UNIVERSITÄT
PADERBORN

PROJECT GROUP AICON: ARTIFICIAL INTELLIGENCE FOR COMPUTER NETWORKS

SEMINAR PAPER: ADAPTIVE PREDICTION MODELS FOR DATA CENTER RESOURCES UTILIZATION ESTIMATION

LAMEYA AFROZE (6869313)

Outline

- Introduction
- Proposed System Methodology
- Proposed Model Evaluation
- Experimental Results
- References

Introduction

- Accurate estimation of resource utilization is important for minimizing the operational cost and maximizing the performance of data center.
- In this paper the authors has presented an adaptive multi-methods approach which automatically selects the most promising machine learning method to estimate resources utilization of data center.

Proposed System Methodology

Workload Prediction

- Resource utilization logs are divided into fixed size sliding windows with specific time interval.
- Four machine learning models for predicting results. The models are Linear Regression, Support vector regression, Gradient boosting and Gaussian Process Regression (in paper mentioned as Kriging).

Adaptive Model Selector (AMS)

- Predicts the best regression model for each time stamp.
- Use 5 classification model to choose the regression model. These models are K-Nearest Neighbors, Naïve Bayes, Multilayer Perception, Random Decision Forest (RDF) and Gradient Boosting.
- Generate workload using the selected regression model.

Proposed Model Evaluation

Datasets

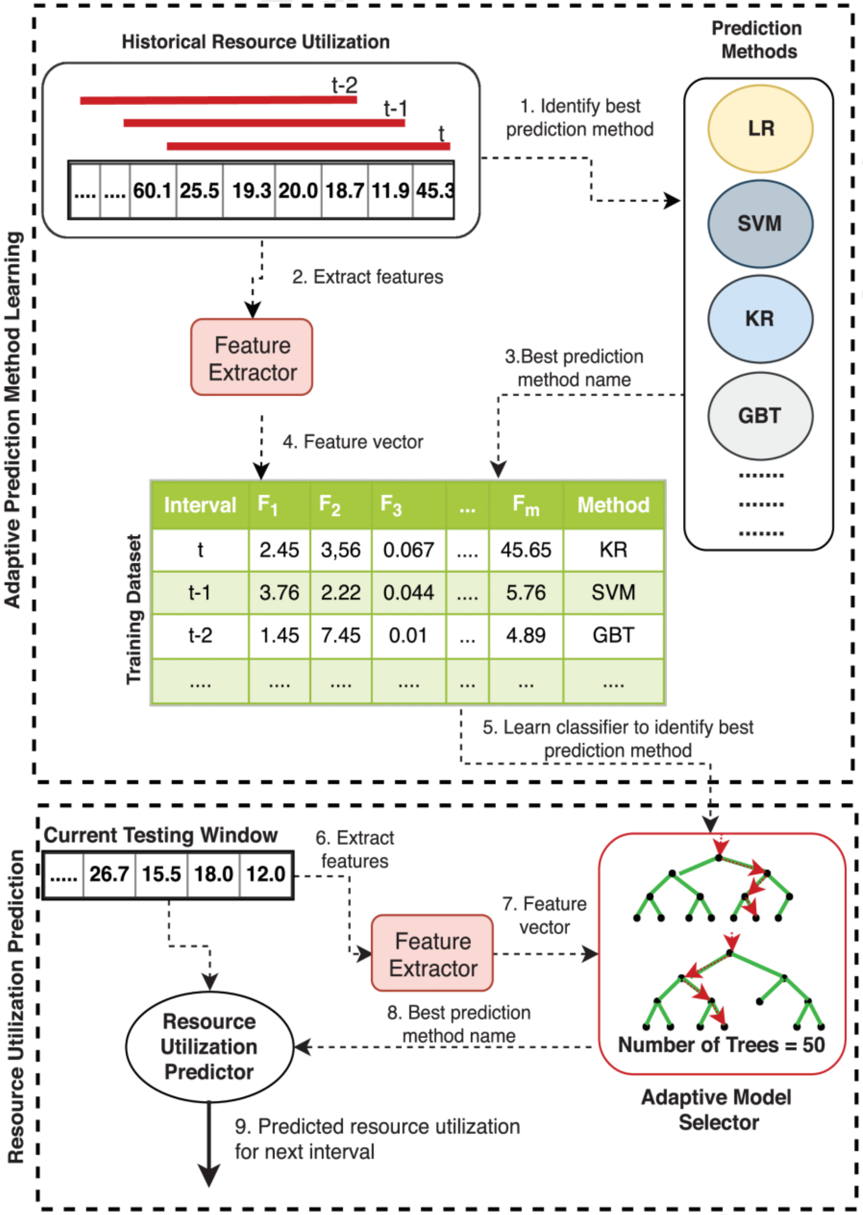
- Alibaba datasets [1]
- Bit brain datasets [2]
- Google cluster traces [3]

