**Paper No: PU-SOE-CSE-13**

**Subspace-Based Aggregation for Enhancing Utility, Information Measures, and Cluster Identification in Privacy Preserved Data Mining on High-Dimensional Continuous Data**

**Shashidhar Virupakshaa**, D.Venkatesulub

[a.](https://www.researchgate.net/institution/Velagapudi_Ramakrishna_Siddhartha_Engineering_College) [Department of CSE, Presidency University, Bangalore, India](https://www.researchgate.net/institution/Velagapudi_Ramakrishna_Siddhartha_Engineering_College)

b. Department of CSE, VFSTR (Deemed to be University), Guntur, India

**Abstract**

Clustering is a data mining technique that has been effectively used in the last few decades for knowledge extraction. Privacy is a major problem while releasing data for clustering and therefore privacy-preserving data mining (PPDM) algorithms have been developed. Aggregation is a popular PPDM technique that has been used. However, in the last few years, certain applications require that data mining be performed on high-dimensional data. The present privacy preservation techniques perform aggregation in a univariate manner along each dimension. This affects the utility measures, information measures, and especially retention of original clusters. This paper proposes a new technique called as subspace-based aggregation (SBA). SBA categorizes the dimensions into dense and non-dense subspaces based on the density of points. Aggregation is performed separately for dense and non-dense subspaces. This approach helps to maximize utility measures, information measures, and retention of clusters. SBA is run on high-dimensional continuous datasets from UCI Machine Learning repository. SBA is compared with related work methods such as SINGLE, SIMPLE, MDAV, and PPPCA. SBA provides an improvement of 66% in utility, 400% in cluster identification, 5% in co-variance, and standard deviation.

**Keywords:**

[Privacy preservation](https://www.tandfonline.com/keyword/Privacy+Preservation), [privacy preserved data mining](https://www.tandfonline.com/keyword/Privacy+Preserved+Data+Mining), [Data privacy](https://www.tandfonline.com/keyword/Data+Privacy)

**Publication Details:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Journal Name** | **Vol.** | **Month & Year** | **Page No.** | **Publisher** | **Scimago Ranking** |
| International Journal of Computers and Applications | - | Nov. 2019 | Published online | [Taylor and Francis Ltd](https://www.scimagojr.com/journalsearch.php?q=Taylor%20and%20Francis%20Ltd.&tip=pub). | Q4 |