Matthias Huber, Simone Schüller, Marc Stöckli und Klaus Wohlrabe

Maschinelles Lernen in der ökonomischen Forschung

Die Anwendung von maschinellem Lernen hat sich in zahlreichen Gebieten etabliert und wird für Unternehmen und den Wissenschaftsbetrieb zunehmend unabdingbar. Auch in der empirischen Wirtschaftsforschung steigt die Anzahl der Publikationen, die mit entsprechenden Methoden arbeiten. Dennoch scheint eine gewisse Skepsis bei manchen Wissenschaftlern zu bestehen. Ein Kritikpunkt ist, dass sich maschinelles Lernen zwar für Vorhersagen eignet, aber nicht dazu, kausale Zusammenhänge zu identifizieren. In den vergangenen Jahren hat sich die Forschung jedoch verstärkt mit diesem Problem auseinandergesetzt, und es wurden zahlreiche Fortschritte erzielt. Maschinelles Lernen hat daher das Potenzial, in Zukunft in der Wirtschaftsforschung an Bedeutung zu gewinnen. Der Beitrag zeigt einige Hintergründe und aktuelle Entwicklungen auf.

Maschinelles Lernen (ML) bezeichnet die Fähigkeit von Computern, aus Erfahrung zu lernen, ohne dass diese spezifisch dafür programmiert worden sind. Algorithmen – definierte Abläufe von Einzelschritten – können verwendet werden, um grundlegende Charakteristika in Datensätzen zu erkennen und diese zu verallgemeinern. ML ermöglicht es, Vorhersagen über neue und bisher unbekannte Datenpunkte zu treffen. Ein Arzt kann beispielsweise anhand des Krankheitsbildes eines Patienten eine Einschätzung darüber treffen, ob es sich bei einer Gewebeveränderung um einen gut- oder bösartigen Tumor handelt. On-Demand-Streaming-Dienste wie Netflix oder Spotify verwenden ML-Algorithmen, um Anwendern aufgrund ihrer Nutzungsgeschichte jene Filme oder Musik vorzuschlagen, die ihren Präferenzen am ehesten entsprechen. Und Unternehmen wie Tesla oder Google entwerfen damit Fahrzeuge, die in Zukunft ohne das menschliche Eingreifen unfallfrei(er) durch den Straßenverkehr manövrieren können.

