

## 分岐予測器における予測性能とエントロピー

横 田 隆 史<sup>†</sup> 大 津 金 光<sup>†</sup>  
古 川 文 人<sup>††</sup> 馬 場 敬 信<sup>†</sup>

予測器は一般に対象の過去の挙動をもとにして、可能性のある選択肢のなかから最尤のものを次の状態として「予測」する任を担う。予測器があらかじめ与えられる情報なしに働くならば、予測対象とする系の性質・挙動が予測器の性能となって現れてくるはずである。我々は、系の挙動に見られる偏りやランダムさに着目し、エントロピーを測度として定量化することで、予測器の性能を表現することを試みた。系を2値のマルコフ情報源と考え、その情報源エントロピーを、予測に関して系が持つ情報量と考えた。また、予測器が内部にテーブル構造を持つとき、テーブルエントリごとの使用の多寡からもエントロピーを定義し検討対象とした。予測器の方式によってエントロピー値と予測成功率の間に明らかな相関性が認められる。

### An Entropy Measure for Branch Predictor Performance

TAKASHI YOKOTA,<sup>†</sup> KANEMITSU OOTSU,<sup>†</sup> FUMIHITO FURUKAWA<sup>††</sup>  
and TAKANOBU BABA<sup>†</sup>

Predictors are responsible for predicting the next state by means of past events. When a predictor works without any knowledge on the system, its prediction performance should be influenced by statistical characteristics of the system at some level. We focus our viewpoint on randomness in the system behavior and introduce an entropy concept so as to quantitatively measure the system behavior, i.e., randomness. We define information source entropy and reference entropy. These entropy values reveal prediction performance in some prediction methods.

#### 1. はじめに

現在のマイクロプロセッサ・アーキテクチャの中で、予測器、特に分岐予測器は、非常に重要な役割を担っており、その予測性能がシステムの処理性能を決める鍵となっている。このため、これまで様々な角度から検討が加えられ、様々な方式の(分岐)予測器が提案されてきた。

予測器は、対象とする系の履歴をもとにして、次の(近未来の)状態として起こり得るものの中から最尤であるものを「予測」する。予測器がプログラム作成時やコンパイル時に与えられる静的な情報を用いず、履歴のみをもとに働く場合は、系の挙動の性質が直接的に予測性能となって現れると期待される。たとえば、系が予測性の低いランダムな挙動をしている場合は、予測器は十分な性能を達成できないことが見込まれるが、逆に、規則的な挙動をする系では予測は容易で

ある。

このように予測器の性能は系の挙動のランダムさにある程度依存するものと推察される。そうであれば、系の挙動のランダムさを測る指標を導入することによって、予測器の性能を表現できるのではないか。本稿は、こうした観点から、ランダムさの測度として情報エントロピーの考え方を導入し、系のランダムさと(分岐)予測器の性能について論じる。

以下、2節で本稿の研究の契機となった2パス限定投機手法におけるパス予測器の挙動に関して説明する。3節では、系のランダムさを定量的に表現するためにエントロピーを導入し、パス予測器に適用する。4節ではさらに各種の分岐予測器に拡大し、評価結果を示す。

#### 2. パス予測器の振舞い

##### 2.1 2パス限定投機手法

高頻度・集中的に実行されるホットループを対象として、ループイテレーションを単位にしてマルチスレッド実行を行う場合を考える。たとえば、ループ内で実行可能性のある3本の実行経路(パス)の中から次に

<sup>†</sup> 宇都宮大学  
Utsunomiya University  
<sup>††</sup> 帝京大学  
Teikyo University

