



Warum sich Machine Learning in der Cloud für Unternehmen lohnt

Computerwoche vom 23.10.2015 von [Michael Grözinger \(Experte\)](#)

Cloud-Plattformen wie Amazon Web Services oder Microsoft Azure erlauben es, Machine Learning in der Cloud zu implementieren und dann in verschiedenen Anwendungen einzusetzen. Davon profitieren etwa Marketing- und E-Commerce-Spezialisten, aber auch Anwender in der Medizin- und der Automotive-Branche.

Egal ob wir bei Amazon oder Zalando einkaufen, Google Mail nutzen oder einfach nur einen Begriff im Internet suchen: es überrascht uns nicht mehr, dass uns die Online-Händler plötzlich weitere Produkte anbieten, die uns interessieren könnten. Auch dass Anzeigen für Scheidungsanwälte am Seitenrand erscheinen, wenn wir im regen Mailverkehr mit dem frischgeschiedenen Freund stehen: die Maschine hat über uns Daten gesammelt und daraus gelernt, uns zu klassifizieren.

In Zeiten exponentiell zunehmender Rechen- und Speicherkapazitäten ist das Prinzip Machine Learning, also maschinelles Lernen, allgegenwärtig. Anwendungsbereiche von Machine Learning findet man aber nicht nur im Marketing. Sprach- und Bilderkennung und autonome Steuerung werden erfolgreich in Fahrzeugen eingesetzt. Die Medizin nutzt Bilderkennung und Mustererkennung in der Diagnostik, bei der Auswertung von hochdimensionalen genetischen Daten und bei der Analyse neuronaler Verknüpfungen im Gehirn, denn dort finden ja die natürlichen Lernvorgänge statt.

Anders jedoch als im Gehirn, das fast von selbst an Kapazität zulegt, rasant wächst und lernt, stellt Machine Learning enorme Anforderungen an Rechnerkapazität, Design und Implementierung der Lösungen. Verbunden ist das mit hohen Kosten für die Anschaffung und den Betrieb von Systemen.

[Cloud Computing](#) steht für flexibel und kosteneffektiv verfügbare Rechenkapazität. Das schließt auch Plattform-as-a-Service ([PaaS](#)) ein, mit dem Unternehmen aus verschiedenen Komponenten, die in der **Cloud** ausgeführt werden, neue Lösungen erstellen können. Verschiedene Anbieter haben jetzt ihr PaaS Angebot um Machine Learning erweitert. Kenntnisse der Grundlagen des Machine Learning vorausgesetzt, können Unternehmen schnell und einfach Modelle in der Cloud trainieren und darauf basierende Anwendungen erstellen.

Machine-Learning-Beispiel: Diagnose-App in der Medizin

Ein einfaches Beispiel verdeutlicht das: Es soll eine einfache und praktikable Diagnose-App entwickelt werden, die es ermöglicht, verschiedene Krankheitskategorien zu unterscheiden. Aus früheren Studien liegt eine Datenbank vor, in der für eine Gruppe von Patienten ("Instanzen") eine große Anzahl von Informationen, wie zum Beispiel Laborwerte, Alter, Erkrankungsparameter ("Variablen") und die Einordnung in die Krankheitskategorie vorliegt. Die Entwickler laden diese Daten als .csv-Datei in den Cloud-Speicher und trainieren Modelle mit Hilfe der von ihnen definierten Algorithmen. Die Modelle werden als Web-Services, geschützt durch Zugriffsschlüssel, in der Cloud abgespeichert. Apps oder Web-Anwendungen können dann auf diese Modelle zugreifen. Der Anwender gibt die ihm bekannten Laborwerte in die App ein und erhält z.B. eine Empfehlung, eine Einordnung oder einen Risiko-Wert.



Was Azure- und AWS-Plattformen bieten

Unter anderen bieten Amazon und Microsoft derartige Cloud Plattformen bereits an. Diese können nach der Einrichtung eines kostenlosen Benutzerkontos direkt genutzt werden und bieten eine nutzungsabhängige Abrechnung. Das umfasst die Nutzung der Entwicklungswerkzeuge und den Betrieb der erstellten Modelle als Web Services.

Die grundsätzlichen Funktionen solcher Plattformen sind:

- **Werkzeuge zum Design und zur Nutzung:** Die Cloud Plattform stellt Anwendern Web-basierte Werkzeuge bereit, mit denen die kompletten Machine Learning Anwendungen erstellt und genutzt werden können. Das macht den Einstieg und den Einsatz sehr einfach. In der Microsoft Azure Cloud steht beispielsweise ein Design Tool zur Verfügung, das komfortabel im Browser bedient werden kann. Die Abbildung unten zeigt, wie ein Experiment in der Azure Cloud mit Azure ML Studio erstellt wird.



