International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 16 (2018) pp. 12747-12752

Dynamic Method to Predict Features for Amazon Spot Instances

Richa Dubey[1], Mrs. Pragya Pandey[2]

Computer Science Department, Medicaps University, Indore, India.

Abstract

Cloud computing is accepted in recent years for a number of applications because it is reliable, efficient and scalable according to the need of applications. In order to execute the applications on server computational resources are required and the needs of resources are varying with the time and load on the resources. Therefore the prices are also time varying in cloud environment. In order to understand the price variation for spot instance prices the prediction algorithms are required. In this context the proposed work introduces a data mining technique based price prediction model. The data mining models provides us the ability to learn with the historical data trends and identify or recognize the similar trends of the data. That ability makes it essential for various new generation applications. The proposed model is a supervised learning model and implemented using the SVR (support vector regression) model. The proposed system works in three phases in first phase the data collection and preprocessing is adopted. In next phase the sliding window protocol is executed for grouping of similar data patterns. Finally the training and testing of SVR model is performed for predicting the price of spot instances. The implementation of the proposed technique is performed using LibSVM library and the experiments are performed using Amazon spot instance dataset. The proposed work provides accurate and efficient results therefore proposed model is acceptable for real world usage.

Keywords: AWS, Spot Instance, Prediction, Cloud Computing, Regression Analysis, Dataset.

INTRODUCTION

Cloud become highly accepted technology in recent years. A number of companies and organizations are now in these days hosting their data and applications over cloud. Basically in order to run or execute applications over cloud expensive resources are required. Additionally the demand of resources is varying continuously, due to variations in load time or demand of computing resources. Therefore the prices of the resources are also varying according to the time, demand and load on resources. Therefore the resource cost cannot be determined directly and bidding process is used to buy the resources according to the needs.

In this context for making effective bids for purchasing the resources in best prices the predictive methodologies can help. Thus in this presented work the approach is proposed using the data mining technique that is used for predicting the price of computational resources. The data mining techniques are applying the computational algorithms on the historical price data. Basically the data mining is a tool or technique by which the prediction, classification, clustering and association rule mining. All these techniques are applied for finding the trends on data or relationship in data attributes. Using these algorithms trends of price increase and decrease is modeled using the data models. And these data models are used to predict the similar trends of the data. In this presented work the SVM (support vector machine) is used for analysis and prediction of data values for the resource prices.

PROPOSED WORK

In order to predict the spot instances price of the cloud computing resources a data mining technique based model is proposed. The details of the proposed approach are reported in this section.

A. System Overview

Resource prices in a particular time domain is termed as spot instance price. Where the spot denoted for time, instance is used for resource object and their price. This price value is varying according to the time, demand and work load appeared on cloud infrastructure. Therefore the bidding process is applied for resource purchasing. In order to known about the variations in price and possible upcoming value of spot instance price the machine learning based technique can be used for predicting the instance price values. The machine learning approaches are basically computational algorithms that analyze data and recover relationships and trends on data. Using the discovered trends of data similar patterns and data trends are identified or recognized.

Therefore in this presented work the machine learning approaches are employed for analyzing the historical spot instance price data which is taken from the Amazon instance dataset. The trends hidden in the training data is analyzed here using SVM (support vector machine). The support vector machine is a supervised learning model that first trained on the examples of similar data and then after training able to predict the similar trends or class labels for new patterns. During training individual instance of data pattern is considered as the point in n-dimensional space. Additionally the attributes or features of data are considered as the coordinates for the samples. For classifying the data into their classes the data model plot a hyper plan between the samples coordinates. That hyper plan makes classifiable the two data instances in n-dimensional space. That model sometime also used with the regression analysis for more accurate learning and increase the accuracy. This kind of SVM based on regression process is termed as SVR (support vector regression). In this presented work the SVR model is implemented using libSVM library. This section provides initial description of the proposed instance price prediction model. In next the detailed model is explained.

B. Proposed Methodology

The proposed data model for spot instance price prediction is demonstrated in figure 2.1. The diagram is grouped with the different components that process information intermediately for completing the required task.

