

# Towards Optimal Resource Utilization in Data Centers using Long Short-Term Memory

Laboratory for  
Software  
Design & Analysis

established in 2005

Kundjanasith Thonglek, Kohei Ichikawa, Keichi Takahashi, Chawanat Nakasan, Hajimu Iida

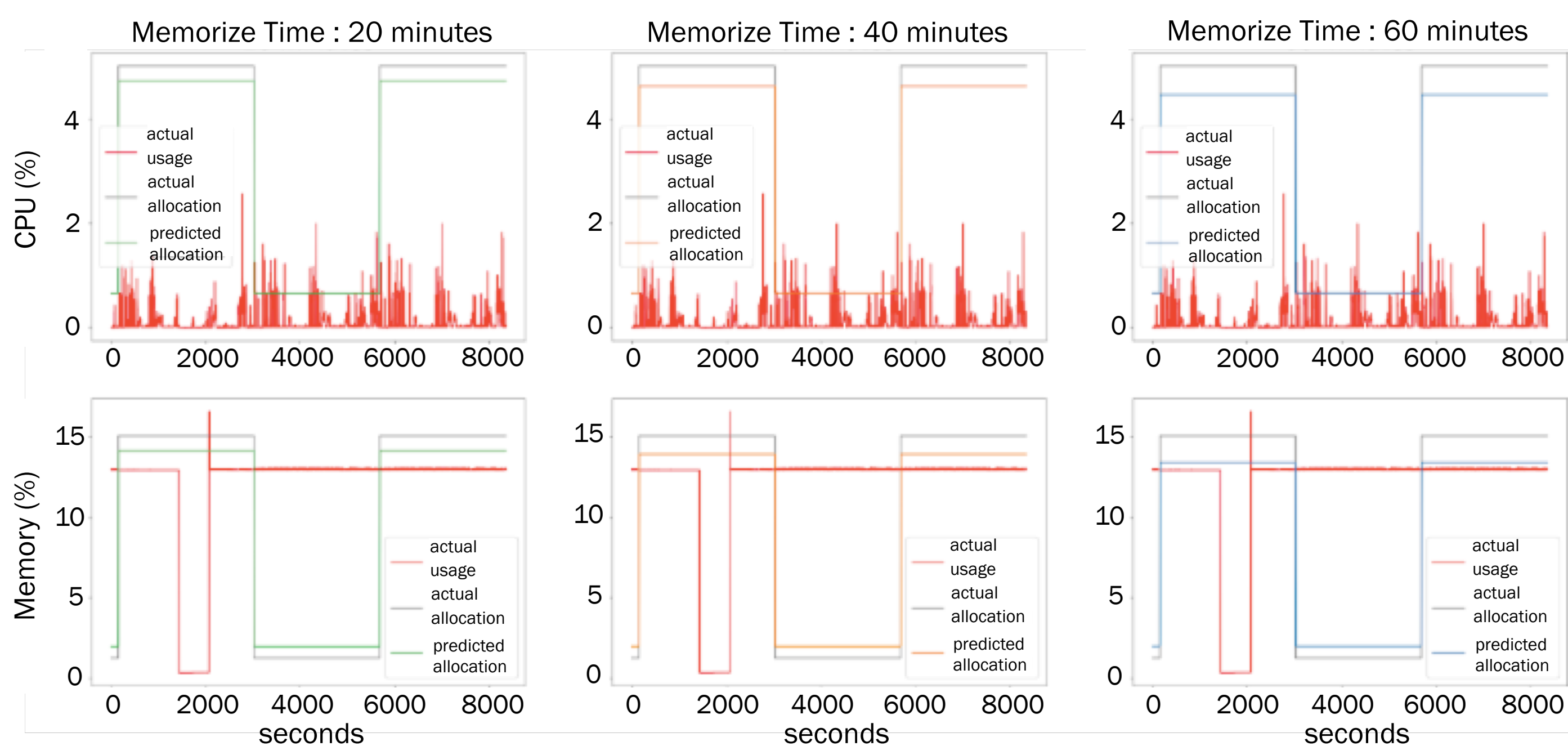
Email: thonglek.kundjanasith.ti7@is.naist.jp