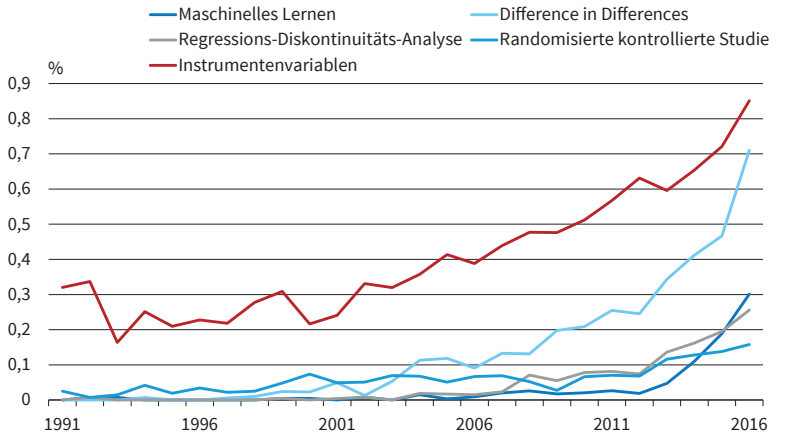
METHODENTRENDS IN DER ÖKONOMISCHEN FORSCHUNG

Wie aber kommt ML in der ökonomischen Forschung an? Um dies beurteilen zu können, vergleichen wir die Verwendung von ML mit den gängigsten ökonometrischen Methoden: Die Regressions-Diskontinuitäts-Analyse, Difference in Differences, Instrumentenvariablenschätzer und randomisierte kontrollierte Studien sind anerkannte Methoden, die standardmäßig zur Werkzeugkiste eines empirischen Wirtschaftswissenschaftlers gehören. Sie kommen in der Kausalanalyse zum Einsatz und sind in den vergangenen drei Jahrzehnten im Zuge der »Glaubwürdigkeitsrevolution« entwickelt worden (vgl. Angrist und Pischke 2010). Im Kern steht die Erkenntnis, dass nur die Anwendung rigoroser empirisch-kausalanalytischer Verfahren exakte und verlässliche Aussagen über die ursächliche Wirkung wirtschaftspolitischer Maßnahmen ermöglicht. Diese Verfahren orientieren sich an experimentellen bzw. quasi-experimentellen Untersuchungsdesigns, die die kontrafaktische »Was-wäre-wenn«-Situation der Nicht-Durchführung der betreffenden Politikmaßnahme mit Hilfe von Kontrollgruppen abbilden, um Ursache-Wirkungs-Zusammenhänge zu erforschen. Abbildung 1 zeigt, wie sich die Verwendung dieser ökonometrischen Methoden in Fachpublikationen anteilig an den Gesamtpublikationen entwickelt hat. Hierfür haben wir alle wissenschaftlichen Arbeiten in referierten Fachzeitschriften nach entsprechenden Schlagwörtern durchsucht und klassifiziert (eine genaue Beschreibung der Klassifizierung ist im Kasten zu finden). In der Abbildung ist zu erkennen, dass die Methoden der Kausalanalyse seit Ende der 1990er Jahre im Mainstream der empirischen Wissenschaften zunehmend Fuß fassen. Instrumentenvariablen kommen dabei am häufigsten zum Einsatz. 2016 arbeitet fast 1% aller Publikationen (inklusive theoretischer Arbeiten) mit dieser Schätzmethode. Sie ermöglicht es – unter bestimmten Annahmen – eine Korrelation zwischen den erklärenden Variablen und dem Fehlerterm auszuschließen, um eine Aussage über kausale Effekte treffen zu können. In Abbildung 2 werden die entsprechenden Jahreswachstumsraten dargestellt. Am aktuellen Rand

Abb. 1

Verwendung von Methoden in Publikationen

In Prozent aller wirtschaftswissenschaftlichen Publikationen



Quelle: EconLit; Berechnungen des ifo Instituts.

© ifo Institut

weist insbesondere Difference in Differences starkes Wachstum auf. Diese Methode wurde in den 2000er Jahren populär. Dabei wird die Zielvariable (z.B. Arbeitslosigkeit) einer Versuchs- und Kontrollgruppe jeweils vor und nach einer Maßnahme (z.B. Einführung eines Mindestlohns) miteinander verglichen. Die Regressions-Diskontinuitäts-Analyse und randomisierte kontrollierte Studien kommen im Vergleich dazu weniger zur Anwendung, es kann aber auch hier ein positives Wachstum festgestellt werden.

Maschinelles Lernen in der Wirtschaftsforschung

Um die Entwicklung der kausalanalytischen Methoden und deren Verwendung in den Wirtschaftswissenschaften zu erfassen, durchsuchen wir die in referierten Fachzeitschriften veröffentlichten Publikationen nach Schlagwörtern (eine sog. Boolesche Suche). Ziel ist es, die Anzahl der jährlich veröffentlichten wirtschaftswissenschaftlichen Publikationen, die inhaltlich die entsprechenden Methoden verwenden, möglichst nahe zu approximieren. Über die EconLit-Datenbank stehen uns Titel, Kurzfassung und Schlüsselwörter von über einer Million veröffentlichten Fachartikeln zur Verfügung, beginnend im Jahr 1900. Die Datenbank umfasst die relevantesten Publikationen innerhalb der Wirtschaftswissenschaften in Fachzeitschriften, Arbeitspapieren führender Universitäten, Dissertationen, Sammelbänden, Tagungsberichten und Buchrezessionen. Für unsere Analyse beschränken wir uns auf die referierten Fachzeitschriften und auf die Jahre 1991–2016. Wir suchen in der Datenbank nach Publikationen, in denen neben maschinellem Lernen folgende Methoden verwendet werden: Differenz von Differenzen (*difference in differences*), Regressions-Diskontinuitäts-Analyse (*regression discontinuity design*), randomisierte kontrollierte Studie (*randomised control trial*) und Instrumentenvariablenschätzer (*instrumental variable estimator*). Eine Publikation wird dabei einer bestimmten Methode zugeordnet, wenn mindestens eines der von uns bestimmten Schlagwörter entweder im Titel, der Kurzfassung oder in einem der Schlüsselwörter vorkommt. Die gesuchten Schlagwörter entsprechen dabei den Methodennamen bzw. den in der Literatur verwendeten Abkürzungen. Im Falle von Differenz in Differenzen wird beispielsweise nach »difference in differences«, »difference-in-differences«, »DID« etc. gesucht. Für ML-Publikationen suchen wir nach »big data« und »machine learning«. Es ist möglich (und oft auch der Fall), dass eine Publikation mehr als einer Methode zugeordnet wird.

