

シミュレーテッドアニーリングプログラミングの温度並列化

松井 勇樹

1 はじめに

近年、コンピュータによりプログラムを自動生成する研究が行われている。そこで我々は、自動プログラミングの手法の 1 つであるシミュレーテッドアニーリングプログラミング (Simulated Annealing Programming:SAP)¹⁾ を研究している。SAP とは、シミュレーテッドアニーリング (Simulated Annealing:SA)²⁾ を木構造が扱えるように拡張させた手法である。この手法の重要なパラメータに温度が存在するが、その決定は容易でない。SA では、温度並列化によってこの問題を解決した有効な手法として、温度並列 SA (Temperature Parallel Simulated Annealing:TPSA)^{3, 4)} がある。TPSA は、複数のプロセッサがそれぞれ異なる一定の温度で探索を行い、ある周期ごとに解交換を行う手法である。温度並列化により、温度スケジュールの自動化だけでなく、局所解に陥りにくい探索が期待できる。本研究では、このアルゴリズムを SAP に適用し、逐次 SAP を用いた並列手法と比較することで SAP における温度並列化の有効性を示す。

2 温度並列 SAP

温度並列 SAP とは、温度並列化によって SAP の温度スケジュールを自動化した手法である。以下に、SAP、及び TPSAP についての詳細を述べる。

2.1 シミュレーテッドアニーリングプログラミング

SAP は、金属の焼き鈍しを模倣した進化的最適化手法である SA を木構造が扱えるように拡張したプログラム探索手法であり、GP における突然変異をベースに探索を行う。SAP のアルゴリズムを以下に示す。

1. 生成処理

現在の解に対してランダムに挿入点を選択し、挿入点を根とする部分木を削除する。その部分にランダムに生成した部分木を挿入し、新しく生成した次解候補の評価値を得る。

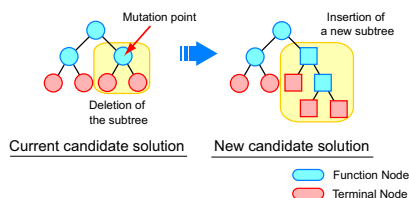


Fig.1 生成処理

2. 受理判定, 状態遷移

新しく生成した解候補を受理するか判定を行う。判定には、温度 T 、および現在の解の評価値 E と新しい解候補の評価値 E' との差分 ΔE を用いた

Metropolis 基準 (式 (1)) を用いる。

$$P = \begin{cases} 1 & \text{if } \Delta E \leq 0 \\ \exp(-\frac{\Delta E}{T}) & \text{otherwise} \end{cases} \quad (1)$$

3. クーリング

1 と 2 を繰り返し行うことをアニーリングと呼び、ある一定期間アニーリングを行った後に、式 (2) を用いて温度を下げる。

$$T_{next} = \gamma T_{current} \quad (0.8 \leq \gamma < 1) \quad (2)$$

2.2 温度パラメータの役割

SAP (SA) の大きな特徴は、次解候補が改悪方向へ生成された場合でも、その解候補への遷移を確率的に認めることである。その確率は式 (1) の Metropolis 基準により決定され、改悪方向への遷移は温度パラメータ T に依存する。そのため、温度パラメータのスケジューリングは、解探索に大きな影響を与える。この温度スケジュールは一般的に、温度を高温から低温にするクーリングと温度を固定した一定温度の 2 通りに分けられる。

