The Journal of Supercomputing
https://doi.org/10.1007/s11227-022-04782-z

Multivariate workload and resource prediction in cloud computing using CNN and GRU by attention mechanism

Javad Dogani[1] · Farshad Khunjush[1] · Mohammad Reza Mahmoudi[2] · Mehdi Seydali[1]

Accepted: 21 August 2022

Abstract

The resources required to service cloud computing applications are dynamic and fluctuate over time in response to variations in the volume of incoming requests. Proactive autoscaling techniques attempt to predict future resource demand and assign resources before the request. Thus, establishing accurate prediction methods optimizes resource allocation and avoids service-level agreement violations (SLA). Most existing methods for predicting resource usage in cloud computing rely on univariate time series prediction models. These models consider only a single resource usage metric, make future predictions based on historical data for the target metric, and provide predictions for a single step ahead. This paper provides a hybrid method for multivariate time series workload prediction of host machines in cloud data centers that predicts the workload for the next few steps. First, a statistical analysis is used to construct the training set. Then, a convolutional neural network (CNN) is employed to extract the hidden spatial features between all correlated variables. Finally, the spatial features extracted by the CNN are fed into a GRU network optimized with the attention mechanism in order to extract the temporal correlation features. Two experiments employing Google cluster data were conducted to evaluate the proposed method. Experimental results reveal that our method improves prediction accuracy by 2% to 28% compared to baseline methods and previous research.

Keywords Cloud computing · Workload prediction · Multivariate time series · Deep learning · Attention mechanism

✉ Farshad Khunjush
khunjush@shirazu.ac.ir

Extended author information available on the last page of the article

Published online: 04 September 2022