

# 可逆グラフアルゴリズム

2016SE085 鳥居大樹 2016SE098 吉田翔亮

指導教員：横山哲郎

## 1 はじめに

本研究では、可逆アルゴリズムの効率化の整備が進んでいないグラフアルゴリズムの可逆化および効率化を目標とする。また、本研究では、グラフアルゴリズムの中でも基本的なアルゴリズムである深さ優先探索と幅優先探索を取り扱う。

### 1.1 背景

計算過程においての可逆とは、直前の状態が高々一意に定まるもので、可逆な場合、状態から状態への遷移が1:1の関係になっている。アルゴリズムは計算機上で問題を解くための手法である。アルゴリズムには線形探索アルゴリズムやバブルソートをはじめとした多くのアルゴリズムが存在し、その中に連結無向グラフの深さ優先探索や幅優先探索が存在する。実際にプログラムを記述して実行させると計算過程で情報を失うことによって、エネルギーを消費し熱を発生させる。これは非可逆計算を行うと必ず発生してしまうことが分かっている。反対に、可逆計算では計算過程で情報を失わないのでエネルギーの消費や熱を発生の下限が存在しない。非可逆なアルゴリズムを可逆化させ、可逆アルゴリズムにする。すなわち、各実行状態に移る計算が単射であり、直前の状態が高々一意に定まるプログラムで記述すると非可逆なプログラムで本来発生するはずだったエネルギーの消費や熱の発生が抑えられる可能性がある。また、基本的な探索アルゴリズムの一つである線形探索では、提唱されている一般解法を用いた可逆化が行われている。加えて、手動で可逆化することで一般解法よりも時間計算量の効率の良いアルゴリズムも提案されている [2]。また、深さ優先探索や様々な比較ソートでも可逆化が行われている [1][3]。しかし、深さ優先探索の可逆化は手動による効率化が不十分であると考えられる。そのため、効率のいい手動の解法の提案が求められている。加えて、似たケースで使われる幅優先探索の可逆化も求められている。

### 1.2 目的

本研究では、はじめにすでに解析された可逆線形探索とか逆深さ優先探索の解析方法の改善を行う。参考文献 [1][2] では、解析を行ごとに行われている。しかし、計算量に多く関わる重要な演算を決めその演算の計算量を求めるほうが解析の効率が良いと考えられる。解析の方法の変更による結果の変化を明確にする。また、連結無向グラフの深さ優先探索の既に可逆化された一般解法と比較してより効率的な解法を提案すること、連結無向グラフの幅優先探索の一般解法による可逆化と手動による効率的な解法を

提案することをを目標とする。可逆化したプログラムと効率化したプログラムの解析を行い、トレードオフの関係を明確にする。

### 1.3 アプローチ

はじめに、C言語のプログラムを用意し、一般解法を用いて可逆化を行う。そして可逆化されたプログラムの解析を行い、時間計算量がどの程度かかるのかを把握する。その後可逆化されたプログラムの改良を行い時間計算量の効率化を図る。最後に改良したプログラムの解析を行い、どの程度効率化できたかを明確化する。

## 2 関連研究

### 2.1 探索アルゴリズム

探索アルゴリズムとは、入力されたデータの集合の中から目的となる値や状態を見つけるアルゴリズムである。探索アルゴリズムの分類としてリスト探索、木探索・グラフ探索、文字列探索といったものが挙げられ、本研究で扱う深さ優先探索及び幅優先探索は木探索・グラフ探索の分類に属する。また木探索・グラフ探索の分類には他にも、双方向探索や反復深化深さ優先探索等が挙げられ、その中で深さ優先探索と幅優先探索は基本的なアルゴリズムといえる。

### 2.2 線形探索

通常線形探索では探したいキーの値と配列の先頭の値が一致するか確認し、一致すれば終了し、一致しなければ、確認した値が配列の最後であるかを確認し、最後の値でなければ配列の次の値を確認し、キーの値と一致するかもしくは配列の最後の値を探索するまでこの探索をつづける。探索成功の場合1、失敗の場合0を返し成否を判定する。通常線形探索の他に番兵法や、整列リストを用いた線形探索がある。番兵法は配列の最後にキーと同じ値を加え、通常線形探索と同じ方法で探索を行い、キーと同じ値を見つけたときに配列の最後に加えた値かどうかの確認を行い、加えた値ならば失敗とし、加えた値でないならば、成功とする。通常線形探索と比べて、キーと同じ値を見つけたときにだけ、配列の最後かどうかの確認を行えばよいので、比較回数を減らすことができる。整列リストを用いた線形探索はあらかじめ配列の値が昇順または降順にソートされている場合に用いることができる。昇順の場合で考えると、まず番兵法のように配列の最後に を加え、先頭から探索を行い、探索している配列の値がキーよりも小さい値ならば次の値を探索し、キー以上の値だった場合キーと一致するかどうかを確認し、一致した場合成功、しなかった場合は失敗とする。

