How can we predict rental price from web data?

Longtian Qiu
University of Shanghaitech
qiult@shanghaitech.edu.cn

Abstract

With the rapid development of Internet industry, the real estate agency has been offering rental service by providing the information of apartments and houses for rent on the website. In this project, we propose a pipeline from collecting data to predicting the rental price for apartments and houses in shanghai. We crawl the room information at lianjia.com by python script and generate feature vector for each room. We mainly explore two categories of algorithm to predict the rental price, neural networks and regression algorithm. Experiments on collected data set shows that given the size of the data set, the regression methods outperform the neural networks. The results are visualized as graphs and heat maps.

1. Introduction

Web crawlers are programs designed to collect the information from Internet. With the prosperity of the Internet, large amounts of data are available on various websites and social media, however, it's not feasible for researchers to collect these information by hands. The automatic scripts collecting data from internet are basic tools for data mining.

Feature generation is of primary importance in any regression or classification tasks, feature vector are used to represent the property of an object. Normally feature vector consist of numerical values which is called structured data. The main purpose of feature generation is to convert the unstructured raw data to structured feature vectors while keep as much as possible original information in the raw data.

Regression algorithms in machine learning are basically categorized into two class, linear and non-linear regression algorithm. The regression algorithms learn the relationship within the training data by finding the weight which leads to minimum loss value and predict a continuous numerical value given new data.

Neural network consist of multiple layers of nodes, where nodes are connected by edges with weights. The neural network learn the relationship within the data by optimizing the weights in the edges during the iteration of forward propagation and backward propagation. Compare with the regression algorithms, neural networks are able to learn more complex relationships within data given a sufficient size of data, however, the regression algorithms may perform a relative good result given a small data set.

2. Data Collection

2.1. Methodology

The *linajia.com* website is a relatively simple one from the perspective of web developer. There is no anti-scraping technology such as encrypted html code or authentication restriction. Most information can be extracted by parsing the page html code with the help of beautiful soup 4, a python library for parsing html code. The information of recommending house and nearby house are passed by a request *https://sh.lianjia.com/zufang/aj/house/similarRecommend* and we make an extra request to get the data. There are many crawl scripts for collecting the rental house in *linajia.com* available on github, however they are not designed to fetch complete information of house for rent, usually only the summary and price of the house are collected. We build crawler on the template of *https://github.com/jumper2014/lianjia-beike-spider* and fetch the information in the house detail page, which is not implemented by the template crawler.

The latitude and longitude information is not provided in the page content, however, we find latitude and longitude information with corresponding house id in the response of recommending house request. After the data from website is collected, we conduct a match between the latitude and longitude information and corresponding house. There are 7564 match found out of 9192 houses. To get the latitude and longitude information of the rest 1628 house, we call the *Baidu Map API: Inverse Geocoder* to get the latitude and longitude of a house by searching the community name of the house and restrict the search area in shanghai.

The pictures of the house are collected as well. The house owner provided the pictures of different rooms and

plane graph of the house. We fetch 70,084 pictures and the total size is 29.8GB.

2.2. Summary

The total number of unique house information collected is 9192. The description for data columns is listed in Table1.

2.3. Future Work

There is one idea that not been implemented, the *Baidu Map API* provided an API which return the ratings of restaurant given a coordinate. This provides the information of the quality of supporting facilities near the house, which intuitively offer more information about the rental price but in this project we didn't to validate the idea.

3. Feature Generation

3.1. Methodology

One hot encoding For straightforward columns in the collected data such as *layout, house info dict, facility info dict*, we use one hot encoding to embedding the raw data.

Images embedding To extract useful information from the pictures provided by the house owner, we use the CLIP(Contrastive Language-Image Pre-Training) model from OpenAI [1] to score the room images. The reason behind CLIP is explained below.

Natural language supervision in computer vision is the idea of using natural language as a training signal during image representation learning. Different from traditional image representation learning where each image is labeled by a specific class, the natural language supervision labels an image with a caption. The image representation may contain more information learned from the caption describing the content in the image.

