

談話室

分類における釣合併合法の利点

西村 恕彦*

磁気テープを使った分類（ソート）の手法は、釣合併合法（ p -way merge sort）か、あるいは不釣合併合法（多相分類法など）によるのがふつうである。磁気テープの台数が3台のときには、後者によるしかない。

磁気テープの台数が4台以上のときには、どちらの手法でも採用できるが、台数 n が一定とすれば、ほとんどつねに多相分類法のほうが、釣合併合法よりも高速であるとされる。それは、多相分類法では $n-1$ 次の併合ができるのに、釣合併合法では $n/2$ 次の併合しかできないからである。

実際に、メーカーによっては、4台以上の磁気テープを備えている客先にたいしても、まず多相分類法によるサービスルーチンを提供している例がある。

ところが、がんとして釣合併合法によるルーチンしか作成しないメーカーもあるらしい。

実は釣合併合法には、低速かもしれないという欠点をカバーする非常に大きな実用上の利点があるのである。分類について論じたり解説したりした文章はかなり多いが、この利点に触れているものを、筆者は見たことがない。自明のことではないし、だれかがすでに述べているようなものであるが、ちょっとさしがあてられなかった。どなたか既出文献をご承知の方は、ご教示いただきたい。

実用上の利点というのは、釣合併合法では、入力ファイルにあった旧配列を保存するようにルーチンを作成できるということである。多相分類法では、これは絶対に不可能なのである。前者の利点は、おそらく一部のメーカーは知っているであろう。ユーザもまた、経験的に知っている可能性がある。

例をあげよう。日付けの順に配列されている取引伝票のファイルがあるとす。これを別のキー、たとえば取引先番号によって再分類する。多相分類法では、

取引先別に分類されたファイルを目的どおり得たときに、1つの取引先のグループのなかでの伝票の配列順は、でたためになっている。

ところが釣合分類法では、1つの取引先のグループのなかでは、伝票は旧配列つまり日付け順にならんでいるようにできる。これが、再分類において旧配列を保存できるという意味である。（もちろん、釣合分類法においても、旧配列を破壊するようにルーチンを作ることとはまったく可能なのである。その場合に、釣合分類法を採用する利点はそれほどないように思われる。）

分類作業においては、分類のキーを多重にしなければならぬことが多い。たとえば、取引伝票を年度別に分け、1つの年度のなかでは営業所別に分け、そのなかでは取引先別に分け、そのなかでは日付けの順にならべるといような場合に、分類のキーは大から小へ、年度—営業所—取引先—日付けとなる。

多相分類法では、これらのキーをまとめて指定し、一挙に作業を行なわなければならない。したがって、多相分類法によるルーチンの外部仕様は、キーの個数や、けた数の制限をゆるやかなものにしておかなければならない。

ところが釣合分類法では、小キーによってまず分類し、できあがったファイルをこんどは中キーによって再分類し、というふうの部分的に分類作業をすすめてゆくことも可能である。したがって、ルーチンの外部仕様は多相分類法の場合にくらべて、キーの個数や、けた数の制限がかなり窮屈であっても、使用に耐える。

のみならず、きわめて大きなファイルの場合には、このことは、作業上あるいは時間上の利益をもたらすように思われる。

分類機能においては、キーの指定にしたがった配列が実現されるのは当然である。指定された以外の観点からみて、レコードがどのように配列されているかは規定されないことが多い。メーカーの作る外部仕様書では、このことは明確に述べるほうがよい。

* 電子技術総合研究所

たとえば釣合併合法のルーチンのユーザが、旧配列が保存されることを、経験的に知って利用していた場合に、多相分類法のルーチンに切り替えたとしたら、作業手順上かなりの混乱をまぬがれないだろう。

旧配列保存型分類法のルーチンを提供するメーカーは

その仕様書に、旧配列が保存されることを明示し、他方、旧配列破壊型分類法のルーチンの仕様書では、旧配列が保存されないことを明示しておけば、そのような混乱は相当防げるだろう。

(昭和45年9月16日受付)