INSTITUT FÜR ARBEITSMARKT- UND BERUFSFORSCHUNG
Die Forschungseinrichtung der Bundesagentur für Arbeit

Online Magazin IAB-Forum

Was kann maschinelles Lernen in der Arbeitsmarkt- und Bildungsforschung leisten? Ergebnisse eines internationalen Workshops

12. Februar 2021 | Regina T. Riphahn, Claus Schnabel, Gesine Stephan

Maschinelles Lernen versetzt Computer in die Lage, mit minimaler menschlicher Intervention aus Daten zu lernen. Mittels dieser Methoden erkennen IT-Systeme in großen Datenmengen Muster und wählen optimale Vorgehensweisen aus. Bei einem internationalen Workshop zeigten Forschende, dass solche Verfahren in der Arbeitsmarkt-, Bildungs- und Gesundheitsforschung spannende Beiträge leisten können.

Die Methoden des sogenannten maschinellen Lernens sind vielfältig und werden in der Forschung dynamisch weiterentwickelt. Was im Einzelfall die jeweils bestgeeignete Methode ist, hängt von der Problemstellung und den Daten ab. Diese Verfahren werden in der Informatik schon seit Jahrzehnten genutzt und weiterentwickelt. In letzter Zeit kommen die Methoden zunehmend in der

Arbeitsmarkt-, Bildungs- und Gesundheitsforschung zum Einsatz. Darüber hinaus werden sie für ökonomische Anwendungsfelder weiterentwickelt.

In dem internationalen Workshop „Machine Learning in Labor, Education, and Health Economics" präsentierten Forscherinnen und Forscher vor einem breiten Publikum online aktuelle Forschungsergebnisse, die sie auf Basis von maschinellem Lernen gewonnen hatten oder mit denen sie die Methoden selbst weiterentwickelt haben. Das IAB richtete die Veranstaltung gemeinsam mit dem Labor and Socio-Economic Research Center (LASER) der Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) aus.

Keynote: Kann maschinelles Lernen Vorhersagen verbessern?

Nach einer Begrüßung und Einführung durch Professorin Regina T. Riphahn von der FAU hielt Professor Jens Ludwig von der University of Chicago, ein hochrenommierter internationaler Experte zum Thema maschinelles Lernen in der Ökonomie, die erste Keynote. Unter dem Titel „Prediction policy problems in economics" vermittelte er einen Überblick über den Einsatz maschinellen Lernens für ökonomische Fragestellungen.

Als wichtigen Aspekt hob er hervor, dass maschinelles Lernen darauf ausgerichtet sei, gute Vorhersagen zu treffen, während die traditionelle Ökonometrie vor allem darauf abzielt, kausale Beziehungen zu identifizieren. Er verdeutlichte dies an folgendem Beispiel: Richter in den USA müssen darüber entscheiden, ob eine Person auf freien Fuß gesetzt wird. Sie müssen dabei abwägen, ob Angeklagte auf Kaution möglicherweise weitere Verbrechen begehen. Eine solche Vorhersage könnte grundsätzlich auch maschinell getroffen werden. Allerdings kann sie ausschließlich auf Daten basieren, die aus früheren Entscheidungen der Richter resultieren. Dies muss in entsprechenden Analysen berücksichtigt werden.

Ein Problem dabei ist, dass Algorithmen in den Daten vorhandene Diskriminierung reproduzieren können. Wurde beispielsweise die Gruppe der 30- bis 35-Jährigen nie auf freien Fuß gesetzt, so lässt sich deren potenzielle Straffälligkeit auch nicht verlässlich beurteilen. Jens Ludwig betonte, dass die Methoden neue Erkenntnisse in der ökonomischen Forschung hervorbringen könnten, aber gleichzeitig der Beitrag der Volkswirte für die korrekte Anwendung der Verfahren und die Auswertung der Ergebnisse wichtig sei.

„LASSO“ bringt neue Erkenntnisse zu Lohnunterschieden zwischen Männern und Frauen

Auf der Online-Konferenz wurden unterschiedliche Methoden präsentiert und diskutiert. Dazu gehörte auch LASSO (least absolute shrinkage and selection operator) – ein Verfahren, das datenbasiert erklärende Variablen zur Vorhersage auswählt. Damit lässt sich die Genauigkeit von statistischen Schätzungen verbessern. Diese Methode wurde auf vielfältige Art weiterentwickelt, um zu verlässlichen Aussagen zum Zusammenhang zwischen erklärenden und abhängigen Variablen zu gelangen.

