

Intelligenz erwünscht: Machine Learning bei Big Data

Bei unstrukturierten Mengendaten, wie sie in Big-Data-Umgebungen anfallen, wird die Analyse und Prognose von Geschäftsentwicklungen zunehmend schwer. Hier bietet sich die Nutzung von intelligenten lernenden Systemen an, die die gewünschten Ergebnisse automatisch aus dem vorhandenen Pool an Daten extrahieren können.

Ein Weg hierfür ist die Nutzung von Machine Learning, also Systeme, die nicht nur aufgrund eines vorgegebenen Lösungspfades agieren, sondern überdies hinaus aus Resultaten lernen und die Vorgehensweise automatisch optimieren können. Kurz gesagt, handelt es sich um nichts geringer als die Generierung von Wissen aus Erfahrung anstelle der Vorhersage aufgrund "auswendig gelernter" Fakten.

Ein solches System erkennt im Gegensatz zu Data Mining und Knowledge Discovery in Databases (KDD) Muster selbstständig. Das kann bei verschiedenen Anwendungen von entscheidender Bedeutung sein. Beispielsweise könnte durch die Erkennung des Kundenverhaltens bei Zahlvorgängen autonom auf Kreditkartenbetrug geschlossen werden. Auch in autarken Systemen, etwa dem [selbständig fahrenden Auto](#) oder in der Sprach- und Texterkennung können diese Systeme weitaus bessere Ergebnisse liefern als solche, bei denen der Lösungsweg fest vorgegeben ist.

Der reinste Daten-Zoo

Nun können in [Big Data](#) unterschiedlichste Arten von Daten vorkommen. Diese unterscheiden sich in Aufkommen und Komplexität. Dabei stellen sich – unabhängig von der Struktur der Daten – meist ähnliche Probleme. Bei großen Datenmengen müssen Ergebnisse schnell vorliegen.

Klassische [Datenbanken](#) können dies nicht liefern, da sie Datensatz für Datensatz sämtliche relevanten Felder durchsuchen müssen. Auch setzen verschiedenartige Informationstypen, etwa Texte, numerische Informationen oder Bilder ein Big-Data-System voraus.

Den Kunden verstehen lohnt sich

Viele Aufgaben sind mit den herkömmlichen Mitteln einer reinen Datenbankanalyse nicht zu bewerkstelligen. Möchte beispielsweise ein Onlineversand den Grad der Beliebtheit seiner im Internet angebotenen Waren anhand der Onlinerezensionen der bisherigen Käufer analysieren, begnügen sich viele mit der quantitativen Erfassung der vergebenen Noten beziehungsweise Sterne.



Die Vergabe von Noten erfasst aber nur eine ungefähre Stimmung. Wie die Kunden einzelnen Kriterien des Produkts oder der Dienstleistung gegenüberstehen, bleibt verborgen, solange der E-Shop-Betreiber nicht in mühevoller Einzelarbeit Kommentare liest und entsprechend einordnet. Big-Data-Anwendungen bieten hier einen Weg, sowohl Kosten zu sparen als auch die Auslastung der beteiligten Teams zu reduzieren.

Sentiment Detection ermittelt, wie der englischsprachige Begriff nahelegt, die automatisierte Ermittlung vorherrschender Stimmungen. Mittels Sentiment Detection können so Texte automatisch sowohl auf ihre Grundtendenz als auch auf häufig genannte und sowohl kritisch als auch lobend herausgestellte Produkt- oder Dienstleistungsmerkmale untersucht werden. Doch auch die Vernetzung einzelner Meinungsbildner ist in Zeiten von Social Media wichtig. So ist es interessant zu sehen, wie Personen, deren Meinung besonderes Gehör findet, miteinander interagieren und wie sie sich über bestimmte Themengebiete austauschen.

Big Data (fast) überall

Auch bei der Mustererkennung kann maschinelles Lernen im Big-Data-Kontext seine Trümpfe ausspielen. So können handgeschriebene Texte oder sogar Captchas effizient ausgelesen werden. Wo herkömmliche Zeichenerkennung (OCR) versagt – beispielsweise in der Logistik – verhilft Machine Learning Unternehmen zu mehr Effizienz. Doch auch bei korrekter Erfassung graphischer oder biometrischer Informationen können Unternehmen und Organisationen auf intelligente Routinen zurückgreifen.

Grundlagen des Machine Learnings

Auf welchem Wege stellt ein intelligentes System die Informationen bereit? Hier kommen verschiedene Methoden ins Spiel: Statistische Analysen, etwa die lineare Regression, Random Forest, bestehend aus mehreren unkorrelierenden Entscheidungsbäumen, oder der k-Means-Algorithmus, der als Verfahren zur Vektorquantisierung zur Clusteranalyse herangezogen wird.



Letztendlich orientiert sich maschinelles Lernen an der Arbeitsweise des menschlichen Gehirns und simuliert Neuronenverbindungen. Am Beispiel einer optischen Mustererkennung lassen sich Parallelen aufzeigen. Nachdem das intelligente System die Pixel eines Bildes gemäß ihrer Helligkeit untersucht hat, geht es dazu über, etwaige Muster zu erkennen. Diese Muster werden anschließend verbunden und ihre Formen identifiziert.

Zu guter Letzt werden die Formen zu komplexen und zusammenhängenden Objekten zusammengesetzt und erkannt. Generell erfolgt die Erkennung zunächst aus dem Detail heraus, ob es nun – wie im Beispiel – ein Gesicht ist oder andere Merkmale wie etwa Kommunikationsverhalten, Zahlungsbewegungen oder Bewegungsdaten in der Smart City.

Aus der Granularität der Ursprungsdaten entspringt die Notwendigkeit einer hoch performanten Big-Data-Lösung. Zwar stammt das Konzept des Machine Learning bereits aus den achtziger Jahren des vergangenen Jahrhunderts, doch erst jetzt stehen die Systeme zur Verfügung, um wahrhaft intelligente Systeme zuverlässig umzusetzen.

Tipp: Was haben Sie über Big Data abgespeichert? [Überprüfen Sie Ihr Wissen – mit 15 Fragen auf silicon.de.](https://www.silicon.de)