Abstract

Document clustering is very important in the field of text categorization. Genetic algorithm, which is an optimization based technique which can be applied for finding out the best cluster centres easily by computing fitness values of data points. While clustering around weighted prototype technique is especially helpful when proper pairwise similarities are available. This technique does not find global solution of the objective function. Experimental result shows that F-measure and Normalized mutual information of genetic algorithm is better than clustering around weighted prototype for 20 Newsgroup dataset. F-measure and accuracy of genetic algorithm is better than clustering around weighted prototype for the Reuter-21578 dataset.

References

2. G. Salton, C. Buckley, Term-weighting approaches in automatic text retrieval, Inf.


Index Terms

Computer Science

Algorithms

Keywords

Clustering, Similarity Based, Genetic Algorithm, Document Categorization, Text mining.