プロダクション規則と局所評価関数にもとづく

計算モデル CCM による各種のソート法

新情報処理開発機構 (RWCP)

つくば研究センタ

金田　泰

目次

- まえおき
- 挿入ソート
- 交換ソート
- 連結ソート
- その他のソートについて
- おわりに (CCM の研究・発表状況)

はじめに

- 自己組織的な計算をめざして計算モデル CCM をかんがえた．
  - CCM = 化学的キャスティング・モデル (Chemical Casting Model)．
  - 自己組織的 = 全体的な計画なしに = 創発的に．
  - CCM ではランダムさをふくむ (非決定論的な) 計算をめざす．
- CCM によって，本来は複雑な開放系の問題をやりたい．
  - しかし，いまのところ目標がきちんとさだまっていない．
    → 古典的な制約充足問題や最適化問題を中心に実験している．
- CCM によって従来的な問題もとけるべきである．
  - 従来のパラダイム / プログラミング言語にかわるものをめざしているから．
- ゆえに，今回は CCM にもとづくソート法について論じる．
- ソート法以外の最近の研究状況についてものべる．

全体計画なしの計算

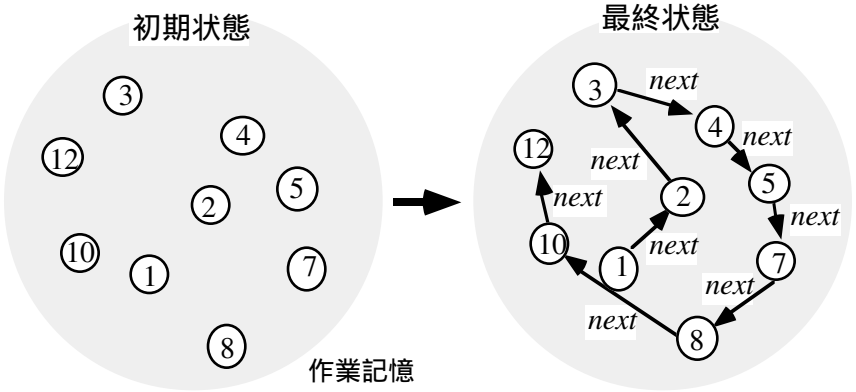
バブル・ソートのばあい

- バブル・ソートを局所的にみると，
  - 隣接要素の交換をおこなう — 局所操作．
  - 隣接要素 a_i, a_i+1 間に $a_i \geq a_i+1$ なる関係がなりたつべき — 局所評価．
- CCM のかんがえかた：局所操作と局所評価だけをあたえてバブル・ソートを実現する．
  - 局所操作を反応規則 (プロダクション規則) として表現．
  - 局所評価を「局所秩序度」(一種の評価関数) として表現．
  - 局所的な情報だけで計算する — 創発的計算．
- 従来のプログラミング言語における手順や論理にかわるもの
  - それはある種のランダムさ (非決定論的なもの，or 確率的なもの)．
  - 局所操作や評価による決定論的なものと非決定論的なものとのからみあいから，それが創発する (?!)
    - 熱力学における散逸構造のようなものをかんがえている．

挿入ソートのためのデータ構造と初期状態・終状態

挿入ソート — 1

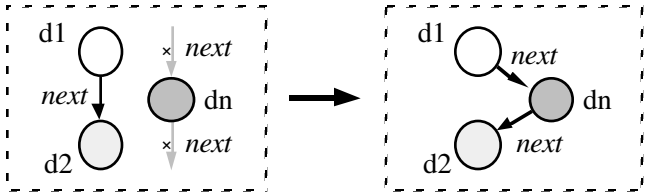
- 個々のデータを原子 (CCM の primitive data) として表現．
- データを昇順または降順にリンクでつなぐ．



反応規則と局所秩序度

挿入ソート — 2

- 反応規則



  - dn へのリンクをもつ原子はない — データを単鎖にするため．
- データ (data 型) に関する局所秩序度 (降順ソートのばあい)

$o_{\text{data}}(d) =$ 0 **when** $d.next =$ **nil**
0 **when** $d.value < d.next.value$
1 **when** $d.value \geq d.next.value$