## Introduction

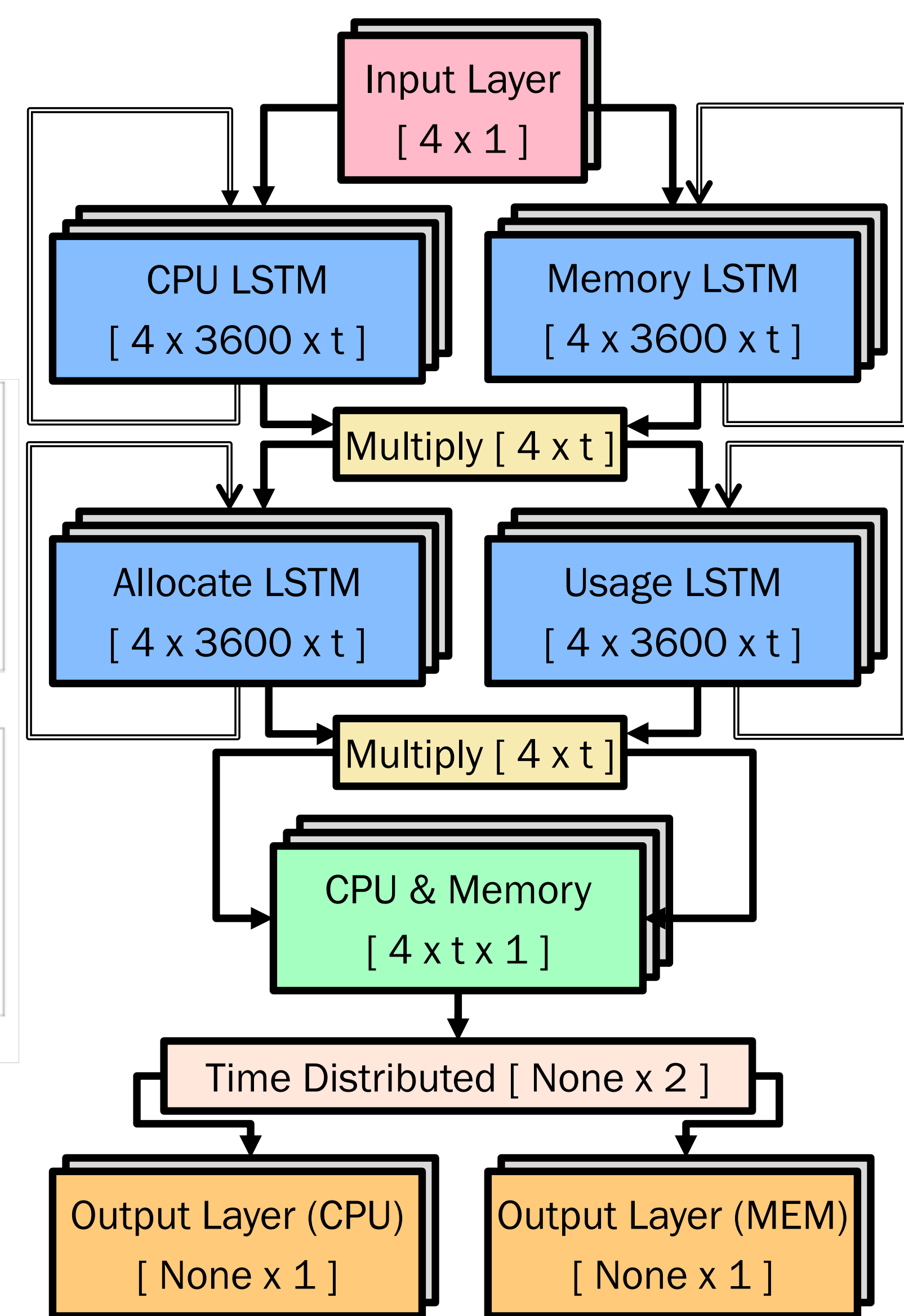
Data Centers are simply centralized resources where computing and networking equipment is concentrated for the several purpose application to handle large amounts of data. Applications have to allocate the resources for execution but almost usually resource usage is less than their allocated resources. This is a common problem in resource utilization. The objective of this work is investigating optimized resource utilization for real workload in the data center. We applied Long Short-Term Memory, is a time-series deep learning technique, to predict the optimized amount of allocated resources. Google's cluster usage data, which is open data of real workload in the data center, is used as training dataset.

## Long short-term memory optimize allocated resource

Long Short-Term Memory or LSTM is a kind of recurrent neural network which capable of learning long-term dependencies. We applied the amount of computing resources is allocated and used to be the model features. CPU and memory are computing resources are used to our approaches. Moreover, we vary the memorize time interval in the memory cell to 20, 40 and 60 minutes to observe memory time in LSTM is significant for predictive allocated resources.