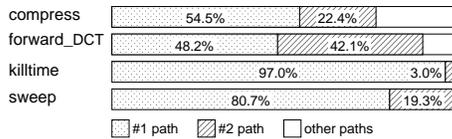


図 1 上位 2 つのパスの実行頻度

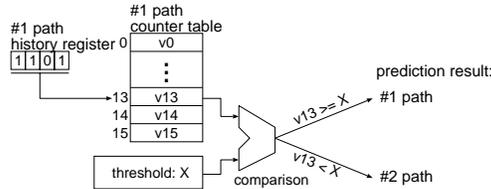


図 2 2 レベルパス予測器

実行されるパスを予測し、投機的に実行する。あらかじめ個々のパスに特化して最適化した実行コードを用意しておき、予測結果に従って投機実行すれば高効率な処理が期待できる。

我々は、ベンチマークプログラムでのパスの出現頻度を調べ、多くの場合、出現頻度の高い 2 つのパスが支配的であることを明らかにした<sup>1)~3)</sup>。SPECint95 のいくつかのプログラムでの出現頻度上位 2 つのパスの頻度を図 1 に示す。以降、頻出パスを頻度順に #1 パス、#2 パスと呼ぶ。

このように、プログラム中に存在するパスの数に比べて実際に実行されるものは僅かであり、実質上、2 つに限定しても大きな損失はない。逆に、投機スレッドコードの準備を省ける、スレッド間での依存の解決が容易になる等のメリットがある。このため頻出パス 2 つに限定して投機を行う 2 パス限定投機方式を提案している<sup>4)</sup>。

この 2 パス限定投機方式に用いる予測器として、2 レベル分岐予測器をベースにした 2 レベルパス予測器を提案している (図 2)。これは、イテレーションごとの #1 パスの実行有無を履歴レジスタに保存し、この履歴レジスタにより飽和カウンタによるカウンタテーブルをアクセスし、#1 の選択の可否を決めるものである。予測の成否により当該カウンタを更新する。

この予測器を用いたときの、履歴のビット数と予測成功率との関係を図 3 に示す。p1only は #1 の出現確率であり、常に #1 パスを予測する予測器の性能に相当する。この図から、複数のアプリケーションにおいて、概ね 10 ビット以上の履歴長で成功率が漸増していることがわかる。

## 2.2 パス予測器の挙動

前節での評価において 2 レベルパス予測器の履歴レジスタの値が変化する様子記録し挙動を解析した結果、(1) 実際に出現する履歴レジスタ値の値が限られていること、(2) 履歴レジスタ値の遷移のしかたに強い法則性が観測されること、が明らかになった<sup>1),2)</sup>。

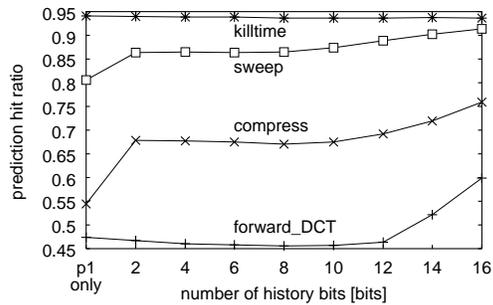


図 3 履歴長と予測成功率の関係

その具体的な例として、表 1 に、compress で観測された履歴レジスタ値を出現頻度順に示す。履歴レジスタ長は 8 ビット、カウンタテーブルは 2 ビット、データセットは test である。出現 50 回未満の履歴レジスタ値は省略しており、出現した履歴レジスタ値は合計 90 通りである。表 1 に示されているように強い法則性が見られる。すなわち、'0' が非常に長く連続すること、'1' は '0' ほど長く連続せず途中にわずかな '0' を含むことが多い、'0' と '1' が入り交じった状態になることは多くない、などの規則性が推測される。また、全  $2^8 = 256$  個の可能性のうち、実際に出現したのは 90 であり、さらに、出現頻度の高いもの (50 回以上) は 31 個となっている。パスの出現のしかたがランダムであれば、個々の履歴レジスタ値の出現頻度に多寡が生じることがあっても、一度も出現しない値はない。

