zurück zum Artikel

Technology Review
DAS MAGAZIN FÜR INNOVATION

"Sollten Roboter lügen, Herr Arkin?"

Der Robotiker Ronald Arkin entwickelt Maschinen, die täuschen und betrügen können. Und hält das trotzdem für einen Fortschritt.

Der Roboter fährt langsam vorwärts, stoppt, setzt ein kleines Stück weiter vor, scheint zu zögern. Dann wendet er sich nach links, stößt einen im Weg stehenden Holzkegel um, dreht sich dann jedoch zurück und fährt geradeaus, hinter einen Pappschirm. Ein zweiter Roboter folgt. Er passiert den umgestürzten Kegel, wendet sich stattdessen aber nach links und fährt hinter den linken Pappschirm.

Was ein wenig bizarr wirkt, ist ein Experiment in Sachen Täuschung und Lüge. Denn der erste Roboter hatte den zweiten planvoll und absichtlich in die Irre geführt. Alan Wagner und Ronald Arkin vom Georgia Institute of Technology, die das Experiment 2010 veröffentlichten, wollten zeigen, dass auch Maschinen dazu in der Lage sind.

Die Grundlagen dafür fanden die Wissenschaftler in der mathematischen Spieltheorie. Ein Roboter berechnet aus den Reaktionen seines Partners auf seine eigenen Aktionen sogenannte "Interaktionsmatrizen" – ein mathematisches Modell des Spiels mit allen möglichen Verläufen. Der suchende Roboter lernte, aus den umgefallenen Kegeln auf das Versteck zu schließen. Im Gegenzug lernte der Versteck-Roboter Strategien, die dem Fänger die Arbeit erschweren. Er begann, absichtlich falsche Spuren zu legen.

Technology Review: Professor Arkin, Sie haben zwei Roboter gebaut, die miteinander Verstecken spielen. Wie sind Sie überhaupt auf diese Idee gekommen?

Ronald Arkin: Der Ausgangspunkt waren Forschungsarbeiten zum Vertrauensverhältnis zwischen Mensch und Maschine. Die meisten Forscher haben bislang untersucht, wann ein Mensch einer Maschine trauen kann. Im Unterschied dazu haben wir uns gefragt: Unter welchen Umständen sollte ein Roboter einem Menschen trauen?

TR: Das klingt für mich sehr akademisch. Warum ist das wichtig?

Arkin: Denken Sie an den Anschlag vom 11. September 2001. Wir wissen, dass die Flugzeugsysteme Alarm gegeben haben, als die Maschine direkt auf die Hochhäuser zusteuerte. Aber wir erlauben auch in solchen Fällen einer Maschine nicht, die Kontrolle von den menschlichen Piloten zu übernehmen. Das hat uns zu der Frage gebracht: Können wir Modelle entwickeln, die dem Roboter erlauben zu verstehen, was der Mensch sich denkt, wenn er bestimmte Befehle gibt? Danach war es nur konsequent, sich der Täuschung zuzuwenden. Lüge, Betrug – das ist die Kehrseite des Vertrauens. Jeder gute Betrüger weiß, dass er zunächst ein Vertrauensverhältnis zu seinem Opfer aufbauen muss, bevor er jemanden übers Ohr hauen kann. In gewisser Weise war das also die Fortsetzung früherer Forschungsarbeiten. Das hat zu anderen Arbeiten geführt. Wir haben zum Beispiel biologische Modelle der Täuschung getestet. Betrug ist weit verbreitet in der Natur.