## 2.3 グラフ

グラフは隣接リストの集合による表現と隣接行列による表現の2種類の表現方法がある。本研究では隣接リストによる表現を用いる。隣接リストによる表現は  $G = (E, V)$  で表現される。  $V$  は頂点の集合  $E$  は頂点間を結ぶエッジの集合を表す。グラフには、エッジが有向と無向の2種類がある。また、エッジに重みを追加することで、最短経路を求めるアルゴリズムに用いられる。このようにグラフを用いることによって、様々なアルゴリズムを表現することができ、本研究で取り扱う深さ優先探索や幅優先探索にも用いられている。本研究では、無向グラフ表現によるグラフアルゴリズムを取り扱う。

## 2.4 グラフアルゴリズム

グラフアルゴリズムには様々な探索方法が存在し、探索方法を分類すると、深さ優先探索、幅優先探索、双方向探索、分枝限定法の4つの基本的な探索方法に分類される。深さ優先探索を応用したのものには、深さ制限探索、反復深化深さ優先探索がある。幅優先探索を応用したのものには、最良優先探索、均一コスト探索、 $A^*$  がある。基本的な探索方法である深さ優先探索と幅優先探索の効率化できれば同様のアプローチもしくは、それに近い方法で応用した探索は効率化できる可能性が高い。そのため深さ優先探索と幅優先探索は、グラフアルゴリズムの中でも基本的なアルゴリズムだといえる。本研究では、グラフアルゴリズムの中の基本的なアルゴリズムの効率化を行うことによって、応用的なグラフアルゴリズムの効率化にも貢献できると考える。

## 2.5 深さ優先探索

深さ優先探索は縦型探索とも呼ばれ、始点からの距離が大きい未探索の頂点を優先的に探索するアルゴリズムである。手順は、はじめに、与えられた始点と隣接する頂点を探索済みにし隣接する頂点がすべて探索済みになるまで行う。隣接済みの頂点がすべて探索済みの場合、ひとつ前の頂点に戻る。これを繰り返し、すべての頂点が探索済みになるか、探索したい頂点が見つかったときに探索が終了する。始点からの距離が無限にあるグラフでは無限の深さまで探索してしまうため、解が必ず見つけられるとは限らない。また始点からの距離が大きい頂点を優先的に探索する性質があるため、解が複数個ある時最小経路が最初に見つかるとは限らない。

## 2.6 幅優先探索

深さ優先探索とは反対に横型探索とも呼ばれ、始点から距離が小さい頂点を優先的に探索する。手順は、始点からの距離が小さい頂点を順に探索する。これを繰り返し、すべての頂点が探索済みになるか、探索したい頂点が見つかったときに探索が終了する。深さ優先探索とは反対に始点からの距離が小さい頂点を優先的に探索する性質がある

ため、解が複数個ある時最小経路が最初に見つかる。

## 2.7 一般解法

可逆化を行う一般解法には幾つかの方法が存在し、その中で Landauer 法と Bennett 法と呼ばれるものがある [4]。Landauer 法は、代入などの非可逆な計算では情報の消失が起こる。そこで非可逆な計算が行われる前に非可逆な計算によって消失する情報を別の値に保存することで可逆化を実現した。しかしこの方法は元の非可逆な計算の失われる情報を保存するため、実行時間に比例したメモリ使用量が必要となる。Landauer 法の欠点として出力が必要以上に出てしまうということが挙げられる。Bennett 法は Landauer 法の欠点を解消する解法で、プログラムを実行した後に出力に有効な情報を保存し、その後逆実行を行う。逆実行を行うことでスタックを空にし、メモリ使用量を抑えることができる。しかし、逆実行を行うことによって実行時間が約2倍になってしまうという欠点がある。また、Lange-McKenzie-Tapp 法と呼ばれるものもあり、チューリング機械を可逆的にシミュレートする方法で、他の計算モデルにおいても一般化できる。この方法のメリットは、構成サイズがある制限の値を超えると探索を中断するので、空間計算量の値を少なくできる。また最悪の場合の漸近的な実行時間が増えない点もある。

## 2.8 Janus

命令型可逆プログラミング言語 Janus で書かれたプログラムは可逆性が保証されている。非可逆な言語において条件分岐文や繰り返し文は可逆性を持たない。しかし、Janus ではこれらの文も可逆性が保証されている。本研究では可逆なプログラムを作成する際に、Janus で実装する。