2.3 温度並列 SAP

温度並列 SA を SAP に適用した温度並列 SAP について以下に詳細を示す。

1. 初期設定

各プロセスに異なる温度を設定する。各温度は、最高温度と最低温度、および最高温度と最低温度の間を使用するプロセス数で等比的に分割した温度とする。

2. 温度を固定した SAP

各プロセスが、初期設定で与えられた温度を基に一定温度を温度スケジュールとした SAP で解の探索を行う。

3. 交換判定, 解交換

アニーリングを一定周期 (解交換周期 k) まで繰り返した後、解の交換を行う。隣接するプロセスが持つ解の評価値 E と E' との差分 ΔE 、および隣接するプロセスが持つ温度 T と T' との差分 ΔT により、プロセス間で解交換を行うかどうかの判定を行う。交換判定には、式 (3) を用いる。

$$P(\Delta T, \Delta E) = \begin{cases} 1 & \text{if } \Delta T \cdot \Delta E \leq 0 \\ \exp(-\frac{\Delta T \cdot \Delta E}{T \cdot T'}) & \text{otherwise} \end{cases} \quad (3)$$

温度並列化では、通常のクーリング処理で温度 T から T' に冷却すること (Fig. 2-a) が、温度 T のプロセスと温度 T' のプロセスの間で解を交換すること (Fig. 2-b) に

相当する。また、温度スケジュールを設定することは、温度並列化ではプロセス間で解の交換をいつ行うかを指定すること (Fig. 2-b) に相当する。

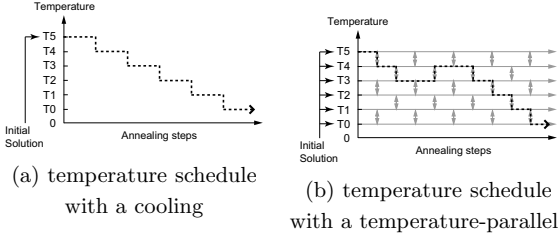


Fig.2 クーリング及び温度並列化を用いた温度スケジュール

3 対象問題

テスト問題は、自動プログラミングの代表的なベンチマーク問題である Santa Fe trail 問題⁵⁾、および Simple Symbolic Regression 問題⁵⁾とする。Santa Fe trail 問題は解の評価に影響を及ぼさないノードが発生する問題である。一方、Symbolic Regression 問題は全てのノードが解の評価に影響を及ぼす問題である。

3.1 Santa Fe trail

Santa Fe trail 問題とは、1 匹の人工蟻が Fig. 3 に示す 32×32 のマス目上に配置された餌を、限られたエネルギー内でできるだけ多く獲得するプログラムを生成する問題である。人工蟻は餌上を通ることにより餌を獲得することができ、終端記号が 1 つ実行されるごとにエネルギーは 1 消費する。人工蟻の初期エネルギーは 400 である。

この問題に用いる非終端記号は {if_food_ahead, progn2, progn3}, 終端記号は {MOVE, RIGHT, LEFT} とした。if_food_ahead は引数を 2 個持ち、人工蟻の 1 マス前方に餌があれば第 1 引数、無ければ第 2 引数を実行する。progn N は引数を N 個持ち、第 1 引数, 第 2 引数, ..., 第 N 引数の順に実行する。

評価関数 E_{val} 式 (4) は、餌の総数である $F_{max}(= 89)$ から人工蟻が獲得した餌の数 F を引いたものであり、すべてのエサを獲得すれば ($E_{val}=0$)、探索が成功したとする最小化問題である。

$$E_{val} = F_{max} - F \quad (4)$$

3.2 Simple Symbolic Regression

Simple Symbolic Regression とは、未知の関数 $y = f(x)$ に対して n 個の入出力データを用いて関数 f を同定する問題である。同定する目的関数は式 (5) に示す f_{obj} である。その概形を Fig. 4 に示す。

$$f_{obj} = x^4 + x^3 + x^2 + x \quad (5)$$

この問題に用いる非終端記号は {+, -, *, %, sin, cos,

exp, rlog}, 終端記号は { x } とした。なお, % は剰余, rlog は自然対数である。

評価関数 E_{val} は、-1 から 1 の間を 0.1 刻みにした 21 個の入力に対する出力誤差の和とし、 $E \leq 0.01$ を最適解とする最小化問題とした。ここで、 $prog$ は生成されたプログラムを示す。