Das Verfahren wurde in zwei auf der Konferenz vorgestellten Studien zur Lohnlücke zwischen Männern und Frauen eingesetzt. Philipp Bach von der Universität Hamburg untersuchte in einer dieser Studien für die USA, welche Faktoren die Lohnlücke zwischen den Geschlechtern beeinflussen. Dieser Studie zufolge spielen hier vor allem familienbezogene und arbeitsplatzbezogene Faktoren eine wichtige Rolle. So haben beispielsweise der wirtschaftliche Sektor und der Beruf, den Männer und Frauen ausüben, einen großen Einfluss auf die Lohnlücke.

In der zweiten Studie nahm Professor Anthony Strittmatter von der Universität St. Gallen die Frage in den Fokus, wie sich der Erklärungsgehalt der verschiedenen Faktoren, die zur Lohnlücke zwischen Männern und Frauen beitragen, mithilfe maschinellen Lernens methodisch verbessern lässt.

Die Potenziale von Analysen mit „Causal forests“

Die Methode der „Entscheidungsbäume“ wird unter anderem zur automatischen Klassifikation von Daten genutzt. Soll zum Beispiel das Risiko der Langzeitarbeitslosigkeit vorhergesagt werden, würde das Verfahren auf Basis der Daten zunächst das am besten geeignete Kriterium zur Bildung von Teilgruppen auswählen (etwa das Alter) und dann in den resultierenden Gruppen nach den dort bestgeeigneten weiteren Kriterien (zum Beispiel Gesundheitszustand) suchen.

Sogenannte Random-Forest-Verfahren kombinieren für eine Klassifikation mehrere Entscheidungsbäume auf Basis von Teilmengen der Daten. „Causal Forests“ sind eine neue Variante dieses Ansatzes, bei der es darum geht, die Wirkung von Maßnahmen für Teilgruppen zu ermitteln. Dieser Ansatz liegt den drei im Folgenden skizzierten Studien zugrunde, die inhaltlich sehr unterschiedliche Fragestellungen bearbeiten.

Professor Peter Elek, Forscher am Centre for Economic and Regional Studies der Eötvös-Loránd-Universität Budapest sowie Associate Professor an der Corvinus-Universität Budapest, untersuchte in seiner Studie regionale Unterschiede bei der Betroffenheit von Diabetes in Europa und

identifizierte entsprechende Risikofaktoren. Auf Basis von Befragungsdaten zeigte er unter anderem, dass das Risiko, an Diabetes zu erkranken, in Süd- und Osteuropa deutlich höher ist als in Westeuropa.

Im Anschluss ging Professor Andreas Gulyas von der Universität Mannheim der Frage nach, welche Faktoren die Höhe der Einkommensverluste bei Massenentlassungen bestimmen. Er zeigte auf Basis österreichischer Daten, dass die Verluste vor allem durch sogenannte Unternehmens-Lohnprämien bestimmt sind. Damit ist der Betrag gemeint, den Unternehmen Beschäftigten über das hinaus zahlen, was diese in anderen Unternehmen erwarten können.

Abgeschlossen wurde die Session durch Martin Kroczek vom Institut für Angewandte Wirtschaftsforschung in Tübingen mit einem Vortrag zu der Frage, wie sich der Anteil von Vollzeitbeschäftigten in Pflegeberufen erhöhen ließe. Die Ergebnisse zeigen, dass für die Attraktivität einer Vollzeittätigkeit sowohl Arbeitsplatzmerkmale wie die Lohnhöhe als auch persönliche Merkmale wie das Alter eine wichtige Rolle spielen.

Keynote: Wie kann maschinelles Lernen die Evaluation aktiver Arbeitsmarktpolitik verbessern?

Zu Beginn des zweiten Workshoptags begrüßte Professor Claus Schnabel von der FAU als zweiten Keynote-Speaker Professor Michael Lechner von der Universität St. Gallen. Lechner ist international als hochrangiger Experte für die Evaluation von Maßnahmen aktiver Arbeitsmarktpolitik bekannt. Er war einer der ersten Arbeitsmarktforscher, der Methoden maschinellen Lernens in seine Arbeit integrierte. Im Fokus seines Beitrags stand die Frage, inwiefern Maßnahmen aktiver Arbeitsmarktpolitik je nach Personengruppe unterschiedlich wirken können. Dies untersuchte er auf Basis flämischer Daten. Seinen Analysen zufolge war etwa die Förderung beruflicher Weiterbildung bei Immigranten besonders effektiv.