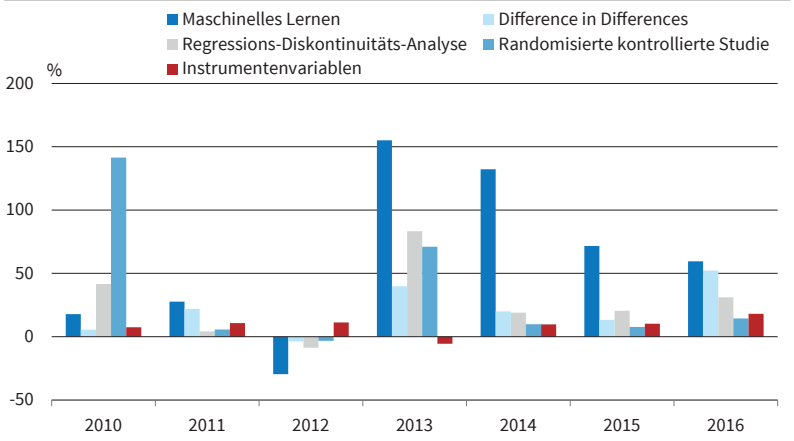
In unserer Stichprobe befinden sich insgesamt 777 122 Publikationen über einen Zeitraum von 25 Jahren. Dabei ist bei sämtlichen Einträgen der Titel vorhanden, jedoch nur bei 624 808 (80,4%) die Kurzfassung und bei 647 170 (83,3%) mindestens ein Schlüsselwort. Die jährliche Anzahl der Publikationen ist seit 1991 stetig gestiegen. 1991 waren es noch 11 864, 2013 bereits 50 690. Es ist davon auszugehen, dass auch in den Jahren 2014–2016 die Anzahl weiter gewachsen ist. Da die Publikationen aber mit einer Verzögerung von mehreren Jahren in der EconLit-Datenbank aufgenommen werden, kann für diese Jahre noch keine eindeutige Aussage getroffen werden.

Gegen eine Boolesche Suche zur Beantwortung unserer Forschungsfrage gibt es zahlreiche Einwände. Zum einen besteht die Gefahr, dass eine Publikation fälschlicherweise einer Methode zugeordnet wird. Dies kann geschehen, wenn nach irrelevanten oder zweideutigen Schlagwörtern gesucht wird. Beispielsweise wird die Suche nach »DID« auch jene Publikationen als Difference in Differences identifizieren, in denen äquivalente Abkürzungen verwendet werden (so wird *dissociative identity disorder* ebenfalls mit DID abgekürzt). Des Weiteren ist es unwahrscheinlich, dass sämtliche Autoren ihre Methoden explizit in dem Titel, der Kurzfassung oder den Schlagwörtern erwähnen. Auf der anderen Seite kann nicht ausgeschlossen werden, dass gewisse Autoren bewusst Modewörter verwenden, um ihre Leserschaft zu vergrößern. Letzteres Problem besteht möglicherweise bei maschinellem Lernen, da dieser Begriff zurzeit große Aufmerksamkeit genießt. Dies würde dazu führen, dass überproportional viele Publikationen mit der entsprechenden Methode identifiziert werden. Unsere Ergebnisse müssen daher mit gebotener Vorsicht interpretiert werden.

