■コンピュータサイエンス教科書シリーズ **2**

データ構造とアルゴリズム

博士（工学）　伊藤 大雄 著

COMPUTER SCIENCE TEXTBOOK SERIES

コ ロ ナ 社

刊行のことば

インターネットやコンピュータなしでは一日も過ごせないサイバースペースの時代に突入している。また，日本の近隣諸国もIT関連で急速に発展しつつあり，これらの人たちと手を携えて，つぎの時代を積極的に切り開く，本質を深く理解した人材を育てる必要に迫られている。一方では，少子化時代を迎え，大学などに入学する学生の気質も大きく変わりつつある。

以上の状況にかんがみ，わかりやすくて体系化された，また質の高いIT時代にふさわしい情報関連学科の教科書と，情報の専門家から見た文系や理工系学生を対象とした情報リテラシーの教科書を作ることを試みた。

本シリーズはつぎのような編集方針によって作られている。

（1） 情報処理学会「コンピュータサイエンス教育カリキュラム」の報告，ACM Computing Curricula Recommendationsを基本として，ネットワーク系の内容を充実し，現代にふさわしい内容にする。

（2） 大学理工系学部情報系の2年から3年の学生を中心にして，高専などの情報と名の付くすべての専門学科はもちろんのこと，工学系学科に学ぶ学生が理解できるような内容にする。

（3） コンピュータサイエンスの教科書シリーズであることを意識して，全体のハーモニーを大切にするとともに，単独の教科書としても使える内容とする。

（4） 本シリーズでコンピュータサイエンスの教育を完遂できるようにする。ただし，巻数の制限から，プログラミング，データベース，ソフトウェア工学，画像情報処理，パターン認識，コンピュータグラフィックス，自然言語処理，論理設計，集積回路などの教科書を用意していない。これらはすでに出版されている他の著書を利用していただきたい。

（5）　本シリーズのうち「情報リテラシー」はその役割にかんがみ，情報系だけではなく文系，理工系など多様な専門の学生に，正しいコンピュータの知識を持ったうえでワープロなどのアプリケーションを使いこなし，なおかつ，プログラミングをしながらアプリケーションを使いこなせる学生を養成するための教科書として構成する。

本シリーズの執筆方針は以下のようである。

（1）　最近の学生の気質をかんがみ，わかりやすく，丁寧に，体系的に表現する。ただし，内容のレベルを下げることはしない。

（2）　基本原理を中心に体系的に記述し，現実社会との関連を明らかにすることにも配慮する。

（3）　枝葉末節にとらわれてわかりにくくならないように考慮する。

（4）　例題とその解答を章内に入れることによって理解を助ける。

（5）　章末に演習問題を付けることによって理解を助ける。

本シリーズが，未来の情報社会を切り開いていけるたくましい学生を育てる一助となることができれば幸いです。

2006年5月

編集委員長　曽和　将容

まえがき

本書はデータ構造とアルゴリズムの重要部分について基礎から最新までを記した本である。対象は主に大学の学部 2〜3 年の講義を想定しているが，大学院でも使用できる進んだ内容も含んでいる。

「データ構造とアルゴリズム」と題した書籍は数多く出版されており，それらの中には良書，好著が数多く含まれる[†1]。それらがあるにもかかわらず私が本書を記したのにはもちろん意味がある。「データ構造とアルゴリズム」の世界は日進月歩であり，つぎからつぎへと新しく重要な技術が生まれてくる。そしてそのうちのいくつかは基礎的な必須のアルゴリズムとして残っていく。そういったアルゴリズムの中で非常に重要であるにもかかわらず，なまじ先進的であるために基礎的記述が中心となる教科書にまったく出現しないものが少なからず存在する。それらを知るためには現状では原著論文か，せいぜい英語の専門書に当たるしかない。しかしわれわれ専門家は別として，学生や一般企業のエンジニアにとって，そういった文献に当たれといわれても難があるだろう[†2]。

本書はそれらを解決することを主眼として書かれている。したがって，類書にはない，最先端のアルゴリズムを解説することに多くのページを割いている。最先端のアルゴリズムはもちろん比較的難解である。しかし前述のように本書の主対象者である学部生でも理解できるよう，極力平易な解説を試みた。それらを完全に理解するには相当な努力が必要とされるかもしれないが，主たる考え方を理解することはさほど難しくないと考えている。

