Design of an Agent-Based Model to Predict Crime (WIP)

Raquel Rosés Brüngger  
Department of Management, Technology and Economics  
ETH Zurich, Switzerland  
rroses@ethz.ch

Cristina Kadar  
Department of Management, Technology and Economics  
ETH Zurich, Switzerland  
ckadar@ethz.ch

Irena Pletikosa Cvijkj  
Department of Management, Technology and Economics  
ETH Zurich, Switzerland  
ipletikosa@ethz.ch

ABSTRACT
In modern societies where fighting crime has a long history, establishing effective methods for crime prevention is of high significance. For this purpose, police departments worldwide undertake efforts to analyze past crime data. They aim to detect the most prolific crime areas and predict their development, in order to direct their prevention efforts. In parallel, scholars investigate possibilities to build crime prediction models, by applying various techniques, from simple regression to data mining. Acknowledging the latest advances in this field, which suggest that Agent-Based Modeling (ABM) is a promising method, in this paper we present the design of an ABM capable of predicting where and when future crimes will most probably happen. We extend the previous work by accounting for offender behavior and integrating a realistic representation of the environment. In contrast to existing models, past crime data will be included to achieve automatic calibration. Furthermore, we will assess how crime data can be used to model agent’s behavior and we will include environmental data stepwise, in order to achieve the optimal balance between prediction accuracy and complexity. The resulting ABM will be developed as a crime prediction tool, and as an experimental environment to test prevention strategies.

Author Keywords
Agent-Based Model; Crime Prediction; Crime Simulation.

1. INTRODUCTION
Crime reduction is a critical challenge around the world [1], with numerous negative consequences ranging from financial losses to decreasing emotional well-being [2]. Still, crime levels can be reduced by implementing prevention strategies [3], e.g. preventive police patrols. Moreover, crime analysis in line with Crime Pattern Theory (CPT) has confirmed that crimes are unequally distributed in time and space [4]. Consequently, it is possible to identify areas at higher risk of crime for a certain period of time, and to predict its development by means of data analysis [5, 6]. Nowadays, the most effective information used to prevent crime is gained from past crime events and their attributes.

The various actors engaging in crime prevention use crime analysis to spot patterns and hot spots in the environment, allowing them to direct their prevention strategies [7]. By doing so, they target the most prolific areas, where the odds of success in preventing crime are highest. Moreover, scientists are trying to go one step further by trying to predict which areas will be more prolific in the future [5]. With the ability to foresee where future crimes are more likely to take place, the effectiveness of crime prevention strategies can be improved [8]. Despite the advances in crime prediction tools, their application in policing operations is still limited due to practical reasons [9] and their accuracy can still be improved to further reduce crime.

One approach towards achieving this goal is through simulation, which is described in literature as a suitable tool not only to better understand complex phenomena, but also for predictive modeling in social sciences [10]. Simulation, as a scientific methodology, can be applied to create models of social systems in order to process complex inputs and produce consequences which result in predictions [11]. Consequently, simulation can be seen as a potential methodology to simulate future crime (e.g. to make predictions about future human behavior [12]). In particular, agent-based modeling (ABM) was shown to be a “powerful simulation modeling technique” [13], which can be applied to simulate social systems and complex social phenomena [14]. Moreover, crime is understood as part of a complex social system, which can be explored using computational techniques [15, 16]. As we will outline in the next section, existing methods to simulate crime applying ABM show a potential to be improved by including a more realistic representation of the environment and of the offenders.

In order to contribute in the direction of crime prevention through ABM, in this paper we present our preliminary work towards developing an ABM to simulate crime.

