

グローバル履歴の時系列を考慮したパイプライン処理によるパーセプトロン分岐予測器の高速化*

小西貴大†

大阪工業大学 情報科学研究科‡

1 はじめに

本研究ではプロセッサにおける分岐予測手法の一つであるパーセプトロン分岐予測器 [1] に着目する。

パーセプトロン分岐予測器では内積の加算処理に時間がかかるという問題がある。この問題に対して精度の劣化と引き換えに加算部分をパイプライン化することで予測速度を上げる研究がある [2]。パイプライン化によって通常の内積計算より過去の分岐結果を使うことになり、予測に用いられる数値の信頼性が低くなる。もう一つの問題点として線形分離不可能な予測パターンは、予測できない問題がある。この問題に対してパーセプトロン分岐予測器の精度を上げる研究として冗長入力を付加する技法がある [3][4]。

本研究では冗長入力を付加する手法とパイプライン化を組み合わせる。提案手法ではグローバル履歴 (GHR) の時間要素に着目して重要性を考慮した分岐予測器を設計した。GHR には直近で実行された分岐命令の結果が時系列順で保持されている。GHR 中の古い方の履歴は予測精度への影響が小さいと考えられ、これに関わる積和計算部分はパイプライン化により高速化する。一方で、新しい方の履歴は影響が大きいと考えられこれに関わる積和計算部分はパイプライン化せずに計算を行う。これに加えてパーセプトロンの重み更新も時系列要素を考慮した更新方法にすることで古い履歴の影響をより低くし、パイプライン化で信頼性が低くなることを防ぐ。

2 提案手法

図 1 は提案手法の分岐予測器全体図である。

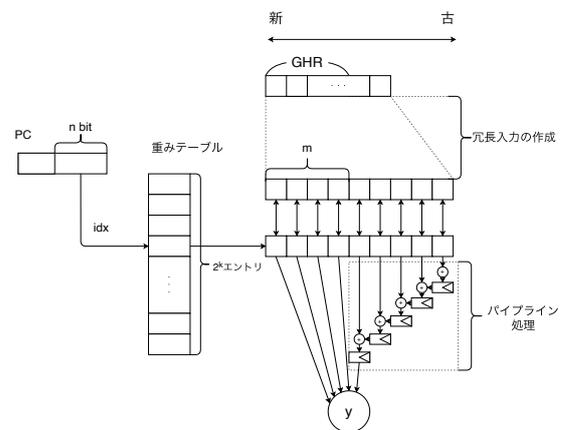


図 1: GHR の時系列を考慮した分岐予測器

この分岐予測器では GHR から冗長入力を作成し [4] 使用する。作成した冗長入力の時系列が新しい部分はパイプライン化せず内積の加算計算を行う。一方で、残りの古い部分の内積の計算をパイプライン処理することで加算一回分の時間で処理する。例として図 1 の GHR を 16bit とした時、生成される冗長入力は 48bit になる。パイプライン化せずに計算する範囲は GHR の新しい方から mbit とする。残りの 48-mbit はパイプライン処理をすることで加算一回分の処理時間にする。古い履歴が予測精度に良い影響を与えることが少ないと考えられ、この手法で精度の劣化を抑えながら計算処理を高速化する事が期待できる。

さらに重み更新にも時系列を考慮した。式 (1) は提案手法の重み更新式である。古い履歴に対応する重みの更新割合を減らす事で予測結果に対する影響を少なくする。

$$W[idx][i] = W[idx][i]/2 + \alpha * t * GHR[i] \quad (1)$$

α は学習係数, t は真の分岐結果である。既存のパー

*Acceleration of Perceptron Branch Predictor by Pipeline Processing Considering Time Series of Global History

†Konishi Takahiro

‡Osaka Institute of Technology

セプトロン分岐予測器の更新方法より半分の数値を更新する。更新の値を半分にする事によってパイプライン化した時に精度の劣化の要因になる古い履歴部分の影響を少なくし、新しい履歴部分の影響を大きくする事ができる。

