AI1103 Reasarch Paper

Akyam L Dhatri Nanda - AI20BTECH11002

Probability Density For Amazon Spot Instance Price

Amazon EC2 (Elastic Compute Cloud)

It is a computing environment provided by AWS. It supports highly scalable compute capacity. Amazon EC2 can be used for moving application.We can configure security and networking in amazon EC2 much easily than our own custom data center.

Instances

1. **On-demand Instances** are virtual servers that run in AWS EC2 or AWS RDS. These are purchased at a fixed rate per hour and are charged only for the seconds for which the instance is in running state. We have full control over it's life cycle.
2. **Spot Instances** are the unused compute capacity AWS has. It is a way to reduce your EC2 On-demand instance cost by about 90 percent.

Abstract

1. Amazon EC2 is a computing environment provided by AWS. Bidders can bid for the spare compute capacity called Amazon Spot Instance (SI).
2. Traditional point prediction algorithms provide an optimised value through error approximation. The resultant point prediction value is close to mean or median and therefore there would only be 50 percent probability of winning the bid.
3. In this paper, we develop a technique to calculate the probability density of the SI price considering both the curve fitting and historical similarities.
4. This probability density helps the bidders in setting a price considering both urgency of the task and condition of the market.

Introduction

Popularly applied error values

1. RMSE - Root Mean Square Error
2. MAPE - Mean Absolute Percentage Error
3. SSE - Sum Squared Error

Gaussian Probability distribution:

1. Heteroscedastic error values with point prediction
2. NN based prediction interval
3. PDF contains all possible values

Current Rules In Amazon EC2 Spot Market

- A bidder gets access to SI when the bid is higher than the price of SI.(equal in some cases)
- The price of SI varies over time.
- When the price of SI becomes higher than the bid, the user is notified with a two-minute warning.
  - Increase the price
  - Save the progress and stay idle and the instance terminates.
  - Stay idle without saving the progress and the instance terminates.
- Partial hour is not charged if the user loses the instance due to price increment.
- Partial hour is charged with the price of closing time as full hour if user releases the instance.
- The user is charged with the price of the hour-end time for running instances

Hibernation

- During price increase SI goes to hibernation.
- User is not charged for partial hour.
- The user needs to pay for the backup storage at standard Amazon Elastic Block Store (EBS) storage rates.
- The user may terminate or cancel bid during hibernation.

Limitations on bidding

- Some SI types are not available in all regions.
- Each user account can bid for roughly 20 Spot instances per region.
- The highest limit of the bid price is ten times the on-demand price.

Spot Fleet

It is a collection of Instances. Through the SF system, a user can bid for thousands of servers with a range of bids. That fleet also has several restrictions.

Key features of SFs

- The number of active SF in a region $\leq$ 1,000
- The target capacity of an SF $\geq$ 3,000
- The target capacity of all SF in a region $\leq$ 5,000
- An SF request cannot span more than one region.
- Users can not bid into any pools with more than the on-demand price.
- When the user can divide instances and bid at different prices, the price change varies the execution speed but the execution continues

https://aws.amazon.com/blogs/aws/ec2-fleet-manage-thousands-of-on-demand-and-spot-instances-with-one-request/

Probability Density Computation

A. Probability density through the curve fitting/correlation

Similar occurrences are searched through the direct correlation between recent samples and the training string.

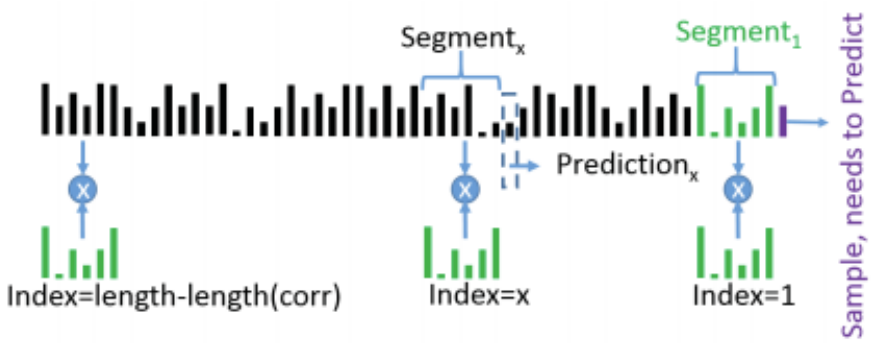

Figure: Correlation of the string with recent samples for finding similar occurrences

It represents search process through indexing.

