Interpretable Knowledge Gain Prediction for Vocational Preparatory E-Learnings

Benjamin Paaßen[1], Malwina Dywel[2], Melanie Fleckenstein[2], and Niels Pinkwart[1,3]

[1]German Research Center for Artificial Intelligence
[2]Provadis Partner für Bildung und Beratung GmbH
[3]Humboldt-University of Berlin

This is a preprint of a paper accepted at tha AIED 2022 Practitioner's Track.

Abstract

Vocational further education typically builds upon prior knowledge. For learners who lack this prior knowledge, preparatory e-learnings may help. Therefore, we wish to identify students who would profit from such an e-learning. We consider the example of a math e-learning for the Bachelor Professional of Chemical Production and Management (CCI). To estimate whether the e-learning would help, we employ a predictive model. Developing such a model in a real-world scenario confronted us with a range of challenges, such as small sample sizes, overfitting, or implausible model parameters. We describe how we addressed these challenges such that other practitioners can learn from our case study of employing data mining in vocational training.

Keywords: Multi-dimensional item response theory, performance modeling, knowledge gain, vocational education, further education

1 AIEd implementation

Vocational further education aims to teach certain, job-related skills to a wide variety of learners. For example, the Provadis trade school offers a two-year course for the Bachelor Professional of Chemical Production and Management (BP-CPM; German: "Industriemeister Chemie", as defined by the chamber of commerce and industry, IHK) to acquire the skills necessary to supervise workers and apprentices in chemical plants[1]. A particular challenge for such further education courses is that incoming students have very different levels of prior knowledge, depending on their prior school education, their amount of job experience, and further individual factors. For example, some may enter the program right after high school education and apprenticeship, while others may have left school after lower secondary education, followed by an apprenticeship and multiple years of work experience before entering the program. Accordingly, students may lack or have forgotten pre-requisite knowledge, which severely impacts their chances for a successful qualification Fulano et al. [2021], Hailikari et al. [2008], Schaap et al. [2012].

Our ultimate educational goal is to provide personalized support in order to maximize learners' chances at succeeding in the program, irrespective of their individual starting point.

[1]`https://www.provadis.de/weiterbildung/fuer-berufstaetige/chemische-produktion/industriemeister/-in-chemie-ihk/`

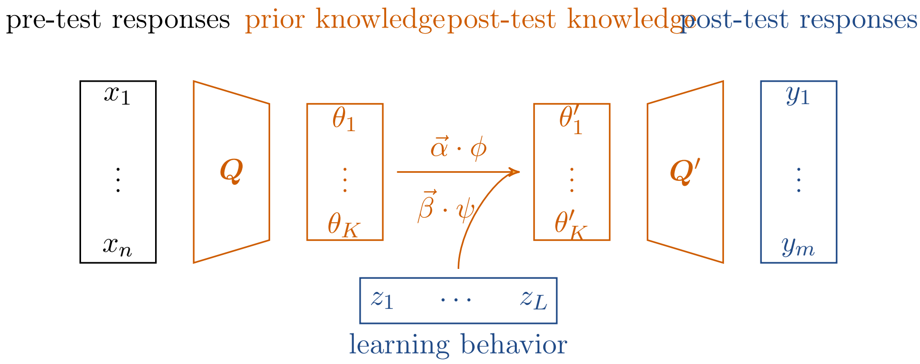

Figure 1: The proposed prediction pipeline. For prediction, we only use pre-test responses (black). For the training data, we also record learning behavior and post-test responses (blue). The predictive model itself is shown in orange.

In this paper, we focus on one specific strategy, namely recommending a preparatory mathematics e-learning to prospective students in the BP-CPM course who lack some pre-requisite math knowledge and would profit from the e-learning. Our recommendation scheme is fully automatic and has the following steps.

