

Gerhard Götz, Mosbach

Evaluation des Einsatzes adaptiver Online-Trainings in einem Inverted-Classroom-Kurs

Im August 2020 fand an der DHBW Mosbach ein virtueller, mathematischer Inverted-Classroom-Vorkurs mit 149 Studienanfänger*innen aufgeteilt auf elf Gruppen statt (Brüstle et al. 2021). In diese Kurse war jeweils zum Tagesausklang die Durchführung adaptiver Online-Trainingseinheiten mit Übungsaufgaben zur Vertiefung elementarer Themen der Sekundarstufen wie Arithmetik, Gleichungen, Potenzen, Wurzeln und Logarithmen (PWL) sowie Funktionen organisatorisch mit eingebunden. Die automatisierten Trainings empfehlen Nutzer*innen abhängig von zuvor gewählten Antworten individuell geeignete Folgeaufgaben.

Empfehlungssysteme basieren im Allgemeinen auf den Ansätzen des kollaborativen Filterns oder inhaltsbasierter Empfehlungen (Zhang et al. 2019). Ersterer überträgt Erkenntnisse über Nutzer*innen auf andere basierend auf Ähnlichkeiten zwischen diesen. Hierbei ist kein Vorabwissen, dafür aber eine große Nutzungszahl nötig. Im Gegensatz dazu modellieren inhaltsbasierte Empfehlungen das Verhalten der Nutzer*innen basierend auf a priori gesetzten Kategorien von Empfehlungsinhalten und Nutzer*innen. Diese Kategorien entsprechen oft nicht genau der Realität, dafür ermöglichen sie Empfehlungen ab der ersten Nutzung. Ziel des vorgestellten Systems ist die Verknüpfung der Vorteile beider Konzepte.

Fachdidaktische Modelle des Wissens und Könnens (Götz et al. 2020; Pinkernell et al., 2017), mittels denen sich Übungsaufgaben einsortieren lassen, sind im Sinne inhaltsbasierter Empfehlungen in den Empfehlungsalgorithmus implementiert, so dass auch zu Beginn didaktisch sinnvolle Empfehlungen gegeben werden. Mit dem Erfahrungswissen über Nutzer*innen (Götz und Wankerl, 2019) soll das System dazulernen, um die Empfehlungen zu verbessern. Einflussfaktoren sind dabei u.a. Bearbeitungsdauer, gewählte Antwort, Fehlertyp und die subjektive Schwierigkeitsbewertung am Ende einer jeden Aufgaben. Durch Analysen und Simulationen auf den gewonnenen, anonymisierten Datensätzen (vgl. Wankerl et al. 2020) soll das Empfehlungssystem dann schrittweise zu einem datengetriebenen System auf Basis des kollaborativen Filterns weiterentwickelt werden.

	Arithmetik	Gleichungen	PWL	Funktionen	Summe
Erfolgreich abgeschlossen	57	24	19	14	114
Nicht erfolgreich abgeschlossen	78	58	30	59	225
Vorzeitiger Abbruch	71	23	9	16	119
Summe	206	105	58	89	458

Tabelle: Übersicht über Anzahl absolvierter Trainingseinheiten

Eine erste Analyse des Empfehlungssystems basiert auf der empirischen Evaluation von insgesamt 10119 beantworteten Übungsaufgaben zu den vier Themengebieten in 458 Trainingseinheiten, von denen 119 vorzeitig abgebrochen wurden. Dies wirft insbesondere die Frage auf, ob sich aus den verfügbaren Daten Rückschlüsse auf die Ursachen für vorzeitige Abbrüche ableiten ließen, was zwischen 15% bei PWL und 34% bei Arithmetik aller Trainingseinheiten betrifft.

