

Stefan KRAUSS, Regensburg, Sven HILBERT, Regensburg,  
Karin BINDER, Regensburg & Alfred LINDL, Regensburg

## Re-Analysen des COACTIV-Datensatzes mit Machine Learning: Die ersten Schritte einer innovativen Analyse­methode

### Einleitung

Mit der COACTIV-Studie ist in den Jahren 2003 und 2004 eine in Bezug auf das Datenvolumen sehr umfangreiche empirische Studie zu professionellen Lehrkraftkompetenzen durchgeführt worden, bei der pro Lehrkraft (Studien­teilnehmende waren Mathematiklehrkräfte von PISA-Klassen) über 1000 Variablen erfasst wurden. Obwohl mittlerweile neben dem Sammelband (Kunter et al., 2011) auch zahlreiche wissenschaftliche Artikel zur COACTIV-Studie publiziert wurden, ist nach wie vor jedoch erst ein Bruchteil der erhobenen Variablen in statistische Analysen eingegangen.

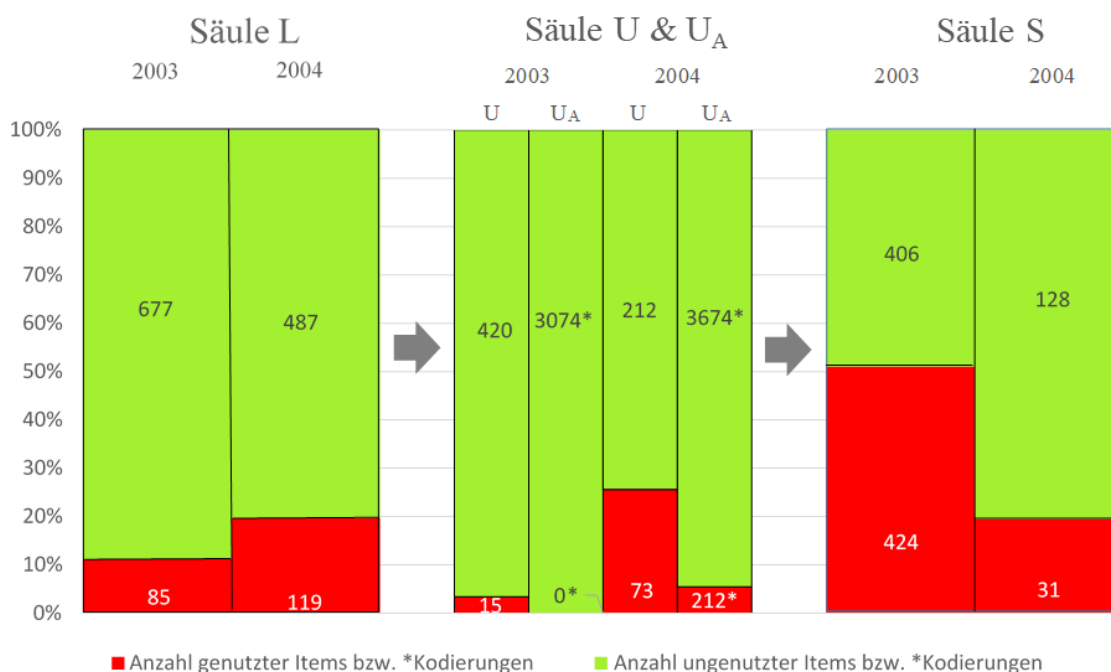


Abb.: Anzahl pro Lehrkraft verfügbarer (genutzter/ungenutzter) Items bzw. Kodierungen bezüglich Lehrkraftmerkmalen (Säule L), Unterrichtsmerkmalen (Säule U bzw. U<sub>A</sub> für die eingesammelten und kodierten Aufgaben) sowie Merkmalen von Schülerinnen und Schülern (Säule S), jeweils getrennt nach 2003 und 2004

Die Abbildung illustriert die prozentuale bisherige Verwendung (rot) von insgesamt pro Lehrkraft erhobenen COACTIV-Variablen in Strukturgleichungsmodellen zur Vorhersage von Lernfortschritten von Schülerinnen und

