Privacy-Preserving K-Means clustering over vertically partitioned data - Review

Shumin Guo

Problem Description

K-means clustering over vertically partitioned data among multiple parties brings challenges for privacy protection. This paper proposes a privacy protection protocol under this scenario.

Contribution of this paper

This paper proposes a privacy protection method by disguising the site component(attribute values) with random values, and using permuted distance comparison result for purpose of communication among different parties.

By utilizing secure circuit evaluation method, distance comparison result is calculated without disclosing anything else.

And in the proposed protocol, three parties are explicitely selected as special parties, one party P_1 is responsible for generating perturbation and permutation matrix, another party P_2 is used for final distance comparison with the third party P_r which is responsible for computing the distance of all parties except P_2 and do distance comparison with P_2. By doing this, all the parties will not be able to have all the information about another party so that privacy can be protected.

Weaknesses of this paper

No experimental results are given by this paper.

The K-means clustering accuracy is not discussed.