Sarah Webster
University of New South Wales

Topic: Increased data distribution awareness within a compiler

Using irregular data in data-parallel programs is a popular method used in scientific computing. These problems are often executed on distributed-memory systems, to take advantage of the favourable computing power/cost ratio.

Current data distribution techniques, combined with load balancing strategies, can be used to influence and improve performance. However these techniques are limited when using irregular data, such as sparse matrices. The main problem is that many strategies can only be implemented during runtime, limiting the optimisations available.

This research proposes a new framework that provides greater awareness of the data distributions into a compiler able to handle irregular data. The result is that the compiler can make more decisions in terms of load balancing strategies, providing more techniques to improve performance.

Specifically, data distribution types are extended, and categorised into known, unknown, ordered, unordered, balanced and imbalanced groups. Common characteristics within a group will lead to similar load balancing strategies being effective for a particular data distribution. Using this information, the compiler can make more load balancing decisions based on the data distribution, removing this burden from the runtime system. In addition, these extended data distributions will affect the behaviour of the primitive data-parallel functions, such as scan and pack, and these changes are described.

Preliminary benchmarks show that using this framework leads to an increase in performance. Finally, the limitations of such a framework, for example in comparing two instances of actual distributions, are also outlined.