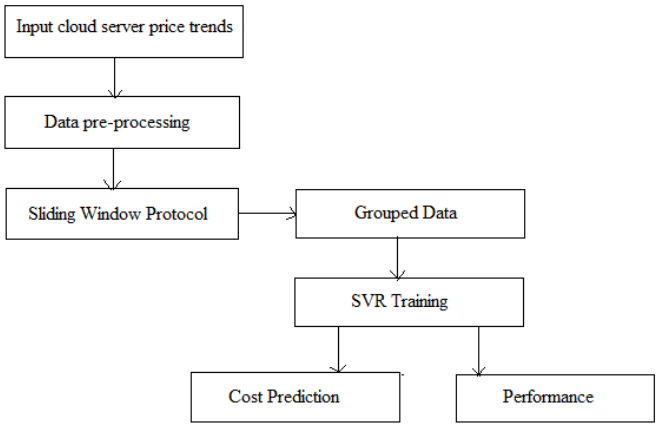

Figure 2.1 Proposed methodology

Input cloud server price trends: Any machine learning data model first need to train with the examples. Therefore the system needs a training dataset by which the modeling of data can be performed. After the learning of system, system is able to predict the similar trend values for new input samples. In this presented work the learning examples or training dataset is taken from the Amazon spot instance price dataset. That data set contains a set of values i.e. server instance, price and time stamp.

Data pre-processing: The data preprocessing is a technique by which the data is enhanced for making data acceptable for the learning algorithms. Therefore in preprocessing the data is cleaned and transformed to reduce noise and unwanted attributes from data. in this presented work the data pre-processing is carried out for removing unwanted attributes from the dataset. After cleaning of data it is moved for next phase for process.

Sliding window protocol: Sliding window protocol is developed in this system for making group of similar data instance. Sliding window one by one scanning the data instances and all the similar attributes are combined in a group.

Grouped data: The generated group of data instances is prepared and now it is ready to produce to the learning algorithm.

SVR training: SVR (support vector regression) is a data mining model which works in supervised manner. Supervised learning algorithms need the set of attributes and possible values which are needed to be predicted. SVR is a variant of SVM (support vector machine). In this model the objective function is optimized on the basis of regression analysis therefore that variant of support vector machine is termed as the SVR (support vector regression). In this system for implementing the support vector regression model LibSVM library is used. That library contains the methods of training and testing with the data sample values.

Cost prediction: After training of the model given by the LibSVM the new instance attributes are passed as input to the system. Additionally according to the previously analyzed data the cost of new instance is predicted.

Performance: After prediction of price of the data sample the performance of system prediction is measured to know the accuracy of predicted price and the resources consumed for prediction. Therefore the different performance factors are computed based on the system's outcomes i.e. accuracy, error rate and time and memory usages.

C. Algorithm Steps

The following processes are taken place for predicting the values of instance price data. The process steps are reported in table 2.1.

Table 2.1 Proposed algorithm

Input: training dataset T, testing samples Ts Output: predicted price values PV
Process: 1. $R = readInputData(T)$ 2. $P_n = preProcessData(R)$ 3. $for(i = 1; i \leq n; i++)$ a. $S = SelectSingleSample(P_i)$ b. $for(j = 1; j \leq n; j++)$ i. $if(S == P_j)$ 1. $SW.Add(P_j)$ ii. $end\ if$ c. End for 4. End for 5. $TrainModel = SVR.Train(SW)$ 6. $PV = TrainModel.Predict(Ts)$ 7. Return PV

RESULT ANALYSIS

In this, we provide the detailed understanding about the evaluated results of the proposed dynamic price prediction of spot market instance. Therefore, here we included different performance parameters and their description on which the proposed system is evaluated on the basis of different experimentation.

A. Accuracy

The accuracy of a predictive system is the unit of closeness of a quantity to that required to be predict in other words the

amount of correctly classified samples that are need to be predict is termed as predictive system accuracy. The accuracy of the system is computed using the following formula:

$$\text{Accuracy} = \frac{\text{Total correctly identified patterns}}{\text{Total Patterns to classify}} X100$$

