



Stasi, Skynet – und Maschinengewehre im Kindergarten

Data Science, Data Analysis, Digitalisierung (D-D-D)... Der Autor Professor Dr. Joachim Grammig hat den Lehrstuhl *Statistik, Ökonometrie und Empirische Wirtschaftsforschung* inne. In diesen WiWi NEWS berichtet er von den spannenden Veränderungen im Bereich D-D-D, seitdem er am Fachbereich 2003 den Lehrstuhl übernommen hat.

von Joachim Grammig

Als ich Mitte der 2000er im Rahmen der Umstellung auf das Bachelor-Master-System auf die Idee kam, die Vorlesung „Statistik I“ (an meiner alten Uni „Stasi I“ genannt) in „Explorative Datenanalyse“ (fancy!) umzubenennen, hätte ich nicht gedacht, dass die WiWi-NEWS „Data Analysis“ einmal als Schwerpunktthema wählen würden. 2003 kam ich nach Tübingen und man sagte mir, dass man sich hier schon mal die Frage gestellt hätte, wozu man eigentlich Statistik (und zwei Statistik-Lehrstühle) braucht. Dank D-D-D und dem Spruch von Googles Chefökonom „I keep saying the sexy job in the next ten years will be statisticians“ stellt man die Frage auch am Neckar nicht mehr. Man hätte neben D-D-D übrigens auch noch die Schlagworte Big Data, Artificial Intelligence und Machine Learning aufnehmen können. Vieles was unter D-D-D et al. läuft, wird an der Schnittstelle von statistischer Datenanalyse und Informatik verhandelt. Machine Learning hieß übrigens ursprünglich „Statistical Learning“, aber das klang den Informatikern wohl etwas zu dröge. Machine Learning, das hat etwas von Skynet, und tönt nicht nur für Fans der Terminator-Filme etwas aufregender.

Innovationen in der digitalen Lehre

In den Vorlesungen meines Lehrstuhls und denen von Professor Biewen spielen D-D-D et al. eine wichtige Rolle und wir sind bestrebt, neue Entwicklungen zeitnah aufzugreifen. Die Studierenden haben uns in den vergangenen vier Jahren dreimal den WiWi-Impuls-Award für verschiedene Innovationen in der digitalen Lehre verliehen. Der diesjährige Universitäts-Lehrpreis, über den WiWi NEWS an anderer Stelle berichtet, zeigt, wie wir versuchen, bereits den Erstsemestern den Einstieg in den Gebrauch von D-D-D Tools wie Python oder R zu ermöglichen. PD Dr. Thomas Dimpfl und unser Doktorand Johannes Bleher haben hier didaktisch einen großen Sprung nach vorne gemacht. Im Rahmen unseres Masterangebots hat Dr. Jantje Sönksen vor Kurzem eine neue Vorlesung konzipiert, „Machine Learning



Noch vor einigen Jahren empfand man bestimmte Methoden als völlig überdimensioniert: „You are handling a machine gun into a kindergarten.“ (Bild: Arnold Schwarzenegger als Terminator, ddp images)

in Econometrics“, mit der sie die Verbindung von Maschinellen Lernen und den dafür benutzten Methoden (z.B. Künstliche Neuronale Netze) mit denen der Mainstream-Ökonometrie aufzeigt.

Die praktische Anwendung wird mit Matlab geübt. Auch im Masterkurs *Statistics of Financial Markets* von Professor Biewen wird diese Schnittstelle behandelt. In den praktischen Teilen meiner Master-Vorlesungen sind statistisch-ökonomische Daten-Analysen am Computer (mit Matlab oder SAS) seit langem üblich. Dieses Training ermöglicht Seminar- und Abschlussarbeiten im Bereich D-D-D et al., die bei uns die Regel sind.

