Research Report: Performance and Power prediction model

Ting Li (SN:2017124076)

May 26, 2019

1 Regression Model

In 2006, Benjamin C. Lee and David M. Brooks claimed that regression model can as an efficient approach for accurately predicting performance and power for various applications executing on any microprocessor configuration in a large microarchitectural design space. This method was used by later scholars and considered to be effective.

In Benjamin's article, he summarize a statistically rigorous approach for deriving regression models that includes (1) variable clustering, (2) association testing, (3) assessing strength of response-predictor relationships, and (4) significance testing with F-tests.

And he also give a baseline model that accounts for predictor interaction and non-linearity, presents model optimizations to improve prediction accuracy by (1) stabilizing residual variance, (2) deriving application-specific models, and (3) deriving regional models with samples most similar in architectural configuration to the predictive query.

The baseline model given:

Once β is determined, evaluating Equation (6) for a given x_i will give the expectation of y_i and, equivalently, an estimate $\hat{y}_i$ for y_i. This result follows from observing the additive property of expectations, the expectation of a constant is the constant, and the random errors have mean zero.

$$\begin{aligned} \hat{y}_i &= E[y_i] \\ &= E\Big[\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\Big] + E[e_i] \\ &= \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} \qquad (6) \end{aligned}$$

Figure 1: result

And finally, they give several experimental results from four different performance and power regression models: (Record detailed descriptions(can skip))

1. Baseline(B): (Shown above) Without variance stabilization and formulated with a naive subset of $n_B < n$ observations.

2. Variance Stabilized(S): Model specified with a square-root transformation on the response and formulated with a naive subset of $n_S < n$ observations.

3. Regional(S+R): Model is reformulated for each query by specifying a naive subset of $n_{S+R} < n$ observations. These observations are further reduced to include the $r_{S+R} < n_{S+R}$ designs with microarchitectural configurations most similar to the predictive query. We refer to r_{S+R} as the region size. Similarity is quantified by the normalized euclidean distance bqetween two vectors of architectural parameter values, $d = \sqrt{\sum_{i=1}^{p} |1 - b_i/a_i|^2}$.

4. Application-Specific(S+A): use a new set of $n_A = 4,000$ of observations for varying microarchitectural configurations, but a fixed benchmark. An application-specific model is obtained by eliminating application-specific predictors from the general model and reformulating the model with a naive subset of $n_{S+A} < nA$. We consider S+A models for six benchmarks: ammp, applu, equake, gcc, gzip, and mesa.

(Detailed result figures or tables are omitted here, and only conclusion shown.)

In conclusion, Application-Specific models are most accurate for performance prediction, Regional models are most accurate for power prediction.

2 Learned and Thought

Researchers pointed out the regression model can be used in predicting microarchitectural performance and power. I think it could be used in both CPU and GPU's performance and power consumption prediction. Benjamin also shown detailed derivation process of models in his article.

However I feel it is not easy to totally understand it. Maybe more time is needed in deriving prediction models.

References

[1] Benjamin C Lee and David M Brooks. Accurate and efficient regression modeling for microarchitectural performance and power prediction. In *ACM SIGOPS Operating Systems Review*, volume 40, pages 185–194. ACM, 2006.

[2] WANG Gui-bin. A gpu power prediction model based on hardware performance counter. In *COMPUTER ENGINEERING SCIENCE*, volume 34, pages 50–54, 2012.