The CLIP is motivated by natural language supervision and it propose a contrastive pre-training method as showed in Figure1. The images are encoded by Vit[2] or ResNet[3].The texts or captions are encoded by encoding part of Transformer[4].

Within a batch of size N, the text encoder and image encoder are training to maximize the cosine similarity of N positive pairs of image and text while minimize the cosine similarity of N*N - N negative pairs. The CLIP learns a multi-modal embedding space after pre-training which can be used to find the similarity of an image and a caption. Another contribution is the training data set of CLIP is collected from Internet which is of large scale and noisy. The openAI team publish the paper with several small-scale models, we pick model *"ViT/32"* with best performance to do a downstream task.

The CLIP model is use to score a room picture with five positive comments and five negative comments. We define each positive comment weight 1 and negative comment weight -1.As illustrated in ViLT[5], the CLIP's Modality Interaction part where the relationship of image and text is evaluated is simply so that CLIP may fail to match complex and abstract captions with corresponding images. We propose a simple and effective way of generating comments. The comments consist of two part, *"The room is"* and *a description of room*. The ten comments we chosen are *'The room is pretty big and glorious', 'The room makes people feel comfortable and relaxed', 'The room is well-lighted and drafty', 'The room is clean and organized', 'The room is modern and well-designed', 'The room is small and unexceptional', 'The room makes people feel mournful and terrible', 'The room is dark and windowless.', 'The room is disordered and dusty', 'The room is bare and lifeless'*. We visualize the result by choose room images with the highest mean score and the lowest mean score in Figure2 and Table2.

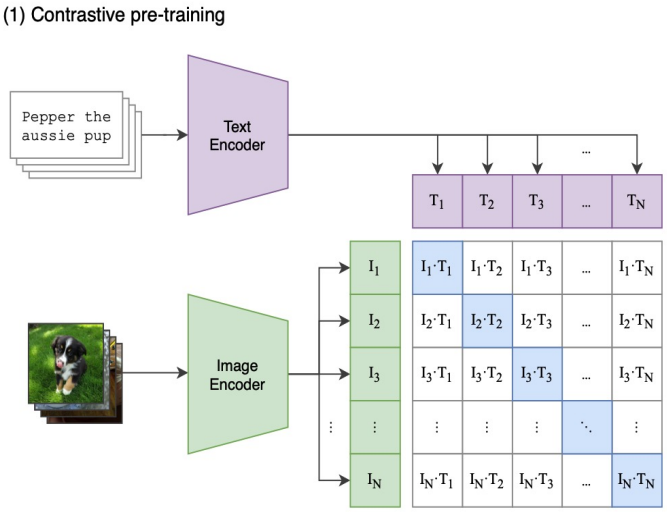

Figure 1. CLIP Pretain Pipeline

3.2. Summary

The final features contain 58 elements. The detail definition is [house layout(2), house size(1), house detail(16), house facility(10), house description(3), agent(1), house tags(10), geo(2), subway(2), image(10), image number(1)]. The agent feature is the number of houses the agent is currently acting. the subway feature contains the number of subway nearby and the minimum distance to a subway station. The house description feature is a simple measurement of how detail the description is and whether there are hospital or commercial center nearby.

3.3. Future Work

There is one idea that have not been implemented. The comments in image embedding may be collect from wiki

	description
fetch_date	The date of fetching, in format "YYYYMMDD"
district	The district of the house
area	A more detail location than district
xiaoqu	Name of the housing estate, in format "{rental type}·{housing estate}"
layout	Layout of the house
size	Floor space
price	Rental price
url	The url of the house in website
house_id	A unique id for one house
relative_image_path	Displayed images in website, in format "[({image url, room type})]"
house_info_dict	Information about "size,facing,maintain,living,floor,elevator,parking,water,electricity,gas,warming"
facility_info_dict	The furniture and household appliances information
house_description	The description of the house provided by the house agent
agent_name	The name of the agent
recommend_house_id_geo	The house id and geographic information of recommended houses.
nearby_house_id	The house id of nearby houses
tag_list	The tags of the house
subway_info	A list of nearby subway station and distance.
geo_lat	The latitude of the house
geo_lng	The longitude of the house

