

Intelligenz am Netzwerkrand

KI am Edge sorgt für mehr Datenschutz und Nutzerakzeptanz

Deep Learning hat sich zu einer praktischen Technologie entwickelt, die zahlreichen Anwendungen mit künstlicher Intelligenz zugrunde liegt. Viele dieser Anwendungen werden in der Cloud auf Servern gehostet, aber mehr Datenschutz und Nutzerakzeptanz sind zwei der Gründe, die KI näher an den Netzwerkrand zu rücken.

Autor:Cliff Ortmeyer

Leistungsfähige Cloud-Server sind in vielen Fällen für die Verarbeitung datenreicher Quellen wie Bilder, Videos und Audiodateien zuständig. Diese Server beziehen oft zusätzliche Leistung von Beschleunigungshardware, die von Grafikbearbeitungseinheiten bis hin zu Sonderanfertigungen reicht. Diese sind besonders wichtig für den numerisch intensiven Vorgang, ein neuronales Netzwerk für neue Daten zu trainieren.

Üblicherweise ist der Inferenzprozess, der ein trainiertes Netzwerk nutzt, um neue Daten zu bewerten, viel weniger rechenintensiv als das Training. Darüber hinaus gibt es Arbeitslasten, die weniger datenintensive Quellen verwenden, etwa Sensormessungen von IoT-Geräten, wobei sowohl das Training als auch der Inferenzprozess auf weniger leistungsstarker Hardware durchführbar sind. Systementwickler sind daher der Meinung, dass es

nicht unbedingt nötig ist, Arbeitslasten in der Cloud zu hosten, sobald das KI-Modell trainiert wurde. Trotzdem geschieht dies oft aus geschäftlichen Gründen. Stattdessen lässt sich das trainierte Modell auf eine lokale Maschine übertragen, damit die Verarbeitung näher an der Datenquelle stattfindet.

Vorteile von KI am Netzwerkrand

Es gibt viele Gründe für die Verlagerung von KI-Modellen näher an den Rand des Netzwerks. Zwei Hauptgründe sind Datenschutz und Nutzerakzeptanz. Beispielsweise haben Konsumenten, die Geräte wie Smart Speaker nutzen, mehr und mehr Bedenken in Bezug auf die routinemäßige Aufzeichnung ihrer privaten Gespräche und das Hochladen in Cloud-Dienste für Leistungen, die lokal unterstützt werden könnten. In industriellen Steuerungsanwendungen, bei denen nun

begonnen wird, KI für die Überwachung von Maschinenzuständen oder zur Prozessoptimierung zu nutzen, erfordern Bedenken im Hinblick auf die Vertraulichkeit von Produktionsdaten ebenso, dass so wenig Datenübertragung wie möglich in die Cloud stattfindet. Für viele industrielle Anwendungen gibt es Probleme in Bezug auf die Zuverlässigkeit und die Übertragungsgeschwindigkeit zur Cloud. Viele Systeme, ob direkt in der Fertigung oder an einem entfernten Standort, verfügen nicht über eine Verbindung mit hoher Bandbreite, die erforderlich wäre, um cloudbasierende Inferenzprozesse zu unterstützen. Eine hohe Kommunikationslatenz beeinträchtigt zudem auch Steuerungssysteme. Wenn KI-Modelle in Regelungssystemen zum Einsatz kommen, führt jede Verzögerung beim Abrufen eines Updates aus der Cloud zu Ungenauigkeiten und zur Entwicklung von Insta-

Eck-DATEN

In industriellen Steuerungsanwendungen bestehen häufig Bedenken hinsichtlich der Vertraulichkeit von Produktionsdaten. Daher ist es sinnvoll, so wenig Daten wie möglich in die Cloud zu übertragen und lokal zumindest mit einer Vorverarbeitung der Daten zu beginnen. Der Beitrag erläutert anhand einer KI-Anwendung zur vorausschauenden Wartung, welche Vorteile eine Verlagerung der KI an den Netzwerkrand bietet, welche Optionen es für die Lerntechnologie gibt und stellt eine Lösung vor, die alle Tools umfasst, die für eine rasche Entwicklung und Bereitstellung von KI-fähigen Systemen erforderlich sind.

