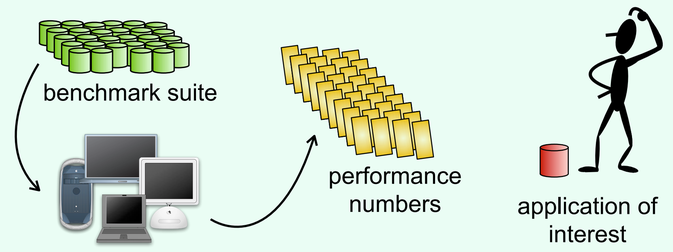
Analyzing Commercial Processor Performance Numbers for Predicting Performance of Applications of Interest

Kenneth Hoste*, Lieven Eeckhout* and Hendrik Blockeel°

* *Ghent University, Belgium* - ° *K.U. Leuven, Belgium*

Current Practice in Benchmarking



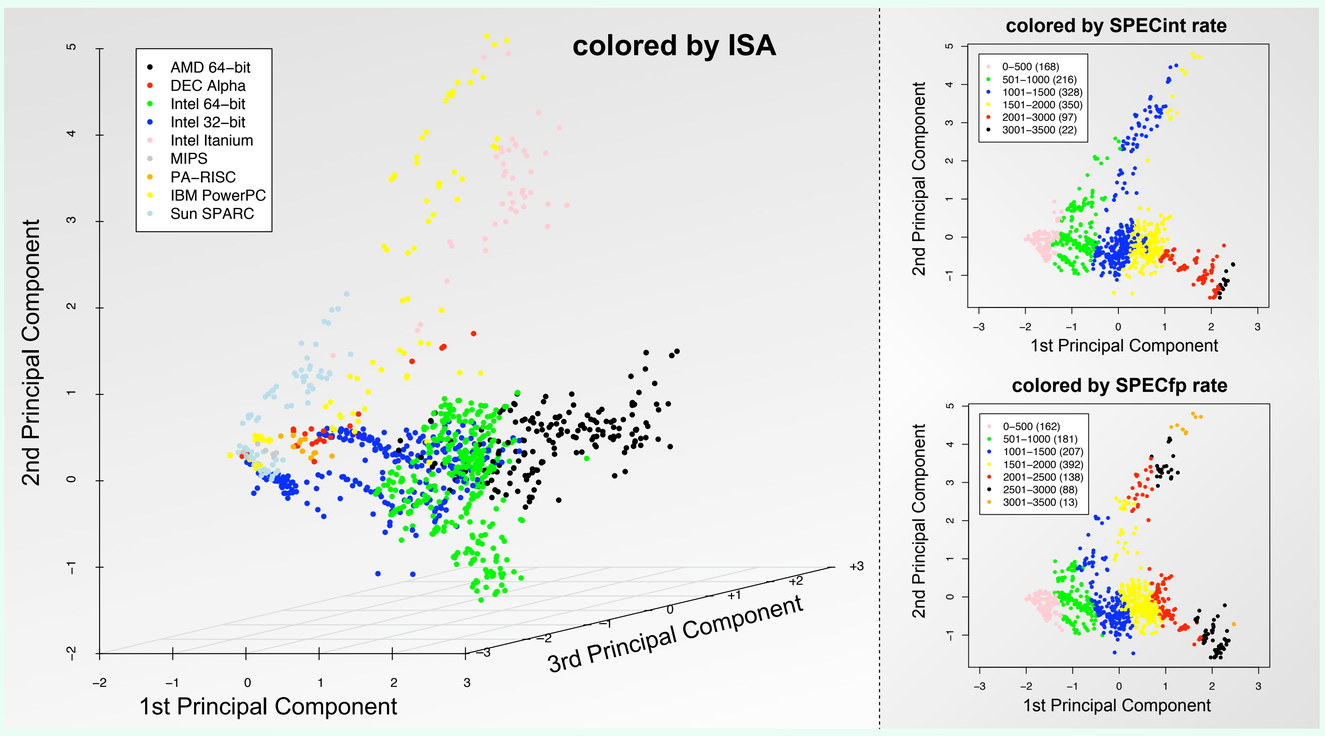
Benchmarking systems is commonly done to *assess system performance*, for example using the well-known *SPEC CPU* benchmarks (`www.spec.org`). Performance numbers for each of the benchmarks are reported by various organizations, including system and processor manufacturers, both for research and marketing purposes. *Two problems* come forward when trying to interpret these performance numbers:

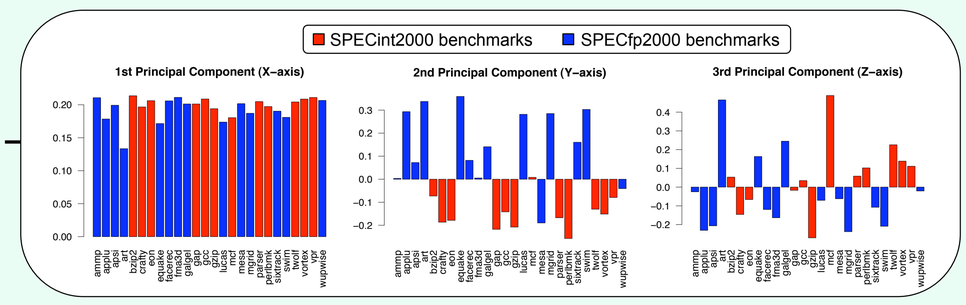
- Because of the *huge amount of data* (speedup numbers for 26 benchmarks for 1000+ machines for SPEC CPU2000), getting *valuable insights* is hard.
- It is unclear *which benchmarks* are *representative* for a given *application of interest*. Using *average performance numbers* across a set of benchmarks (for example SPEC rates) to determine which system is best suited for a given task can be *misleading*, and may lead to suboptimal purchase decisions.