こうした法則性は、履歴レジスタ値の遷移の偏りとしても観測される。履歴レジスタの値は実際に出現したパスによって (1 ビットのシフト後) 最下位ビットに 0 または 1 が与えられる。このため履歴レジスタ値の遷移先は常に 2 つ (以下) に限られるが、実際の遷移は一方に著しく偏る場合と、ほぼ均等になる場合が多く観測されている。偏り方が著しいほど予測成功率が向上するものと考えられる。出現した全履歴レジスタ値に対して遷移確率を求め、90%以上の割合で偏った遷移をするものの比率を求めた。履歴長 4, 8, 12, 16 に対して求めた結果を表 2 に示す。偏りの割合が履歴長 10 以上で著しく増す傾向が分かる。これは図 3 で履歴長 10 以上で予測成功率が上昇する傾向と符合する。

## 3. 情報エントロピーの導入

### 3.1 参照エントロピー

上記評価においてパス予測器内で生じているランダムさを、情報エントロピーを用いることで表現する。エントロピーが大きいとき系はランダムであり、小さければランダムでなくすなわち偏りが大きいことになる。

2 レベルパス予測器において、履歴のビット数を  $n$ 、カウンタテーブルの各エントリ  $i$  の参照回数を  $r_i$  と

表 1 compress での頻出履歴レジスタ値

出現頻度	履歴レジスタ値
975	00000000
300 ~	11101111 11110111
200 ~	11011111 11111011 11111111
100 ~	10111111 11111101 00100000 01111111
	11111110 00000100 00001000 00010000
50 ~	01000000 00000001 01000001 10000010
	00000010 10000000 10111011 11011101
	11011110 01111011 10111101 10111110
	01111101 01110111 01111110 01001000
	10010000

表 2 著しく偏った履歴レジスタ値遷移をする割合

関数	履歴長			
	4	8	12	16
compress	0.0	4.1	22.5	48.3
forward_DCT	0.0	0.0	1.3	10.3
killtime	95.9	95.9	95.9	95.9
sweep	64.4	60.0	65.2	76.0

(偏り 90%以上) 単位: %

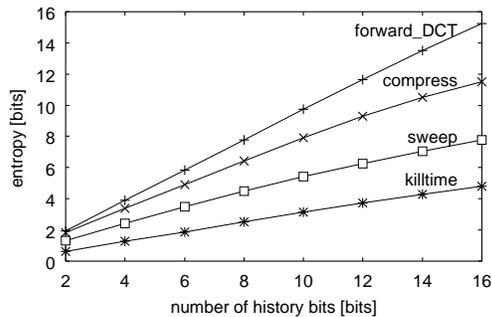


図 4 履歴長と参照エントロピーの関係

する．参照回数 $r_i$ の総和を  $N$  としたとき、 $N = \sum_i r_i$  ( $0 \leq i < 2^n$ ) である．ここで  $q_i = r_i/N$  として、 $q_i \neq 0$  なる  $q_i$  について情報エントロピーを

$$H(\bar{R}^n) = \sum_i -q_i \log_2(q_i) \quad (1)$$

のように定義する．カウンタテーブルの参照状態をもとに算出したものであるから、以降、参照エントロピーと呼ぶことにする．

2 節で行った評価において、履歴レジスタのビット長に対する参照エントロピーを求めた結果を図 4 に示す．図からわかるように、参照エントロピーは履歴レジスタ長の増加とともに単調増加するが、アプリケーションによって傾きが異なっている．パスの生成挙動がランダムであれば、エントロピーは履歴レジスタのビット長を示す．逆に、履歴レジスタ長を増加してもエントロピーが大きく変わらない場合は、系の偏りが大きいことを示している．図 4 のグラフの傾きの大小と、図 3 に示された予測成功率を見比べれば、エントロピー値が小さいほど成功率が高いことがわかる．

