

Christian BÜSCHER, Dortmund

Qualitative Fragestellungen in großem Maßstab – Potenziale maschinellen Lernens für mathematikdid. Forschung

Eine häufig genutzte Kategorisierung mathematikdidaktischer Forschung ist die Trennung in qualitative und quantitative Forschung. Qualitative Methodologien zielen dabei auf Fragestellungen nach detailliertem Beschreibungs- und Erklärungswissens ab, während quantitative Methodologien eher vergleichende Fragestellungen beantworten. Als eng damit verknüpft betrachtet wird der *Maßstab* der Forschung: Qualitative Analysen sind derart aufwändig, dass sie nur für wenige Fälle durchgeführt werden können, während quantitative Forschung in großem Maßstab stattfindet.

Diese Verknüpfung von kleinem Maßstab mit hoher Detailliertheit wirft die bekannte Frage nach der Verallgemeinerbarkeit qualitativer Ergebnisse auf. Möglicherweise ist die Gleichsetzung von kleinem Maßstab und qualitativen Fragestellungen dabei allerdings durch die Instrumente bedingt, die Forschenden in der Mathematikdidaktik zur Verfügung stehen: Für tiefgehende Analysen, etwa von Transkripten, ist viel Auswertungspersonal notwendig. Neueste Entwicklungen im Feld des Maschinellen Lernens (ML) könnten ein Werkzeug liefern, welches Teile aufwändiger Analysen übernehmen könnte. So nutzen etwa Kersting, Sherin und Stigler (2014) maschinelles Lernen, um offene Antworten von Lehrkräften zu gezeigten Unterrichtsvignetten automatisch zu bewerten.

Sollte es möglich sein, qualitative Analysen zumindest teilweise zu automatisieren, könnte dies qualitative Fragestellungen in großem Maßstab ermöglichen, und somit einen Beitrag zur Verallgemeinerbarkeit qualitativer Ergebnisse liefern. In diesem Beitrag wird ein kurzer Überblick über maschinelles Lernen und die Potenziale für mathematikdidaktische Forschungspraxis aufgezeigt, welche in einem ersten Projekt zu realisieren versucht wurden.

Qualitative Fragestellungen der Mathematikdidaktik und ML

Maschinelles Lernen ist ein Teilbereich der Künstlichen Intelligenz, einem Forschungsfeld, welches sich damit beschäftigt, automatisierte Lösungen für Probleme zu finden, welche bisher nicht von Maschinen, sondern nur von Menschen gelöst werden konnten (vgl. Rich, Knight, & Nair, 2009). Im Gegensatz zu traditioneller Programmierung besteht der Ansatz dabei nicht darin, Algorithmen zu entwickeln, die bestimmte Probleme lösen. Stattdessen

werden Architekturen von selbstständig lernenden Programmen (auch ‚Modelle‘) entwickelt, die auf Grundlage von einer großen Menge von Beispiellösungen Muster erkennen, mit denen auch unbekannte Probleminstanzen mit einer hohen Wahrscheinlichkeit gelöst werden können.

In den letzten Jahren kam es dabei zu großen technologischen Sprüngen im maschinellen Lernen. Einige der automatisierbaren Probleme lassen sich dabei gut auf die Probleme mathematikdidaktischer Forschung übertragen (Tab.1, Büscher, im Druck). So wird maschinelles Lernen etwa genutzt, um die Inhalte von Texten automatisch zu erkennen – eine Aufgabe mit großer Ähnlichkeit zu einer Inhaltsanalyse in qualitativer Forschung.

Problem	ML Beispiel	MD Beispiel
Textklassifizierung	Gegeben ein Text, ist es eine Zeitungsmeldung?	Gegeben eine Äußerung, ist es eine Begründung?
Textanalyse	Gegeben ein Text, was sind die Inhalte?	Gegeben ein Gespräch, was sind die Inhalte?
Sentiment-Analyse	Gegeben ein Tweet, ist es eine implizite Aufforderung zu Gewalt?	Gegeben eine Äußerung, was sind implizite Konzepte darin?
Sequenz-zu-Sequenz Zuordnung	Gegeben ein Satz, was ist eine Übersetzung mit der gleichen Bedeutung?	Gegeben eine Begründung, was sind Argument und Schlussfolgerung?

Tab. 1: Probleme in Maschinellern Lernen (ML) und Mathematikdidaktik (MD)

Potenzial: Unterstützung durch Vorverarbeitung reichhaltiger Daten

Ein Potenzial, das sich durch die Nutzung von Methoden des maschinellen Lernens ergeben kann, ist die praktische Unterstützung von Forschenden durch Vorverarbeitung reichhaltiger Daten. Ein trainiertes Modell, welches in beliebigen Transkripten etwa Äußerungen erkennt, in denen Begründungen vorkommen, könnte von Forschenden genutzt werden, um aus großen Datenmengen die potenziell interessantesten Äußerungen zu filtern. Nach einer solchen Vorauswahl könnten Forschende sich auf detaillierte Analysen der Begründungen fokussieren, anstatt nach diesen Begründungen erst suchen zu müssen. Somit kann sich durch maschinelles Lernen eine rein praktische Zeitersparnis ergeben, die es eventuell schon ermöglicht, den Maßstab qualitativer Forschung zu vergrößern.

