Similarity Measurement for Proxy Application Fidelity

Si Chen si.chen2@emory.edu
Omar Aaziz oaaziz@sandia.gov
Jeanine Cook jeacook@sandia.gov
Avani Wildani avani@mathcs.emory.edu

EMORY COLLEGE OF ARTS AND SCIENCES
Sandia National Laboratories

Motivation and Goal

- Quantitatively understand the correspondence of proxy to parent applications with respect to their underlying node and memory hardware behavior.
- Ensure the system co-design and procurement that proxy application are high fidelity models of the target applications
- Find an **efficient** similarity measurement.

Methodology

1. Application representation
   - A vector with 700 dimensions (features), each of which is an average rank value for one hardware event.
2. Feature set
   - Overall (700 features)
   - Subset (subgroup features selected by experts):
     - Cache-related, e.g., longest_lat_cache.miss
     - Instruction mix
     - Pipeline
3. Distance measurement
   - Euclidean distance:
     - Cosine similarity

$$\cos\theta = \frac{\sum_{i=1}^{n} x_i y_i}{\|x\|\|y\|}$$

   - Statistic distance:
     - Kullback–Leibler divergence

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

     - Jensen-Shannon divergence

$$\mathrm{JSD}(P \parallel Q) = \frac{1}{2} D(P \parallel M) + \frac{1}{2} D(Q \parallel M)$$

     - Wasserstein distance

$$W_2(p,q) = \sqrt{\min_{P_{XY}} E_{P_{XY}}\left[\|x-y\|_2^2\right]} \text{ s.t. } P_X \sim p, P_Y \sim q$$

4. Cluster similar applications
   - Hierarchical Clustering
     - Get the similarity by tree branch height to rank applications.
5. Evaluation
   - Similarity ranking
   - Self similarity

ACKNOWLEDGEMENTS

Experiment

Applications

- ECP Applications in scientific domain of cosmology, seismic modeling, molecular dynamics, and thermal transport.
- 16 apps in proxy/parent pairs and 5 unpaired.

System platform

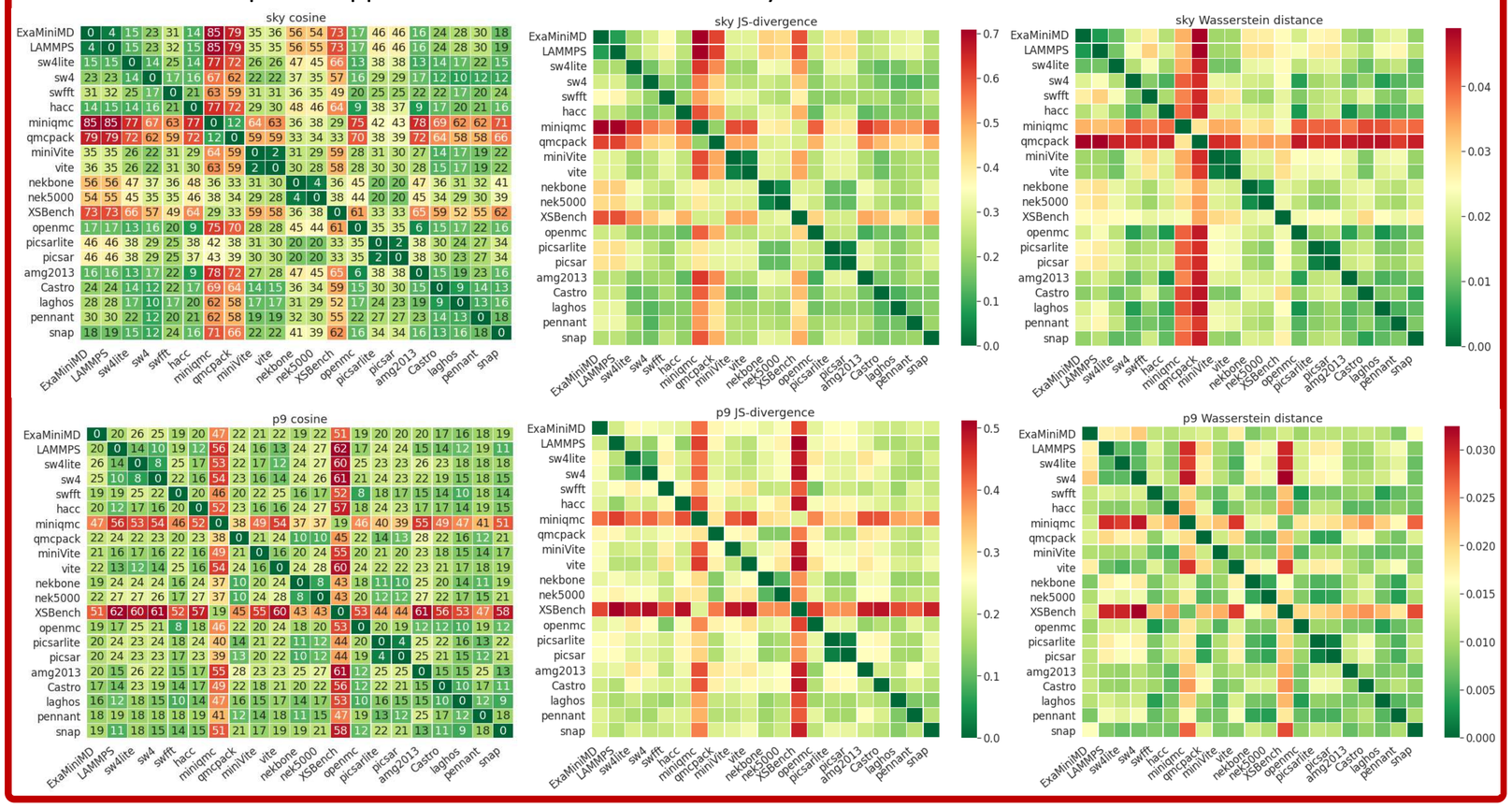
- **Intel Skylake** (sky)
- **IBM Power9** (p9)
- Configuration: MPI-only mode, 128 ranks, 1 rank/core, on four nodes.

Data collection

- Collection infrastructure: LDMS
- PAPI sampler collect 700 hardware events performance for each application.
- Standard input and parameters across applications.

Result

- Feature results of three distance methods look alike within each system!
- Two pairs (nek5000/nekbone and vite/miniVite) show high similar performance in cosine similarity, and relatively high similar performance in other two methods.
- The last five unpaired applications show relative similarity.



Conclusion

- Overall rankings show similar results for three distance methods.
- Cross-platform similarities show different results.
- Most proxies are good representations of their parents, only with few divergences.
- Cosine similarity is simple, quick and interpretable by geometric angle
- Since other distance methods show similar results, cosine similarity is sufficient for similarity measurement.

Future work

- Calculate self-similarity for time series data with sliding windows to further validate the similarity methods.
- Validate similarity through measures of latency or some other environmental / real metric.