

When individual data points matter: Interactively analysing classification landscapes

Bruno Schneider, Sebastian Mittelstädt and Daniel Keim

University of Konstanz, Germany

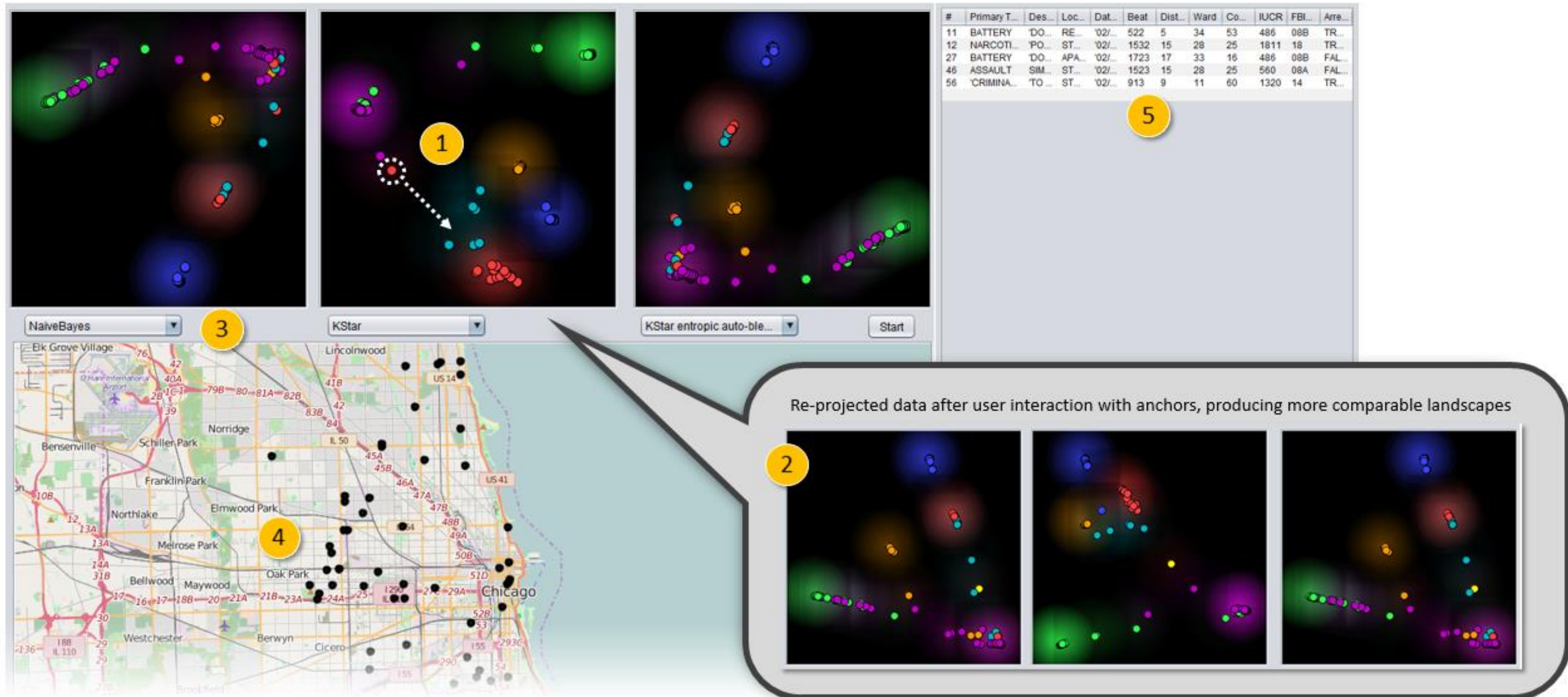


Figure 01: Our system provides visual model comparison by (1) showing side-by-side the classification landscapes of different models in linked panels after applying MDS projection on the probabilities estimates for each class of each data instance, (2) enabling interactive anchor-points selection, dragging and data reprojection for producing more comparable landscapes, (3) allowing the selection of other previously evaluated models, (4) letting the user collaborate with domain knowledge through direct selection of geo-referenced events on a map, and also (5) giving detailed information of each data instance in a text table.