2012 untersuchten die Forscher um Arkin eine verfeinerte Version des Versteckspiels, die sie sich bei Eichhörnchen abgeschaut hatten. Die Tiere schützen ihre Nussvorräte vor anderen Eichhörnchen, indem sie die Konkurrenten zu leeren Scheinverstecken lotsen. Arkin und sein Team konnten zunächst in Simulationen zeigen, dass sich diese Strategie auch für Roboter umsetzen ließ. In einem realen Experiment ließen sie zwei autonome Roboter gegeneinander antreten. Der erste, der "Sammler", musste farbige Pappscheiben einsammeln und sie zu "Verstecken" bringen, die durch blaue Eimer markiert waren. Hatte er genug "Futter" gesammelt, schaltete das Steuerprogramm um. Der Roboter patrouillierte zwischen den Verstecken, um seine Vorräte zu kontrollieren. Der "Räuber" irrte zunächst zufällig auf dem Spielfeld herum, bis er den Sammler entdeckte, dem er dann zu den Verstecken folgte. Entdeckte der Sammler wiederum, dass ihm Vorräte gestohlen wurden, steuerte er nur noch Scheinverstecke an. Fand der Räuber nach einiger Zeit keine Beute, machte er sich auf die Suche nach einem neuen Opfer. In dem Paper räumen die Forscher allerdings ein, dass das Experiment stark vereinfacht ist – in weiteren Experimenten wollen sie eine größere Gruppe von Robotern einbeziehen, die den Ort ihrer Verstecke frei wählen können.

TR: Waren Sie denn die ersten, die sich mit diesen Fragen beschäftigt haben?

Arkin: Ja und nein. Nehmen Sie Alan Turing. Sein berühmter Test, bei dem eine Maschine dann als intelligent gilt, wenn sie so tun kann, als ob sie ein Mensch sei, beruht im Kern auf einer Täuschung. Das Konzept der Täuschung ist in gewisser Weise von Anfang an Teil der Forschung zu künstlicher Intelligenz gewesen. Aber nur wenige Wissenschaftler haben sich bislang mit dem physischen Aspekt beschäftigt, also mit Robotern, die lügen. Einer der ersten war mein Schweizer Kollege Dario Floreano. Er hat sich mit der evolutionären Entstehung der Täuschung in Roboterschwärmen befasst.

Wissenschaftler um Dario Floreano von der Eidgenössischen Technischen Hochschule in Lausanne ließen 2009 eine Gruppe von Robotern in einem Testareal auf die Suche nach "Nahrung" gehen. Fand ein Roboter ein Nahrungsfeld, wurde ihm pro Spielrunde ein Punkt gutgeschrieben. Abzüge gab es hingegen dafür, sich an eine Stelle im Raum zu begeben, die die Forscher als "Gift" gekennzeichnet hatten. Die erfolgreichsten Roboter durften ihre Steuerungsalgorithmen den anderen weitervererben – angereichert mit einigen zufälligen Änderungen. Die kleinen Maschinen besaßen zudem Lämpchen. Zunächst leuchteten sie zufällig, deshalb war es an Futterstellen, an denen sich viele von ihnen sammelten, hell. Weil die Roboter Licht auf kurze Entfernung auch sehen konnten, lernten sie schnell, nach Helligkeit zu suchen, um Futter zu finden. Weil der Platz an den Futterstellen jedoch nie für alle Roboter ausreichte, blinkten einige Maschinen schon nach 50 Generationen nicht mehr, wenn sie Futter gefunden hatten.

TR: Wie erfolgreich ist die Täuschung?

Arkin: Das kommt darauf an. In Computersimulationen haben wir beispielsweise die Strategie von Antilopen untersucht, die auf der Flucht vor Raubtieren sind. Manche tun so, als seien sie sehr fit. Sie springen sehr hoch und laufen sehr schnell, um dem Löwen zu signalisieren: Du kannst mich nicht fangen. Ich bin zu schnell für dich. Obwohl sie dieses Tempo vielleicht gar nicht lange durchhalten würden. In einigen Fällen ist das tatsächlich erfolgreich. Der Löwe versucht gar nicht erst, die Antilope zu fangen. Der Punkt ist: Wenn alle die Strategie versuchen, funktioniert der Bluff nicht mehr. Die Frage ist also: Wo ist der sweet spot? Wie viele Antilopen können bluffen, ohne den evolutionären Mechanismus zu zerstören? Die Antwort hängt ganz von den Ausgangsbedingungen ab: Von der Fitness des Raubtiers, dem Zustand der Beute und dem Verhalten der anderen Herdentiere.

