

DLisp

Differential List Processor

Jin SANO

October 10, 2021

Contents

1 概要

2 背景

3 メモ

4 2021/10/10

1 概要

破壊的操作を試みる際に、事前にオブジェクト全体を複製するのではなく、操作の履歴を発行することで、過去のデータを復元できる手段を残しながら、最新のオブジェクトを手にいれるコストを最小化する手法を提案する。

この手法が顕著に活かせる例として、アトムと差分リストのみからなる Lisp 派生言語 DLisp を実装した。DLisp は append が $O(1)$ でできる（より正確にはアッカーマン関数の逆関数）。

2 背景

従来のプログラミング言語におけるデータの持ち方の考え方

C/C++ などの古典的な手続き型言語 破壊的更新が標準

- `const` が明示的に記述されていないと破壊的更新の可能性を否定できない
- 共有されているオブジェクトでもお構いなしに破壊的更新を行う
- どうしても「元のオブジェクト」が欲しくなる場合は、ユーザが明示的にその実装を行う必要がある
 - 再帰的にコピーを行う関数を実装するものと思われる

Rust などの比較的新しい手続き型言語 破壊的更新を行うが、所有権解析により共有されている可能性を排除することもできる

- そもそもオブジェクトの共有自体が推奨されない
- 実行効率は良いし安全性も高いが、制限が強すぎて、一部の低レイヤ好きにしか好まれない

Python/JavaScript などの比較的新しいスクリプト言語 関数型の思想を言語仕様・ライブラリなどでどんどん取り入れている

- 基本的に実行効率はお構いなし
- ガンガン複製する

関数型言語 immutable が標準

- append などを行う際は複製を行う必要がある
- そもそも append などのコストがかかる作業を避けるプログラムを意識的に書く傾向にある
- データ構造をめっちゃめっちゃ工夫して「償却時間は」 $O(1)$ のアルゴリズムを発明する傾向にある
 - ただし、あくまで「償却時間」であり、毎回それが保証されるわけではない
 - 「工夫」はあまり自明でない
 - 「工夫」のために無駄な中間データ構造を要求する場合が多い
 - * 定数倍で性能が悪化
 - * メモリの消費（メモリ消費は GC のタイミングを早めるため速度にも影響するはず）

論理型言語 未具現化変数

- append は一回しかできない
- 一度のみ具現化可能というのは Rust の所有権解析と少し似ている

LMNtal link による強い制約

- link は 2 頂点を繋ぐことしかできない
- hyperlink を用いてデータを共有することもできるが、共有物を純粋に保つのが面倒で（共有物を純粋に保たないプログラミングが人類に可能だとは思えない）、静的な検査もないため、めっちゃめっちゃバグりやすい
- 結局 ground を使うことになる
 - C などのように再帰を回してオブジェクトを clone するコードを自前で書かなくて良いというメリットはある
 - が、とてもコストが大きい
 - のにめっちゃめっちゃ頻繁に使われている

まとめると、既存の言語では、基本的には「後で元のオブジェクトも使うかもしれないとき」は事前にそのオブジェクト全体を複製する

- 関数型言語、スクリプト系、LMNtal (with ground), ...

しかし、それにはとてもコストがかかる

- オブジェクトの大きさに比例したコスト

しかも、後で「使うかもしれない」ということは使わない可能性もある

- つまり、既存のパラダイムでは
 1. 「事前に莫大なコストを払い」、
 2. 「後で古いものを使うときのコストはゼロ」としていた
- しかし、基本的には **新しいものを使う可能性が高い**と考えられるので、
 1. 「事前のコストは最小化」した上で、
 2. 後で「古いものを使いたくなったら（多少の）コストを払って戻す」ようにしたい

そこで、

1. 「後で元のオブジェクトも使うかもしれない」のに、破壊的更新を行うときは、「破壊的操作の履歴」をコミットすることにする
 - このコミットは単にアドレスとそこに代入した値のペアさえあればよく、「オブジェクト全体の複製」などよりも遥かに低コストである
2. 新しいものを使い続ける場合は上記の履歴のコミット以外のコストは払わない
 - 最新であるかどうかのチェックはビットが立っているかどうかなどで $O(1)$ で行う
3. 古いものに戻したい場合は、「(多少の) コストを払って復元する」
 - 基本的に新しいものを使い続けることの方が多くはずなので、古いものを欲しがる人にコストを払わせる
 - ただし、このコストはあくまで破壊的更新の回数 (append の回数など) であり、純粹にできる部分は純粹に保っていたらそこまで大きくはならないはず

