

On Visualization, Screening, and Classification of Cell Cycle-Regulated Genes in Yeast

Henry Horng-Shing Lu¹
hslu@stat.nctu.edu.tw

Han-Ming Wu²
hmwu@stat.sinica.edu.tw

- ¹ Institute of Statistics, National Chiao-Tung University, Hsinchu 30050, Taiwan, R.O.C.
² Institute of Statistical Science, Academia Sinica, Taipei, Taiwan 11529, R.O.C.

Keywords: visualization, microarray, cell cycle-regulated genes, multidimensional scaling, dimension reduction, wavelet transform, support vector machines

1 Introduction

We propose a new and integrated approach for visualization, screening, and prediction of gene functions using microarray data, based on multidimensional scaling (MDS), nonlinear dimension reduction and wavelet transform (DWT). This approach is applied to analyze the cell cycle of yeast in Spellman *et al.* [2].

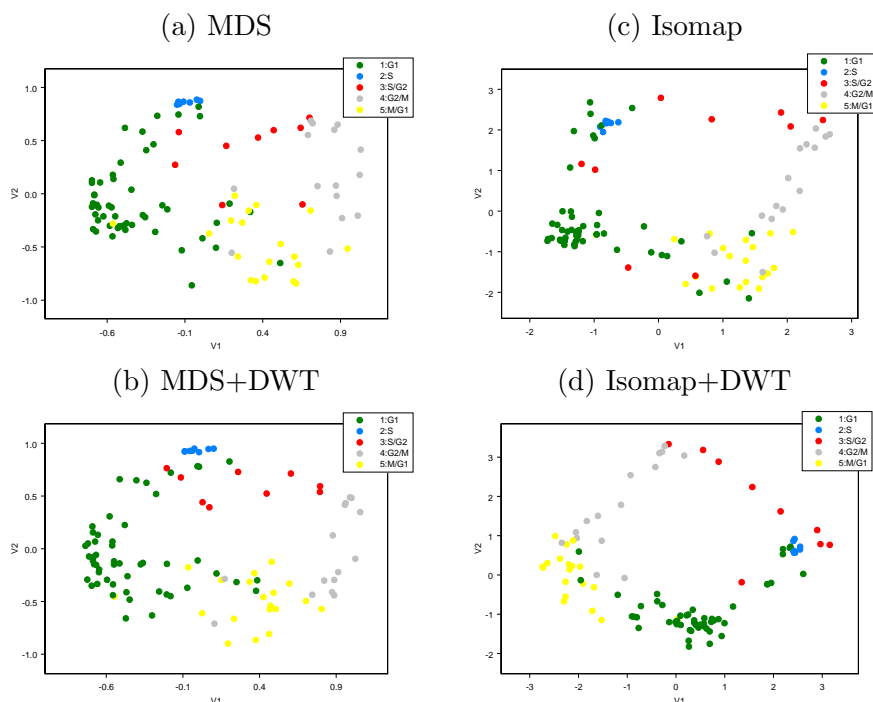


Figure 1: The 2D representation of cell 103 cycle-regulated genes are displayed.

2 Method and Results

Tenenbaum *et al.* [3] have proposed a global geometric framework for nonlinear dimensionality reduction, called Isomap. The MDS projections in Figure 1 (a) of 103 known genes reflect the majority order of cell cycle in clockwise rotation. The projection plot of Isomap with DWT in Figure 1 (d) are more distinguishable. We formulate a simple test for screening. One can investigate those genes that are not involved in the sets of Spellman *et al.* [2] and Cho *et al.* [1] for further studies. The prediction errors by Jackknife in 2D MDS space are showed in Table 1. The KDE method performs better against others except Isomap. This may due to the wavelet features make the clusters in 2D MDS space more compact and so that KDE can perform significantly well than others.

Table 1: Error rates of prediction of 103 known genes by jackknife in 2D MDS space are reported, where KNN (k nearest neighbors) uses $K = 3$, KDE (kernel density estimation) uses the Gaussian kernel with equal prior and SVM Torch uses polynomial kernel.

Method		G1	S	S/G2	G2/M	M/G1	Total
MDS	KNN	0.1154	0.0000	0.6667	0.3333	0.3158	0.2233
	KDE	0.1346	0.0000	0.6667	0.2000	0.2632	0.2039
	CART	0.1538	0.0000	0.6667	0.4667	0.3158	0.2621
	SVM Torch	0.1346	0.1250	1.0000	0.2000	0.4211	0.2718
MDS with DWT	KNN	0.1154	0.0000	0.4444	0.1333	0.4211	0.1942
	KDE	0.0963	0.0000	0.2222	0.2000	0.2632	0.1456
	CART	0.1731	0.2500	0.4444	0.2667	0.4211	0.2621
	SVM Torch	0.1154	0.0000	1.0000	0.2000	0.3684	0.2427
Isomap	KNN	0.2115	0.1250	0.7778	0.2000	0.4211	0.2913
	KDE	0.1731	0.3750	0.7778	0.2000	0.2105	0.2524
	CART	0.1538	0.2500	0.5556	0.2667	0.2106	0.2233
	SVM Torch	0.1154	1.0000	0.8889	0.2000	0.2632	0.2913
Isomap with DWT	KNN	0.0962	0.1250	0.3333	0.4000	0.3158	0.2039
	KDE	0.1154	0.0000	0.2222	0.2667	0.1053	0.1359
	CART	0.1154	0.2500	0.3333	0.3333	0.2632	0.2039
	SVM Torch	0.0577	1.0000	0.2222	0.2667	0.1053	0.1845

3 Discussion

The Isomap plot in 2D space with wavelet transformation can reveal more distinguishable patterns. This research makes a step toward developing an effective approach for discovering new functionality of known and unknown genes for the mechanism of cell cycle by microarray data.

References

- [1] Cho, R.J., Campbell, M.J., Winzeler, E.A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T.G., Gabrielian, A.E., Landsman, D., Lockhart, D.J., and Davis, R.W., A genome-wide transcriptional analysis of the mitotic cell cycle, *Mol. Cell*, 2(1):65–73, 1998.
- [2] Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D., and Futcher, B., Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization, *Mol. Biol. Cell*, 9(12):3273–3297, 1998.
- [3] Tenenbaum, J.B., de Silva, V., and Langford, J.C., A global geometric framework for nonlinear dimensionality reduction, *Science*, 290(5500):2319–2323, 2000.