

Apriori Versions Based on MapReduce for Mining Frequent Patterns on Big Data

Abstract:

Pattern mining is one of the most important tasks to extract meaningful and useful information from raw data. This task aims to extract item-sets that represent any type of homogeneity and regularity in data. Although many efficient algorithms have been developed in this regard, the growing interest in data has caused the performance of existing pattern mining techniques to be dropped. The goal of this paper is to propose new efficient pattern mining algorithms to work in big data. To this aim, a series of algorithms based on the MapReduce framework and the Hadoop open-source implementation have been proposed. The proposed algorithms can be divided into three main groups. First, two algorithms [Apriori MapReduce (AprioriMR) and iterative AprioriMR] with no pruning strategy are proposed, which extract any existing itemset in data. Second, two algorithms (space pruning AprioriMR and top AprioriMR) that prune the search space by means of the well-known anti-monotone property are proposed. Finally, a last algorithm (maximal AprioriMR) is also proposed for mining condensed representations of frequent patterns. To test the performance of the proposed algorithms, a varied collection of big data datasets have been considered, comprising up to $3 \cdot 10^{18}$ transactions and more than 5 million of distinct single-items. The experimental stage includes comparisons against highly efficient and well-known pattern mining algorithms. Results reveal the interest of applying MapReduce versions when complex problems are considered, and also the unsuitability of this paradigm when dealing with small data.

SHIELD

SHIELD TECHNOLOGIES,

2232, 3RD FLOOR, 16TH B CROSS, YELAHANKA NEW TOWN, BANGALORE-64

Mail us: shieldtechnobl@gmail.com / manager@shieldtechno.com

Contact: 9972364704 / 8073744810