Knowledge Support for Parallel Performance Data Mining

Kevin A. Huck
Performance Research Laboratory
Computer and Information Science Department
University of Oregon
Motivation

• Parallel performance tool research has been limited in its contributions in:
  • Large-scale performance data management
  • Large-scale performance data analysis
  • Automated performance investigation
  • Knowledge-based performance problem reasoning
Contributions

• Parallel performance data and metadata definition and management
• Design of a systems framework to support performance data mining
• Data mining algorithms applied to parallel performance analysis
• Techniques for flexible, extensible automation of data analysis and data mining
• Techniques to incorporate expert knowledge through inference rules
Implementation

• Development of prototype tools for performance data / metadata management and performance data mining

• PerfDMF

• PerfExplorer v1 / v2

• Application of PerfDMF and PerfExplorer to real parallel performance analysis studies and evaluation
General Purpose Parameter Recommendation System

Offline Analysis

Runtime Recommendation
Classifier Construction

1. Find "unique" tuples of properties
2. Find which tuple parameters are "significant"
3. Method for each tuple
4. Find "optimal" method for each tuple
5. Significant Parameters
6. Weka Classifier Construction
7. Trained Classifier
PETSc driven cavity flow

- Non-linear solution of sparse matrices
- Selection of solver and preconditioner affects runtime, success
- Need recommendation for solver and preconditioner
- 3684 training instances created, 521 and 314 unique tuples found (9 and 10 parameters)
- 4*2 classifiers constructed
  - J48
  - Multilayer Perceptron
  - Naïve Bayes
  - Support Vector Machine
- 64% of default runtime
Thanks

TAU, PerfDMF, PerfExplorer:
http://tau.uoregon.edu
email: khuck@cs.uoregon.edu

Acknowledgements:
US Department of Energy, Office of Science
National Science Foundation