Potenzial: Automatische Analyse offener Antwortformate

Offene Antwortformate bieten in Tests die Möglichkeit, auch Denkwege von Lernenden zu erfassen. Müssen Lernende etwa die Passung eines Terms zu einer Situation begründen, kann diese Begründung häufig einen Einblick in das konzeptuelle Verständnis ermöglichen. Gleichzeitig sind solche offenen Items mit traditionellen Methoden nicht schnell auswertbar, was ihren Nutzen für Forschung in großen Maßstäben stark verringert. Hier bietet maschinelles Lernen nicht nur das Potenzial einer Vorverarbeitung, sondern sogar das Potenzial einer automatischen Analyse, und damit einen echten Gewinn an möglichem Maßstab.

Hindernis und Potenzial: Charakter einer Blackbox

Ein ernstzunehmendes Hindernis beim Nutzen von maschinellem Lernen ist die fehlende Sichtbarkeit der gelernten Zusammenhänge. Aktuelle Methoden erlauben nahezu keinen Einblick darin, *wie* ein trainiertes Modell zu einer bestimmten Entscheidung gelangt. Trainierte Modelle bilden eine Blackbox; die Frage, ob ein trainiertes Modell auch in einem anderen Kontext und mit anderen Daten valide Analysen liefert, ist nach aktuellem Forschungsstand nur experimentell beantwortbar. Trainierte Modelle müssen nicht zwangsläufig dieselben Zusammenhänge in den Blick nehmen wie die Forschenden. Gleichzeitig bietet die Blackbox damit aber auch ein Potenzial, möglicherweise Zusammenhänge zu entdecken, die für das Forschungsteam in einem theoretisch ‚blinden Fleck‘ liegen.

Potenzial: Artifizielle Objektivität

Neben den praktischen und methodischen Potenzialen zeigt sich noch ein weiteres, theoretisch interessantes Potenzial. Ist ein Modell erstmal dazu trainiert worden, etwa Darstellungsvernetzung in offenen Antworten zu bewerten, so kann es auch in anderen Kontexten und von anderen Personen genutzt werden – es kann in einer Datei gespeichert werden, die anderen Forschenden zur Verfügung gestellt werden kann. Zwar geht der Nutzen solcher Modelle mit gewissen Einschränkungen einher, wie einer gewissen Kontextabhängigkeit durch die Wahl der Trainingsdaten. Zentral wird aber hier, dass die Forschenden sich dadurch auch diese Einschränkungen teilen, was die Forschungsergebnisse vergleichbar macht. Damit garantiert der Nutzen eines zwischen Forschenden geteilten Modells eine Art Objektivität: eine *artifizielle*, d.h. an ein Artefakt (dem trainierten Modell) gebundene Objektivität, die Ergebnisse verschiedener Forschender dadurch vergleichbar macht, dass sie aufgrund des gleichen genutzten Modells denselben Einschränkungen unterliegen.

Beispiel: Automatische Analyse von mathematischem Gehalt

Zum Ausloten des Potenzials maschinellen Lernens für qualitative Forschung wurde in einem Projekt ein Textanalyse-Ansatz gewählt, um in Äußerungen von Lehrkräften in Fortbildungen die Bezüge auf mathematische Inhalte automatisch zu erkennen (Büscher, im Druck). Hierzu wurden aus Fortbildungstranskripten 1343 Sätze extrahiert und von Hand durch einen Experten in Hinblick auf das Auftreten von mathematischen Inhalten kodiert. Ein Teil dieser kodierten Sätze wurde als Trainingsdaten für ein neuronales Netzwerk genutzt, um ein Modell zu trainieren, welches automatisch mathematischen Gehalt in Äußerungen identifiziert. Die restlichen Sätze wurden zur Überprüfung der Kodierqualität des Modells genutzt.

Es zeigt sich, dass auf dem Test-Set die Übereinstimmung von menschlichem Kodierer und Modell, gemessen durch Cohens Kappa, bei einer moderaten Übereinstimmung von 0,52 liegt. Da nur einfachste Mittel des maschinellen Lernens mit einem sehr kleinen Datenkorpus genutzt wurden, zeigt sich damit ein vielversprechendes erstes Ergebnis. Gleichzeitig ist nicht auszuschließen, dass die Ergebnisse durch die Stichproben beeinflusst wurden. Eine weitere Beurteilung der Güte dieser automatischen Analyse könnte nun erfolgen, indem es von anderen Forschenden genutzt wird, um die artifizielle Objektivität zu überprüfen; Modell und Trainingsdaten sind frei verfügbar unter www.ai.cbuescher.eu.

Fazit

Methoden des maschinellen Lernens bieten vielfältige Potenziale, von praktischer Vorsortierung von Daten über Vergrößerung des Maßstabs durch automatische Analysen hin zur Sicherung der Vergleichbarkeit durch artifizielle Objektivität. Dass dies nicht unbedingt utopische Potenziale bleiben müssen, zeigt eine erste, wenn auch eingeschränkte Realisierung einer automatischen Kodierung. Eine größere Ausnutzung des Potenzials benötigt eine breitere Umsetzung und, aufgrund der Vielzahl benötigter Daten, einen engen Austausch in der mathematikdidaktischen Community.

Literatur

- Büscher, C. (im Druck). Scaling up qualitative mathematics education research through Artificial Intelligence methods. *For the Learning of Mathematics*.
- Kersting, N. B., Sherin, B. L. & Stigler, J. W. (2014). Automated Scoring of Teachers' Open-Ended Responses to Video Prompts: Bringing the Classroom-Video-Analysis Assessment to Scale. *Educational and Psychological Measurement*, 74(6), 950–974. <https://doi.org/10.1177/0013164414521634>
- Rich, E., Knight, K. & Nair, S. B. (2009). *Artificial intelligence*. New Delhi: Tata McGraw-Hill.