2. BACKGROUND AND PREVIOUS WORK
2.1. Understanding Crime
With the aim of designing an ABM for crime simulation, the phenomenon of crime, and the elements influencing its appearance and development first need to be outlined and understood from a criminological point of view. According to the Routine Activity Theory (RAT) [17], there are three
elements, which need to be present in time and space for a crime to happen: (1) a motivated offender, (2) a suitable target, and (3) a lack of guardianship. Following this idea, an offender evaluates a situation and only acts if the three mentioned elements are present and come together in time and space. Furthermore, in the Rational Choice Theory (RCT), the offender is seen as a rational decision-making individual, which evaluates a situation and tries to maximize the outcome from his actions [18]. As opportunities (e.g. vulnerable targets) and motivated offenders are not randomly distributed in time and space, individual criminal behaviors result into certain recognizable patterns in crime data, as described by CPT [4]. As a result, crime occurs in the context of complex interactions between an offender and its environment, whereas characteristics of time and place can contribute to explain the appearance of this phenomenon. In detail, there are specific environmental characteristics of places with higher crime rates. In the literature those places are described as criminogenic places, which can be characterized as crime attractors or crime generators [19]. Crime attractors are places which attract criminals, because there are known opportunities in those areas. As a consequence, the probability of a crime happening in those places is higher compared to other places (e.g. night life district). In turn, crime generators are places in which 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). In those situations, crime occurs as a side effect, due to a high volume of people leading to more criminal opportunities. In conclusion, to simulate crime using ABM, we need to account for specific environmental factors influencing the appearance of crime, as well as for the offenders and their characteristics.

2.2. Related Work
In the next section we introduce work related to the purpose of this paper. Particularly, we introduce existing crime prediction methods and the challenges they face, followed by the opportunities that ABM offers to resolve these issues. Furthermore, we present previous ABM to simulate crime, with an emphasis on potentials to improve those models.

2.2.1. Crime Prediction Methods and Challenges
Current crime prediction techniques include statistical methods, geographical information systems (GIS) methods and miscellaneous data mining techniques. For a review on current crime prediction techniques refer to [9] and [20].

One of the drawbacks of the above mentioned methods is their failure to account for a realistic representation of the environment [15], which has an important influence on whether crime happens in a certain point in time and space [7]. Furthermore, they also fail to represent the dynamic interactions between the individual actors, such as the criminals and the environment in which they act [22]. Thus, predictive models based on statistical methods and GIS methods will always be limited in their prediction accuracy [21]. Moreover, miscellaneous data mining techniques are starting to account for the various factors influencing crime, but not for the interactions between those elements [19]. Most techniques used to predict crime don’t allow to include both of these important factors in one model. Yet, an ABM has a potential to address the previously outlined challenges of commonly used crime prediction methods, making it a promising approach for studying and understanding crime. An ABM allows to simulate reality and model individual behaviors in a new and dynamic way [21], thus providing the possibility to study individual’s behaviors within its context [22]. This kind of simulations allow replication of dynamic processes involving single agents (representing individuals), and their interaction with the (changing) environment.

2.2.2. Previous ABM to Predict Crime
A number of researchers from various disciplines have presented models to simulate crime by means of ABM. Still, the aim of the majority of those simulations has been to test crime theories or relevant hypotheses drawn from theory [23], without aiming to represent the phenomenon of crime in a realistic manner. Nonetheless, research in this direction has sheered light on the use of ABM to better understand crime, with the latest advances in computational criminology using this technique to predict crime.

Researchers have attempted to simulate crime in many different ways, starting with simple models, not aiming to represent the environment with real data. Examples of such work include [24] and [25], who created a basic multi-agent model of crime, based on the RCT with a goal to examine the equilibrium of crime and punishment. These preliminary models showed that dynamics of crime can be studied using simulation techniques. In the same direction, ABM was used to test and validate criminological theory, arguing for the potential of this method to simulate crime dynamics [26]. In this paper, the authors, presented two models, one for general offending, and one for burglary. Furthermore, [27] proposes an ABM to study strategies of preventive policing by simulating criminals and police patrols, and [28] proposes an ABM combined with Cellular Automata (CA) based on social disorganization and RAT, where agents decide where to commit a crime based on certain characteristics of the environment, but without the ability to navigate the environment. Likewise, [10] built a first ABM to predict physical and cybercrime, using clustering methods to estimate the number of offender agents in the environment, based on incident attributes. In contrast to those simpler models, more advanced models integrate spatio-temporal characteristics. They include GIS or other concrete features of the environment, which offer a more realistic representation of the crime phenomena. In this sense, [16, 22, 29] present a framework for modeling crime, integrating theories of crime analysis and prediction, by means of multi-agent modeling combined with abstract state machines. This model includes a more realistic urban environment with a road network, which the agents can navigate, and a basic
temporal component, for a hypothetical analysis of crime. Furthermore, [30, 31] simulate crime using CA integrating a GIS environment to study the interactions between offenders and targets in a criminal incident, representing the impact of crimes on targets by tension (as a state variable in the CA) and including the reactions of the targets to a crime. This model includes real crime data for visual model calibration. Similarly, [32, 33] aim for a spatio-temporal model integrating GIS to better understand crime events and to test criminological theories such as RAT, with basic offenders and parameters.

