Visual Analysis of Origin/Destination Time Patterns of Travellers

Katre Karja

Supervised by Dr. Jahnke Mathias

Thesis Defence Presentation

2016

Outline

- Introduction
- Background
- Methods
- Results
- Conclusion

Introduction

Objective of Thesis

- Taxi FCD to detect popular places in the city
  - High density O/D hotspots on street scale
- Identify places near the detected hotspots
- Exploring temporal patterns for these hotspots
- With the support of created visualisation application

Background

Clustering Methods

Partitioning	**Hierarchical**	**Density-based**
Distance based	*Distance based*	*Density based*
• Knowledge of resulting number of clusters • Assigns all points into clusters • Spherical-shaped cluster • (*k*-means) (Agrawal et. al, 1998)	• Hierarchical decomposition • Compact cluster • Also clusters with only a few points • Spherical-shaped cluster • (AGNES)	• Detects noise points • Detects natural shaped clusters • (DBSCAN) (Esther et al. 1996)

Visualising Spatio-Temporal Data

- Complexity
  - Multidimensional
  - dynamic (time)

- Classification (Andrienko et al.,2003)
  - Existential changes
  - Spatial changes
  - Thematic changes

Visualising Spatio-Temporal Data

- Methods
  - Different dimensional representations (3D, 4D, 2D)
  - Multiple views for different dimensions
  - Interactivity
  - Animations
  - Plots of different time moments