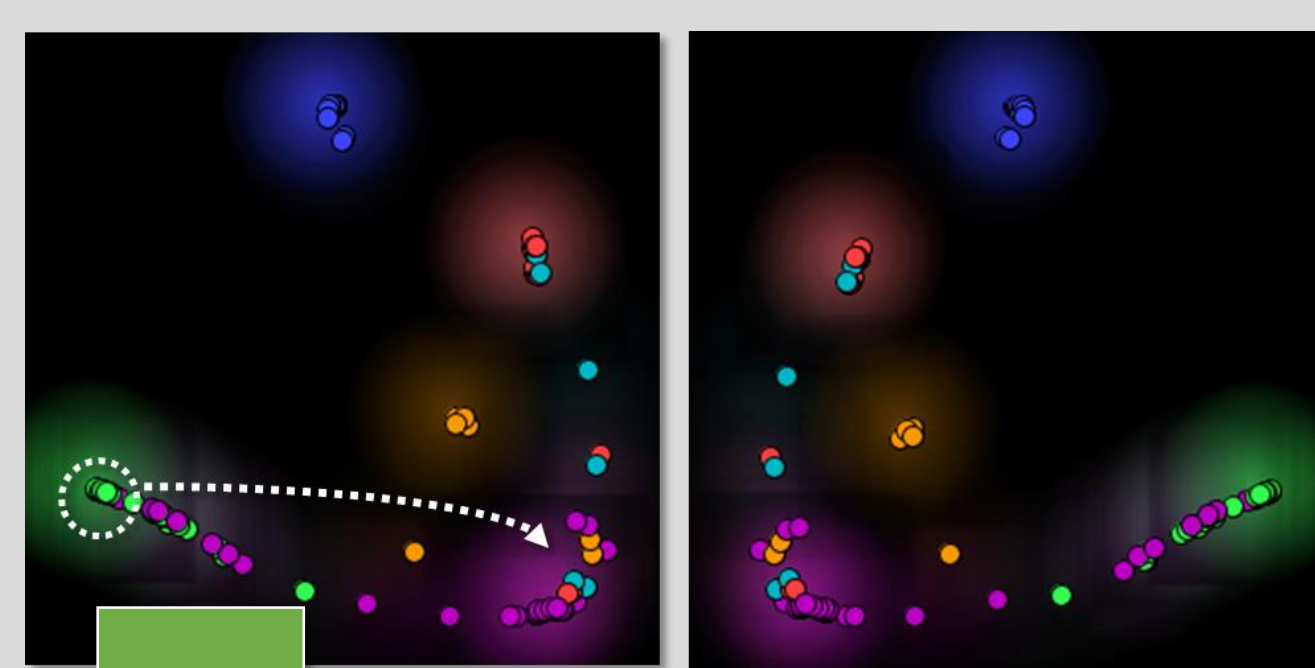
Abstract

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. 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 reproject the model results in order to get more comparable classification landscapes. We provided a use case with crime data, for crime intelligence analysis.