Table 1. Description of fetched data

Captions	Best room	Worst room
The room is pretty big and glorious	1.4899	0.2871
The room makes people feel comfortable and relaxed	25.326	0.5549
The room is well-lighted and drafty	9.4641	1.3511
The room is clean and organized	0.3662	0.0304
The room is modern and well-designed	31.511	1.2198
The room is small and unexceptional	-10.734	-1.994
The room makes people feel mournful and terrible	-1.7917	-1.413
The room is dark and windowless	-0.5192	-1.694
The room is disordered and dusty	-1.0322	-1.356
The room is bare and lifeless	-0.7646	-3.096
Final score	53.3159	-6.113

Table 2. Score of best room picture and worst room picture. The score of a caption shows how related is the picture and the caption. The pictures of rooms are in Figure2.

English corpus by searching key words *"room is"* or other short words describing a room. We may collect large amounts of comments and enlarge the feature vector size of a house. This may leads to better results.

4. Prediction

4.1. Methodology

Regression algorithm There are basically two category of regression algorithms, linear regressor and non-linear regressor. Given the feature size is 58, we choose three non-linear regression algorithms, *kernel rigid regression, SVR, decision tree regression*. The kernel rigid regression and SVR need a kernel function to project the input data points into a hyperspace where the data points are more separable than original space.

Neural Network We propose two architecture of neural network, multiple layer perception and multi-head self attention as Figure3. The MLP architecture consist of one hidden layer, one input layer and one output layer. The multi-head self attention architecture consist of four self attention blocks, where each self attention consist of 10 heads. The self attention architecture was first introduced in Transformer[4] and the reason we used it is that self attention layer connects all position with a sequentially executed operations which make it possible for network to capture the relationship between different input features. The multi-head mechanism of self attention layer enables attention layer to capture multiple input features relationship.

Optimizer The optimizer we chosen are stochastic gradient descent with weight decay and Adam with weight

Figure 2. The upper four pictures are the room with lowest score.The lower four pictures are the room with highest score.

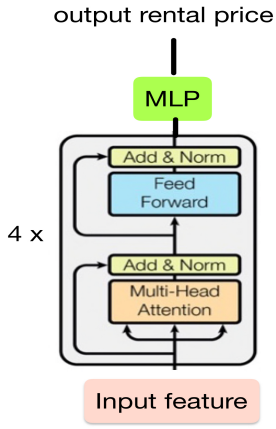

Figure 3. Attention Model Architecture

decay[6]. The SGD optimizer is simple but may be problematic when the learning rate or gradients are too large near a local optimum. The Adam alleviate this condition by introducing the first moment estimate and second raw estimate. However, under large scale data set, SGD optimizer may still be a better option for cheaper computational cost than Adam. The optimizer scheduler we choose is cosine annealing and warm restart[7].

5. Experiments

5.1. Experiment Settings

We spit collected data into 90% training set and 10% testing set. Given that the public house rental data won't be noise, the loss function is a modified mean square loss named MMSE for short where the ground truth price is divided by 10^{-4}. We consider this modification will help us observe the loss value more directly and accelerate convergence speed. Here is the MMSE formula:

$$MMSE = \sum_i (pred_i - label_i * 10^{-4})^2$$

5.2. Regression Algorithm

We test kernel rigid regression, SVR and decision tree regression, for SVR and kernel rigid regression, we test kernel functions **[linear, polynomial, rbf, laplacian, sigmoid, cosine]**. The result of MMSE loss is in Table5.

