

---

ISyE 6416 – Computational Statistics – Spring 2016  
Project: “Big” Data Analytics Proposal

---

- Team Member Names: Yi Wen; Xi Yang; Xue Zhang
- Project Title: Better utilization of the room (meeting room, classroom etc.)
- Project proposal:

- Problem Statement

As the development of information technology, the operation and management efficiency of room reservation has been improved significantly. People can reserve a certain room for a specific time slot in advance which reduces scheduling conflicts and waiting time. In such way, the room utilization seems to be improved, however, it also comes with couple of big drawbacks for the operation. For instance, some room reservation was made in advance, and the length of reservation is also determined at that time, which is not flexible for any changes. Some reservation can be made instantly as for needed if there are available rooms. Most of the room reservation systems was designed into time slot with certain length (e.g 1hr), and the room can only be reserved based on number of time slots. However, the length of a meeting is hardly fitted in to certain amount of time, and has a big variation and effected by many factors, for example, number of participants, contents and progress of the meeting. In order to avoid lack of time, people may easily reserve a time slot that is much longer than they actually need. (e.g The meeting only takes 1hr and 10min to finish, however, for room reservation people may need to reserve the room for 2hrs, and the in the rest 50mins, the room will not be used, and either available for reservation). This causes a lower room utilization and operation efficiency. In order solve this problem, we would like to collect real time information of room occupancy. With this data, the information of room occupancy can be provided in real time, and rooms that are not in used can be available for reservation immediately even though it is reserved under previous reservation.

In this project, we used the data that can be a factor to determine room occupancy, and using several classification methods to develop a effective model for room occupancy determination.

- Data Source

<http://archive.ics.uci.edu/ml/datasets/Occupancy+Detection+>



datatest.txt



datatest2.txt



datatraining.txt

Data attribute:

date time year-month-day hour:minute:second

Temperature, in Celsius

Relative Humidity, %

Light, in Lux

CO2, in ppm

Humidity Ratio, Derived quantity from temperature and relative humidity, in kgwater-vapor/kg-air

Occupancy, 0 or 1, 0 for not occupied, 1 for occupied status

- Methodology

We will explore below methods learnt in class and compare the results to make final recommendation.

- 1) Logistic Regression

Logistic regression is commonly used for classification problems. Since our expected forecast result is binary (1 for occupied, 0 for unoccupied), logistic regression is considered to be an appropriate method to be used. It can also help us study the probability of room occupancy under certain conditions.

- 2) LDA/QDA

Discriminant analysis is a powerful method to characterize or separate 2 or more classes of objects or events. The method works for data that is normally distributed within each class, which is applicable to our case. LDA/QDA also works closely with principle component analysis and factor analysis in that both look for best linear/quadratic combination of variables which best explain the data.

- 3) Random forest

Random forest can also be applied for classification problem. This method will construct a multitude of decision tree and output the mode of the classes (classification). Random forest method can correct for decision trees' habit of overfitting to their training set.

- Expected Result

The outcome of this project is a model with the capability to detect the room occupancy based on the input data (could be collected by monitors or other information technology). So that the status of the room is able to be updated in the reservation system in real time. This will provide a more efficient usage of the room and better fulfillment of customers need.