

**CHENNAI – PONDICHERRY**

**PPHOPCM: PRIVACY-PRESERVING HIGH-ORDER POSSIBILISTIC C-MEANS ALGORITHM FOR BIG DATA CLUSTERING WITH CLOUD COMPUTING**

**Abstract:**

As one important technique of fuzzy clustering in data mining and pattern recognition, the possibilistic c-means algorithm (PCM) has been widely used in image analysis and knowledge discovery. However, it is difficult for PCM to produce a good result for clustering big data, especially for heterogenous data, since it is initially designed for only small structured dataset. To tackle this problem, the paper proposes a high-order PCM algorithm (HOPCM) for big data clustering by optimizing the objective function in the tensor space. Further, we design a distributed HOPCM method based on MapReduce for very large amounts of heterogeneous data. Finally, we devise a privacy-preserving HOPCM algorithm (PPHOPCM) to protect the private data on cloud by applying the BGV encryption scheme to HOPCM, In PPHOPCM, the functions for updating the membership matrix and clustering centers are approximated as polynomial functions to support the secure computing of the BGV scheme. Experimental results indicate that PPHOPCM can effectively cluster a large number of heterogeneous data using cloud computing without disclosure of private data.