  - 反応 (反応規則の適用) にかかわるデータに関する局所秩序度の和が増加するときだけ，実際に反応する．

記号処理研究会での前回の発表との関係

- 前回 (3 月) には制約充足問題の解法について発表した．
- ソートも制約充足問題の一種なので，この解法をあてはめられる．
  - この解法では，局所秩序度をつぎのように定義する．
    - 局所的な制約がみたされるとき 1，そうでないとき 0．
  - ソートの局所秩序度は実際このように定義されている．

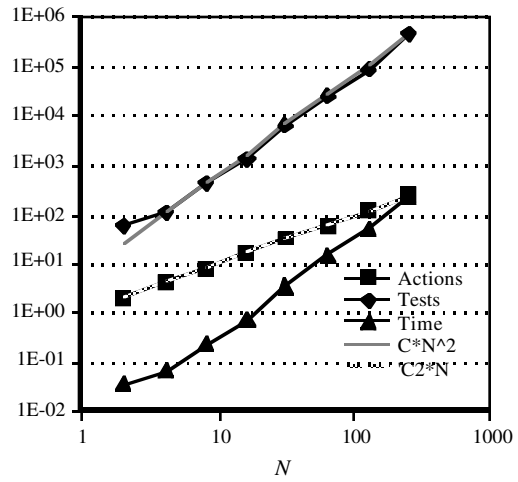
実装の詳細

挿入ソート — 3

- SOOC という言語で反応規則と局所秩序度とを記述する．
  - SOOC = Self-Organization-Oriented Computing.
- ダミー・データを導入する．
  - 前記の規則を適用するには，初期状態においてもリンクされたデータがすくなくとも 1 組は必要．
  - “最大の数” と “最小の数” をあらわす 2 個のダミー・データをリンクしたものを作業記憶にいれる．
  - この「ひも」にデータを挿入していくことによりソートがすすむ．
- データがリンクによってさされているかどうかを知るために，
  - 双方向リンクを使用する，または
  - 単方向リンクを使用するばあいには参照フラグ (リンクによってさされていることをしめすフラグ) を併用する．
    - SOOC の単方向リンクは逆向きにはたどれない．

平均実行時間などの実測データ

挿入ソート — 4

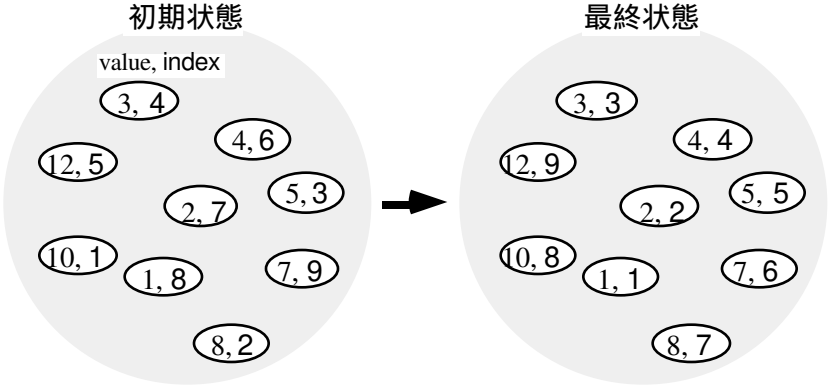


- 測定条件
  - ランダムデータを使用．
  - 反応規則とデータの選択は疑似乱数による．
    - つまり「ランダム戦略」を使用．
    - 「ふかさ優先戦略」でもおなじ傾向．
- 結果
  - マッチング回数 (Tests) と計算時間の平均は $O(N^2)$．
  - 平均反応回数は $O(N)$．

交換ソートのためのデータ構造と初期状態・終状態

交換ソート — 1

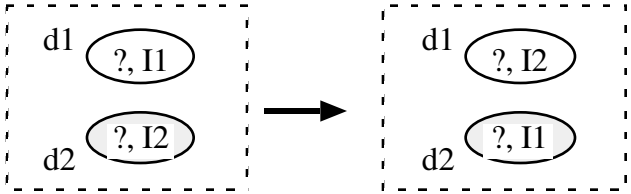
- インデクスをつかってデータの順序を表現する
  — リンクをつかうと繁雑になるため．



