polimaps: Supporting Predictive Policing with Visual Analytics

Florian Stoffel\textsuperscript{1}, Hanna Post\textsuperscript{2}, Marcus Stewen\textsuperscript{2} and Daniel A. Keim\textsuperscript{1}

\textsuperscript{1} University of Konstanz, Germany
\textsuperscript{2} State Office for Criminal Investigation of North Rhine-Westphalia, Germany

Figure 1: Main user interface of polimaps showing example data (\cite{Nat17}). A and B present the current layers as well as contextual information. C denotes the main visualization canvas, D the current spatial selection for the aggregated information shown in B, E an exemplary heat map with point overlay. F indicates the colors used to indicate point density in E.

Abstract

Recently, predictive policing has gained a lot of attention, as the benefits, e.g., better crime prevention or an optimized resource planning are essential goals for law enforcement agencies. Commercial predictive policing systems commonly visualize predictions on maps but provide only little support for human analysts in the technical and methodological processes that constitute corresponding implementations. In this paper, we report on a project of bringing visual analytics to the field of predictive policing. We introduce a process model that includes machine learning as well as visualization and has been developed together with experts from a law enforcement agency. We also showcase a visual analytics tool, called polimaps, that is part of a real-world predictive policing project and implements elements of the proposed process.

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [User Interfaces]: Graphical user interfaces (GUI), User-centered design— I.3.8 [Computer Graphics]: Applications—

1. Introduction

Police forces all over the world are using databases to store information about offenses, e.g., the date and time of the incident, the location, and various other metadata that typically depends on the type of crime. These vast amounts of data are used primarily for spatiotemporal analysis, for example, to identify trends or hotspots [MRH\textsuperscript{10},LMGH15].

Besides standard data analysis tasks that refer to the past, predictive modeling and in consequence predictive policing is gaining more and more attention [MHR\textsuperscript{11},MMT\textsuperscript{14}], as it promises to anticipate offenses or (new) crime hotspots appearing in the future [PMP\textsuperscript{13}]. Consequently, strategic decisions based on predictive policing methods are designated to lead to better resource allocation, and ultimately to a reduction of crime. Despite the fact that there are commercial applications available [Aza18,Pre18], positive effects of their application have not been proven yet [HSH14, MSM\textsuperscript{15}, SHH16]. A major problem of evaluating predictive polic-
ing is that in general there are spatial and temporal correlations of the criminal incidents and the deployment of predictive policing methods. These correlations are caused by effects such as sparse or incomplete data, global influences such as weather, large events or other external influences that are hard to include in the pool of data utilized for a quantitative evaluation. There is no objective indicator of the effectiveness of predictive policing, as a clear and definitive causal relationship has to be identified first, which has not been done convincingly yet. Although, the expected benefits make the development and implementation of such systems attractive for police forces.

This work is part of a predictive policing pilot study of a German law enforcement agency (LEA) on the state level. Analytically, the goal of the study was to develop and deploy a method to predict the risk of a specific offense (domestic burglary) for a set of predefined geographical regions, the so-called prediction areas. The prediction models are created with state of the art data mining methods by the LEA’s analysts, as another goal was to understand the technical and analytical background of the procedures that constitute predictive policing implementations. The subject of this work was to develop and provide a visual interface for the prediction results for different user groups. The interface should enable users the visual inspection of the prediction areas and the assigned risk scores concerning a variety of different datasets, e.g., to contrast the prediction with the spatial distribution of offenses from the last week. This is what we call the context of a prediction.

In the following, we introduce the predictive policing process, which provides a generic platform for similar experiments or applications, not only concerning visualization. Afterward, we describe polimaps, a visualization system designed and implemented specifically for predictive policing, and some of the design implications induced by the predictive policing application problem.

### 2. Predictive Policing Process

Together with the domain experts from the LEA, we did a concept of a simplified, process-oriented view of the application problem, shown in Figure 2. The primary goal of the predictive policing process is to structure the contained methods of data wrangling, machine learning/statistics, and visualization. Most notably, the last two steps of the process include the end-users for visualizing and the eventual deployment of the predictions.

**Figure 2: Predictive Policing Process. Steps illustrated in color,** 1, 4 and 5 profit from visualization in various ways. Steps 2 and 3 are the points to develop and apply methods from visual analytics or aspects of human-centered machine learning. The steps 4 and 5 are executed by local police forces and therefore include the targeted end-users.

**Step 1: Data selection, Collection, and Preparation.** The subject of this step is the selection of the utilized datasets, identify and exploit reliable sources for these datasets, and develop a data preparation process, if needed. Almost any geospatial visualization can help to visually identify correlations, e.g., by overlaying data from police databases that indicate offenses on the map with other datasets from different sources. Naturally, it is also possible to inspect the effects of the data preparation with a visualization of the datasets before and after the data is processed, for example, geographic coding of addresses or point of interests that come without spatial coordinates.

