

1 Summary

Dynamic Branch Prediction with Perceptrons (2001): The authors propose as a way to improve the accuracy of branch prediction to replace traditional pattern history table (PHT) with a neural network, perceptrons, as the branch prediction mechanism. Perceptrons are chosen because of their simplicity and lower cost compared to most other complex neural networks. They compare their method to dynamic global predictors, gshare and bi-mode, which are generally limited in the lengths of the history register, as well as a hybrid gshare/perceptron predictor. The perceptron predictor can accommodate longer histories. The authors find that their perceptron predictor performed particularly well for linearly separable branches (which the authors classify as a pattern where inputs map to either true or false outputs) and branches requiring longer history lengths, and has more accurate results than other existing branching global branch predictors for the SPEC 2000 integer benchmarks (misprediction improved by 10.1%).

2 Strengths

- Section 3 is very thorough in explaining the new concepts and mechanisms the authors introduce.
- Explanation of methodology is detailed in how they tune the simulation, which leads to an interesting conclusion on the linear relationship between history length and threshold.

3 Weaknesses

- The authors only benchmarked using the SPEC 2000 integer benchmarks. Their breadth of experiments does not nearly encompass the types or variety of workloads seen in real application.
- Some of the justifications for using perceptrons is that they are easier to implement and less expensive compared to other neural networks, and use cost less in terms of computation hardware compared to the table sizes with the other branch prediction mechanisms, gshare and bi-mode; however, the authors did not discuss if perceptrons are overall more cost-effective compared to gshare and bi-mode, and whether or not if the improved misprediction was worth and that cost differences.

4 Rating: 4

5 Comments

The authors identified an unexploited area in the dynamic branch predictors of the time of the paper and found an approach to extract more performance gains where those mechanisms fell short. Techniques with complementary strengths outperforming older dynamic branch predictions and the perceptrons alone also supported their hypothesis that prediction mechanisms could be more accurate if they also accommodated longer history lengths. Their implementation also showed it was feasible to incorporate a machine learning based techniques in hardware and was computationally feasible despite how prohibitively complex neural networks were known to be in that area. Their work was not only successful in inspiring future research in the area of perceptron prediction, but also actually implemented in a few commercial architectures, like [piledriver](#), [ryzen](#). It is likely that in the future when hardware and training costs are less expensive, hybrid prediction mechanisms such as the one the paper implements are going to be more ubiquitous given the better performance.