Trust-but-Verify: Verifying Result Correctness of Outsourced Frequent Item set Mining in Data-Mining-As-a-Service Paradigm

Cloud computing is popularizing the computing paradigm in which data is outsourced to a third-party service provider (server) for data mining. Outsourcing, however, raises a serious security issue: how can the client of weak computational power verify that the server returned correct mining result? In this paper, we focus on the specific task of frequent item set mining. We consider the server that is potentially untrusted and tries to escape from verification by using its prior knowledge of the outsourced data. We propose efficient probabilistic and deterministic verification approaches to check whether the server has returned correct and complete frequent item sets. Our probabilistic approach can catch incorrect results with high probability, while our deterministic approach measures the result correctness with 100 percent certainty. We also design efficient verification methods for both cases that the data and the mining setup are updated. We demonstrate the effectiveness and efficiency of our methods using an extensive set of empirical results on real datasets.