

Vortrag am: 10.11.2005

Vortragende: Gernot D. Kleiter, intern

Thema: Zur Komplexität multivariater Modelle

In den Verhaltens- und Sozialwissenschaften bestehen Modelle aus einer Menge von Variablen und den Beziehungen zwischen den Variablen. Beispiele sind Regressionanalysen, Pfadanalysen, log-lineare Modelle, Structural Equations, oder Bayes'sche Netzwerke.

Eine elementare Beziehungen zwischen zwei (Mengen von) Variablen ist, ob sie (probabilistisch) unabhängig sind oder nicht – genauer, ob sie bedingt unabhängig sind oder nicht. Entsprechende Modelle werden conditional independence models genannt. Dies ist eine sehr allgemeine und umfassende Klasse von Modellen. Die populärste Teilklassen sind Bayes'sche Netzwerke. Sie können anschaulich mit Pfeildiagrammen dargestellt werden. Unglücklicherweise können jedoch Bayes'sche Netzwerke, die verschieden aussehen, identisch sein. Daher können Bayes'sche Netzwerke grob irreführend sein. Essential Graphs sind eine Weiterentwicklung, die diese Schwierigkeit nicht haben.

Welche Modelle sind auf grund von Daten die besten oder wahrscheinlichsten? Um diese Frage angehen zu können, muß man zunächst etwas über die Menge der möglichen Modelle (den Modellraum) wissen – wie viele Modelle gibt es für eine gegebene Zahl von Variablen überhaupt, wie kann man den Modellraum irgendwie ordnen, usw.?

Der Vortrag berichtet über eigene Arbeiten zur Untersuchung des Problemraumes von Essential Graphs. Dazu gehören Algorithmen zur Aufzählung von unlabeled und von labeled Essential Graphs und zur Untersuchung der Frage, wann ein Modell ein Teilmodell eines anderen ist – eine Eigenschaft, die für die Strukturierung des Modellraumes nützlich ist.

Die Untersuchungen zeigen, daß die Zahl der möglichen Modelle selbst für eine relativ kleine Zahl von Variablen unvorstellbar groß ist und daß es sehr schwer ist, eine intuitiv ansprechende Ordnung in dem Modellraum zugänglich zu machen. Der Umgang mit der Komplexität des Modellraumes stellt ein grundsätzliches und schweres Problem dar. Praktische Folgen zeigen sich zum Beispiel bei der teilweise schweren Replizierbarkeit von Ergebnissen in der Verhaltensgenetik oder bei bildgebenden Verfahren. Die Überlegungen zeigen (wieder) wie wesentlich das theoriegeleitete Bemühen um einfache Modelle ist.