Usage of the canopy clustering algorithm

Technology, Artificial Intelligence

In general, the PSO algorithm runs on holding a collection(usually known as swarms) of candidate solutions(called particles). The particles roam around the search space with respect to some designated formulas. Finding their best known positions, they render the movements of these particles in a search space and also the whole swarm’s best position. On finding better known positions, they guide the movements of the swarm. This process continues again and again until a perfect and satisfied solution is obtained finally. Along with the GA, PSO is superior than GA as they are very easy to implement and just has very little adjustments to be made. Major applications of PSO are: optimization of functions, fuzzy system control, and network training using artificial neural and other applications where GA can be implemented. The main Pros of Particle swarm optimization are: it depends on swarm-intelligence, It provides approximate results and is also a nondeterministic optimization method. The PSO handles several potential solutions at a single time. When the algorithm is iterated every time, each solution is calculated by an function to arrive at its best fitness. Every solution is shown as a particle in its fitness landscape (i. e.) its search space. The particles either “ fly” or “ swarm” through a search space and determines the highest measure returned by the corresponding objective function. Every particle beholds the following: it’s position in the search space, it’s velocity, the best Individual position and mainly maintains its global best. There are only Three primary stages in the PSO algorithm: to determine the fitness of all the particles, to update the individual an global bests and finally update the velocity and the particle’s position. The velocity of an individual particle is updated using the following equation:

x a(y+1)= zx a(y)+ s 1 r 1[^p a(y)-p a(y)+ s 2 r 2[q (y) – p a(y)] where

x a(y) is the particle’s velocity at any time t

p a(y) is the particle’s position at any time t

p^a(y) is the best individual solution at any time t

q(y) is swarm’s best solution at any time t

i is the index value

z is the inertial coefficient

s 1, s 2 are acceleration coefficients, 0 ≤ s 1, s 2 ≤ 2

r 1, r 2 are random values (0 ≤ r 1, r 2 ≤ 1)

Canopy Clustering algorithm:

The Canopy clustering algorithm is a pre-clustering algorithm that remains unsupervised and is conducted before a K-means clustering also known as the hierarchical clustering. The primary objective of this algorithm is to boost the rate of the clustering process when large data sets are taken into account, where a direct use of the main algorithm may become impractical because of the huge data set size.

Algorithm:

Initialize with a list of data points and take into consideration two distances D1> D2 for easier processing:

1. Choose a random point from the given data point list in order to develop a canopy cluster.
2. Distance from the selected point must be approximated to every other point in the list.
3. Place all the points that lie within the given distance threshold of D1 into a canopy.
4. Pop out all the points that lie within threshold D2 from the initial data list. The main purpose is that these points should not be the centre and develop new canopies.
5. Execute steps 1 to 4 until all the data points in the initial list is processed.