$$E_{val} = \sum_{i=0}^{20} |prog(x_i) - f(x_i)| \quad (6)$$

$$x_i = 0.1i - 1 \quad (7)$$

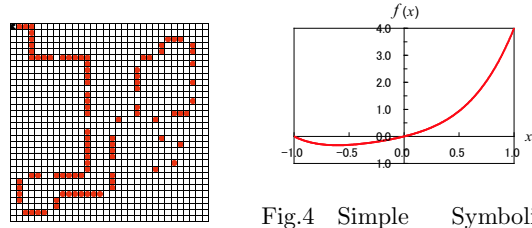


Fig.3 Santa Fe trail 問題 Fig.4 Simple Symbolic Regression 問題の目的関数の概形

4 温度並列化による性能比較

4.1 実験概要

SAP における温度並列化の有効性を検証するために、3 章の各問題に対して比較実験を行う。比較する手法を以下に述べる。

- 温度並列 SAP (以下, TPSAP)
 n 温度の温度並列 SAP を, プロセス数 n で実行する。解交換は解交換周期 k 毎に行う。
- 解交換を行わない温度並列 SAP (以下, TPSAP-NoExchange)
 n 温度の温度並列 SAP を, プロセス数 n で実行する。解交換は行わない。
- 並列 SAP (以下, Parallel SAP : PSAP)
クーリングの温度スケジュールを用いた逐次 SAP を, プロセス数 n で同時に実行する。解交換は行わない。

4.2 パラメータ

表 1 には、Santa Fe trail 問題、Simple Symbolic Regression 問題に用いた各パラメータを示す。

Table1 用いたパラメータ

Parameter	Santa Fe trail	Symbolic Regression
Number of evaluation	400,000	200,000
Number of processes	32	32
Maximum temperature	129	1579
Minimum temperature	0.28	7.9×10^{-3}

なお、最高温度は最大改悪を 50% は認める温度、最低温度は最小改悪を 1 回は認める温度とする。TPSAP、及

び TPSAP-NoExchange の各プロセスに用いた温度は、最高温度、最低温度、およびその間を等比的に分配した温度、PSAP の各プロセスにおける初期温度は最高温度とした。また、解交換周期は一般的に用いられる値である 40 回⁴⁾とし、PSAP に用いられるクーリング数は、用いるプロセス数と同じ 32 とした。

4.3 実験結果

Santa Fe trail 及び Simple Symbolic Regression において、各手法を 50 試行を行った時の成功率の履歴を Fig. 5 に示す。なお、Fig. 5 は、横軸に探索回数 (Annealing steps = Number of evaluation / Process number)、縦軸に成功率を示す。ただし、各手法における最良値は、全プロセス中の探索において得た最も良好な解の評価値とする。

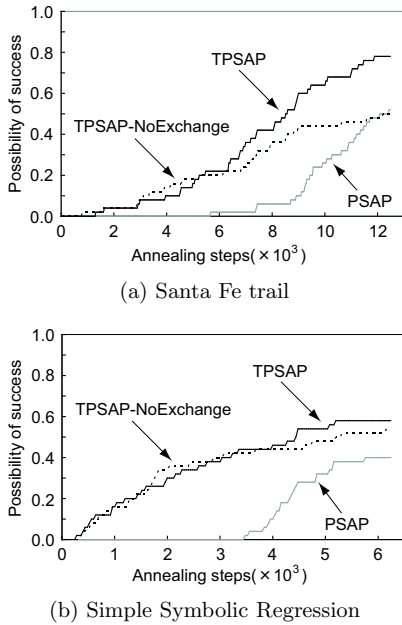


Fig.5 探索成功率

Fig. 5 より、Santa Fe trail 及び Simple Symbolic Regression 共に TPSAP が良好な解を得られることがわかる。また、構文的イントロンを含まない Simple Symbolic Regression に比べ、構文的イントロンを含む Santa Fe trail の方が、探索性能の差が大きく出る結果となった。以上のことから、TPSAP は逐次 SAP を用いた並列手法と比べ、良好な結果を得られることがわかった。

