Wolfgang E. Nagel
© Foto: Robert Gommlich

Thomas Ludwig

Data Analytics

Data Analytics ist ein Teil von Data Science und befasst sich mit der Analyse von Ausgangsdaten, um zu neuen Einsichten und Handlungsempfehlungen zu gelangen. Als Data Science oder Datengetriebenes Forschen und Entwickeln bezeichnen wir dagegen wissenschaftliche Erkenntnis, die aus allen Aspekten des Lebenszyklus von Daten, von ihrer Entstehung über die Analyse und Archivierung bis zu ihrer Weiterverbreitung, gewonnen wird. Beide Schlagworte sind in der modernen Welt ständigen Anpassungen unterworfen, und wir wollen in diesem Themenheft einen Eindruck des Spektrums von Data Analytics vermitteln.

Data Analytics befasst sich also mit der Extraktion von Informationen aus großen Datenmengen. Methodisch kommen hier insbesondere Verfahren der Statistik zum Einsatz. Eine moderne Ausweitung dieser Methoden finden wir im Maschinellen Lernen und insbesondere im Deep Learning. Beide Ansätze versprechen schnelle und effiziente Erkenntnisgewinne, ihr Einsatz ist aber stets auch problematisch, und ihre Aussagekraft muss sorgfältig analysiert werden.

Im vorliegenden Themenheft betrachten wir Data Analytics zunächst klassisch vom Standpunkt eines Statistikers. Ihren Einsatz finden solche Methoden sowohl in der Forschung als auch in der Industrie. Der Leser findet Beispiele aus den Lebenswissenschaften und auch aus verschiedenen Bereichen des industriellen Umfelds. Neben den numerischen Varianten der Datenanalyse gibt es auch eine visuelle Analyse. Am Beispiel der Klimaforschung zeigen die Autoren, wie mit großen Datenmengen gearbeitet wird und welche Herausforderungen auf diesem Gebiet zu finden sind. Dabei erzählen Bilder auch Geschichten – eine eigene Variante der Data Analytics, die im Data Story Telling ihren Ausdruck findet.

Die Bedeutung von Data Analytics in der Forschung und Entwicklung erfordert auch neue Ausbildungsformen. In München nimmt die Munich School for Data Science dieser Tage ihre Arbeit auf. Ihr Koordinator erläutert die Konzepte dieses Ansatzes.

Data Analytics stützt sich auf vorhandene Daten ab. Das gesamte Ökosystem ihrer Erzeugung, Analyse, Archivierung und Verteilung nutzt bestehende Forschungsinfrastrukturen. An zwei ausgewählten Beispielen wird die Einbindung in Data Science dargestellt. Fachgebietsübergreifende Bemühungen auf nationaler Ebene dienen dabei der Strukturierung der Forschungslandschaft und sollen eine vielseitige und langfristige Datennutzung ermöglichen.

Die Ideen zu einer datengetriebenen Wissenschaft gehen unter anderem auf eine Veröffentlichung aus dem Jahr 2009 zurück: „The Fourth Paradigm: Data-intensive Scientific Discovery“. Ihr Herausgeber Tony Hey blickt im Themenheft zurück auf 10 Jahre Entwicklungsgeschichte dieser Strömung und analysiert den aktuellen Stand.

Wolfgang E. Nagel
Thomas Ludwig

Big Data Analytics am Centrum für Systembiologie Dresden (CSBD).

Das Bild zeigt ein Foto aus der CAVE des CSBD, wo Embryos (hier das einer Fruchtfliege) mittel virtu-

eller Realität immersiv betrachtet und mittels Computersimulation (kleines Bild, Simulation von Proteindiffusion in einer Zelle) modelliert und verstanden werden können.

i
Zum Titelbild