

Proposal for Machine Learning Project – Apartment Rental Price Prediction

Hao Ge, Zizhuo Liu, Xu Wang

April 14, 2016

1 Motivation

Nowadays, there are millions of students leaving their hometown either internationally or domestically to continue their education. In most cases, they rent apartments around the campus during the academic year. But sometimes students might leave for other places temporarily for exchange or internship. Under these conditions, students might consider to sublease their original apartments. However, it might not be that easy to determine the price for sublease. They have the incentive to increase the rate while a high rate may attract few renters.

In this project, we want to use machine learning method to help students in Chicago area determine a reasonable price to sublease their apartment.

2 Project Plan

To do the task, first we need to acquire data for the apartment rental in Chicago area from Internet. Then we should determine the attributes we want to use in the machine learning algorithm. At last we should find the best machine learning approach to give us the results.

2.1 Data Acquisition

Data acquisition is the most important part in this project. We can find apartment rental data from the following sources

- Download data on Zillow Research (<http://www.zillow.com/research/data/>) or Chicago data portal (<https://data.cityofchicago.org/>)
- Collect those rental data on Craigslist on our own.

2.2 Attributes Determination

There are many attributes that may affect the price for apartment sublease. Apparently any attributes that may affect the apartment rental price are also the attributes for sublease. Besides those, the sublease period, living habits of the subleser may also have impacts on the final price. We will start with some basic attributes and add others if time permits.

2.3 Machine Learning Approach

Initially, we want to try the decision tree method, which we need to implement in the second assignment. It is also possible to use the nearest neighbour method. We will use cross validation to avoid overfitting problem. To evaluate the result, we may compare our results with the price on real estate management website.