Betrachtet wurden hierfür drei a priori definierte Trainingsgruppen: vorzeitig abgebrochene Trainings, nicht erfolgreich abgeschlossene Trainings (Lösungswahrscheinlichkeit von unter 60%), die Empfehlungen für eine weitere Aufarbeitung beinhalteten und erfolgreich abgeschlossenen Trainings. Diese Gruppen wurden zunächst bezüglich folgender drei Aspekte miteinander verglichen: durchschnittliche, subjektive Bewertung der Aufgabenschwierigkeit, durchschnittliche Lösungswahrscheinlichkeit und durchschnittliche Anzahl absolvierter Aufgaben. Wichtig ist hierfür, dass die Anzahl der insgesamt vorgeschlagenen Aufgaben bis zum Trainingsabschluss zwischen 20 und 30 variieren konnte, je nachdem wie hoch die durchschnittliche Lösungswahrscheinlichkeit der Nutzenden war. Dieser Ansatz fußt auf folgender Hypothese: Für schwache Nutzer*innen wird erwartet, dass neben dem Üben mittels Aufgaben auch eine zeitnahe, ergänzende Beschäftigung mit weiteren Lernelementen, wie Texten hilfreich und nötig ist. Sie erhalten daher insgesamt weniger Aufgaben. Auch gute Studierende erhalten weniger Aufgaben, da bei ihnen eine zu große Anzahl eine gewisse Langeweile aufkommen lassen könnten. Bei mittelstarken Studierenden hingegen wird erwartet, dass sie durchaus direkt von zusätzlichen Übungen im Lernprozess profitieren könnten. Diese Hypothese gilt es in zukünftigen Analysen zu überprüfen.

Vergleicht man die Mittelwerte der durchschnittlichen subjektiven Schwierigkeitsbewertung durch Studierende (Abb. 1), so ergibt sich über alle vier Themengebiete dieselbe Tendenz: erfolgreich Abschließende liegen mit 0,3 deutlich unter nicht erfolgreich Abschließenden und diese liegen mit maximal 0,1 nur leicht unter den vorzeitigen Abrecher*innen. Dies deutet darauf hin, dass dieses Kriterium allein nicht ausschlaggebend für einen vorzeitigen Trainingsabbruch sein dürfte.

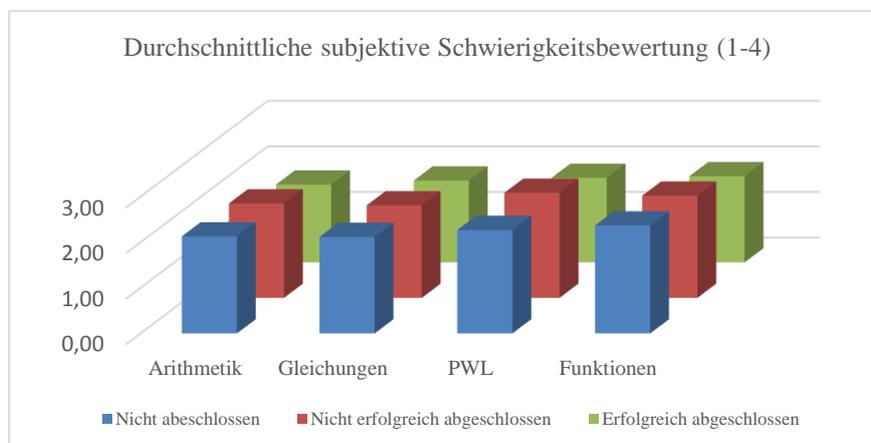


Abb. 1: Abhängigkeit des Erfolgs von der subjektiven Schwierigkeitsbewertung

Ein ähnlicher Eindruck ergibt sich bei der Betrachtung der durchschnittlichen Lösungswahrscheinlichkeit für die drei oben genannten Gruppen (Abb. 2), die über alle Themengebiete ein zu erwartendes Ergebnis liefern: Die niedrigste mit knapp über 50% ist gegeben bei nicht abgeschlossenen Trainings, leicht überboten mit knapp 60% bei nicht erfolgreich abgeschlossenen Trainings mit einem deutlichen Abstand zu erfolgreich absolvierten Trainings mit gut 80%.