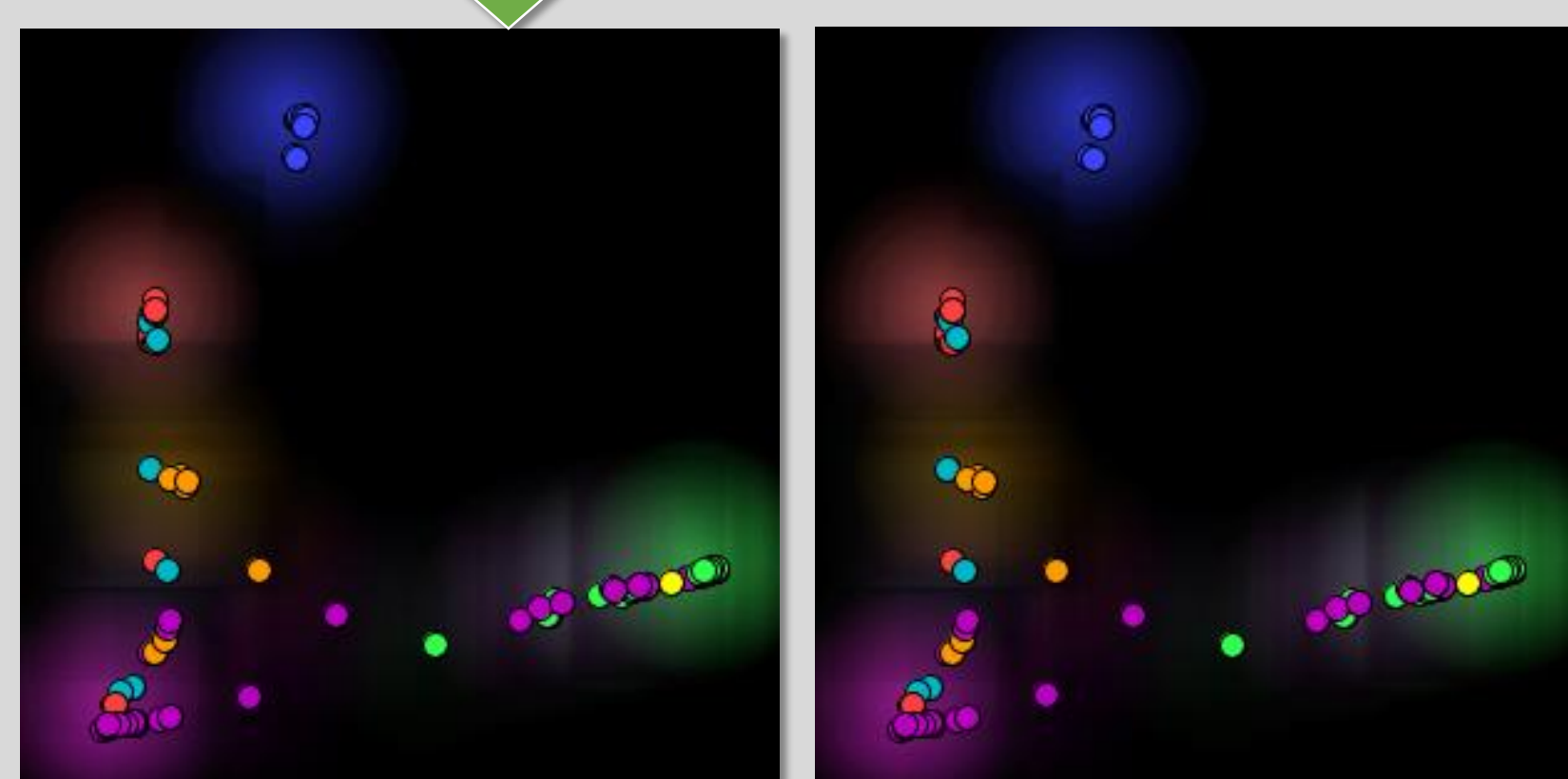
Main features

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.

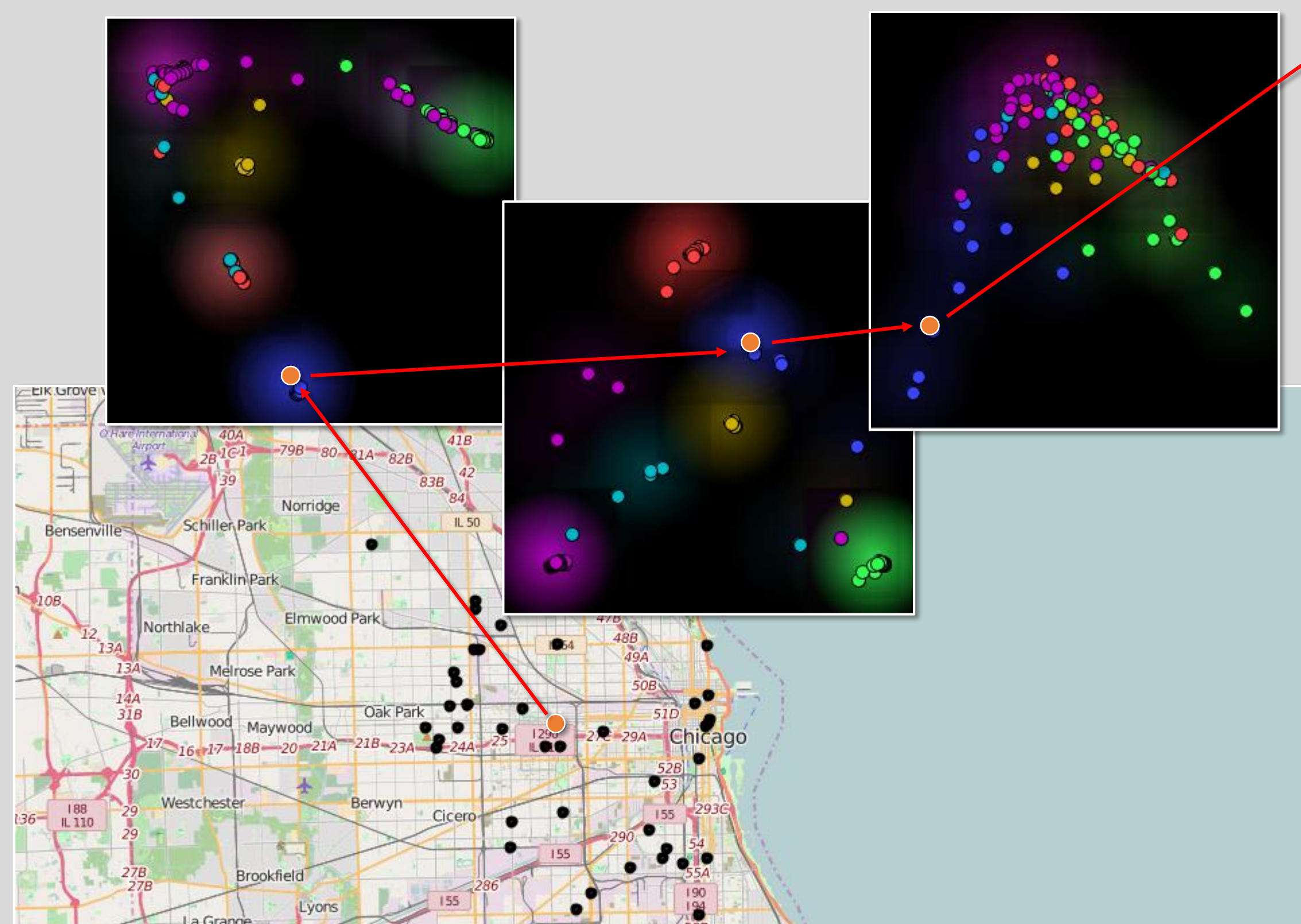


Updated projections after user interaction



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.



#	Primary T...	Des...	Loc...	Dat...	Beat	Dist...
11	BATTERY	DO. RE.	'02/	522	5	
12	NARCOTI...	PO. ST.	'02/	1532	15	
27	BATTERY	DO. APA.	'02/	1723	17	
46	ASSAULT	SIM. ST.	'02/	1523	15	
56	'CRIMINA...	TO. ST.	'02/	913	9	