

Automatische Auswertung von Concept Maps: Wie kann Machine Learning helfen?

Theoretischer Hintergrund

Das Fach Physik wird von Schüler:innen als besonders schwierig und äußerst unbeliebt eingeschätzt (Williams et al., 2003; Winkelmann et al., 2021). Dies ist unter anderem darauf zurückzuführen, dass in naturwissenschaftlichen Fächern ein hoher persönlicher Aufwand betrieben werden muss, um ein Verständnis und Wissen über die verschiedenen Zusammenhänge zu entwickeln (Winkelmann et al., 2021). Eine Möglichkeit, dieses Problem anzugehen, kann das formative Assessment sein, da es seit langem als eines der wirksamsten Instrumente zur Optimierung des schulischen Lernens gilt (Black und Wiliam, 1998; Hattie, 2009). Eine gute Methode für ein formatives Assessment sind Concept Maps (Ruiz-Primo und Shavelson, 1996). Concept Maps sind gerichtete strukturierte Wissensdarstellungen, in denen die Knoten genau definierte Konzepte und die Kanten ihre Beziehungen darstellen (Novak und Cañas, 2008). Concept Maps decken ein breites Spektrum an Anwendungsfällen ab, da sie z.B. zur Vorwissensevaluation oder zum Aufdecken von Fehlvorstellungen genutzt werden können (Ley, 2015; Stracke, 2004). Damit Concept Maps für ein formatives Assessment genutzt werden können, müssen diese jedoch ausgewertet und analysiert werden. Eine qualitative Analyse der Concept Maps bietet einen detaillierteren Einblick in den aktuellen Wissensstand der Schüler:innen als z.B. das bloße Zählen richtig verwendeter Begriffe. Aus diesem Grund wird in dieser Arbeit ein qualitativer Ansatz gewählt, der sich auf die Analyse der einzelnen Propositionen und deren Qualität fokussiert. Allerdings ist eine qualitative Analyse sehr zeitaufwändig und deshalb im Schulalltag nur schwer umsetzbar (Hartmeyer et al., 2018). Um dieses Problem zu lösen, erforscht diese Arbeit eine neue Art der qualitativen Auswertung von Propositionen einer Concept Map zum Thema Mechanik unter Verwendung überwachter

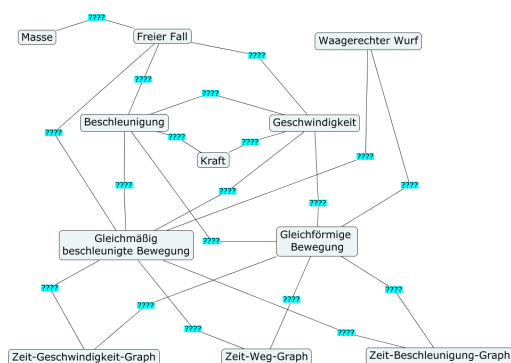


Abb. 1 Eingesetzte Concept Map zum Thema Mechanik

maschineller Lerntechniken. Durch Machine Learning soll so eine zeitnahe automatische Auswertung erfolgen, die den Lehrkräften Informationen liefert, welche sie für den weiteren Unterrichtsverlauf und individuelle Rückmeldung für ihre Schüler:innen nutzen können.

