Abstract

Clustering technique has been broadly used in numerous disciplines, such as science, statistic, software engineering and other social sciences in order to identify natural groups in large amounts of data. K-means is one of the most generally used partitioning clustering algorithms that tries to locate a user specific number of clusters (k), which are represented by their centroids, by minimizing the square error function. There are two straightforward approaches to cluster center initialization i.e. either to choose the initial values arbitrarily or else to choose the first k samples of the data points. Both approaches cause the algorithm to converge to sub optimal solutions. In contrast Genetic algorithm is one of the most frequently used transformative calculations which perform worldwide research to discover the result to a clustering issue. The algorithm normally begins with an arrangement of haphazardly developed individuals called the populace and design consecutive, new eras of the populace by genetic operations for example population selection, fitness function, crossover and mutation. This paper compares K-means and genetic algorithm based data clustering. A new algorithm is proposed known as genetic algorithm K-means (GAKM). Comparison was done of the basis of
external, internal and time complexity.

References

2. BARAHATE SACHIN R and SHELAKHE VIJAY M. 2012. A survey and future vision of data mining in educational field, 978-0-7695-4640-7/12,IEEE.
12. MUKHERJEE D. 2011. Reducing out-of-school children in India, National University of Educational Planning and Administration, Delhi.
15. BOKOVA., and BUSH. 2012. Literacy is key to unlocking the cycle of poverty, at http://www.chron.com/ opinion/outlook/article /Literacy-is-key-to-unlocking-the -cycle-of-poverty-3848564. php (2012)

Index Terms

Computer Science Algorithms
Keywords

Clustering, K-means, Genetic Algorithm, Dropout, never enrollment.