Schülern (Säule S) durch Lehrkraftkompetenzmerkmale (Säule L) beziehungsweise Unterrichtsqualitätsaspekte (Säulen U bzw.  $U_A$ ). Während die Säulen U (jeweils getrennt nach den beiden COACTIV-Messzeitpunkten 2003 und 2004) dabei die im Rahmen der Studie erhobenen Skalen von Schülerinnen und Schülern bzw. Lehrkräften visualisieren, beziehen sich die beiden entsprechenden Säulen  $U_A$  auf die von den Lehrkräften eingesammelten etwa 47.000 Aufgaben, die diese in ihren PISA-Klassen eingesetzt haben und die jeweils bezüglich 29 Dimensionen kodiert wurden (beispielsweise ob und auf welchem Niveau mathematische Argumentationen oder Modellierungen erforderlich waren oder ob es sich um rein technische Rechenaufgaben handelte; Jordan et al., 2008).

Gerade die bislang publizierten einflussreichen Strukturgleichungsmodelle zur Vorhersage des Leistungszuwachses der PISA-Schülerinnen und -Schüler unterliegen aber den Restriktionen traditioneller Auswertungsmethodik, da die maximale Anzahl an schätzbaren Parametern eng begrenzt ist, weshalb in den bisherigen Modellen meist nur ein sehr beschränkter Ausschnitt von COACTIV-Variablen gleichzeitig implementiert werden konnte.

Im Vortrag sollen die Grundzüge eines geplanten Projekts vorgestellt werden, nämlich die *gemeinsame* Analyse aller in COACTIV erhobenen Variablen zur Vorhersage des Lernzuwachses von Schülerinnen und Schülern mithilfe von Machine Learning-Verfahren. Da derartige Methoden noch ein Novum in der deutschsprachigen empirischen Bildungsforschung darstellen, sollen im Vortrag sowohl die Verfahren selbst als auch deren Eignung für den COACTIV-Datensatz illustriert werden.

### **Machine Learning-Verfahren**

In jüngster Zeit sind unter dem Begriff Machine Learning in vielen Forschungs- und Anwendungsbereichen Algorithmen prominent geworden, die eigens zur Analyse großer Datenmengen entwickelt wurden, um auf der Basis einer effektiven Kombination von Prädiktoren möglichst genaue Vorhersagen treffen zu können. Zu Zeiten der Administration der COACTIV-Studie waren solche Verfahren jedoch noch nicht unter vertretbarem Aufwand durchführbar. Mit dem Anstieg der Rechenleistung in der letzten Dekade sind die computationalen Möglichkeiten statistischer Analysen heute aber weitaus vielfältiger geworden, weshalb sich die Verwendung von Machine Learning-Verfahren für eine exhaustive Analyse des COACTIV-Datensatzes nun empfiehlt.

Der entscheidende Vorteil von Machine Learning-Methoden liegt in der praktisch unbegrenzten Variablenzahl, die im Verbund analysiert werden kann, wohingegen klassische statistische Ansätze bei einer großen Anzahl

an Variablen häufig mit Kollinearitätsproblemen zu kämpfen haben (deshalb konnte in COACTIV-Strukturgleichungsmodellen bislang meist nur eine überschaubare Menge von Konstrukten implementiert werden). Darüber hinaus sind Machine Learning-Verfahren weitgehend unabhängig von Modellannahmen und Parametrisierungen, während klassische statistische Verfahren wie Strukturgleichungsmodelle auf starken (und selten erfüllten) Voraussetzungen wie Verteilungsvorgaben, Linearität und Unabhängigkeiten von Fehlertermen basieren. Ein weiterer Vorteil ist, dass Machine Learning-Verfahren durch eine strikte Unterscheidung von Trainings- und Testdatensets eine realistische Einschätzung der Vorhersagekraft und Generalisierbarkeit erlauben, ganz im Gegensatz zur klassischenweise berichteten Varianzaufklärung in Form des Determinationskoeffizienten ( $R^2$ ), der die Prädiktivität in der Regel aufgrund zu großer Anpassung an die Stichprobe („overfitting“) deutlich überschätzt.

