

... denn sie wissen, was wir tun¹

Axel Ockenfels

Die meisten Menschen glauben, dass sie sich unter Kontrolle haben. Doch unsere Entscheidungen, Gefühle und Gedanken können immer besser prognostiziert und gesteuert werden. Die Vermessung des Menschen und die Berechnung seiner Zukunft bergen auf der einen Seite großartige Chancen, unser Leben erfolgreicher zu meistern. Sie erzeugen aber auch Befürchtungen vor Missbrauch und Ausbeutung. In diesem Beitrag skizziere ich ausgesuchte Beiträge der Erforschung von Rationalität und Künstlicher Intelligenz bei der Entschlüsselung des Menschen. Ich beschränke mich dabei auf das Verhalten in strategischen Situationen, in denen verschiedene Akteure Konflikte austragen, sich koordinieren müssen oder Kooperation anstreben.

A. Intelligenz, Rationalität und Prognose

Die wichtigste Voraussetzung für kluge Entscheidungen in unsicheren, strategischen Situationen ist die Fähigkeit, das Verhalten und die Reaktionen der anderen Spieler vorherzusagen. Eine gute Prognose ist der Schlüssel zum Erfolg in so unterschiedlichen Kontexten wie Schach oder Poker, auf Finanzmärkten, in der Politik, beim Management von Unternehmen, in Auktionen und auf Märkten, bei Verhandlungen und der Kindererziehung.

Doch wie können wir wissen, wie andere Akteure ticken, was sie als nächstes vorhaben und wie sie auf unsere Strategie reagieren werden? Zwei Prognosemodelle haben sich in vielen Situationen als erfolgreich erwiesen und ergänzen sich in wunderbarer Weise. Die *Spieltheorie* benötigt keine Daten, basiert auf der Annahme rationalen Verhaltens und erfreut sich großer Beliebtheit bei Wirtschaftswissenschaftlern. Die *künstliche Intelligenz* benö-

1 Ich bedanke mich beim European Research Council (ERC, im Rahmen des European Union's Horizon 2020 research and innovation programme, GA No 741409 – EEC) und der Deutschen Forschungsgemeinschaft (DFG, im Rahmen der Exzellenzstrategie des Bundes und der Länder – EXC 2126/1- 390838866) für die Unterstützung meiner Forschung.

tigt dagegen sehr viele Daten, kommt dafür aber ohne Verhaltensannahmen aus und erfreut sich großer Beliebtheit in der Computerwissenschaft.

Obwohl die beiden Ansätze auf den ersten Blick nur wenig gemein haben, stehen beide Ansätze im Kern auf derselben Konzeption kluger Entscheidungen: Intelligent beziehungsweise rational ist, wer sich so verhält, dass das eigene Ziel aller Voraussicht nach bestmöglich erreicht werden kann.² Da der Schlüssel für eine erfolgreiche Zielerreichung eine gute Prognose ist, spielt die Frage, wie man Verhalten möglichst gut prognostizieren kann, in beiden Forschungsrichtungen die zentrale Rolle.

B. Rationalität und Superrationalität bei Schere – Stein – Papier

Schere – Stein – Papier ist das vielleicht einfachste Beispiel für eine strategische Situation. Jeder Spieler überlegt, was wohl der andere Spieler vorhat, und versucht, dem anderen einen Schritt voraus zu sein. Die bessere Prognose gewinnt.

Wie ist es möglich, in einer solchen Situation, ganz ohne die Hilfe von Daten, Verhalten vorherzusagen? Die Spieltheorie erklärt, wie man mit purer Logik und allein auf Basis von Rationalität zu einem Ergebnis kommt. Man erkennt schnell, dass rationale Spieler von Schere – Stein – Papier versuchen müssen, sich selbst unvorhersehbar zu machen. Daraus folgt, dass die Spieler in jeder Runde ganz unabhängig davon, was bisher passiert ist, jede der drei möglichen Optionen mit der gleichen Wahrscheinlichkeit auswählen. Jede Abweichung von dieser Strategie – jede systematische Bevorzugung einer der drei Optionen – könnte von dem Gegenüber ausgebeutet werden und so die Wahrscheinlichkeit erhöhen, dass man verliert. Bei Rationalität gibt uns die Spieltheorie so eine klare und eindeutige Vorhersage über das Verhalten der Spieler. Mit dieser Strategie kann jeder Spieler für sich garantieren, dass er keinen systematischen Nachteil besitzt und mit mindestens die gleiche Gewinnchance besitzt.