bilität. Einige Systeme verwenden unter Umständen eine Mischung aus Cloud- und lokaler Verarbeitung. Überwachungskameras beispielsweise erhalten die Netzwerkbandbreite, indem sie unmittelbare Bedrohungen lokal identifizieren, sich aber dann an die Cloud wenden, um zusätzliche Verarbeitungsleistungen in Situationen auszuführen, mit denen die lokalen Modelle nicht zurechtkommen. Auch die Akkulaufzeit steigt deutlich, wenn KI am Rand des Netzwerks platziert ist, da viel weniger Datenversand über das Netzwerk geschieht. Das reduziert sowohl Netzwerk- als auch Cloud-Kosten.

Vorausschauende Wartung

Die vorausschauende Wartung ist derzeit der beliebteste Anwendungsfall in der vernetzten Industrie. Sie bietet einen hohen Investitionsgewinn, da sie die Häufigkeit von Inspektionen vor Ort reduziert. Ebenso lässt sich die Ausfallzeit von Maschinen senken, indem die verbleibende Betriebszeit korrekt berechnet wird und die Lebensdauer so ohne Ausfallrisiko vollständig ausnutzbar ist. Bei der vorausschauenden Wartung können die Sensoren in einer Werkzeugmaschine Temperaturveränderungen feststellen, die mit gestei-

gerten Geräuschen oder Vibrationen einhergehen. Die Erkennung geschieht durch eine Kombination von Mikrofonen und Beschleunigungsmessern und deutet auf ein potenzielles Problem hin. Bei traditionellen Algorithmus-Techniken kann es äußerst schwierig sein, Kombinationen verschiedener Echtzeit-Sensormessungen bestimmten Problemen zuzuschreiben. KI-basierende Modelle können Zeitreihendaten zusammen mit Echtzeiteingaben interpretieren, um die Gesundheit des Systems akkurat zu bestimmen. Derartige Modelle können Technologien wie die Sensorfusion nutzen, um festzustellen, ob die Situation eine Wartung erfordert oder nicht. Die Komplexitäts- und Latenzanforderungen des Modells bestimmen, ob das Modell lokal ausführbar ist.

Ansätze für lokale Verarbeitung

Es gibt zwei Ansätze für die lokale Verarbeitung. Der erste besteht darin, die Verarbeitungsleistung, die am Gerät selbst vorhanden ist, zu nutzen. Ob dies möglich ist, hängt von der Komplexität des KI-Modells und der verfügbaren Verarbeitungsleistung ab. Für leistungsarme Sensorknoten ist es nicht möglich, große Deep-Learning-Modelle auch nur für Inferenzprozesse auszuführen. Da Algorithmen des maschinellen Lernens allerdings viele Formen annehmen können, ist es unter Umständen nicht notwendig, eine derart gewichtige Aufgabe vollständig lokal auszuführen.

Die zweite Option ist, die Verarbeitung entweder vollständig oder teilweise auf ein anderes Gerät auszulagern. Das Gerät kann beispielsweise ein vereinfachtes KI-Modell ausführen, das eine erste Analyse der Daten durchführt. Im Fall eines Überwachungssystems, das Aktivität erkennt, kann dieses Modell beispielsweise die Aufgabe übernehmen, festzustellen, ob es sich bei dem Signal um Hintergrundgeräusche wie Wind handelt oder um einen Eindringling beziehungsweise das Geräusch eines zerbrechenden Fensters. Komplexere Funktionen lassen sich auf ein Gateway übertragen, das Daten aus mehreren Kameras und anderen Security-Geräten vereint. Das Gerät kann eine gewisse Vorverarbeitung durchführen, um die Arbeit des Gateway-Modells zu erleichtern. Leistungsstarke Hardware wie die