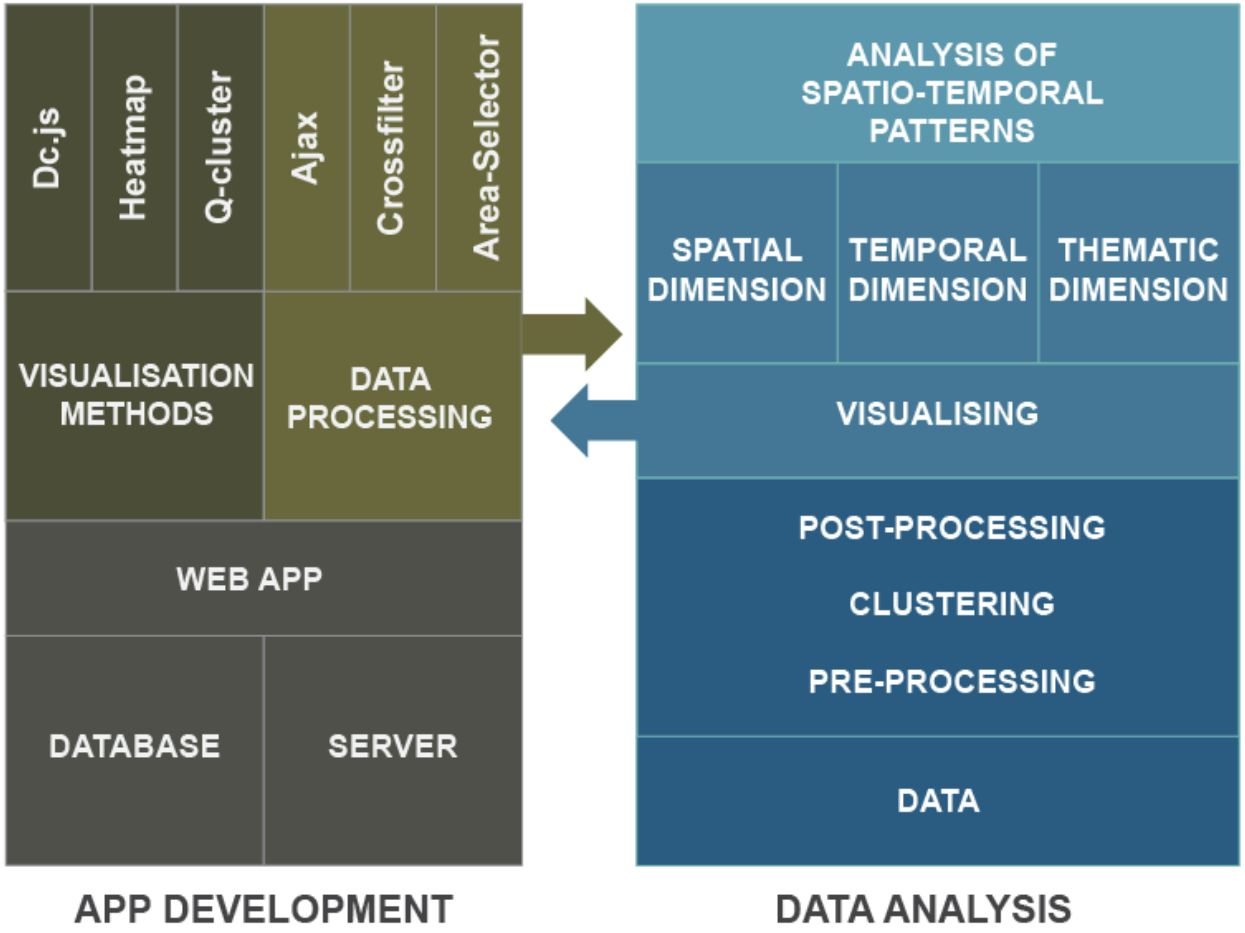
Visualising Spatio-Temporal Data

- Interactivity
  - Focusing
  - Linking views
  - Linked brushing
  - On-demand-visualisation
  - Scale-based data sub-selection
  - Time sliders

(Andrienko et al., 2003),
(Resch, 2013)

Methods

Overview of the Study



Data and Pre-Processing

- Database
  - Importing
  - indexing
  - Sorting
  - Cleaning
    - filtering
  - O/D extraction


```
{
  "_id" : ObjectId("53319c066536032l7fbdd010"),
  "timestamp" : ISODate("2010-05-22T04:49:16.000Z"),
  "car_id" : 10003,
  "lon" : 121.4239120,
  "lat" : 31.3497069,
  "car_status" : 0
}
```


Figure 8 The relevant variables for this study

Clustering

- High intensity areas on street level

- Spatial clustering
- 3-hour interval basis data collections

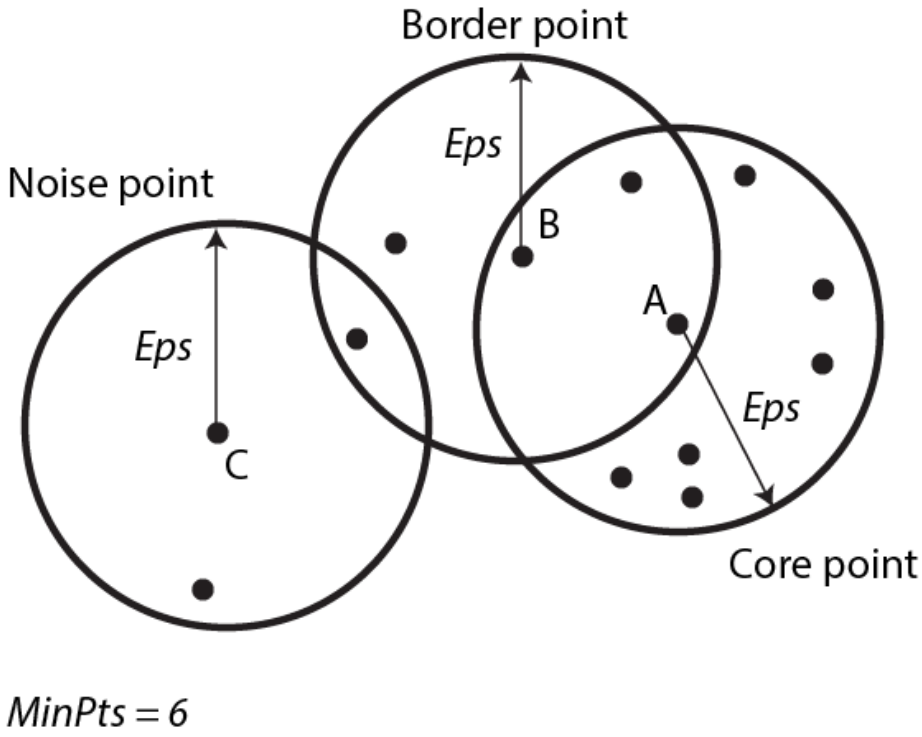
- Each interval with different distribution
    - Different number of resulting clusters expected
- Only a subset of data items could form clusters
    - High density areas