$$linear : k(x, y) = x^T y + c$$

$$polynomial : k(x, y) = (\alpha x^T y + c)^d$$

$$rbf : k(x_i, x_j) = exp(-\gamma ||x_i - x_j||^2)$$

$$laplacian : k(x, y) = exp(-\frac{||x_i - x_j||}{\sigma})$$

$$sigmoid : k(x, y) = tanh(\alpha x^T y + c)$$

$$cosine : k(x, y) = \frac{x * y}{||x|| * ||y||}$$

5.3. Neural Network

In all experiment with neural network, we use SGD and Adam optimizer with 0.0001 initial learning rate, 0.00001 weight decay and 0.00001 momentum for SGD. The optimizer scheduler is cosine annealing with warm start where restart begins at 1000iteration. The batch size is 200 and for MLP and Attention architecture, the hidden layer size is 256. We train the 4 combination of model and optimizer for 300 epoch each. The result for MLP is in Figure6,7 and result for Attention is in Figure4,5. The lowest evalution MMSE for Attention network is 73.85 and 42.27 for MLP network. The reason for poor performance is the lack of training data set. We consider data augmentation techniques may alleviate this condition to some extend.

	linear	poly	rbf	laplacian	sigmoid	cosine
KR with Regularization strength 0.5	0.553718	2.21813	1.275188	0.3284159	1.440662	0.729343
KR with Regularization strength 1.0	0.555052	2.21256	1.371862	0.340766	1.440655	0.7339337
KR with Regularization strength 1.5	0.555854	2.210651	1.451493	0.3524404	1.440649	0.731567
SVR with Regularization 0.5	1.177661		0.7433953		0.998917	
SVR with Regularization 1.0	3.206519		0.69805		0.877105	
SVR with Regularization 1.5	9.913045		0.679994		0.828925	
decision_tree	1.325817					

Table 3. The MMSE of different regression algorithm and different combination of parameters.

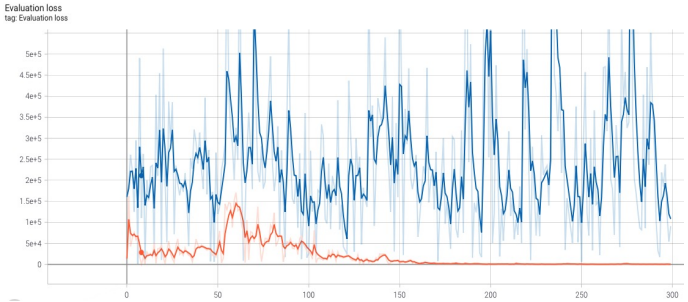

Figure 4. Attention Model Evaluation MMSE Loss. The red line represents the model trained by AdamW optimizer. The blue line represents the model trained by SGDW optimizer

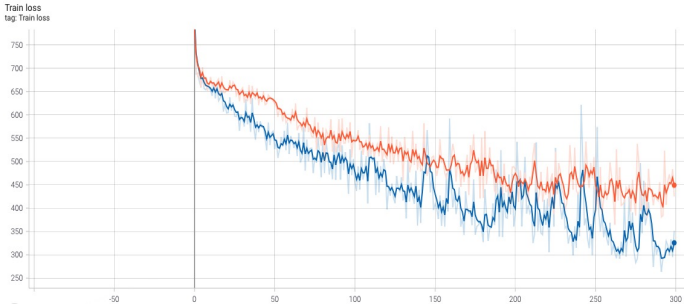

Figure 5. Attention Model Training MMSE Loss. The red line represents the model trained by AdamW optimizer. The blue line represents the model trained by SGDW optimizer

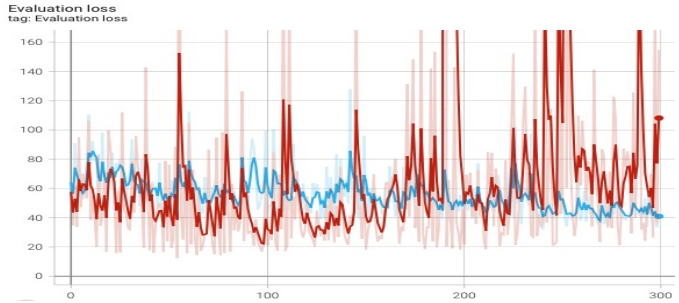

Figure 6. MLP Model Evaluation MMSE Loss. The red line represents the model trained by AdamW optimizer. The blue line represents the model trained by SGDW optimizer