3 メモ

差分リストはリストよりも強力なデータ構造

- append が $O(1)$ でできる
- 他の操作はリストと同様

ただし、差分リストの append

4 2021/10/10

提案手法は、単に共有物に対して破壊的操作を行っている場合は履歴を保持するというだけ。

- つまり、これはグラフに限らず、例えば配列などに対しても適用可能ではある。
- ただし、配列はめちゃめちゃ破壊的操作を行うため、履歴が大量発生する & 戻すのに操作の数だけ逆操作するため、あんまり嬉しくはない。
- (単方向の) 差分リスト (もどき) が嬉しいのは、「末尾の破壊的更新」しかできないということであった。

- 末尾の破壊的更新以外は純粹にできるため、それらの履歴の保持が不要であり、「履歴のコストが比較的小さい」というメリットがあった。
- これは先週の段階ではぼんやりとしか理解していなかった (ので説明ができなかった)
- こう言った性質をグラフ (の shapetype のような型) において自動的に導出できるのかは不明。
- 双方向リストにしてしまうと、Head に cons するだけでも破壊的更新をする必要があり、この履歴も管理せざるを得なくなるため、履歴のコストが無視できなくなる (かも)
- 現実装は、すごく安直で、全ての操作を逆実行する (Nil (未具現化変数) の更新 (具現化) をした部分もわざわざ戻す) が、こういった部分に関しても差分リストの場合は最適化が可能
 - * 現実装は一般のグラフへの適用を考えた (差分リストに最適化されていない) 素朴な手法
- 差分リストの場合は move 可能な部分は履歴を管理する必要がないというのが僕の直感的な理解 (あまりきちんと説明できないのでちゃんと例題を書く必要がある)
 - * ただし、一般のデータ構造に対してはこれは保証できないことに気づいた (配列の破壊的代入など)

まとめると、

- 提案手法が差分リスト (もどき) において有効なのはほぼ確信している。
 - これをあまり理解してもらえなかったのは純粹に説明が悪かったのだと思う。
- 提案手法がより一般のグラフにおいて最適化可能なかはよくわからない。
 - 操作の数だけ復元にコストがかかる可能性がある。
- が、仮に最適化できなかったとしても、「純粹 (風に) にグラフ (破壊的データ構造) を扱う」という「今まで人類ができなかったこと (調べ学習が足りていない感はある)」を実現しているのでこれは価値があると思っている。
 - つまり、「今までできていたことをより良くする研究」ではないということ。
- 提案手法が一般にはコストゼロで途中の move と組み合わせられない (move できる部分は履歴の保持がいらないというのには一般には保証できない) のは痛手であったが、

所有権解析を取り入れている言語は「徹頭徹尾」move させるようにしているので、途中の move があまり最適でないというのは仕方ないことだと言える。

- ただし、差分リストの場合はこれがおそらく可能で、どういうパターンのときにそうなるのかはもっと考える必要がある (考える価値があると思う)

(木だけではなく) グラフ (特に差分リスト) を扱うメリット:

- グラフを扱えるとより時間効率の良い実装ができる場合がある。
 - キューなど

- 末尾再帰化してループへ変換できる関数がある.
- グラフを扱えるにより空間効率の良い実装ができる場合がある.
 - 従来の `append` は第一引数のリストを複製するため、その分メモリを消費する.
 - 末尾再帰化可能でない関数はスタックを消費する.
- グラフを扱えるにより直感的に記述できる可能性がある.
 - キュー. 2本スタックを用意するのはすごく直感的というわけではない.

そもそも所有権解析により、常に `move` させるようにするのはダメなのか？

- 常に `move` させるのでは困るという明確な例題は正直あまり思い付いていない.
- だが、例えば `python` ユーザに所有権解析を押し付けるのはどうかと思う.
 - という非形式的な感想しかないと言われればそこまで.
- `alim` さんの `ground` を用いた `hypergraph` による `lambda` も `ground` を用いていて、しかもラムダ式は木にちょっと毛が生えたくらいだから提案手法で（効率的に）扱えないと困る.
 - この場合、`ground` のように丸々コピーするよりも遥かに安価である（あって欲しい）
 - しかし、そもそも `lambda` のエンコードの価値が実はあまりよくわからない（失礼！）から、これが最適化されることの意味もよくわからない.