DBSCAN (Esther et al. 1996)

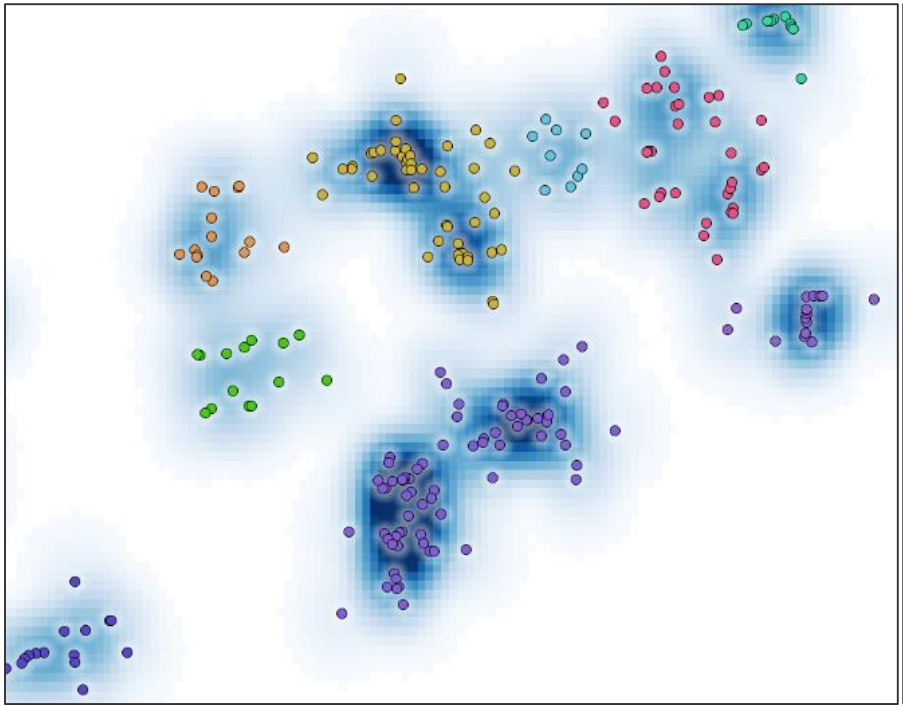
Density-based clustering algorithm DBSCAN

- *Eps* and *MinPts*

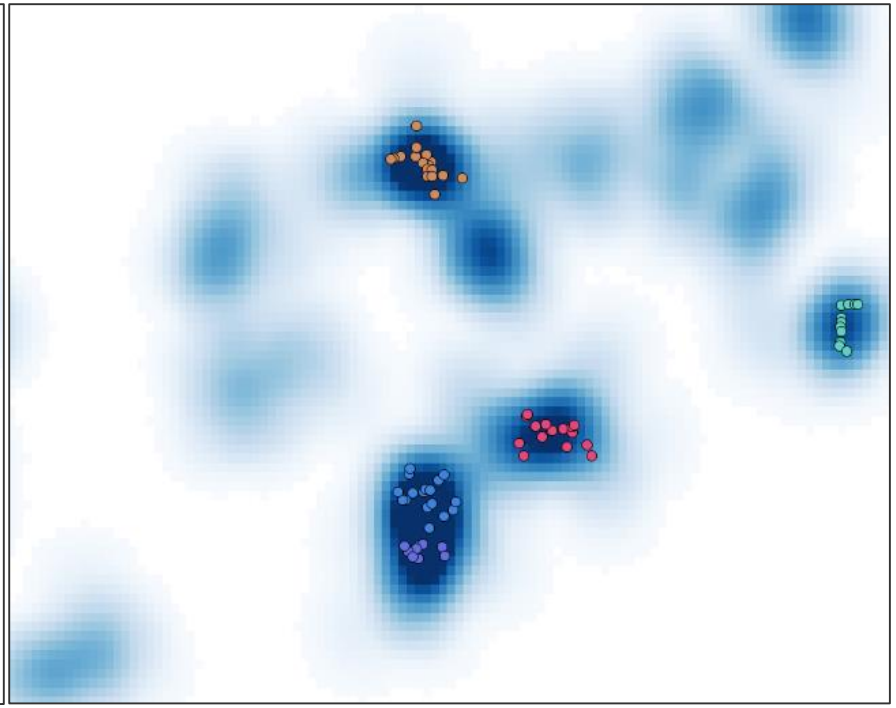
eps	*MinPts*
300 m	10
200 m	10
100 m	10
25 m	6



