

**CHENNAI – PONDICHERRY**

**RASP-Boost: Confidential Boosting-Model Learning with Perturbed Data in the Cloud**

**Abstract:**

Mining large data requires intensive computing resources and data mining expertise, which might be unavailable for many users. With widely available cloud computing resources, data mining tasks can now be moved to the cloud or outsourced to third parties to save costs. In this new paradigm, data and model confidentiality becomes the major concern to the data owner. Data owners have to understand the potential trade-offs among client-side costs, model quality, and confidentiality to justify outsourcing solutions. In this paper, we propose the RASPBoost framework to address these problems in confidential cloud-based learning. The RASP-Boost approach works with our previous developed Random Space Data Perturbation (RASP) method to protect data confidentiality and uses the boosting framework to overcome the difficulty of learning high-quality classifiers from RASP perturbed data. We develop several cloud-client collaborative boosting algorithms. These algorithms require low client-side computation and communication costs. The client does not need to stay online in the process of learning models. We have thoroughly studied the confidentiality of data, model, and learning process under a practical security model. Experiments on public datasets show that the RASP-Boost approach can provide high-quality classifiers, while preserving high data and model confidentiality and requiring low client-side costs.

**Existing System:**

The exported data may contain private information. The Netflix prize, as a successful example of outsourced mining, has to be suspended due to privacy breach of the shared data.

The data ownership is not protected. Once published, the dataset can be accessed and used by anyone. As data has become an important property for many companies, protecting data ownership is in top priority.

The ownership of the mined models is not protected. Curious cloud service providers can easily learn the unprotected models, and make profits by sharing them with other parties, which may seriously damage the data owner’s interests.

**Proposed System:**

We identify the urgent need and the challenges of confidential classifier learning in the cloud or with datamining service providers. We also propose four measures: data confidentiality, model confidentiality, model accuracy, and client-side costs to holistically evaluate a confidential learning method.

The proposed approach can utilize the RASP perturbation method to protect data and model confidentiality, which provides much stronger confidentiality guarantee than other perturbation methods such as geometric data perturbation and random projection perturbation.

As we originally design the RASP perturbation method for outsourced database services (i.e., range queries and kNN queries) it was unknown whether it can be used for confidential cloud mining or not. We design algorithms to generate weak classifiers with random half-space range queries, and then use the boosting framework to construct high-quality models from these weak classifiers. Our experiments show that two of the four candidate methods can generate high-quality models with accuracy very close to the optimal boosting models.