First, before entering the course, students perform a (voluntary) pre-test with twenty-one math questions regarding six skills, namely basic algebra, fractions, equation solving with a single variable, text tasks with two variables, powers, and linear functions[2]. Second, we diagnose gaps in math knowledge via classic test theory, that is, counting the rate of correct responses for each skill. Third, we predict how much the rate could improve if the student would visit the preparatory e-learning. Finally, we recommend the e-learning if the model predicts a sufficiently high gain (at least 5%, averaged over all skills) and we communicate the prediction to the student.

1.1 Predictive Model

To predict knowledge gain, given prior knowledge as measured by a pre-test, we use a predictive model. Developing such a model posed a challenge, particularly due to the small sample size in our educational setting. Given that the BP-CPM course is highly specialized, only very few students sign up per year (often less than ten). Accordingly, it is impossible to accumulate enough data from this population to train a very data-hungry model. On the other hand, a very simple, linear model is likely insufficient because we expect nonlinearities. In particular, we expect a bell-shaped relation between prior knowledge and knowledge gain, because very low prior knowledge means that a preparatory e-learning is insufficient (the skills would need to be learned from scratch) and high prior knowledge means that nothing is left to be learned.

Accordingly, we need a nonlinear model which requires as little training data as possible. To achieve such a model, we applied three assumptions: First, we assume that the relation between prior knowledge and knowledge gain can be modeled well by a linear combination of (few) bell curves. Second, we assume that knowledge gain is always non-negative, that is, there is no forgetting during the e-learning. This makes sense because the e-learning is fairly short (roughly five hours), such that forgetting is unlikely. Third, we assume that there is no interaction between skills, that is, prior knowledge in one skills does not help with acquiring

[2] `https://projekte.provadis.de/showroom/provadis/Mathematik_Orientierungstest/online/#p=32`

another skill. This last assumption is most likely false but crucial to reduce the number of free parameters and, thus, the need for data.

In more detail, our model has the following structure. Let θ_k be the prior knowledge of a student for skill k and θ'_k be the knowledge *after* visiting the e-learning. Further, let $\phi_1, \ldots, \phi_L$ be bell-curves, centered at different knowledge values. Then, we assume that the following, non-linear relationship holds:

$$\theta'_k = \theta_k + \sum_{l=1}^{L} \alpha_{k,l} \cdot \phi_l(\theta_k) + \text{auxiliary feature influences} \tag{1}$$

where $\alpha_{k,l}$ is a model parameter that weighs the influence of the lth bell-curve for skill k. We also include influences of auxiliary features in our model, especially the number of tasks a student has worked on, the number of correctly solved tasks, and the time spent on skill k, but these features are unknown for new students and, hence, we treat these influences as constant in the prediction.

We fit our parameters $\alpha_{k,l}$ by maximizing the log-likelihood on observed data. In particular, a sample of learners completed the pre-test, the e-learning, and a post-test. For the post-test, we assume an item response theory model Baker [2001], Hambleton and Swaminathan [1985], Hambleton and Jones [1993], where Equation 1 describes the ability of each student for skill k. The item-to-skill assignments for pre- and post-test $\boldsymbol{Q}$ and $\boldsymbol{Q}'$ where manually designed by experts. Using this model, we performed a maximum likelihood estimation of the difficulty parameters for each post-test question and the parameters $\alpha_{k,l}$. As such, our approach bears some resemblance to performance factors analysis, which also replaces student-specific ability parameters with an expression of prior knowledge plus learning behavior Pavlik et al. [2009]. Our model is visualized in Fig. 1.

1.2 Experiments and Results

For model training and evaluation, we recorded the data of $N = 30$ learners in the final year of their Chemical Production Technician or Chemical Laboratory Assistant apprenticeship (ages 16-19). This represents the population of learners who may later become BP-CPM. While 30 is a small sample size, it reflects the overall small population: A full cohort at the Provadis trade school consists of roughly 120–140 learners. We recruited all learners who were currently preparing for their final exam and were not sick or otherwise excused. Our sample included students both with prior high school education as well as students with lower secondary education. The study was performed as part of in-classroom teaching and lasted for four to five hours, including the pre-test, the e-learning, and the post-test (time varied depending on each learner's individual speed). This time span might appear short for a full e-learning, but bear in mind that the e-learning is meant to refresh knowledge, not necessarily teach from scratch.