TR: Wenn das Roboter könnten...

Arkin: Schon der chinesische Gelehrte Sun Tsu hat in seinem Buch "Die Kunst des Krieges" geschrieben, dass alle

Kriegsführung im Wesentlichen auf Täuschung beruht. Wenn ein militärischer Roboter beispielsweise nur noch wenig Energie hat und die Gefahr besteht, dass er gefangen oder zerstört wird, was sollte er tun? Könnte er einen anderen, feindlichen Roboter davon überzeugen, dass er stärker und gefährlicher ist, als das in Wirklichkeit der Fall ist? Die Arbeit daran förderte unter anderem das Office of Naval Research der Marine. Aber es ist reine Grundlagenforschung, die nicht unter Geheimhaltung fällt.

TR: Können wir bald also nicht einmal mehr unseren Maschinen trauen?

Arkin: Das ist eine gute Frage. In all meinen Veröffentlichungen bemühe ich mich stets, eine Diskussion über diese Frage anzustoßen. Ich möchte aber daran erinnern, dass dieses Problem natürlich auch bei zwischenmenschlicher Kommunikation auftritt. Und schließlich kann die Täuschung für den Getäuschten durchaus positiv sein.

Zumindest halten Menschen, die einem betrügerischen Roboter gegenübersitzen, diesen nicht nur für intelligenter, sondern auch für interessanter. Brian Scassellati und seine Mitarbeiter von der Yale University entwickelten 2010 eine Maschine, die beim Stein-Schere-Papier-Spiel gelegentlich betrügt. "Nico" spielte gegen einen Menschen. Wenn er beim Vergleichen der Handgesten feststellt, dass er eigentlich verloren hätte, wechselt er seine Geste nachträglich. Wahlweise verkündet er auch einfach fälschlicherweise: "Ich habe gewonnen!" Mithilfe des betrügerischen Roboters wollten die Forscher herausfinden, wie Menschen solch eine Maschine beurteilen. Tatsächlich gab eine große Mehrheit der Testspieler an, den Roboter im betrügerischen Modus am intelligentesten und interessantesten gefunden zu haben.

TR: Dennoch ist es riskant, Roboter lügen zu lassen, oder?

Arkin: Die Frage ist nicht, ob Roboter die Fähigkeit besitzen sollten, andere zu täuschen. Denn dieses Verhalten ist in der Natur ohnehin universell. Die Frage ist: Wann ist Täuschung akzeptabel? Das ist ein Punkt, über den man diskutieren kann. Nehmen Sie Kant. Er hat, basierend auf seinem kategorischen Imperativ, gezeigt, dass es unethisch ist zu lügen. Auf der anderen Seite haben wir andere philosophische Schulen wie den Utilitarismus. Die betonen, dass eine Handlung, die das allgemeine Wohl vermehrt, gebilligt werden kann, auch wenn sie gegen moralische Grundsätze verstößt.

TR: Und was denken Sie? Sollte ein Computer lügen?

Arkin: Im militärischen Kontext gibt es im Großen und Ganzen mehr Situationen, in denen das Verhalten angebracht ist.

Aber auch hier existieren natürlich Grenzen. Sich tot zu stellen beispielsweise oder Kampfunfähigkeit vorzutäuschen, um den Gegner dann anzugreifen, verstößt gegen die Genfer Konventionen. Ich habe daher keine allgemein gültige Antwort auf Ihre Frage. Deshalb rate ich, besonders in militärischen Situationen, immer dazu, vorsichtig zu sein. In Forschungsprojekten zu bewaffneten Robotern haben wir Methoden entwickelt, mit denen wir die Einhaltung ethischer Prinzipien technisch erzwingen können. Ein Software-Modul beurteilt die Aktionen nach ethischen Gesichtspunkten. Im Zweifelsfall deaktiviert es die Waffen. Ich gehe davon aus, dass man so etwas auch auf Täuschung anwenden kann.