Studiendesign und ML-Modell

Für diese Arbeit wurde eine Concept-Map-Aufgabe zum Thema Mechanik entwickelt, bei der die Schüler:innen bis zu 19 Beziehungen zwischen 11 bereits vorher ausgewählten Begriffen ausfüllen mussten (Abb. 1). Es wurde sich bewusst für dieses eher geschlossene Format entschieden, damit zum einen vergleichbare Concept Maps entstehen und zum anderen keine Schüler:innen aufgrund kognitiver Überforderung ausgeschlossen werden. Die Concept-Map-Aufgabe wurde in 14 verschiedenen 11. Klassen an fünf verschiedenen Gymnasien in Norddeutschland eingesetzt. Insgesamt wurden 203 Concept Maps von den Schüler:innen mit der IHMC CmapTools Software (Daley et al., 2007) bearbeitet (Durchschnittsalter 16,9 Jahre; w 49%/ m 48%/ d 3%). Nachdem die Concept Maps bearbeitet wurden, wurden sie von zwei verschiedenen Bewertern (dem Erstautor und einem Studenten der Nanotechnologie) anhand einer Vier-Punkte-Skala bewertet. Zu diesem Zweck wurde ein allgemeines Bewertungsschema entwickelt, das aus drei Abstufungen von richtig (einfache Beziehung, einfache gerichtete Beziehung, detaillierte gerichtete Beziehung) und einer Kategorie für falsche Beziehungen bestand (Fischler & Peuckert, 2000; Friege, 2001). Außerdem wurde im Vorfeld für jeder der 19 Propositionen Codierbeispiele erstellt, sodass eine Übereinstimmung von 87 % zwischen den Bewertern erreicht werden konnte. Für die automatische qualitative Bewertung wurde ein separates Machine Learning Modell für jede der 19 Propositionen entwickelt. Das bedeutet, dass insgesamt 19 verschiedene Textklassifikationsmodelle entwickelt wurden, die als Input die Propositionen in der Form "<Concept_A> <SuS_Eingabe> <Concept_B>" und die Labels 0, 1, 2, 3 analog zu den Kategorien im Bewertungsschema erhalten. Die Antworten der Schüler:innen wurden dann mit Hilfe einer 10-fachen Kreuzvalidierungsstrategie in Trainings- und Testsets aufgeteilt (90% Training, 10% Test). Der verwendete Algorithmus ist ein Support-Vector-Machine-Klassifikator (SVC) mit einem Radial-Basis-Kernel (Noble, 2006). Für jeden der 10 Folds wurden die Hyperparameter des SVC mit Hilfe einer Gittersuche optimiert. Anschließend wurde der SVC mit den optimierten Parametern trainiert und die Antworten im Testset klassifiziert. Um die Leistung der Modelle zu bewerten, wurden Cohen's Kappa und der gewichtete F1-Score (harmonisches Mittel zwischen Recall und Precision) berechnet. Um die Ergebnisse zu vergleichen, wurde ein Zufalls-Klassifikator verwendet, der Vorhersagen unter Berücksichtigung der Klassenverteilung des Datensatzes erstellt (Pedregosa et al., 2011).

Ergebnisse

Die Verteilungen der 19 Propositionen der Concept Maps zeigen, dass der Ansatz einer annähernd gleichmäßigen Verteilung der Antworthäufigkeit gut funktioniert hat. Wenn man jedoch die vier Ratingkategorien betrachtet, erkennt man ein leichtes Übergewicht in der Kategorie 0 (falsche Antworten) mit 32 %. Danach folgt Kategorie 2 mit 26 %, Kategorie 3 mit 23 % und schließlich Kategorie 1 mit 19 % der Schüler:innenantworten. Es ist anzumerken, dass dies innerhalb der einzelnen Propositionen stark variiert. Die Gesamtergebnisse der Machine Learning Modelle zeigen, dass diese durchaus in der Lage

war, die studentischen Propositionen in dieselbe Kategorie einzuordnen wie die menschlichen Bewerter. Das Cohen's Kappa mit $\kappa = .80 \pm .06$ über alle 19 Propositionen (Abb. 2) weist eine gute Übereinstimmung auf (Vergleich Zufalls-Modelle $\kappa = .03 \pm .05$). Auch der gewichtete F1-Score von $.87 \pm .04$ (Abb. 2) zeigt, dass die Modelle deutlich besser waren, als ein Modell welches nur zufällig entscheidet (Vergleich Zufalls-Modelle $.36 \pm .07$).

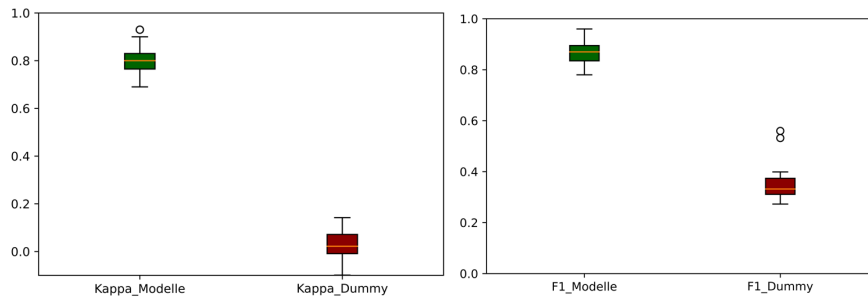


Abb. 2 Cohen's Kappa und gewichteter F1-Score der Modelle und der Zufalls-Klassifikatoren