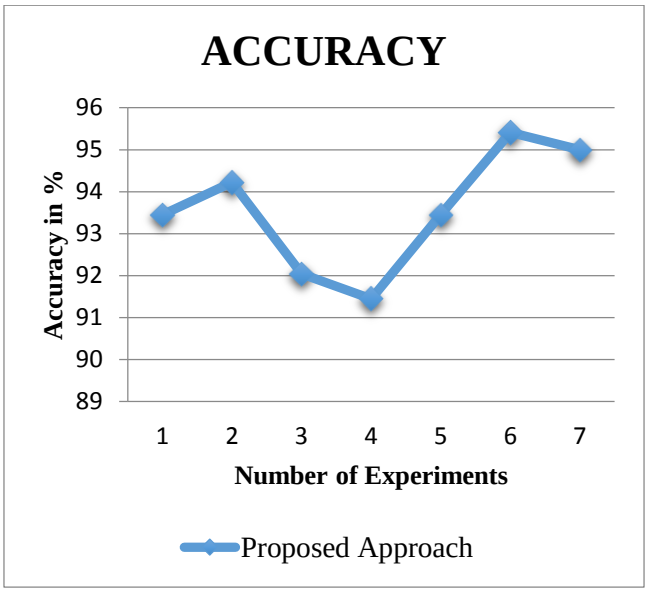

Figure 3.1 Accuracy

The figure 5.1 shows the proposed prediction model of spot instances implemented using support vector regression method. Additionally the obtained performance data is also given in table 5.1. In this diagram X-axis contains the different experiments to be performed and Y-axis shows accuracy in terms of percentage. In order to represent the performance of implemented prediction model the blue line shows the performance of the proposed approach. According to the graph and table results the proposed technique provides high accurate results for market price prediction of spot instances.

Table 5.1 Accuracy Performance

Number of Experiments	Proposed Prediction Approach
1	93.45
2	94.22
3	92.04
4	91.45
5	93.45
6	95.41
7	95

B. Error Rate

Error rate is the measurement of the performance in terms of misclassification rate. In other words the amount of data which is not correctly recognized using the trained classifier is termed as the error rate of the system. That can be evaluated using the following formula:

$$\text{Error Rate} = 100 - \text{Accuracy}$$

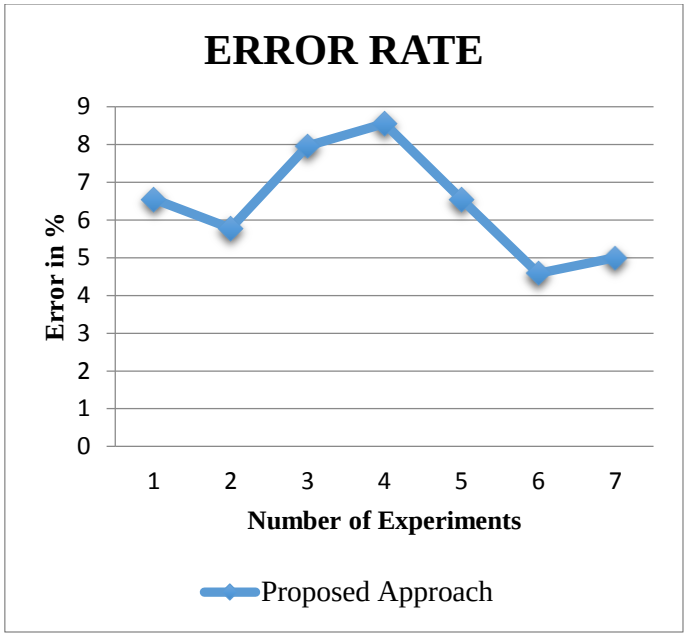

Figure 3.2 Error Rate

The graph 3.2 represents the Error rate of implemented proposed regression analysis based spot price prediction. In this graph the blue line represents the error rate of the proposed technique. The X-axis of the given graph demonstrates the number of experimentation performs and the Y-axis shows error rates are obtained by implement to algorithm. According to the obtained result error in percentage is very fewer to and the proposed system delivers the minimization of data error.

Table 3.2 Error Rate Performance

Number of Experiments	Proposed Prediction Approach
1	6.55
2	5.78
3	7.96
4	8.55
5	6.55
6	4.59
7	5

C. Time Consumption

The amount of time required to learn the given patterns from input training samples using the implemented prediction algorithm is known as the time consumption or time complexity of the implemented algorithms.

