Applied Geography 86 (2017) 255–261

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context

Anneleen Rummens, Wim Hardyns*, Lieven Pauwels

Institute for International Research on Criminal Policy (IRCP), Department of Criminology, Criminal Law and Social Law, Ghent University, Belgium

A R T I C L E I N F O

Article history:
Received 1 November 2016
Received in revised form
23 January 2017
Accepted 6 June 2017
Available online 20 June 2017

Keywords:
Predictive analysis
Predictive policing
Crime forecasting
Spatiotemporal modeling
Crime mapping

A B S T R A C T

Police databases hold a large amount of crime data that could be used to inform us about current and future crime trends and patterns. Predictive analysis aims to optimize the use of these data to anticipate criminal events. It utilizes specific statistical methods to predict the likelihood of new crime events at small spatiotemporal units of analysis. The aim of this study is to investigate the potential of applying predictive analysis in an urban context. To this end, the available crime data for three types of crime (home burglary, street robbery, and battery) are spatially aggregated to grids of 200 by 200 m and retrospectively analyzed. An ensemble model is applied, synthesizing the results of a logistic regression and neural network model, resulting in bi-weekly predictions for 2014, based on crime data from the previous three years. Temporally disaggregated (day versus night predictions) monthly predictions are also made. The quality of the predictions is evaluated based on the following criteria: direct hit rate (proportion of incidents correctly predicted), precision (proportion of correct predictions versus the total number of predictions), and prediction index (ratio of direct hit rate versus proportion of total area predicted as high risk). Results indicate that it is possible to attain functional predictions by applying predictive analysis to grid-level crime data. The monthly predictions with a distinction between day and night produce better results overall than the bi-weekly predictions, indicating that the temporal resolution can have an important impact on the prediction performance.

1. Introduction

Big data and predictive analysis are relatively new concepts in criminology, while they have become standard practice in disciplines such as finance, marketing, business intelligence, and biomedical sciences. It has recently been stated that the use of big data will also inadvertently impact social sciences and humanities in general (Kitchin, 2014) and criminology in particular (Chan & Moses, 2015). In the context of crime analysis, the large amount of crime data available in police databases can be considered a valuable source of big data, which we can use to gain useable new insights and knowledge on current and emerging crime trends and patterns. The application of advanced statistical methods to obtain this intelligence from big data is commonly referred to as predictive analysis. The use of predictive analysis in criminological applications is often referred to as predictive policing (Perry, McInnis, Price, Smith, & Hollywood, 2013). It can be defined as: *"the use of historical data to create a spatiotemporal forecast of areas of criminality or crime hot spots that will be the basis for police resource allocation decisions with the expectation that having officers at the proposed place and time will deter or detect criminal activity"* (Ratcliffe, 2014, p. 4).

As a policy strategy, predictive policing can be situated within the scope of intelligence-led policing (ILP). The main objective of ILP is to apply crime data analysis to objectively inform policy, policing strategies, and tactical operations in order to reduce and prevent crime (Ratcliffe, 2016). The emphasis lies on a proactive use of police resources, in contrast to reactive crime response strategies. Ratcliffe (2016) developed the 3-i (interpret, influence, and impact) conceptual model to summarily visualize the relevant actors and their role in ILP: crime intelligence analysis interprets the criminal environment, using this intelligence to influence the decision-making process, which in turn leads to the development of strategies to impact the criminal environment. Similarly, it can be situated within the broader 5Is framework (intelligence, intervention, implementation, involvement and impact) (Ekblom, 2011),

* Corresponding author. IRCP, Ghent University, Universiteitstraat 4, B-9000, Gent, Belgium.
E-mail address: wim.hardyns@ugent.be (W. Hardyns).

http://dx.doi.org/10.1016/j.apgeog.2017.06.011
0143-6228/

which emphasizes knowledge-based prevention.

The potential to predict crime follows from the empirical observation that crime does not happen randomly, but tends to be concentrated in time and space in so-called crime hotspots. Therefore, there are contextual factors and patterns influencing crime opportunity that can be identified (Brantingham, 2013; Kinney, Brantingham, Wuschke, Kirk, & Brantingham, 2008). Hotspot analysis can be considered the main precursor to predictive policing. The general principle of hotspot analysis is that crime events are mapped to find high concentrations of crime during a certain period. However, for the purposes of predicting crime this technique is considered retrospective, in the sense that patterns from the past are merely extrapolated to the present. To counter this problem, prospective hotspot analysis was proposed by Bowers, Johnson, and Pease (2004). In prospective hotspot analysis, hotspots are formed not by the areas with the highest concentration of crime, but by the aggregation of risk zones surrounding each incident. These risk zones are temporary, making the hotspots more dynamic. Another related development is risk terrain modeling (RTM) (Caplan & Kennedy, 2010), which creates a risk map of locations sensitive to high crime rates, based on their spatial properties and the interactions of those properties. Both prospective hotspot analysis and RTM result in a heat map charting the risk of crime for each area. The purpose of these prospective methods is to obtain a more dynamic picture of future crime trends (Groff & La Vigne, 2002; Ratcliffe, 2010). Predictive policing can thus be considered a step forward in the crime mapping evolution because of its specific focus on spatiotemporal predictions of crime, thus enabling a more accurate estimation of future crime patterns. This process is fully in line with the recent evolution of using micro geographic levels of analysis (street segments) in criminological research to explain the unequal distribution of crime (e.g. Groff, Weisburd, & Yang, 2010) and has been suggested as the new standard in crime mapping and analysis (Weisburd, Groff, & Yang, 2012). The micro geographic level is considered to be more suitable and accurate as it better reflects the existing variability at that level of both crime and socio-economic variables and provides more predictable crime patterns compared to higher geographic units of analysis such as census tracts, neighborhoods or districts (Weisburd et al., 2012).