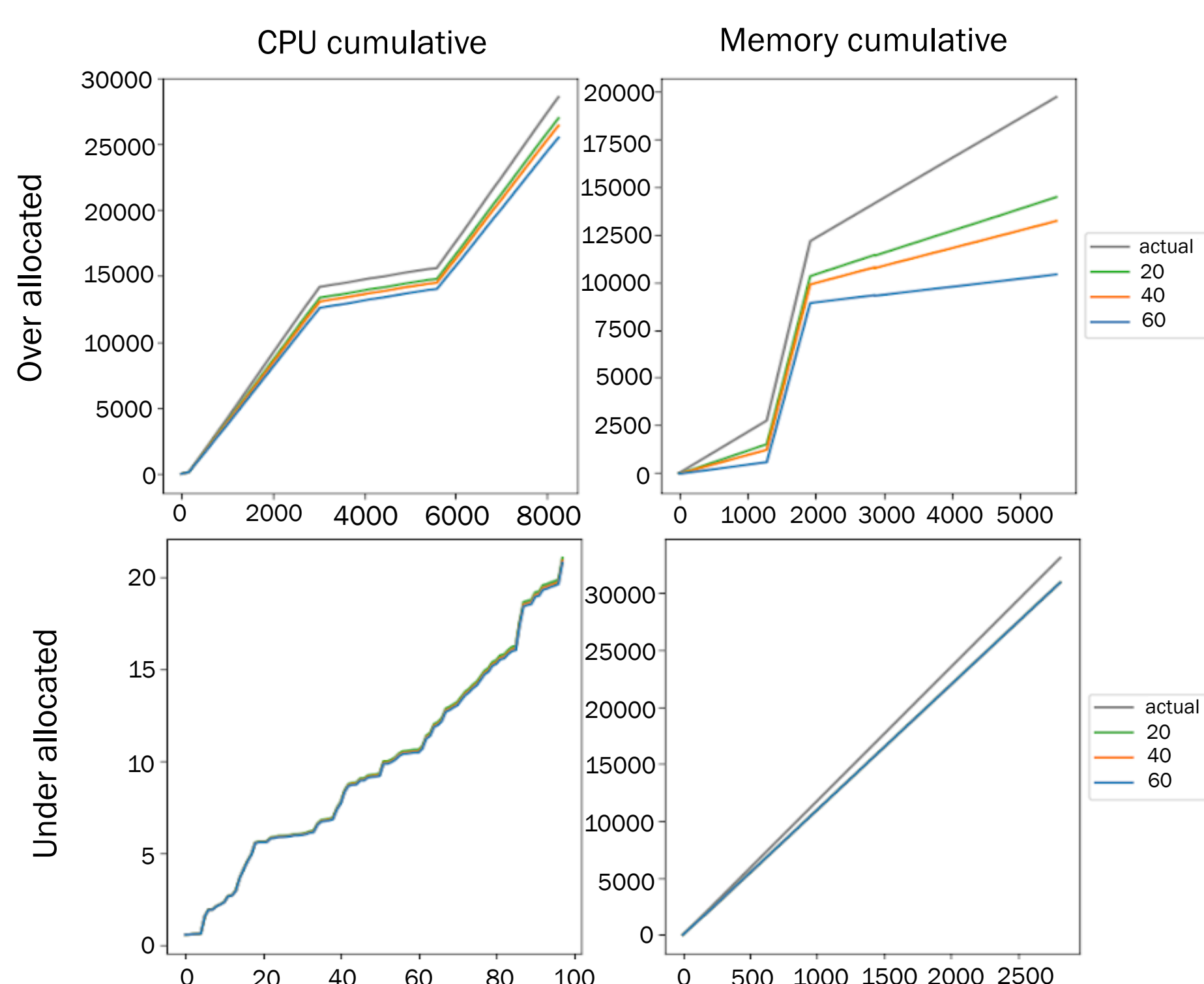


The above graph shows the actual resource usage and allocated resources using LSTM model with the difference of memorizing time interval. It indicates that LSTM is able to predict the allocated resource by reducing the wasted resources of the allocated resource without using.



Long Short-Term Memory model architecture

## Resource utilization in data center



We divided resource allocation into two allocation type. (1) Allocated resources are more than resource usage (2) Allocated resources are less than resource usage. The left graph shows the cumulative wasted resource from start the job until complete executing.

As the result, we use absolute value of the difference value between allocated resources and resources usage to be the evaluated matrix of resource utilization.

The resource utilization rate of difference memorize time interval in memory cell is not a significant deviation. It indicates that varying the memorize time is not much effect to predict allocated resources but the maximum memorize time is better resource utilization rate than the lower memorize time.

Our future work is applying the other time-series forecasting techniques to optimize the resource utilization.