

{tag}

{/tag}

International Journal of Computer Applications

© 2012 by IJCA Journal

Volume 43 - Number 1

Year of Publication: 2012

Authors:

Yogesh Kumar Meena

Shashank

Vibhav Prakash Singh

10.5120/6067-8221

{bibtex}pxc3878221.bib{/bibtex}

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

- Wei Jian-Xiang, Liu Huai, Sun Yue-hong, Su Xin-Ning. "Application of Genetic Algorithm in Document Clustering", International Conference on Information Technology

and Computer Science, Vol. 01, pp. 145-148, 2009.

- George Karypis, Eui-Hong(Sam)Han, Vipin Kumar. "CHAMELEON: A Hierarchical Clustering Algorithm Using Dynamic Modeling", IEEE Computer Society, Vol. 32, Issue. 8, pp. 68-75, 1999.
- A. Casillas, M. T. Gonz´alez de Lena, and R. Mart. "Document Clustering into an unknown number of clusters using a Genetic Algorithm", Lecture Notes in Computer Science, Vol. 2807/2003, pp. 43-49, 2003.
- Calinski & Harabasz. "A Dendrite Method for Cluster Analysis", Communications in Statistics, Vol. 3(1), pp. 1-27, 1974.
- K. Premalatha, A. M. Natarajan. "Genetic Algorithm for Documents Clustering with Simultaneous and Ranked Mutation", Modern Applied Science, Vol. 3, No. 2, 2009.
- Sheng ZHONG, Zhiwei LIN, Beihai ZHANG, Chengcheng YU. "Genetic Algorithm on Documents Clustering", Journal of Computational Information Systems, Vol. 3, pp. 1063-1068, 2008.
- Quan-Ke Pan, M. Fatih Tasgetiren, Yun-Chia Liang. "A Discrete Differential Evolution Algorithm for the permutation flowshop scheduling problem", GECCO, pp. 126-133, 2007.
- Chong Su, Qingcai Chen, Xiaolong Wang, Xianjun Meng. "Text Clustering Approach Based on Maximal Frequent Term Sets", IEEE International Conference, pp. 1551-1556, 2009.
- Jeffrey L. Solka, "Text Data Mining: Theory and Methods", Statistics Surveys, Vol. 2, (2008), pp. 94-112 (electronic), 2007.
- Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques", 2nd Edition, Elsevier, 2008.
- David E. Goldberg, "Genetic Algorithms in Search, Optimization, and Machine Learning", 1st edition, Pearson, 2008.
- <http://kdd.ics.uci.edu/databases/reuters21578/reuters21578.html>.
- <ftp://ftp.cs.cornell.edu/pub/smart/> (accessed on April 13, 2006)
- C. Xiaohui, T. E. Potok, P. Palathingal Document Clustering using Particle Swarm Optimization, IEEE Swarm Intelligence Symposium, The Westin Pasadena, Pasadena, California, 2005.
- S. Das, A. Konar, U. K. Chakraborty, Two Improved Differential Evolution Schemes for Faster Global Search in ACM-SIGEVO Proceedings of Genetic and Evolutionary Computation Conference (GECCO-2005), Washington DC, June, 2005.
- K. Deb, A. Anand and D. Joshi (2002). A Computationally Efficient Evolutionary Algorithm for Real-Parameter Optimization, Evolutionary computation, 10(4), pp. 371 – 395.
- A. Ratnaweera, K, S. Halgamuge, : Self organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients. IEEE Trans. on Evolutionary Computation (2004) 8(3): 240-254.
- TREC. 1999. Text Retrieval Conference. <http://trec.nist.gov> ((accessed on April 13, 2006).

Index Terms

Computer Science

Artificial Intelligence

Keywords

Genetic Algorithm Discrete Differential Evolution Document Clustering