葉数最適整列法LOASの実現法

4Ｊ－９

二村 良彦　　平井 利治
早稲田大学 理工学部

1．はじめに

葉数最適整列法LOAS(Leaves-Optimal Adaptive Sort)[4,5]は，基本的にはマージに基づいた整列法である[1]．それは与えられた数列をまず葉数個の部分列に分割し，次に分割された部分列を2つづつマージする(ただし数列の葉とは，数列において隣人よりも値の小さい要素である[4,5])．従ってLOASを実現するに際して，適正なマージ戦略を選択することが重要である．本稿に於いては，我々がLOASを実現する際に検討した下記3種類のマージ戦略とその性能について報告する：

(1) 単純マージ．

(2) 2分マージ(binary merge)[8]．

(3) 基本的には単純マージであるが，マージする2つの数列の特徴に応じて2分法を利用したもの．例えば，一方の数列の長さが1のときのみその挿入個所を探索する為に2分法を利用したもの．

葉数を制御したランダムな順列[2]を用いて各種方式のキー比較回数および実行時間の比較評価をした．その結果，単純マージを利用した最も素朴な方法を陵駕する方式は，現在までには見つかっていない．LOASとその他の整列法との性能比較については[3,5]で報告した．以下では整列すべき数列は配列z(1),...,z(n)に格納されており，これを昇順に整列する場合について考える．

2．LOASにおけるマージの基本戦略

LOASは例えば数列X=〈7, 1, 2, 6, 5, 3, 4〉を次のような2つのフェーズにより整列する：

フェーズ1：まずXを左からスキャンし下降列と上昇列のペア列(〈7, 1〉, 〈2, 6〉), (〈5, 3〉, 〈4〉)に分ける．次に各ペアを(単純)マージし葉数個(この場合は2個)の上昇列を作る：(〈1, 2, 6, 7〉), (〈3, 4, 5〉)．

フェーズ2：上で得られた葉数個の上昇列に対してマージを繰り返し，最終的に１つの上昇列を得る．

このアルゴリズムがO(数列の長さ*log葉数)であることは明らかである．また高々葉数個の部分数列の管理に[(n+1)/2]，そしてマージのために[n/2]，合計nのスペースが必要である．

マージ戦略の詳細を説明する為にまず2つの数列の重複と適応マージ[1,7]について述べる．以下ではzのb番地から始まりe番地で終わる上昇列を(b, e)で表す．

数列の重複：隣接する2つの上昇列(b1, e1)および(b2, e2)の重複とは次の2つの上昇部分列(ob, e1)および(b2, oe)である(b2=e1+1に注意)．ただしz(ob-1)≦z(b2)〈z(ob)かつz(oe)〈z(e1)≦z(oe+1)．例えば〈1, 2, 3, 4, 6, 7, 9〉と〈5, 8, 10〉の重複は〈6, 7, 9〉と〈5, 8〉である．

適応マージ：2つの上昇列をマージする際に重複以外の要素を移動させないマージ法を適応マージと呼ぶ．配列を用いるマージ法ではデータのコピーが大きなオーバーヘッドである．適応マージはそれを減らす為の方法である．

LOASのマージは常に重複を持つ数列に対して行われる([4]の性質1および2)ので重複の有無の判定が不要である．これにより整列に必要なマージの回数(約葉数回)分のキー比較が節約できる．そして次のようにすれば効率良く適応マージをすることができる．

LOASにおける適応マージ：マージすべき2つの上昇列を(b1, e1)および(b2, e2)とする．ここでまず2者のうちで短い方，例えば(b2, e2)，の中の重複の端点oeを見つける．この時(oe+1, e2)は移動不要であることが分かる．従って(b2, oe)のみをコピー領域

Merging in Leaves-Optimal-Adaptive-Sort(LOAS)
Yoshihiko FUTAMURA and Toshiharu HIRAI
School of Science and Engineering, Waseda University.