Several types of statistical models for applying predictive policing have been proposed in the literature (among others Mohler, Short, Brantingham, Schoenberg, & Tita, 2011; Wang & Brown, 2012). Near-repeat models focus on a specific phenomenon recognized mainly in home burglaries: when a burglary takes place, other homes within a certain distance have, within a specific time frame, a higher risk of also being burglarized (Bowers & Johnson, 2005). Several models have been developed based on this phenomenon, also taking inspiration from similar natural phenomena. For example, Townsley, Homel, and Chaseling (2003) use the Knox method, an epidemiological model used to simulate the spread of a contagious disease and Mohler et al. (2011) make use of self-exciting point process models, based on the modeling of earthquake aftershocks. The main advantage of these types of model is that they only need time, place and type of crime. Their effectiveness derives from the aforementioned near-repeat phenomenon and the fact that the near-repeats act as proxies for the underlying factors that make a particular area attractive for crime. However, this is also their main disadvantage: they only focus on this near-repeat phenomenon and are usually not flexible enough to incorporate other information as well. Statistical models such as logistic regression and the more complex machine learning algorithms such as neural networks are therefore used to allow demographic, socio-economic, land use and crime opportunity variables representing the environmental background to be incorporated. These models also have the advantage of being well documented in other domains in which predictive analysis is applied.

Applications of predictive policing are currently being used by law enforcement in the US, the UK, the Netherlands, Germany and Switzerland (Hardyns & Rummens, 2016). Other countries, such as Austria (Glasner, 2015), are currently exploring the possibility of implementing predictive policing. Although the number of applications of predictive policing keeps increasing, only a few randomized field experiments (Mohler et al., 2016 and Temple University, n.d.; study still in progress) have been conducted.

The aim of this study is to explore the potential for applying predictive analysis in an urban context, specifically in a large city (population > 250.000) in Belgium, and to explore the possibilities and limits of this method in spatial criminology. The following research questions will be addressed:

1) What is the predictive performance for home burglary, street robbery and battery?;
2) Are there differences between daytime and nighttime predictions, and how do they compare to the predictions made without such a distinction?

To answer these questions, predictive analysis is applied to the available crime data and its predictive performance is evaluated. The three crime types of home burglary, street robbery, and battery[1] are compared. Home burglary and street robbery were chosen as they are high impact-crimes prioritized by the local police services. Battery was included to enable a comparison between violent crime and property crime. The comparison between day and night is included, first, because there is often a practical need for differentiated maps depending on police shifts, and second, to test the intuitive assumption that there are important differences in crime patterns that could influence the prediction performance.

2. Data and methods

The available crime data are spatially aggregated to grids of 200 by 200 m and retrospectively analyzed. The approach taken in this study is based on the Crime Anticipation System (CAS), developed by the Amsterdam Police Department in the Netherlands. CAS collects historical data on more than 200 variables, consisting of crime, demographic, socio-economic and land use indicators. These variables are modeled using neural networks and the 3% highest risk cells identified by the model are then mapped using a grid with a resolution of 125 by 125 m (Hardyns & Rummens, 2016). A similar method to CAS was chosen because from all currently known cities implementing predictive policing, Amsterdam is most comparable to the city under consideration in this study.

2.1. Variables included in the model

The total number of variables is deliberately kept low by selecting relevant indicators based on previous empirical research (Hardyns, Vyncke, Pauwels, & Willems, 2015). Where possible, the variables were collected at the grid level, but some of the variables could only be collected at the higher level of the statistical sector (the smallest administrative level in Belgium), meaning that each

[1] Home burglary is defined as: ‘theft from a home by breaking and entering or illegally trespassing, including attempts’; street robbery is defined as: ‘theft in a public place using violence, the threat thereof or a deadly weapon, including snatch thefts and attempts’; battery is defined as: ‘the intentional use of force or violence, causing bodily harm’.