Die Vorzüge von Machine Learning-Verfahren im Vergleich zu klassischen statistischen Methoden hinsichtlich ihres Erkenntnispotenzials wurden bereits in vielen Bereichen offenbar: Erste Beispiele aus der Bildungsforschung sind hierfür die Sekundäranalysen des TIMSS-Datensatzes für Korea (Yoo, 2018), jedoch konnte in diesen Analysen (im Vergleich zu COACTIV) nur auf einen wesentlich kleineren Datensatz zurückgegriffen werden. Da Machine Learning-Verfahren in der deutschsprachigen empirischen Bildungsforschung bislang kaum genutzt werden, sollen im Vortrag neben den Verfahren selbst auch deren Vorteile, die Eignung für den COACTIV-Datensatz sowie die damit verfolgten Forschungsziele und erwartete Ergebnisse verdeutlicht werden.

### **Forschungsziele**

In der geplanten COACTIV-Sekundäranalyse sollen im Wesentlichen vier Machine Learning-Verfahren eingesetzt werden, nämlich *Random Forest* und *Gradient Boosting*, beides baumbasierte Verfahren, sowie *Least Absolute Shrinkage and Selection* und *Elastic Net*, beides regularisierte Regressionsmodelle, die besonders in Kombination mit mehrschleifigem Resampling robuste Ergebnisse liefern und die zu den wichtigsten gegenwärtigen Machine Learning-Algorithmen zählen (die nicht zum Bereich des Deep Learning gehören; siehe Efron & Hastie, 2016).

Dabei soll mit Methoden des „Interpretable Machine Learning“ (IML; Molnar, 2018), welche ein Teilgebiet des Machine Learnings darstellen, a) ein Set aus der vollständigen Menge der COACTIV-Variablen identifiziert werden, welches verlässlich und robust Lernzuwächse in Mathematik vorher-

sagt, und b) eine „Rangfolge“ der Bedeutung der Variablen für den Lernzuwachs abgeleitet werden (sog. „variable importance“). Zur Sicherung der Robustheit sollen hier geschachtelte Resampling-Verfahren durchgeführt werden, welche für das Tunen der Modellparameter in Kombination mit einer verlässlichen Schätzung der Vorhersagegenauigkeit notwendig sind (Hastie et al., 2005).

Eine zentrale Aufgabe wird im Anschluss an die (dem Wesen nach explorativen) Analysen vor allem auch die Re-Interpretation der Ergebnisse im Sinne einschlägiger Lehr-Lern-Theorien sein. Da in der COACTIV-Studie versucht wurde, ein möglichst umfassendes Bild zu erhalten (die Lehrkräfte benötigten für die Bearbeitung der Instrumente – über ein Jahr verteilt – insgesamt etwa 12 Stunden) und im COACTIV-Datensatz zahlreiche Skalen zu den verschiedensten (auch zu COACTIV-Modellen alternativen) Theorien implementiert wurden, bietet sich dieser Datensatz für das Vorhaben im Besonderen an.

### **Förderhinweis**

Die vorgestellte Studie wird gefördert von der Deutschen Telekom Stiftung.

### **Literatur**

- Efron, B. & Hastie, T. (2016). *Computer Age Statistical Inference*. Vol. 5. Cambridge: University Press.
- Hastie, T., Tibshirani, R., Friedman, J. & Franklin, J. (2005). The Elements of Statistical Learning: Data Mining, Inference and Prediction. *The Mathematical Intelligencer*, 27, 83–85.
- Jordan, A., Ross, N., Krauss, S., Baumert, J., Blum, W., Neubrand, M., Löwen, K., Brunner, M. & Kunter, M. (2006). *Klassifikationsschema für Mathematikaufgaben: Dokumentation der Aufgabenkategorisierung im COACTIV-Projekt* (S. 73–74). Berlin: Max-Planck-Institut für Bildungsforschung.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Molnar, C. (2018). *Interpretable machine learning. A Guide for Making Black Box Models Explainable*, 7.
- Yoo, J. E. (2018). TIMSS 2011 student and teacher predictors for mathematics achievement explored and identified via Elastic Net. *Frontiers in psychology*, 9, 317.