

Das Institut für Informatik lädt im Rahmen der Kolloquiumsveranstaltungen ein zu folgendem Vortrag:

Reverse Engineering in Systems Biology

Dr. Reinhard Guthke,

Hans-Knöll-Institut HKI, Jena, www.hki-jena.de

Donnerstag, 3. November 2005, 16.00 Uhr c.t.

Magnus-Hörsaal, Informatik-Gebäude, Informatik-Gebäude,
Robert-Mayer-Str. 11-15

Abstrakt

Die Erforschung und das Verständnis komplexer molekularer Interaktionen, die für lebende Organismen wesentlich sind, sind zentrale Aufgabe der modernen Molekularbiologie, Systembiologie und Bioinformatik. Es wird im Vortrag eine neue, optimierte Methode zur Rekonstruktion von dynamischen genregulatorischen Netzwerken, dem sog. Reverse Engineering, vorgestellt. Diese bioinformatische Methode basiert auf der Analyse von Mikroarray-Daten und nutzt dabei verfügbares biologisches Fachwissen aus. Die wesentlichen Kinetiken der Prozessantwort nach einem gesetzten Störimpuls werden mittels der Fuzzy-Clusteranalyse von Genexpressionsprofilen (Zeitreihen) identifiziert. Die Klassenanzahl wird anhand verschiedener Gütekriterien bestimmt. Für jedes Cluster wird ein repräsentatives Gen mit hoher Klassenzugehörigkeit ausgewählt. In diese Auswahl fließt das verfügbare physiologische Wissen (aus der Fachliteratur) mit ein. Anschließend werden hypothetische Netzwerkstrukturen identifiziert, indem Systeme gewöhnlicher Differentialgleichungen simuliert und an die Expressionsprofile (Zeitreihen) der für die Cluster-repräsentativen Gene angepasst werden. Resamplingverfahren werden eingesetzt zur Validierung der Netzwerkhypothesen und zur Verbesserung der Generalisierungseigenschaft der erhaltenen Modelle.

Der neue Reverse-Engineering-Algorithmus wurde auf verschiedene Mikroarray-Datensätze angewandt: erstens zur Charakterisierung der Stressantwort des Mikroorganismus *Escherichia coli* nach Induktion der Produktion rekombinanter Proteine, zweitens zur Analyse der Immunantwort von humanen Blutzellen auf eine Infektion mit pathogenen *E. coli* und drittens auf mit LiCl stimulierte Hepatozyten zum Studium der räumlichen Zonierung der Leber.

Zum vergleichenden Nachweis der Überlegenheit der neuen Methode zur Konstruktion hypothetischer Netzwerkstrukturen wurden auch Methoden eingesetzt, die auf der Singular Value Decomposition (SVD) basieren. Es konnte gezeigt werden, dass die neue heuristische Methode überlegen ist: Es wurden bessere Anpassungsgüten an die gemessenen Daten mit einfacheren Modellen, d.h. mit weniger Modellparametern, erzielt.

Abstract

Discovering and understanding the complex molecular interactions that make up living organisms is one of the most interesting and challenging problems of modern molecular biology, systems biology and bioinformatics. We present an optimized reverse engineering strategy aimed at a reconstruction of dynamic gene interaction networks. The proposed approach is based on both, microarray data and available biological knowledge.

The main kinetics of the response after a perturbation are identified by fuzzy clustering of gene expression profiles (time series). The number of clusters are optimized using various evaluation criteria. For each cluster a representative gene with a high fuzzy-membership is chosen in accordance with available physiological knowledge. Then hypothetical network structures is identified by seeking systems of ordinary differential equations, whose simulated kinetics could fit the gene expression profiles of the cluster-representative genes. Resampling methods are applied to validate the obtained dynamic network models and to improve their generalization strength.

This novel reverse engineering algorithm was applied to different sets of DNA microarray data, such as characterizing the stress response of *Escherichia coli* after induction of recombinant protein synthesis or the immune response after infection of human blood cells by pathogenic *E. coli*. For the construction of hypothetical network structures Singular Value Decomposition (SVD) based methods and the novel heuristic Network Generation Method were compared.

It turned out that the proposed novel method could find sparser networks and gave better fits to the experimental data.