Project title
Combining Machine-learning and Learning Analytics to provide personalized scaffolding for computer-supported learning activities.
Senior Researcher
Irene-Angelica Chounta

Project Summary

The **goal** of this project is to propose a methodological framework along with a computational, machine-learning approach for providing personalized guidance and feedback to students who engage with computer-supported learning activities.

To that end, we propose the use of a computational student model to assess the student's knowledge and cognitive state and to adapt the level of scaffolding accordingly. The student's knowledge state is assessed with respect to prior practice. To monitor and to model the student's prior practice, we employ Learning Analytics. Learning Analytics is defined as "*the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs*" (Siemens & Gasevic, 2012). In order to provide the appropriate kind and level of scaffolding with respect to student's specific needs, we follow the Vygotskian construct of the Zone of Proximal Development (ZPD) (Vygotsky, 1978) and the theoretical concept of Contingent Tutoring (Wood, 2003).

The adoption of technology in education has provided us with new tools and methods to support learning and teaching. Data richness offers new opportunities for testing analytical approaches to facilitate and inform new or revisited paradigms. Thus, there is the unique opportunity to employ cutting-edge computational approaches for addressing fundamental pedagogical challenges that remain open: how to adaptively guide students and how to provide appropriate scaffolding to facilitate learning and to improve the learning outcome.

However, despite the technological advances and the richness of resources, fundamental pedagogical questions remain open:

> what makes some students benefit from scaffolding and learn more than others, why some students give up even when they are given feedback and support and how we can provide students appropriate support in order to prevent failures (for example, dropouts or unsuccessful assignments) and turn them into successful learning episodes?

So far, machine-learning and data-mining approaches focus on assessing student performance using indicators of student's activity during a learning task. Based on these assessments, the learning environments that support the task – typically learning management systems, such as Moodle[1], or study information systems – will generate and send messages to students who are identified to be "in risk". These messages usually are reminders of deadlines that the student might have overlooked, notifications that the student failed to reply to some message by the

[1] https://moodle.com/

instructor and others. However, empirical research suggests that the Learning Analytics methods used to provide feedback are not based on established pedagogical strategies for instruction and scaffolding. On the contrary, they are commonly data-driven and have limited theoretical grounding (Gašević, Dawson, & Siemens, 2015). The lack of theoretical grounding can lead to providing inappropriate support that fails the purpose of scaffolding but instead inhibiting learning instead of enabling it: for example, providing the wrong amount of support (too much or too little), providing support on the wrong time (too late or too soon) or even providing the wrong kind of support (giving away the answer to a question when providing hints would be advisable).

The **open challenge** is to bridge the gap between pedagogical theory and practice when it comes to scaffolding. That is, to bring together established theoretical frameworks and modern practice that builds on computational and technological advances in order to create appropriate conditions for learning to take place and to help students learn by addressing their specific needs.

The Zone of Proximal Development (ZPD) is one of the best known theories in educational psychology, defined by Vygotsky as: "*the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers*" (Vygotsky, 1978). The definition of the ZPD indicates the importance of appropriate assistance in relation to the learning and development process. Thus, it can be stated more simply as "*the difference between what a learner can do without help and what he or she can do with help*"(Reber, 1995). Research suggests that when students are presented with material they already know (that is, the material is below students' ZPD), they tend to get bored. On the other hand, when students are presented with material that they are not able to understand (that is, the material is beyond students' ZPD), they get frustrated. Therefore, deriving ways to identify and formally describe the ZPD is an important step towards understanding the mechanisms that drive learning and providing appropriate pedagogical interventions to sustain students' engagement and promote achievement (Chaiklin, 2003).

The main objective of the research project is to study how we can combine computational models informed by learning analytics with established pedagogical theories in order to provide personalized scaffolding to students targeting their specific needs.
In particular, this project aims to provide personalized guidance and feedback to students by adapting scaffolding to their background knowledge and cognitive state. We propose the use of machine-learning student models to assess students' knowledge and cognitive state with respect to prior practice. To monitor prior practice, we propose the use of learning analytics. In order to maintain the most up-to-date representation of student's knowledge and cognitive state, the student models will be dynamically updated during students' practice. In order to provide guidance and feedback with respect to student's specific needs, we follow the Vygotskian construct of the Zone of Proximal Development and we adapt scaffolding with respect to the principles of Contingent Tutoring.