反応規則と局所秩序度

交換ソート — 2

- 反応規則



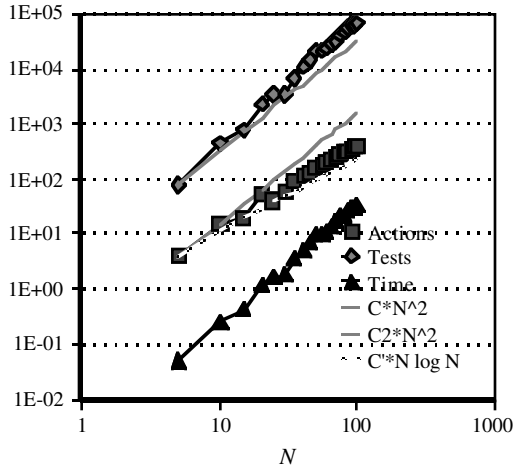
- データに関する局所秩序度 (相互秩序度)

$o_{data}(d1, d2) =$ 1 **when** $d1.value = d2.value$ **or** ($d1.value < d2.value$ **and** $d1.index < d2.index$) **or** ($d1.value > d2.value$ **and** $d1.index > d2.index$)

0 **otherwise**

平均実行時間などの実測データ

交換ソート — 3

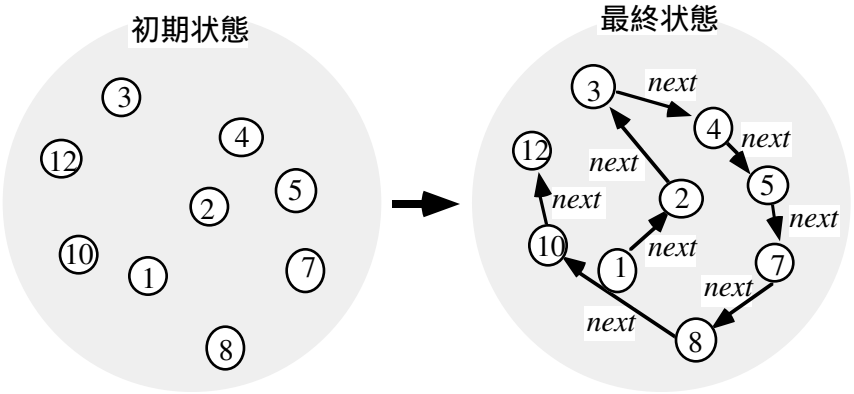


- 測定条件は挿入ソートのときとおなじ．
- 結果
  - マッチング回数と計算時間は $O(N^a)$ $(2 < a < 3)$．
    - 手続き型言語よりややおおきい．
  - 反応回数はほぼ $O(N \log N)$．

連結ソートのためのデータ構造と初期状態・終状態

連結ソート — 1

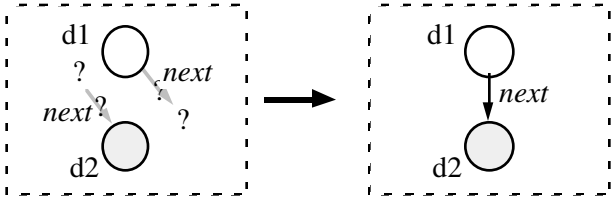
- 連結ソートは CCM 特有のソート法．
- 使用するデータ構造は挿入ソートとおなじ (リンクを使用)．



反応規則と局所秩序度

連結ソート — 2

- 反応規則



- 局所秩序度

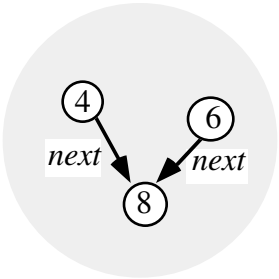
$o_{\text{data}}(d) = 0$ **when** $d.next = \textbf{nil}$

