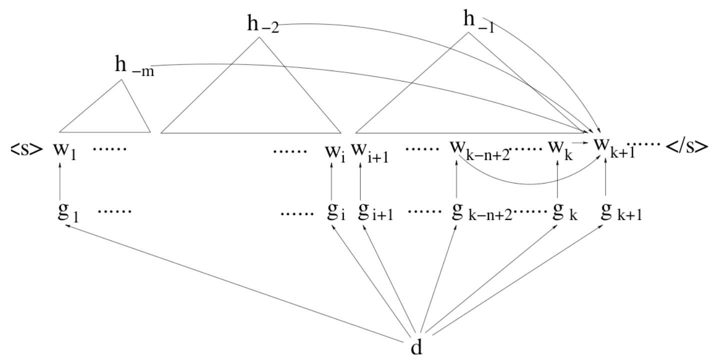
Building a better statistical language model



The highly ambiguous nature of natural language presents many challenges to researchers who design software to analyze, understand and generate languages that humans use naturally. Shaojun Wang, Ph.D., an assistant professor of Computer Science and Engineering at Wright State University, is leveraging the resources of the Ohio Supercomputer Center to build a statistical language model that will capture various kinds of regularities of natural language to improve the performance of a range of natural language processing applications.

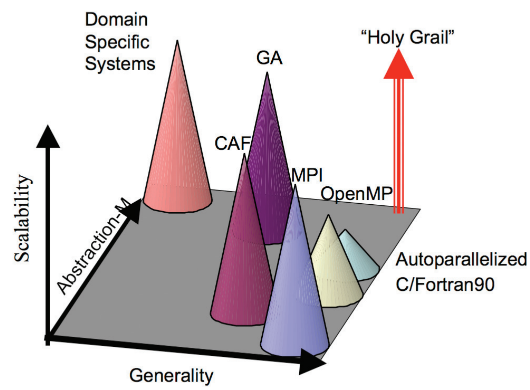
"It has long been a challenge in statistical language modeling to develop a unified framework to integrate various language model components to form a more sophisticated model that is tractable and also performs well empirically," said Wang. "Natural language encodes messages via complex, hierarchically organized sequences. The local lexical structure of the sequence conveys surface information, whereas the syntactic structure, which encodes long-range dependencies, carries deeper semantic information."

By exploiting the particular structure of various composite language models, Wang can decompose the seemingly complex statistical representations into simpler ones; this enables the estimation and inference algorithms for the simpler composite language models to become internal building blocks for the estimation of complex composite language models, thus finally solving the estimation problem for extremely complex, high-dimensional distributions.

To evaluate the performance of the composite language models in a real-world scenario, Wang and his colleagues will embed their new models in large-scale machine translation systems to remove language barriers to processing human communication. ■

Project lead: Shaojun Wang, Wright State University
Research title: Exploiting syntactic, semantic and lexical regularities in language modeling
Funding source: National Science Foundation, Google, Wright State University

Promoting PGAS model for petascale computers



The Partitioned Global Address Space (PGAS) programming model has attracted considerable attention in HPC circles, primarily because it offers application programmers the convenience of globally addressable memory along with the locality control needed for scalability. This makes the PGAS models an attractive alternative for petascale computing systems, compared to the hybrid MPI/OpenMP model.

To further develop the PGAS model, Ponnuswamy Sadayappan, Ph.D., professor of computer science and engineering at The Ohio State University, is developing a collaborative research group involving his team and staff members of the Ohio Supercomputer Center (OSC).

"Although some significant production applications like NWChem have been implemented using the Global Arrays PGAS model, the overall use of PGAS models for developing parallel applications has been very limited," said Sadayappan.

"Several advances are needed for PGAS to become more widely adopted, including compiler/runtime support for enabling existing sequential codes to be transformed and new parallel applications to be developed using PGAS models. The research collaboration with OSC provides significant opportunities to work with the developers of production applications to implement and evaluate our research advances in compiler/runtime systems for PGAS computing."

The group also will explore the integration of several popular HPC programming languages with the PGAS model, according to Ashok Krishnamurthy, interim co-executive director of OSC. Sadayappan has worked with OSC on a number of funded projects, including development of high-performance parallel software for electronic structure calculations and compiler/runtime optimization techniques for multicore CPUs and GPUs. ■

Project lead: Ponnuswamy Sadayappan, The Ohio State University
Research title: A Parallel Global Address Space framework for petascale computing
Funding source: The Ohio State University, Ohio Supercomputer Center