GOERGEN INSTITUTE FOR DATA SCIENCE

FOCUS AREA: **METHODS, TOOLS, AND INFRASTRUCTURE**

A 15-month-old participates in a study led by Assistant Professor Celeste Kidd at the University of Rochester Baby Lab. The Baby Lab is a facility focused on early human development where researchers observe children while they watch movies or play and then analyze their eye-gaze patterns and brain activity to study how they think, learn, and make decisions.

▶ One of the University of Rochester's priorities within its data science initiative is to develop the computational methods, tools, and infrastructure that will drive discovery in science, business, medicine, the humanities, and virtually every other field. This computational framework will allow researchers to efficiently collect large amounts of data, manipulate it, and then surround it with the right computing resources and statistical expertise to make sense of it.

The University is uniquely positioned for success due to having faculty members who can build the systems and others who can apply it to their fields of research.

The University at Work

Computer scientists, electrical and computer engineers, and statisticians are behind much of the data science research being done in a variety of fields.

Parallel Computing

Sandhya Dwarkadas, a professor and chair of computer science and a professor of electrical and computer engineering; Michael Scott, a professor of computer science; and Srinivas Aluru, a professor at the Georgia Institute of Technology, are developing computational systems that will drive a variety of clinical outcomes and processes, including gene sequence and protein structure discovery.

Their focus is on designing software and hardware techniques that are portable and scalable to effectively execute tasks in parallel. This is challenging work—the more processors there are working together on a problem, the more difficult it becomes to coordinate their activities and run them in parallel efficiently. Without sufficient care, the processors tend to "step on each other's toes."

Together, these researchers are building systems that know how to balance data and computation across a large number of processors and memories. IBM's Blue Gene/Q supercomputer at the University's Health Sciences Center for Computational Innovation is an important platform for their work.

Statistical Analysis

Robert Strawderman, a professor and chair of biostatistics and computational biology, is laying important statistical groundwork to facilitate medical research. He and his team of 25 faculty members engage in both methodological and collaborative research and educate others at the University on the use of statistics in their work.

"This department is involved in all stages of data collection and analysis," says Strawderman. "This includes developing clinical trials of new neurological therapies and in cardiovascular disease, predicting suicide risk among men and women in the United States Air Force, examining how genomic influences can affect human health in those with cancer and for other diseases, and much more."

By designing studies and developing state-of-the-art methods and computational tools, Strawderman and his team are helping medical researchers here successfully query, analyze, and understand the increasing amounts of data now available.

Research Applications

Those applying the data science analytics to their research are experiencing new levels of understanding and contributing new knowledge to their fields.

In Business

After examining Twitter data from major U.S. airlines, Simon Business School researchers Huaxia Rui and Abraham Seidmann found that companies are more likely to engage with customers who have a high number of followers. The study, which was featured in *USA Today*, shows that airlines answered less than 40 percent of complaints and compliments they received through Twitter and that they responded more often to users with a high number of followers, although the degree of favoritism varied by airline.

In Medicine

Kelly Conn from the School of Nursing, Wendi Heinzelman from the Department of Electrical and Computer Engineering, and several undergraduates are developing an Android app for parents of children with asthma. The app will provide resources and support services tailored to each family's circumstances and needs. For instance, it could point people to the nearest place to receive a HEPA filter and a map to get there. Heinzelman says that the framework of providing resources and support based on the specific context of the user and the current disease conditions could be applied to other medical conditions as well. Eventually, she says, the team plans to submit the project to the National Institutes of Health for a grant to further explore the app's potential.

In Child Development

Celeste Kidd, an assistant professor of brain and cognitive sciences, combines computational modeling with traditional methods to study how infants and children allocate their attention to the world. Kidd runs behavioral tests in her lab with children, and she uses eye-trackers for infants to identify what mechanisms they employ to learn. For instance, she might show infants a display featuring three boxes that pop up at different times. The eye-trackers measure where infants focus their attention. Kidd then analyzes this data to understand, predict, and even optimize learning environments.

How You Can Help

Endowment support for professorships, research, graduate fellowships, and undergraduate scholarships enable the continued development of computational methods, tools, and infrastructure, the critical building blocks needed for further exploration and advancement in a multitude of fields. Please consider supporting this important aspect of data science.

For more information, contact
Stephen A. Dare
Senior Associate Vice President
University Advancement
(585) 275-7530
stephen.dare@rochester.edu

Learn, Discover, Heal, Create—And Make the World Ever Better
www.rochester.edu/data-science