DISASTER RELIEF

Schuler's team discovering grammar rules of lesser-known languages

It is estimated there are more than 7,000 languages worldwide. For those involved in disaster relief efforts, that breadth and variety can be overwhelming, especially when addressing areas with meager resources.

William Schuler, Ph.D., linguistics professor at The Ohio State University, is part of a project called Low Resource Languages for Emergent Incidents (LORELEI), an initiative through the Defense Advanced Research Projects Agency (DARPA). LORELEI's goal is to develop technology for languages about which translators and linguists know nothing.

Schuler and his team use the Ohio Supercomputer Center's Owens Cluster to develop a grammar-acquisition algorithm to discover the rules of lesser-known languages so disaster relief teams can react quickly.

"We need to get resources to direct disaster relief and part of that is translating news text, knowing names of cities, what's happening in those areas," Schuler said. "It's figuring out what has happened rapidly, and that can involve automatically processing incident language."

Schuler's team is working to build a Bayseian sequence model based on statistical analysis to discover a given language's grammar rules. It is hypothesized this parsing model can learn a language and make it syntactically useful.

"The computational requirements for learning grammar from statistics are tremendous, which is why we need a supercomputer," Schuler said.

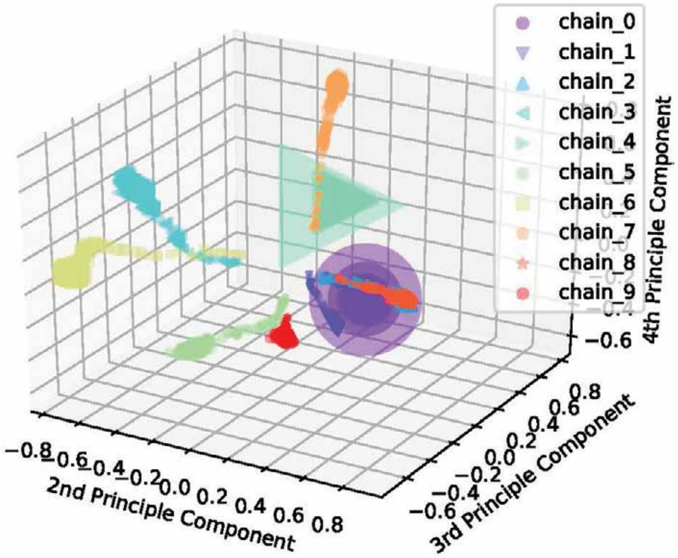
On a powerful single server, Schuler's team can analyze 10 to 15 categories of grammar, according to Lifeng Jin—a Ph.D. student who oversees the computational aspects of the project. GPUs on the Owens Cluster allow Jin to increase the number of categories greatly.

GPUs–graphics processing units—are a complementary processing unit to CPUs—central processing units, composed of hundreds of cores that can handle thousands of threads simultaneously.

"We can increase the complexity of the model exponentially," Jin said. "It's a more realistic scenario of imitating what humans are doing."

In August, DARPA organized a trial run to simulate two disasters in Africa. Schuler's group used 60 GPUs on Owens for seven days for four grammars of two languages, illustrating the importance of OSC's resources to the project.

"We're answering fundamental questions about what it means to be the animal that talks to each other," Schuler said. "The ability to ask these questions and get answers is a relatively recent innovation that requires the high performance computing infrastructure OSC gives us. It's a game-changer." ◀



This graph displays an algorithm that explores the space of possible probabilistic grammars and maps out the regions of this space that have the highest probability of generating understandable sentences.

PROJECT LEAD // WILLIAM SCHULER, PH.D., THE OHIO STATE UNIVERSITY RESEARCH TITLE // COGNITIVELY BASED UNSUPERVISED GRAMMAR INDUCTION FOR LOW-RESOURCE LANGUAGES FUNDING SOURCE // DARPA WEBSITE // LING.OHIO-STATE.EDU/~SCHULER.77