To gauge the accuracy of our proposed model (refer to Fig. 1), we compared it to a direct logistic regression which predicts post-test answers from pre-test answers (logreg), a deep learning model which learns $\boldsymbol{Q}$ and $\boldsymbol{Q}'$ (deep), and a variation of our proposed model which uses a different encoder for pre-test abilities which is based on ranking the difficulties of items (rank). As hyperparameters, we used $L = 8$ bell-curves, regularization strength $C = 1$ for the difficulty parameters in the post-test, and regularization strength $C = 10^{-3}$ for the parameters $\vec{\alpha}$. For deep learning, we used 10,000 epochs of Adam optimization with a learning rate of 10^{-3}.

The results of our experiments are shown in Table 1. As we can see, our proposed pipeline performs best on the test set, although logreg is far superior on the training set. This highlights

Table 1: Mean accuracy ($\pm$ std.) in leave-one-out crossvalidation over $N = 30$ students.

	logreg	deep	rank	proposed
train	$\mathbf{0.99} \pm 0.00$	0.82 ± 0.01	0.83 ± 0.00	0.83 ± 0.00
test	0.81 ± 0.14	0.80 ± 0.18	0.81 ± 0.17	$\mathbf{0.83} \pm 0.16$

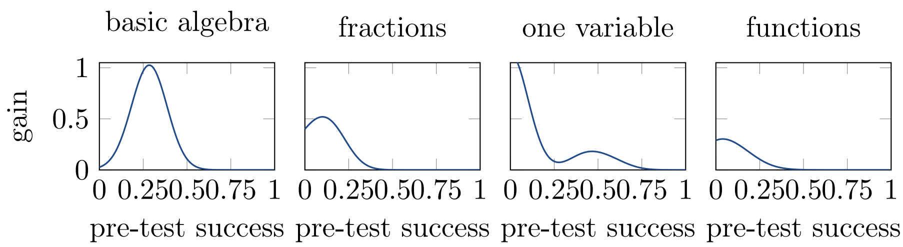

Figure 2: The predicted gain $\sum_{l=1}^{L} \alpha_{k,l} \cdot \phi(\theta_k)$ (y-axis) versus pre-test success rate θ_k (x-axis) for the skills 'basic algebra', 'fractions', 'equation solving with a single variable', and 'functions' as learned by our model.

the very real risk of overfitting for a small data set. The deep model does worst, indicating that learning the encoder does not offer a big advantage for this data.

Finally, we re-trained our proposed model on the data of all $N = 30$ students to obtain a model for practical application. Fig. 2 shows the learned curves for four example skills. Note that the curves differ for different skills, which may indicate that the e-learning is—on average—more effective in teaching some skills (such as basic algebra) compared to other skills (such as functions) in our sample. Nonetheless, given the small sample size and the interaction with other factors, we perform such interpretations with care.

2 Reflection of the challenges and opportunities

Developing artificial intelligence applications for vocational education is a challenging endeavor: The subjects are, typically, highly practical and thus difficult to translate to a virtual setting. Skillsets are, typically, highly specialized which means that cohorts are small and data is scarce. Finally, vocational education topics, typically, evolve quickly such that any AI teaching tool for a specific topic bears the risk of being out-of-date soon. Such practical issues might explain why artificial intelligence has is still much less common in vocational education compared to university or school education Seifried and Ertl [2021].

In this paper, we considered math knowledge, which is relevant across a broad range of vocational topics and less subject to change. Nonetheless, we need to develop a model which works specifically for the population of chemistry (lab) technicians who might be interesting in becoming BP-CPM—which is a small population. Thus, we only had a small sample size. We tried to address this challenge by a combination of expert knowledge and model design. In particular, experts manually assigned each pre- and post-test question as well as each part of the e-learning to one of six skills which enabled us to make dedicated predictions for each skill. Regarding model design, we assumed that the relationship between prior knowledge and knowledge gain is bell-shaped, that knowledge gain is never negative, and that knowledge gain only depends on factors related on the same skill, not other skills.

