SEG 100 Conference: Celebrating a Century of Discovery

F5
Machine Learning in Earth Science Is Hard; Now What?

Matt Hall
Agile Scientific, Mahone Bay, NS, Canada

Practitioners are discovering that applying machine learning to complex systems like the earth is difficult, but what does that imply about technical communities and earth science education? I think we have three tasks:

- Raise our community's digital literacy level.
- Build a toolkit around open code and open data.
- Grow a culture of open collaboration and entrepreneurship.

We must urgently raise the level of digital literacy in subsurface science and engineering, both among practitioners and students and among the people who train, manage, QC, and regulate. Our community needs formal educational efforts, both at universities and in the workplace, but we also need accessible short courses, tutorials, and hackathons. Technical societies have a big role to play in this retooling effort, if they want it. One or two are already making a difference.

Alongside education, we must grow the ecosystem of open technology and open data with which people are building, training, and evaluating new tools, especially machine learning tools. Awareness and understanding of open-source software, experimental replicability, open data licensing, and almost every aspect of this domain are quite low, especially in the corporate domain (as the continual modification of Equinor's Volve data set licence illustrates). This growth must happen among academics, corporate practitioners, entrepreneurs, and government agencies.

Digital literacy and an open toolkit are necessary but insufficient conditions for a high-functioning community of digital innovators. They also need to be connected with others with similar and complementary goals. We must seek out ways to connect people at every scale, from teams to global communities.

Building on the foundation afforded by the previous three points, we will be able to more effectively deepen our scientific ideas, the developing theoretical framework, and our applied practice in industry. These ideas include

- Novel neural network architectures;
- Ways to develop data-driven models alongside physics-driven models;
- The relative importance of natural and synthetic data in training sets;
- The design of cost functions that reflect what we know about the earth.

Participants at a subsurface hackathon, Copenhagen, 2018.


#surfhack