Interactive Data Mining Frameworks for Understanding the Impact of Funding Portfolios

Krishna Madhavan
School of Engineering Education
Purdue University

The past two decades have seen tremendous national investments by the National Science Foundation (and other federal agencies) in STEM research and educational programs at various scales. This funding is expected to deliver major scientific and learning breakthroughs. While common knowledge derived from relatively fragmented impact data indicates that such results are forthcoming, it is incredibly difficult to understand the true nature of the impact delivered by federal funding. Instrumenting communities in meaningful ways to derive deep insights about the return on such major national investments remains a frontier challenge that calls for a highly interdisciplinary and focused effort.

A significant amount of data derived from federal funding are not appropriately archived, indexed, and utilized to make operational or policy decisions. These data are also known as "dark data". This "dark data" are at extremely large scale (also known as big data), longitudinal, unstructured, and is growing at extremely high velocity. Furthermore, "dark data" hold a significant competitive key to not only scientific knowledge about how learning research and new programs can be designed and perform, but more importantly about how research productivity can be enabled and accelerated. Understanding the impact of federal investments is not just an academic exercise. Rather, we see this as a process that can have direct policy and operational impacts on how the larger community is funded in the future.

In this talk, we discuss – DIA2 – Deep Insights Anytime, Anywhere – a central resource for the NSF Education and Human Resources (EHR) community. Our approach combines theories of community formation, user-centered design, ultra large-scale data mining, social network analysis, and interactive visualization theories. We enable users to mine massive amounts of data and make sense of it using a highly intuitive process. The talk includes a full demo of the DIA2 Alpha system that interactively mines nearly 50 years of NSF-wide funding data. Parts of DIA2 are now deployed inside the National Science Foundation (NSF) firewall with initial user testing currently underway.

Speaker Bio:

Dr. Krishna Madhavan is an Assistant Professor in the School of Engineering Education at Purdue University. He is Co-PI and Education Director of the NSF-funded Network for Computational Nanotechnology (nanoHUB.org which serves over 325,000 global researchers and learners annually). Dr. Madhavan also served as Visiting Researcher at Microsoft Research (Internet Services Research Center). He was the Chair of the IEEE/ACM Supercomputing Education Program 2006. In January 2008, he was awarded the US National Science Foundation (NSF) CAREER award for work on learner-centric, adaptive cyber-tools and cyber-environments. He was one of 49 faculty members selected as the nation's top engineering educators and researchers by the US National Academy of Engineering to the Frontiers in Engineering Education symposium. Dr. Madhavan leads a major NSF funded effort called Deep Insights Anytime, Anywhere (DIA2) that attempts to characterize the impact of NSF and other federal investments in the area of science, technology, engineering, and mathematics education using interactive knowledge mining and visual analytics for non-experts in data mining. DIA2 is currently deployed inside the NSF and is already starting to affect federal funding policy. Dr. Madhavan's research has been published in Nature Nanotechnology, IEEE Transactions on Computer Graphics and Applications, IEEE Transactions on Learning Technologies, and several other top peer-reviewed venues. Dr. Madhavan currently serves as PI or Co-PI on federal and industry funded projects totaling over $20M.