EDM and the 4th Paradigm of Scientific Discovery - Reflections on the 2010 KDD Cup Competition

John Stamper, Carnegie Mellon University, USA

Technology advances have made the ability to collect large amounts of data easier than ever before. These massive datasets provide both opportunities and challenges for many fields and education is no different. Understanding how to deal with extreme amounts of student data in the EDM field is a growing problem. The 2010 KDD Cup Competition, titled "Educational Data Mining Challenge", included data for over 10,000 students. The students completed over 30 million problem steps collected over a year long courses from Carnegie Learning Inc.'s Cognitive Tutors. We believe these are the largest educational dataset at this level of granularity to be released publicly. The competition drew broad interest from the data mining community, but it was also clear that many in the research community could not handle datasets of this size. In this talk, John will discuss the 2010 KDD Cup and the impact of larger and larger amounts of data coming available for educational data mining and how this will drive the direction of educational research in the future.