FYP Week3 Review

Using Beamer

Lenox Low

Multimedia University Cyberjaya Campus

December 1, 2016

Outline

X-Mean - Introduction

It is an algorithm that able to search the space of the cluster locations efficiently and the optimal number of clusters (the value of K). The Bayesian Information Criterion (BIC) or Akaike Information Criterion (AIC) measurement is used for determining the value of K.

X-Mean - Procedures

It goes into action after each run of K-means, making local decisions about which subset of the current centroids should split themselves in order to better fit the data. The splitting decision is done by computing the Bayesian Information Criterion.

X-Mean Procedures(Cont.)

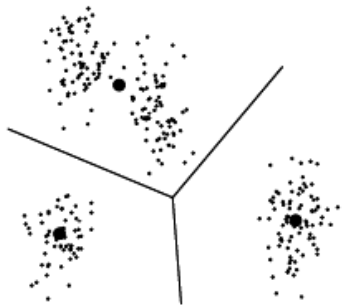

Figure 1. The result of running K-means with three centroids.

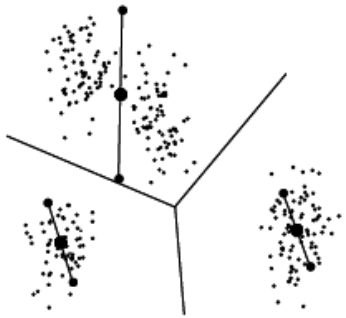

Figure 2. Each original centroid splits into two children.

X-Mean Procedures(Cont.)

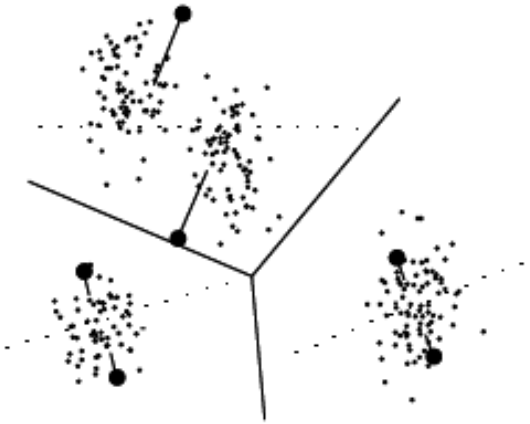

Figure 3: The first step of parallel local 2-means. The line coming out of each centroid shows where it moves to.

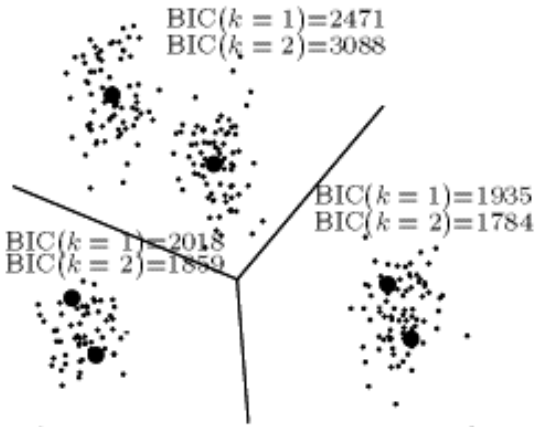

Figure 4: The result after all parallel 2-means have terminated.

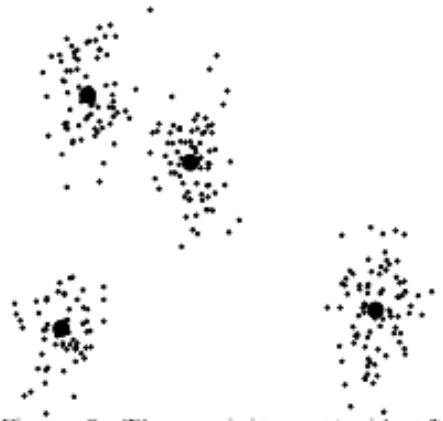

Figure 5: The surviving centroids after all the local model scoring tests.

Data Collection

- Few features is selected and collected using Torque(Lite)
- Currently 9 drivers file is available.

Data Collection (Cont.)

The features include

- GPS coordinate
- Speed (GPS)
- GPS Bearing
- GPS Altitude
- Engine RPM
- Engine Load
- Throttle Position
- Engine Coolant Temperature
- Acceleration Sensor (Total)

Data Preprocessing

- Eliminate the first few rows and last few few rows of the data.
- Road Structure Labelling
  1. Manual label the road structure of a single data set using Google Earth
  2. Develop scripts to label the remaining data sets using Python
  3. Check the result of the scripts using the Tableau.
- Data Fusion

Data Exploring

- Explore the data distribution by using K Means Algorithm
- Find the optimal number of clusters (The result is 2)

Driver Profiling

Propose method

Data Classification

Naive Bayes will be used as Machine Learning Technique to train the classifier

Thank you