Iterative Solver Selection Techniques for Sparse Linear Systems

Kanika Sood[1], Boyana Norris[1], Elizabeth Jessup[2]

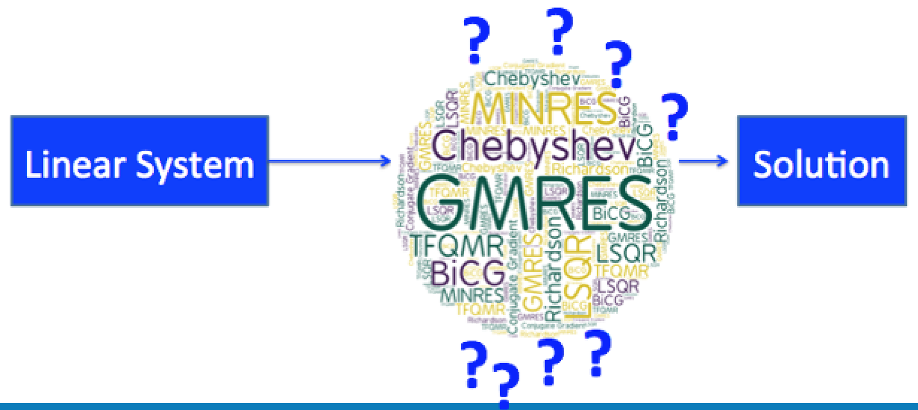
University of Oregon[1], University of Colorado Boulder[2]

UNIVERSITY OF OREGON

University of Colorado Boulder

PROBLEM STATEMENT

High-performance numerical frameworks rely on optimized packages (e.g. PETSc [1]) for the numerical solution of nonlinear PDEs. The growing number of potential solution methods makes the selection of a good solver configuration increasingly challenging.



GOAL

Enable users to choose solver configurations that are likely to perform well for a given sparse linear system.

CONTRIBUTIONS

- A machine learning (ML) workflow for classifying preconditioned Krylov solvers.
- An analytical communication model for large scale distributed-memory resources.
- Low-overhead linear system feature computation.
- A new approach to computing matrix features in matrix-free PDE applications.

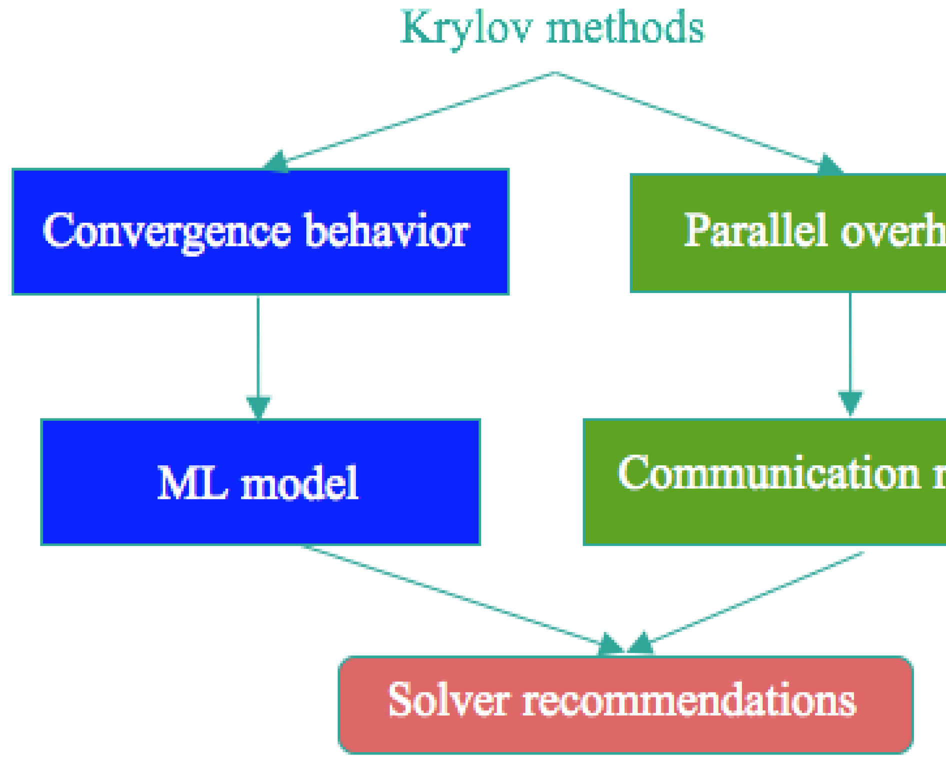
MATRIX FEATURES

- Linear System characteristics (e.g., number of non-zeros, matrix norm).
- Categories: Structural, size-based, spectral.
- The features are the input to the convergence model.
- Explicit matrix: Entire matrix is stored.
- Matrix-free: Matrix is not stored explicitly, instead matrix-vector product approximations are approximated using function evaluations.
- Compute features using PETSc.

