Abstract

Clustering in data mining is a discovery process that groups a set of documents such that documents within a cluster have high similarity while documents in different clusters have low similarity. Existing clustering method like K-means is a popular method but its results are based on choice of cluster centers so it easily results in local optimization. Genetic Algorithm (GA) is an optimization method which can be applied for finding out the best cluster centers easily. But sometimes it takes more iteration for finding best cluster centers. In this paper, we use features of GA with the features of Discrete Differential Evolution (DDE) to solve text documents clustering problem. To test the efficiency of our algorithm we have taken sample database of Reuters-21578. From the experimental results, it is clear that our algorithm performs better than GA and DDE.

References

- Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques", 2nd Edition, Elsevier, 2008.

**Index Terms**

Computer Science

Artificial Intelligence
Keywords
Genetic Algorithm  Discrete Differential Evolution  Document Clustering