Efficient Genetic K-means Clustering Algorithm and its Application to Data Mining on Different Domains

Ahmed Alsayat
Department of Computer Science, Bowie State University

November 8, 2016 @ 3:30 – 4:45 PM, CSB - Room 210

Ahmed Alsayat is pursuing his Ph.D. in the Department of Computer Science at Bowie State University. In 2009, Mr. Alsayat graduated from Florida Institute of Technology with a Master's of Science in Computer Science. From the same school, he received his second master in Engineering Management in 2009. Prior to coming to Bowie State University, Mr. Alsayat worked as a lecturer in the department of computer science, Aljouf University, Saudi Arabia, who sponsors his post graduate studies. He was also working as a manager of the network and operating systems department in the IT of Aljouf University.

Abstract:
Owing the massive increase in the amount of data streams currently available and still being continuously produced, the areas of data mining and machine learning have become increasingly popular, as companies, organizations and industries scramble to determine the most optimal methods and techniques for processing these large data sets. Data mining offers several techniques for processing this large data in order to determine some previously unknown pattern or trend from the data. This study aims at presenting an enhanced K-means clustering algorithm applied to the domains of social media and healthcare. We present a novel method called Optimize Cluster Distance (OCD) to maximize the distance between clusters and minimize it within a cluster by pair-wise clustering. Furthermore, the Genetic algorithm was applied with K-means to solve the possibility of getting bad positions for the centroids. Finding the optimal number of k in the K-means was considered where we applied the Self Organizing Map (SOM). In both domains, compared to the conventional K-means and other clustering techniques, the analysis shows that the proposed K-means is accurate and achieves better clustering performance along with valuable insights for each cluster. The approach is unsupervised, scalable and can be applied to various domains.

Contact Dr. Soo-Yeon Ji (sji@bowiestate.edu) if you have any question.