International Conference on Hydroinformatics

2014

Visual Analytics Clarify The Scalability And Effectiveness Of Massively Parallel Many-Objective Optimization: A Groundwater Monitoring Design Example

Patrick M. Reed

11th International Conference on Hydroinformatics
HIC 2014, New York City, USA

VISUAL ANALYTICS CLARIFY THE SCALABILITY AND EFFECTIVENESS OF MASSIVELY PARALLEL MANY-OBJECTIVE OPTIMIZATION: A GROUNDWATER MONITORING DESIGN EXAMPLE

PATRICK REED (1) AND JOSHUA KOLLAT (2)
(1): School of Civil and Environmental Engineering, Cornell University, 211 Hollister Hall, Ithaca, NY 14853, USA
(2): DecisionVis, LLC, 315 S. Allen St., Suite 321, State College, PA 16801, USA

In this study, we contribute a comprehensive assessment of massively parallel multiobjective evolutionary algorithm (MOEA)-based search using a highly challenging optimization—assimilation application. The application focuses on a four objective groundwater monitoring application in which parallel scalability is tested across compute core counts ranging from 64 to a maximum of 8192. Our many-objective visual analytics framework clarifies how to assess and attain highly effective search on large scale high-performance computing systems. Moreover, our results also refute the common assumption that adding objectives in many-objective evolutionary optimization will always increase the computational difficulty of a problem. Our results agree with prior theoretical analysis demonstrating several instances where the overall four objective groundwater monitoring problem formulation is actually easier to solve than lower dimensional formulations composed of subsets of the original formulation's objectives. Although a groundwater application is used to demonstrate our parallelization, the visual analytics and metrics utilized to characterize the parallel scalability of MOEA-based search are broadly applicable in water resources and beyond.

OVERVIEW OF KEY CONCLUSIONS

Our results show perfect naïve speedup efficiency for up to 1024 cores and as would be expected from theory, algorithm overhead and communication costs begin to degrade speedup for higher core counts. This work emphasizes that while MOEAs should not be seen as "embarrassingly" parallel, their scalability does directly benefit from solving more computationally demanding applications. The term naïve speedup refers to scalability assessments that only consider how parallelization increases the number of function evaluations that can be completed in a given period of wall clock time without assessing how well the MOEA exploits them. Our results demonstrate how search metrics can used to move beyond naïve speedup assessments and more rigorously benchmark the quality of search. Our results show several instances where it is highly doubtful that the serial version of the algorithm could ever be tractably benchmarked for our highest attained metric values. This observation implies that a classical evaluation of speedup relative to a serial baseline would yield superlinear speedups due to the evolutionary search dynamics.

This study also demonstrates the value and importance of augmenting metrics-based evaluations of parallel MOEA search with interactive visual analytics. Our results demonstrate how visualization can clarify when an MOEA's search shifts from "translating" the approximation set to "diversifying" its coverage over the extent of the objectives. This is an important observation. If shorter run durations are required, the rapid early translation of the set may yield a reasonable approximation of the Pareto approximate set where further search is unnecessary. Although in this study we are showing visual analytics of the approximate set's evolution as a post run analysis, in real operational contexts our visual analytics can be used to show intermediate movies of search progress. Users can exploit these movies to understand how the set is translating the space and diversifying across the extents of the Pareto approximate surface. It can then be an expert driven judgment if further computational effort is justified or decision relevant.

State-of-the-art MOEAs have the potential to address up to a theoretical limit of 10-objectives. We show analytically that solving a single 10-objective formulation implicitly addresses the tradeoffs for 1,022 subproblems. Given that compute time on national HPC resources are allocated on a competitive basis, many-objective formulations serve as a highly efficient mechanism for exploring a maximal number of tradeoffs for increasingly complex water resources systems problems. Additionally, when many-objective problem formulations are coupled with effective visual analytics, decision makers can avoid "myopic" decision errors that result from highly aggregated and narrow definitions of optimality. Our results demonstrate how many-objective visual analytics has value for assessing the value and challenges posed in exploiting HPC in optimization—assimilation or other computationally intensive optimization efforts. Unlike prior approaches to water resources planning and management, massively parallel many-objective visual analytics benefits in its parallel scaling and range of tradeoffs explored with increases in application complexity (i.e., function evaluation times and the number of objectives considered). Our study contributes a promising computational platform for providing rapid and evolving feedbacks between science, engineering, and decision makers for problem complexities relevant in a nonstationary and rapidly changing world.