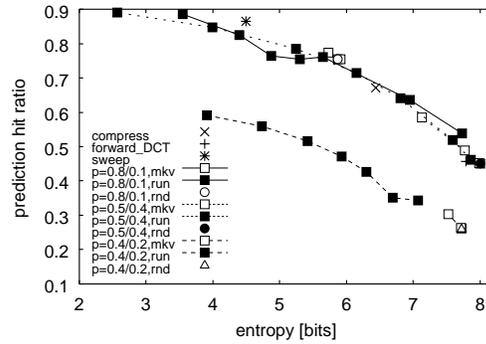


図 5 参照エントロピーと予測成功率の関係

### 3.2 参照エントロピーを用いた評価指標

さらに、パスの出現確率が同じであっても、出現のしかたに偏りがあることでパス予測器の予測精度が変動することを確認する．ここで、人為的に偏りのあるパスの出現パターンを作成し、その出現系列を 2 レベルパス予測に適用して、前述の方法で求めた参照エントロピーと予測成功率との関係を求めた．履歴レジスタのビット数は 8 である．

結果を図 5 に示す．図中  $p=P1/P2$  は、#1、#2 パスの出現確率  $p_1, p_2$  を、それぞれ P1、P2 に設定していることを表している．また、図中の各グラフに付された mkv, run, rnd はパスの生成方法の違いを表す．さらに比較のため、2 節の評価で求められた compress, forward\_DCT, sweep の予測成功率とエントロピー値をプロットしている．人為的にパスを生成する mkv, run, rnd では、1 回の試行ごとに 1,000,000 個のパスを出現させている．

mkv は、出現パス間の遷移確率を表した遷移確率行列を求め、それによってランダムにパスの出現を遷移させるやり方である．エントロピーを変えるために、アプリケーション固有の遷移確率行列を、各パスの出現頻度を変えないように変更している．パス  $i$  からパス  $j$  への遷移確率  $a_{ij}$  は、#1、#2 パスの出現確率  $p_1, p_2$  と揺動パラメータ  $\delta$  から

$$A = \begin{pmatrix} p_1^2 + \delta & p_1 p_2 + \delta & p_1 p_3 - 2\delta \\ p_1 p_2 + \delta & p_2^2 + \delta & p_2 p_3 - 2\delta \\ p_1 p_3 - 2\delta & p_2 p_3 - 2\delta & p_3^2 + 4\delta \end{pmatrix} \quad (2)$$

としている．ただし、 $p_3 = 1 - p_1 - p_2$  であり、 $\delta$  は  $A$  の各要素  $a_{ij}$  について、 $0 \leq a_{ij} < 1$  が成立する範囲で変化するものとする．

run では各パスをできるだけ連続して出現させるが、その時々パスの出現頻度を計算し、その結果が各々の設定値（#1 パスであれば  $p_1$ ）を越えないことを条件とする．設定値を越えた場合、パラメータで決められた割合で他のパスに遷移する．これにより、#1、#2 パスが連続して出現しやすい、エントロピーの低い状態を作ることができる．たとえば、このパラメータを 0.01 (1%) とすると、#1、#2 パスがほぼ (99%) 一定

の回数で繰り返し出現する状況が得られる。

rnd は、上記の 2 方法とは異なり、パスの出現頻度に従って単純にランダムに出現パスを選択する方法である。[0.0:1.0] の範囲の実数値を返す乱数生成関数 rand() を用い、rand() < (p1/(p1+p2)) のとき #1 を生成し、そうでないとき #2 を生成する。これにより、パスの出現確率が与えられた条件下でもっともランダムな状態 (エントロピー最大の状態) を得ることができ。

図 5 より、まず、2 レベルパス予測器の予測成功率は情報エントロピーの値により大きく変動することがわかる。情報エントロピー値が小さいほど成功率が高く、情報エントロピーが十分に小さいとき、予測成功率は #1, #2 パス両者の出現確率の和に近くなる。

また、p=0.8/0.1 と p=0.5/0.4 の場合でほぼ同じ軌跡となること、これらの軌跡は p=0.4/0.2 のそれとは明らかに異なること注目したい。2 レベルパス予測器では、#1, #2 パス両者の出現確率の和によって予測性能を表せることがわかる。

