Crime Hotspot Detection For Interesting Patterns

Anubhav Panda
panda047@umn.edu
5509727

Sanjit Dash
dash0030@umn.edu
5509721
Problem Statement

- **Input:**
  - A set of crime location
  - The date and time of their occurrence
  - Type of Crime and their respective frequency

- **Output:**
  - Hotspot patterns for associated crimes

- **Goal:**
  - Finding out important features in the dataset
  - To find the interesting patterns in crime(association)
  - Combine the Associated crimes using Neighbor Graph
  - Find the Minimal Enclosing Circle of the Neighbor graph
  - Compute the hotspot using above two features
Problem Statement

Traditional Algorithm

A’s hotspot

B’s hotspot

Proposed Algorithm

- B associated with A
- C is eliminated
Challenges

- Difficulty of enumerating all potential hotspot areas. (computational complexity)
- Testing for statistical significance to reduce chance patterns.
- Find the patterns of crime occurrence within a certain time interval and radius
- Hyper-parameter tuning
- Non-existent addresses caused by typographical error.
- Address duplication problems that are caused by dozens of streets with the same name across a city.
- SaTScan for Polygons
Impact: Societal Importance

- Researchers and analyst can develop interesting crime theories based on hotspot of associated patterns.
- Detecting crime hotspots to focus the deployment of police enforcement.
- Can create societal awareness about the frequent crimes in a locality.
- Predicting the potential residence of a serial criminal.
Current Methods

- Spatial scan statistics and its applications (i.e. SaTScan)
- Naïve Ring Shaped Hotspot Detection Algorithm
- Linear Hotspot Detection (use for the crime in road-networks)
Proposed Approach

- Crime Dataset: NYC Crime Dataset
- Clean the Data set: Remove redundant rows and rows having missing features
- Exploratory Data Analysis: To find interesting features
- Identify Associated Patterns: Apriori algorithm
- Compute Neighbourhood Graph of associated Crimes
- Compute Minimal Enclosing Circles and its center
- Apply hotspot detection algorithms (SaTScan) for circles
Toy Example

Crime data in a certain location.
Apriori Algorithm

- It is used for generating association rules from datasets
- It is based on the principle “subsets of frequent itemsets are also frequent itemsets”
- **Input:** set of crimes with Borough, premise Description, Support, Confidence, Max Length
- **Output:** Associated crimes
- Crime A associated with Crime B
Neighbor Graph

- A neighbor graph needs to be computed using BFS and borough information.
- It considers all the associated crimes that are present within a certain threshold distance.
Minimal Enclosing Circle

- Find Minimal Enclosing Circle (MEC) using Welzl’s Algorithm
- **Input:** A set of points with X and Y coordinates
- **Output:** Center and radius of MEC
- It has Time and Space complexity of $O(n)$
SatScan

- It is used to measure spatial, temporal or space-time clusters using a spatial scan approach
- **Input:** Set of crimes with x and y coordinates
- **Output:** Hotspot
- **Constraints:** It works for points only
Exploratory Data Analysis
Exploratory Data Analysis
Results (Associated Crimes)

- A transaction consists of borough name, premise description, offense description
- Grand Larceny and Burglary in Queens and Bronx (Near ATM Location)
- DANGEROUS_DRUGS and CRIMINAL_MISCHIEF_AND_RELATED_OF in Brooklyn and Manhattan (Near Cemetery)

\[
\begin{array}{lllll}
\text{lhs} & \text{rhs} & \text{support} & \text{confidence} & \text{lift} \text{ count} \\
1 \{ & \text{Borough=QUEENS, premiseDescription=ATM} & \Rightarrow \{\text{offenseDescription=GRAND_LARCENY}\} & 4.695076e-05 & 0.8461538 \ 9.42514 & 22 \\
1 \{ & \text{Borough=QUEENS, premiseDescription=ATM} & \Rightarrow \{\text{offenseDescription=BURGLARY}\} & 1.067063e-05 & 1.00 \ 31.2384 & 5 \\
\end{array}
\]
Results

SATScan of associated crimes in Queens (Proposed Method)

SATscan of all crimes in Queens
Future Work

- Implement this approach with Colocation patterns
- Replace Minimum Enclosing Circle with Convex hull Approach
- Extended this work for other city Datasets (LA, Boston, etc)
- Extend this approach for Linear Hotspot Detection Algorithms