

WORKLOAD PREDICTION METHODS AND APPARATUSES FOR SERVICE IN SERVICE CLUSTER

ABSTRACT

Embodiments of this specification provide a workload prediction method and apparatus for a service in a service cluster. The method includes: obtaining n+1 indicator sequences respectively corresponding to services in a same historical time period, which include n status indicator sequences corresponding to n system status indicators and a load indicator sequence corresponding to a workload indicator; calculating a correlation coefficient of a same indicator between different services based on the n+1 indicator sequences corresponding to the services, and constructing a similarity graph corresponding to the same indicator, to obtain n+1 similarity graphs; processing the n+1 similarity graphs by using an attention mechanism, to obtain an aggregated similarity graph; performing, by using a first convolutional network, graph convolutional processing on an initial representation of each service in the aggregated similarity graph, to obtain an aggregated representation corresponding to each service; and performing workload prediction based on the aggregated representation.

FIG. 2