またさらに、実際のプログラムを実行した場合のエントロピーおよび予測成功率のプロットが、グラフの曲線上に乗っていることから、上で求めた情報エントロピーが予測器の性能指標として有効であることが示される。

#### 4. 分岐予測器のエントロピー

前節で、参照エントロピーが 2 レベルパス予測器の性能の指標になることを示した。この結果がよく似た挙動を示すことが期待される分岐予測器に適用できるのかを本節で検討したい。

##### 4.1 エントロピーに関する議論

前節で扱った 2 レベルパス予測器の参照エントロピーは、以下のように解釈することができる。

パスを通った/通らなかったを、1 ビットの情報源として考える。これにより、0 または 1 の値をとる 1 ビットの情報が、時系列で生成される 2 値情報源をなす。2 レベルパス予測器の履歴レジスタは、この 1 ビット情報の最近傍  $n$  ビットを時系列に保持したものである。また、カウンタテーブルの各エントリの参照数は、履歴レジスタの各値が現れた回数に等しい。

ここで、上記の 2 値情報源をマルコフ情報源として考えてみる。 $n$  ビット履歴レジスタの出現回数をもとに式 (1) により定義した参照エントロピーは、 $n$  次の拡大随伴情報源 ( $\bar{S}^n$ ) のエントロピー  $H(\bar{S}^n)$  で近似できる。つまり、参照エントロピーは、系がマルコフ情報源として持つ情報量 (エントロピー) を近似的に表現したものであり、前節で得られた結果 (図 5 等) が妥当であることがわかる。

$n+1$  次の拡大随伴情報源エントロピー  $H(\bar{S}^{n+1})$  が求めれば、もとの情報源  $S$  のエントロピーの  $n$  次

近似  $H^n(S)$  は、

$$H^n(S) = H(\bar{S}^{n+1}) - H(\bar{S}^n) \quad (3)$$

で求められる。 $n \rightarrow \infty$  で、もとの情報源の真のエントロピー  $H(S)$  となる。

2 レベルパス予測器に関する上述の議論で、参照エントロピーが拡大随伴情報源エントロピーを近似しているのであれば  $H(\bar{S}^{n+1}) - H(\bar{S}^n)$  により、もとの情報源のエントロピー (の近似値) を求めることができる。この値は、図 4 の傾きであり、このグラフの傾きが小さいプログラムほど情報源のエントロピーが小さい、すなわち予測性が高いことが確認できる。

##### 4.2 分岐予測器への適用

2 レベル分岐予測器では、 $n$  ビットの履歴レジスタを  $n$  次拡大随伴情報源のシンボルとして (近似的に) 扱うことができたが、一般の分岐予測器では、方式の差異からこの前提をそのまま適用することはできない。このため、参照エントロピーはカウンタテーブルを対象に式 (1) により定義されるものとし、さらに、ここで、条件分岐の実行の結果得られる分岐の有無の情報を 2 値マルコフ情報源として考え、 $n$  次拡大随伴情報源エントロピーを系の性質を表す指標として扱う。なお、以下では ( $n$  次) 情報源エントロピーと称することにする。

すなわち、プログラムの実行中の条件分岐の有無から情報源エントロピーが、予測器のカウンタテーブルの参照状況から参照エントロピーが、それぞれ求められる。

##### 4.3 SimpleScalar による評価

プログラム実行中の分岐予測器の挙動を調べるため、SimpleScalar toolset (ver.3.0d)<sup>5)</sup> の sim-bpred を一部改変し、参照エントロピー、情報源エントロピーとヒット率を求められるようにした。これらは、プログラムの実行全体を通しての値ではなく、指定されたウィンドウ時間内の値を求め、時間経過とともに変化する様子を観測できるようにした。本稿での評価に用いたウィンドウ時間は、シミュレーション実行中に検出した分岐命令の数をもとにしており、本評価でのウィンドウサイズは 1,000 万分岐命令である。予測器には 2 level, gshare, bimod を用いた。