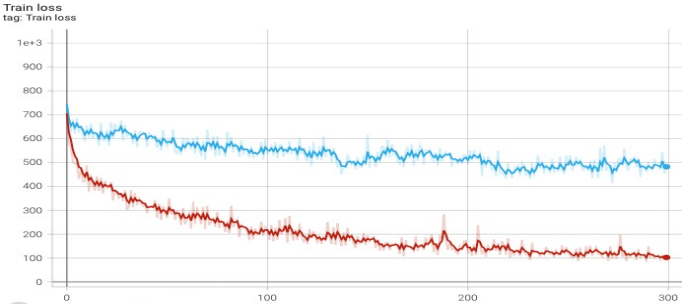

Figure 7. MLP Model Training MMSE Loss. The red line represents the model trained by AdamW optimizer. The blue line represents the model trained by SGDW optimizer

5.4. Feature Analysis

In this section, we analyze the importance of each feature elements. Here we use the sequential forward selection algorithm. The SFS is a greedy search algorithm which takes the whole feature set as input. The SFS maintains a chosen feature list and will add the element which output highest score when tested with chosen elements. We choose Decision Tree Regressor algorithm as the base testing algorithm. The result of feature importance is in Table6,7. The result show that the most importance feature is the room size. The coordinate of the room is of secondary importance. Intuitively, the room size and the location are the most important factor of rental price and our result demonstrate that. We consider that the latitude and longitude will split the room into different areas, which greatly influence the predicted rental price. Since the coordinate of a room includes the information of nearby subway station, the fact that extracted subway features are of low ranking is reasonable. Surprisingly, we find that the image feature 6, "The room is small and unexceptional" is of third importance. We come to a conclusion that the general negative description of a room will split the rental price in decision tree regressor.

We also run the experiments with top k features in the rank to see how much features gives best performance. The result is in Table4. The result shows that when we choose top 50 features, the algorithm gives the best performance. The conclusion is that top 50 features are meaningful and will give algorithm useful information while the rest 8 features will confuse the algorithm. However, top 15 and top 35 also give relatively good performance. These two combinations are available options when we are need to find a

Top k	Number of features
5	1.7921922284033818
10	1.8748175900326087
15	0.9727805556630434
20	1.7625101513152177
25	1.1438755001195655
30	1.203894725021739
35	0.8164614535326087
40	1.1416450491413044
45	0.743628090673913
50	0.647765078978261
55	0.910040187554348
58	2.0233932191847828

Table 4. The MMSE of top k features in decision tree regressor

trade off between model accuracy and computational cost.

5.5. Ablation Study of Visual Feature

We investigate how image feature affect our prediction result. We test feature vector with and without image feature by regression algorithms. The result of performance difference is showed in Table3. The positive values in table mean the MMSE loss without image feature is higher than the MMSE loss with image feature. The result shows that the image features plays a positive effect in simple algorithms such as decision tree regressor and kernel rigid regressor with linear, cosine and sigmoid kernel function while image features plays a negative effect in kernel rigid regressor with laplacian, rbf and polynomial kernen function.We think the reason is that in the embedding space of laplacian or polynomial kernel function, the positive and negative image features are not separated, which leads to a drop in performance. From the perspective of feature analysis, we can tell that except for the image feature 6 rank 3^{th} in the feature importance, other image features rank behind 30^{th}. In conclusion, we think the image features do have some positive effects on predicting the rental price, however, the ten descriptions of room pictures should be improved given that some image description features are of low ranking.

