When individual data points matter:
Interactively analysing classification landscapes

Bruno Schneider, Sebastian Mittelstädt and Daniel Keim

University of Konstanz, Germany

Abstract
The selection of classification models among several options with similar accuracy cannot be done through purely automated methods, and especially in scenarios in which the cost of misclassified instances is crucial, such as criminal intelligence analysis. To tackle this problem and illustrate our ideas, we developed a prototype for the visualization and comparison of classification landscapes. In our system, the same data is given to different classification models. Classification landscapes are shown in the scatter plots, together with their geographical location on a map and detailed textual description for each data record. To enhance model comparison, we implemented interactive anchor-points selection in classification landscapes. Using those anchors, the user can manipulate and reprojection the model results in order to get more comparable landscapes. We provided a use case with crime data, for crime intelligence analysis.

Main features
- Anchor-points interactive selection
- Linkage of textual, geographic and projected classification data

Anchor-points interactive selection
Each color in the scatter plots represents one different class predicted by the classifier, and each plot represents the classification output produced by the same model with distinct parameter setting. In the example above, by interacting with the scatter plots the user can select anchor points and move them in order to get more comparable landscapes.

Linkage of textual, geographic and projected classification data
The scatter plot panels are updated together after selecting and dragging the anchor points in any of them. The user can select the anchors based on his/her domain knowledge, because a map with the geographical location of each crime record and a text table with the details for each record are linked with the other panels. Also, the linked panels support the inspection of how an individual data point was classified by each model and give insights about where to drag anchor points.