Driving Style Clustering

Data Mining and Text Mining
Project
Spring 2009

Filippo Sironi and Matteo Villa

Exploration

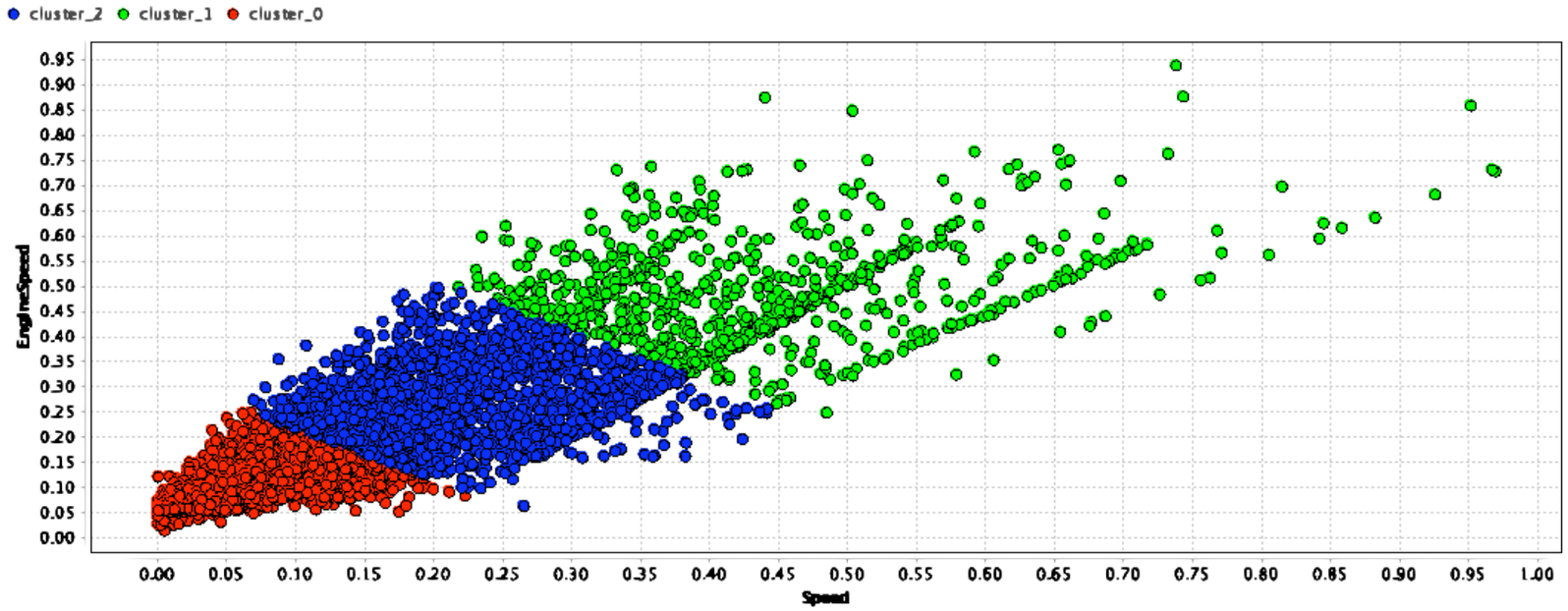
- Find meaningful attributes
  - Speed
  - Engine speed
  - GPS information (latitude, longitude, and date)
- Discard non meaningful attributes
  - Acceleration
  - Torque
  - Fuel consumption
  - GSM information