Feature Extraction & AMS Evaluation

- Use several features (prediction time, size, accuracy, precision, recall, F-measure, positive rate etc.) to extract datasets.
- Create training dataset with best prediction method for each time interval.
- Predict best model using AMS.

Resource Estimation

- Perform random split 80% / 20% of training / validation sets to estimate resource.
- Calculate Root-Mean Square Error (RMSE) & Mean Absolute Error (MAE) to measure the error of each regression model.


Adaptive Prediction Method Learning
Historical Resource Utilization
t-2
t-1
t
1. Identify best prediction method
Prediction Methods
LR
SVM
KR
GBT
....
....
....
2. Extract features
Feature Extractor
3.Best prediction method name
4. Feature vector
Training Dataset
5. Learn classifier to identify best prediction method
Resource Utilization Prediction
Current Testing Window
6. Extract features
Feature Extractor
7. Feature vector
8. Best prediction method name
Resource Utilization Predictor
Number of Trees = 50
Adaptive Model Selector
9. Predicted resource utilization for next interval

....	....	60.1	25.5	19.3	20.0	18.7	11.9	45.3

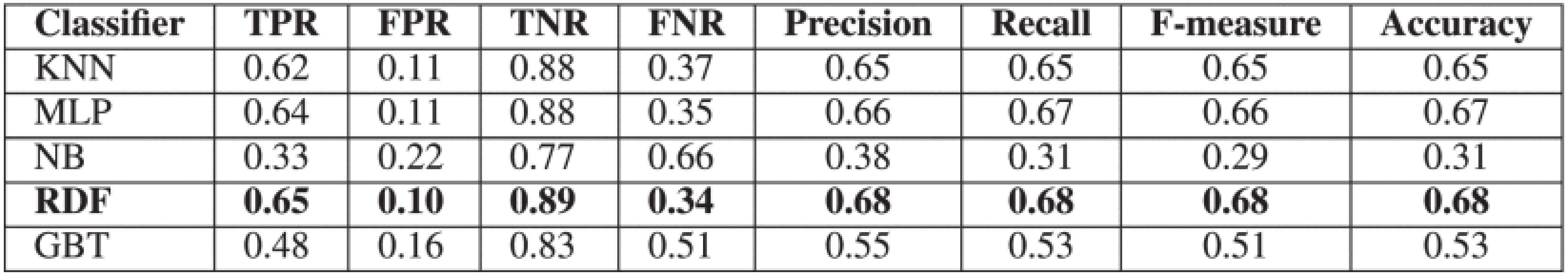
Interval	F_1	F_2	F_3	...	F_m	Method
t	2.45	3,56	0.067	....	45.65	KR
t-1	3.76	2.22	0.044	....	5.76	SVM
t-2	1.45	7.45	0.01	...	4.89	GBT
....	....	....	....	...	...	....