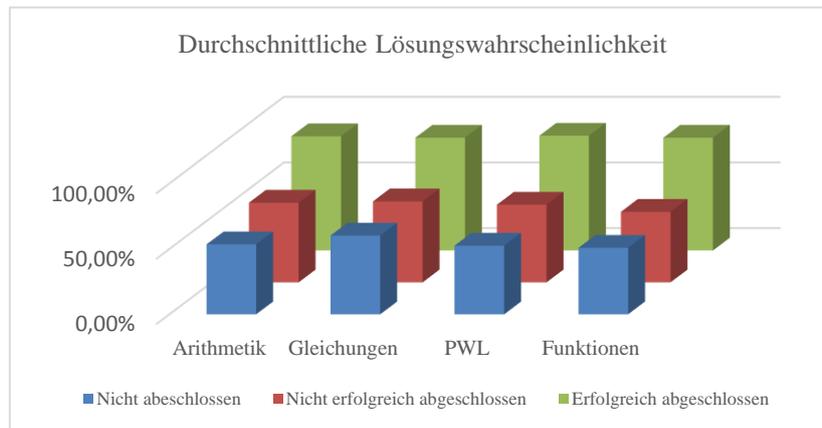


Abb. 2: Abhängigkeit des Erfolgs von der durchschnittlichen Lösewahrscheinlichkeit

Als dritte Frage wird untersucht, wie viele Aufgaben jede der drei Gruppen im Schnitt bearbeiten hat (Abb. 3). Der Unterschied zwischen erfolgreich und den nicht erfolgreich Abschießenden ergibt sich aus dem dynamischen Trainingsumfang. Deutlich spannender ist vielmehr die Frage, warum Studierende die Trainings nach ungefähr acht Aufgaben (hohe Varianz) abbrechen, obwohl sie diese kaum schwerer einschätzen und eine ähnlich hohe Lösewahrscheinlichkeit aufweisen wie die Gruppe, welche die Trainings nicht erfolgreich abschließen.

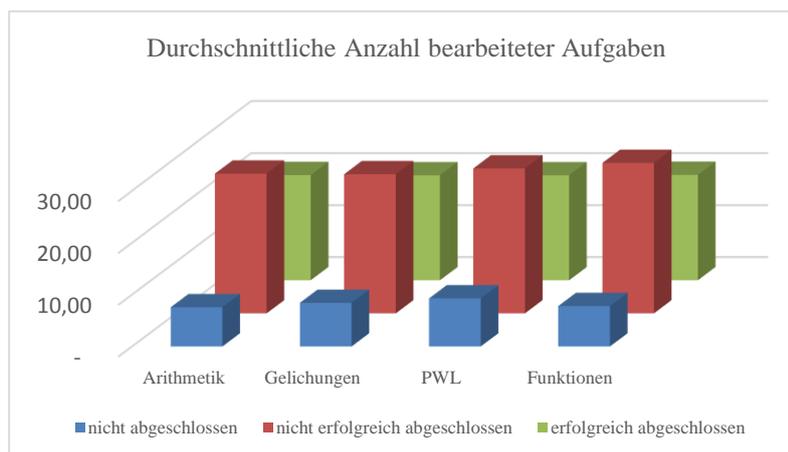


Abb. 3: Zusammenhang zwischen Erfolg und Anzahl bearbeiteter Aufgaben

Wirft man zusätzlich einen Blick auf die durchschnittliche Bearbeitungsdauer pro bearbeitete Aufgabe, so verleitet deren hoher Varianz zur Hypothese, dass dies an unterschiedlichen Typen von Trainingsabbrucher*innen liegen mag, was in weiteren Analysen noch überprüft und verifiziert werden soll. Sieht man die Reduktion der Abbruchquote als ein wichtiges Ziel an, dann deutet diese Wahrnehmungen darauf hin, dass die drei Gruppen noch anders bzw. feiner unterteilt werden

müssten, um daraus konkrete Schlüsse ziehen zu können, wie man das System für jede einzelnen Gruppe verbessern kann. Hierfür werden aktuell Ansätze der Korrelations- und Clusteranalyse verfolgt. Ebenso wird der Frage nachgegangen, ob sich aus der Analyse der Aufgabenpfade innerhalb der Trainingseinheiten Erkenntnisse bezüglich der Lernerfahrung ableiten ließen. Darüber hinaus werden Methoden des Machine Learnings angewandt, um auf Basis des Datensatzes Vorhersagen für das Empfehlungssystem treffen zu können (vgl. Wankerl et al. 2019).