Parameter tuning for DBSCAN



a) *eps* = 200m, *MinPts*=10



b) *eps* = 100m, *MinPts*=10

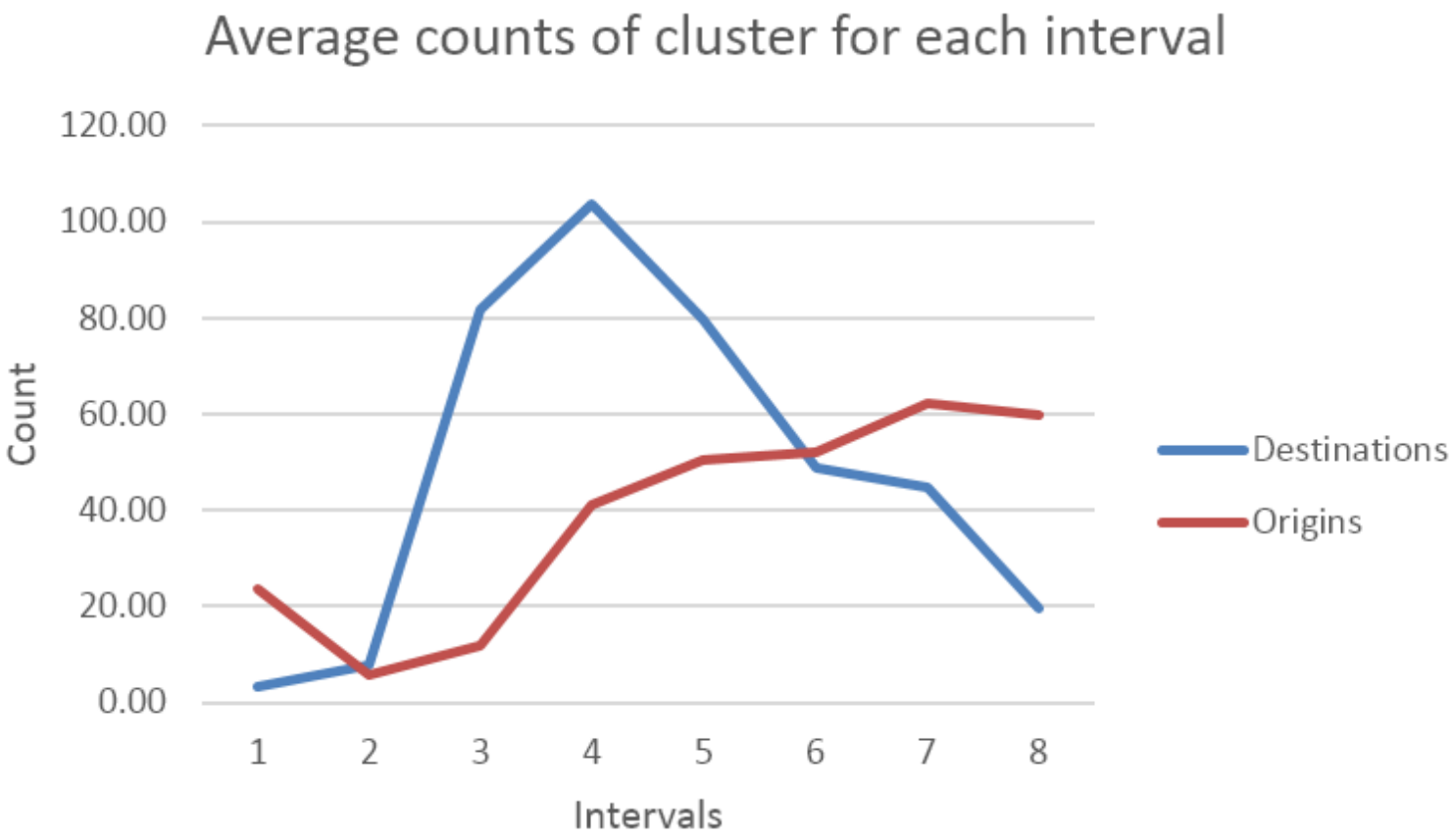
Parameter tuning for DBSCAN

Density-based clustering algorithm DBSCAN

- *Eps* and *MinPts*

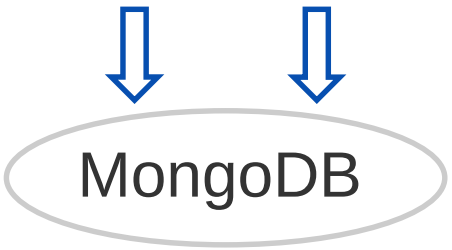
eps	*MinPts*	scale
300 m	10	Blocks level
200 m	10	Blocks level
100 m	10	Blocks level
25 m	6	Street level

Results


Average counts of cluster for each interval
120.00
100.00
80.00
60.00
40.00
20.00
0.00
Count
1
2
3
4
5
6
7
8
Intervals
Destinations
Origins

Post-processing and Semantic tagging

- Joining clustering results
- POIs from OpenStreetMap + Internet search



- Geo-queries and semantic tagging

Methods

APP DEVELOPMENT

Square

Crossfilter

Fast Multidimensional Filtering for Coordinated Views

Leaflet

dc

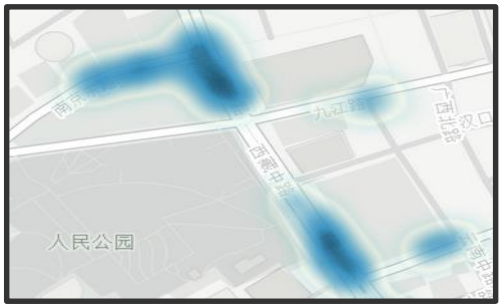
mongoDB

Flask

web development, one drop at a time

Map Views

- The locations of hotspots
  - Heat layer



- The proportions of tagged POIs for each hotspot
  - Donut symbols with counts

- The proportions of occurring time intervals for each hotspot
  - Pie-charts

Case Study: Shanghai

- Time Period
  - Seven days of a week 17.05.2010 – 23.05.2010

- Four sample areas
  I. Shanghai New International Expo Centre (NIEC)
  II. Shopping street West Nanjing Road
  III. Shanghai Exhibition Center
  IV. Century Commercial and Trade Plaza

The visualisation application

Results

- Different number of clusters for O/D
- The Application enables
  - Visual queries on three dimensions