So wird ein

Experiment mit Microsoft Azure Machine Learning erstellt.

Foto: Microsoft

- **Datenspeicherung in der Cloud:** Als ersten Schritt beim Nutzen der Services laden die Entwickler der Anwendung die Daten in den Cloud-Speicher. Dabei können die Speicherorte des Rechenzentrums ausgewählt werden. Datenspeicher können unter anderem eine NoSQL-Datenbank oder eine einfache .csv Datei sein.
- **Visualisierung der Daten:** Die Visualisierung der Daten vor und nach dem Training der Modelle ist ein wichtiges Hilfsmittel beim Debugging und beim Optimieren. Bei manchen Modellen benötigt man Eingangsdaten mit einer bestimmten statistischen Verteilung. Die Visualisierungsfunktionen ermöglichen dazu eine schnelle Beurteilung der Daten.
- **Datentransformation:** Beim Training und der Optimierung von Modellen ist es oft notwendig, die so genannten Features, also die Eingangswerte der Modelle, zu transformieren oder fehlende Werte zu



ergänzen. Dazu stehen in der Cloud Module oder Funktionen bereit, die das automatisiert durchführen. Es lassen sich Programme in der Programmiersprache R integrieren, die es ermöglichen, komplexere eigene Datentransformationen durchzuführen.

- **Verschiedene Modelle und Algorithmen:** Abhängig von der gewählten Cloud-Plattform stehen verschiedene Modelle, zum Beispiel für Klassifizierung oder Clustering, zur Verfügung. Jedes Modell verfügt über die typischen Parameter, die entsprechend zur Optimierung der Ergebnisse eingestellt werden können.
- **Beurteilung der Qualität der trainierten Modelle:** Bei der Nutzung von Machine-Learning-Modellen werden je nach Art der Aufgabe verschiedene Parameter zur Beurteilung der Qualität des trainierten Modelles eingesetzt. Beispiele sind Precision, Recall oder F1 Score. Diese stehen auch in der Cloud zur Verfügung.
- **Programmierschnittstellen, so genannte APIs, und Web Services für die operative Nutzung:** Nachdem die Modelle trainiert und optimiert sind, können diese als Web Service über das Internet genutzt werden. Die Erstellung und Veröffentlichung der Web-Services erfolgt komplett automatisiert.
- **Nutzungsabhängige Bezahlung:** Bei allen Plattformen erfolgt die Abrechnung abhängig von der konkreten Nutzung. Für jedes entworfene Experiment kann der Entwickler die Kostenkomponenten analysieren und damit die Kosten für eine Machine-Learning-Anwendung abschätzen.

Wann rentiert sich Machine Learning in der Cloud?

Die oben beschriebenen Cloud Dienste lassen sich vor allem dann gut einsetzen, wenn die Daten, die ausgewertet werden sollen, kontinuierlich entstehen und die Datenquellen sich an verschiedenen Orten im Internet befinden. Der Einsatz einer Cloud-basierten Lösung lohnt sich umso mehr, wenn die Machine-Learning-Services von vielen Nutzern verwendet werden. Wenn die Daten in einer großen Datenbank lokal gespeichert sind und die Ergebnisse nur intern genutzt werden, ist es gegebenenfalls sinnvoller, die Lösung im eigenen Rechenzentrum zu implementieren.

Fazit

Grundsätzlich bieten Cloud-basierte Machine Learning Services den Vorteil, dass die Erfahrungen und Kenntnisse großer Dienstleister und Softwareentwickler in die Lösungen eingeflossen sind. Der Einstieg in Machine Learning wird den Anwendern durch kostenlose Test-Abonnements leicht gemacht. Diese sind zeitlich und bezüglich der Datenmengen und Rechenkapazität begrenzt, ermöglichen aber erste Erfahrungen.

Ohne dass es uns bewusst ist, hat Machine Learning unser tägliches Leben durchdrungen und wurde unverzichtbar. Bei der Verarbeitung von Daten mit Machine-Learning-Anwendungen in der Cloud ist Datensicherheit und Datenschutz eine wichtige Aufgabe. Dazu sollten Datenschutzbeauftragte beim Design frühzeitig eingebunden werden.