PARALLEL AND DISTRIBUTED COMPUTING (CSE4001)

Abstract

Submitted to-

Prof. Harini S

School of Computer Science and Technology

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

Team Members

Shourya Gupta 19BCE1704

Ridhika Sahni 19BCE1697

Sam Methuselah 19BCE1698

Mihir Antwal 19BCE1641

Project Title: **"Parallel Implementation of K-Means Clustering algorithm"**

ABSTRACT

Parallel computing has turned into the predominant worldview in PC design, essentially as multi- center processors. OpenMP is an Application Programming Interface which facilitates shared memory multiprocessing. Parallel code with OpenMP marks, through a unique order, segments to be executed in parallel.

K-Means algorithm is a basic clustering algorithm which belongs to the cluster with the nearest mean cluster. The problem is computationally difficult (NP-hard). The algorithm has a loose relationship to the k-nearest neighbor classifier, a popular machine learning technique for classification that is often confused with k-means due to the name. Applying the 1-nearest neighbor classifier to the cluster centers obtained by k-means classifies new data into the existing clusters. This is known as nearest centroid classifier or Rocchio algorithm.

OBJECTIVE

K-Means algorithm always asks for the number of clusters that are required to be formed. This requirement is given by the user only. Moreover, in a sequential manner the runtime is high for large data set because it is performing the same function only once in a given time frame. The objective of the project is to decrease the run time of the algorithm and also automatically select how many clusters are required.

PROPOSED METHODOLOGY

We plan to create sub datasets of the dataset and then, run them in parallel with each other and perform k- means there. In that we will later try to cover the centroids and repeat the process until the inertia calculated is found.