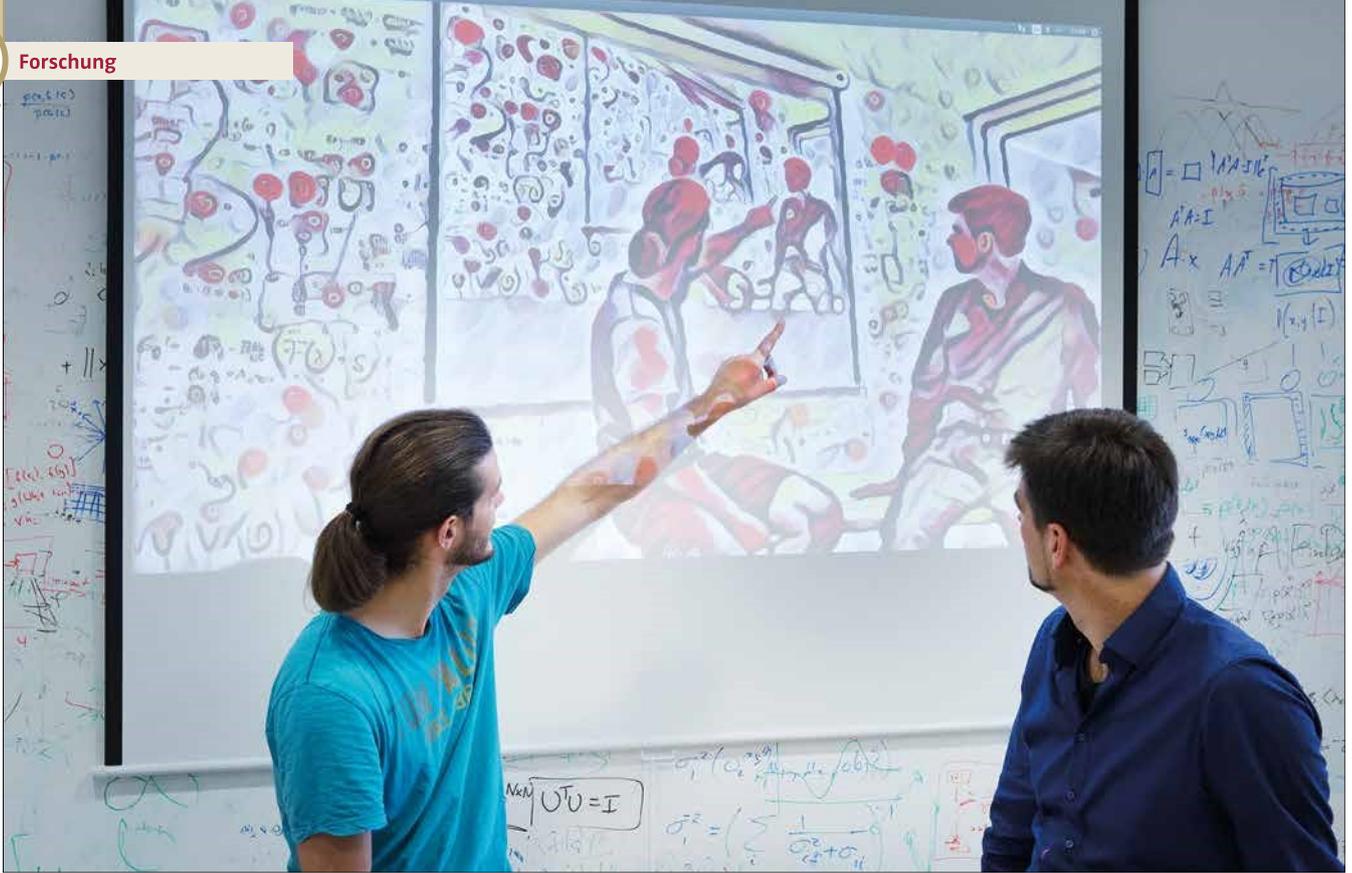
Forschungsfrage: zukunfts- oder vergangenheitsorientiert?

Auch in der Forschung interessieren wir uns für den Einsatz neuer Methoden der Datenanalyse. In einem gemeinsamen Projekt mit Professor Christian Schlag von der Uni Frankfurt untersuchen Jantje Sönksen, Constantin Hanenberg und ich das Potential von Methoden Maschinellen

Lernens zur Analyse von Preisprozessen auf Finanzmärkten. Wir vergleichen diese mit einem diametral entgegengesetzten Ansatz, welcher die in Optionspreisen enthaltenden Markterwartungen nutzt. Das ist sehr spannend, weil jede Data-Science-Methode im Grunde versucht, die sich wiederholenden Muster in historischen Daten zu erkennen. Die in Optionspreisen enthaltene Informationen sind aber inhärent vorausschauend. Was ist der bessere Blickwinkel, der zukunfts- oder der vergangenheitsorientierte? Und ist eine Kombination möglich?

Besonderes Skillset für unsere Studis

Bei aller Euphorie über die Möglichkeiten, die mit D-D-D verbunden sind: Eine Methode, welche die Nutzung künstlicher Intelligenz verspricht, ist nur so klug wie ihr/e Anwender/in. Meine Berufstätigkeit hat mir gezeigt, dass gerade in der Wirtschaftswissenschaft immer wieder einmal komplexe Methoden gehyped werden. Nach meiner Promotion habe ich bei der Unternehmensberatung Roland Berger



