

International Conference on
COMPUTATIONAL BIOLOGY AND BIOINFORMATICS
September 05-06, 2018 Tokyo, Japan

Principal component analysis/tensor decomposition based unsupervised feature extraction applied to bioinformatics

Y-H Taguchi
Chuo University, Japan

In the usual data mining problems performed in bioinformatics, there are a greater number of features than that of samples. Typically, it is 10-20 features (genes) vs. a few 10's samples. It is a typical large p small n problem that was hard to resolve. One of these problems is identification of genes associated with distinct expression between several conditions. Although the combination of statistical analysis and fold change is often used for this purpose, how significant the results should be and how large fold change should be are not often hard to decide. To address this problem more systematically, there is a proposed Principal Component Analysis (PCA)/Tensor Decomposition (TD) based unsupervised Feature Extraction (FE) and applied to various bioinformatics problems ranging from biomarker identification, identification of disease causing genes and even *in silico* drug discoveries from gene expression profiles. PCA/TD based unsupervised FE can also be used for integrated analysis of multi-omics data set including gene expression, proteomics, metabolomics, various epigenetic profiles including promoter methylation and histone modifications and single nucleotide polymorphism. Since this methodology is unsupervised and linear method, it can also be applied to wide range of unlabeled data set without losing interpretability that is often missing in non-linear methods like deep learning and kernel tricks. The presentation includes the introduction of mathematical basis of this methodology as well as the recent applications to bioinformatics.

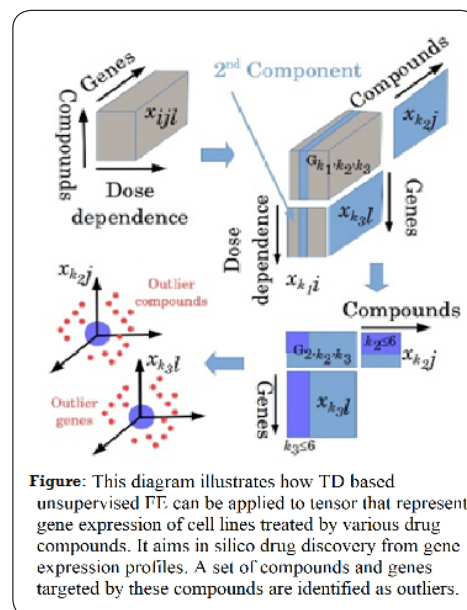


Figure: This diagram illustrates how TD based unsupervised FE can be applied to tensor that represent gene expression of cell lines treated by various drug compounds. It aims *in silico* drug discovery from gene expression profiles. A set of compounds and genes targeted by these compounds are identified as outliers.

Biography

Y-H Taguchi is currently a Physics Professor at Chuo University, Tokyo, Japan. He has obtained his PhD at Department of Physics, Tokyo Institute of Technology (1988). He has spent nine years as an Assistant Professor and nine years as Associate Professor and Professor at Department of Physics, Chuo University. He has his experience in theoretical physics and bioinformatics. His interest mainly focuses feature extraction when more variables are available than samples, since it is very usual situation in the bioinformatic analysis.

tag@granular.com

Notes: