Abstract

A lot of recent research works have pointed out that metric learning is far better as compared to using default metrics such as Euclidean distance, cosine similarity, etc. Moreover, similarity learning based on cosine similarity has been proved to work better for many of the data sets, which are not necessarily textual in nature. Nevertheless, similarity learning in nearest neighbor algorithms has been inherently slow, owing to their $O(d^3)$ complexity. This short-coming is addressed in this research and a similarity learning algorithm for many core architectures is proposed; whereby, Similarity Learning Algorithm (SiLA) is parallelized. The resulting algorithm is faster than the traditional one on many data sets because of its parallel nature. The results are confirmed on UCI data sets.

References


Index Terms

Computer Science  Information Systems
Keywords

Similarity learning, SiLA algorithm, Parallel computing, Supervised machine learning, CUDA programming