$1 / (d.value - d.next.value)$ **when** $d.next \neq \textbf{nil}$

  - 作業記憶に同一の値をもつデータが存在しないことを仮定する．
  - この局所秩序度は前回の制約充足問題の解法にはしたがっていない．
    - 0, 1 だけでなく，任意の正の実数値をとりうる．

挿入ソートとおなじ局所秩序度をつかわない理由

- 連結ソートの反応規則を挿入ソートの局所秩序度とくみあわせると，ただしくソートされない．
- ひとつの原子が複数の原子からリンクでささされた状態においても局所秩序度の値はひとしいため反応は停止する．

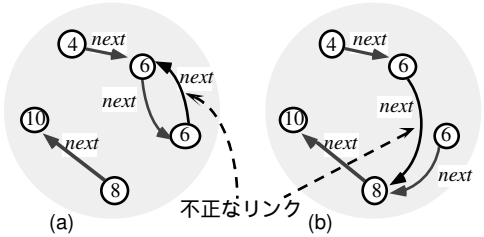


連結ソートの特徴

- あらかじめみたされるべき制約がゆるい (?)
  - 初期状態でつぎのようなデータが存在しても，ただしくソートされる．
    - 逆順のリンク．
    - 複数のデータからリンクされたがデータ (単鎖でないデータ)．
  - ソートの途中または完了後に上記のようなデータが追加されても同様．
  - 挿入ソートでは，上記のようなばあいはただしくソートされない．
    - 「単鎖である」という制約を保持するために特別のくふうが必要．
    - 事前にこの条件がみたされていないと，みたすことはできない．
- 動的に秩序度関数を昇順ソート用から降順ソート用，あるいはその逆にとりかえても，ただしく動作する．
- これらの特徴は探索グラフの強連結性なる性質からくる．
  - 探索グラフの強連結性：秩序度に関する条件を無視すれば，適当なデータに関して反応規則の適用をくりかえすことにより，任意の状態から任意の状態にいけるという性質．

連結ソートの致命的な問題点

- 同一値が複数個存在するとただしくソートされない．
- その理由
  - 局所秩序度定義で同一値が存在しないと仮定 — 主要な理由ではない．
  - 同一値をもつ原子をむすぶ輪や分岐構造が生成されるのが主要な理由．



  - この問題は巡回セールスマン問題において輪が分裂する問題と同根．
- 解決策はまだない．
  - この問題を解決するには，いまのところ非局所的なデータ参照が必要．
  - 局所情報だけによる計算という原則をやぶらずに解決する方法は不明．

実測のための条件

連結ソートの実測 — 1

- ランダムデータ，昇順データ，降順データのそれぞれについて実験．
  - 昇順データとは，反応時にデータが昇順に選択されるように，データを生成したもの．
  - 降順データも同様．
- 反応規則とデータの選択法
  - 「ふかさ優先戦略」のばあいをしめす．
- 初期状態においてリンクは存在しない．

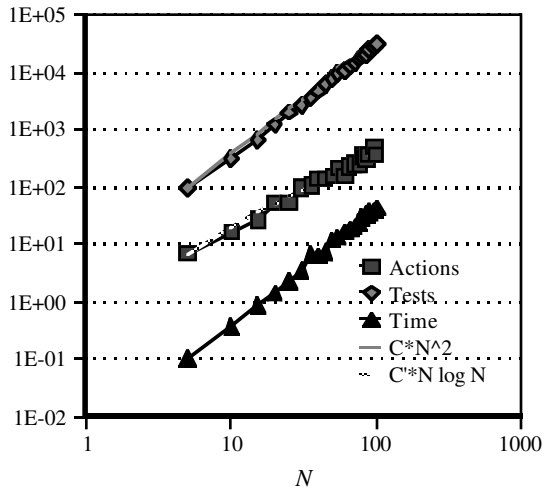
実測結果の要約

連結ソートの実測 — 2

- ランダムデータのばあい
  - マッチング回数と計算時間の平均は $O(N^2)$．
  - 平均反応回数はほぼ $O(N \log N)$．
- 昇順データのばあい
  - マッチング回数と計算時間はランダム・データとほぼひとしい．
  - 反応回数はほぼ $O(N)$．
- 降順データのばあい
  - 反応回数はほぼ $O(N^2)$．

