

Construction of spatio-temporal data mining system for time-series satellite imagery using Hadoop

NISHIMAE, Kou¹ ; MIYOSHI, Tomoya¹ ; HONDA, Rie^{1*}

¹Kochi University

A large number of spatio-temporal data have been stored in various fields of science, such as remote sensing, numerical simulation, and astronomical observation, in which data often appears as time-series images. To extract spatio-temporal knowledge from spatio-temporal data including time-series images, spatio-temporal cross section relevant to a target task has to be extracted from a mass of data. Since these data are stored as a large number of files, utilization of distributed processing framework such as Hadoop or Gfarm is promising.

We constructed distributed data mining system for time-series satellite images using 53 nodes (3 masters and 50 slaves at maximum) of iMac and Hadoop which enables distributed file system and distributed processing using MapReduce. We evaluated the scalability and performance of the system for the task extracting time-series data from a large number of images carefully and found that partitioning the images into optimum numbers and reducing the data between map phase and reduce phase is essential.

The system was then applied to two different tasks focusing on time-series data analysis extracted from satellite imagery: statistical modeling of seasonal changes in vegetation index and spatio-temporal correlation analysis of weather satellite images. The tasks were successfully implemented on the system and the computational time was decreased in inverse proportion to the number of slave nodes, thus usefulness of the distributed system to spatio-temporal data mining for time-series images.

Keywords: distributed processing, Hadoop, MapReduce, data mining, spatio-temporal, satellite imagery