**ENTSCHLÜSSELN
SIE DAS
AUSGANGSSIGNAL
IHRER ROGOWSKI-SPULE**

**CIRED
2019
Halle 9, Stand D19A**

AI-PMUL-Serie

Sie suchen einen vielseitigen Digital-Signalkonditionierer für das Sekundärsignal von Rogowski-Spulen mit der Möglichkeit von verschiedenen Analogausgängen? Zusammen mit den Rogowski-Spulen ART & ARU von LEM bietet der AI-PMUL bis 5000 A eine universelle Mess- und Überwachungslösung mit hoher Messgenauigkeit.

- Linearitätsfehler unter 0,1 %
- Einfache Auswahl zwischen 12 Strombereichen, 6 Rogowski-Spulen-Empfindlichkeiten und 6 möglichen Ausgängen (Effektiv- und Momentanwerte, Strom und Spannung)
- Erfüllt die Anforderungen von MS/NS-Anlagen, Leistungsmessung, Submetering in Gebäuden und Überwachung der Netzqualität
- DIN-Schienenmontage

www.lem.com

At the heart of power electronics.

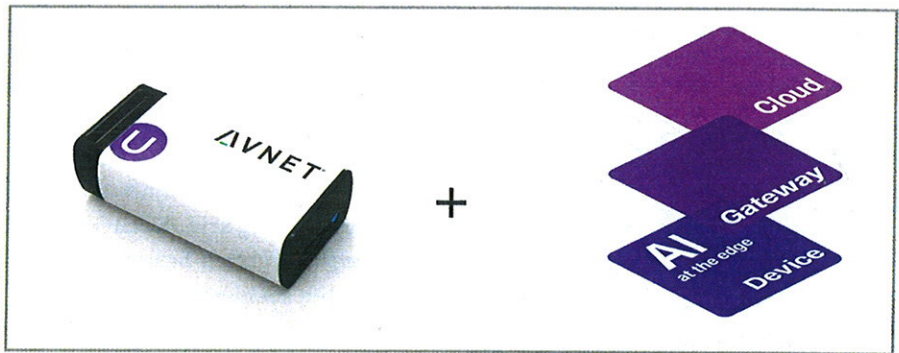
 **LEM**[®]

Xilinx-Geräte mit programmierbarer Logik, die mit der Ultra96-Plattform von Avnet im Angebot sind, können Inferenzprozesse basierend auf fortgeschrittenen Deep-Learning-Modellen und anderen komplexen Algorithmen des maschinellen Lernens ausführen. Das lokale Gateway kann das Modell updaten, indem Daten gesammelt und das Training schubweise ausgeführt wird, wenn die Inferenzarbeitslast gering ist. Alternativ ist das Gateway in der Lage, Daten zu erheben, die wichtig für das erneute Training sind, und diese täglich, wöchentlich oder monatlich auf einen Cloud-Server zu übertragen.

KI am Gerät bietet die geringste Kommunikationslatenz. Die gesteigerte Verarbeitungsgeschwindigkeit eines lokalen Gateways kann jedoch unter Umständen den Geräteprozessor mit schwächerer Leistung überholen und die besten Latenz- und Durchflussparameter bieten.

Optionen für Lerntechnologie

Der Entwickler hat je nach spezifischer Anwendung eine Auswahl an verschiedenen Modelltypen. Ein Random Forrest ist ein Beispiel für ein überwachtes Lernsystem. Es setzt Daten, die von den Systementwicklern zur Verfügung gestellt werden, mit den Inputs in Beziehung, woraus das maschinelle Lernsystem Verbindungen lernt. Deep Learning ist ein weiteres Beispiel für eine überwachte Lerntechnologie, da es Trainingsdaten verwendet, die zuvor mit Labels gekennzeichnet wurden. In einem Bildklassifikationssystem beispielsweise nimmt das Modell Daten aus dem Bild auf und wendet sie auf eine Schicht simulierter Neuronen an. Die Resultate aus



