DOCUMENT OF CURRENT WORKS

2016-9-6

CONTENTS

LIST OF TABLES

1 INTRODUCTION

Our work is estimating Amazon's EC2 spot instance price. We analyzed five types instances which is as follow, d2.8xlarge, g2.xlarge, m3.medium, m4.xlarge, r3.large. The data begins at June 29th and ends at August 29th in useast. And the operating system is Linux/Unix. We try to use two methods to estimate the price. First we use mean values. Second we use Linear regression to estimate price.

2 METHODS

2.1 Data preprocessing

The data get from Amazon is formed as the chart. The next two methods only use the column price to estimate the 'Price'.

Table 1: Data from Amazon

Type	Price	Timestamp	InstanceType	ProductionDescription	AvailabilityZone
SPOTINSTANCEPRICE	0.010800	2016-08-29T07:59:36+0800	m3.medium	Linux/UNIX	us-east-1a
SPOTINSTANCEPRICE	0.010700	2016-08-29T07:43:09+0800	m3.medium	Linux/UNIX	us-east-1a
SPOTINSTANCEPRICE	0.010600	2016-08-29T07:25:29+0800	m3.medium	Linux/UNIX	us-east-1a
SPOTINSTANCEPRICE	0.010700	2016-08-29T07:10:22+0800	m3.medium	Linux/UNIX	us-east-1a
SPOTINSTANCEPRICE	0.010800	2016-08-29T07:09:54+0800	m3.medium	Linux/UNIX	us-east-1a
SPOTINSTANCEPRICE	0.010700	2016-08-29T07:07:11+0800	m3.medium	Linux/UNIX	us-east-1a
SPOTINSTANCEPRICE	0.010800	2016-08-29T06:57:13+0800	m3.medium	Linux/UNIX	us-east-1a
SPOTINSTANCEPRICE	0.010700	2016-08-29T06:56:44+0800	m3.medium	Linux/UNIX	us-east-1a

The set of explanatory variables for our predict methods is summarized in Table II

Table 2: Variables in our methods

Variables	Description
P	The price vector for all regions
P_i	The price vector for part of region a,b,c,d,e
Num_i	
$\rho_{benchmark}$	The benchmark of high corrlative of two variables
X	The matrix of input data
V	The vector of each day's price
m	The number of the data
m_train	The number of training data
m_test	The number of test data
$price_{estimate}$	The price that we estimate
$price_{real}$	The real price at that time

First we pick up the column 'Price' into the vector P, and then we find the Num_i for each region. To complete this work, we use the funtion 'autocorr' in MATLAB. The function returns the correlation coefficients of the current price and previous prices. In math, if the correlation coefficient is greater than 0.5, we think these two variables is strong correlative. So

$$Num_i = sum(autocorr(P_i) > \rho_{benchmark})$$

For example the result of $autocorr(P_a) = [1.0000, 0.8038, 0.8296, ..., 0.5749]$. The num of element which is greater than $\rho_{benchmark}$ is 19. So Num_a is equals 19. And then according to Num_i we construct a matrix X_{m*num_i}. Each row in this matrix is an input data. colunm i is the price at i times

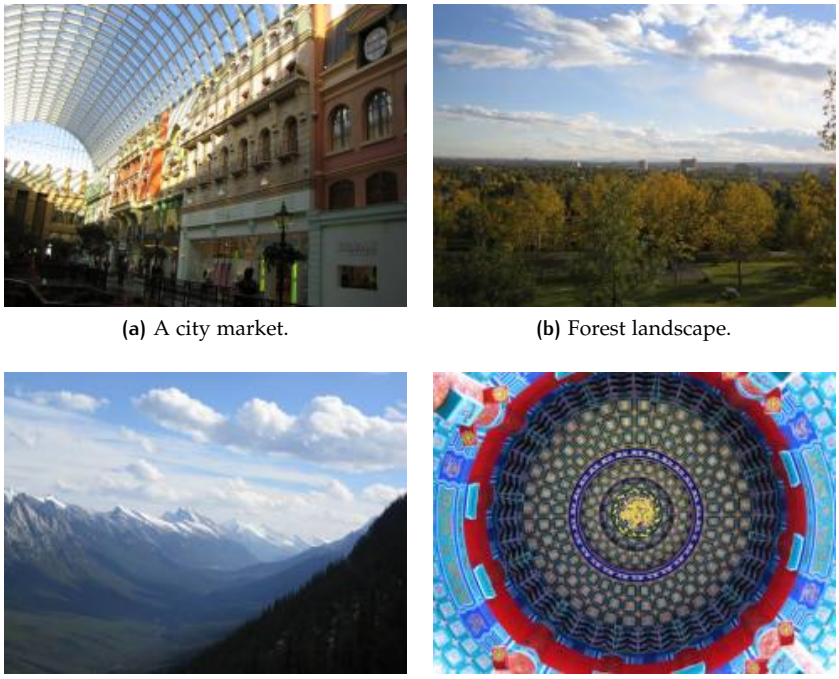
before current. We use current price to construct a vector V_{m*1}. And V is a label of X.

2.2 Use mean values to estimate the price

For each row of X, we get it's mean values. And then we use $e = \frac{abs(price_{estimate} - price_{real})}{price_{real}}$ to evaluate our predict. The predict results show the figure 1. And I list the numbers of data which $e \geqslant 0.1$ and the total number of this type instance.

2.3 Use Linear regression to estimate the price

First we divide the data X into training set and testing set. Then we use the function 'regress' in MATLAB to fit the training set and got coefficient vector β. At last compare βX_{train} with $price_{real}$.



(a) A city market.

(b) Forest landscape.

(c) Mountain landscape.

(d) A tile decoration.

Figure 1: A number of pictures with no common theme.