Importantly, while our model is nonlinear, it can still be visualized and interpreted by human experts in the form of knowledge gain curves (refer to Fig. 2), which enables our domain experts to judge whether the learned model is plausible or whether further changes need to be applied. In fact, we used such visualizations repeatedly during model development to arrive at our current version. Prior versions of our model did not use the bell-curve assumption or the non-negativity assumption and yielded visibly implausible models.

Our work provides some insight into the challenges and opportunities of applying educational data mining and artificial intelligence for vocational education settings. In particular, we conclude that it is necessary to record data and expertise that is applicable to the specific educational context and population at hand, for example, chemistry technicians who may be interested in becoming BP-CPM. In many cases, this means small sample sizes and highly specialized knowledge, which means that very data-efficient models with strong inductive bias have to be applied. We hope that our own case provides an example how to develop such a model for a practically relevant task—namely recommending a math preparatory e-learning.

3 Description of future steps

In future work, we intend to apply the learned model in practice for recommending the preparatory math e-learning to potential learners in the BP-CPM course. We will also monitor learners' feedback. We will also generalize the approach to other e-learnings, concerning topics such as chemistry, pharmacy, or process engineering, for courses such as the Bachelor Professional of Pharmaceutics (CCI).

Acknowledgements

Funding by the German Ministry for Education and Research (BMBF) under grant number 21INVI1403 (project KIPerWeb) is gratefully acknowledged.

References

Frank Baker. *The basics of item response theory.* ERIC Clearinghouse on Assessment and Evaluation, College Park, MD, USA, 2001. URL `https://eric.ed.gov/?id=ED458219`.

Celso Fulano, Paula Magalhães, José C. Núñez, Simone Marcuzzo, and Pedro Rosário. As the twig is bent, so is the tree inclined: Lack of prior knowledge as a driver of academic procrastination. *International Journal of School & Educational Psychology*, 9(sup1):S21–S33, 2021. doi:10.1080/21683603.2020.1719945.

Telle Hailikari, Nina Katajavuori, and Sari Lindblom-Ylanne. The relevance of prior knowledge in learning and instructional design. *American journal of pharmaceutical education*, 72(5), 2008. doi:10.5688/aj7205113.

Ronald Hambleton and Hariharan Swaminathan. *Item response theory: Principles and applications.* Springer Science+Business Media, New York, NY, USA, 1985. doi:10.1007/978-94-017-1988-9.

Ronald K. Hambleton and Russell W. Jones. Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3):38–47, 1993. doi:10.1111/j.1745-3992.1993.tb00543.x.

Philip I. Pavlik, Hao Cen, and Kenneth R. Koedinger. Performance factors analysis – a new alternative to knowledge tracing. In *Proceedings of the International Conference on Artificial Intelligence in Education (AIED)*, page 531–538, 2009. URL `https://eric.ed.gov/?id=ED506305`.

Harmen Schaap, Liesbeth Baartman, and Elly De Bruijn. Students' learning processes during school-based learning and workplace learning in vocational education: A review. *Vocations and learning*, 5(2):99–117, 2012. doi:10.1007/s12186-011-9069-2.

Jürgen Seifried and Hubert Ertl. Forschungsrichtungen zur künstlichen intelligenz in der beruflichen bildung. In Sabine Seufert, Josef Guggemos, Dirk Ifenthaler, Hubert Ertl, and Jürgen Seifried, editors, *Künstliche Intelligenz in der beruflichen Bildung: Zukunft der Arbeit und Bildung mit intelligenten Maschinen?!*, pages 341–347. Franz Steiner Verlag, Stuttgart, Germany, 2021. German.