Eingesetzt werden sollen die Trainings ab Sommer 2021 an allen DHBW Standorten mit einer insgesamt größeren Nutzer*innenbasis, die am Ende auch zur Trainingsnutzung befragt werden soll. Explizite Rückfragen könnten insbesondere über die konkreten Bedingungen und Ursachen aufklären, welche zu einem vorzeitigen Abbruch führten. Dem Ziel einer Verringerung der Abbruchquote soll zukünftig auch durch eine erhöhte Flexibilität und Anpassung der Trainings an geeignete Nutzer*innencluster näher gekommen werden. Dies geschieht maßgeblich auf Basis einer expliziten Einbindung oben erwähnter Analyse- und Simulationsergebnisse. Mittelfristig ist darüber hinaus beabsichtigt, übergreifendes „Repräsentationswissen“ über Formeln (vgl. Wankerl et al. 2020) ebenso in den Empfehlungsalgorithmus zu integrieren wie die Erweiterung um eine Auswahlfunktion zwischen zwei Aufgaben, die den Nutzer*innen gewisse Entscheidungsfreiheiten zugesteht, welche Aufgabe sie als nächstes bearbeiten möchten und ihnen dadurch die Möglichkeit bietet, den Lernprozess selbst mit steuern zu können.

Literatur

- Brüstle M., Götz G. und Hirnickel K. (2021). Brückenkurse im virtuellen Inverted-Classroom-Format. Erscheint in: *Synchron und asynchron: Berichte, Erfahrungen und Beispiele zur Lehre 2020. #DUAL. ZHL-Schriften-reihe für die DHBW. Band 5. DHBW CAS ZH.*
- Götz G. (2020). Automatisierte, adaptive Aufgabentrainings. In: *Beiträge zum Mathematikunterricht 2020* (S. 341–344).
- Götz G. & Wankerl S. (2020). Adaptives Online-Training für mathematische Übungsaufgaben. In: *Digitale Kompetenzen und Curriculare Konsequenzen. Tagungsband der Herbsttagung des Arbeitskreises Mathematikunterricht und digitale Werkzeuge vom 27. bis 28. September 2019 an der Pädagogischen Hochschule Heidelberg.* (S. 85-96)
- Götz, G., Hamich, M., Pinkernell, G., Schönwälder, D., Ullrich, D., & Wankerl, S (2020). Adaptives Üben, adaptive Aufgabentrainings, Modelle grundlegenden Wissens und Könnens. In *Selbststudium im digitalen Wandel* (pp. 93-126). Springer Spektrum, Wiesbaden.
- Pinkernell G., Düsi C. & Vogel M. (2017). Aspects of proficiency in elementary algebra. In: *Proceedings of the Tenth Congress of the European Society for Research in Mathematics Education (CERME10), 2017 464-471*
- Wankerl, S., Götz, G. & Hotho, A. (2020). f2tag - Can Tags Be Predicted Using Formulas? In: *19th IEEE International Conference On Machine Learning And Applications (ICMLA), pp. 565-571.*
- Wankerl, S., Götz, G. & Hotho, A. (2019). Solving Mathematical Exercises: Prediction of Students' Success. In: *Proceedings of the Conference on „Lernen, Wissen, Daten, Analysen“ (LWDA 2019), Vol. 2454, 190-194.*
- Zhang, S., Yao, L., Sun, A., Tay, Y.: Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys (CSUR)* 52(1), 5 (2019)