

WARUM DIE DATENWISSENSCHAFTEN DIE SOZIALWISSENSCHAFTEN BRAUCHEN

| VON JOHANNES MÜLLER (CAUSE&EFFECT UG)

Es ist das Kernversprechen der künstlichen Intelligenz: Wenn wir nur genug Daten haben, können mathematische Modelle – viel besser als der Mensch – Muster erkennen, Resultate vorhersagen und Entscheidungen treffen. Doch war dieses Kernversprechen nicht eigentlich der Markenkern der Sozialwissenschaften? Komplexe Sachverhalte verstehen, Muster abstrahieren, Folgen abschätzen, um anschließend bessere Entscheidungen treffen zu können? Stellt sich die Frage, ob die Sozialwissenschaften im Kontext von KI einzig deren Auswirkungen auf Gesellschaft und Politik untersuchen und begleiten werden.

Ganz im Gegenteil: sozialwissenschaftliche Methoden werden nicht nur eine wichtige Rolle in den Datenwissenschaften spielen, vielmehr sind sie elementar für den Erfolg von Datenanwendungen im öffentlichen und dritten Sektor.

VON DER KORRELATION ZUM VERSTÄNDNIS

Bei all den großen Versprechen von maschinellem Lernen und künstlicher Intelligenz basieren deren Methoden am Ende des Tages auf linearer Algebra und Statistik. Ein Beispiel sind die viel besprochenen künstlichen neuronalen Netze: Sie erkennen Muster in großen Mengen von Daten und bieten so die Möglichkeit, neue Datenpunkte einzuordnen. Diese Muster basieren aber fast ausschließlich auf Korrelationen. Bei einigen Anwendungsfällen ist das vollkommen ausreichend – bei der Bilderkennung zum Beispiel. An anderer Stelle sind sie unterkomplex, denn sie helfen uns

nicht, die Daten zu verstehen und die Ergebnisse zu interpretieren. Ein Beispiel: Ein neues Regelwerk soll helfen, Langzeitarbeitslose in den Arbeitsmarkt zu reintegrieren. Dazu werden mithilfe eines neuronalen Netzwerks und historischer Daten speziell Menschen identifiziert, bei denen eine erfolgreiche Reintegration wahrscheinlich ist.

Gemessen an der Modellgüte scheint das neuronale Netz hervorragend zu funktionieren, allerdings bleiben entscheidende Fragen unbeantwortet: Funktioniert das Regelwerk? Wie funktioniert es? Warum funktioniert es und unter welchen Umständen (nicht)? Funktioniert die Software? Wie funktioniert die Software? Warum funktioniert sie beziehungsweise wann und warum funktioniert sie nicht?

Antworten auf diese Fragen liefert das Modell – da es auf Korrelationen basiert und die Kausalität ausblendet – nicht. An dieser Stelle bieten die Sozialwissenschaften das passende Instrumentarium, nämlich kausale Wirkungsmodelle, experimentelle und quasi-experimentelle Methoden und insbesondere eine theoretische Fundierung.

KOMPLEXITÄT

Eine Grundannahme von Statistik und maschinellem Lernen ist die sogenannte „Independent and identically distributed“-Annahme: Man geht davon aus, dass sich das System, in dem die



Daten generiert werden, selbst nicht ändert. Nur so lässt sich ein gelerntes Muster auf neue Daten anwenden. Dynamische Komplexität ist damit der wahrscheinlich größte limitierende Faktor von künstlicher Intelligenz. Bei manchen Anwendungsfällen ist sie praktisch vernachlässigbar (zum Beispiel bei Anwendungen im Internet-of-Things-Bereich), in anderen ist sie elementar.

Genau das ist die Stärke von sozialwissenschaftlicher Betrachtung. Sozialwissenschaften analysieren Daten fast ausschließlich im Kontext. Sie untersuchen Veränderungen dieses Kontexts und deren Auswirkungen. Denn Daten entstehen nicht in einem Vakuum – sie werden mal mehr, mal weniger bewusst generiert. Um auf dieser Basis Einblicke zu generieren und Entscheidungen zu treffen, ist ein qualitatives Verständnis dieses Datengenerierungsprozesses, des Kontexts, unbedingt geboten.

DER MENSCH IM MITTELPUNKT

Nun zur zentralen Frage: Sollen algorithmische Systeme Entscheidungen für uns treffen? Auch hier kommt es sehr stark auf den Kontext an. In manchen Situationen sind die Risiken begrenzt. Wenn zum Beispiel ein Unternehmen mithilfe eines Algorithmus Windelwerbung gezielt an Nutzer ausspielen möchte, ist die Fehlerwahrscheinlichkeit nicht wirklich entscheidend. Finanziell ja, aber eben nur das. Wenn ein mathematisches Modell aber darüber entscheiden soll, ob ein Antrag bei einer Behörde angenommen wird oder nicht, steht viel mehr auf dem Spiel.

Das Ziel sollte daher eine datenbasierte Entscheidungsfindung sein – und nicht die Entscheidung durch Daten. Dafür müssen Daten kritisch eingeordnet und Handlungsempfehlungen zielgerichtet abgeleitet werden.

FAZIT

Der öffentliche Sektor kennt viele komplexe Herausforderungen. Datenanwendungen können helfen, diese Komplexität zu reduzieren und neue Chancen zu nutzen. Ein rein technischer Ansatz wird dabei aber scheitern, ein rein qualitativer Ansatz ebenfalls. Für den verantwortungsvollen und wirkungsorientierten Umgang mit datenbasierten Systemen und daraus abgeleiteten maschinellen Entscheidungen im öffentlichen und sozialen Sektor brauchen wir ein sozialwissenschaftliches Verständnis der Daten und der Methoden der Datenwissenschaften. ●



Johannes Müller ist Data Scientist und Sozialwissenschaftler. Er ist Mitgründer und Chief Data Scientist bei der Datenberatung cause&effect UG und Gründer und Vorsitzender von CorrelAid e.V. Er hält einen Master of Science in „Evidence-based Social Intervention and Policy Evaluation“ von der University of Oxford.

