Mutually Enhancing Human Strengths and Computational Power in Future Scientific Methodologies

Chad A. Steed
Computer Science and Mathematics Division
Oak Ridge National Laboratory

"Nothing – not the careful logic of mathematics, not statistical models and theories, not the awesome arithmetic power of modern computers – nothing can substitute here for the flexibility of the informed human mind … Accordingly, both [analysis] approaches and techniques need to be structured so as to facilitate human involvement and intervention."

-John W. Tukey & Martin B. Wilk, Data Analysis & Statistics, 1966

With growing data volumes, faster computational resources, and expanding user facility capabilities, the opportunity for data-driven discovery today and in the near future is staggering. These advances have enabled new advances in artificial intelligence, quantum computing, and edge computing, which are drastically changing the nature of scientific methodologies. At the same time, the role of scientists during data analysis is evolving as redundant processes are automated and computational results provide scent that guides information foraging tasks. However, the probability of uncovering scientific breakthroughs ultimately lies in the ability of empowered scientists to pursue questions, uncover domain-specific insights, and effectively communicate knowledge. Thus, the sentiments of Tukey and Wilk are still relevant today: we must "facilitate human involvement and intervention" during all stages of scientific analysis. Designing efficient ways to preserve human involvement in data-drive scientific investigations will be a significant topic in the near future.

Although computing advances and the availability of more data have made previously theoretical approaches to automatically analyze data possible, the exploratory nature of most scientific investigations highlights the importance of ensuring humans are empowered to direct analytical workflows. Successful evolution to a scientific workflow that integrates human interactions will help preserve the chance of making unexpected discoveries, which often lead to the most impactful breakthroughs, and help avoid potential bias in algorithms that are susceptible to repeatedly finding the same patterns. Fortunately, methods to display information and capture human interactions are also increasing. New human-centered technologies, such as augmented reality and mobile platforms, offer more natural interactions that far exceed the low bandwidth interfaces of traditional workstation setups. Efforts to bring these advances into scientific workflows are in their infancy, but the demand continues to grow as barriers for entry are removed and new generations of scientists who are daily immersed in them take the reins. As these technologies continue to mature and permeate our everyday lives, scientific analyses can benefit from new interactive experiences that exceed even the futuristic visions portrayed in science fiction.

As scientific methodologies evolve at Department of Energy Office of Science national laboratories over the new three decades, we should consider effective ways to capitalize on new human interaction techniques to couple and mutually enhance the strengths of humans and computational machinery. In addition to techniques that leverage new modes of interaction for asking questions of data and augmenting algorithmic processes, investments into techniques that actively mine human interactions to automatically learn features of interest and guide scientists to similar features, especially in unexamined portions of the data, will be critical to enabling comprehensive investigations that take full advantage of the data volumes at hand. These technologies will be crucial to validating computational results, wrangling raw data, and communicating scientific insights. By building on foundational research in human-computer interaction, perceptual and cognitive psychology, and interactive data visualization, new degrees of human machine cooperation will help scientists magnify the value of both data and harness new advances in computing to address grand national challenges in a myriad of domains.