2015 NSF Workshop On Intelligent Systems For Geosciences

Victor Pankratius Massachusetts Institute of Technology, Haystack Observatory
victorpankratius.com, pankrat@mit.edu

Computing trends that may have an impact on geosciences research

1. Parallelism Everywhere: Multicore and Cloud Computing Today's laptops, servers, and mobile devices are all parallel computers. Technical and economic factors are limiting increases in processor clock speeds, so performance scalability of scientific applications has to come through parallelism. Scientific applications need to be able to take advantage of multicore/manycore CPUs and GPUs. On a larger scale, cloud computing now provides affordable on-demand, configurable infrastructures for high-performance computing that can be seamlessly integrated into geoscience analysis environments. The growing complexity of scientific applications requires intelligent software tools that make the engineering process more efficient, so scientists have more time to do science. Interesting directions include automated performance tuning (auto-tuners), platform-specific code generation (for numerical codes), testing and validation (can we trust scientific results that were computed on a variety of nodes that may change in every run? What happens if floating point precisions are mixed up?).

2. Machine Learning Sensor cost is decreasing due to technological progress, miniaturization, and economies of scale. However, insight generation from large data collections is now a bottleneck and becomes more expensive. Machine leaning can help, particularly in early explorative analysis. Another challenge is data diversity: interesting phenomena can be observed with different instruments (GPS, radar, satellite, and others). Feature detection algorithms should therefore be able to process a variety of data and leverage cross-correlations to minimize false positives/negatives. Geosciences provide an fruitful ground for experiments with large-scale machine learning on real-world data, with important societal implications.

3. Mobile Computing Phones and tables have become versatile computing platforms that integrate multicore CPUs/GPUs, storage, and network connectivity (WiFi, cell data network, Bluetooth). Compared to individual embedded systems developed for sensor networks, platforms like Android are better standardized, increase development productivity through software reuse (e.g., libraries), and are more likely to be supported over a longer period of time. Using extensible commodity hardware, if possible, can free up budget for scientific activities. Mobile devices should be considered for data collection and pre-processing in the field, as well as for data transport to cloud environments that handle more sophisticated processing. Crowdsourcing data collection is another new avenue enabled by mobile phones.

Research trends with relevance to workshop goals

1. Mobile Devices in Scientific Data Collection, Computing, and Discovery Mobile devices are enablers for new avenues in geoscience. For example, phones can be easily extended with additional sensors; they offer respectable computing power that can be leveraged for real-time data compression, pre-processing, noise reduction, signal filtering, trend analysis, and more. There are opportunities for crowdsourcing and engagement of citizen scientists (e.g, to help with classification or other tasks). All of these aspects represent additions to cyberinfrastructure capabilities. For a current example of such a project, see the NSF-funded Mahali project at http://mahali.mit.edu

2. Computer-Aided Discovery Enhanced environments for computer-aided discovery will support humans in the search process of making new discoveries while optimizing the division between machine-supported and human-supported tasks. This is currently ongoing work at MIT Haystack, sponsored by the NSF.

Creating new trends; NSF influence

Many of the current challenges require greater commitment to interdisciplinary research. Funding agencies could request that proposers have teams with interdisciplinary backgrounds (e.g., computer science and geoscience). Furthermore, universities need to commit to hiring more faculty with interdisciplinary backgrounds; this can be incentivized through appropriate rules in CAREER awards, center awards, or through other types of awards. Education science awards could focus on interdisciplinary teaching and curricula for next-generation (geo)scientists.