



Parallel Processing

Technology Summary

This algorithm uses a divide-and-conquer strategy in a novel parallel algorithm to reduce the computational complexity of D2 clustering. The goal is to parallelize the centroid update in D2 clustering by: dividing the data into segments based on their adjacency, computing some local centroids for each segment in parallel, and combining the local centroids to a global centroid. This parallel algorithm achieves significant speed up with minor accuracy loss. The computational intensiveness of D2 clustering limits its usage to only relatively small scale problems.

Application & Market Utility

Scalability has emerged as a problem with D2 clustering because unknown variables grow with the number of objects in a cluster. As a result, it takes several minutes to learn each category by performing D2 clustering on 80 images, and takes days to complete the modeling of thousands of images. With emerging demands to extend the algorithm to large-scale datasets (online image datasets, video resources, biological databases) this invention exploits parallel processing in a cluster computing environment in order to overcome the inadequate scalability of D2 clustering.

Next Steps

Seeking research collaboration and licensing opportunities.

TECHNOLOGY READINESS LEVEL

4-7

Seeking

Investment | Licensing |

Keywords

- D2 clustering
- parallel algorithm
- image
- big data

Researchers

James Wang

Professor of Information Sciences and Technology

[Online Bio](#)

[Website](#)

Jia Li

Professor of Statistics

[Website](#)

Yu Zhang

Associate Professor of Statistics

[Website](#)

Originating College

College of Information Sciences and Technology

Office of Technology Management Contact

Swope, Brad

bas101@psu.edu

814-865-6277