International Conference on Parallel Processing (ICPP) 2018

APPROACH SUMMARY

1. For any problem size, first use the convergence model [2] to classify solvers.
2. For large problems:
   - Analyze the communication overheads of Krylov methods to generate a communication-based solver ranking.
   - Combine the convergence model and the communication-based ranking [3].
3. Generate solver recommendations.



RESULTS

Dataset: MOOSE [4] features are extracted with the matrix-free approach. Full feature set contains 27 features and reduced set 1 (RS1) has only 7 features.

Size: 4,845 data points

Training set: 3,875 data points

- 2,035 `good` class labels
- 1,844 `bad` class labels

Test set: 970 data points

- Best convergence model: J48 [5]
- Accuracy: 82.0%

RS1 features
Minimum non zeros/row
Lower Bandwidth
Non Zero Pattern Symmetry
Infinity Norm
Column Variance
Diagonal Non Zeros
Diagonal Average

Confusion matrix for J48 (C4.5 decision tree) (RS1)

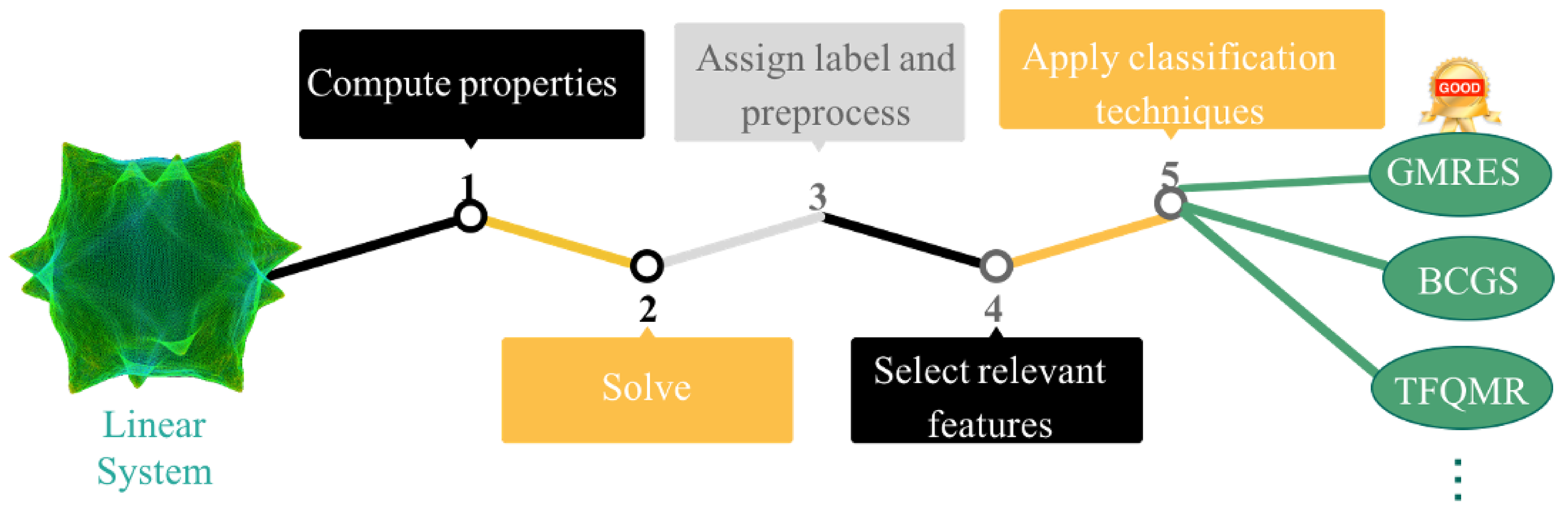
-----> Predicted Labels; -----> Actual Labels