Sehende Maschinen: Die KI-Forscher an der Uni Tübingen Alexander Ecker (links) und Matthias Bethge haben einen Algorithmus entwickelt, mit dem sich der Stil eines Kunstwerks auf ein beliebiges Bild übertragen lässt. (Bild: Verena Müller Fotografie)

gearbeitet und eines unserer Projekte war der Einsatz eines künstlichen neuronalen Netzes zur Bestimmung des optimalen Preises für Hotels auf den Kanaren. Der Kunde war ein großes Tourismus-Unternehmen. Das ist schon 23 Jahre her, schon damals waren D-D-D hoch im Kurs – es war ihre erste Blütezeit. Dieses Projekt war für mich sehr lehrreich, und die Bemerkung eines Mitarbeiters des Kunden über die möglichen Folgen unseres Projekts, klingt mir heute noch im Ohr: „You are handing a machine gun into a kindergarten“. Was wir in wiederkehrenden Finanzkrisen erlebt haben, geht in eine ähnliche Richtung. Der Einsatz unzureichend verstandener komplexer Methoden kann gravierende Folgen haben. Aufgrund dieser Erfahrungen weiß ich, gerade als jemand, der ein Faible für Methoden der Datenanalyse hat, wie wichtig es ist, ein gutes Verständnis für die Prozesse zu entwickeln, welche die Daten erzeugen. Um dieses Verständnis zu entfalten, hilft den Studierenden im wirtschaftswissenschaftlichen Kontext eine gute Ausbildung im Bereich theoretischer Grundlagen, sowohl der ökonomisch-theoretischen als auch der statistisch-ökonomischen. Ein solches Skillset, ergänzt um D-D-D-Kompetenzen, wird Absolventen der Wirtschaftswissenschaft von denen der Informatik abheben, die über diese Dinge in ihrem Studium sehr viel weniger erfahren.

Neue Hardware trifft gewachsene Anforderungen

Am Fachbereich Wirtschaftswissenschaft arbeiten zahlreiche Forschende und fortgeschrittene Studierende seit Jahren mit großen Datensätzen und rechenintensiven Methoden, die erhebliche Rechenkapazität benötigen. Eine Projektgruppe hat daran gearbeitet, die (Groß-)Rechenkapazitäten des Fachbereichs mit Blick auf große Datensätze und die Data Science Initiativen weiterzuentwickeln. Dieses Konzept ruht auf drei Säulen.

1. Für den PC-Pool des Fachbereichs wurden 40 neue Rechner beschafft, die mit Intel i7-Prozessoren der neuesten Generation und 64 GB Arbeitsspeicher ausgerüstet sind. Damit verfügt der Fachbereich über einen PC-Pool, der – was die Rechenstärke angeht – seinesgleichen sucht und beste Bedingungen für Studierende bietet.

Für die großen und rechenintensiven Projekte von Forschenden auf allen Ebenen (einschließlich beispielsweise anspruchsvoller Masterarbeiten) ist die Herausforderung, dass sich die Anforderungen an Rechenkapazitäten in mehreren Dimensionen unterscheiden. Manche Projekte brauchen lange Laufzeiten bei geringem Ressourceneinsatz, andere brauchen einen Rechenkern, aber großen Arbeitsspeicher (z.B. umfangreiche Datensätze mit nicht-numerischen Variablen), andere brauchen zahlreiche Rechenkern bei langer Laufzeit (z.B. multi-chain Bayesian

Estimation mit Markov Chain Monte Carlo Verfahren), andere brauchen Kombinationen aus all diesem. Um diesem Mix an Anforderungen gerecht zu werden, bietet der Fachbereich nun zwei Komponenten:

2. Forschende können – wie bisher – das High Performance Cluster des Landes Baden-Württemberg (bwHPC) nutzen, insbesondere den Standort in Karlsruhe und Mannheim. Hier sollen in Zukunft verstärkte Schulungen angeboten werden, um Interessierte damit vertraut zu machen. Dieses eignet sich vor allem für sehr ressourcenintensive Berechnungen mit begrenzter Laufzeit, die im Batch-Modus ausgeführt werden. Für interaktive Arbeit an Daten und sehr lang laufende Berechnungen, ist das System weniger geeignet.

3. Für interaktive, langlaufende Berechnungen mit proprietärer Software wie z. B. Stata bietet der Fachbereich nun einen unkomplizierten und komfortablen Zugriff auf die sogenannte bwCloud, bei der Forschende sich mit weniger als zehn Klicks einen Linux-basierten virtuellen Rechner in der Cloud einrichten können. Der virtuelle Rechner kann auch individuell eingerichtet werden und steht den Nutzern bis zu 40 Tage zur Verfügung.

Mit diesem Paket hat der Fachbereich ein ausgezeichnetes Portfolio an Lösungen, das eigentlich für jede Art von Rechenvorhaben etwas bietet.

Dominik Papies