Smart-Edge-Agile-Geräte von Avnet enthalten eine von Octonion entwickelte Softwarelösung. Sie erlauben es Entwicklern, KI-Lösungen mit umfassender Edge-Verarbeitung zu entwickeln.

der ersten Neuronenschicht gibt das System sukzessive durch viele Schichten weiter. Einige kombinieren die Outputs aus mehreren Neuronen in der vorhergehenden Schicht, um einen einzigen Wert zu erstellen, der an die nächste Schicht übergeben wird. Auf diese Art führen tiefe, neuronale Netzwerke eine Dimensionsreduktion aus: ein wichtiger Schritt bei der Konvertierung von komplexen, multidimensionalen Daten wie Bildern oder Audiodateien in eine Form, die für eine finale Klassifizierung zum Einsatz kommt.

Das neuronale Netzwerk kann eine Klassifizierung ausführen, da es lernen kann, wie verschiedene Arrangements von Koeffizienten in den neuronalen Schichten auf unterschiedliche Bilder reagieren. Daraus erzielt es Resultate, die den Trainingslabels entsprechen. Das Training eines maschinellen Lernmodells muss nicht ausschließlich auf mit Labels gekennzeichneten Daten aufbauen. Unüberwachte Algorithmen maschinellen Lernens wie Clustering können Muster in Daten ohne zusätzliche Hilfe finden. Ein derartiger Prozess kann in industriellen Steuerungs-

systemen sehr hilfreich sein, wo mehrere Sensoren zum Einsatz kommen oder das Zeitreihenverhalten der Inputs wichtig ist. Im Fall der Zustandsüberwachung einer Werkzeugmaschine deutet die Größenordnung der Vibration nicht unbedingt auf ein Problem hin – es kann sich dabei auch einfach um eine Konsequenz des Prozesses handeln. Allerdings kann ein Bewegungsmuster in den Zeitreihendaten in Verbindung mit schnellen Temperaturveränderungen auf ein Problem hindeuten, das einer Wartung bedarf. Diese Unterscheidungen lassen sich treffen, indem eine Aufteilung der Daten in Cluster stattfindet, die einfach zu unterscheiden sind, wenn die Quelldaten bei direkter Nutzung kein klares Muster erkennen lassen.

Verlagerung an den Netzwerkrand

Ein wichtiger Faktor beim maschinellen Lernen als Technologie ist, dass eine Implementierung oft fundiertes Wissen und Erfahrung erfordert, um das Optimum aus den verschiedenen verfügbaren Formen der KI herauszuholen. Octonion hat für dieses Problem eine Softwarelösung entwickelt, die in Smart-Edge-Agile-Geräte von Avnet integriert ist. Sie erlaubt es Entwicklern, KI-Lösungen für IoT-Systeme zu entwickeln, die Edge-Verarbeitung umfassen, ohne teure Experten zuzuziehen oder sich selbst umfassend einarbeiten zu müssen. Die Lösung läuft auf drei Ebenen: Gerät, Gateway und Cloud. Die Gerätehardware Avnet Smart Edge Agile ist ein autonomer Sensorknoten mit geringer Leistung. Das Gerät bietet Zugang zu verschiedenen Sensortypen, die Beschleunigungsmesser, Gyroskope, Magnetometer und zusätzliche Sensoren umfassen. Ein Mikrofon dient dazu, Audio-Inputs aufzuzeichnen. Die Smart-



Die Vorteile von KI am Netzwerkrand liegen vor allem in der Datensicherheit und der höheren Nutzerakzeptanz. Die Sicherheit lässt sich erhöhen, indem eine lokale Vorverarbeitung der Daten stattfindet und so wenig Daten wie möglich an die Cloud übertragen werden.