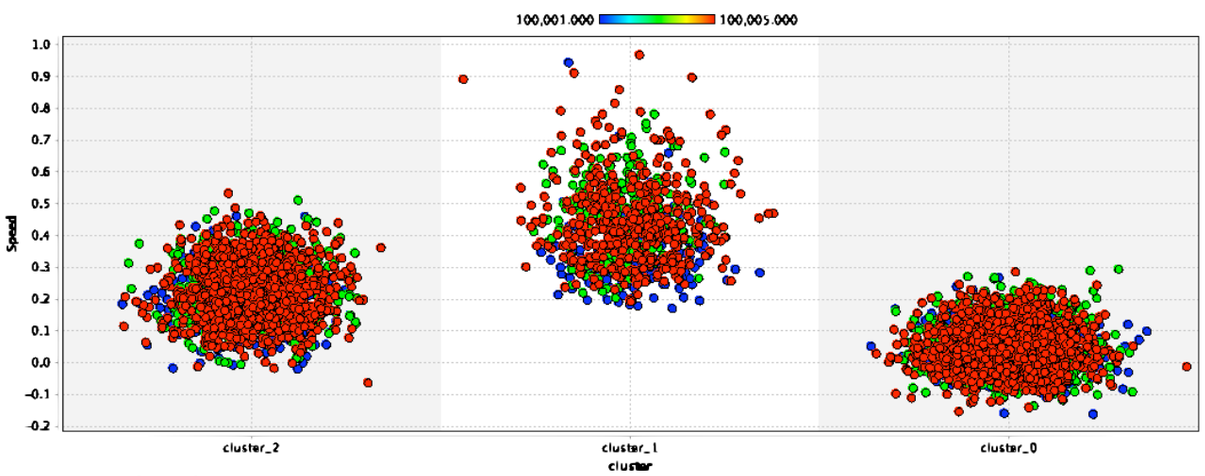
Clean and Preprocess

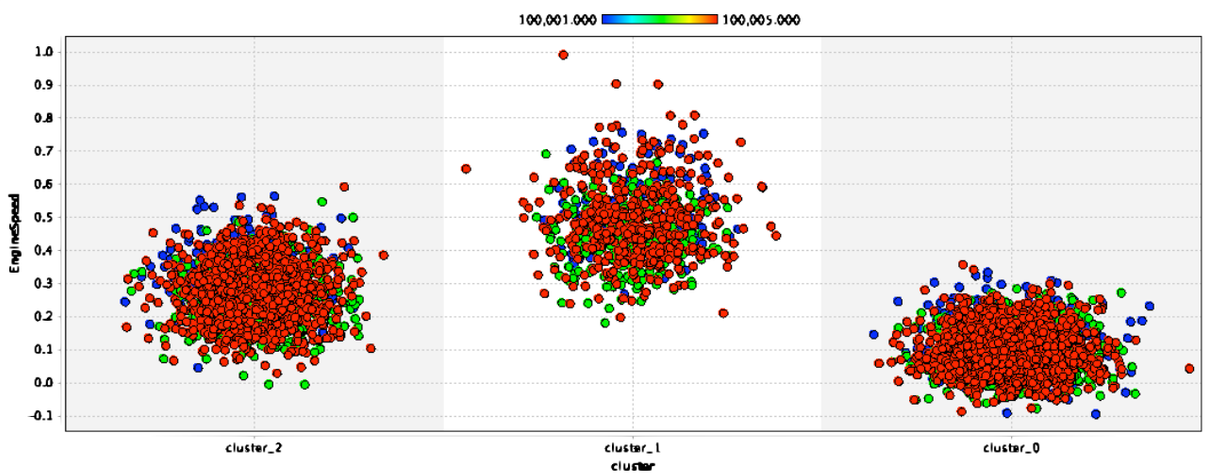
- Get rid of:
  - Duplicates examples
  - Non meaningful attributes
  - Non meaningful values
- Smooth peak values
- Eliminate dependence from time
- Normalize attributes to set equal weights before the clustering process

Mine (1/2)

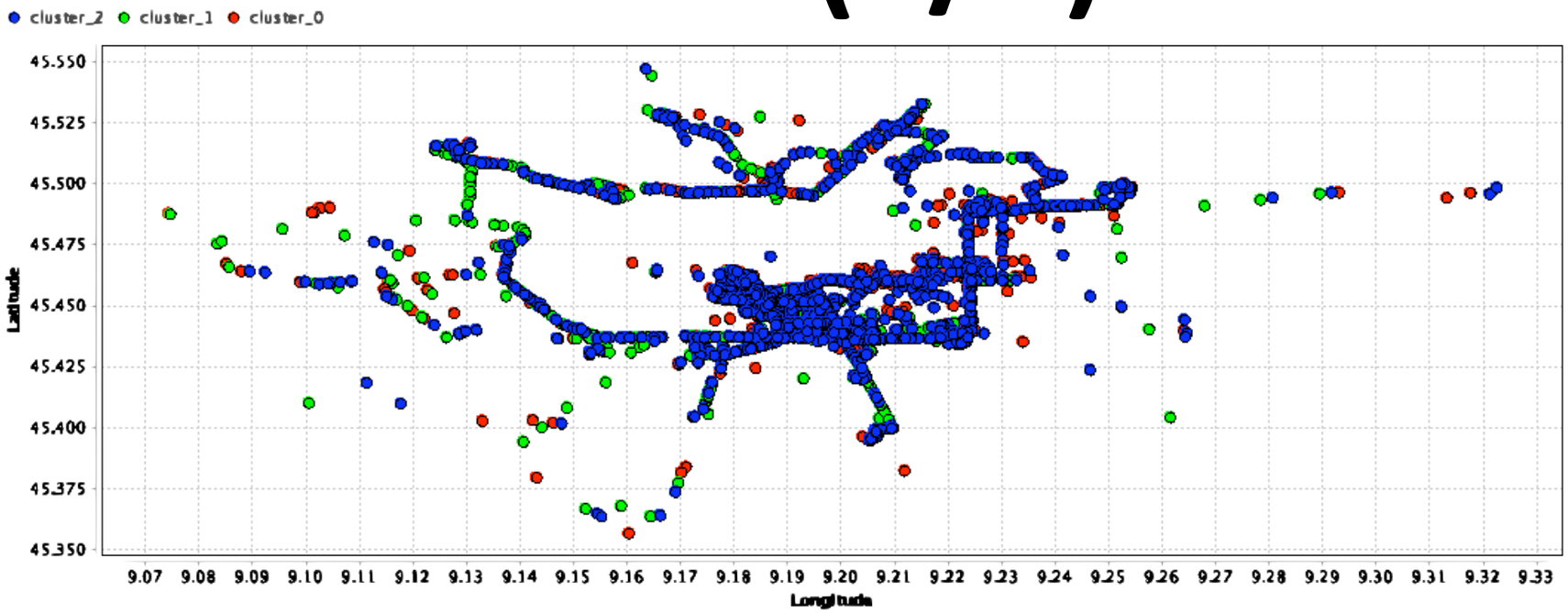
- Apply K-Mean (unsupervised learning) to find clusters

