Spatio-temporal clustering methods

AUTHORS: MATEJ SENOŽETNIK, LUKA BRADEŠKO, BLAŽ KAŽIČ, DUNJA MLADENIĆ, TINE ŠUBIC

LJUBLJANA, 10/10/2016
Motivation

- Analyzing user paths or behavior, measuring traffic congestion
- Application which can track coordinates from GPS
- Raw data doesn’t provide any useful information
- Stay points
Motivation (2)

- Process is called clustering
- Detects: Frequently visited locations, next place prediction
Outline

• Type of clustering data
  • Appropriate for clustering spatio-temporal data?
• Comparison of algorithms
• Find the most appropriate algorithm for clustering spatio-temporal data
Type of clustering data

Clustering methods are, in general, separated into the following categories:

• Partitioning methods
• Hierarchical methods
• Density-based methods
## Algorithms: A comparison

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Spatio-temporal</th>
<th>Noise sensitivity</th>
<th>Returning stay points/ paths</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBSCAN [2]</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>ST-DBSCAN [5]</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>SMoT [1]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CB-SMoT [6]</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>SPD [7]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>OPTICS [8]</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
Density Based Spatial Clustering of Application with Noise (DBSCAN)

Density-based clustering algorithm which identifies arbitrary-shaped objects and detects noise in a dataset

Advantages:
- Robustly detects outliers
- Appropriate for large databases

Drawbacks:
- Works on spatial elements
- Doesn’t work on different densities.
Density Based Spatial Clustering of Application with Noise (DBSCAN (2))

Original Points [3]

Clusters [3]

Original Points

(MinPts=4, Eps=9.92)

(MinPts=4, Eps=9.75)
Spatio-Temporal Density Based Spatial Clustering of Application with Noise (ST-DBSCAN)

Improves DBSCAN in three ways:

- Takes into account temporal data
- Identifies noisy objects if there are various densities of the input data
- More accurately differentiates adjacent clusters
Ordering Points to Identify the Clustering Structure (OPTICS)

- Finding density-based clusters in spatio-temporal data
- Improves on DBSCAN's biggest weakness, the failure to detect clusters when density of the data varies
Stops and Moves of Trajectory (SMoT)

- Stop candidates $R_{C1}$, $R_{C2}$, $R_{C3}$
Clustering-Based Stops and Moves of Trajectory (CB-SMoT)

• SMoT drawback is incapable of detecting stay points that are not predefined by user
• Idea behind of this method:
  • We move slower than when we are traveling from one place to another
• Less incorrect stop compared to SMoT algorithm
Stay Point Detection (SPD)

Based on time and distance thresholds.

Advantages:
• Not need of predefined structures
• Computationally inexpensive

Drawbacks:
• Sensitive to noise (can be partially reduced by adjusting the parameters)
Discussion

- Algorithm needs to have the following properties:
  - Is able to cluster spatio-temporal data
  - Is noise insensitive
  - Returns stay points and paths

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Spatio-temporal</th>
<th>Noise sensitivity</th>
<th>Returning stay points/ paths</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBSCAN</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>ST-DBSCAN</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>SMoT</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CB-SMoT</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>SPD</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>OPTICS</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
SMoT vs. CB-SMoT vs. SPD

- **SMoT:**
  - Uses predefined regions

- **CB-SMoT:**
  - Predefined ratio of stay points and paths

- **SPD:**
  - Sensitive to noise
  - False stay points/paths
Conclusion

• SPD algorithm
  • Suffers from detecting false stay points and paths
  • This problem can be alleviated by running multiple iterations of the algorithm on resulting dataset.

• Insensitive to noise (Wi-Fi, activity recognition)
Thank you for your attention!
References