Die Sache wird zwar schnell kompliziert, aber theoretisch gibt es keine Grenzen für die Anwendbarkeit der Spieltheorie. Für (fast) alle strategischen Situationen können im Prinzip kluge Strategien berechnet werden und dadurch die Verhaltensweisen rationaler Akteure prognostiziert werden. Das ist nicht bloß theoretisch interessant; bei zahlreichen Anwendungen in der

2 S. J. Russell, *Human Compatible, Artificial Intelligence and the Problem of Control*, New York City, New York: Viking Press 2019.

Praxis lassen sich durch logische Schlussfolgerungen wichtige konzeptionelle Einsichten ableiten. So spielt das Rationalitätskonzept beispielsweise eine gewichtige Rolle bei der Verhaltensprognose in Spielen wie Poker, in Verhandlungen und Auktionen, bei Wettbewerbsstrategien unter Konkurrenten, bei Handelskriegen, politischen Wettkämpfen und militärischen Konflikten, beim Finanzmarkt- und Onlinehandel, bei Kooperationsfragen und bei der Verkehrssteuerung.

Die künstliche Intelligenz beschäftigt sich oft mit ähnlichen Herausforderungen, geht aber anders vor. Sie prognostiziert nicht auf Basis logischer und kausaler Schlussfolgerungen, sondern durch datenbasiertes Lernen. Hier gilt es vornehmlich, mithilfe maschinellen Lernens in großen Datensätzen statistische Zusammenhänge zu erkennen, die dann ihrerseits Vorhersagen erlauben. Autonomes Fahren prognostiziert zum Beispiel auf der Basis von sehr großen Datenmengen, wie sich menschliche Verkehrsteilnehmer verhalten würden.

Die Spieltheorie besitzt den Vorteil, dass sie auch Vorhersagen für Situationen treffen kann, für die es noch keine Daten gibt. Dafür kann datenbasierte Intelligenz ein Menschenmodell bauen und so oft bessere Strategien entwerfen als rationale Strategien im Sinne der Spieltheorie – man könnte diese Strategien als „superrational“ bezeichnen. Wenn beispielsweise große Datenmengen über das Verhalten in Schere – Stein – Papier und über die zugehörigen Spieler zur Verfügung stünden, dann könnte eine künstliche Intelligenz allerlei Abweichungen vom Rationalverhalten feststellen. Aus der Verhaltensforschung wissen wir zum Beispiel, dass Menschen Schwierigkeiten haben, unabhängig von vergangenem Verhalten Entscheidungen zu treffen. Wenn in den vergangenen Runden bereits Schere gewählt wurde, wird es etwa unwahrscheinlicher, dass auch in der nächsten Runde nochmals Schere gewählt wird. Dieses Verhaltensphänomen ist als „Gesetz der kleinen Zahl“ bekannt. Es ist irrational und kann daher durch eine kluge datengetriebene Prognose über die nächste Aktion des Gegenübers ausgebeutet werden.

Die Strategie eines menschlichen Spielers kann aber auch von soziodemographischen und anderen Charakteristiken abhängen, etwa wie viel Erfahrung der Spieler in diesem Spiel oder in ähnlichen Spielen besitzt, von der Tageszeit, ob Freunde zuschauen, wie viel Alkohol getrunken wurde, von dem Gemütszustand, vom Wetter und vielem mehr. Eine künstliche Intelligenz, die Zugriff zu einem reichen Datenschatz hat, kann solche Zusammenhänge identifizieren und dadurch Prognosen erstellen, die dabei helfen, Schere – Stein – Papier *systematisch* zu gewinnen. Wenn die Maschine

etwa prognostiziert, dass Stein am wahrscheinlichsten gewählt werden wird, wird sie mit Papier antworten.

Die perfekte Strategie weicht in diesem Fall von der rationalen Strategie (ein Drittel Wahrscheinlichkeit für jede Option) ab. Die Abweichung von der Rationalität des Menschen steckt die Maschine an, die aber ihrerseits nicht irrational, sondern superrational wird. Dadurch macht sich zwar auch die Maschine verwundbar. (Wenn nämlich der Mensch die Maschine durchschaut und erwartet, dass sie Papier spielt, dann kann er mit Schere den Gewinn erzwingen.) Doch spätestens seit Menschen Maschinen beim Schach hoffnungslos unterlegen sind, dürfte offenbar sein, dass Menschen an die Denktiefe einer mit vielen Daten gefütterten und mit enormer Rechenkapazität ausgestatteten intelligenten Maschine in der Regel nicht herankommen.