そして本書の最も重要な特長は，定数時間アルゴリズムの紹介に多くのペー

---

†1 参考文献リストに挙がっている教科書はどれもそうだが，数冊選ぶのなら，例えば文献 39), 43), 44), 50) などであろうか。

†2 実際，原著論文にはしばしば間違いが含まれており，その修復は専門家でないと無理であろう。

ジを割いていることである。**定数時間アルゴリズム**とは，入力のうちのほんの一部，定数個のデータだけを見て判定，計算するアルゴリズムの枠組みである。20 年ほど前に現れ，今世紀になって急激に発展した。当初はほぼ理論的興味だけで研究が行われてきたが，最近ビッグデータとの絡みで実用面での研究も行われている[†1]。いまや国際会議のメイントピックの一つとなっているが，それを解説する専門書は非常に少ない。まして日本語で解説したものはほとんどなく，あるのは筆者やその共同研究者の書いた短めの解説[47),48),65)] ぐらいで，これだけきちんと解説した日本語の教科書は初めてといってよかろう。定数時間アルゴリズムは 21 世紀のアルゴリズムの標準技術となることは間違いなく，これからアルゴリズムを学ぶ者はこの機会に是非とも学んでおいてほしい。

本書で使用する変数や関数などの表記法（notation）について，その主なものを “記号表” という形で目次の後にまとめた。また，本書で使用する数学に関する基本的な用語および基本公式については，8 章に簡単にまとめたので，適宜必要に応じて参照してほしい。さらに，各章末に演習問題を，また必要に応じてプログラム演習を掲載したので，本書の理解の助けになるだろう[†2]。

本書の使い方についての一例を記しておく。学部の基礎講義として用いる際には，前半で 1 章から 3 章「整列」までをまずしっかりと教え，その後理解度に応じて 4 章「集合に関する操作」と 5 章「平衡二分探索木」に進む。そして余裕があれば 6 章「古典的アルゴリズム」のうちの可能な項目を教えればよい。その際には，付録に回した部分はやるとしても紹介程度で，解説している時間はないであろう。もし通年で教える場合には，演習問題などをじっくりとやら

†1　例えば「JST CREST『ビッグデータ時代に向けた革新的アルゴリズム基盤』研究代表者：加藤直樹 関西学院大学教授」など。

†2　内容的にやや高度と思われる本書の一部（アステリスク（$*$）のある項タイトルの内容）を付録として Web ページに回した。また，本書各章末の演習問題の解答も併せてそちらに掲載しているので（プログラム演習の解答は掲載していない），コロナ社 Web ページの本書の紹介ページ

`http://www.coronasha.co.jp/np/isbn/9784339027020/`

を必要に応じて参照してほしい。なお，PDF データの入った Zip ファイルを展開する際のパスワードは以下のとおりである。

せ，付録の部分も解説することができると思う。その上で 6 章のすべてを教えられると思うが，さらに 7 章「定数時間アルゴリズム」もざっと教えられれば理想である。ただし，正則性補題はそれを理解するだけでも（一部の優秀な学生を除き）学部生には難しいので，証明は省いてアルゴリズムのみ説明することで十分である。

大学院の講義に用いる場合には，前半で 1 章〜6 章のトピックから選んで講義し，後半は 7 章を中心とするのがよい。付録に回した正則性補題の証明まできちんと理解させようとすると，実は 7 章全部で半期の講義にできるぐらいの内容はある。

本書を執筆するにあたって多くの方に助けていただいた。電気通信大学名誉教授で先進アルゴリズム研究ステーション教授でもある富田悦次先生には，本書の執筆へお誘いいただいたのをはじめとして，さまざまな機会にご助言をいただき，大いに助けていただいた。ここに深く感謝したい。また，他の方々にもたいへんご多忙であるにもかかわらず拙著の原稿に目を通していただき，間違いのご指摘や改善案の提示などいろいろご教授いただいた。たいへんありがたく思い，深謝している。それは大阪府立大学の宇野裕之准教授，京都大学学術メディアセンターの宮崎修一准教授，国立情報学研究所の吉田悠一准教授，京都大学大学院情報学研究科の玉置卓助教である。それにもかかわらずまだ不完全な部分やもしかしたらまだ間違いも残っているかもしれない。しかしそれらの責任はすべて筆者である私にある。また，コロナ社には私の遅筆でたいへんご迷惑をおかけした。ここにお詫びを申し上げる。最後になったが，分野は違えどわが尊敬する学者であり，最愛の妻である，愛知県立大学日本文化学部国語国文学科教授の伸江に心から感謝の気持ちを表したい。ありがとう。

