



A feedback component that leverages counterfactual explanations for smart learning support

Sebastian A. Günther, Felix Haag, Konstantin Hopf,
Philipp Handschuh, Maria Klose und Thorsten Staake

Abstract

The growing prevalence of digital learning in higher education is accompanied by challenges regarding students' self-regulated learning. While there is a plethora of behavioral interventions that aim at supporting students' self-regulated learning, they often do not consider the heterogeneity of students in their intervention design. This paper presents a novel feedback intervention that leverages the potential of machine learning and counterfactual explanations for providing personalized feedback to support students' learning. Ultimately, this approach could automatically adapt to different courses and thereby empower scalable and effective feedback.

S. A. Günther (✉) · F. Haag · K. Hopf · T. Staake
Department of Information Systems and Applied Computer Sciences, University of Bamberg, Bamberg, Germany
E-Mail: sebastian.guenther@uni-bamberg.de

F. Haag
E-Mail: felix.haag@uni-bamberg.de

K. Hopf
E-Mail: konstantin.hopf@uni-bamberg.de

T. Staake
E-Mail: thorsten.staake@uni-bamberg.de

P. Handschuh · M. Klose
Leibniz Institute for Educational Trajectories, Bamberg, Germany
E-Mail: philipp.handschuh@lifbi.de

M. Klose
E-Mail: maria.klose@lifbi.de

1 Introduction

Higher education is a demanding environment that poses new challenges for the learning behavior of students. In contrast to school, in which students' learning is structured and supported in more detail by their teachers (Vosniadou 2020), university students have to actively plan, monitor, and control their learning behaviors largely on their own to achieve their goals. Yet, many students are struggling in this process, potentially explaining severe dropout rates from study programs (see, e.g., Heublein et al. 2022). The resulting negative consequences accumulate among individuals, leading to major challenges also on a societal level (e.g., loss of time and funds).

With the growing role of digital learning environments in higher education, more and more learning data from students is automatically captured that can be leveraged to actively support them in their learning process. For instance, log data can serve to identify students at risk of dropping out of a course (Foster and Siddle 2020), allowing instructors to proactively support them. In a similar vein, digital learning environments offer scholars and practitioners novel opportunities to implement a wide range of behavioral interventions to automatically support students' online learning. Recent examples are components that provide feedback on performance (Leung et al. 2022), that help students to monitor their learning progress (Yoon et al. 2021), or that illustrate the online learning time spent (Günther 2021).

Despite the potential of behavioral interventions within digital learning environments, existing research in this area has neglected that students' personality and learning strategies are inherently different. More precisely, self-regulated learning theory (see, e.g., Pintrich 2004) implies that students employ different learning strategies and therefore might need personalized guidance for their learning. However, such personalized interventions require high efforts for human instructors (Hogan und Pressley 1997). Thus, personalized interventions are so far hardly scalable for a wide range of courses.

We argue that digital learning environments can empower such personalized guidance at great scale for university courses using online content: Through combining vast amounts of user activity data with students' course performance data (from previous runs of a course), the learning platform can identify patterns which learning actions have been influential to master a course. Ultimately, when these patterns are deployed in a digital learning environment, a corresponding feedback component can provide current students with personalized instructions on how to improve their online learning and potentially so their course performance.

Against this backdrop, we developed a feedback component that leverages the potential of digital learning environments, which we present in this paper. Specifically, we will briefly summarize our technical approach to initialize the feedback component, show its feedback design, and shed light on our experimental approach to test its effects on learners. The paper concludes with a brief description of the anticipated contributions of the study and the planned next steps.

2 Research Design

For our feedback component, we have instantiated machine learning (ML) models that learn relationships between students' digital learning actions and their overall course performance from past runs of a corresponding course. These models, each of them is solely used for feedback provision in a specific week of the course, are embedded into the learning platform. For each course participant, a week-specific ML model predicts a participant's performance in the final exam of the course, based on the participant's past behavior (log data, time tracking data) and their characteristics (socio-demographic backgrounds yielded from the registration page). To subsequently provide personalized feedback, we employ counterfactual explanation methods, which are a recent technical innovation in the field of explainable ML. Counterfactual explainers estimate how model input parameters (i.e., features) need to change in order to achieve a desired model outcome. Embedded in our feedback component, the explainer method infers what additional actions a learner has to perform (that is the change in input parameters) to improve their exam performance (that is the ML model's output), which Fig. 1 illustrates. The feedback component displays the obtained changes in input parameters as actions for exam improvement. By contrast, the ML model's predicted exam performance and the potential for exam improvement are not displayed (we treat these just as internal technical metrics).