C. Mensch versus Maschine

Herbert A. Simon, Wirtschaftsnobelpreisträger und einer der Begründer der modernen Wissenschaft der Künstlichen Intelligenz, erklärte vor Jahrzehnten die Wirtschaftswissenschaft mit ihrem Fokus auf Rationalität zu einer Wissenschaft vom Künstlichen ("Economics is one of the sciences of the artificial").³ Wenn künstliche Intelligenzen mit enormer Rechenkapazität aufeinandertreffen und miteinander interagieren, ist Rationalität ein nützliches Konzept, um Strategien und Ergebnisse zu prognostizieren. Es gibt auch eine Welt der Mathematiker und Spieltheoretiker, in denen das Rationalitätskonzept unmittelbar weiterhilft. In großen Auktionen der Telekommunikationsindustrie etwa, auf Finanzmärkten oder bei militärischen Konflikten, wenn Experten unablässig analysieren, wie die Wettbewerber ticken und welche Strategie wohl rational ist. Doch in einer Welt, in der Menschen aus Fleisch und Blut mitmischen, mit all ihren kognitiven und motivationalen Beschränkungen, führt die Rationalprognose für sich genommen selten zur besten Prognose. Datenbasierte Ansätze wie die experimentelle Forschung und, bei großen Datenmengen, die künstliche Intelligenz können dann die Prognosegüte verbessern helfen und so zu klügeren Entscheidungen führen.

Es ist vielleicht weniger überraschend, dass uns die künstliche Intelligenz durch neue Technologien und den Einsatz massiver Computerpower in Spielen wie Schach überlegen ist. Diese Intelligenz, richtig angewendet, eröffnet neue Möglichkeiten bei der Bewältigung unserer Herausforderungen. Doch

3 H. A. Simon, *Models of Bounded Rationality*, Cambridge, Mass., MIT Press 1976.

eine besondere Faszination für die künstliche Intelligenz entsteht, weil die menschliche Intelligenz so beschränkt ist. Intelligente Maschinen verstehen die Menschen mit ihren Irrationalitäten und Schwächen in mehr und mehr ökonomischen und sozialen Kontexten besser und können uns daher besser prognostizieren, als wir uns selbst verstehen und unsere eigenen Handlungen vorhersagen können. Dies führt dazu, dass die künstliche Intelligenz Menschen selbst in einfachsten Situationen (wie Schere – Stein – Papier) gut einschätzen kann und uns dadurch systematisch überlegen ist. Dazu kommt, dass Akteure, die gelernt haben, wie Menschen auf soziale oder ökonomische Impulse, auf Entscheidungsarchitekturen oder auf Informationsfeedback reagieren, auch Verhalten in eine gewünschte Richtung lenken können. Das muss nicht negativ sein. Viele Herausforderungen erfordern ein verändertes Verhalten – man denke etwa an den menschengemachten Klimawandel, Pandemien, Verkehrskollaps, Selbstkontrollprobleme und Finanzmarktexzesse. Doch auch die Sorge vor ökonomischem Missbrauch, Ausbeutung und vor einer Aushöhlung der Demokratie durch Eingriffe und Manipulationen ist berechtigt.

Eine breite Diskussion darüber, wie und unter welchen Umständen moderne Formen der Verhaltenslenkung eingeschränkt werden sollen, ist notwendig. Hier steht die Forschung erst am Anfang. Ein zu Regulierung und Gesetzgebung ergänzender Weg, solchen Herausforderungen zu begegnen, ist die Entwicklung von Regeln und Algorithmen, die wünschenswerte Ergebnisse aus sich heraus erreichen, selbst wenn sich alle (algorithmischen und menschlichen) Spieler bestmöglich im jeweiligen individuellen Eigeninteresse verhalten. Viele Beispiele im Forschungsfeld „Behavioral Market Design“ zeigen, dass dieser Weg vielversprechend ist. Auch hier spielen Spieltheorie und künstliche Intelligenz eine zentrale Rolle.

