Abstract

The function of all living cells is controlled by the process of *gene expression*; the translation of the genetic codes (genes) in the DNA into proteins that regulate the cell, as well as the expression of genes themselves [Str92, SR96]. This indirect interaction of genes composes a gene regulatory network regulating the cell [DLS00]. Due to the development of microarrays it is now possible to (indirectly) measure the expression of genes and thus to reverse engineer these gene regulatory networks. These can give insight into many genetically related diseases and disorders [WMSB98, ZHZ+04].

In this thesis we extend on a method from [EPdJS04]. For this we use the Spellman [SSZ+98] time-series gene expression data. Firstly we remove noise by applying a scale-space smoothing, after which the expression data is discretized based upon the local extrema in the expression pattern. A similarity function is applied to the discretized expression values of the genes, grouping them in a pairwise fashion. We remove chance findings by combining the findings of two datasets. We validate the results by a gold standard protein-protein interaction database, showing that local extrema are a significant feature for identifying gene relations.

We also experiment with learning dynamic Bayesian networks from the discretized expression data, for finding causal gene interactions.