## 2.9 O 記法

計算量を漸的に表す記法で  $O$  で表される。 $f(x) = 3x + 2(x)$  で計算量を表せるプログラムがあるとすると、 $O(x)$  と表す。 $O$  は関数  $f(x)$  を実行するのに、 $x$  の定数倍の時間程度しかかからず、関数  $f$  は漸的に  $x$  によって上から抑えられることを表す。反対に  $\Omega$  記法と呼ばれる  $\Omega(x)$  では、漸的に  $x$  によって下から抑えられることを表す。

# 3 グラフアルゴリズムと可逆化

## 3.1 アルゴリズムの解析

実際にアルゴリズムをプログラムで実装すると、同じアルゴリズムでも実装の方法によってプログラムの動き方が変わってしまう。そのため、プログラムをある指標を用いて数値に表し、同じ指標を用いて数値に表した別のプログラムと比較することでそれぞれのプログラムの良さを把握することができる。このプログラムの良さを見積もることをアルゴリズムの解析と呼ぶ。本研究ではプログラムの良さを表すための指標として時間計算量、空間計算量、ゴミ出力量の3つを用いる。

### 3.2 時間計算量

時間計算量はアルゴリズムを実行した際にどの程度時間がかかるかを表す。しかし、アルゴリズムの実行時間はプログラムや実装する計算機、入力されたデータの量によって変化する。そのため、一般的にはオーダー記法という考え方をういてアルゴリズムそのものの効率を漸近的に考える。具体的にはそのアルゴリズムにとって最も重要とされる演算を基本演算と呼び、基本演算が実行される回数を求めるだけで良いとされる [5]。アルゴリズム同士ではオーダーを用いて比較することが望ましいが、同じアルゴリズムを別々のプログラムで実装した場合だと、大抵は同じ結果が出てきてしまう。そのため、本研究では、可逆プログラムは全ての行を同一の時間で計算するものとし、行毎に実行回数を評価して時間計算量を求める。

### 3.3 空間計算量

空間計算量はプログラムを実行した際にどの程度記憶領域を必要とするかを表す。ここでは変数を 1 つ新たに用意すると記憶領域を 1 使用するものとする。すなわち変数や配列を宣言するか、スタックに push した場合に空間計算量が増える。

### 3.4 ゴミ情報

ゴミ情報は可逆アルゴリズム特有の考え方で、出力データの中で問題解決に必要な情報を保持している出力データ以外の本来は必要でない出力データをゴミ情報と呼ぶ。非可逆アルゴリズムでは必要な情報のみを出力すれば良いが、可逆アルゴリズムでは逆実行する際の入力として必要な情報も出力される場合があるためゴミ情報と呼ばれるデータが出てくる。また、実行途中で出てくるゴミ情報を中間ゴミ、実行終了時に出てくるゴミ情報を最終ゴミと呼ぶ。

## 4 アルゴリズムの解析方法

本章では可逆線形探索 [1] を元にアルゴリズムの解析方法の説明を行う。

### 4.1 線形探索アルゴリズム

線形探索アルゴリズムとは、配列やリスト等のデータの中からあるデータを探し出すための探索アルゴリズムの一つである。データの先頭から順に値を比較し、キーと呼ばれる探索したい値と一致するか、データの末尾まで比較を行うと探索を終了する。入力としては具体的な動きは以下のようなになる。

A1. 先頭の要素を見る // A2. キーと見ている要素が一致するか比較を行い、一致するなら成功として終了する // A3. 次の要素を見る // A4. 今見ている要素が存在するなら S2 へ戻り続行する、そうではなくこれ以上探索する要素がないなら失敗として終了する //

探索終了した時に探索した要素の場所を出力とする場合

と、探索が成功したか否かを出力する場合がある。本稿では後者の探索が成功したか否かを出力するものとして考える。上記が線形探索アルゴリズムの中で最も基本的なアルゴリズムと考えられ、基本的な方法とする、更に拡張したものとして番兵法と整列リストを用いた方法を挙げる。番兵法は A4 の動作を減らすために考えられた方法で、入力データとして用意されたデータの末尾にキーをあらかじめ付け加えるというものである。これによって、例えば最初のデータ内にキーがなかったとしても最後に必ずキーを見つけることができ、そのキーの位置が最後に判定することで探索成功か否かを判定する。具体的な動きは以下のようなになる

B1. データの末尾にキーを入力する // B2. 先頭の要素を見る // B3. キーと見ている要素が一致するか比較を行い、一致したら成功か判定を行い終了する // B4. 次の要素を見て、B3 へ戻る //

