Pervasive, Personalized and Precision (P^3) analytics for massive bio-social systems

A white paper for the Big Data and Extreme-scale Computing workshop, November 2018

Madhav V. Marathe, Christopher L. Barrett
Biocomplexity Institute and Initiative & Department of Computer Science.
Email: {marathe,clb5xe}@virginia.edu

Motivation: We are concerned with the range of conceptual, theoretical, technical and use issues related to providing actionable information to individuals related to at-scale, massively interacting, bio-social/technical systems. These co-evolving environments are often very large but not large enough so that they can be approximated using mean field theories. They are comprised of intricately networked biological, social, informational, technological and infrastructural components. Examples abound: from natural and artificial ecosystems, including mega-cities and natural ecologies to human and animal immune systems. Such systems are characterized by decentralized coordination, adaptation and memory, heterogeneity and non-stationarity with various time-varying interactions. Perhaps most importantly, these networked systems co-evolve as a result of the dynamical interactions within and amongst them. At any time system knowledge is provisional and subject to revision, whether this knowledge involves historical, current or projected system behavior. As a result "best" estimates or "best" actions, in general, will change over time, experience and are sensitive to predecessor state histories. For example, the structure of a mega-city is not just dictated by legal framework and centralized governing bodies, but also and largely dictated by the various simultaneously co-defined functional networks of its inhabitants. This sort of decentralized, yet agentic, functionality and structure is constantly evolving at multiple time, space and social scales. Technologies ranging from pharmaceuticals to transportation and communication create ever new layers for functional embodiments, interaction and phenomenological innovations. Advances in computing, information sciences and sensor systems have led both to the emergence of entirely new social phenomena arising from massive interaction with the cognitive, biological and physical environments in the ICT layers and to the possibility of pervasive, personalized and precision analytics and decision making within, for, and among these systems. Such services need to be understood and supported at various levels of organizational hierarchy – from individual constituent elements to systemic level wherein certain centralized or understandable coordination of decentralized decision making might be possible.

Computational challenges: Interestingly, while advances in computing, communication and information technology has fueled the desire and vision of P^3 analytics --- this vision brings forth new challenges far greater than simple implementation and fielding of the individual computational technologies. Many of the challenges are beginning to be perceived and articulated in the so-called Big Data and extreme-scale computing worlds and certainly are applicable here. Some salient ones that are worth noting here include: (*i*) distributed, pervasive decision making in real-time and multiple stake holders – leading to systems that constantly sense, act and adapt in response to the bio-social system; (*ii*) inability to conduct typical experiments at scale that can often be done for physical systems (e.g. large scale disaster scenarios) – leading to the need for developing at-scale computational models and experiments to study such systems (*iii*) lack of fundamental and well accepted theories that describe how such systems work (one could argue that there is really no universal law describing human mobility – indeed movements will coevolve with the social structures we create) – leading to development of models that are informed by theories but also by observed data; and (*iv*) notion of validity and reproducibility is likely to be substantially different than physical sciences – leading to design and development of systems that are done so for a reason. Additionally we believe it is centrally important that we must actively develop in adversarial information environments, adopt a provisional knowledge perspective, and take very seriously the natural properties of spontaneously appearing and engineered decentralized agency.

An abductive multi-scale systems to support P^3 analytics and decision making: One critical aspect of provisional knowledge systems is the problem of imputing unclear and multi-scale complex system state for purposes of system management and the problem of *adapting action related knowledge*. We envision the need for *abductive systems* that support state and action reasoning for decision making in massive biosocial systems environments. Such systems, will need to constantly sense the environment , and modify their internal representation of the system based on the observations. They will also need to revisit their actions against current knowledge. In other words, decision making, sensing and analytics is done based on a provisional model and knowledge of the underlying system that evolves in time. We do not use the term incomplete knowledge (or information) since in many cases, the system itself is adapting and hence there is no clear notion of complete knowledge. Indeed there is no asymptotically true state over time in general for these rather radically nonstationary environments,. This view is inspired by a cognitive view of how natural decision making systems work. Such a view immediately reconciles with the fact that information is inherently noisy, incomplete and often manipulated. The notion of validity and best description of the system is based on the notion of abduction. It acknowledges that larger outcomes are really the result of individual decisions and outcomes and in this sense predictive or retrospective validity cannot be the only measure of success. One will have to extend the notion of abduction to distributed, local and hierarchical abduction: abduction by distributed agents for estimate their environment; abduction by agents based largely on local information and finally abduction at various organizational levels.