ベンチマークプログラムは SPEC CPU2000 の各アプリケーションを用いた。データサイズは train で、シミュレーションの最初から最後まですべての実行を対象にしている。参照エントロピーは履歴長  $n$  に対し  $n = 4, 6, 8, 12, 14, 16$  の 6 通りを求め、 $n$  次情報源エントロピーは  $1 \leq n \leq 20$  の各場合について求めた。

前後の時間を除外したり、SimPoint などによる抽出などは一切行っていない。

4.4 参照エントロピーと予測成功率の時系列変化  
参照エントロピーと予測成功率の時間経過の様子を  
図 6 に示す。同図は、履歴長  $n = 16$  の gshare 予測  
器の場合である。参照エントロピー ( $H(\bar{R}^{16})$ ) と予測  
成功率とを同一グラフ上にプロットしている。横軸が  
時間経過を表し、縦軸左座標が予測成功率、右座標が  
エントロピーである。比較を容易にするため、いずれ  
の図も同一のスケールで表示している。いずれの  
グラフも、エントロピーが小さいときに高い予測成功  
率となっており、両者が相補的な関係にあることがわ  
かる。

4.5 情報源エントロピーと予測成功率との関係  
 $n$  次および  $n - 1$  次の情報源エントロピーから、  
式 (3) により、分岐命令による 2 値情報源のエント  
ロピーを求める。本稿では、 $H(\bar{S}^{20}) - H(\bar{S}^{19})$  と  
 $(H(\bar{S}^{20}) - H(\bar{S}^{18}))/2$  を求め、小さい方を近似的な  
情報源エントロピーとした。こうして求めた情報源  
エントロピー  $H(S)$  と、予測成功率 ( $n = 16$  の場合)  
との関係を表すグラフを図 7 に示す。

このエントロピー ( $H(S)$  が『次に来る分岐がとる  
値 (分岐する: 1, 分岐しない: 0)』の情報量を表して  
いる。この情報量の値から、確率を逆算すれば、予測  
成功率の期待値を求めることもできる。

図 6 で観測されたように、参照エントロピーと予測  
成功率の間にはあきらかに相関性がある。図 7 は、  
こうした観測結果を裏付けており、全体的に見れば、  
相関性が認められる。

ただし、プログラムによっては、 $x$  軸値によらず  $y$  軸  
値がほとんど変化しない場合があるなど、必ずしも相  
関が厳密に守られているわけではない。また、bimode  
では、いくつかのプログラムで、奇妙な軌跡を描いて  
いることが確認されている。これは、bimod 予測器の  
予測性能が、必ずしも情報源の性質 (マルコフ情報源  
としての性質) に依らないことを意味している。

#### 4.6 参照エントロピーと予測成功率との関係

前節と同様にして、参照エントロピーについて式 (3)  
をもとにエントロピーを求める。本評価では、シミュ  
レーションによって得られた 16 次および 12 次の参照  
エントロピーから  $(H(\bar{R}^{16}) - H(\bar{R}^{12}))/4$  によって求  
めている。結果を図 8 に示す。

2 level, gshare の場合には、図 7 よりも良好な相関  
性を示している。逆に、bimode の場合は相関性が全  
く見られない。

#### 4.7 考察

個々のプログラムについて見ると、 $n$  次参照エント  
ロピーと予測成功率の時間経過のグラフ (図 6) のよ  
うに、相関関係を見出すことができる。これは図 7、  
図 8 によって確認できるが、その一方で、必ずしも明  
確な相関性が得られるケースばかりではないことも明

らかになった。2 level, gshare の場合は、分岐履歴か  
ら直接的に参照テーブルのエントリを求めることから、  
本稿で扱ったエントロピーと予測成功率の相関性が強  
い。一方 bimod では、予測成功率は各分岐命令での  
分岐確率に大きく依存し、マルコフ情報源としての性  
質 (エントロピー) があまり反映されない。図 7、図 8  
の結果は、この事も示している。