10 fold cv	good	bad
good	1843	188
bad	462	1382

-----> Predicted Labels; -----> Actual Labels

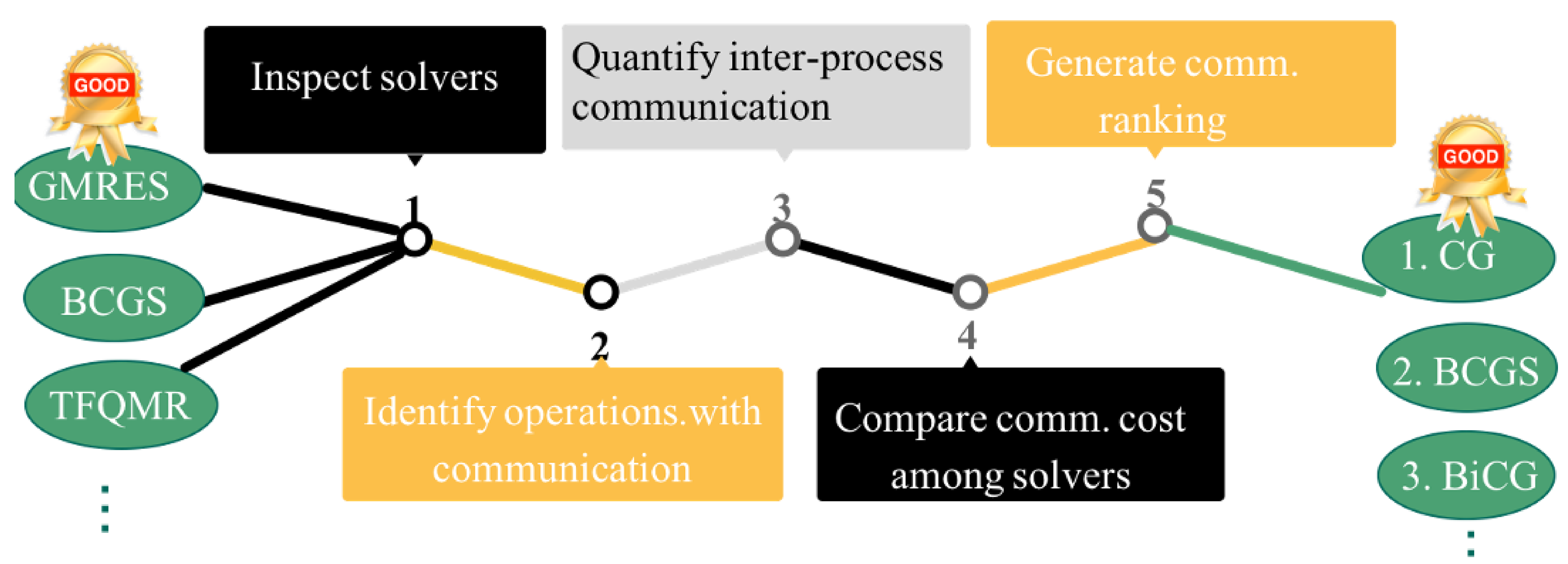
66%-34%	good	bad
good	620	58
bad	169	470

APPROACH: SMALL SCALE



A low-overhead technique for selecting solution methods effectively, by training the model and suggesting solvers based on the input problem characteristics.

APPROACH: LARGE SCALE



A new communication model for capturing performance at different parallelism scales that generates a ranked list of solver suggestions.

CONCLUSION

- Demonstrate a matrix-free approach for ML-based selection of preconditioned Krylov methods in the context of PDE applications.
- **Enable solver recommendations for sparse linear systems at different scales.**
- Future work: Expand our convergence model dataset to include more matrices, use in more applications.

REFERENCES

[1] Balay, Satish, et al. PETSc web page, http://www.mcs.anl.gov/petsc/ , 2018.

[2] Jessup, Elizabeth, et al. Performance-based numerical solver selection in the Lighthouse framework. SIAM Journal on Scientific Computing, 2016.

[3] Sood, Kanika et al. Comparative Performance Modeling of Parallel Preconditioned Krylov Methods, IEEE HPCC 2017.

[4] Gaston, Derek, et al. "MOOSE: A parallel computational framework for coupled systems of nonlinear equations.", 2018.

[5] Chauhan, Harvinder, et al. "Implementation of decision tree algorithm c C4. 5.", 2013.

U.S. DEPARTMENT OF ENERGY

CCF- 1219089, CNS-0821794 DE-AC02-06CH11357