.....	26.7	15.5	18.0	12.0

Figure 1: Purposed system overview to learn adaptive model selector and using it to estimate the data center resource utilization. Source:[4], Figure 2

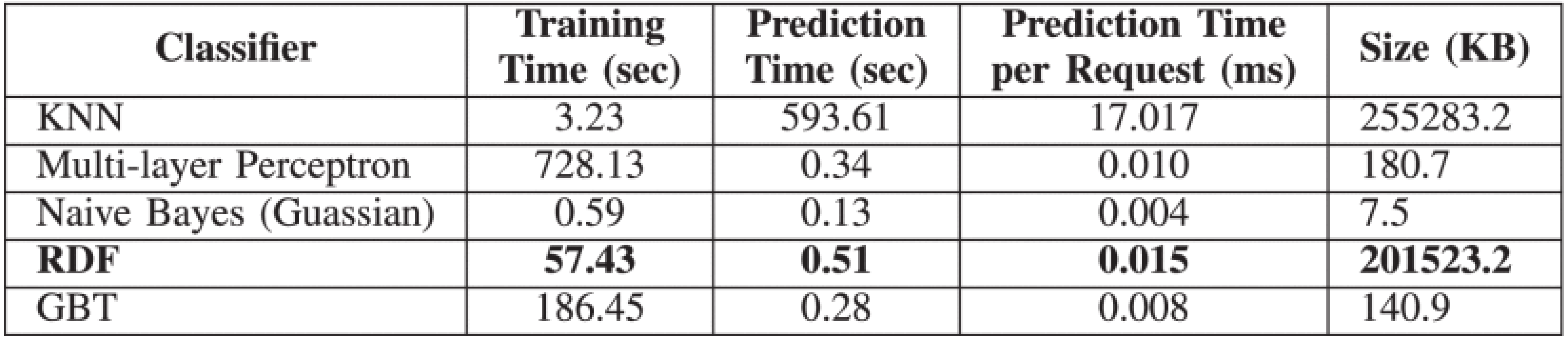
Experimental Results (1)

AMS Evaluation



Classifier	TPR	FPR	TNR	FNR	Precision	Recall	F-measure	Accuracy
KNN	0.62	0.11	0.88	0.37	0.65	0.65	0.65	0.65
MLP	0.64	0.11	0.88	0.35	0.66	0.67	0.66	0.67
NB	0.33	0.22	0.77	0.66	0.38	0.31	0.29	0.31
RDF	**0.65**	**0.10**	**0.89**	**0.34**	**0.68**	**0.68**	**0.68**	**0.68**
GBT	0.48	0.16	0.83	0.51	0.55	0.53	0.51	0.53

Table 1: AMS evaluation results using different classifiers for Alibaba dataset[4].

