also to model offenders path preference according to real fluctuations [30]. This is in line with RAT, as an offender tends to act along his usual path while engaging in daily activities [6] and makes a rational decision according to RCT to decide whether to offend or not in a certain situation [52]. An important issue, which will be addressed throughout this project, is the question on how to account for past crime data in the model. As outlined above, past crime events give information on likely future occurrences of crime and could therefore significantly improve crime predictions, if incorporated as a prior probability in the model.

The output of the model will be predicted crime points, which can be aggregated to identify the areas at highest risk of future crime. To determine the accuracy of the model, I will proceed to: (1) a visual comparison of simulated vs. actual crime points; (2) calculating the Predictive Accuracy Index (PAI) and the Predictive Efficiency Index [47].

5 First Results

A basic environment simulating offenders randomly moving along the road network for NYC has been setup in Repast to conduct first experiments and inform the future
development of the ABM. This first setup studies different implementations for offender mobility using location-based venue data from Foursquare in New York City (NYC), and human dynamics insights from previous studies to simulate offender movement. I hypothesize, that offender agents follow similar patterns to those of the general population. First, I compare three approaches for step size calculation informed by human dynamic insights, i.e. a static, a uniform and a Lévy-flight distance, with the goal of determining the optimal strategy for calculation of the search area radius (determining the distance which the offender attempts to travel). Second, I leverage user-generated content from Foursquare (i.e. venues and check-ins) [5] to inform offender target selection (target is selected within the distance of the search radius). The aim of this comparison is to assess if unknown offenders can be realistically modeled in such manner. The setup results in nine offender mobility designs, combining search distances strategies and target selection algorithms. The offender behavior performance was measured using the ratio of crime locations passed vs average distance traveled by each offender, inspired (PAI). The first results (see Table 1) for this implementation, show no difference in the different search selection strategies, while the target selection algorithm for POI is performing best, followed by popular POI. The search selection strategies could be improved by informing the initial position of the offender (e.g. placed with higher probability in residential areas) and by assuming the initial position as the home location, to which the offenders returns after a certain number of steps.

<table>
<thead>
<tr>
<th></th>
<th>Static strat.</th>
<th>Uniform strat.</th>
<th>Lévy strat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road al.</td>
<td>0.324</td>
<td>0.382</td>
<td>0.267</td>
</tr>
<tr>
<td>POI alg.</td>
<td><strong>0.439</strong></td>
<td><strong>0.519</strong></td>
<td><strong>0.425</strong></td>
</tr>
<tr>
<td>Popular POI alg.</td>
<td>0.347</td>
<td><strong>0.523</strong></td>
<td>0.283</td>
</tr>
</tbody>
</table>

6 Discussion and Implications

In summary, I have proposed a doctoral thesis addressing to what extent ABM can be applied for crime prediction. The aim is to model offender behavior using novel data, improving state of the art ABM for crime. I argue, that novel data and human mobility insights, can inform more realistic offender mobility and decision-making. First steps in this direction have already been made in the first prototype.

The proposed project will not only advance our understanding on crime and on how complex social systems can be simulated using agent-based modeling, but it will also provide insights into crime patterns, crime development, environmental influences on crime as well on offenders behaviors (mobility and decision making). Hence, this research will add scientific value to criminology, computational social science, in detail about informing offender mobility and decision-making using novel data.

In terms of practical implications, the crime prediction model resulting from this thesis could be applied to inform evidence based crime prevention as well as to provide
the basis for an experimental platform to support policy makers and city planners in their decisions. This research is expected to have a positive impact on the national security for the following main two reasons. First, the ability to predict crime to a certain extent and on an aggregated level provides relevant information to the police and inform their resources allocation. On the other hand, an agent-based crime prediction model can be developed to an experimental and intelligent decision support system. Such a system allows performing experiments in the context of the simulation, meaning that for example environmental factors can be changed in the simulation to assess how this affects future crime. Therefore, policy makers, urban planners and other actors involved in shaping the environment can implement new planned actions in the simulation before implementing them in real life, to assess their impact on crime. Not only does such a platform support decisions on an empirical basis, but it also allows carrying out preliminary experiments on a large number of subjects with little costs.

References


