定例進捗報告 2020年12月01日(火) 15:15〜

修士2年 M2019SE009 渡邉将匡

〜変える〜

研究テーマ：命令型プログラミング言語におけるプログラム可逆化

背景：

目的：平均ゴミ出力量の最適化

可逆性の定義：

可逆プログラムの定義：

状態：

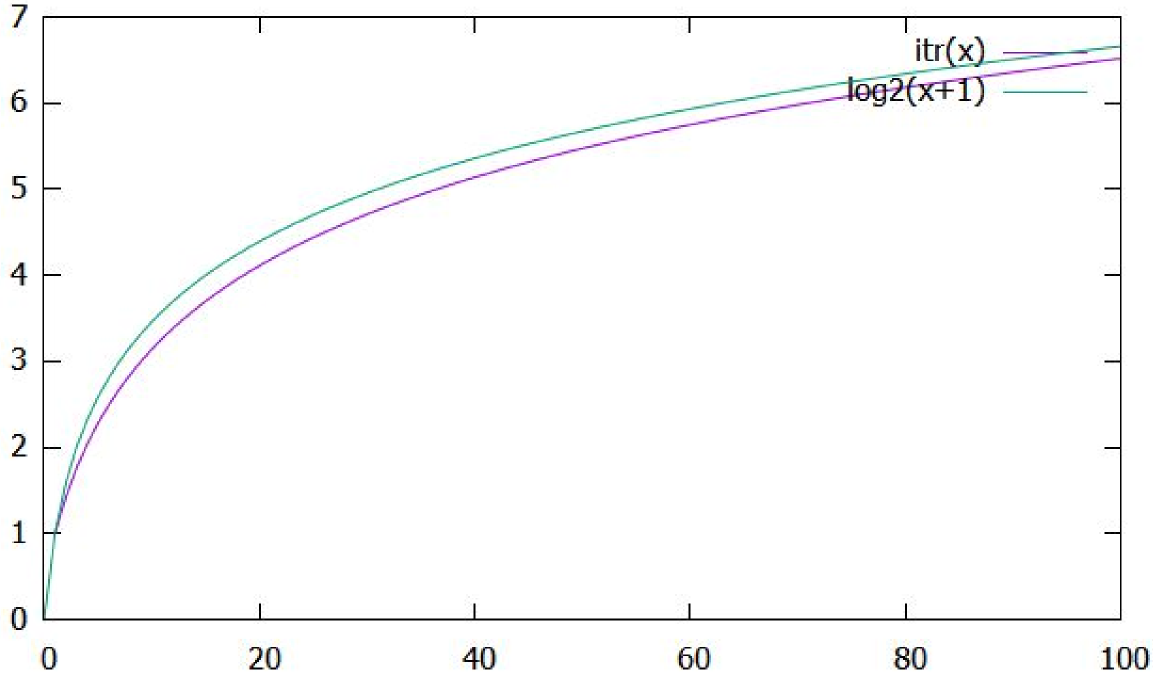
〜ここまで〜

先週からの進捗状況：

1. 前回の報告で要チェックとなった部分を再び調べた

T3 繰り返し回数が最大n回の繰返し
ゴミ出力量　|~ log2(n+1) ~|
・繰り返し回数が1回の場合も1ビット必要なので|~ log2(n) ~|だと足りない

エントロピー　1/n Σ^n_{i=1} i*(-(1/(i+1))log2(1/(i+1)) - (i/(i+1))log2(i/(i+1)))
・繰り返し回数がi回の場合のエントロピー
　・繰返し1回ごとのエントロピーは-(1/(i+1))log2(1/(i+1)) - (i/(i+1))log2(i/(i+1))
　・i回の繰返しが起こるのでエントロピーにiを掛ける
・繰り返し回数の確率分布は一様分布とする
　・繰り返し回数がi回になる確率は1/n
・グラフの形はlog2(n+1)とだいたい同じになる



gnuplotのコード

```
set samples 101
set xrange [0:100]
log2(x) = log(x)/log(2)
f(n) = (n > 0 ? -n*log2(1/(n+1))/(n+1) - n*n*log2(n/(n+1))/(n+1) + f(n-1) : 0)
itr(n) = f(n)/n
plot itr(x), log2(x+1)
```

T4 関数呼出し（末尾再帰に関して）

ゴミ出力量
- 末尾再帰では，1回の再帰ごとに2分岐と同じ量のゴミが出力される
- n回の再帰が起こるとゴミ出力量はnビットになる

エントロピー
- われわれの手法では繰返しと同じになる
  - 再帰が起こらないのは最後の1回だけ
- 1/n Σ^n_{i=1} i*(-(1/(i+1))log2(1/(i+1)) - (i/(i+1))log2(i/(i+1)))
  - だいたいlog2(n+1)になる

線形探索

前提
- 事前に確率の見積もりが可能
- 値は重複しない
- データ数はn個

データが見つかる場合

確率
・m番目で見つかる　1/(n-m+1)
・m番目で見つからない　(n-m)/(n-m+1)
エントロピー
・Σ^m_{i=1} -((n-i)/(n-i+1))log2((n-i)/(n-i+1)) - (1/(n-i+1)log2(1/(n-i+1)))
　・第1項 < 第2項 < ... < 第m項
　　・ただし，m=nの場合は除く
　・m=n-1の場合，第m項=1
Perumallaの手法
・データを1回比較するごとに1ビット必要
・m番目で見つかった場合はmビットのゴミ出力量
Perumallaの手法との比較
・Perumallaの手法のゴミ出力量をエントロピーと同じ形に書き直すと
　・Σ^m_{i=1} 1
・エントロピーはすべての項が1以下なので，Perumallaの手法よりも常にエントロピーの方が小さい
算術符号を用いた場合，平均ゴミ出力量はエントロピーに近づく．よって算術符号を用いると平均ゴミ出力を減らすことができる

データが見つからないことがある場合
確率
・m番目で見つかる　1/(100億*n)
・m番目で見つからない　(100億*n-1)/(100億*n)
　・mに比べて100億*nが十分に大きいため，mを無視する
見つからない場合の情報量　-log2((100億*n-1)/(100億*n)) ≒ ほぼ0
　・(100億*n-1)/(100億*n)はほぼ1
見つかる場合の情報量　-log2(1/(100億*n))
　・巨大な情報量になる
エントロピー　-((100億*n-1)/(100億*n))log2((100億*n-1)/(100億*n)) - (1/(100億*n))log2(1/(100億*n)) ≒ ほぼ0
・-((100億*n-1)/(100億*n))log2((100億*n-1)/(100億*n)) ≒ ほぼ0
・log2(100億*n) < 100億*n
　・-(1/(100億*n))log2(1/(100億*n)) << 1
Perumallaの手法の平均ゴミ出力量　n
・ほとんど見つからないため，見つかった場合のゴミ出力量を無視する
算術符号を用いた場合，Perumallaの手法と比べて平均ゴミ出力量は下がるが，データが見つかった場合はゴミ出力量が極端に増える

次回までに進めること
1.