Abb. 2

Verwendung von Methoden in Publikationen

Wachstumsraten gegenüber dem Vorjahr



Quelle: EconLit; Berechnungen des ifo Instituts.

© ifo Institut

In den vergangenen Jahren ist das stärkste Wachstum aber bei ML zu erkennen. 2013 steigt dessen prozentualer Anteil an allen Publikationen sprunghaft an. Es scheint so, als ob sich ein neuer Trend in der Wirtschaftsforschung anbahnt. Bei genauerer Betrachtung zeigt sich aber, dass es sich bei den meisten dieser Publikationen um Arbeiten aus dem Bereich der Konjunkturprognose handelt.[1] Wirtschaftsexperten sind in der Tat nach wie vor skeptisch, inwiefern ML außerhalb der Prognose überhaupt einen Mehrwert für die Wirtschaftswissenschaft darstellt. Es wird kritisiert, dass ML zwar die Prognosegüte verbessern kann, sich aber nicht eignet, um kausale Effekte zu identifizieren. Kleinberg et al. (2015) illustrieren dies in ihrer Unterscheidung zwischen »Regenschirm«- und »Regentanz«-Problemen: Für die Entscheidung, einen Regenschirm einzupacken oder nicht, benötigt man lediglich eine möglichst präzise Prognose der Regenwahrscheinlichkeit. Im Gegensatz dazu stehen »Regentanz«-Probleme. Ist angesichts einer Dürre die Entscheidung zu fällen, ob in einen Regentanz investiert werden soll, ist das Wissen um den kausalen Zusammenhang zwischen Tanz und Regen unabdingbar. ML-Methoden sind, so die Kritik, ausschließlich auf »Regenschirm«-Probleme anwendbar und führen hier auch zu beachtlichen Effizienzsteigerungen. Zur Lösung von »Regentanz«-Problemen können sie aber keinen Beitrag leisten. Es ist insofern nicht verwunderlich, dass ML hauptsächlich bei den Prognostikern Anklang findet. Diese sind vorrangig an einer möglichst genauen Vorhersage interessiert, weniger an den genauen Wirkungskanälen.

MASCHINELLES LERNEN UND KAUSALE INFERENZ

Die in der Computerwissenschaft seit Mitte der 1980er Jahre entwickelten ML-Methoden wurden keineswegs mit der Absicht der Erforschung von Ursache-Wirkungs-Zusammenhängen entwickelt. Vielmehr dienen sie dem Zweck, möglichst präzise Voraussagen zu treffen. Mit dem aufkommenden Internet fanden diese Techniken beispielsweise Verwendung in Algorithmen von Suchmaschinen oder bei der Erstellung von Produktvorschlägen auf Einkaufsportalen. Kausale Inferenz wird in der Tat erst in der jüngeren Literatur adressiert. Zu den Pionieren auf diesem Gebiet zählen etwa Susan Athey, Stefan Wager, Guido W. Imbens (alle Stanford University) oder Stephen Ryan (Washington University in St. Louis). Athey und Imbens (2016) zeigen, dass ML insbesondere für die Verbesserung von bewährten kausalanalytischen Verfahren eingesetzt werden kann. Um beispielsweise den durchschnittlichen Effekt einer wirtschaftspolitischen Maßnahme mittels eines Instrumentenvariablenschätzers zu identifizieren, schlagen Belloni et al. (2013) ein doppeltes Selektionsverfahren vor. In einem ersten Schritt werden mit Hilfe eines ML-Algorithmus jene Variablen in einem Datensatz ausgewählt, die besonders hoch mit der Zielgröße korreliert sind. Im zweiten Schritt wird der gleiche Algorithmus verwendet, um die Variablen auszuwählen, die mit zu untersuchenden Maßnahmenvariable korreliert sind. Anschließend werden die Parameter der so ausgewählten Variablen in einem gemeinsamen Regressionsmodell geschätzt. Verglichen mit einem einfachen Regressionsmodell können dadurch die Schätzereigenschafter verbessert werden. Athey und Wager (2017) entwickeln eine Methode zur Schätzung heterogener Behandlungseffekte. Oftmals haben Maßnahmen eine unterschiedliche Wirkung auf verschiedene Personengruppen. So hat der Mindestlohn vermutlich einen anderen Beschäftigungseffekt im Verarbeitenden Gewerbe als im Dienstleistungssektor. Es besteht aber die Gefahr, dass Wissenschaftler iterativ nach Untergruppen mit hohen Behandlungseffekten suchen und anschließend in der Publikation nur jene Gruppen angeben, bei denen besonders starke Effekte gefunden wurden. Ist die Stichprobe groß genug, kann aus statistischen Gründen immer Heterogenität gefunden werden, auch wenn diese nur zufällig zustande gekommen ist (eine sog. Alphafehler-Kumulierung). Athey und Wager verwenden ein Regressionsbaummodell, ein in der ML-Literatur häufig verwendetes Verfahren, und passen dieses so an, dass für die Alphafehler-Kumulierung kontrilliert wird und dadurch kausale Aussagen getroffen werden können.