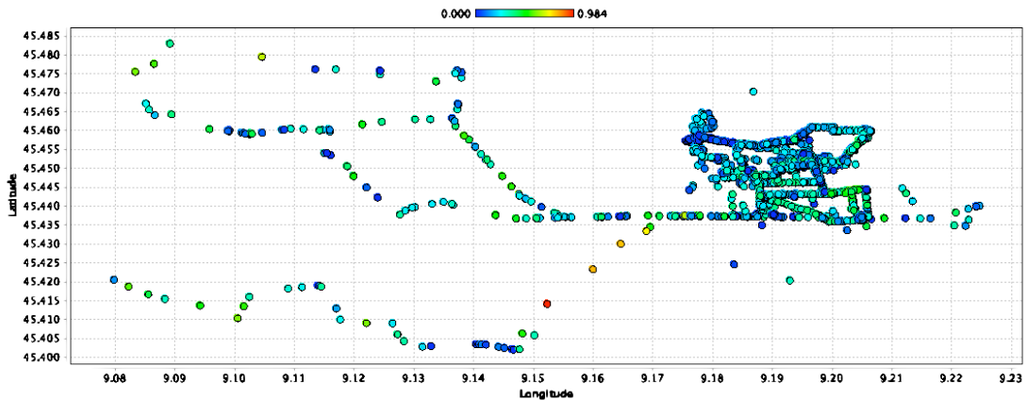


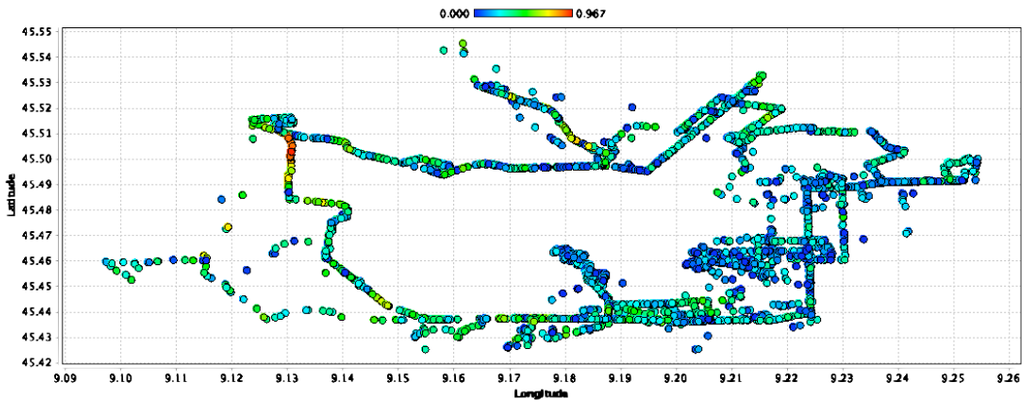


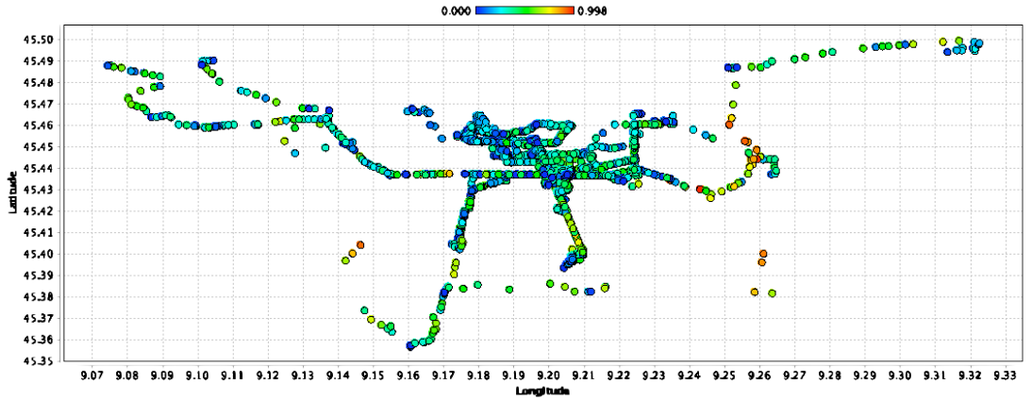


Mine (2/2)


cluster_2
cluster_1
cluster_0
Latitude
Longitude


0.000
0.984
Latitude
Longitude


0.000
0.967
Latitude
Longitude


0.000
0.998
Latitude
Longitude

Conclusions

- Speed and Engine speed are obviously correlated hence diagrams that shows clusters distribution over these attributes are very similar
- Three clusters basically represent a logical subdivision considering the real life
- Every vehicle has an equal distribution over all three clusters
- Driving style is not affected by the position of the vehicle over the city since all clusters are equally spread over GPS coordinates