Review of *Visualizing Data using t-SNE*

Grace Yi Chen

In this paper, the authors introduced "t-SNE", which is a technique that visualize high-dimensional data that could be compressed to a low-dimensional, non-linear manifold. Over the last few decades, a variety of visualization techniques for high-dimensional data have been proposed. The goal of dimension reduction is to preserve the important structure of the high-dimensional data in the low-dimensional map. t-SNE starts by converting the high and low-dimensional Euclidean distances between datapoints into conditional probabilities representing similarities respectively. The goal here is to find a low-dimensional data representation that minimizes the mismatch between high and low-dimensional conditional probabilities. The authors used a symmetrized single Kullback-Leibler divergence measure as the cost function which focuses on keeping the local structure of the data in the map. It also uses a student-t distribution, a heavy-tailed distribution to compute the similarity between two points in the low-dimensional space, which could avoid both the crowding problem and the optimization problems. Comparing with other non-parametric visualization tools, t-SNE is significantly better under a variety of scenarios.

I think this paper is well-written and the authors clearly explained the rationale and details about t-SNE. I am more familiar with some traditional dimensionality reduction techniques such as Principal Components Analysis. These techniques could compress high-dimensional data that lies on low-dimensional linear manifold. For high-dimensional data that map to a non-linear manifold, it seems there were not many good dimension reduction tools before t-SNE. With the development of new technologies like single-cell sequencing, there are more and more high dimensional data arising and the development of new techniques like t-SNE could support the analysis. However, as the authors mentioned, there are some weaknesses in t-SNE. One major weakness of t-SNE is that the cost function is not convex while there are several optimization parameters. So the optimized parameter might be different each time t-SNE is run from an random initial points. Although the authors mentioned that the quality of the optima does not vary much from run to run, it is not good for reproducibility and may cause confusions for users and readers.

Question:
The authors mentioned about methods using autoencoder that could reduce the dimension of data with high intrinsic dimensionality. Will this method with autoencoder work well with low intrinsic dimensionality?

UMAP is another popular high dimensional data visualization tool and I am wondering how it is compared with t-SNE?