4.4 考察

4.3 節に示した通り、TPSAP の方が逐次 SAP を用いた並列手法と比べ、高性能であることがわかった。TPSAP の高性能である要因として、TPSAP では各温度におけるプロセス間で解交換が行われているため、良好な探索が行われたと考えられる。また、4.3 節での実験では、用いた温度数は表 1 に示した様に 32 温度としたが、この温度数が適当であったか検討する必要があると考えられる。このことから、以下では TPSAP における解のプロセス間の遷移、及びプロセス数における比較について考察を

行う。

4.4.1 最良解のプロセス間の遷移

Santa Fe trail 及び Simple Symbolic Regression において、TPSAP で得たある試行の最良解の遷移履歴を Fig. 6 に示す。なお、Fig. 6 は、横軸に探索回数、左縦軸に評価値、右縦軸にプロセス番号を示す。プロセス番号は最も高いものが最高温度、最も低いものが最低温度とする。

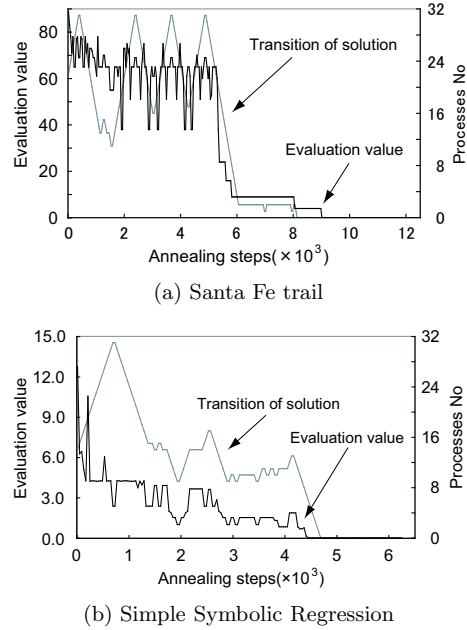


Fig.6 最良解の遷移履歴

Fig. 6 - (a) より、探索序盤に注目すると解は高温のプロセス付近を遷移しており、大域的探索を行っていることがわかる。探索中盤において、評価値が大幅に改良されるに伴い、解は低温のプロセスに遷移していることから、局所的探索に切り替わっている。探索終盤では、解のプロセス遷移、及び評価値の更新が頻繁に行われなくなるが、徐々に評価値が低くなり、最適解を得ることが出来ている。

一方、Fig. 6 - (b) より、Fig. 6 - (a) と同様に探索序盤では、評価値が高いため、解は高温のプロセス付近を遷移しており、大域的探索を行っていることがわかる。探索中盤では、評価値の更新は行われているが、解が低温のプロセスに遷移せず、中盤の温度領域を行き来していることから、他のプロセスに比べ、探索は停滞していると言える。しかし、終盤では良好な解が得られたため、評価値が更新され、それに伴い、解も低温のプロセスに遷移しており、最適解を得ることが出来ている。

4.4.2 プロセス数における比較

TPSAP における用いるプロセス数に関する検証実験を行った。用いたプロセスは 2 から 64 を等比的に分割したプロセス数とし、用いた最高、最低温度は表 1 に示した温度と同様とする。また、Santa Fe trail 問題における評価計算回数は 20 万回と 40 万回、Simple Symbolic Regression 問題における評価計算回数は 10 万回と 20 万

回とした。

Santa Fe trail 及び Simple Symbolic Regression において、各プロセス数の 50 試行を行った時に得られた成功率を図 7 に示す。なお、図 7 は、横軸にプロセス数、縦軸に成功率を示す。