本書が一人でも多くの方の助けとなることを祈りつつ筆を置くことにする。

2017 年 8 月

伊藤　大雄

目　　次

1 はじめに

2 基本的データ構造

3　整　　列

4 集合に関する操作

5 平衡二分探索木

6 古典的アルゴリズム

7 定数時間アルゴリズム

8 数学用語の解説

記　号　表

$\deg_G(v)$，$\deg(v)$	頂点 $v \in V[G]$ の次数	p.31
$E[G]$	グラフ G の辺集合	p.30
$E_G(U,W)$，$E(U,W)$	U と W におのおの端点をもつ辺集合	p.32
$e_G(U,W)$，$e(U,W)$	$\lvert E_G(U,W)\rvert$，$\lvert E(U,W)\rvert$ のこと	p.32
$E_G(U)$，$E(U)$	カット（$E_G(U,V-U)$，$E(U,V-U)$ のこと）	p.32
$\mathrm{Ex}[X]$	確率変数 X の期待値	p.195
$\Pr[*]$	事象の起きる確率	p.195
$P_T(v)$	根（s とする）付き出木 T の辺のみを使った s–v 路	p.136
$G(W)$	頂点部分集合 W によるグラフ G の誘導部分グラフ	p.32
K_n	n 頂点完全グラフ	p.35
$K_{n_1,\ldots,n_k}$	$(n_1,\ldots,n_k)$ 頂点完全 k 部グラフ	p.36
ln，lg，log	前の二つは，前から順に底が e，2 の対数，最後のものは，底が略されている場合は，底が 2 の対数	p.6, 7, 193
$\exp(x)$	指数関数 e^x	p.193
$O(*)$	関数のオーダー表記	p.5
$o(*)$	関数のオーダー表記	p.8
$\omega(*)$	関数のオーダー表記	p.8
$\Omega(*)$	関数のオーダー表記	p.8
$V[G]$	グラフ G の頂点集合	p.30
$\chi(G)$	グラフ G の染色数	p.156
$\Delta(G)$	グラフ G の最大次数	p.156
$\Gamma_G(v)$，$\Gamma(v)$	頂点 $v \in V[G]$ の隣接点集合	p.32
$\Gamma_G^-(v)$，$\Gamma^-(v)$	頂点 $v \in V[G]$ に入る有向辺による隣接点集合	p.32
$\Gamma_G^+(v)$，$\Gamma^+(v)$	頂点 $v \in V[G]$ から出る有向辺による隣接点集合	p.32

$\times$	（集合の場合）直積，別名デカルト積	p.193
$\lfloor\cdot\rfloor$, $\lceil\cdot\rceil$	床関数，天井関数。それぞれ順に，それ以下の最大，あるいはそれ以上の最小の整数を対応付ける関数	p.29, 30
$\vee$	論理和 OR の意味。$\cup$ とも書く	p.12
$\rightleftharpoons$	データの中身を入れ替える操作	p.43
$\oplus$	二つの n 頂点グラフに対する辺集合の対称差集合	p.161
2^A	A の冪集合	p.192
$\binom{n}{k}$	n 要素からなる集合から，異なる k 個の要素を選び出す組合せの数（二項係数）	p.192
$\rightarrow$	対応	p.193
$\preceq$	多項式時間帰着可能	p.15
$\preceq$	順序または半順序	p.194
$(A, \preceq)$	A は順序付き集合	p.194
$A(m, n)$	アッカーマン関数	p.80
$\alpha(m, n)$	アッカーマン関数の逆関数	p.79, 80
$e(A, B)$	AB 間の辺の本数	p.200
$d(A, B)$	(A, B) の密度	p.201

1 はじめに

1.1 アルゴリズムとデータ構造の重要性

n（> 0）人から成るグループにおいて，各人の 100 メートル走のタイムがわかっている。その情報を基に，いろいろな基準に従って k（$< n$）人のグループを選び出すことを考える。簡単のため，各人には 1 から n までの ID（識別番号）が割り当てられていて，i さんのタイムは t_i だとする（**表 1.1** 参照）。

表 1.1 100 メートル走のタイムの表

ID	1	2	3	4	5	6	7	$\cdots$	n
タイム〔秒〕	13.5	15.2	14.1	17.0	16.5	12.7	14.4	$\cdots$	16.6

問題 1 k 人の平均タイムが最も早くなる様に選ぶ。

このとき誰でも考える当り前のやり方は，タイムの速い順から k 人選ぶ方法で，実際これで目的は果たせる。最も速い人を一人選ぶのは全員のタイムを一通り眺めればわかる。2 番目，3 番目，$\cdots$，k 番目も同じなので，計算の手間は大体 kn に比例する時間になる。これを $O(kn)$ 時間と書くことにする（厳密な定義は **1.2.3 項** 参照）[†]。これはかなり大きな n に対しても十分速い。

