Proposal: Parallel t-Distributed Stochastic Neighbor Embedding

Yuhao Zhou & Fengxu Tu

EC 526

Spring 2020

t-Distributed Stochastic Neighbor Embedding (tSNE)

Introduction

t-Distributed Stochastic Neighbor Embedding (tSNE) [1] is a dimensionality reduction algorithm which is used to visualize high-dimensional data.

Method

- Calculate the pairwise similarity high-dimensional data.
- Construct a low-dimensional pairwise similarity map.
- Minimize Kullback-Leibler divergence

t-Distributed Stochastic Neighbor Embedding (tSNE)

Measure of Similarity

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\sigma_i^2)}$$

where σ_i is the variance of the Gaussian that is centered on data point x_i.

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N}$$

t-Distributed Stochastic Neighbor Embedding (tSNE)

Low-dimensional Similarity

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum^{k \neq i}(1 + \|y_i - y_k\|^2)^{-1}}$$

t-Distributed Stochastic Neighbor Embedding (tSNE)

Cost function

Kullback-Leibler divergence

$$C = \sum_i KL(P_i||Q_i) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}}$$

Gradient of cost function:

$$\frac{\partial C}{\partial y_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j)(1 + \|y_i - y_j\|^2)^{-1}$$

So we use gradient descent with momentum to optimize the cost function.

t-Distributed Stochastic Neighbor Embedding (tSNE)

Visualization of MNIST dataset

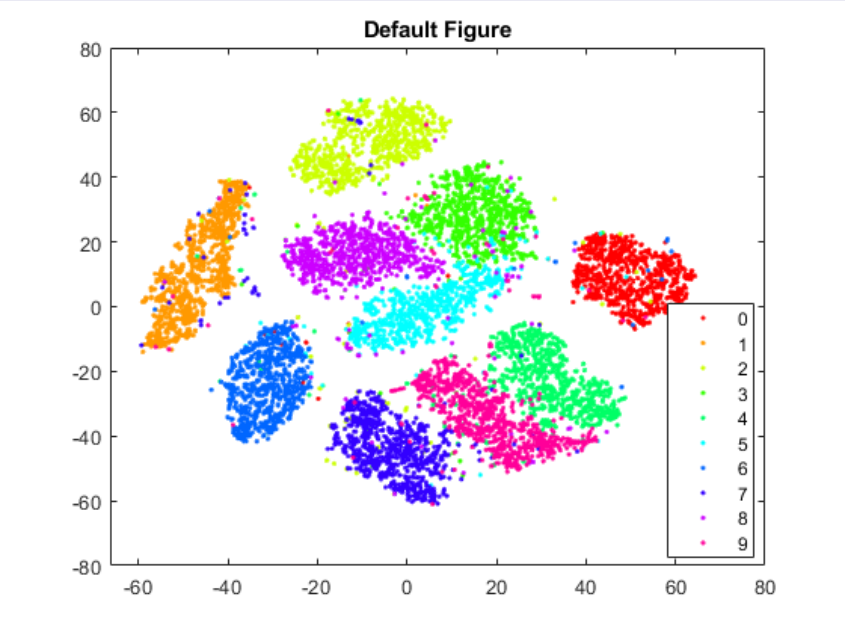

Figure: MATLAB built-in tSNE

t-Distributed Stochastic Neighbor Embedding (tSNE)

Where to use MPI/openACC

- Pairwise Similarity Calculation
- Gradient Descent
- PCA is used in the implementation. We can implement a parallel version of SVD of eigen-decomp as well

Reference

 Laurens van der Maaten and Geoffrey Hinton. "Visualizing data using t-SNE". In: *Journal of machine learning research* 9.Nov (2008), pp. 2579–2605.