

Adaptive Prediction Models for Data Center Resources Utilization Estimation

Lameya Afroze

Summer term 2020

This paper presents a new technique that adaptively and automatically selects the most promising machine learning method to estimate the data center resources utilization.

1 Introduction

The introduction describes why accurate estimation of resource utilization is important, why this accurate estimation is challenging, the main contributions of this paper.

2 Proposed System Methodology

2.1 Workload prediction using machine learning

How workload is predicted using different machine learning classifier.

2.2 Adaptive Model Selector(AMS) using regression models

How various regression model is used to build AMS model for predicting the best model for given sliding window data.

3 Proposed model evaluation

The section describes the evaluation of proposed model step by step.

3.1 Datasets

Presents three different datasets that has been used to conduct the experiment.

3.2 Feature Extraction

- How features are selected manually and open source library(TSFRESH).
- Use three methods from TSFRESH to filter data

3.3 AMS Evaluation

- Trained datasets by selecting features and using regression several model
- validate proposed model using three different datasets
- selects the best predictor for recent resource utilization

3.4 Resource Estimation

Evaluate AMS with the help of Root-Mean Square Error (RMSE) and Mean Absolute Error (MAE)

3.5 Window Size Sensitivity

Test window size of 20, 40, 60, 80 and 90 minutes to train and validate AMS model

4 Experimental Results

Discuss several evaluation results that has been produced during the time of experiment .

4.1 AMS Evaluation

4.2 Resource Estimation of Three Datasets

4.3 Window size sensitivity

5 Discussion

5.1 Identify Best Prediction Model

5.2 Comparison of Proposed Model with Existing Works

6 Conclusion

References

- [1] M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr. “Time series feature extraction on basis of scalable hypothesis tests (tsfresh—a python package)”. In: *Neurocomputing* 307 (2018), pp. 72–77.
- [2] 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.
- [3] T. K. Ho. “Random decision forests”. In: *Proceedings of 3rd international conference on document analysis and recognition*. Vol. 1. IEEE. 1995, pp. 278–282.