



RESEARCH ARTICLE

Hierarchical Document Clustering Using Correlation Preserving Indexing

L. Prabhakar¹, Srikanth. B²

¹M.Tech 2nd year, Dept. of CSE, ASCET, Gudur, India

²Associate Professor, Dept. of CSE, ASCET, Gudur, India

¹ lakkuprabhakar09@gmail.com; ² BSrikanth@yahoo.com

Abstract— This paper presents a spectral clustering method called as correlation preserving indexing (CPI). This method is performed in the correlation similarity measure space. Correlation preserving indexing explicitly considers the manifold structure embedded in the similarities between the documents. The aim of CPI method is to find an optimal semantic subspace by maximizing the correlation between the documents in the local patches and simultaneously correlation in the patches outside are minimized. Correlation is a similarity measure can capture the intrinsic structure in high dimensional data. In an effort to reduce the computational cost of CPI method, we propose to apply the bi-iterative least square method to reduce the dimensions. On comparison of the effectiveness of the CPI method with other clustering methods, using the Bi-iterative least square method there has been a considerable reduction in the time computation.

Key Terms: - Document clustering; Correlation preserving indexing; Singular value decomposition; Dimensionality reduction; Correlation measure; QR decomposition

Full Text: <http://www.ijcsmc.com/docs/papers/August2013/V2I8201352.pdf>