Poster J-38

Nonlinear dimesnionalilty reduction in apoptosis signalling networks

Authors:
Sergii Ivakhno (*University of Edinburgh; School of Informatics*)
Douglas Armstrong (*University of Edinburgh; School of Informatics*)

Short Abstract: System wide modelling and analysis of signalling networks is essential for understanding complex cellular behaviour, such as biphasic responses to different combinations of growth factors. We applied unsupervised learning approach via non-linear dimensionality reduction using ISOMAP for analysis and clustering of apoptosis signalling network activities.

Long Abstract:

Introduction, methods and datasets

System wide modelling and analysis of signalling networks is essential for understanding complex cellular behaviour, such as biphasic responses to different combinations of growth factors. Cellular signalling networks can be considered as multivariate non-linear functions that map various perturbations and extracellular cues into different cellular outputs. In order to build predictive models of cellular responses it is first necessary to characterise and catalogue activities of signalling networks based on the measurement of molecular signals (network nodes) along. However, most machine learning approaches reported in the literature for analysis of signalling networks explicitly employ supervised learning for predictive modelling, while value of unsupervised learning have not been sufficiently investigated. One application of predictive modelling is a systems wide analysis of combinatorial mechanisms, which lead to generation of multiple and sometimes antagonistic cellular responses. For instance, the proapoptotic and proinflammatory cytokine tumour necrosis factor (TNF) can not only activate caspases to promote cell death but also induce prosurvival mechanisms and cellular proliferation. Our approach uses Cytokine compendium dataset of Janes et al [1] and nonlinear dimensionality reduction algorithms ISOMAP [1] to find low dimensional projection of apoptosis signalling network represented by activities of 19 molecular signals where groups of different cellular responses such as apoptosis versus survival can be easily identified and visualized. In the Cytokine compendium dataset human colonic adenocarcinoma cells were stimulated with 10 combinations of saturating or subsaturating concentrations of TNF with epidermal growth factor (EGF) or insulin to investigate crosstalk between proapoptotic and prosurvival signaling pathways. This was followed by sequential measurements over 24 hours of 19 different intracellular molecular signals known to be associated with TNF, EGF, or insulin signalling. Each cytokine elicited multiple classes of signals - for instance, high concentrations of TNF elicited signals associated with both TNF and EGF signalling. To test whether apoptosis could be connected to the measured signals in the network, cell-death phenotype was assayed for each combinatorial cytokine stimulus using four distinct apoptosis assays.

We applied unsupervised learning approach via non-linear dimensionality reduction using ISOMAP for analysis and clustering of apoptosis signalling network activities. Several considerations call for application of nonlinear dimensionality reduction techniques for

studying signalling networks rather then classical PCA or MDS. First, activities of molecular signals in signalling networks often lie on the non-linear manifold, which reflect nonlinear interactions between signals in the network. For example, in the apoptosis signalling network TNF receptor forms trimer complexes before trandusing signalling into the cytoplasm, therefore introducing non-linear effects into the network. Second, ISOMAP ere already successfully applied for dimensionality reduction in gene networks [3].

Results

By comparing low dimensional maps of signalling network found by ISOMAP and PCA we show superiority of non-linear dimensionality reduction approaches for network analysis. PCA was unable to find low dimensional embedding: not only first three leading eignevalues accounted for only 70% variance, but the reconstructed map itself showed no distinct patters or clusters representing different activity states of apoptosis signalling network. On the contrary, ISOMAP identified clearly differentiable clusters of network activity corresponding to different cytokine treatments lying on a three dimensional submanifold. Clusters of apoptosis signalling network activities corresponding to different treatments with TNF, EFG and Insulin were identified, showing that ISOMAP can successfully find dimensions of nonlinear submanifolds corresponding to combination of original molecular signals and also pseudoaxes of autocrine feedback loops. In particular, ISOMAP found distinct groups related to insulin 500 ng, TNF 100 ng, mixed insulin-TNF and EFG-TNF high doses treatments. It is noteworthy that ISOMAP was able to reconstruct not only low dimensional embedding of insulin-induced apoptosis network, but also temporal profiles of evolution of network activity, with early time points occupying far end of the map and late time points approaching the centre.

To compare algorithms in an unbiased way we have used both exploratory analysis and supervised classification approach. To perform supervised classification it was necessary to chose class labels for each time course treatment, for instance activity of signalling network 5 minutes after stimulation with of TNF. Unfortunately we could not have used results of apoptosis assays explicitly as they were used only twice during the 24 hours time course. Two approximate class labels we made the assumption that entire time course corresponding to distinct cytokine treatments represents only one class. This assumption is justified if we consider cellular response to be deterministic function of individual treatment, and it allowed us to consider 12 apoptosis measurements as signatures for class labels. EM clustering was used to assign class labels to each treatment based on apoptosis assays data, resulting in 3 different classes which represent strong, medium or absence of apoptosis response.By applying supervised learning approach with automatic assignment of class labels (apoptosis responses) by EM clustering, we showed that low dimensional representation of signalling network activities achieve classification accuracy comparable to original 19-dimesnional dataset (69% vs 71% and 72% vs 78% for k-nearest neighbours and quadratic discriminant analysis respectively). We conclude that unsupervised learning through non-linear dimensionality reduction can facilitate analysis of signalling networks activities and categorize their responses to external stimuli.

Litertaure

[1] Janes, K.A., Albeck, J.G., Gaudet, S., Sorger, P.K., Lauffenburger, D.A. and Yaffe, M.B. (2005) Systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis, Science, 310, 1646-1653.

[2] Tenenbaum, J.B., de Silva, V. and Langford, J.C. (2000) A global geometric framework for nonlinear dimensionality reduction, Science, 290, 2319-2323.
[3]Dawson, K., Rodriguez, R.L. and Malyj, W. (2005) Sample phenotype clusters in high-density oligonucleotide microarray data sets are revealed using Isomap, a nonlinear algorithm, Bmc Bioinformatics, 6.