Sensor-Hardware kommuniziert mit einem lokalen Gateway über BLE. Versionen, die andere IoT-orientierte Netzwerktechnologien wie LoRaWAN, Sigfox, Mobilfunk und WLAN unterstützen, befinden sich in Entwicklung. Das Gateway-Modul lässt sich in Form eines Android- oder iOS-Smartgerätes nutzen. Die Gateway-Software kann auf einem Raspberry-Pi oder einem ähnlichen Linux-Plattformgerät laufen, das Zugang zur Cloud bietet. Cloudschicht lässt sich auf AWS, Microsoft Azure oder individuellen Serverlösungen bereitstellen.

Die Koordination der verschiedenen Schichten erfolgt durch die Brainium-Softwareumgebung von Octonion, die Edge-to-Cloud-Unterstützung in einer Umgebung bietet, die keine Programmierung erfordert. Sicherheit ist bei allen IoT-Bereitstellungen von unermesslicher Bedeutung und ein wichtiges Element des Systems, um den Schutz der Daten zu gewährleisten, die während des Inferenzprozesses und zum Trainieren von Modellen Verwendung finden. Das für Brainium

entwickelte System nutzt AES-Verschlüsselung zusammen mit einem Hardware-basierenden manipulationssicheren Speicher für kryptographische Schlüssel. Die von den Geräten verwendeten Firmware-Bilder zur Ausführung von KI-Funktionen sind ebenso mit digitaler Signaturvalidierung verschlüsselt. Wenn Daten gesendet werden, setzen Kommunikationskanäle TLS-Verschlüsselung ein, um Nachrichten auf dem Weg vom Netzwerkrand in die Cloud vor dem Ausspionieren durch Kontrahenten zu schützen.

Die Softwareumgebung bietet eine Kombination aus unüberwachtem und überwachtem maschinellen Lernen. Beispielsweise lässt sie sich auf Anomalien in der Zeitreihenrepräsentation von rohen Sensordaten trainieren. Zusätzlich werden Daten aus verschiedenen Szenarien geclustert, um gemeinsame Datenmuster zu identifizieren. Diese Datenmuster lassen sich in einen flexiblen Modell-Builder innerhalb der KI-Studio-Software ziehen, sodass Entwickler das Modell anpassen

können. Während des Lernvorgangs sammelt das Gerät Datenproben, sichert und verschlüsselt diese, bevor die Übertragung in die Software stattfindet, die in der Cloud läuft. Diese Software lernt aus den Proben und erstellt KI-Modelle, die sie für Inferenzprozesse zurück an das Edge-Gerät überträgt. Jedes Gerät innerhalb der Umgebung des Kunden erhält dieses KI-Modell und führt in Standalone-Funktion Überwachung und Analysen aus. Die verbleibende Umgebung ist vom Prototyp bis hin zur vollständigen Produktion einheitlich. Für die Bereitstellung können Kunden Standard-Smart-Edge-Agile-Hardware von Avnet nutzen oder das Design ihrer eigenen Umsetzung anpassen. (na) ■

Autor

Cliff Ortmeier

Global Head of Technical Marketing and Solutions Development bei Premier Farnell

all-electronics.de 

infoDIREKT

802ei0619



Grenzen Sie Ihre Anwendung von anderen ab

Passen Sie Ihr Embedded-Design mit flexiblen, integrierten Building-Blöcken individuell an

Microchip weiß, dass Ihre Designanforderungen einzigartig sind. Deshalb bieten unsere Produkte robuste, integrierte Peripherie für eine Vielzahl von Anwendungen:

- Fügen Sie wirkungsvolle Displays und Touchscreens für eine intuitive Benutzerinteraktion hinzu
- Vernetzen Sie Ihre Anwendung weltweit – mit und ohne Kabel
- Regeln Sie Ihren Antrieb oder Ihre Leistungswandlung
- Schützen Sie die Daten Ihrer Anwendung

Passen Sie Ihr Produkt mit integrierter Peripherie an und sparen Sie Entwicklungskosten und Zeit.

Gestalten Sie Ihr Design unter www.microchip.com/FlexibleFunctions

