

## Abstract

In the biological and medical sciences more and more genetic data are gathered in order to unravel the complex genetic machinery of living cells. The extraction of the biological meaningful information from among these large data sets is tedious work that cannot be done without the help of computers. This thesis illustrates by two examples from gene expression data analysis how formal concept analysis can contribute to this effort. The first example is the identification of combinatorial biomarkers in breast cancer. In this application, formal concept analysis complements the classical approach of differentially expressed gene detection by focussing on groups of genes and their concerted activity. The second example concerns biochemical networks. Here we show how members of a gene regulatory network can be inferred from gene expression data using a concept lattice. Furthermore the potential application of formal concept analysis in the description and deduction of central network properties is considered. The discussion of our results with respect to the current biological knowledge affirms formal concept analysis as a useful tool in genetic data analysis.

## Kurzzusammenfassung

In den biologischen und medizinischen Wissenschaften werden mehr und mehr genetische Daten angehäuft, um die komplexe genetische Maschinerie lebender Zellen aufzudecken. Das Auffinden biologisch bedeutsamer Informationen in diesen großen Datensätzen ist aufwändig und bedarf des Einsatzes von Computern. Diese Arbeit zeigt anhand zweier Beispiele aus der Analyse von Genexpressionsdaten wie die Formale Begriffsanalyse hierzu beitragen kann. Das erste Beispiel ist die Bestimmung kombinatorischer Biomarker für Brustkrebs. In dieser Anwendung ergänzt die Formale Begriffsanalyse durch ihre Konzentration auf Gengruppen und deren gemeinschaftliche Aktivität die klassische Herangehensweise der Detektion differenziell exprimierter Gene. Das zweite Beispiel betrifft biochemische Netzwerke. Hier zeigen wir, wie Mitglieder eines genregulatorischen Netzwerks mit Hilfe eines Begriffsverbandes aus Genexpressionsdaten gewonnen werden können. Des Weiteren wird das Potenzial der Formalen Begriffsanalyse für die Beschreibung und Ableitung zentraler Netzwerkeigenschaften betrachtet. Die Diskussion unserer Ergebnisse vor dem Hintergrund des aktuellen biologischen Kenntnisstands bestätigt die Formale Begriffsanalyse als nützliches Werkzeug für die Analyse genetischer Daten.