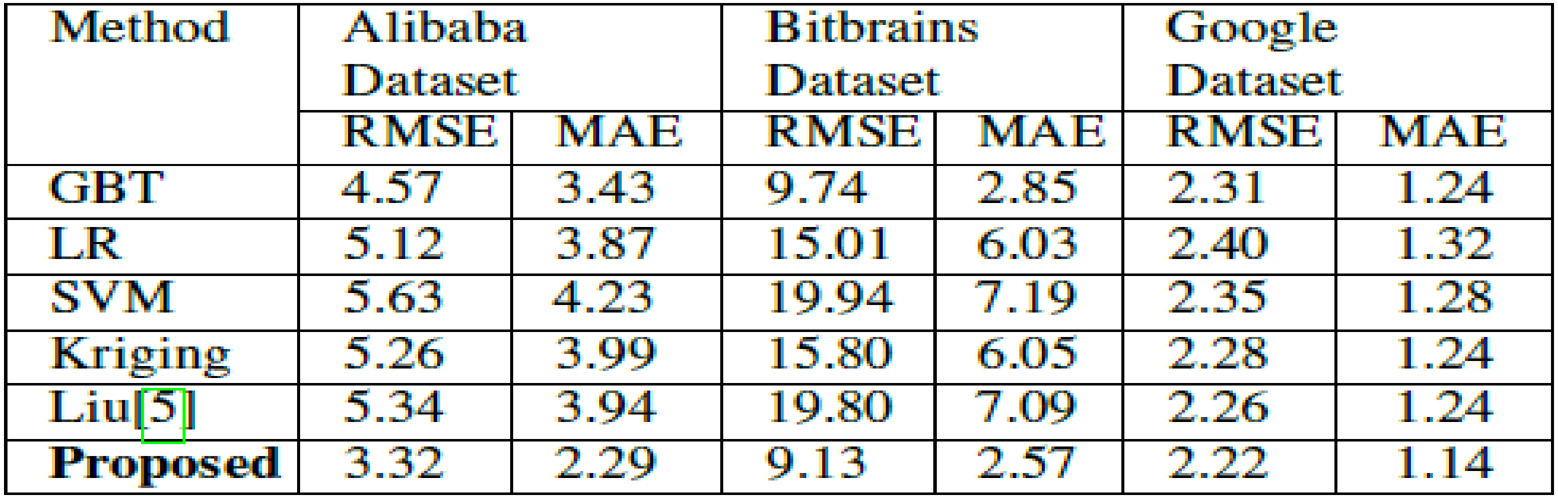


Classifier	Training Time (sec)	Prediction Time (sec)	Prediction Time per Request (ms)	Size (KB)
KNN	3.23	593.61	17.017	255283.2
Multi-layer Perceptron	728.13	0.34	0.010	180.7
Naive Bayes (Guassian)	0.59	0.13	0.004	7.5
RDF	**57.43**	**0.51**	**0.015**	**201523.2**
GBT	186.45	0.28	0.008	140.9

Table 2: Time & space efficiency of AMS using different classifiers for Alibaba dataset[4].

Experimental Results (2)

Dataset Resource Estimation & Window Size Sensitivity



Method	Alibaba Dataset		Bitbrains Dataset		Google Dataset	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
GBT	4.57	3.43	9.74	2.85	2.31	1.24
LR	5.12	3.87	15.01	6.03	2.40	1.32
SVM	5.63	4.23	19.94	7.19	2.35	1.28
Kriging	5.26	3.99	15.80	6.05	2.28	1.24
Liu[5]	5.34	3.94	19.80	7.09	2.26	1.24
Proposed	3.32	2.29	9.13	2.57	2.22	1.14

Table 1: RMSE and MAE for resource estimation using proposed system for three different datasets[4].

Experimental results (3)