TR: Was machen Sie denn, wenn Ihnen die Entscheidung des Roboters nicht gefällt?

Arkin: Bei den Kriegsrobotern beispielsweise können Menschen die Deaktivierung der Waffen von Hand umgehen. Allerdings nur, wenn zwei Menschen unabhängig voneinander so entscheiden. Die Aktion wird zudem aufgezeichnet und anschließend juristisch überprüft. Wir hoffen, dass diese Sicherung Menschen davon abhält, verbrecherisch zu handeln.

TR: Und wenn der Roboter nur so tut, als hätte er die Waffen deaktiviert?

Arkin: Wir wollen sicherlich keine Situation wie in dem Film "2001", wo der Computer Hal die Besatzung aus dem Raumschiffs gelockt hat und sich dann weigert, die Luftschleuse wieder zu öffnen. Wie wir mit dieser Ambivalenz umgehen, ist noch nicht gelöst. Sie beschäftigt die Leute. Auf der einen Seite ist unsere Arbeit mit dem sich versteckenden Roboter 2010 vom "Time Magazine" als eine der 50 bedeutendsten Innovationen des Jahres ausgezeichnet worden. Es hat aber auch sehr viele geradezu hysterische Artikel in den Medien gegeben, nach denen das Ende der Welt bevorsteht, wenn wir Robotern erlauben würden, über Tricks und Täuschung auch nur nachzudenken.

TR: Was durchaus verständlich ist...

Arkin: Wir müssen aufmerksam gegenüber potenziell dystopischen Entwicklungen sein. Aber ein guter Teil der Angst, die die Leute haben, ist verursacht durch düstere Science-Fiction-Visionen. Die finde ich unterhaltsam. Aber man muss sich bewusst darüber sein, dass es nur Geschichten sind – keine Realität.

TR: Sie können nicht Realität werden?

Arkin: Wir haben gerade erst angefangen, diese Dinge zu erforschen. Die Robotik schreitet schneller voran, als sie es je zuvor getan hat. Wir sollten also in der Tat schon jetzt über lügende Roboter diskutieren. Diese Technologie verändert das

Leben grundsätzlich. Es ist besser zu fragen: Was sollten wir tun? Statt zu fragen: Was haben wir getan?

Wissenschaftler der Universität Lissabon entwickelten einen lügenden virtuellen Software-Agenten, der sich besonders gut in seine Mitspieler hineinversetzen kann. Die Forscher ließen den Agenten in einem einfachen Spiel gegen Menschen antreten. Bei diesem Spiel ist einer der Mitspieler der "Werwolf", die anderen seine potenziellen Opfer. Die Rollen werden verdeckt zugewiesen. Die Spieler müssen durch Fragen herausfinden, wer der Werwolf ist – ohne gefressen zu werden. Die portugiesischen Forscher ließen nun jedes Mal den Agenten den Werwolf spielen, ohne die Menschen davon in Kenntnis zu setzen. Der Agent musste versuchen, durch Täuschung möglichst lange unentdeckt zu bleiben, indem er irreführende Antworten gab. Die Forscher konnten zeigen, dass er dann am erfolgreichsten war, wenn er nicht nur versuchte zu berechnen, was die anderen Spieler wissen. Zur erfolgreichen Strategie gehörte auch, darüber zu spekulieren, was die anderen Spieler dachten, was er weiß.

TR: Sollten wir derartige Technologien gesetzlich regulieren? Es wäre ja durchaus möglich, dass Unternehmen zum Beispiel betrügerische Chatbots programmieren, oder?

Arkin: Ja, absolut. Es gibt Situationen, in denen Lüge und Täuschung für Maschinen unmoralisch sind – und daher illegal. Wie Sie wissen, hält das Menschen nicht davon ab zu lügen. Bei Computern wird das hoffentlich etwas einfacher.

URL dieses Artikels:
http://www.heise.de/-2543665