Abstract

Data mining is defined as a techniques of solving data warehouse problem, in order that data collection which doesn't have anything information can be mined and analyzed. Data colection can't be processed directly, data with high dimensionality must be preprocessed so the output is truly maximal when data is mined.

ICA (Independent Component Analysis) which can minimize dimension and then new dimension is named IC (Independent Component) is a reduction of numeric data. Reduction in ICA is finding dimension which is independent each other to be a new dimension. This condition is stronger than finding uncorrelation between dimension.

Curse of dimensionality phenomenon can be solved by minimizing dimension using ICA. Evaluation using ICA can reduce data colon tumor with 2000 dimension extracted to be 60 IC and DLBCL dataset with 4026 dimension extracted to be 46 IC. Both datasets have the best clustering evaluation in one IC. As seen in the value of SSE in data colon tumor is 0 and the value of SSE in DLBCL is 0.97. ICA gives better clustering evaluation than PCA (Principal Component Analysis) because ICA has smaller SSE values than PCA.

Key Word: Data mining, ICA, PCA, numeric data, high dimension, independent, clustering, SSE (Sum of Squared Error)