Das Fraunhofer-Institut für Naturwissenschaftlich-Technische Trendanalysen berichtet über Neue Technologien

Automatisiertes maschinelles Lernen

Künstliche Intelligenz ist als Schlagwort derzeit in aller Munde. Die heute besonders leistungsstarken Anwendungen, wie z. B. die großen generischen Modelle GPT zur Sprachverarbeitung und Stable Diffusion zur Bildgenerierung, finden mittlerweile in vielen Bereichen Verwendung. Diese, wie auch eine Vielzahl an weiteren kleineren, spezialisierten Modellen basieren überwiegend auf maschinellem Lernen (ML), indem sie Daten nutzen, um Aufgaben zu lösen. Die dabei zugrundeliegenden Verfahren sind heute noch zum großen Teil auf die Mitwirkung menschlicher Experten angewiesen. Hier könnte eine zunehmende Automatisierung in Zukunft zu entscheidenden Leistungssteigerungen führen.

In aktuellen ML-Prozessen sind einerseits Fachleute im Anwendungsgebiet erforderlich, die über ein Verständnis für die konkret zu lösenden Aufgaben und die dafür vorhandenen, relevanten Daten verfügen. Daneben werden andererseits hoch spezialisierte Experten im Bereich der Datenanalyse benötigt, die geeignete Daten aufbereiten und den ML-Prozess aufsetzen. Diese ML-Experten gestalten den Prozess größtenteils manuell und arbeiten dabei eng mit den Fachleuten zusammen, um aufgabenspezifische Lösungen effektiv bereitstellen zu können. Das Durchlaufen des ML-Prozesses erfordert damit eine erhebliche Menge an Ressourcen und Zeit. So kann u. a. der Zeitraum von der Auswahl der zu lösenden Aufgaben bis zur Erstellung eines fertigen ML-Modells im Bereich von mehreren Wochen bis zu einigen Monaten liegen.

Das automatisierte maschinelle Lernen (AutoML) soll den derzeit hohen manuellen Aufwand reduzieren, indem es einerseits ML-Experten unterstützt und deren Effizienz steigert, um so leistungsfähigere ML-Lösungen zu erhalten. Andererseits soll es Anwendern ermöglichen, auch ohne datenwissenschaftlichen Hintergrund automatisiert ML-Lösungen für den eigenen Bedarf zu erstellen. AutoML ist als aktuelles Forschungsgebiet in ziviler und militärischer Hinsicht von großem Interesse, weil damit Anwendungen der künstlichen Intelligenz wesentlich schneller, kostengünstiger und somit flexibler erstellt und eingeführt werden können.

Mit AutoML befassen sich derzeit drei Arten von Akteuren. So besteht aufseiten der Anbieter von Cloud-Infrastrukturen, wie Google, Microsoft, Amazon und IBM, ein kommerzielles Interesse, da die besonders rechenintensiven AutoML-Ansätze große Cloud-Infrastrukturen erfordern. Den Kunden wird damit ermöglicht, eine eigene künstliche Intelligenz kostengünstig und schnell erstellen zu können. Weitere Akteure sind Unternehmen im Bereich der Datenanalyse, wie Data Robot, DotData und H2O. Diese bieten Software an, um ML-Experten bei der bislang manuellen Durchführung des ML-Prozesses zu unterstützen. Ihre Motivation, sich mit AutoML zu beschäftigen, liegt in der weiteren Verbesserung ihrer erstellten Lösungen. Als dritte Gruppe von Akteuren sind Hochschulen und Universitäten zu nennen, die auf Grundlage ihrer Forschungsaktivitäten nutzbare AutoML-Lösungen bereitstellen und mit der zunehmenden Anzahl ihrer Publikationen zu einer dynamischen Forschungslandschaft beitragen.

Die aktuell verfügbaren AutoML-Lösungen, wie auch die Forschung dazu, sind noch weit davon entfernt, eine vollautomatische Durchführung zu ermöglichen und konzentrieren sich im Wesentlichen auf die Phasen Modellerstellung und Evaluation. In der Phase der Modellerstellung wird aufbauend auf grundlegenden Algorithmen, wie Entscheidungsbäumen, künstlichen neuronalen Netzen usw., das eigentliche ML-Modell erstellt. Hierfür sind auch die sogenannten Hyperparameter zu optimieren, die im Gegensatz zu den herkömmlichen Parametern nicht während des Lernprozesses aus den Daten ermittelt, sondern bereits vor dem Lernen festgelegt werden müssen. Insgesamt ist die Automatisierung der Modellerstellung weit fortgeschritten. Die hierzu verwendeten Ansätze reichen von klassischen Optimierungsalgorithmen über die systematische Suche bis zu evolutionären Algorithmen und finden sich auch in aktuellen AutoML-Lösungen wieder. In der anschließenden Phase der Evaluation erfolgt die Bewertung des Modells auf Grundlage der vorliegenden Daten und der gestellten Aufgabe, um die statistische Signifikanz der Ergebnisse zu ermitteln. Hierfür sind ausreichend viele Lern- bzw. Trainingsprozesse auf dem gesamten Datenbestand durchzuführen. Auch die Automatisierung dieser Phase ist Bestandteil derzeitiger AutoML-Lösungen.

Weit weniger Beachtung findet heute die Phase der Datenerfassung. Aktuelle AutoML-Lösungen erfordern, dass Anwender ihre Daten selbst bereitstellen und sicherstellen, dass diese für die gestellten Aufgaben relevant sind. Sie führen selbst keine oder nur eine sehr beschränkte Erfassung, wie die Suche nach nützlichen Daten im Internet, durch. Auch die Möglichkeiten, Daten synthetisch zu erzeugen sind auf einfache Ansätze, wie das Beschneiden, Spiegeln, Auffüllen und Drehen von Bilddaten oder das Einfügen von Synonymen bei Textdaten, beschränkt. Ebenfalls wenig Beachtung finden die beiden Phasen der Datenaufbereitung und der Merkmalserzeugung. Bei diesen wird Fachwissen bezüglich der Aufgabenstellung nicht mit einbezogen. So können aktuelle AutoML-Lösungen nur allgemeine Merkmale, wie Uhrzeit, Datum oder Adressen, automatisiert extrahieren. Fast gar keine Beachtung in der wissenschaftlichen Literatur finden die Aufgabenformulierung und die Empfehlung. Nutzer bestehender AutoML-Lösungen müssen ihre zu lösenden Problemstellungen erst geeignet aufbereiten und Empfehlungen aus den erhaltenen Ergebnissen selbst ableiten.

Grundsätzlich stellt AutoML einen vielversprechenden Ansatz dar, um ML-Lösungen effizienter, schneller und kostengünstiger zu verwirklichen. Die aktuelle Forschung mit ihren bisher verfügbaren AutoML-Lösungen konzentriert sich allerdings auf einige wenige, leicht zu automatisierende Phasen des ML-Prozesses. Der Mangel an Forschungsaktivitäten in Bezug auf die anderen Phasen verzögert derzeit ihre Weiterentwicklung zu einem vollständig automatisierten ML-Prozess.

Dr. Dirk Thorleuchter