NACHRICHTEN AG

02. August 2024

Neue Geometrie-Methoden erhöhen Sicherheit von KI-Systemen in Würzburg

Wissenschaftler der Universitäten Würzburg und München erforschen geometrische Methoden zur Verbesserung der Robustheit neuronaler Netze.

Die Fortschritte im Bereich der künstlichen Intelligenz und des maschinellen Lernens sind in den letzten Jahren unübersehbar geworden. Technologien, die autonomes Fahren ermöglichen oder medizinische Diagnosen unterstützen, sind längst keine Zukunftsmusik mehr. Dennoch stehen diese Systeme vor einem massiven Problem: ihre Anfälligkeit gegenüber Manipulationen. Ein neues Forschungsprojekt des Verbunds zwischen der Julius-Maximilians-Universität Würzburg und der Technischen Universität München zielt darauf ab, diese Schwachstellen anzugehen und die Robustheit neuronaler Netze zu verbessern.

Die Bedeutung der Forschung für die Gesellschaft

Das bevorstehende Projekt, das mit 565.000 Euro von der Deutschen Forschungsgemeinschaft (DFG) gefördert wird, behandelt ein hochaktuelles Thema. Die Anfälligkeit von KI-Systemen für gezielte Attacken kann gravierende Auswirkungen auf die Sicherheit in verschiedenen Bereichen haben, von der Verkehrssicherheit bis hin zur medizinischen Diagnostik. Die Wissenschaftler Leon Bungert und Dr. Leo Schwinn setzen sich deshalb dafür ein, die Zuverlässigkeit dieser Systeme zu erhöhen und dadurch das Vertrauen der Gesellschaft in innovative Technologien zu stärken.

Das Projekt GeoMAR: Geometrische Methoden im Einsatz

Das Forschungsprojekt trägt den Titel „GeoMAR: Geometric Methods for Adversarial Robustness“. Dabei geht es darum, neuronale Netze so zu trainieren, dass sie weniger anfällig für feindliche Angriffe sind. Ein zentrales Element ist das Verständnis der sogenannten Entscheidungsgrenze, die ein Netzwerk zieht, um Objekte zu klassifizieren. Diese Grenze ist oft der kritische Punkt, an dem Angriffe ansetzen können. Die Mathematiker setzen geometrische Methoden ein, um diese Grenzen besser zu definieren und damit zu schützen.

Robust gegen Angriffe: Ein spielerischer Ansatz

Um neuronale Netze robuster zu machen, verfolgen Bungert und Schwinn einen innovativen Ansatz: Sie wollen die Netzwerke während des Trainings absichtlich mit fehlerhaften Daten konfrontieren, um sie auf reale Angriffsszenarien vorzubereiten. Hierbei wird ein zweites Netzwerk trainiert, das als potenzieller Angreifer fungiert. Dieser Ansatz könnte dazu führen, dass die Systeme kreativer in ihrer Verteidigung werden und sich besser an unerwartete Manipulationen anpassen können.

Die Herausforderungen der Genauigkeit

Eine der zentralen Herausforderungen dieses Projekts ist die Balance zwischen Robustheit und Genauigkeit. Bisher haben robustere Systeme in der Regel an Genauigkeit eingebüßt. Das Ziel der Forschung ist es, neue mathematische Ansätze zu finden, die es ermöglichen, die Werte für Robustheit und Genauigkeit zu optimieren, sodass kein Verlust an Verlässlichkeit auftritt.

Ein Beispiel from der Praxis

Um die Gefahren von Manipulationen zu verdeutlichen, führt Bungert ein einfaches Beispiel an: Ein Bild von einer Geige wird durch minimal hinzugefügtes Rauschen von der Software fälschlicherweise als Seelöwe identifiziert. Diese Verwechslung mag trivial erscheinen, aber im Straßenverkehr oder in der Medizin könnten solche Fehler fatale Folgen haben. Ein Fehler im autonomen Fahren, bei dem ein Fußgänger übersehen wird, oder im medizinischen Bereich, wo ein Tumor fälschlicherweise nicht erkannt wird, sind potenzielle Risiken, deren Vermeidung oberste Priorität hat.

Fazit: Die Bedeutung von geometrischen Methoden

Die Forschung, die Leon Bungert und Dr. Leo Schwinn durchführen, könnte einen bedeutenden Schritt in Richtung sichererer KI-Systeme darstellen. Durch den Einsatz geometrischer Methoden wollen sie neuronale Netze nicht nur widerstandsfähiger machen, sondern auch unsere Abhängigkeit von diesen Technologien sicherer gestalten. Der Erfolg dieses Projekts könnte weitreichende Auswirkungen auf die Entwicklung und den Einsatz von KI-Anwendungen in verschiedenen Lebensbereichen haben. Das Ziel ist klar: Ein robustes KI-System, das auch in kritischen Situationen zuverlässig bleibt.

Für wissenschaftliche Informationen:
Prof. Dr. Leon Bungert, Professur für Mathematik III (Mathematik des Maschinellen Lernens),
Tel: +49 931 31-82849,
E-Mail: leon.bungert@uni-wuerzburg.de

– **NAG**

Details