3 検証結果

既存の分岐予測器と提案手法で予測精度の評価、比較を行う。

本研究ではベンチマークプログラムとして SPEC-Cint2000 内にあるプログラムを使用する。比較検証した分岐予測器は ① パーセプトロン分岐予測器, ② 単純なパイプライン化, ③ 冗長入力の付加, ④ 冗長入力の付加+単純なパイプライン化, ⑤ 提案手法である。GHR の長さを 16bit, 冗長入力を付加した入力の長さは 48bit とする。提案手法におけるパイプライン化しない m に当たる部分は 16bit に設定する。重みテーブルは 512 エントリとする。図 2, 図 3 が提案手法と既存の分岐予測器の正答率を示したものである。

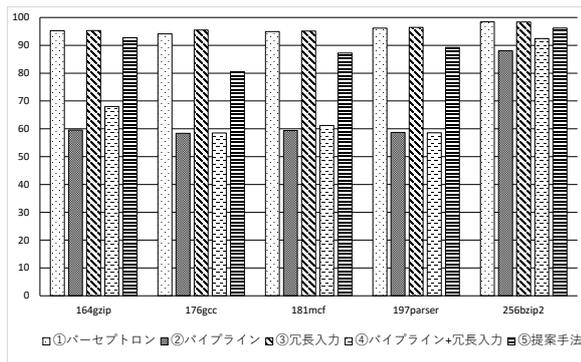


図 2: 提案手法の比較

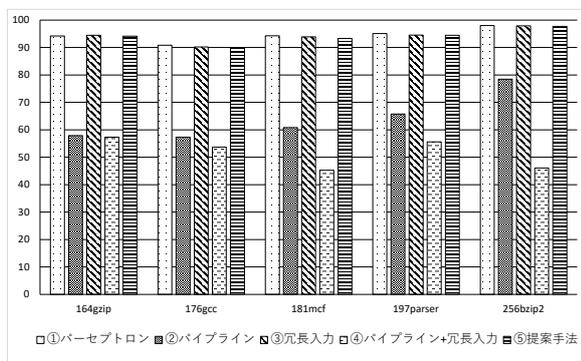


図 3: 提案手法を式 (1) の更新方法にした場合の比較

図 2 が通常の重み更新した場合の予測正答率, 図 3 は式 (1) を適応した重み更新した場合の予測正答率である。縦軸が正答率で横軸がベンチマークの種類とする。

提案手法では図 2, 図 3 で②と④の分岐予測器より良い精度を示す事がわかった。図 2 の通常の重み更新をした場合を見ると, 164gzip, 256bzip2 の 2 つのベンチマークではパイプライン化していない①パーセプトロン分岐予測器と③冗長入力と同等の精度を示す。しかし, 176gcc, 181mcf, 197parser の 3 つのベンチマークでは①, ③と比較して精度が悪い。

図 3 の式 (1) の重み更新をした場合では, ②単純なパイプライン化したものと④冗長入力を付加と単純なパイプライン化した分岐予測器よりも 30%以上の精度向上が確認できた。①パーセプトロン分岐予測器と③冗長入力と比較しても全てのベンチマークで同等の精度を示す。時系列を考慮したパイプライン化と重み更新方法を適応することによってパイプライン化による精度の劣化を抑える事ができる。

4 まとめ

GHR の時系列を考慮してパイプライン化と重み更新を実装することでパイプライン化による予測精度の劣化を抑える事ができた。今後の展望は提案手法の重み更新の条件を考え直す必要がある。今回は無条件に重みを半分にする事で影響を少なくしたが重みの更新する大きさや条件を加えることでより劣化を抑えるか検証したい。さらに, 単位命令あたりの消費電力やハードウェア資源の検証を行う必要がある。

参考文献

- [1] D. A. Jimenez and Calvin Lin. : Dynamic Branch Prediction with Perceptrons, *Proc. the Seventh International Symposium on High Performance Computer Architecture (HPCA'01)*, pp.197-206(2001)
- [2] D. A. Jimenez. : Fast path-based neural branch prediction, *Proc. 36th Annual IEEE/ACM International Symposium on Microarchitecture*, pp.243-252(2003)
- [3] A. Sez nec: Revisiting the perceptron predictor, *Technical Report 1620*, IRISA, (2004)
- [4] 澁川誠, 二ノ宮康之, 阿部公輝, 小林聡 : パーセプトロン分岐予測における冗長入力付加の効果, *IPSJ SIG Technical Report*, (2005)