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

Neurosymbolische künstliche Intelligenz

Die aktuellen Fortschritte in der Künstlichen Intelligenz (KI), z. B. bei der Erkennung von Objekten auf Bildern oder der Verarbeitung von natürlicher Sprache, sind ganz wesentlich auf der Basis von künstlichen neuronalen Netzwerken (KNN) erreicht worden. Derzeitige KI-Systeme weisen allerdings noch einige Einschränkungen im Vergleich zu den Fähigkeiten eines Menschen auf. So benötigen sie u. a. sehr viele Trainingsdaten, um neue Sachverhalte zu erlernen, und ihre Ergebnisse sind häufig nicht nachvollziehbar bzw. erklärbar. Weitere Fortschritte bei KI-Systemen erwartet man in Zukunft durch die Kombination der KNN-basierten Verfahren mit Methoden aus dem Bereich der sogenannten symbolischen KI. Diese beschäftigt sich mit der expliziten Darstellung und regelbasierten Verarbeitung von Wissen, z. B. auf der Basis von formaler Logik. Zusammen entsteht dann das Gebiet der neurosymbolischen KI, welche die jeweiligen Stärken der beiden unterschiedlichen KI-Ansätze miteinander kombinieren soll. Diese Stärken liegen beispielsweise bei KNN auf dem Gebiet des Lernens anhand von Trainingsdaten und bei symbolischer KI im Ziehen von Schlussfolgerungen aus vorhandenem Wissen.

KNN-basierte KI und symbolische KI stellen generell die zwei wesentlichen Ansätze im KI-Bereich dar. KNN bestehen aus einem Netzwerk aus künstlichen Neuronen, die in unterschiedlichen Schichten angeordnet sind, wobei jeder Verbindung zwischen den einzelnen Neuronen jeweils ein einstellbarer Parameter zugeordnet ist. KNN sind ein Ansatz im Bereich des maschinellen Lernens, das es Computern ermöglicht, selbstständig anhand von Trainingsdaten bestimmte Sachverhalte zu erlernen, z. B. welches Objekt auf einem Bild abgebildet ist. Während der Trainingsphase eines KNN werden dabei die Parameter des Netzwerks im Hinblick auf die Trainingsdaten optimiert. Die aktuellen Fortschritte im KI-Bereich sind ganz wesentlich durch den Einsatz von KNN erreicht worden, die über viele Schichten von Neuronen und eine sehr große Anzahl von Parametern verfügen. Dieser Ansatz wird als Deep Learning bezeichnet. Das mithilfe von Deep Learning erlernte Wissen liegt nicht in Form eines explizit geschriebenen Programms vor, bei dem die Ergebnisse des KI-Systems mithilfe der relevanten Programmschritte vergleichsweise einfach nachvollzogen werden können. Stattdessen ist es in den Parametern des KNN enthalten und daher aufgrund ihrer großen Anzahl typischerweise nicht verständlich. Deshalb wird bei Deep Learning vielfach auch von einer Black Box gesprochen.

Im Rahmen der symbolischen KI soll menschliche Intelligenz nachgebildet werden, indem Symbole verwendet werden, die Dinge in der Welt repräsentieren. Der Denkprozess basiert hier darauf, dass diese Symbole nach bestimmten Regeln verarbeitet werden. Beispielsweise lässt sich so neues Wissen mithilfe von logischen Schlussfolgerungen aus bereits bekanntem Wissen ableiten. Ein klassisches Beispiel für eine solche logische Schlussfolgerung besteht darin, dass aus den Aussagen „alle Menschen sind sterblich“ und „Sokrates ist ein Mensch“ folgt, dass Sokrates ebenfalls sterblich ist. In diesem Beispiel stellen die Wörter „alle“, „Menschen“ usw. die entsprechenden Symbole dar. Ein Vorteil von symbolischer KI liegt darin, dass ihre Ergebnisse erklärbar sind, da der betreffende Verarbeitungsprozess auf der Grundlage von Regeln erfolgt und für einen Menschen im Prinzip nachvollziehbar ist.

Ein wichtiges Anwendungsgebiet von neurosymbolischer KI wird vermutlich im Bereich der Verarbeitung von natürlicher Sprache liegen. Hier könnte z. B. das Verständnis von natürlicher Sprache mithilfe von logischen Schlussfolgerungen in Verbindung mit Alltagswissen gesteigert werden. Auf dem Gebiet der Bildverarbeitung könnte die Nutzung von neurosym-bolischen KI-Systemen u. a. ein verbessertes Verständnis von Bildern durch geeignetes Hintergrundwissen erlauben. In der Robotik könnte mit Hilfe von neurosymbolischer KI z. B. sichergestellt werden, dass autonome Fahrzeuge zur Unfallvermeidung bestimmte Sicherheitsvorgaben einhalten, indem gewissermaßen der symbolische Teil des KI-Systems den KNN-basierten Teil überwacht.

Im medizinischen Bereich wäre beispielsweise die potenziell verbesserte Erklärbarkeit der Ergebnisse von neurosymbolischen KI-Systemen von großem Interesse. Dies gilt ebenso für einen Einsatz derartiger Systeme in der Wissenschaft, etwa zur Vorhersage von chemischen Reaktionen. Im Rahmen der Unterstützung von Programmierern bei der Softwareentwicklung könnten neurosymbolische KI-Systeme im Vergleich zu aktuellen KNN-basierten KI-Systemen Vorteile hinsichtlich der Einhaltung der formalen Regeln bei der Programmierung von Computern besitzen. Das Gebiet der IT-Sicherheit könnte u. a. von einer potenziell höheren Robustheit von neurosymbolischen KI-Systemen gegenüber sogenannten Adversarial Examples profitieren. Hierunter werden Eingabedaten für KI-Systeme verstanden, die von einem Angreifer speziell mit der Absicht entworfen wurden, Fehler bei den betreffenden KI-Systemen hervorzurufen, wie z. B. eine fehlerhafte Erkennung von Objekten auf Bildern.

Neurosymbolische KI befindet sich zurzeit noch in einem relativ frühen Entwicklungsstadium. Gegenwärtig werden verschiedene Ansätze zur Verwirklichung von neurosymbolischen KI-Systemen verfolgt. In diesem Zusammenhang gibt es jedoch noch keinen dominierenden Ansatz. Unter Fachleuten besteht allerdings Einigkeit darüber, dass das grundsätzliche Ziel bei neurosymbolischen KI-Systemen darin bestehen sollte, Konzepte auf einer hohen Abstraktionsebene zu verarbeiten und fundierte Schlussfolgerungen zu ziehen. Die verschiedenen aktuellen Ansätze auf diesem Gebiet unterscheiden sich u. a. dadurch, wie eng der symbolische Teil des KI-Systems und der KNN-basierte Teil miteinander gekoppelt sind. Die Stärken von existierenden neurosymbolischen KI-Systemen liegen dabei typischerweise im Bereich des Lernens oder im Ziehen von Schlussfolgerungen, aber nicht in beiden Bereichen gleichzeitig.

Dr. Klaus Ruhlig