

32 Big Data und der „Fluch der Dimensionalität“

Big Data ist das große Thema in der IT-Branche. Fast täglich gibt es Berichte in den Medien über Ge- und Missbrauch umfassend erhobener Datenbestände. Die Angaben über die Menge und Struktur der betreffenden Daten ist nur sehr unklar angegeben. Das Einzige was man mit Sicherheit sagen kann ist, dass das Volumen ungeheuer groß sein dürfte.

Seriöse Quellen sprechen davon, dass sich derzeit das gesammelte Datenvolumen weltweit alle 2 Jahre verdoppelt. Die angesehen Zeitung „Die Welt“ schrieb 2014: „Die Menschheit erstickt an ihren Daten“ und nennt die Zahl 10^{21} Bits, also eine Zahl mit 21 Nullen.

Aber nicht die schiere Anzahl der Bits ist das Problem, sondern die Dimensionalität der notwendigen Datenbank- „Würfel“. "Es ist wichtig, die Methodologie zu verstehen: wie die Anzahl von Informationen gemessen werden kann, um sozioökonomische Entwicklungen besser nachvollziehen zu können", sagt Martin Hilbert, Kommunikationswissenschaftler an der University of Southern California. Zitate aus <http://www.welt.de/wirtschaft/webwelt/article12507803/Die-Menschheit-erstickt-an-ihren-Daten.html>

Fluch der Dimensionalität ist ein Begriff, der von Richard Bellman eingeführt wurde, um den rapiden Anstieg im Volumen beim Hinzufügen weiterer Dimensionen in einen mathematischen Raum zu beschreiben. Quelle Wikipedia

Wie viele Dimensionen muss nun ein Datenraum aufweisen? So viele als notwendig und so wenig als möglich lautet die Antwort der IT-Pragmatiker.

Aber wie stellen wir fest wie hoch die Dimensionalität sein muss? Für Firmen wie BEKO ist die Frage einfach zu beantworten. Genauso viele Dimensionen wie der Kunde will. Und der Kunde will immer mehr. Je komplexer die gewünschte Darstellung der „Wirklichkeit“ desto höher werden die Dimensionalitäten. Längst sind sie auch nicht mehr ganzzahlig, sondern zunehmend auch gebrochen im Sinne der Hausdorff-Dimensionen.

Da diese Materie nur schwer erklärbar ist hat das IHI seit Jahren auf die Methode der sogenannten Venn-Diagramme gesetzt, die viele Aspekte der Dimensionalitäten-Problematik visualisieren kann. Inzwischen sind zahlreiche finanziell mehr oder weniger erfolgreiche Projekte auf Basis dieser empirisch/theoretischen Experimente abgewickelt worden und haben ein reiches Wissen über die tieferen Zusammenhänge in Datenbanken als Mengenkonstrukte gebracht.

Nun können diese Knowhow-Bestände zunehmend dazu dienen, Entwicklungen innerhalb der IT-Szene besser als der Wettbewerb einzuschätzen und am Kapitalmarkt für die Portfoliopflege einzusetzen.

Wenn man halbwegs fundiert erkennen kann ob ein neuer oder bestehender Player am Markt auf eine Big-Data Strategie setzt, die auch strukturtheoretisch Sinn macht, ist das ein eindeutiger Wissensvorsprung, der sich in Wertsteigerungen des Investments ausdrücken wird.

Die moderne Physik macht es uns vor. In dieser unzweifelhaften Leitwissenschaft der modernen Welt hat die Dimensionen-Frage höchsten Stellenwert. Die heftigen Fachdiskussionen um das holografische Prinzip das vom Nobelpreisträger t´Hooft, Susskind in Stanford und vielen anderen vertreten wird, beweisen wie wichtig Fragen der Dimensionalität sein können. Als Pionier der Verwendung gebrochener (nicht ganzzahliger Dimensionen) ist unbestritten Mandelbrot mit seinen Fraktalen, der nachgewiesen hat, dass ganze Industriezweige auf Basis fraktaler Strukturen geschaffen wurden. So ist zum Beispiel das moderne Smartphone ohne fraktale Sendeantenne nicht mehr denkbar, weil nach der alten Antennenbauweise das Handy wegen der gängigen Multifrequenztechnik wie ein Igel aussehen müsste.

Auch der Hochfrequenzhandel an dem Weltbörsen wäre ohne fundierte Dimensionalitätskenntnisse nicht programmierbar. Im Hause BEKO sind die Mitarbeiter deshalb so wenig mit Dimensionalitätsfragen befasst, weil diese in den Programmttools, die wir verwenden bereits im

algorithmischen Code eingearbeitet sind und vom Nutzer gar nicht mehr als Problem wahrgenommen werden.

Ganz anders ist es bei personenbezogenen Daten. Hier gibt es immer mindestens einen unmittelbaren Betroffenen. Dieser wird von der autokreativen Wucht multidimensionaler Datensätze heutzutage so überrollt, dass er manchmal gar nicht mehr die Wirklichkeit seines eigenen Zustandes erkennen kann. Das gilt beispielsweise bei Zuordnungsfehlern in der Gerätemedizin, wo es immer wieder zu tragischen Fehlbehandlungen kommt, bis hin zur Terrorprevention wo Kampfdrohnen völlig unbeteiligte Menschen umbringen.

Das vom IHI schon lange erforschte Phänomen der autokreativen Transformation zeigt, dass in genügend komplexen multidimensionalen Datenwürfeln je nach Drilling-Algorithmus zwangsläufig polymorphe Muster entstehen, die dann in Auswertungsprogrammen so behandelt werden wie echte Urdaten. Je länger die transformatorischen Ketten werden, in denen die Daten immer weitergereicht werden und dann vor einem vielleicht ganz anderen kontextuellen Hintergrund maschinell interpretiert werden ist es unvermeidbar, dass sogenannte Artefakte entstehen, die fälschlicherweise als reale Sachverhalte eingestuft werden. Das kann sehr fatal werden. Es sind sogar Kreisläufe von Transformationsketten denkbar und wahrscheinlich, wo die Artefakte die Realdaten überschreiben und so zu scheinbaren objektiven Befunden werden ohne dass jemand etwas merkt. Viele Fehlentscheidungen der letzten Zeit könnten darauf zurückzuführen sein. Besonders anfällig für solche Phänomene ist die Finanzindustrie und die zentralisierte Medizin.

Steven Wolfram hat mit seinen einfachen zellulären Automaten bewiesen, dass unter den vielen Kurzalgorithmen die er getestet hat eine Anzahl anzutreffen ist die zu unwiederholbaren Mustern führen. Das gleiche ist Mathematikern in der Zahl Pi wohlbekannt, wo ebenfalls jede denkmögliche Sequenz von Ziffern darstellbar ist, wenn man nur genügend viele Dezimalstellen für Pi errechnet und durchfiltert.

In letzter Zeit häufen sich Anzeichen, dass die Sicherheitsindustrie einen Punkt erreicht hat, wo Daten-Artefakte buchstäblich über Leben und Tod entscheiden.

55. IHI Bericht, 2015