## 5. おわりに

本稿では分岐予測器を主な対象として、系に現れる  
ランダムネスに着目しエントロピーの考えを取り入れ、  
エントロピーと予測器の性能について関係を明らかに  
した。本稿では、対象とする系を 2 値マルコフ情報源  
とみなした場合の情報量 (エントロピー)、また、予測  
器内のテーブル構造の使用状況の偏りを対象としたエ  
ントロピーの 2 つを考えた。SPEC CPU2000 ベンチ  
マークでの評価の結果、エントロピーと予測性能は概  
ね相関関係にあることを明らかにした。

一般に予測器では資源を多く用意するほど全体性能  
が向上する。しかし、本稿で行ったエントロピーの観  
点からは、偏りが顕著でエントロピーが小さいほど予  
測性能は高くなる。すなわち、予測器の持つ資源の中  
で有効に使われている正味の量はわずかである。様々  
な予測方式が提案されるなか、今後、本稿で行ったよ  
うな観点から、効率よく高性能な予測方式を求めたい。  
謝辞 本研究は、一部日本学術振興会科学研究費補助  
金 (基盤研究 (B)14380135, 同 (C)16500023, 若手研  
究 (B)17700047) の援助による。

## 参 考 文 献

- 1) 斎藤 盛幸, 古川 文人, 大津 金光, 横田 隆史,  
馬場 敬信, “投機的マルチスレッド実行のための  
限定的 2 パス予測方式の検討,” 情報処研報, ARC-  
158-2, 2004 年 5 月。
- 2) T. Yokota, M. Saito, F. Furukawa, K. Ootsu  
and T. Baba, “Prediction and Execution  
Methods of Frequently Executed Two Paths  
for Speculative Multithreading,” *Proc. 16th  
IASTED PDCS 2004*, pp.796–801, Nov. 2004.
- 3) 増保 智久, 道口 貴史, 斎藤 盛幸, 大津 金光,  
横田 隆史, 馬場 敬信, “MediaBench ホットルー  
プ並列化のためのパスプロファイリング,” 情報処  
大, pp.1-203-1-204, 2005 年 3 月。
- 4) 横田 隆史, 斎藤 盛幸, 大津 金光, 古川 文人,  
馬場 敬信, “2 パス限定投機方式の提案,” 情報論  
ACS, 2005. (掲載予定)
- 5) SimpleScalar LLC, <http://www.simplescalar.com/>