Diskussion und Ausblick

Das Ziel dieser Arbeit war es, einen ersten Schritt in Richtung einer automatischen qualitativen Bewertung von Concept Maps mit Hilfe von Machine Learning Techniken zu machen. Die ersten Ergebnisse zeigen eine gute Übereinstimmung zwischen den menschlichen Bewertungen und den durch maschinelles Lernen generierten Bewertungen. Vergleicht man die Ergebnisse mit anderen Forschungen im Bereich der naturwissenschaftlichen Bildung und des maschinellen Lernens (z.B. Krüger et al., 2020; Wulff et al., 2020; Zhai et al., 2020), so kann man feststellen, dass die Mehrheit der Propositionen eine zufriedenstellende Übereinstimmung zwischen menschlichen und maschinellen Bewertungen aufweist. Dennoch gibt es weitere Schritte, um die Ergebnisse und die Modelle zu verbessern. Es wird außerdem wichtig sein, einen genauen Blick auf die (falsch) klassifizierte Antworten zu werfen, um zu verstehen, warum das Modell welche Kategorie gewählt hat und so die Entscheidungen verständlicher und interpretierbar zu machen. So soll auf Grundlage der Modelle ein Feedback-Tool für Lehrkräfte entwickelt werden, mit denen sie in der Lage sein sollen, einerseits in kurzer Zeit einen Überblick über den Leistungsstand der Klasse zu erhalten und andererseits ihren Schüler:innen ein individuelles Feedback geben zu können. Wie das Feedback-Tool verwendet wird und wie die Lehrkräfte mit dem Feedback umgehen, sind weitere spannende Forschungsfragen in diesem Projekt.

Literatur

- Black, P. & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Daley, B. J., Cañas, A. J. & Stark-Schweitzer, T. (2007). CmapTools: Integrating teaching, learning, and evaluation in online courses. *New Directions for Adult and Continuing Education*, 2007(113), 37–47. <https://doi.org/10.1002/ace.245>
- Fischler, H. & Peuckert, J. (Hrsg.). (2000). *Studien zum Physiklernen. 1. Concept mapping in fachdidaktischen Forschungsprojekten der Physik und Chemie*. Logos-Verl.
- Friege, G. (2001). *Wissen und Problemlösen - eine empirische Untersuchung des wissenszentrierten Problemlösens im Gebiet der Elektrizitätslehre auf der Grundlage des Experten-Novizen-Vergleichs. Studien zum Physiklernen. 19*. Logos Verl.
- Hartmeyer, R., Stevenson, M. P. & Bentsen, P. (2018). A systematic review of concept mapping-based formative assessment processes in primary and secondary science education. *Assessment in Education: Principles, Policy & Practice*, 25(6), 598–619. <https://doi.org/10.1080/0969594X.2017.1377685>
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement (2009. Aufl.)*. Educational research. Routledge Taylor & Francis Group.
- Krüger, D. & Krell, M. (2020). Maschinelles Lernen mit Aussagen zur Modellkompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 26(1), 157–172. <https://doi.org/10.1007/s40573-020-00118-7>
- Ley, S. L. (2015). *Concept Maps als Diagnoseinstrument im Physikunterricht und deren Auswirkung auf die Diagnosegenauigkeit von Physiklehrkräften* [, Duisburg, Essen]. BibTeX. <https://duepublico.uni-due.de/servlets/DocumentServlet?id=38141>
- Noble, W. S. (2006). What is a support vector machine? *Nature biotechnology*, 24(12), 1565–1567. <https://doi.org/10.1038/nbt1206-1565>
- Novak, J. D. & Cañas, A. J. (2008). *The Theory Underlying Concept Maps and How to Construct and Use Them*. Technical Report IHMC CmapTools 2006-01 Rev 01-2008, Florida. <http://cmap.ihmc.us/Publications/>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Ruiz-Primo, M. A. & Shavelson, R. J. (1996). Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching*, 33(6), 569–600. [https://doi.org/10.1002/\(SICI\)1098-2736\(199608\)33:6<569::AID-TEA1>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1098-2736(199608)33:6<569::AID-TEA1>3.0.CO;2-M)
- Stracke, I. (2004). *Einsatz computerbasierter concept maps zur Wissensdiagnose in der Chemie: empirische Untersuchungen am Beispiel des chemischen Gleichgewichts [Dissertation]*.
- Williams, C., Stanisstreet, M., Spall, K., Boyes, E. & Dickson, D. (2003). Why aren't secondary students interested in physics? *Physics Education*, 38(4), 324–329. <https://doi.org/10.1088/0031-9120/38/4/306>
- Winkelmann, J., Freese, M. & Strömmer, T. (2021). Schwierigkeitserzeugende Merkmale im Physikunterricht. Vorab-Onlinepublikation. <https://doi.org/10.25321/prise.2022.1168> (6-23 Pages / *Progress in Science Education (PriSE)*, Vol. 5 No. 2 (2022)).
- Wulff, P., Buschhüter, D., Westphal, A., Nowak, A., Becker, L., Robalino, H., Stede, M. & Borowski, A. (2020). Computer-Based Classification of Preservice Physics Teachers' Written Reflections. *Journal of Science Education and Technology*, 30(1), 1–15. <https://doi.org/10.1007/s10956-020-09865-1>
- Zhai, X., Yin, Y., Pellegrino, J., Haudek, K. C. & Shi, L. (2020). Applying machine learning in science assessment: a systematic review. *Studies in Science Education*, 56(1), 111–151. <https://doi.org/10.1080/03057267.2020.1735757>