World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:14, No:07, 2020

Power Iteration Clustering Based on Deflation Technique on Large Scale Graphs

Authors : Taysir Soliman

Abstract : One of the current popular clustering techniques is Spectral Clustering (SC) because of its advantages over conventional approaches such as hierarchical clustering, k-means, etc. and other techniques as well. However, one of the disadvantages of SC is the time consuming process because it requires computing the eigenvectors. In the past to overcome this disadvantage, a number of attempts have been proposed such as the Power Iteration Clustering (PIC) technique, which is one of versions from SC; some of PIC advantages are: 1) its scalability and efficiency, 2) finding one pseudo-eigenvectors instead of computing eigenvectors, and 3) linear combination of the eigenvectors in linear time. However, its worst disadvantage is an inter-class collision problem because it used only one pseudo-eigenvectors which is not enough. Previous researchers developed Deflation-based Power Iteration Clustering (DPIC) to overcome problems of PIC technique on inter-class collision with the same efficiency of PIC. In this paper, we developed Parallel DPIC (PDPIC) to improve the time and memory complexity which is run on apache spark framework using sparse matrix. To test the performance of PDPIC, we compared it to SC, ESCG, ESCALG algorithms on four small graph benchmark datasets and nine large graph benchmark datasets, where PDPIC proved higher accuracy and better time consuming than other compared algorithms.

Keywords : spectral clustering, power iteration clustering, deflation-based power iteration clustering, Apache spark, large graph

Conference Title : ICCSSS 2020 : International Conference on Computer Science and Service System
Conference Location : Berlin, Germany
Conference Dates : July 23-24, 2020

Open Science Index, Computer and Information Engineering Vol:14, No:07, 2020 publications.waset.org/abstracts/120360.pdf