Our **research hypothesis** is that we can use learning analytics to design a student model that will describe the student's knowledge cognitive state. Then, we can use the output of the student model as a proxy to assess whether the student is - or, is not - in the Zone of Proximal Development. The core rationale is that if the student model cannot predict with acceptable accuracy whether a student will answer a question correctly, then it might be the case that the student is in the ZPD. Based on the student model's outcome - that is, whether the student is in the ZPD, above the ZPD, or below the ZPD - we can further plan the teaching strategy: what tasks to assign to the student, whether the student needs scaffolding and what kind of scaffolding is appropriate.

This rationale (known as the "Grey Area" approach) has been proposed and studied by the PI in the context of Intelligent Tutoring Systems(Chounta, Albacete, Jordan, Katz, & McLaren, 2017; Chounta, McLaren, Albacete, Jordan, & Katz, 2017). The "Grey Area" approach has been used by the Institute of Education Sciences (IES)-funded project Rimac[2], to guide students who used a natural-based dialogue system to practice conceptual physics (Albacete et al., 2018; Katz et al., 2018).

Novelty and impact

To the best of our knowledge, this is a novel approach for providing scaffolding in technology-enhanced learning environments by combining computational machine-learning methods with grounded pedagogical theories.

A key broader impact of this work is that it can support complex pedagogical decision making necessary for providing effective scaffolding. Once the proposed approach has been developed and vetted through efficacy testing it can be widely used in various contexts, such as online courses, MOOCs and collaborative learning environments. Furthermore, we envision that this approach will impact how we design learning activities and material, taking into account students' characteristics.

The proposed project could also contribute to the ever-present *assistance dilemma* (Koedinger & Aleven, 2007)—that is, the challenge of providing the right amount of help to the student so that the student is challenged but not frustrated.

Acknowledgements

This work is supported by the Estonian Research Council grant PSG286.

References

Albacete, P., Jordan, P. W., Lusetich, D., Chounta, I.-A., Katz, S., & McLaren, B. M. (2018). Providing Proactive Scaffolding During Tutorial Dialogue Using Guidance from Student Model Predictions. In *Artificial Intelligence in Education (AIED 2018)*.

Chaiklin, S. (2003). The zone of proximal development in Vygotsky's analysis of learning and instruction. *Vygotsky's Educational Theory in Cultural Context*, *1*, 39–64.

[2] https://sites.google.com/site/rimacsite/home

Chounta, I.-A., Albacete, P., Jordan, P., Katz, S., & McLaren, B. M. (2017). The “Grey Area”: A computational approach to model the Zone of Proximal Development. In *European Conference on Technology Enhanced Learning* (pp. 3–16). Springer.

Chounta, I.-A., McLaren, B. M., Albacete, P., Jordan, P., & Katz, S. (2017). Modeling the Zone of Proximal Development with a Computational Approach. In *International Conference on Educational Data Mining (EDM) 2017* (pp. 356–357).

Gašević, D., Dawson, S., & Siemens, G. (2015). Let’s not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64–71.

Katz, S., Albacete, P., Jordan, P. W., Lusetich, D., Chounta, I. A., & McLaren, B. M. (2018). What Does it Mean to Provide the Right Level of Support During Tutorial Dialogue? In *Proceedings of the 14th International Conference on Intelligent Tutoring Systems*.

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, *19*(3), 239–264.

Reber, A. S. (1995). *The Penguin dictionary of psychology*. Penguin Press.

Vygotsky. (1978). Interaction between learning and development. *Readings on the Development of Children*, *23*(3), 34–41.

Wood, D. (2003). The why? what? when? and how? of tutoring: The development of helping and tutoring skills in children. *Literacy, Teaching and Learning*, *7*(1/2), 1.