    - Spatial, temporal, thematic
  - Exploration in different zoom levels
    - Map symbols adapt to zoom levels
  - Filtering:
    - Time (days, hours, intervals)
    - Space (selected area)
    - Thematic (Districts)
- A demonstration of the use of application with 4 sample areas

Future work

- Future work
  - Addressing strategies for uncertainty issues in semantic labelling and visualisation
  - Detecting unusual hotspots
  - Usability studies

Thank you!

References

- ANDRIENKO, N., ANDRIENKO, G. & GATALSKY, P. 2003. Exploratory spatio-temporal visualization: an analytical review. *Journal of Visual Languages & Computing,* 14**,** 503-541.
- CHANG, H., TAI, Y., CHEN, H. AND HSU, JANE Y. iTaxi: Context-Aware Taxi Demand Hotspots Prediction Using Ontology and Data Mining Approaches. The 13th Conference on Artificial Intelligence and Applications (TAAI 2008), 2008.
- RESCH, B., HILLEN F., REIMER, A., SPITZER, W., 2013. Towards 4D Cartography - Four-dimensional Dynamic Maps for Understanding Spatio-temporal Correlations in Lightning Events. *The Cartographic Journal,* 50**,** 266-275.
- HALLAHAN, N. 2014. *q-cluster* [Online]. [Accessed 16.11 2015].
- AGAFONKIN, V. 2011. *Leaflet an open-source JavaScript library for mobile-friendly interactive maps* [Online]. [Accessed 22.01 2016].
- AGAFONKIN, V. 2014. *Leaflet.heat* [Online]. [Accessed 10.10 2015].
- ESTER, M., KRIEGEL, H.-P., SANDER, J. & XU, X. A density-based algorithm for discovering clusters in large spatial databases with noise. Kdd, 1996. 226-231.