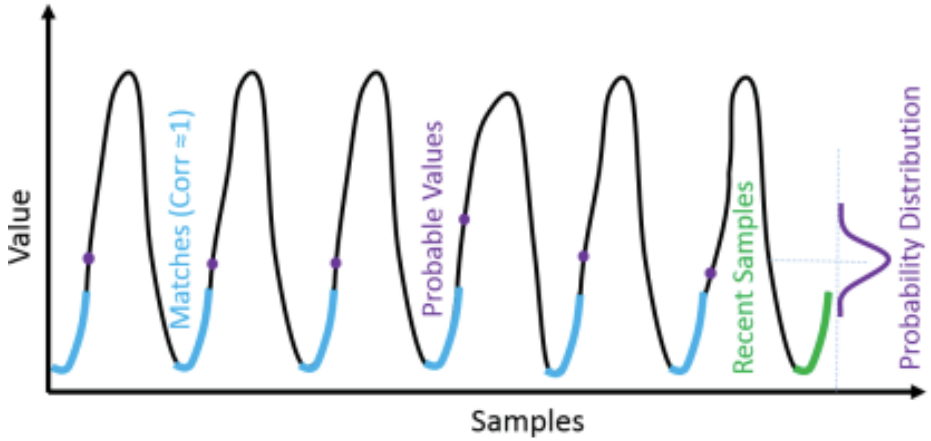

Figure: A graph illustrating locations of good matches, values of the next samples after each match.

The values of next samples after each match are represented by purple dots and the these create a probability distribution which is shown to be Gaussian with no skewness.

$$Corr_{index} = \frac{\sum_{i=1}^{m} segment_1(i) \times segment_{index}(i)}{rms(Segment_1) \times rms(Segment_{index})} \tag{1}$$

Eq (1) represents the process of calculating the correlation

In (1), m is the length of each segment. $m = 10$ is chosen for the balance between accuracy and execution speed.
The value of normalized correlation stays between -1 to +1 inclusive

1. +1 means the exact match
2. 0 means no match
3. -1 means the exact inverse match

After the search of similarities, the indexes which correspond to the best matches are selected for the formation of the cumulative probability distribution

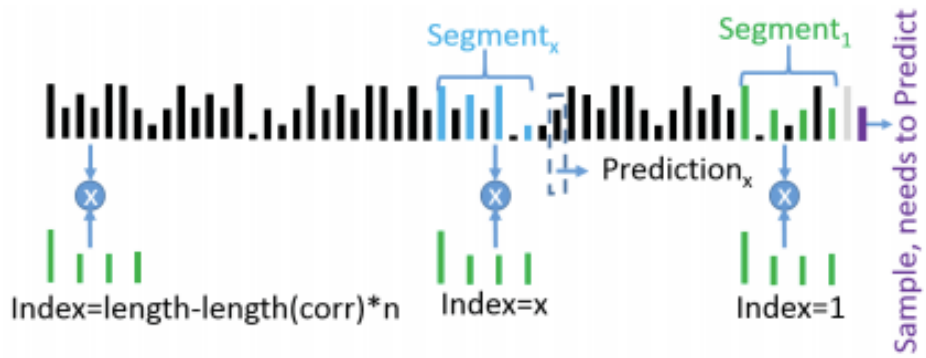

Figure: Correlation based long term prediction distribution calculation

This system is designed in such a way that the prediction time needs to be equal or the integer multiple of the sampling period

- Short-term prediction = Value after 5 mins
- Long-term prediction = Value after 1 hr
- A down sampling of factor twelve is applied for the long-term (hourly) prediction.

However, the downsampling decreases the string size and results in fewer matches.Thus, the segments are downsampled by a factor (n)

The normalising ratio :

$$R_n = \frac{rms(Segment_1)}{rms(Segment_{index})} \tag{2}$$

$$Prediction_{index} = Prediction'_{index} \times R_n \tag{3}$$

Here, $Prediction'_{index}$ is the value of the next sample of the correlated segment ($Segment_{index}$).