Dataset Resource Estimation & Window Size Sensitivity

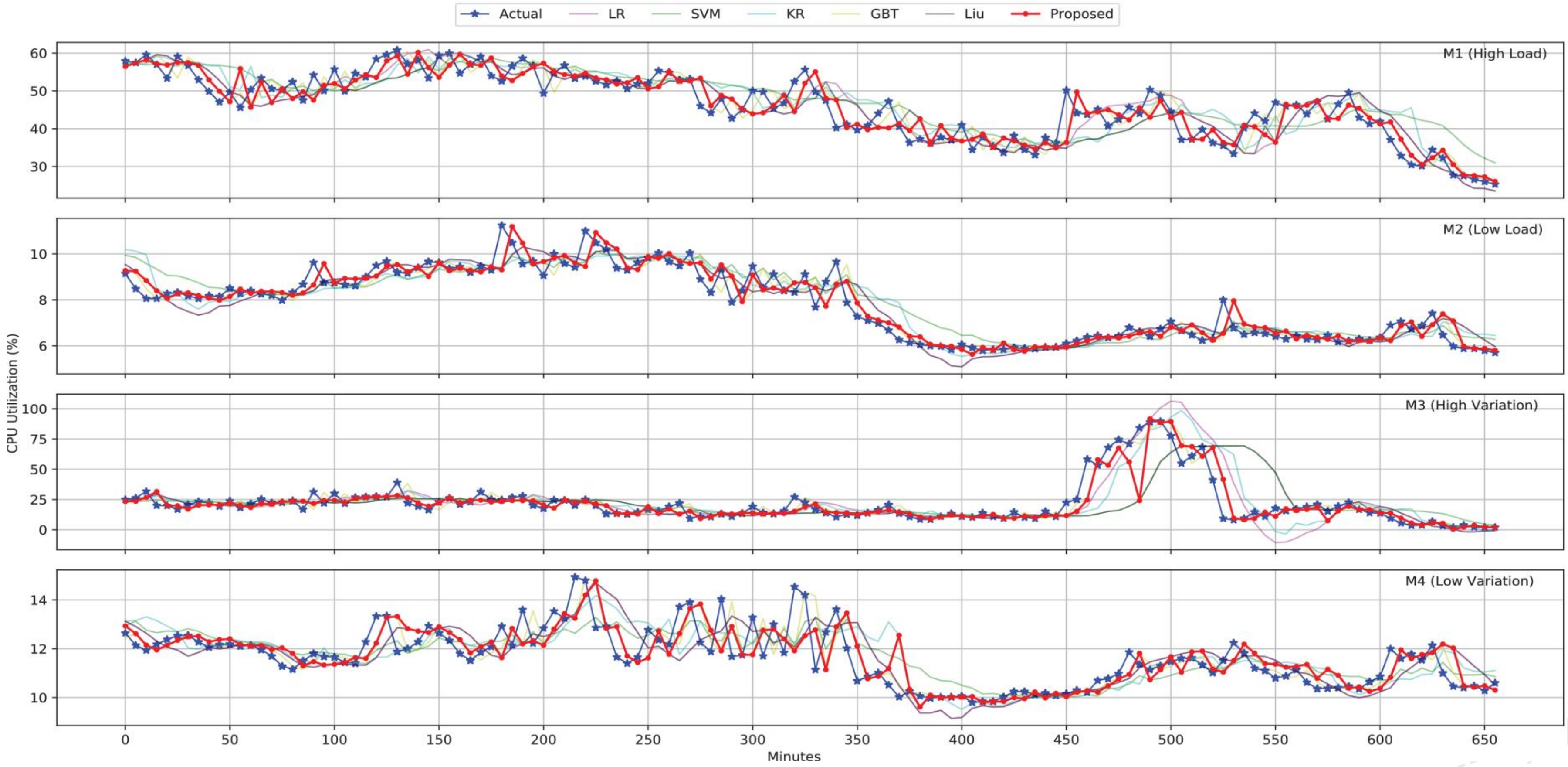


Figure 2:Actual vs proposed method CPU prediction for Alibaba data set for 4 selected machines. M1 =Heavy workload, M2 = Low workload, M3 = High variation, M4 = Low variation. The window size used to train the prediction model is 60 minutes. Source:[4], Figure 10

References

[1] *Alibaba Cluster Log*. Available url =https://github.com/alibaba/clusterdata. Accessed:16-May-2020.

[2] *Bitbrains Cluster Log*. Available url =http://gwa.ewi.tudelft.nl/datasets/gwa-t-12-bitbrains. Accessed: 16-May-2020.

[3] *Google Cluster Log*. Available url =https://github.com/google/cluster-data. Accessed:16-May-2020.

[4] S. Baig, W. Iqbal, J. L. Berral, A. Erradi, andD. Carrera. “Adaptive Prediction Models for Data CenterResources Utilization Estimation”.In:IEEE Transactions on Network and Service Management16.4 (2019), pp. 1681–1693.

[5] C. Liu, C. Liu, Y. Shang, S. Chen, B. Cheng, and J. Chen. “An adaptive prediction approach based on workload pattern discrimination in the cloud”. In: *Journal of Network and Computer Applications* 80 (2017), pp. 35–44.