では今度はつぎの条件ではどうだろうか。

† あるいは全員をタイムの順に並べ替えるのは $O(n \log n)$ 時間でできるので（詳細は 3 章 参照），そうしておいて上から k 人選ぶ方法をとれば $O(n \log n)$ 時間でできる。$k > \log n$ の場合はこちらのほうが有効である。

問題 2　k 人のグループの平均タイムがグループ全体の平均タイムに最も近くなる（差の絶対値が最小になる）ように選ぶ。

グループの平均タイム $\bar{t}$ は n 個の数値を足して n で割るだけなので，たかだか $n+1$ 回の演算で簡単に（すなわち $O(n)$ 時間で）求められる。しかし平均タイムが $\bar{t}$ に最も近いように k 人選ぶのは簡単ではない。この場合，「t_i が $\bar{t}$ に近い k 人を選ぶ」というヒューリスティクスがまず思い付くが，これは最適になるとはかぎらない。

例えば，表 1.1 が見えている 8 名だけのグループ（すなわち $n=8$）だったとする。このときの全体の平均タイムは 15.0 秒となる。$k=2$ だったとすると，平均タイムに近い二人は 2 番（15.2 秒）と 7 番（14.4 秒）だが，その平均は 14.8 秒である。しかし最適解は 1 番（13.5 秒）と 5 番（16.5 秒）の組で，平均はちょうど 15.0 秒に等しくなる。

この問題を確実に解く方法はある。k 人からなるすべての部分集合を列挙して，その一つ一つに対して平均タイムを計算して，比較し，$\bar{t}$ との差が最小のものを選ぶ方法である。これならば確実に目的のものが得られるが，n 人から k 人を選ぶ選び方は

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

通りあり，指数関数であるので，あまり大きくない n と k であっても非常に巨大になってしまい†，少し大きな n と k だと，この方法では，たとえ計算機を使っても解くことはできないだろう。

ではつぎの問題はどうか。

問題 3　タイムのバラツキ，すなわち分散が最小になるように k 人選ぶ。

これも問題 2 のように部分集合を総列挙すればもちろん求められるが，この場合はもっとうまい方法がありそうだ。タイムが近いもの同士を組ませればよいは

† 例えば $n=40$，$k=10$ で 3.0×10^{15}（3 000 兆）以上。

ずなので，全員をタイムの順で並べ替え，そこで連続する k 人のグループのすべてに対して確かめればよい。並べ替えで $O(n\log n)$ 時間（詳しくは 3 章 参照），グループの数はたかだか $n-k+1$ なので，ナイーブにやっても $O(nk+n\log n)$ 時間[†1] で済む。これは問題 1 の解法とほとんど変わらない。

上で見たように，三つの問題に対して，それぞれいろいろな解法があった。これらが「**アルゴリズム**（algorithm）」である。アルゴリズムの定義は「与えられた問題を解くための，機械的操作からなる有限の手続き」などとされる。しかしそういわれてもよくわからない読者も多いであろう。具体的に問題とアルゴリズムのペアをいくつも見ていくことによって，その概念を体得していくのが一番である。

アルゴリズムにはよし悪しがある。それを測る物差しはいろいろあるが，最も標準的なのは「計算時間」である。それはすなわち，上でときどき述べていた「計算時間が $O(*)$」というもののことである。この関数の増加率が少ないものほど，よいとされる[†2]。

例えば問題 3 の解法として，問題 2 のように指数関数的に時間のかかるアルゴリズムをつくってしまうと，ごく小さい n と k の場合は別として，ほとんど役に立たない[†3]。もし君がシステムエンジニアでそんなプログラムをつくったとしたら，会社の業務は進まず，君は早晩クビになるだろう。しかし少し工夫すれば，問題 1 を解くのとあまり変わらない計算時間でできることになり，会社も繁栄し，君のボーナスも増えるかもしれない。

これがアルゴリズムの重要性である。

アルゴリズムをプログラム上で実現するにはデータ構造が重要である。データ構造とはデータを表現する方法であり，それが適切かどうかでアルゴリズムの性能が大きく変わる。データ構造の重要性を説明するには，もう少し突っ込

†1 これの意味は，nk と $n\log n$ のうち大きいほうに比例する時間がかかる，ということ。詳しくは 1.2.3 項 参照。
†2 データ量が増えても計算時間が爆発しにくいからである。
†3 一時的に使うだけで，しかもごく小さい n と k に対してだけしか使用しないとわかっている場合は，サッと書けるのでこういうナイーブなアルゴリズムもあり得る。