And $Prediction_{index}$ is the normalized version of that value.

$$Relevance_{index} = Corr^{\eta}_{index} \times \frac{2}{R_n + 1/R_n} \tag{4}$$

Here, η is the weight parameter. The correlation parameter is given a higher weight with $\eta = 5$.

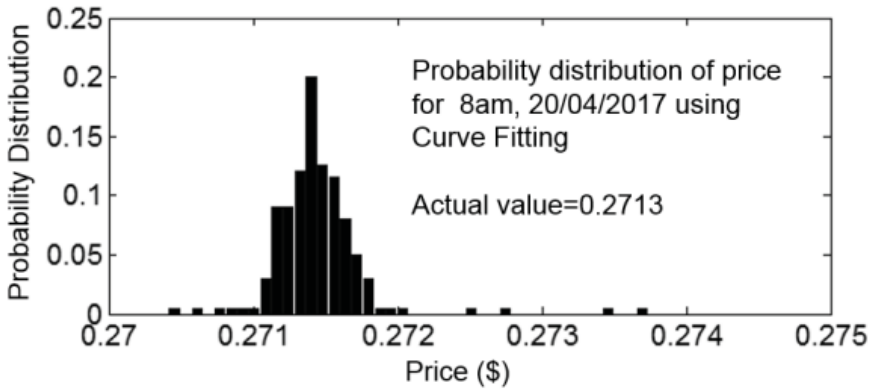

Figure: Weighted distribution of the Amazon EC2 c4.4xlarge SI price prediction using correlation based curve fitting.

B. Limitations of Curve Fitting based Prediction

- The price of SI usually increases at the start of the peak hour and correlation-based prediction density function usually has a denser part at lower values due to the limitations of the search length.
- These upward trends can mislead them and potentially increase the market price.
- Similar for the downward trend.
- Therefore, a daily and weekly pattern based prediction with holiday consideration is required to know the exact value of the resource in order to solve these issues.

