A Two-Stage Dimension Reduction Method and its Applications on Highly Contaminated Image Sets

I-Ping Tu
Institute of Statistical Science, Academia Sinica, Taiwan

Abstract

Principal component analysis (PCA) is arguably the most popular dimension reduction method for vector type data. When applied on image data, PCA demands the images to be portrayed as vectors. The resulting computation is heavy because it would solve an eigenvalue problem of a covariance matrix whose size equals the square of the pixel number. To mitigate the computation burden, multi-linear PCA that uses column and row basis with a Kronecker product to compose the matrix structure was proposed, for which the success was demonstrated on face image sets. However, when we apply MPCA on the particle images of the single particle cryo-electron microscopy (cryo-EM) experiments, the results are not satisfying. Here, we propose a dimension reduction method called Two Stage Dimensional Reduction (2SDR) where we first apply MPCA to extract its projection scores, and then apply PCA on these scores to further reduce the dimension. Tests using single particle cryo-EM benchmark experimental data sets demonstrate that 2SDR reduce huge computation costs compared to PCA, and show 2SDR can reconstruct better quality images than MPCA. Further application of 2SDR on a cryo-EM micrograph data set significantly reduces the noise to clearly reveal the individual particles. Remarkably, the de-noised particles boxed out from the micrograph allow subsequent structural analysis to reach a high-quality 3D density map. This is a joint work with Szu-Chi Chung, Po-Yao Niu, Su-Yun Huang and Wei-Hau Chang.