Der Einsatz von LASSO-Analysen in Studien zu Lohnstrukturen

Die beiden in der darauffolgenden Session vorgestellten Studien nutzten jeweils LASSO-Verfahren zur Analyse spezieller Aspekte der Lohnstruktur. IAB-Forscher Erik-Benjamin Börschlein präsentierte eine Studie zur Mindestlohneinführung in Deutschland. Er zeigte zunächst, dass sich durch LASSO-Analysen genauer hätte voraussagen lassen, wer damals von einem Mindestlohn potenziell profitiert hätte. Im zweiten Teil seines Vortrags präsentierte Börschlein die Ergebnisse einer von ihm durchgeführten nachträglichen Analyse der Reformeffekte des Mindestlohns. Diese bestätigte die auch in anderen Studien gefundenen positiven Effekte auf die Lohnentwicklung am unteren Ende der Einkommensskala.

Dr. Marina Bonaccolto-Töpfer von der FAU nutzte LASSO-Methoden, um in Befragungsdaten für Deutschland und Italien Lohnunterschiede, zum Beispiel nach Geschlecht, zu identifizieren.

In der letzten Session des Workshops stellten Professor Martin Biewen von der Universität Tübingen, Dr. Helmut Farbmacher vom Munich Center for the Economics of Aging des Max-Planck-Instituts für Sozialrecht und Sozialpolitik, und Professor Michael Knaus von der Universität St. Gallen schließlich verschiedene methodisch ausgerichtete Beiträge vor.

Maschinelles Lernen hat auch Grenzen – nicht zuletzt ethische

Maschinelles Lernen trägt dazu bei, die Potenziale großer Datenmengen (Big Data) besser zu nutzen. Wie Gesine Stephan, Leiterin des Forschungsbereichs „Arbeitsförderung und Erwerbstätigkeit“ am IAB und Professorin für Volkswirtschaftslehre, insbesondere Empirische Mikroökonomie, an der FAU, in ihrem Schlusswort zusammenfasste, zeigte der Workshop verschiedenste Anwendungsmöglichkeiten und Potenziale unterschiedlicher Methoden auf. Der Schwerpunkt lag dabei auf LASSO und „Causal Forests“.

In Zukunft werde die ökonomische Forschung voraussichtlich stärker untersuchen, inwieweit die entsprechenden Methoden sinnvoll zur Unterstützung von Entscheidungen herangezogen werden können – zum Beispiel, um zu bestimmen, wer aus einer Gruppe ein bestimmtes Gut erhalten soll, sagte Stephan. Dabei müssten mögliche ethische und datenschutzrechtliche Probleme sehr ernst genommen werden. Da datenbasierten Verfahren die menschliche Beurteilungsfähigkeit fehle, könne der Einsatz von Klassifizierungstechniken etwa zur Diskriminierung bestimmter Gruppen führen.

Verfügbare Daten reichen dabei nach Einschätzung von Stephan nicht notwendigerweise aus, um eine gute Prognosegüte zu erreichen. Auch maschinelles Lernen beruhe auf Daten für die Vergangenheit, sodass Veränderungen möglicherweise zu spät erkannt werden. Deshalb sollten datenbasierte Verfahren in der Praxis zur Unterstützung herangezogen werden, aber nicht, um menschliches Handeln zu ersetzen.

Zitationshinweis

Regina T. Riphahn; Claus Schnabel; Gesine Stephan (2021): Was kann maschinelles Lernen in der Arbeitsmarkt- und Bildungsforschung leisten? Ergebnisse eines internationalen Workshops , In: Online Magazin IAB-Forum 12. Februar 2021, https://iab-forum.de/was-kann-maschinelles-lernen-in-der-arbeitsmarkt-und-bildungsforschung-leisten-ergebnisse-eines-internationalen-workshops/, Abrufdatum: 27. May 2026

Lizenzhinweis

Diese Publikation ist unter folgender Creative-Commons-Lizenz veröffentlicht: Namensnennung – Weitergabe unter gleichen Bedingungen 4.0 International (CC BY-SA 4.0): https://creativecommons.org/licenses/by-sa/4.0/deed.de