---

履歴ビット長の取り方によりバラつくことがあるためである。

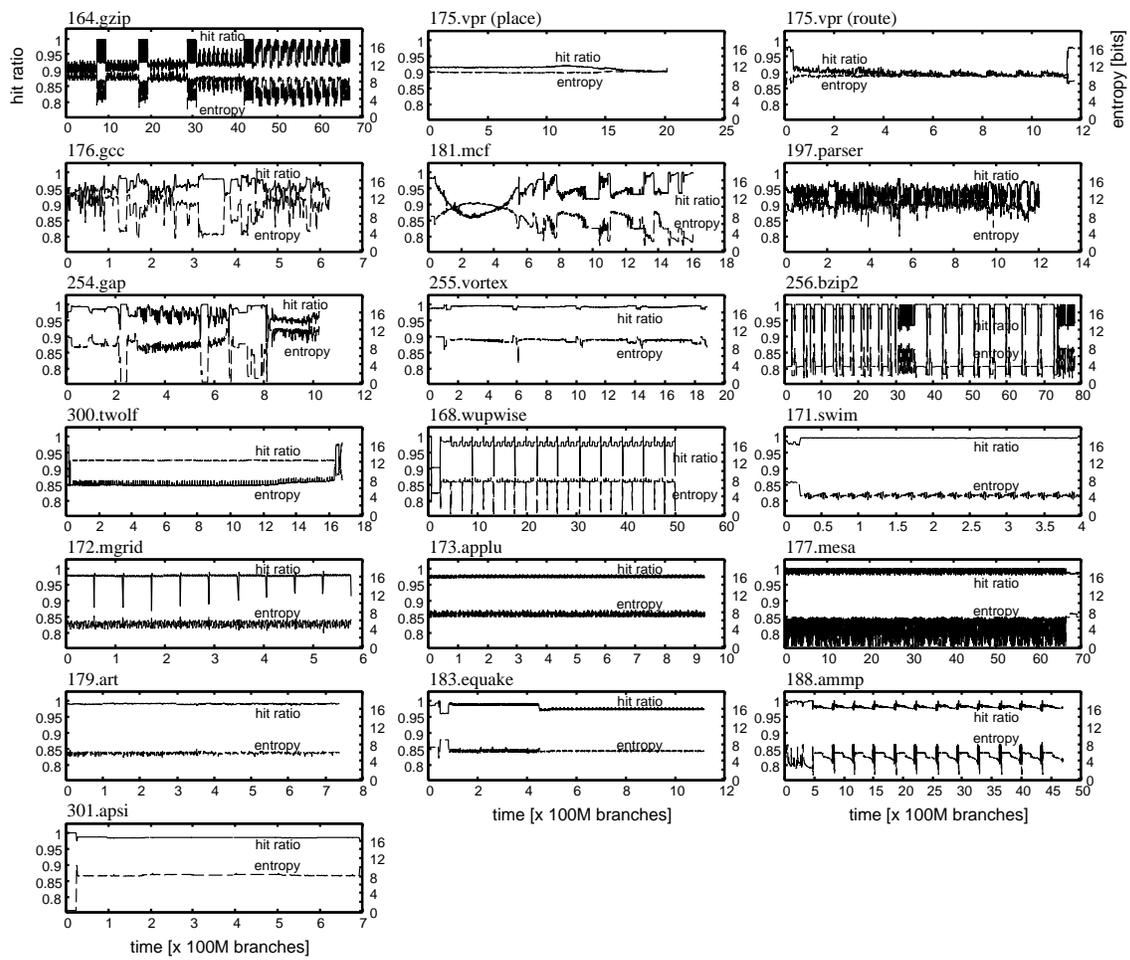


図 6 エントロピーと予測成功率の時系列変化

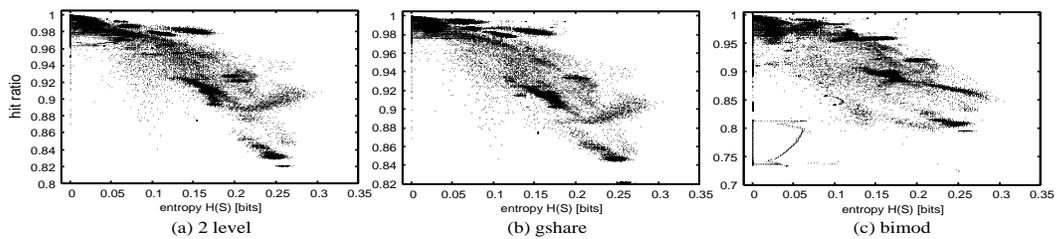


図 7 情報源エントロピーと予測成功率の関係

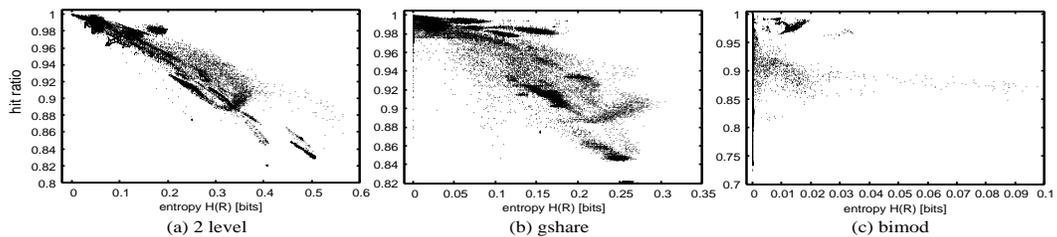


図 8 参照エントロピーと予測成功率の関係