The most recent and complete work, [34–39] attempts to simulate burglary in the city of Leeds by means of ABM, accurately representing the environment and offender behavior. The agents can navigate the virtual environment and decide whether to burglar a property or not, depending on inner drives and believes, as well as on the characteristics of the property. Again, real crime data has been used to manually calibrate the model. Finally, [23] simulate burglary for Beijing, using street and subway network and a statistic-based human mobility pattern, suggesting to include more details about the environment to increase accuracy of the predictions.

To summarize, the most promising ABM built to this date is the one from [34–39], being the only one with a rather realistic representation of the environment and with detailed offender architecture. Still, this work lacks a realistic implementation of agents in general, and does not use automated calibration of the model. Thus, there is a potential to improve crime prediction tools and their utility, by pursuing and extending the research stream on applications of ABM for simulation of crime [23, 34, 35]. In detail, more realistic representation of the environment can include crime generators and crime attractors (which shape crime distribution) [23], as well as a more accurate representation of the offender’s behavior [34]. Additionally, including crime data in the model, as opposed to using them only for calibration purposes, could improve the predictive accuracy of the model. Finally, automating the calibration process of the simulation parameters could lead to a self-learning model which would adapt to the new types of criminal behavior.

3. CRIME PREDICTION MODEL ARCHITECTURE

This project aims to build an ABM to predict crime. We focus on property crimes, representing the wide majority of crimes committed in Switzerland [40]. The architecture of the system is represented in Figure 1.

To build the system, in the first step of the process, real crime data for one Swiss canton will be used. In addition, environmental data will be gathered. In a second step, the initial simulation parameters will be derived from the input data and implemented in the ABM model. The predictions resulting from running the model will be compared to real crime occurrences for the same area, in the fourth step. This automated comparison will help to determine whether the initial simulation parameters should be adjusted (step 5), and if additional environmental data should be added to improve the accuracy of the predictions (step 6). Thus, the model will be assessed iteratively, before more complexity is added to it, with a goal of building the simplest model, which results in highest prediction accuracy [15].

3.1. Data Gathering and Preparation

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 for this purpose is real historical crime data [41]. We have been granted access to past crime data for 5 years. The data requested includes the following attributes for the offences: type, place, date, time, stolen objects, damaged objects, total loss estimation, modus operandi, offence instruments, description of the offence, vehicle used, related offences, case link, type of link, arrests. Additional to the data on crime incidents, data on the offenders has also been requested. This data would enable to define more accurate offender agents within the simulation model. In detail, demographic data about the offenders, as well as the location of residence have been requested. The aim is to model offenders in a data driven manner, including information about past crimes in their behavior or memory. Apart from modeling the offenders, demographic census data will be added, to model guardian agents, which will represent the general population of the areas of interest.

As outlined in the background section, to create an accurate model for crime simulation, it is also important to account for the environment in which crimes take place. As described by the RAT, offenders, as well as victims, navigate the real environment to fulfill their routine activities, referring to normal everyday activities, which influence criminal opportunities [17]. To account for this, we will collect open source data on the road network, public transportation network as well as data on POI (Points of

![Figure 1. ABM architecture](image-url)
Interest) from OpenStreetMaps (OSM)\(^1\) and Foursquare\(^2\), to model the environment in which the agents circulate. To model the POI, we will pay special attention to places which either attract or generate crime [19]. As such we plan to build a web crawler to extract specific social events information (e.g. type, capacity, opening time, closing time, date, location, etc.). Finally, human dynamics data could be added to the simulation for a more accurate representation of agents flow in the environment, in line with research using other approaches to predict crime [42]. For this purpose, proxies will be selected to represent real human activity, for example by means location-based social network data [43]. Past crime and environmental data will be used as input to the initial model of the simulation to shape the agents as well as the environment. Moreover, a big challenge is to identify the relevant data to be included in the model and to determine how much complexity is needed to maximize the accuracy of the model, without unnecessarily increasing the complexity. Therefore, we aim to build an efficient crime prediction model with the right balance between complexity and accuracy.

