

データの扱い

山本昌志*

2007年11月16日

概要

データの扱いについて簡単に学ぶ。

1 本日の学習内容

データの取り扱い，とくにデータのモデルを記号 (図) で表す方法を学習する．本日のゴールは，以下のとおり．

- データのモデル化とその性質がわかる．
- 代表的な論理モデルの表し方が分かる．

教科書 [1] の pp.71-95 が本日の範囲である．ほとんど教科書に沿って，講義を行う．プリントは補足説明，あるいは簡単にまとめたものと考えよ．したがって，諸君は自分で教科書を十分読まなくてはならない．

また，執筆者のひとりである東京大学の玉井先生の講義の記録が WEB で公開¹されている．非常に，参考になるので一度見ると良い．

注意 このプリントは完全ではない．すなわち，説明不足が多く，論理が通っていない．教科書とあわせると，このプリントは完全になる．したがって，このプリントを見るときには，教科書とともに見なくてはならない．講義では，教科書とともに説明する．

2 データモデル

2.1 データとデータモデル

教科書 [1]

「コンピュータの処理対象となる符号化された情報」の意味でデータを使う

*独立行政法人 秋田工業高等専門学校 電気情報工学科

¹<http://lecture.ecc.u-tokyo.ac.jp/~tamai/joho/2006/chapter04.pdf>

と書いている．符号化された情報とはなんだろうか？

符号化とは，情報を記号によって表現することを言う．コード化と言うこともある．この記号で表現されたものを符号（コード）と呼ぶ．コンピューターは0と1で符号を表す．ようするに，コンピューター内部で0と1のビットで取り扱うことができるものをデータととして取り扱う．

逆に符号化できない情報があるだろうか？人間の感情はどうだろうか？人間の感情を言葉で表せば符号化は可能である．しかし，そこに言葉で言い表せない何か（情報）があるとするならば，符号化は非常に困難になる．このようなとき，人間の状態—例えば心拍数，血圧，体温，脳波など—を可能な限り測定器で測って符号化しても無駄であろう．このような情報は，コンピューターで取り扱わないことが得策である．

これ以降は，符号化できる情報のみを取り扱う．コンピューターでは，そのようなものしか取り扱うことができないからである．

2.2 データモデルのレベル

符号化できる情報をコンピューターで取り扱う場合，対象をモデル化する必要がある．このモデル化の善し悪しによって，取り扱う効率が著しく異なる．良いモデルだと，アルゴリズムやデータ構造が簡単になる．

モデル化する方法は，一意に決めることができない．同じ対象でもモデル化の方法は，千差万別である．対象の本質を見抜き，記号で表し，コンピューターで取り扱える形にしなくてはならない．情報処理の達人は，これが上手にできるのである．

データ（対象）をモデル化する課程には，3つのレベルがある．

概念モデル (Conceptual Model)	処理する対象を人間が認識しているが，記号で表すことができないモデル．人間の頭の中にあるだけのもの．
論理モデル (Logical Model)	処理する対象を記号で表したモデル．
物理モデル (Physical Model)	処理する対象をコンピューターで取り扱うことができるモデル．

モデル化の第一段階（概念モデル）は，モデル化を行う人間（プログラマー）の観察力，あるいは理解力に依存する．これは，学校の講義で教えることは難しい．実際の仕事で訓練を積まなくてはならない．

モデル化の第二段階（論理モデル）は，ある程度学校の講義で教えることが可能である．これまでの人類が経験した知識として，いくつかのモデル化の方法があり，それを図（記号）で表すことができる．代表的なモデル化の方法については，この後の2.4節での中心的な話題となる．

モデル化の第三段階（物理モデル）は，実際のプログラミングの講義，とくにアルゴリズムとデータ構造の中心的な内容である．データベースの演習があれば，それも含まれる．他にも対応するものもあるだろう．これらは，学校で基本的な内容を教えることはできる．

2.3 データモデルの性質

このあたりは，教科書の通りに説明する．まとめると，以下の通り．

完全性	対象の全ての要素に対応するモデルの要素がある．全ての対象が表現可能と言うこと．対象の要素に対して操作 (演算) を行った場合，その結果の要素 (対象に含まれる) に対応するモデルの要素も存在する必要がある．
一意性	一つのモデルの要素には，一つの対象の要素がある．モデルの一つの要素に対して，対象の要素があるときあいまいと言う．
忠実性	対象の 1 つの要素が，モデルの 1 つの要素のみに対応している．
整合性	対象の要素の間で成り立つ規則 (演算) が，モデルの要素間で成り立つ．
無冗長性	同じ情報が表現の複数箇所で表れない．
拡張性	データモデルを変更したとき，既存のデータの表現を変更せずに新しいデータを表現できること．

2.4 代表的なデータモデルと演算

教科書 [1] のこの節 (pp.79-93) は，データモデルの 3 つのレベルの論理モデルに関する話である．ここでは，処理すべき対象を図 (記号) で表す方法を示す．

論理モデルは，次のレベルである物理モデル，すなわちコンピューターでの実際の処理につながる．また，論理モデルは人間の頭の中にある考え (概念モデル) から来るものである．したがって，この段階は人間の概念とコンピューターをつなぐ部分となり，情報処理を行う上で極めて重要なである．

幸いなことに，論理モデルはいくつかの代表的なパターンに分けることができる．これまで，人類が情報処理を行ってきて経験的に学んだ方法である．諸君が解くほとんどの問題 (情報処理) は，これらのいずれかのパターンが応用できるであろう．

2.4.1 集合モデル

このあたりは，教科書のとおり説明する．

- 集合について学んでいれば，ベン図と演算は簡単であろう．
- ベン図とオイラー図の違いは，理解してほしい．ベン図は必ず重なるが，オイラー図は重なるとは限らない．一般的には，オイラー図の方が良く使われ，諸君もしばしばいろいろな説明の時に使っているはずである．あのオイラーが考案したから，オイラー図と言うらしい．こんなもので，名前がつくとは …．偉い先生は得をするものだ．

2.4.2 ネットワークモデル

このあたりは，教科書のとおり説明する．

2.4.3 階層モデル

このあたりは，教科書のとおり説明する．

2.4.4 関係モデル

このあたりは，教科書のとおり説明する．

2.4.5 オブジェクト指向モデル

このあたりは，教科書のとおり説明する．

参考文献

[1] 河合慧（編）．情報．東京大学出版会，2006．