に移動する．またz(e1)をoe番地に移動する．次に(b1, e1-1)と(b2, oe)を後尾からb1番地からoe-1番地にマージする．この方法により(b1, ob-1)も移動されないことに注意されたい．(b1, e1)の方が短い場合はマージを数列の先頭から行う．短い方を移動させるので，コピー領域はn/2あれば十分である．

なお，複数の数列をマージする際に数列が奇数個ある場合には，1番目または3番目の数列で長い方をそのまま残しておき，残りの偶数個の数列をマージする．これは一番最後の数列が短いままで残るのを避けるための方策である．

3．マージ戦略の比較

上述の基本マージ戦略に対して下記の2点を変更する方式を各種開発し，基本戦略と性能比較した．

(1)基本戦略では数列の重複の一端，例えばoe，を順次探索で求めるのに対し，i)2分探索法で求める方式．およびii)マージすべき数列の一方が長さ1のとき，他方への挿入個所を2分探索法で求める方式．

(2)基本戦略では適応マージを行うのに対し，i)2分マージ[8]を行う方式(基本戦略との比較を図1中のLABMにより示した．図中ではLOASが基本方式の性能を表す．参考のために普通のマージソート[8]の評価データも図に含めた)．およびii)適応マージを行うが，重複の両端obとoeを見つけてから，(ob, e1-1)と(b2+1, oe)を2分マージする方式(基本戦略では重複の一端のみ見つければよい)．

これらの方式は2分法の威力を期待して開発されたものであるが，我々の実験結果では(1)，(2)共に基本戦略の改良にはならなかった．特に(2)では大幅に性能が低下した(図1のLABM参照)．実験の際に使用されたテストデータは文献[2]の方式2によって生成した．今後より精密なテストをし，その詳細について報告する予定である．

4．おわりに

LOASの基本的マージ戦略およびそれがより複雑な各種戦略より劣らないことを示した．一様な乱数列の葉数は平均約n/3[4]であるので，葉数n/3の近辺でも他の整列法より速くすることを目標に，LOASの新たなマージ戦略および対称整列法[6]を検討中である．

5．参考文献

[1]ESTIVILL-CASTRO AND WOOD:A survey of adaptive sorting algorithms, ACM Computing Surveys, Vol.24, No.4, December, 1992, 441-476.

[2]二村，青木，大谷，遠藤，白井:整列法評価のためのランダム順列の生成，情報処理学会50回大会6F-5, 95年3月．

[3]二村，遠藤，平井，青木:適応整列法の評価，情報処理学会50回大会4J-10, 95年3月．

[4]二村，二村:葉数適応整列法LOASの最適性，情報処理学会50回大会4J-8, 95年3月．

[5]二村，二村，遠藤，平井:LOAS:葉数に関して最適な適応整列法，日本ソフトウェア科学会，94年11月．

[6]二村，二村，筧:対称整列法，情報処理学会アルゴリズム研究会28-8, 92年7月．

[7]浜村，若林，宮尾，吉田:入力列のプリソーテッドネスを考慮した最適並列ソーティングアルゴリズム，COMP-90-37.

[8]KNUTH:The Art of Computer Programming. Vol.3., Addison-Wesley, Reading, Mass., 1973.

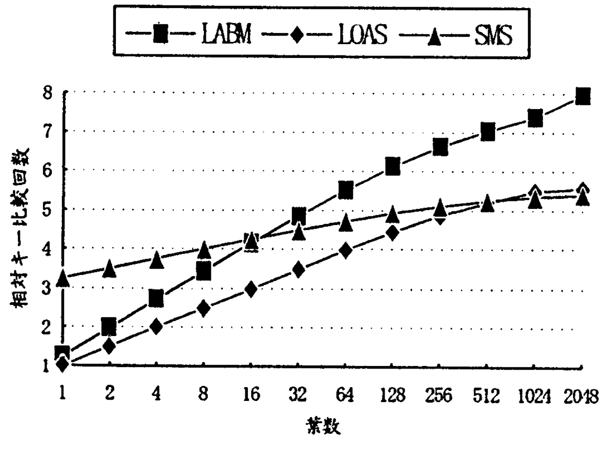

図1:長さ4095の数列の整列に要するキー比較回数

LOASは基本戦略，LABMはマージに2分マージを利用したもの，そしてSMSは普通のマージソートを表す．葉数1に対するLOASの比較回数(約8190)により，全ての観測データを正規化して示した．