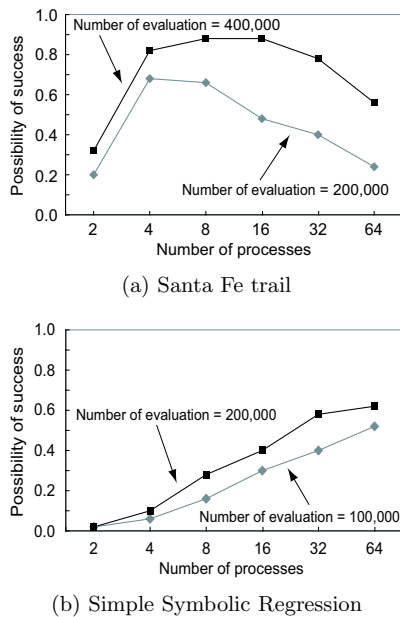


Fig.7 プロセス数における比較

Fig. 7 - (a) より、Santa Fe trail では用いるプロセス数に最適な数があることがわかった。評価計算回数 20 万回ではプロセス数 4, 8 付近、40 万回ではプロセス数 8, 16 付近で最も探索性能が高く、プロセス数が多すぎると成功率が低下する傾向がある。また、プロセス数 16 において、評価計算回数が 10 万回ではピーク性能より低下しているのに対し、評価計算回数 40 回ではピーク性能に達していることから、Santa Fe trail では探索性能向上のためには 1 プロセスにおける計算回数が一定回数以上必要と考えられる。

一方、Fig. 7 - (b) より、Simple Symbolic Regression では用いるプロセス数が多いほど探索性能が向上することがわかった。また、評価計算回数 10 万回に対し、20 万回の方が探索性能は向上しているが、プロセス数を増やす方が探索性能の向上が顕著であることから、Simple Symbolic Regression では、探索性能向上のためには一定数のプロセス数が必要と考えられる。

以上のことから、Santa Fe trail における探索性能の向上のためには、プロセス数を変更するより、評価計算回数を増やす方が探索性能の向上が見込める。一方、Simple Symbolic Regression における探索性能の向上のためには、評価計算回数を増やすよりも、プロセス数を増やす方が探索性能の向上が見込めることがわかった。

4.5 まとめ

本論文では、シミュレーテッドアニーリングプログラミングの温度並列化についての有効性を検討した。数値実験の結果、温度並列 SAP は、逐次 SAP を用いた並列手

法に比べ、良好な解を得ることができた。これは、温度並列 SAP の解の遷移から、解が温度を適応的に変化しているためと考えられる。また、各対象問題において、Santa Fe trail ではプロセス数 8、Simple Symbolic Regression ではプロセス数 64 で最も良い探索性能が出ることが判った。このように、対象問題によって、最も探索性能が向上するプロセス数が存在し、また、評価計算回数、またはプロセス数を変更することによって、効率的に探索性能が向上できることが判った。以上のことから、温度並列 SAP は SAP の平行アルゴリズムとして、非常に有効であると言える。

参考文献

- 1) 藤田佳久, 三木光範, 橋本雅文, 廣安知之. シミュレーテッドアニーリングを用いた自動プログラミング. 情報処理学会論文誌, Vol. 48, No. SIG15, pp. 88-102, 2007.
- 2) Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E. Equation of state calculation by fast computing machines. *Journ. of Chemical Physics*, Vol. 21, pp. 1087-1092, 1953.
- 3) 小西健三, 瀧和男, 木村宏一. 温度並列シミュレーテッドアニーリング法とその評価. 情報処理学会論文誌, Vol. 36, No. 4, pp. 797-807, 1995.
- 4) 三木光範, 廣安知之, 笠井誠之. 連続最適化問題への温度並列シミュレーテッドアニーリングの応用. 情報処理学会論文誌, Vol. 41, No. 5, pp. 1607-1616, 2000.
- 5) John R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, 1992.