$$\text{Time Consumption} = \text{End Time} - \text{Start Time}$$

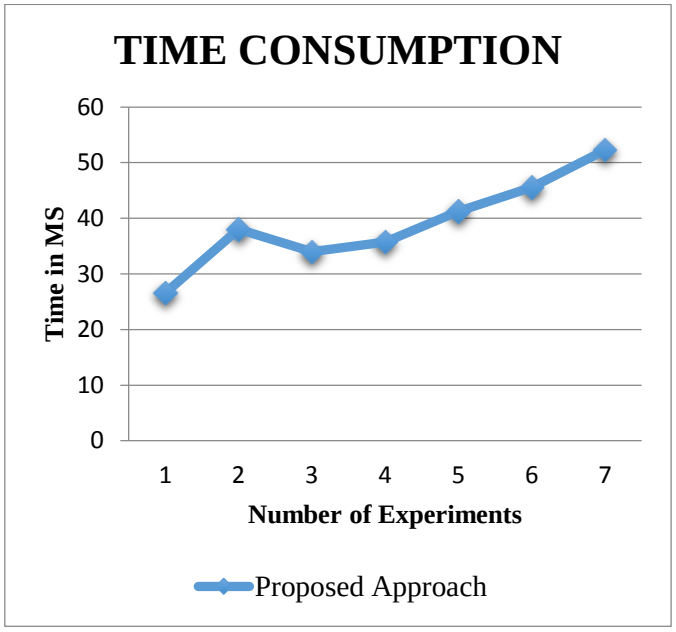

Figure 3.3 Time Consumption

The figure 3.3 and table 3.3 contains the time consumption of the proposed SVR based algorithm in terms of milliseconds. To represent the performance of the classifier the proposed technique is demonstrated using the blue line shows the performance of the proposed approach. According to the given results the proposed technique consume less time for training of input of spot instances and produce efficient result. Therefore the proposed system is less time consuming for prediction of the data patterns.

Table 3.3 Time Consumption Performance

Number of Experiments	Proposed Prediction Approach
1	26.55
2	38.01
3	33.98
4	35.74
5	41.23
6	45.54
7	52.25

D. Memory Consumption

The quantity of main memory required to implement the algorithm is termed as the space complexity of the system. That is sometimes also called the memory consumption of the algorithms.

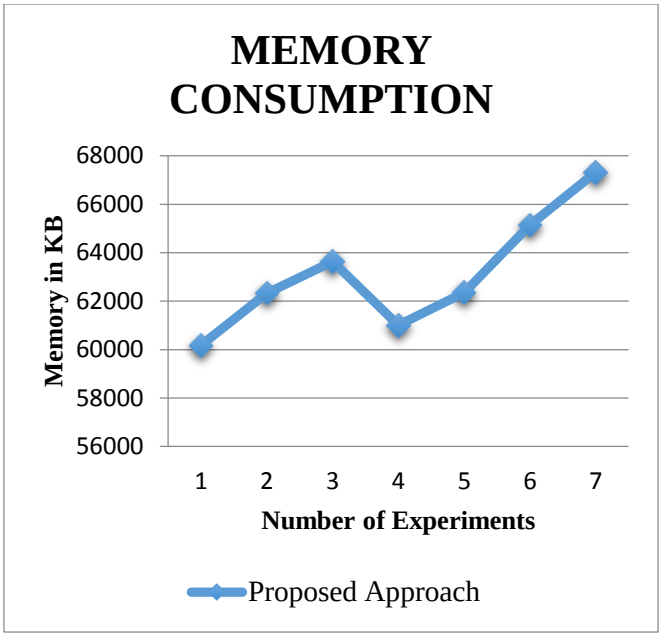

Figure 3.4 Memory Consumption

The performance of the proposed spot instance prediction using data mining approach is given using data mining technique demonstrate using figure 3.4 and table 3.4. In this diagram the X-axis includes different run of project execution, and Y-axis shows the amount of memory consumed during processing of the data. The given memory consumption is provided here in terms of kilobytes. According to the evaluated results the performance of the proposed technique is more efficient for predicting market price of spot instances.

Table 3.4 Memory Consumption Performance

Number of Experiments	Proposed Prediction Approach
1	60186
2	62332
3	63636
4	61002
5	62351
6	65147
7	67321

CONCLUSION

This chapter includes the conclusion and the possible future extension of the proposed work. Therefore the obtained results and the experiments based summary of the work are given and the observation based future work plan is proposed in this chapter.

A. Conclusion

Data mining and machine learning approaches are help us for processing large amount of data and obtain the data trends according to the needs of applications. The machine learning models are can be supervised and unsupervised in nature. The

supervised techniques are basically much accurate then the unsupervised approaches therefore the proposed work includes the study and implementation of supervised learning model for Amazon spot instance price prediction. The need of spot instance prediction is because the cloud server computational resources are required for executing the applications hosted in cloud servers. Additionally the resource requirements of the applications are varying according to the time and work load on the applications.