**Step 2 and Step 3: Predictive Modeling and Computation.** In this two steps, the actual model is created (Step 2) and applied to compute predictions in space and time (Step 3). As prediction models are the methodological core of any predictive policing implementation, they can benefit most from augmenting methods, such as human-centered machine learning [SSZ´17]. Reason for that is that there exists no model of criminal incidents that is exhaustive and sufficiently accurate to be useful to a practical extent. Therefore, the incorporation of expert knowledge to select appropriate models or model parameters can play an important role in the modeling phase.

**Step 4: Prediction Visualization.** Supporting this step is the core of any visualization system that should be used in the context of predictive policing, although the visualization techniques depend on the actual model and prediction subject. Prediction results, in our case the predicted risk scores, could be visualized together with the corresponding prediction area. As it is subject to our work, adding visualizations with data from the past helps to create a visual context of a prediction for plausibility checks, or further model adjustments. Additionally, polimaps implements an interactive prediction area visualization that allows local police forces to remove or include the prediction areas, based on local expert knowledge or specific strategic plans.

**Step 5: Prediction Deployment.** Visualization systems that support predictive policing should be able to export views, for example to an image, for further use. The exported views could be used as a handout for personnel working on the streets, but also to visually document the prediction of a specific model. Additionally, if possible, a tool supporting this step could be the method for prediction deployment itself, for example on personal devices of police officers. polimaps creates pdf files (or images) that are suitable for such deployment scenarios.
3. polimaps

polimaps is a visualization tool that visually combines area based risk prediction with different datasets, and enables analysts to set them into mutual context. Analysts can combine domain knowledge, e.g., from police officers that know the prediction area(s) very well, with results from the automated risk prediction models and geospatial data, e.g., the location of restaurants and bars, using visualization. To integrate into the predictive policing process, polimaps allows the visual, interactive selection of the prediction areas that should or should not contribute to a single prediction referring to a particular time span, e.g., a week, besides the general geospatial visualization support. The accordingly modified prediction can be exported in various ways, e.g., PNG or PDF, or be saved as a session file to be loaded later by another polimaps user.

The collaboration followed the process proposed by Sedlmair et al. [SMM12]. The precondition and core stages were driven by mutual on-site visits and subsequent discussions, sketching ideas and fast feedback rounds with the domain experts. To define the capabilities of polimaps, we started to identify the desired visualization techniques and must-have features, such as prediction area visualization or hotspot maps for different data sources, in a half day workshop. Technical details and the actual implementation were worked off-site, typically in two to three-month cycles. The first, usable version of polimaps was available after three months and was subsequently extended and tested in the target environment by the analysts from the LEA. To support the feedback process, we established a build environment that produced over ten months 46 stand-alone, runnable distributions of polimaps that were subsequently shared with our collaboration partners.

3.1. Tasks, Goals, and Users

In discussion with analysts from the LEA, we identified two fundamental tasks that polimaps is required to support.

Prediction Visualization & Prediction Area Selection. The primary goal of polimaps is to allow analysts to inspect the prediction outcomes visually. To do so, we map the risk scores onto the fill of the related prediction area with a set of sequential color maps [HB03] the analyst can choose from. The result is a classic choropleth map indicating the predicted risk over the set of all prediction areas. The lowest layer that is the foundation for all visualizations in polimaps is fixed to a map visualization that, by default, displays OpenStreetMap tiles [Ope17], and is capable of offline tile rendering. Users are free to choose a virtually unlimited number of layers on top of the map and to combine them arbitrarily. This task refers to Step 4 of the predictive policing process (Figure 2). The prediction visualization is interactive. Each polygon in the choropleth map (prediction area) can be removed or restored from the visualization with a single click. This task is part of supporting Step 4 and Step 5 of the predictive policing process (Figure 2).

Visualization of Prediction Context. The creation of the context for a prediction is supported by two different visualization techniques. First, the visualization of spatial primitives, and second heat maps. The visualization of spatial primitives, such as points, rectangles or lines, supports custom fills, strokes, and some other parameters, e.g., custom labels of the primitives, contributing to their appearance (see Figure 1 E). The heat map, based on kernel density estimation [SiI86], is capable of showing point densities or the aggregated value of an attribute and can be normalized with any data value that is available in the dataset. Therefore, polimaps is capable of contrasting a prediction visualization with a classic hotspot map [CTU08] (see Figure 1 E). Besides hotspot mapping, the heat maps can also be used to visualize information such as the population density or average household income as available from commercial data providers. This task refers to Step 4 of the predictive policing process (Figure 2), but also fits the deployment scenario in Step 5.

