

Big Data

Das Volumen aller weltweit in einem bestimmten Zeitraum produzierten Daten verdoppelt sich heute ungefähr alle zwei Jahre. Die mit der Nutzbarmachung der entstehenden Datenberge zusammen hängenden Aspekte firmieren unter dem Begriff Big Data und sind wesentlich dadurch gekennzeichnet, dass die betreffenden Daten mit herkömmlichen Methoden nicht mehr sinnvoll verarbeitet werden können und daher neue Verfahren und Technologien zu ihrer Verarbeitung notwendig sind. Dabei ändern sich die zu betrachtenden Informationsmengen mit wachsendem technischen Fortschritt. So denkt man im Zusammenhang mit Big Data heute typischerweise an Datenmengen im Bereich von Peta- bzw. Exabyte (10^{15} bzw. 10^{18} Byte), wobei davon auszugehen ist, dass innerhalb der nächsten Jahre die Grenze zu Zettabyte (10^{21} Byte) überschritten wird. Neben der reinen Menge (Volume) rücken neuerdings vermehrt auch die Geschwindigkeit (Velocity), die Vielfalt (Variety) und darüber hinaus auch noch die Nützlichkeit (Value), die Komplexität (Complexity) sowie die Vollständigkeit bzw. Richtigkeit (Veracity) der Daten in den Fokus der Betrachtungen. Alle diese Aspekte zusammen charakterisieren inzwischen das Feld Big Data. Im Fokus stehen dabei insbesondere unstrukturierte Daten, wie z.B. Text-, Audio- und Videodaten, die insbesondere beim Online-Handel, in sozialen Netzwerken, bei Internet-Suchmaschinen sowie beim zunehmenden Einsatz von Überwachungssensoren generiert werden.

Eine wesentliche Technologie von Big Data ist die parallele Verarbeitung von Daten. Seit einiger Zeit gebräuchlich ist das sogenannte MapReduce-Verfahren, das die Firma Google 2004 vorgestellt hat. Dabei werden die zu bearbeitenden Daten auf die einzelnen Rechner eines sogenannte Computer Clusters verteilt und dort parallel verarbeitet. Eine weit verbreitete Open-Source-Implementierung von MapReduce stellt Hadoop von der Firma Yahoo dar, für das aktuell technologische Verbesserungen entwickelt werden. Eine vieldiskutierte und bislang offene Fragestellung ist, welche Probleme aus dem Bereich Big Data

überhaupt für MapReduce bzw. Hadoop zugänglich sind.

Bei der Erkenntnisgewinnung aus dem umfangreichen Datenmaterial wird man sich verstärkt auf geeignete Methoden der Visualisierung abstützen, die die hochentwickelten Fähigkeiten des Menschen bei der Mustererkennung nutzen. Der Auffindung von Mustern in großen Datenbeständen durch Anwendung bestimmter statistischer Verfahren dient weiterhin das sogenannte Data Mining, bei dem Verfahren der sogenannte Künstlichen Intelligenz zum Einsatz kommen werden. Innerhalb des Data Minings wird das Auffinden von Mustern in kontinuierlichen Datenströmen (Data Stream Mining), in Freitexten (Text Mining) und in semistrukturierten Daten (Graph Mining) immer wichtiger sein. Beim Text Mining wird die automatische Verarbeitung von in natürlicher Form vorliegender Sprache, das Natural Language Processing, eine Schlüsselstellung einnehmen.

Ein besonderes Interesse bei der Auswertung von Datenmengen besteht in der Erkennung von Mustern, die zukünftige Entwicklungen bzw. zukünftiges Verhalten, wie z.B. das Käuferverhalten, vorausagen. Diese Analysemethoden werden als Predictive Analytics bezeichnet und im Zusammenhang mit Big Data deutlich an Bedeutung gewinnen. Aufgrund der Vielzahl der bei Big Data vorliegenden Daten ist zu erwarten, dass die hiermit generierten Voraussagen zutreffender sein werden als bei schmaler Datenbasis. Eine interessante Technologie für Big Data stellt auch das Complex Event Processing (CEP) dar, bei dem aus den vorliegenden Daten einzelne Ereignisse abgeleitet werden und innerhalb dieser Ereignisse nach komplexen Zusammenhängen gesucht wird.

Von großem Interesse ist bei Big Data das umfangreiche aus sozialen Netzwerken stammende Datenmaterial. Daher wird die Bedeutung und Leistungsfähigkeit geeigneter Verfahren zur Analyse dieser Daten (Social Network Analysis) deutlich zunehmen. Sowohl im unternehmerischen als auch im öffentlichen Bereich kann Big Data die Entscheidungsfindung unterstützen. In Unternehmen kann Big Data neue Geschäfts-

ideen bzw. Produkte oder Vorschläge für Produktverbesserungen generieren. Ebenso können Geschäfts- und Produktionsprozesse verbessert werden. Stark profitieren dürfte vor allem der Bereich Marketing und Vertrieb. Im öffentlichen Bereich kann die Abstützung auf große Datenmengen unter anderem bei der Städteplanung sowie einem Verkehrsmanagement, das auf aktuelle Ereignisse sofort reagiert, erfolgreich sein. Im Gesundheitswesen kann Big Data einerseits bevorstehende Pandemien zu erkennen helfen und andererseits den Übergang zu einer präventiven Medizin weisen. Einen sehr großen Beitrag verspricht Big Data für den wehr- und sicherheitstechnischen Bereich. Die Auswertung großer Datenmengen, insbesondere aus sozialen Netzwerken, wird bei der Kriminalitätsbekämpfung, speziell der Terrorismusabwehr, als erfolgversprechend angesehen. Zur Erhöhung der IT-Sicherheit kann Big Data bei der Erkennung von Angriffen auf Netzwerke (Network Attack Monitoring) gewinnbringend eingesetzt werden. Auch darüber hinaus ergeben sich vielfältige wehrtechnische Anwendungsmöglichkeiten für Big Data. So kann die Aufbereitung von Daten innerhalb der Vernetzten Operationsführung (NetOpFü) deutlich verbessert werden, insbesondere im Hinblick auf die Generierung eines gemeinsamen, rollenbasierten Einsatzlagebildes (GREL). Im militärischen Sanitätswesen schließlich kann Big Data die Leistungsfähigkeit der Überwachung des Soldaten im Einsatz (War Fighter Monitoring) stark verbessern. Bei allen Chancen sind mit Big Data aber auch Risiken verbunden, die durchaus zu reglementierenden Eingriffen und damit zu einer Verlangsamung der Entwicklungsgeschwindigkeit führen könnten. Dabei liegen die Möglichkeiten des Missbrauchs insbesondere im Bereich des Datenschutzes. Abhilfe könnten hier sogenannte Anonymisierungsverfahren schaffen, die persönliche Daten in eine große Menge anderer Daten integrieren, so dass die Privatsphäre des Einzelnen gewahrt bleibt. Die Entwicklung solcher Verfahren ist derzeit jedoch noch eine offene Forschungsfrage.

Thomas Euting