んだ知識が必要となるので，ここでは述べないが，例えば上のアルゴリズムの計算時間が $O(kn)$ 時間などと述べたが，データ構造が不適切だとその性能は出ない場合がある。

1.2　計算モデルと計算量

アルゴリズムと計算時間を理論的に扱うためには，厳密なモデル化が必要となる。

1.2.1　問題と問題例

上でも「問題 1」などのように「問題」を扱ったが，アルゴリズムと計算量の理論研究の分野で扱う**問題**（problem）とは，通常無限個の**問題例**（problem instance）から成っている。例えばつぎの問題を考える。

分 割 問 題

入力　n 個の整数 $a_1, \ldots, a_n$

要請　$a_1, \ldots, a_n$ を総和が等しい二つの部分に分けられる（すなわち $\displaystyle\sum_{i\in B} a_i = \sum_{i\notin B} a_i$ である $B \subseteq \{1, \ldots, n\}$ が存在する）か否かを判定せよ。

この問題において，n や $a_1, \ldots, a_n$ を具体的な数値として与えたものを問題例と呼ぶ。問題はこうした問題例が無限個集まったものとして定義される（演習問題【1】（p.19）参照）。

1.2.2　計算機モデル

計算機モデルは基本的に RAM（random access machine，ランダムアクセスマシン）と呼ばれる，「ランダムアクセス記憶をもつ，逐次型計算機」を想定する。理論的に扱いやすくするために，以下の仮定を用いる。

- メモリの 1 セルとレジスタには，任意に大きい数を格納できる。

- 各命令（四則演算・比較・メモリ 1 セル分のアクセスなど）は単位時間で実行できる。
- メモリのセル数は無限

これはすなわち実在のコンピュータの理論抽象化である。アルゴリズムの計算時間は，この計算機における計算時間，すなわち命令が実行された数（計算ステップ数）で評価される。

1.2.3　アルゴリズムのオーダー表記

（1）　ビッグオー表記の定義　　アルゴリズムの計算時間（計算ステップ数）は入力の大きさとの関連下で評価されなければならない。すなわち，問題例の入力に必要なデータ量 n の関数 $T(n)$ として，アルゴリズムの計算時間を表す。

ただし，その計算時間をあまり細かく算出しても意味はない。プログラムの組み方によって，定数ステップ程度の違いは出てくるし，それらは当然本質的ではない。そのため，アルゴリズムの計算時間を評価する際には前述の $O(n \log n)$ などという表記を用いるのが普通である。このような表記を「オーダー表記」あるいは「ビッグオー表記」などという。ここで，この表記について正確に定義を与えておく。なお，多くのアルゴリズム開発者にとっては，ここで記する正確な定義は覚えておく必要はなく，その後で述べる「大まかなやり方」を知っておけば十分である。

● **定義 1.1**　　二つの関数 $f(n)$，$g(n)$ について $f(n) = O(g(n))$ であるとは，ある整数 $m > 0$ と実数 $c > 0$ が存在して，任意の $n \geqq m$ に対して $f(n) \leqq cg(n)$ となることである。

この定義に従って $f(n) = 3n^3 + 10n^2 + 5n \lg n - 30n + 100 = O(n^3)$ であることを確かめてみよう。例えば，100 以上の n に対して明らかに $f(n) \leqq 100n^3$ であるので，$m = 100$，$c = 100$ とすれば，定義 1.1 の条件を満たすので，

索　　　　　引

—— 著 者 略 歴 ——

1985年　京都大学工学部数理工学科卒業
1987年　京都大学大学院修士課程修了（数理工学専攻）
1987年　日本電信電話株式会社基礎研究所
1990年　日本電信電話株式会社通信網総合研究所
1995年　博士（工学）（京都大学）
1996年　豊橋技術科学大学講師
2001年　京都大学助教授
2007年　京都大学准教授
2012年　電気通信大学教授
　　　　現在に至る

データ構造とアルゴリズム
Data Structures and Algorithms　　© Hiroo Itoh 2017

2017 年 9 月 28 日　初版第 1 刷発行

検印省略

著　　者	伊　藤　大　雄	
発 行 者	株式会社	コ ロ ナ 社
	代 表 者	牛 来 真 也
印 刷 所	三 美 印 刷 株 式 会 社	
製 本 所	有限会社	愛 千 製 本 所

112–0011　東京都文京区千石 4–46–10
発 行 所　株式会社　コ ロ ナ 社
CORONA PUBLISHING CO., LTD.
Tokyo Japan
振替 00140–8–14844・電話(03)3941–3131(代)
ホームページ　http://www.coronasha.co.jp

ISBN 978–4–339–02702–0　C3355　Printed in Japan　（金）