Status Report of Machine Learning Project – Apartment Rental Price Prediction

Hao Ge, Zizhuo Liu, Xu Wang

1 INTRODUCTION OF TASK

1.1 The definition of the task

Our task is to help students in Chicago area determine a reasonable price to sublease their apartment or find a sublease via machine learning approach. In this project, the input are attributes of the apartment subleased, along with other factors such as sublease period. A simple illustration is shown in Figure 1.1.
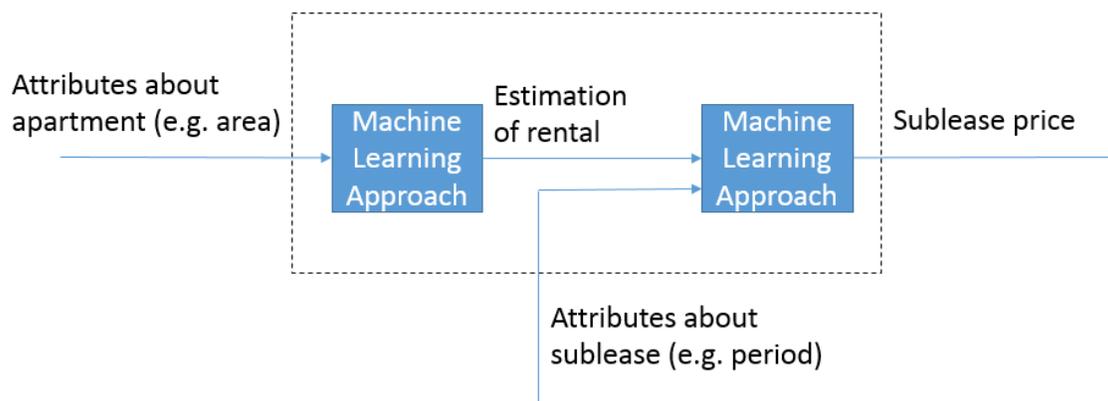

Figure 1: A simple illustration of our task. The inputs are attributes about the apartment and the sublease. The output is the price.

1.2 The meaning of the task

Nowadays especially in the summer, a great number of students might leave for other places temporarily for exchange or internship. In these cases, students often consider to sublease their own apartments and look for a sublease in the city they will move to. Our task is significant for these students, who need to determine the reasonable price and find a suitable sublease.

2 INFORMATION OF DATA-SET

We collect data about rental from Craigslist. A web spider is written to help us find and organize the data on the Craigslist website. Our program can get the house/apartment rental data in a particular area. The attributes we use are listed in Table 1. Since the data on Craigslist may not be complete for each of the posts, we use '?' to denote the missing attributes. We can use the same method in PS2 to deal with the missing attributes. We grab about 4100 set of data from Craigslist for the project in total.

Attribute	Data Type	Attribute Explanation
Price	Float	The listed rental price
Bedroom	Float	The number of bedrooms
Bathroom	Float	The number of bathrooms
Area	Float	The area of the listed house/apartment
House Type	Nominal	$0 \rightarrow$ Apartment, $1 \rightarrow$ Condo, $2 \rightarrow$ House
Cat	Nominal	$0 \rightarrow$ No cats, $1 \rightarrow$ Allow cats
Dog	Nominal	$0 \rightarrow$ No dogs, $1 \rightarrow$ Allow dogs
Parking	Nominal	$0 \rightarrow$ Street parking, $1 \rightarrow$ Garage
Dishwasher	Nominal	$0 \rightarrow$ No dishwasher, $1 \rightarrow$ Has dishwasher

Table 1: Attributes list

Moreover, we collect data about sublease from WildcatPad (http://www.wildcatpad.com/) and BBS of Northwestern (http://bbs.nwucssa.org/). Based on the advertisements, we can get information about attributes listed in Table 1 except the rental (most people post sublease price only), therefore we sent surveys to students who posted the sublease advertisements and asked them for the original rental. Till now more than half have responded to us. In addition to attributes mentioned above, we consider more attributes such as move-in and move-out date, utility and number of roommates.

3 PRELIMINARY RESULT

We've tried a couple of basic methods in Weka to build the model for the data. The overall results showed some promising aspects as well as issues we need to address later.

The intuitive linear regression model with 10-fold cross-validation outputs the formula with the form: $price = 296.6014 * bedroom + 694.7589 * bathroom - 0.03 * area + 384.2259 * cat + 138.4232 * dog - 212.2785 * housetype + 44.3042 * parking + 373.4674 * dishwasher + 76.0938$. The root relative squared error was 60.9023%, which was far to be used as a good model. The similar performance was showed when the multi layer perceptron is utilized, of which the root relative squared error was 67.3446%. Things worked better when we composed the model with multiple choices and rules. In M5Rules model, 9 rules were modeled according to different influence of the attributes. The root relative squared error became 55.5765%.

Still we need to find a better model to work with the data. One thing we considers now is that probably we should group the price into several groups with either $10 or other amount as the step. When dealing with the price of the housing, it is not necessary to specify the values to one dollar. And these trivils will influence the efficiency of the models. Another thing we have in mind is that in the preliminary models, we haven't consider all the import factors for the rent. We'd like to include other attributes later, such as the locating area, year of building and the furniture.

4 FUTURE WORK SCHEDULE

- Find a better estimation of rental price.

- Learn the relationship between the original rental and sublease price. To get more data about sublease, we plan to search on other websites (http://chicago.sublet.com/, http://apartments.oodle.com/)

- Connect the two estimation parts together and create an online interface to help people make estimation on sublease.