Neben Konjunkturprognose und Kausalanalyse sind ML-Methoden auch für weitere Anwendungen

[1] So verwendet das ifo Zentrum für Makroökonomik und Befragungen u.a. Boosting-Methoden für die Prognose konjunktureller Zeitreihen (vgl. u.a. Wohlrabe und Buchen 2014 oder Lehmann und Wohlrabe 2016).

in der Wirtschaftsforschung interessant. Beispielsweise können Methoden aus der Textanalyse große Datensätze durchsuchen und verwendet werden, um Indices über die Wirtschaftsaktivität zu erstellen (vgl. z.B. Thorsrud 2016). Klassifizierungsalgorithmen können in der Datenaufbereitung hilfreich sein, um Zusammenhängen sichtbar zu machen. Grundsätzlich gilt es festzuhalten, dass maschinelles Lernen in der Wirtschaftsforschung an Relevanz gewinnt. Und mit der Entwicklung von kausalanalytischen Methoden ist zu erwarten, dass deren Verwendung auch in Zukunft zunehmen wird.

LITERATUR

Angrist, J. und J.-S. Pischke (2010), »The credibility revolution in empirical economis: how better research design is taking the con out of econometrics«, *Journal of Economic Perspectives* 24(2), 3–30.

Athey, S. und G. Imbens (2016), »Recursive Partitioning for Estimating Heterogeneous Causal Effects«, *Proceedings of the National Academy of Sciences* 113(27), 7353–7360.

Athey, S., G. Imbens und S. Wager (2016), »Efficient Inference of Average Treatment Effects in High Dimensions via Approximate Residual Balancing«, *arXiv:* 1604.07125.

Belloni, A., V. Chernozhukov, I. Fernández-Val und C. Hansen (2013), »Program Evaluation and Causal Inference with High-Dimensional Data«, *arXiv:* 1311.2645.

Kleinberg, J., J. Ludwig, S. Mullainathan und Z. Obermeyer (2015), »Prediction Policy Problems«, *American Economic Review: Papers & Proceedings* 105(5), 491–495.

Lehmann, R. und K. Wohlrabe (2016), »Looking into the black box of boosting: the case of Germany«, *Applied Economics Letters* 23(17), 1229–1233.

Thorsrud, L.A. (2016), »Nowcasting Using News Topics. Big Data versus Big Bank«, *Norges Bank Working Paper*, 20.

Wohlrabe, K. und T. Buchen (2016), »Assesing the Macroeconomic Forecasting Performance of Boosting – Evidence for the United States, the Euro Area, and Germany«, *Journal of Forecasting* 33(4), 231–242.