5.6. Result Analysis

The result show that given data set size of 9192 and feature vector size of 58. Regression algorithm outperforms neural networks. For regression algorithm, the best model is kernel rigid regression with laplacian kernel function. There are more interesting detail to explore considering different kernel functions have different loss. For neural network, Adam optimizer is more efficient than SGD in training data but leads to a worse evaluation loss in MLP model. However, the Adam optimizer perform better in evaluation loss while the train loss decrease more slowly but more steady than SGD optimizer. The reason may be the convergence speed of Attention model is much slower than MLP model while it can learn more information. However, current data is too little for training both models. Inspired by the success of laplacian kernel function, we believe Graph Neural Network may perform better than MLP and Attention models.

6. Visualize

We generate two heat map by *pyechart library* and *baidu map api*. The first heat map is the rental price heat map in shanghai as Figure6. The second heat map is the difference between predicted price and ground truth price as Figure7. For better visualization result, we divided the delta value into several stages.The result of second heat map show that predictions of downtown area are less accurate than other area. The html file of the heat maps are interactive.

7. Conclusion

In this project, we investigate the probability of predicting the house rental price from web data. We propose some interesting way of generating feature from raw data. The prediction algorithm are not fine tuned and only used to demonstrate the feasibility of the algorithm. Visualization are made to show directly the trends of data.

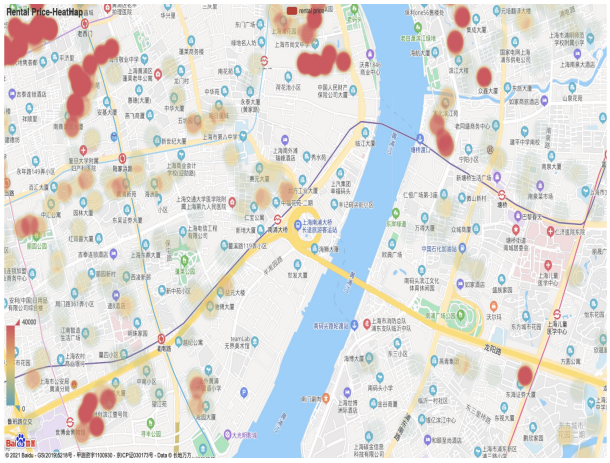

Figure 8. Rental Price Heat Map.The darker the red is, the more expensive the rental price is.

References

[1] Radford, A. , et al. "Learning Transferable Visual Models From Natural Language Supervision." (2021).

[2] Dosovitskiy, A. , et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." (2020).

[3] He, K., Zhang, X., Ren, S., and Sun, J., "Deep Residual Learning for Image Recognition"(2015).

	linear	poly	rbf	laplacian	sigmoid	cosine
KR with Regularization strength 0.5	0.0754	-0.9941	-0.06648	-0.0004279	4.8E-08	0.0004767
KR with Regularization strength 1.0	0.075248	-0.94053	-0.07086	-0.00209	5.6E-07	0.00030531
KR with Regularization strength 1.5	0.075146	-0.931651	-0.073293	-0.00316424	0	0.000242
SVR with Regularization 0.5	0.085339	0.0002763	0.0004		-0.000117	
SVR with Regularization 1.0	3.20651	0.000868	0.00058491		0.000250538	
SVR with Regularization 1.5	-1.2403	0.0009	0.0007		9.78E-05	
decision_tree	0.009283					

Table 5. The result of ablation experiment. The values are the MMSE of a algorithm without image features minus the same algorithm with image features. If the value is positive, the MMSE of algorithm without image features is higher than the one with image features and images features play a positive effect.

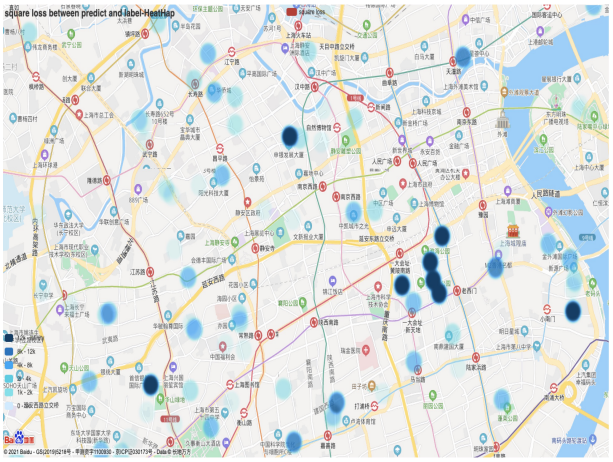

Figure 9. Prediction Loss Heat Map.The darker the blue is, the greater the prediction loss is.

