

PROJECT SUMMARY

Overview:

Online educational systems, and the large-scale data streams that they generate, have the potential to transform education as well as our scientific understanding of learning. Computer Science Education (CSE) researchers are increasingly making use of large collections of data generated by the click streams coming from eTextbooks, interactive programming environments, and other smart content. However, CSE research faces barriers that slow progress: 1) Collection of computer science learning process and outcome data generated by one system is not compatible with that from other systems. 2) Computer science problem solving and learning (e.g., open-ended coding solutions to complex problems) is quite different from the type of data (e.g., discrete answers to questions or verbal responses) that current educational data mining focuses on.

We propose efforts to build community and capacity among CSE researchers, data scientists, and learning scientists toward reducing these barriers and facilitating the full potential of data-intensive research on learning and improving computer science education. We bring together CSE tool build communities with learning science and technology researchers toward a software infrastructure that supports scaled and sustainable data-intensive research in CSE that contributes to basic science of human learning of complex problem solving. We propose a set of community-building and infrastructure capacity-building support whose ultimate goal is to develop and disseminate infrastructure that facilitates three aspects of CSE research: (1) development and broader re-use of innovative learning content that is instrumented for rich data collection, (2) formats and tools for analysis of learner data, and (3) best practices to make large collections of learner data and associated analytics available to researchers in CSE, data science, or learning science. To achieve these goals, we propose to engage a large community of researchers to define, develop, and use critical elements of this infrastructure toward address specific data-intensive research questions. We will host workshops, meetings, and online forums leveraging existing communities and building new capacities toward significant research outcomes and lasting infrastructure support.

Intellectual Merit:

Our project is the first attempt to design an infrastructure that can support various kinds of research in CSE domain as a one-stop-shop, and is the first to focus on full-cycle educational research infrastructure in any domain. If successful, CSE tool developers and educators will become more productive at creating and integrating advanced technologies and novel analytics. Learning researchers will have better tools for analyzing the huge amounts of learner data that modern digital education software produces. Data scientists will have rich new datasets in which to explore new machine learning and statistical techniques.

Collectively, these efforts can reduce barriers to educational innovation and support scientific discoveries about the nature of complex learning and how best to enhance it. We will support scientific investigations through community meetings and mini-grants to others addressing questions such as: What is the optimal ratio of solution examples and problem-solving practice? How do computational thinking skills emerge? In what quanta are programming skills acquired? Can automated tutoring of programming be effective at scale in enhancing student learning?).

Broader Impacts:

This proposal represents the first step toward building a community of practice that will broadly impact both computer science education and research on it. We aspire to have direct impact on enhancing scientific productivity of at least 20 computer science education researchers and at least 10 learning science researchers even at this early stage of the proposed work on the research infrastructure. Their discoveries and technological innovations will in turn help tens of thousands of students in the strategically important field of computer science. Many of the innovations proposed can directly impact learning in any discipline. Educational software will more quickly be developed in the future, that more easily generates meaningful learner data, which in turn can be more easily analyzed.