The displayed actions for learners are based on learning strategies and are presented as instructions for learning behavior in the digital learning environment. To put that into context, instructions such as "Watch the video of lecture 3 again" or "Do the quizzes of Lecture 1" should encourage students to catch up with the learning content, monitor their knowledge, or deepen their understanding of specific topics of the lecture. Figure 2 displays the feedback component with examples of personalized actions for learners to improve their performance. The component is embedded into the main course page of the associated digital learning environment (i.e., open edX) so that it is salient to the learners.

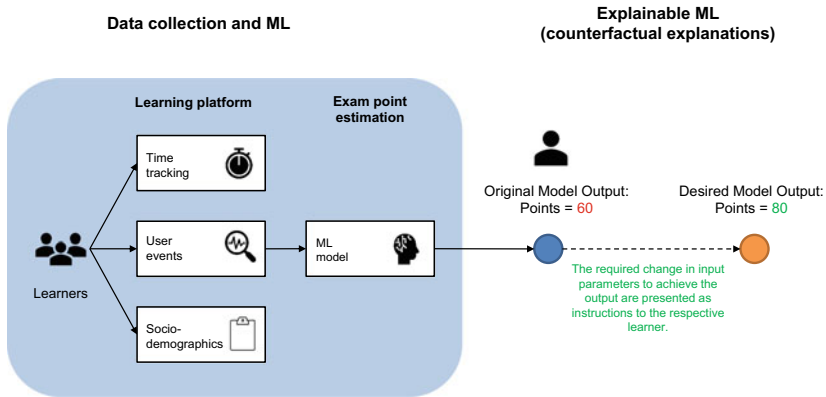


Fig. 1 Technical approach

1 This component has identified relationships between online learning behavior and course success of high-performing students from previous IITP courses. Considering your online learning behavior, you should take the following crucial steps, in addition to your regular online learning, to improve your performance in the exam. (last updated on 2022-08-22 00:00)

Done?

- You have already spent valuable learning time on the edX platform. By spending more online learning time in the following learning section, you are catching up on the learning content and deepen your knowledge: [Validation](#)
- By watching the following lecture video for the first time, you are catching up on the learning content: [L13: Why projects fail: Why projects often fail anyway](#)
- By downloading and working through the following slides, you are catching up on the learning content: [IIT-PM_T06_questions.pdf](#)

Fig. 2 Feedback based on counterfactual explanations

To evaluate the effects of the feedback component, we are running two experimental studies: One in a bachelor level course (summer semester 2022) and one in a master level course (winter semester 2022/2023). Each study follows a difference-in-differences experimental design. More precisely, we provide a group of students with feedback after a baseline phase, while the control group does not receive any feedback. We hypothesize that our feedback component will have desirable effects on students' course success (for an overview of learning techniques see Dunlosky et al. 2013 and for the effectiveness of feedback see Hattie und Timperley 2007). We investigate course success in terms of exam taking rate and points in the exam. To better understand the effects of the feedback intervention, we additionally conduct a pre- and post-survey that capture influential psychological constructs on online learning. These constructs relate for example to students' self-regulated learning skills, their procrastination behavior, and feedback acceptance. Additional analyses using these constructs will allow us to

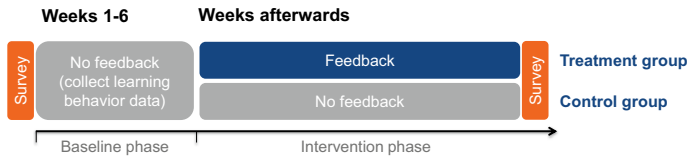


Fig. 3 Experimental setup

understand which students implement the instructions from the feedback component and benefit most from them in terms of study behavior, exam participation, and exam grade. Figure 3 shows the experimental setup for each of the two studies.

3 Conclusion

In this paper, we have presented a novel component that applies counterfactual explanations for providing personalized feedback. Even though we still have to statistically evaluate the effects of this feedback component on learners, its underlying technical approach is promising. The approach has the potential to unite learner characteristics (e.g., self-regulation skills, susceptibility to procrastination, etc.) and behavioral data to derive personalized guidance for learners on how to improve educational success. In doing so, the feedback component, operating in a digital sphere, can potentially promise more benefits than those that can be expected from a human instructor: Providing personalized feedback at scale. Given the relevance of learning strategies on academic success in higher education (Broadbent und Poon 2015), this paper encourages practitioners and scholars to consider such scalable approaches for empowering personalized learning support.

