A-tSNE

Approximated and user steerable for progressive visual analytics

Nicola Pezzotti et al.
Presented by: Lovedeep Gondara

March 9, 2017

Outline

Outline

High dimensional data

- Most real world datasets are high dimensional.
- High dimensional data vis is hard.
- Dimensionality reduction to the rescue.

Outline

tSNE

Introduction

- A tool for dimensionality reduction/vis of high dimensional data.
- Converts similarities between data points in high dimensional space to joint probability distribution P.
- Computes a joint probability distribution Q, describing similarity in low dimensional space.
- Goal: Represent P faithfully using Q.

tSNE

Introduction

- Minimize Kullback-Leibler divergence between P and Q.
- Use gradient descent for minimization.
- Each point attracts or repels all other points with a force F.

Outline

Barnes-Hut SNE

- Original tSNE uses brute force approach for F.
- Computation and memory complexity of $O(n^2)$.
- Barnes-Hut SNE is an evolution of tSNE.

Barnes-Hut SNE

- Uses two approximations.
- Approximation 1: Similarities between data points are computed by only taking set of nearest neighbours N.
- Approximation 2: Uses Barnes-Hut algorithm.
- Reduces computational and memory complexity to $O(N \log(N))$ and $O(N)$ respectively.

Outline

A-tSNE

Introduction

- Evolution of BH-SNE.
- Uses approximations to generate useful intermediate results.
- Approximation defined by user.

A-tSNE

Introduction

Figure: Progressive visual analytics using tSNE

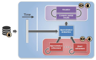

Figure: Progressive visual analytics using A-tSNE

A-tSNE

Introduction

- Improves BH-SNE using approximated KNN computations for approximated P.
- Using a precision parameter ρ, describe the average percentage of points in approximated neighbourhood that belong to the exact neighbourhood.
- ρ is user defined, large values of ρ means better approximations but more computational overhead.
- These approximations make A-tSNE computationally steerable.

A-tSNE

Introduction

Figure: BH-SNE: 3191.8 s

Figure: A-tSNE ($\rho = 0.23$): 20.4 s

Figure: A-tSNE ($\rho = 0.34$): 30.1 s

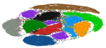

Figure: A-tSNE ($\rho = 0.07$): 13.0 s

Outline

A-tSNE
User driven refinement

- User selection: Select a subset of points for immediate refinement.
- Breadth first search: If only a few points are selected, include the neighbourhoods.
- Density based refinement: Global overview, user defined selection or whole dataset.

A-tSNE
Visualization and interaction

- Density based: Simple points increase clutter, use KDE.
- Visualizing approximations: Precision of high dimensional similarities is gradually refined until exact, requested precision can be visualized while refinement is ongoing.
- Use magic lens to show approximations

A-tSNE
Data manipulation

- Inserting points
- Deleting points
- Dimensionality modification

A-tSNE
Interface

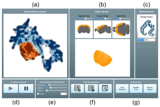

Figure: Interface

Outline

1. Inroduction
   - High dimensional data vis
   - tSNE
   - Barnes-Hut SNE
2. A-tSNE
   - Introduction
   - Interactive analysis
   - Case studies
3. Critique

A-tSNE
Mouse brain gene expression

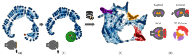

Figure: Analysis of the gene expression in the mouse brain using A-tSNE

A-tSNE
Real-time analysis of high-dimensional streams

Figure: Initial embedding

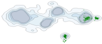

Figure: Evolution of (a)

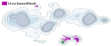

Figure: New cluster indicates the creation of a set of different readings

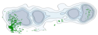

Figure: The cluster that identifies miscalibrated readings is removed

Critique

- Enhanced performance.
- User selective refinement.
- Too many moving parts.
- Not sure if all are helpful.

For Further Reading I

Pezotti et al. 2016
Approximated and user steerable tsne for progressive visual analytics.
IEEE Transactions on Visualization and Computer Graphics, 2016.