This work presents a *solution for both problems*; we *analyze published performance numbers of commercial machines* to obtain valuable insights and propose a methodology to *estimate performance* for a given application of interest.

Analyzing Commercial Processor Performance Numbers

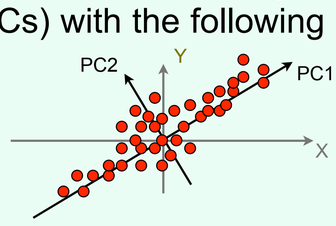
Performance numbers are projected into a three-dimensional space using principal components analysis (PCA). Each point represents one of the *1181 machines*, on which the *SPEC CPU2000 benchmark suite* was run (data obtained from `www.spec.org`). Choosing colors according to processor characteristics yields valuable insights.





Principal Components Analysis (PCA)

Principal components analysis is a *statistical technique* which computes new dimensions (PCs) with the following properties:

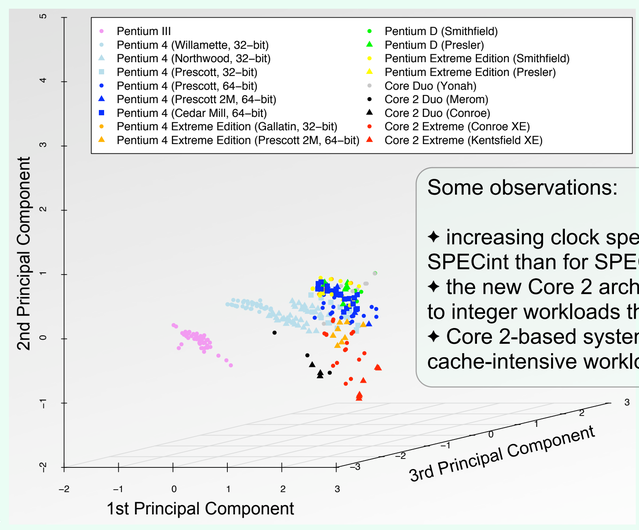


$Var[PC_1] \geq Var[PC_2] \geq ... Var[PC_n]$

$Cov[PC_i,PC_j] = 0 \ (i \neq j)$

Each principal component (PC) is a *weighted linear sum* of input dimensions. Retaining a limited number of PCs allows *reduce the dimensionality of data* while controlling the amount of information lost.
We retain 3 PCs, which explain *92.5% of total variance*.

Case study: Intel desktop processors



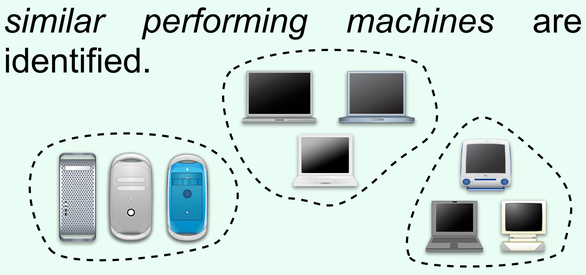
Some observations:

- ✦ increasing clock speed is more beneficial for SPECint than for SPECfp
- ✦ the new Core 2 architecture is more targeted to integer workloads than previous systems
- ✦ Core 2-based systems are well suited for cache-intensive workloads

Performance Estimation for Applications of Interest

Cluster analysis

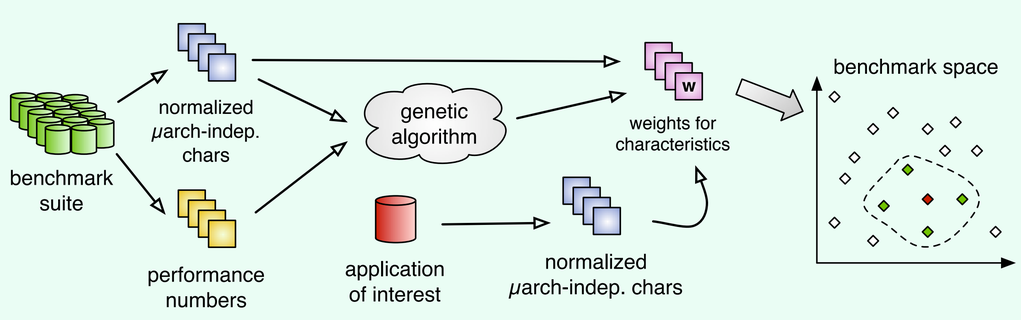
Using cluster analysis (k-means) in the three-dimensional space described above, various *groups of similar performing machines* are identified.



For each cluster of similar machines, a *performance model* is constructed. Simply grouping machines by ISA or average performance is insufficient, because of the large *variety among machine models* (for example, Intel Pentium 4 and Intel Xeon).

Performance model

For each cluster of machines, a performance model is constructed using microarchitecture-independent program characteristics and a genetic algorithm. For a given application of interest, a *performance estimation* is made by making a weighted harmonic average of the performance numbers of *similar benchmarks*.



Experimental results

For each of the SPEC CPU2000 benchmarks, performance numbers (speedups compared to a reference machine) were estimated for *617 machines* divided in *26 clusters* (by means of cross-validation). This results in an *estimated machine ranking*. We evaluate the *performance deficiency* of the estimated best and top 3 machines for each benchmark.