An important goal was to keep polimaps task specific and not become too broad in terms of the contained feature set and the corresponding user interface in order to make sure, that all potential users can work with polimaps. The user-centered approach of polimaps was ensured by involving police officers and data analysts with different backgrounds in the collaboration. Also, with new features and parameters, there was always a discussion about their necessity, appropriate default parameter settings, and overall suitability for the targeted application. All features available via the graphical user interface are tailored to one of the goals and use a task specific terminology, which was also subject to subsequent discussions with the domain experts.

The tasks span over three different user groups. The prediction area visualization can be of use for analysts that compute the prediction (our collaborators), analysts from local police departments that prepare the predictions for distribution, as well as police officers on the street for informational purposes. The prediction area selection is tailored to analysts that prepare the predictions for use in the field, as they typically have the expert knowledge to exclude or include prediction areas. It is also possible that strategic reasons contribute to the decision to select prediction areas, which typically follow a local agenda. The visualization of the prediction context is of interest for all involved users, the state-level, and local analysts as well as police officers, as it can be useful to judge the predictions in the light of crime hotspots of the past.

3.2. User-motivated Features

During development of polimaps, some of the features to include were directly motivated by feedback from our collaboration partners or end-users from local police forces. In the following, we elaborate on two noteworthy functionalities of polimaps, as well as some of the involved design decisions.

polimaps integrates an interactive data aggregation feature, called Info-Mode, to facilitate analytical reasoning in the spirit of Thomas and Cook [TC05]. If this mode is enabled, a rectangular area can be selected by the user, see Figure 1 D. Simultaneously, on the left Figure 1 B, aggregated information from the spatial selection is shown and updated whenever the selected area is changed. For standard datasets that are known to come from the police databases, polimaps contains defaults created by data analysis experts from the LEA to select potentially interesting attributes, as well as their order in the info mode display. For standard data sources and datasets, meaningful aggregated statistics are shown without requiring the user to intervene. Still, the included data attributes and their display order can be adjusted in a visual interface, if required.
were important goals during the development of the tool, based
on Lewis [Lew95] and Lund [Lun01]. The third area contained
questions about the general visualization capabilities. Lastly, we had
a section dedicated to the heat map (hotspot) visualization to learn
whether the users think they understood the visualization technique
and the available parameters such as kernel sizes, interpolation areas
or available color scales. The questionnaire had been answered by
six users of polimaps, two data scientists/analysts and four trained
police officers.

The heat map/density map view is judged as useful; the users
have the feeling that they understood the parameters, color scales,
and polimaps provides a satisfactory level of performance. Concern-
ing the available techniques in general, we see potential for further
enhancements of the Info-Mode. The answers to the usability ques-
tions indicate weaknesses and potential for improvement with the
supported tasks, the overall judgment in this aspect shows at least
a satisfactory level for four out of the six users. Although, the most
potential for improvement seems to be in the aspect of learnability,
which has to be investigated in more detail. polimaps got a good
overall assessment. Still, there are aspects such as memorability,
the learning curve, dialogs and indicators that need to be improved.

5. Future Work and Concluding Remarks

In discussions with the end users, it was always clear that there are
two major data dimensions that the analysts are interested in: space
and time. The visualization capabilities of polimaps are tailored to
the spatial dimension of the data; the filter facilities give practical
support to work with the time dimension as well. Therefore, the next
step is to extend the available visualization techniques to support the
combination of these two dimensions, for example, by adjusting the
visual mapping (fill color, value, maybe texture) accordingly, or by
animating the data along the time dimension. Encouraged by recent
findings in a similar application domain [JSM*17, JSS*18], we want
to integrate abstract visualization techniques, for example, based on
dimension reduction techniques such as time curves [BSH*16], or
link the current visualizations with data projection views [JSS*18]
to reveal clusters in space and time. We see potential in extending
the Info-Mode, e.g., by automatic selection of data attributes based
on data characteristics in the current spatial selection. Finally, we
want to examine the areas of improvement highlighted in Section 4
in more detail to identify the potential for further refinements.

polimaps is a success story. The tool enables state-level analysts,
data analysts of local police departments and field personnel such as
police officers working on the streets to utilize data visualization to
enhance their work routines. During development, as well as after
deployment of polimaps, We experienced tremendous demand for
such highly tailored, application specific visualization solutions. The
developed predictive policing process is used by our partners beyond
our collaboration, which illustrates the impact of the presented
research in the application domain further. The informal response
concerning polimaps was very positive, from the state-level LEA,
but also from local police departments. Simple deployment as a ZIP
file, sensible defaults, a task-driven user interface and department
and data specific defaults contribute to an exceptional adoption. To
the time of writing, polimaps is in the process of state-level approval
to be used by any of the more than 40,000 employees of the LEA
and its local departments.

Table 1: Overview of the questionnaire consisting of four categories.

Two months after the final version of polimaps has been delivered
and used in practice, we handed out a questionnaire with four areas
of questions, see the overview in Table 1. The first and second
area contains questions about the learnability and usability, as this
were important goals during the development of the tool, based