Acknowledgements This work was funded in part by “Stiftung Innovation in der Hochschullehre“ within the project “Digitale Kulturen der Lehre entwickeln (DiKuLe).“

References

Broadbent, Jaclyn and Walter L. Poon. 2015. Self-Regulated Learning Strategies & Academic Achievement in Online Higher Education Learning Environments: A Systematic

- Review. *The Internet and Higher Education* 27: 1–13. <https://doi.org/10.1016/j.iheduc.2015.04.007>.
- Dunlosky, John, Katherine A. Rawson, Elizabeth J. Marsh, Mitchell J. Nathan, and Daniel T. Willingham. 2013. Improving Students' Learning with Effective Learning Techniques: Promising Directions from Cognitive and Educational Psychology. *Psychological Science in the Public Interest* 14 (1): 4–58. <https://doi.org/10.1177/1529100612453266>.
- Foster, Ed, and Rebecca Siddle. 2020. The Effectiveness of Learning Analytics for Identifying At-Risk Students in Higher Education. *Assessment & Evaluation in Higher Education* 45 (6): 842–54. <https://doi.org/10.1080/02602938.2019.1682118>.
- Günther, Sebastian A. 2021. The Impact of Social Norms on Students' Online Learning Behavior: Insights from Two Randomized Controlled Trials. In *LAK21: 11th International Conference on Learning Analytics & Knowledge*, Ed. Association for Computing Machinery, 12–21. New York: Association for Computing Machinery. <https://doi.org/10.1145/3448139.3448141>.
- Hattie, John, and Helen Timperley. 2007. The Power of Feedback. *Review of Educational Research* 77 (1): 81–112. <https://doi.org/10.3102/003465430298487>.
- Heublein, Ulrich, Christopher Hutzsch, and Robert Schmelzer. 2022. Die Entwicklung der Studienabbruchquoten in Deutschland. Hannover: Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW). <https://doi.org/10.34878/2022.05.dzhwbrief>.
- Hogan, Kathleen, and Michael Pressley. 1997. *Scaffolding Student Learning: Instructional Approaches and Issues*. Cambridge: Brookline Books.
- Leung, Alvin Chung Man, Radhika Santhanam, Ron Chi-Wai Kwok, and Wei Thoo Yue. 2022. Could Gamification Designs Enhance Online Learning Through Personalization? Lessons from a Field Experiment. *Information Systems Research*. <https://doi.org/10.1287/isre.2022.1123>.
- Pintrich, Paul R. 2004. A Conceptual Framework for Assessing Motivation and Self-Regulated Learning in College Students. *Educational Psychology Review* 16 (4): 385–407. <https://doi.org/10.1007/s10648-004-0006-x>.
- Vosniadou, Stella. 2020. Bridging Secondary and Higher Education. The Importance of Self-Regulated Learning. *European Review* 28 (S1): S94–103. <https://doi.org/10.1017/S1062798720000939>.
- Yoon, Meehyun, Janette Hill, and Dongho Kim. 2021. Designing Supports for Promoting Self-Regulated Learning in the Flipped Classroom. *Journal of Computing in Higher Education* 33 (2): 398–418. <https://doi.org/10.1007/s12528-021-09269-z>.

Sebastian A. Günther is a researcher and doctoral student at the University of Bamberg, Germany. Email: sebastian.guenther@uni-bamberg.de

Felix Haag is a researcher and doctoral student at the University of Bamberg, Germany. Email: felix.haag@uni-bamberg.de

Konstantin Hopf is a postdoctoral researcher at the University of Bamberg, Germany. Email: konstantin.hopf@uni-bamberg.de

Philipp Handschuh is a research assistant at the Leibniz Institute for Educational Trajectories in Bamberg, Germany. Email: philipp.handschuh@lifbi.de

Maria Klose is a research assistant at the Leibniz Institute for Educational Trajectories in Bamberg, Germany. Email: maria.klose@lifbi.de

Prof. Dr. Thorsten Staake holds the chair of Information Systems and Energy Efficient Systems at the University of Bamberg and is co-director of the Bits to Energy Labs, a research initiative of ETH Zurich and the University of St. Gallen. Email: thorsten.staake@uni-bamberg.de

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