In this context all the application server providers are not maintaining the fixed resources due to their cost. Therefore to find the variation on the computational resource price changes according to the time the prediction of spot instance price is required. in this presented work a model using data mining technique is proposed for implementation and design. That model works in three phases, in first phase the data is collected and their pre-processing is performed. The preprocessing improves the quality of data and makes it suitable to accept for the algorithms. In next the jobs are categorized according to the method of sliding window protocol. The grouping of data helps to understand the total amount different pattern available in datasets. Finally the learning and testing of the predictive model is performed for predicting the current spot instance price.

The proposed model is implemented using the JAVA technology. And to implement the system NETBEANS IDE is used. In addition of that for implementation of SVR algorithm the LibSVM library is used. Finally for experimentation with the system the Amazon spot instance dataset is used. After successfully implementation the proposed technique is evaluated on different performance parameters. The performance based summary is reported using table 4.1.

Table 4.1 Performance summary

S. No.	Parameters	Remark
1	Accuracy	The accuracy of the proposed system is acceptable and provides high accurate results i.e. 90-96%
2	Error rate	The error rate of the system is also acceptable and it produces 5-10% of misclassification results
3	Memory usages	Memory consumption is not much vary according to the different datasets therefore it is acceptable for real world use
4	Time consumption	Time consumption of the proposed model is varying according to the input dataset size for processing with the system.

According to the performance demonstrated in summary table the performance of the proposed spot instance price prediction model accurate and less resource consuming. Therefore the proposed technique is acceptable for future extension and more optimization.

B. Future Work

The main aim of the proposed work is to obtain high performance and accurate prediction for the spot instance price. Therefore in this work a data mining technique based model is reported and results shows their effectiveness for price prediction. In near future the following future extension is suggested for work.

1. The current approach only implements the traditional SVR model for predicting the prices of spot instances. In near future it is planned to implement a hybrid classifier for optimize the performance of existing technique
2. It is also suggested to improve the current technique using the possible ensemble learning technique such as bagging and boosting

REFERENCES

[1] Singh, Vivek Kumar, and Kaushik Dutta, "Dynamic price prediction for Amazon spot instances", In 2015 48th Hawaii International Conference on System Sciences (HICSS), pp. 1513-1520, IEEE, 2015.

[2] Karandeep Kaur. "A Review of Cloud Computing Service Models", International Journal of Computer Applications (IJCA), Volume 140 – No.7, April (2016)

[3] Kelvin, "Basic Overview on Cloud Computing" [Accessed on 20 June 2017], available Online at: https://www.hostdepartment.com/blog/2014/08/05/cloud-computing/

[4] "What is cloud computing? A beginner's guide", [Accessed on 25 June 2018], available online at: https://azure.microsoft.com/en-in/overview/what-is-cloud-computing/

[5] "The Benefits of Cloud Computing", [Accessed on 25 June 2018], available online at: https://www.ibm.com/ibm/files/H300444G23392G14/13 Benefits_of_Cloud_Computing_634KB.pdf.

[6] Karger D, Stein C, Wein J. Scheduling Algorithms. Algorithms and Theory of Computation Handbook: special topics and techniques, Chapman & Hall/CRC; (2010).

[7] Magalhães, Deborah, et al. "Workload modeling for resource usage analysis and simulation in cloud computing", Computers & Electrical Engineering 47 (2015): 69-81.

[8] The State of the Amazon Ec2 Spot Market Research, Conclusions And Opportunities", 2016, available online at: https://spotinst.com/files/white-papers/spotinst-the-state-of-the-amazon-spot-market-white-paper.pdf

[9] "What Are Spot Instances?" available online at: https://spotinst.com/what-are-ec2-spot-instances/

[10] "Spot Instances", available online at: https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-spot-instances.html

[11] Armbrust, Michael, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, "A view of cloud computing", Communications of the ACM 53, Number 4 (2010), pp. 50-58.

[12] Lampe, Ulrich, Melanie Siebenhaar, Apostolos Papageorgiou, Dieter Schuller, and Ralf Steinmetz, "Maximizing cloud provider profit from equilibrium price auctions", In Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on, pp. 83-90, IEEE, 2012.