3.2. Crime Prediction Model

The prepared data for the ABM will guide the initial simulation parameters for the model. For this purpose, the first 4 years of the past crime data will be used to train the model and the 5\(^{th}\) year will be used to assess the prediction accuracy of the model and to decide whether the initial simulation parameters of the model should be further adjusted and more complexity added to the model. Furthermore, environmental data will be added to the model iteratively to determine the impact it has on the accuracy of the model. As a result, the model will be built stepwise increasing the complexity of the environmental factors.

The agents will be modeled as intelligent behavior based agents (see [26] and [44]), following the belief-desire-intention (BDI) approach [34, 45]. In detail, two types of agents will be modeled: offender agents and (simpler) guardian agents. Both agent types will bear the similar inner architecture, although offender agents will have a stable degree of propensity to engage in criminal behavior over the course of the simulation [46]. According to the General Theory of Crime, whether an individual offends or not is related to its personality traits such as self-control, which remain relatively stable over an individual’s life [47]. Moreover, several offender types will be modeled by varying their preferences. As a general norm, guardian agents will have a deterrent effect on offenders willing to commit a crime, when in a certain distance. In general, the agent’s population will roughly represent the real population of the simulated area. Furthermore, the agent’s behavior will be based on their internal architecture, which will determine the activities each agent engages in during the simulation.

Moreover, agents engage in routine activities such as being at home, working, and socializing. To fulfill their inner drives or goals the agents navigate the environment with certain routes and at certain times [26] (e.g. human dynamics data can contribute to model the routes according to real fluctuations [42]). As described in chapter 2.3. and according to RAT, an offender tends to offend along the routes he or she takes while engaging in routine activities [17] and makes a rational decision according to RCT to decide whether to offend or not in a certain situation [18]. As a consequence, an agent decides whether to offend based on its perception of the environment and its inner drives [26].

For the agents to take a decision to act, they need to perceive the environment, which needs to represent reality to a certain degree. Therefore, the simulated area will be represented using the data sources mentioned above, starting with the most basic environmental characteristics and increasing complexity of the model gradually. Hence, data available on OSM will be added in the first step, including the road and public transportation network, followed by more detailed POI. Additionally, specific crime generator and attractors (e.g. night clubs with opening hours) will be included in later steps, to assess their impact on the accuracy of crime predictions.

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.

For this project the evaluation process will be done iteratively, to improve and maximize the models accuracy in each iteration. For that purpose, the models predictive-output will be compared to real data, i.e. to the test data for the last year, to automatically calibrate the model. To be able to do so, the model will be run the equivalent time of one year, to aggregate the crime data. In detail, we will assess to what degree the following evaluation criteria are met: (1) the simulated crime points and the actual crime data points will be displayed and visually compared [34], (2) the patterns will be compared using mean center and standard deviational ellipse to describe the central and directional tendency [23], (3) a density surface map will be created to study hot spots of crime using kernel density estimation method (KDE) [34] and its distribution in space [23], and finally, (4) local Moran’s I will serve to identify hotspots at a significant statistical level [23].

4. CONCLUSION AND FUTURE WORK

In this paper we have introduced the motivation for this project, reviewed previous attempts to simulate crime using ABM, and described the novel design for a crime prediction

---

\(^1\) See http://www.openstreetmap.org

\(^2\) See https:// foursquare.com/
tool. Future work will implement a working prototype of the model, with crime data and environmental data in the Recursive Porous Agent Simulation Toolkit (RePast). The prototype will include basic environmental data (e.g. road network and public transportation network) and the agent architecture. Using the knowledge obtained while implementing the prototype, more complete and accurate crime prediction tool will be implemented. The resulting agent based simulation will have two main applications: (1) to predict probable future development of crime levels for the simulated area (2) to provide an experimental environment to test the effect of prevention strategies on crime. Additionally, the crime prediction tool can be expanded to include other types of crime in the future and can be applied by various actors in the public and private sector to prevent crime.

REFERENCES