C. Probability Density Through the Daily and Weekly Patterns and Holiday Considerations

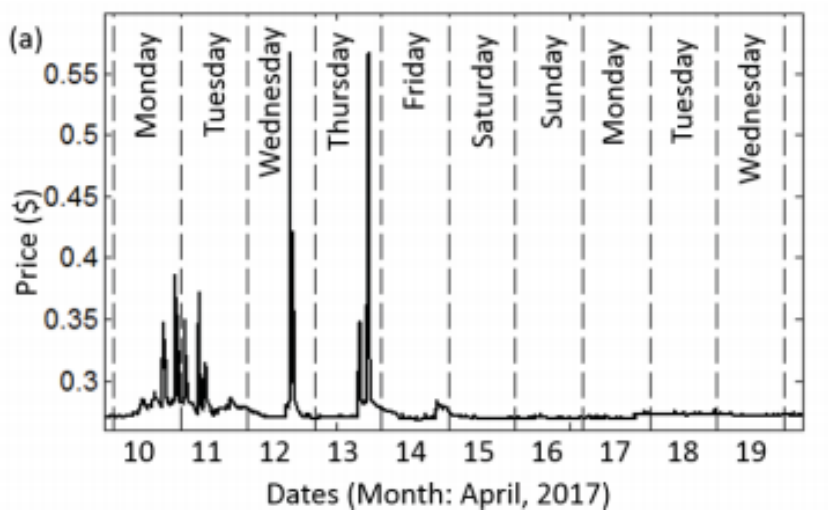

Figure: Daily and weekly patterns of the Amazon EC2 c4.4xlarge SI price

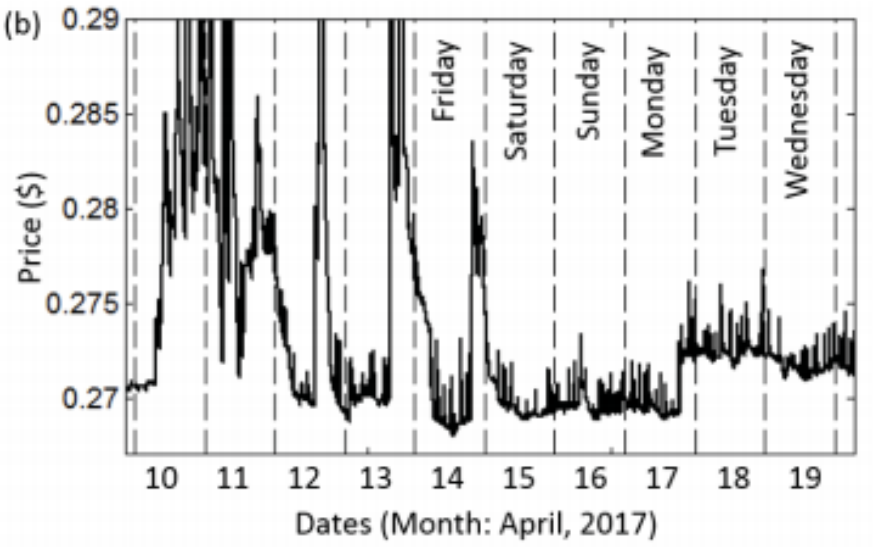

Figure: vertically magnified version of subplot (a)

D. Limitations of the Daily and Weekly Patterns and Holiday Considerations

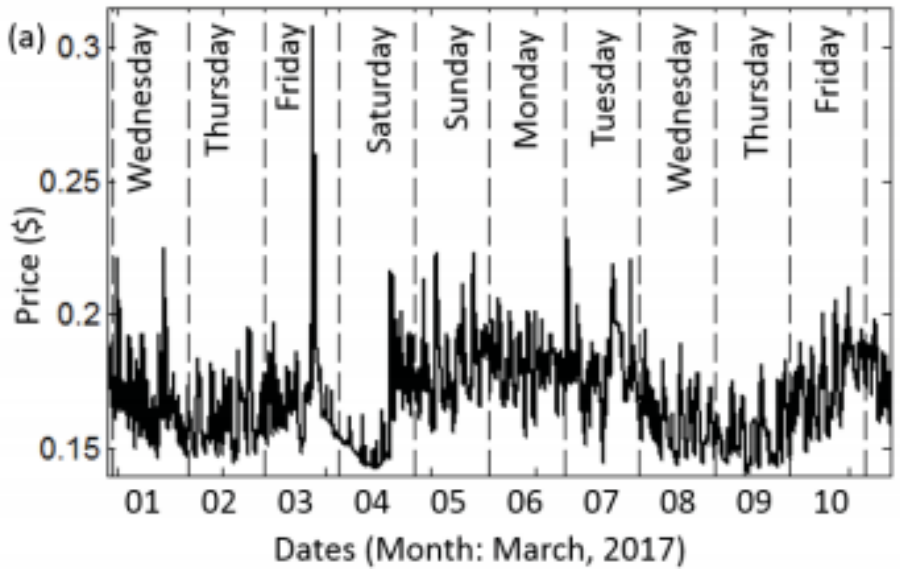

Figure: Change in daily and weekly patterns

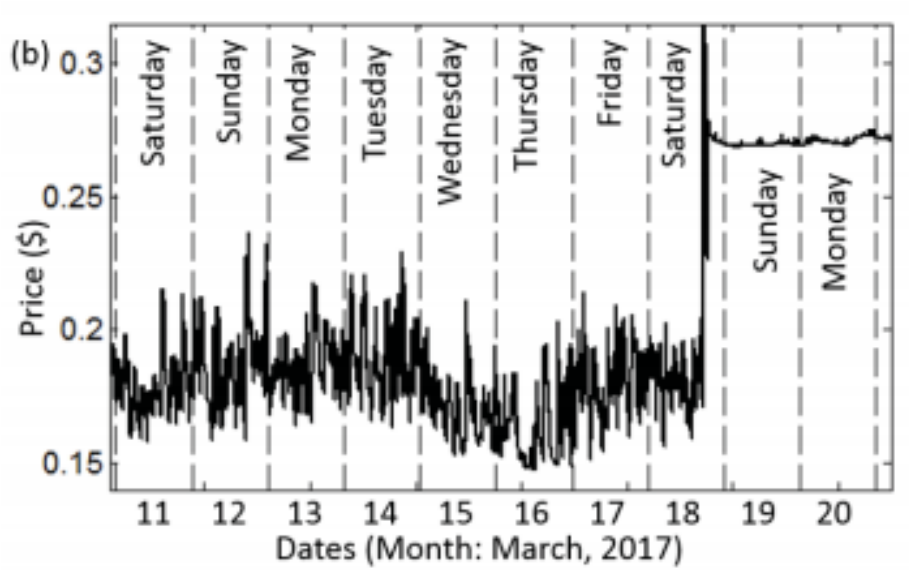

Figure: unexpected price change for the Amazon EC2 c4.4xlarge SI

$$Relevance_{sample} = \frac{1 + same_{day}}{2} \times (1 - \frac{|\Delta T|}{12}) \quad (5)$$