ランダムデータの連結ソートの実測

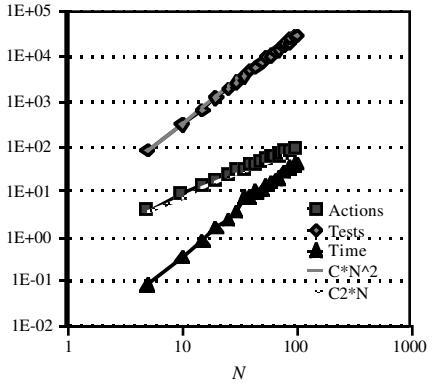
連結ソートの実測 — 3



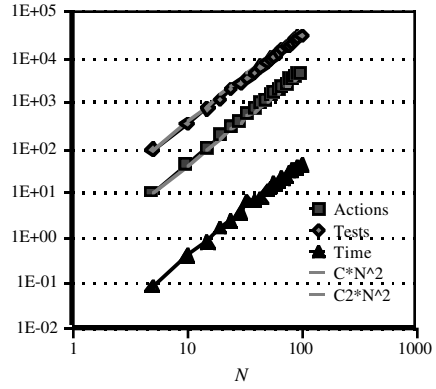
ソートずみデータの連結ソートの実測

連結ソートの実測 — 4

昇順データのソート



降順データのソート



他のソートの実現可能性

- 選択ソートは局所的な情報だけにもとづく計算という原則にあわない．
  - 最大値 / 最小値の選択においては大域的なデータ参照が必要だから．
  - 不在条件の記述をゆるせば選択ソートを記述できる．
    - 不在条件とは，作業記憶中にある条件をみたす原子がないときになりたつ条件．
    - 不在条件の使用は上記の原則に反している．
- 計算時間が $O(N \log N)$ のソートができるかどうか不明．
  - たとえば併合ソートのプロセスをまねても現在の SOOC 処理系のもとでは $O(N^2)$ の時間がかかる．
- 計算時間が $O(N \log^2 N)$ のソートは実現可能性がよりたかい (?)
  - たとえば bitonic sort．
- 基数ソート (radix sorting) は比較的実現可能性がたかい (?)

CCM の研究状況

おわりに — 1

- これまで，つぎのような問題解決をこころみてきた．
  - 制約充足問題： N クウィーン問題，彩色問題．
  - 最適化問題： 巡回セールスマン問題 (TSP)，線形計画問題．
- 最近，ノイズのはいったセル・オートマタの実験をしている．
  - 人工生命的な (実用的ではない !) 方向への展開．
  - 計算過程を見たい．
  - いまのところ，かならずしも CCM の利点をいかした計算にはなっていない (局所秩序度をつかっていない)．

CCM の研究発表先

おわりに — 2

- 92.1 プログラミング・シンポジウム
  (コンピュータによる自己組織化をめざして，例題は N クウィーン)．
- 92.7 夏のプログラミング・シンポジウム
  (自己組織系としてのソフトウェア，例題は彩色)．
- 93.2 SICE システム工学分科会研究会「最適化とその周辺」
  (CCM による最適化，例題は TSP)．
- 93.3 IPSJ 記号処理研究会 (制約充足問題の解法，例題はまた N クウィーン)．
- 93.3 IPSJ プログラミング研究会 / 信学会コンピュテーション研究会
  (確率過程としての計算，例題はまた彩色)．
- 93.8 SWoPP-AI (問題解決 / 解探索法としての CCM・触媒，例題はまた彩色)．
- 93.10 IPSJ 全国大会 (線形計画問題，反応規則の合成)．
- 93.11 IPSJ 記号処理研究会 (ソート)．
- 94.1 Hawaii International Conference on System Sciences
  (自己組織的計算パラダイム，触媒と反応規則の合成)．

挿入ソートの補足*

- 漸進的なソートができる．
  - いったんソートが終了したあとでも，データを追加すればソート列にとりこまれる．
- 局所秩序度を動的に変化させることも可能．