

One step ahead of the burglars

May 3 2019, by Michael Walther



Variables such as time of day, place and population density help to classify a certain plot of land as at risk or not at risk of burglary at any given time. Credit: ETH Zurich

A new machine-learning method developed by ETH scientists makes it possible to predict burglaries even in sparsely populated areas.

Break-ins do not happen everywhere all the time. Certain communities, neighbourhoods and streets, as well as seasons of the year and times of the day, have a lower or higher risk of a burglary taking place. Using break-in statistics, machine learning techniques can identify patterns and

predict the risk of a break-in at a specific location. Computer programs can thus help the police to identify burglary hotspots – places at particularly high risk of a break-in – on any given day, enabling them to deploy patrols accordingly.

Class imbalance makes learning more difficult

To date, such warning systems work only in densely populated areas, primarily in cities. That's because [computer programs](#) need sufficient data in order to recognise patterns, and crime is less frequent in sparsely populated areas. This is referred to as a "class imbalance" in statistics. Specifically, this means that for every section of road that has a burglary, there are several hundred or even a thousand that do not.

Algorithms work in parallel

Cristina Kadar is a computer scientist and doctoral student in the Department of Management, Technology, and Economics. She has developed a method that can make reliable forecasts despite imbalanced data. Her research has just been published in the journal *Decision Support Systems*. She tested numerous machine learning methods with a large data set of burglaries in the Swiss canton of Aargau, combined them and compared the hit rates. A method that uses ensemble learning and combines analyses of different algorithms proved to be the most accurate.

Machine learning is when an algorithm uses large data sets to train itself to classify data correctly. In this example, it takes variables such as time of day, place, population density and much more and learns from them whether to classify a certain plot of land as at risk or not at risk of burglary at any given time.

The challenge lay in training the classification algorithms despite the small number of burglaries in the data set. Kadar preprocessed the data set by randomly removing data units without burglaries until she arrived at the same number of units with burglaries as units without. This statistical method is called "random undersampling". Kadar trained numerous classification algorithms with this reduced data set in parallel, and their aggregated forecasts produced the burglary forecast. Kadar took grid cells of 200 by 200 meters on a given day as her individual data units.

While conventional warning systems mainly use burglary data, Kadar also fed the classification algorithms with impersonal aggregated [population data](#), such as population density, age structure, type of building development, infrastructure (presence of schools, [police stations](#), hospitals, roads), proximity to national borders, as well as temporal information including day of the week, public holidays, hours of daylight and even the phase of the moon.

Hit rate better than in cities

With the new method, Kadar was able to significantly improve the hit rate compared with conventional methods. She directed the computer to use her method in predicting hotspots where burglaries were likely to occur within the canton. A review showed that around 60 percent of actual break-ins were committed in the predicted hotspots. By comparison, when the hotspots were predicted using the traditional method employed by the police, only 53 percent of actual burglaries occurred in the predicted area. "With imbalanced data, the method achieves at least equally good and in some cases better hit rates than conventional methods in urban areas, where the data is denser and more evenly distributed," says Kadar.

The findings are useful first and foremost for the police, as the method

can also be used to predict regions and times with an increased risk of [burglary](#) in less densely populated areas. However, there's no reason why the method couldn't be used to predict other risks: health risks, for example, or the probability of emergency calls to the ambulance service. The [real estate](#) industry could also use it to forecast the development of property prices on the basis of spatial factors.

More information: Cristina Kadar et al. Public decision support for low population density areas: An imbalance-aware hyper-ensemble for spatio-temporal crime prediction, *Decision Support Systems* (2019). [DOI: 10.1016/j.dss.2019.03.001](#)

Provided by ETH Zurich

Citation: One step ahead of the burglars (2019, May 3) retrieved 23 April 2024 from <https://techxplore.com/news/2019-05-burglars.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.