- $Relevance_{Sample}$ is the relevance of the sample.
- $Same_{Day}$ is a function which returns 1 when the sample is taken on the same day of the week.
- T is the time difference in hours with a range of -12 to 12.

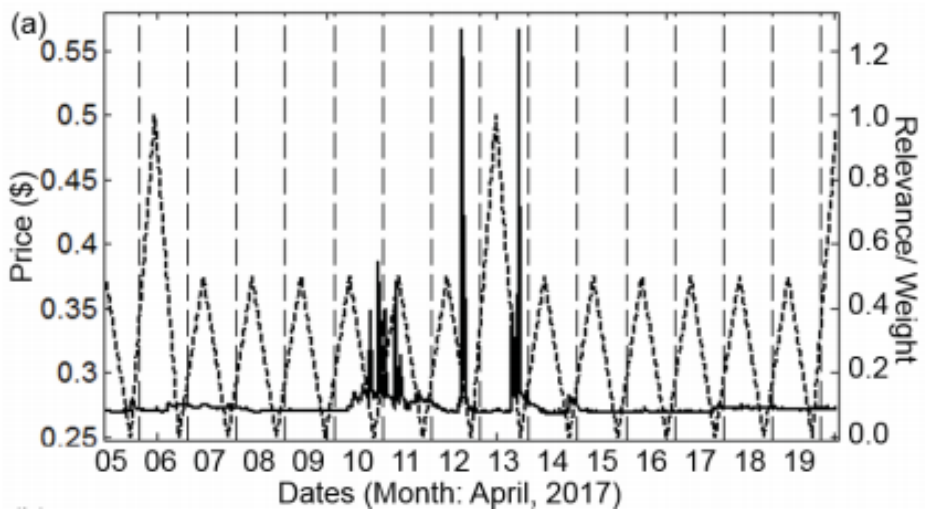

Figure: The price of last 15 days presented by the solid line, thinner dotted lines are presenting day transition and thicker dotted lines are presenting corresponding weight calculated.

E. Combined Probability Density for Bidding

Both of the curve fitting and the daily and weekly based predictions have advantages and limitations. Two distributions are added in order to calculate the overall probability distribution.

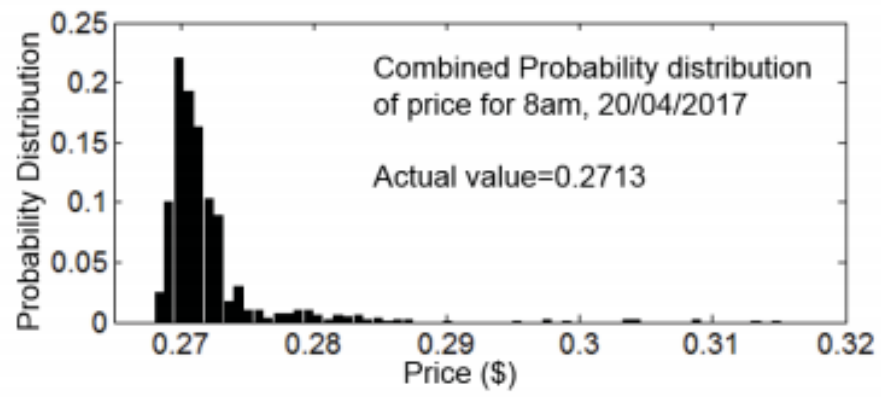

Figure: Bar chart of the combined probability density.

Conclusion

The probability density of price can potentially help bidders to bargain with the spot price. The user can easily understand uncertainty and take risks by picking one uncertainty bound based on the urgency of his task.