

Exploring biases in spatial clustering algorithms for crime mapping in Milwaukee

Background

Machine learning has been increasingly used to make policing decisions

Municipal police departments use crime mapping to attempt to predict where crime will occur

Clustering algorithms are used to inform police parol routes

Criminologists don't understand how this software works or how to use it, but the results have huge implications

Comparing Clustering Algorithms



K Means clustering, the algorithm currently utilized by the Milwaukee Police Department. K centroids are randomly placed. Each point is assigned to the closest centroid, and then the position of the centroid is recalculated. This process repeats until the centroids no longer move. Params: K (number of clusters), distance metric

Laura Schultz

Department of Computer Science Mills College lschultz@mills.edu

Process

We scraped crime data from the Milwaukee Municipal Court website

Address data was sanitized so that it could be interpreted by the Google Maps API

Address data was geocoded to obtain the longitude and latitude of each data point

Using geocodes, this data could be plotted and clustered

Agglomerative clustering, a hierarchical algorithm. Each point begins as its own cluster. In each iteration, the two closest clusters are merged. This continues until there are a specified amount of clusters.

Params: number of clusters, cluster proximity metric (how distance between two clusters is calculated), distance metric

DBSCAN, an algorithm that clusters based on density. Points are designated as core points if they have at least *minPts* points within a radius of *Eps*. Core points are grouped together to form clusters. Params: Eps (the radius of a point's neighborhood), minPts (number of points in neighborhood of a core point), distance metric

Dr. Shion Guha

Department of Mathematics, Statistics & Computer Science Marquette University shion.guha@marquette.edu

give different results

crime database:

Clustering this biased data results in large, high-density clusters in black communities Police department will allocate more officers to this area Cycle of overpolicing continues

Each of the many different clustering algorithms provide a wide array of results

The current algorithm used by the Milwaukee police department may not be best for the situation

Building a better predictive crime model requires acknowledging and correcting bias

Acknowledgements

Special thanks to my team members: Dominique Green, Quinci Henry, Chris Supinger, and Katy Weathington. This work was sponsored in part by NSF #ACI-1461264.

Biases

Changing the parameters of a clustering algorithm will

Additionally, certain communities are oversampled in the

Takeaways