基本的な方法と手順の数は変わらないように見えるが、繰り返し行う手順が基本的な方法だと A2,A3,A4 の 3 つあるのに対して番兵法では B3,B4 の 2 つとなっており、最終的な手順の数では番兵法の方が少なくなっている。しかし番兵法は最初にキーを付け加えるため、データ量が多くなってしまうという欠点もある。次に整数リストを用いた方法はあらかじめデータの値がソートされているものを扱い、昇順ならば比較時に要素がキーの値以上だった場合に終了する。この方法では仮にデータの中にキーがなかったとしてもキーに近い値まで探索を行い探索を終了するため、実行するのにかかる時間が少なくなるというものである。また、データの末尾の次は無限と考えることによって終了条件の手順も繰り返さなくてよくなる。

### 4.2 既存の可逆線形探索

文献 [1] の可逆線形探索では前節で述べられた 3 種類の線形探索アルゴリズムを配列とスタックの 2 通りのデータ構造で可逆化と効率化が行われている。その中でも本稿では通常の線形探索の一般解法と提案解法、番兵法の一般解法の 3 つを取り扱う

### 4.3 既存の解析方法と問題点

文献 [1] での時間計算量の解析では、すべての行のステップ数を 2 種類の変数を用いて表していた。○プログラムコードの引用 //

変数 M は、キーと同じ値がどの位置に存在しているのかを表しており、データの中に存在しない場合はデータの大きさを表す。この M は他の指標と比べて大きな値になりやすく、大きな値になりやすいということはこのアルゴリズムの実行時間に対して与える影響が大きいということである。また、今回の探索アルゴリズムではキーの発見が通常の線形探索で失敗する場合以外での終了条件となっており、実行時間もキーの発見によって影響されやすく、ある程度入力データのキーの位置に法則性があるなら平均的な実行時間を導きだしやすくなっている。したがって、こ

の  $M$  という指標は妥当なものだと考えられる。変数  $S$  は、探索が成功したか否かを表しており、通常の場合は探索が成功したか否かで場合分けをして時間計算量を計算しなければならないのを一つにまとめて表すために用いられている。すべての行のステップ数を数えるとされているが、実装時にはプログラムになってしまうため、同じ動きを記述する方法がほんの少し変化するだけで実行時間に大きな影響を与えてしまう場合がある。ここで、アルゴリズムとプログラムの違いについて考えたい。アルゴリズムとは問題解決のための手順を示すもので、プログラムはコンピュータを働かせるために手順に従ってコンピュータへの命令を書いたものである。アルゴリズムの計算量が優れていればそれを元にしたプログラムすべてが優れていると言えるが、プログラムの計算量が優れていてもそのプログラムだけが優れているとは言いえない。よって、効率化を行うならばプログラムよりもアルゴリズムの方がいい。しかし、すべての行のステップ数を数えるという方法はプログラムの計算量での比較になってしまう。

#### 4.4 別の解析方法

全ての行のステップ数を数えるという指標をキーと比較する演算。探索している要素が最後の要素かを確認する演算の2つをそれぞれ数えるという指標にし各演算を  $key$ 、 $index$  と呼ぶ。これらは基本演算とし、アルゴリズムの中で重要な動きをする部分を表している具体的なプログラム上でどの部分を指すのかは図1に示す。

```

procedure linear1(int k[], int key, int n, int f, stack g)
  local int i = 0
  push(1, g)
  from top(g) = 1 loop
    if k[i] = key then
      push(f, g)
      f ^= 1
      push(1, g)
    else
      push(0, g)
    fi top(g) = 1
    i += 1
  push(0, g)
  until i >= n || f != 0
  push(i, g)
  delocal int i = 0
  
```

図1 基本演算の説明

以上の条件で解析を行った結果を表1に示す。

通常の線形探索の一般解法と提案解法では  $key$  と  $index$  のどちらの値も大体  $M$  改善出来たのに対して  $key$  と  $index$  以外の部分では  $(14M - 5M) - M - M = 7M$  となり、 $7M$  ほど改善できていたことがわかった。この  $key$  と  $index$  によって改善された合計  $2M$  という値は、アルゴリ

表1 基本演算別に見た時間計算量

基本演算	全ての行	key	index
通常的一般解法	$14M + 18S + 22$	$2M$	$2M - 2S + 2$
通常のプロポーザル解法	$5M + 6S + 6$	$M + S$	$M + S$
番兵法の一般解法	$8M + 4S + 32$	$2M - 2S + 2$	$2$

ズムにおいて重要な演算を行う部分であり、プログラムではなくアルゴリズム自体の効率が変化した部分であると考えられる。また、通常の線形探索の一般解法と番兵法の一般解法では  $key$  の値はあまり変化がないのに対して  $index$  の値が非常に小さくなっているという特徴が見ることができた。これは3.1節でも述べた番兵法の特徴であり、こういったアルゴリズムの特徴を確認することができた。基本演算を定めて解析を行うことはアルゴリズムの特徴を発見することができ、アルゴリズム自体の改善にも繋がるということが確認できたため必要なことだと考えられる。