[4] Vaswani, Ashish, et al. ”Attention is all you need.” Advances in neural information processing systems. 2017.

[5] Kim, Wonjae, Bokyung Son, and Ildoo Kim. ”Vilt: Vision-and-language transformer without convolution or region supervision.” arXiv preprint arXiv:2102.03334 (2021).

[6] Loshchilov, Ilya, and Frank Hutter. ”Decoupled weight decay regularization.” arXiv preprint arXiv:1711.05101 (2017). 2

[7] Loshchilov, Ilya, and Frank Hutter. ”Sgdr: Stochastic gradient descent with warm restarts.” arXiv preprint arXiv:1608.03983 (2016). 2

2

Rank	Feature name	Description
1	size_feature	The size of room
2	geo_feature_1	Latitude
3	image_feature_6	”The room is small and unexceptional”
4	geo_feature_2	Longitude
5	tag_list_4	Well remodelling
6	facility_info_feature_8	Central heat
7	house_info_dict_park_no	Whether the parking place is available
8	house_info_dict_west	The room face west
9	facility_info_feature_4	Whether there is TV in the room
10	house_description_2	Whether a commercial center is mentioned in the description
11	tag_list_5	Newly updated
12	facility_info_feature_1	Whether there is washing machine
13	facility_info_feature_5	Whether there is refrigerator
14	house_info_dict_east	The room face east
15	tag_list_2	Check at anytime
16	house_info_dict_park_rent	Whether there is parking place for rent
17	house_info_dict_south	the room face south
18	facility_info_feature_10	Whether there is natural gas
19	tag_list_1	Strongly recommended house
20	house_info_dict_heat	Whether there is central heat
21	house_info_dict_elevator	Where there is elevator
22	tag_list_6	Double washrooms
23	house_info_dict_elect	Whether there is commercial electricity
24	house_info_dict_medium	Medium floor
25	house_info_dict_water	Commercial water
26	house_info_dict_high	High floor
27	house_info_dict_floor	Exact floor number

Table 6. The rank of feature importance and feature description part1.

Rank	Feature name	Description
28	house_info_dict_checkin	Checkin at any time
29	layout_feature_hall	Number of halls in room
30	facility_info_feature_3	Wardrobe
31	agent_feature	Number of house a agent is representing
32	layout_feature_room	Number of rooms
33	image_feature_9	”The room is disordered and dusty”
34	tag_list_9	Verified by authority
35	image_feature_2	”The room makes people feel comfortable and relaxed”
36	image_num_feature	Number of pictures provided by house owner
37	house_description_1	Number of section in description
38	tag_list_8	Pay one deposit one
39	facility_info_feature_6	hot water heater
40	house_info_dict_park_free	Whether the parking place is free
41	facility_info_feature_9	Network
42	house_description_3	Whether hospital is mentioned in description
43	image_feature_8	”The room is dark and windowless”
44	house_info_low	Low floor
45	tag_list_10	Monthly rental
46	image_feature_5	”The room is modern and well-designed”
47	facility_info_feature_7	Whether there is bed
48	facility_info_feature_2	Whether there is air conditioner
49	tag_list_7	near subway station
50	image_feature_1	”The room is pretty big and glorious”
51	tag_list_3	Recommended by house owner
52	image_feature_3	”The room is well-lighted and drafty”
53	subway_feature_num	Number of nearby subway lines
54	subway_feature_dis	Minimum distance to subway station
55	image_feature_10	”The room is bare and lifeless”
56	house_info_dict_north	The room face north
57	image_feature_7	”The room makes people feel mournful and terrible”
58	image_feature_4	”The room is clean and organized”

Table 7. The rank of feature importance and feature description part2.