

モデリング雑感

土谷 隆

1. はじめに

私は理想的にはモデリング・アルゴリズム・数理の研究者でありたいと思っている（あくまでも願望であるが）、普段モデリングモデリングと気軽に口にする割に、モデリングについて書くことがかなり大変であるということには依頼を引き受けてしまってから気がついたのだがもう遅い。ここでは、かなり断片的ではあるが日頃考えることを綴って何とか責を果たしたい。

2. モデリングと数学と数理科学

一言でいうと、モデリングとは「現象を再現できる法則を見い出すこと、あるいは法則を見い出そうとする行為」だと思う。その際の法則がモデルである。ここでのキーワードは「現象」と「再現」である。法則であるから、適用できる状況がかなり明確に特定されていることが必要であると思うし、法則を書き下す手段も問題となる。我々はその手段として数学を思い浮かべがちであるが、生物学や化学などにおけるモデルを考えると分かるように、モデルを記述する言語が数学であるとは限らない点にも注意したい。そしてデータをモデルを用いて説明することが科学である、と考えるならば、少々大げさではあるがモデリングは科学そのものであるとさえいえよう。

とはいえ、我々にとって身近なのは、数学を記述手段とした所謂「数学モデル」である。我々が研究しているORの数学や最適化は、いわゆる「純粋数学」と対比して「応用数学」といわれることが多い。しかし、上述の科学とモデリングは不可分のものであるという立場に立つと、より適切な言葉は「数理科学」ということになる。よりものを作るという工学的な立場からみれば「数理工学」ということになる。「純粋数学」と「数理科学」の違いは、数式という同じ「言語」を用いていても、外界（自然や社会）をモデリングして

記述しようとする意識があるか否かということのように思える。これが言いすぎであれば、意識の強弱と言おうか。

3. モデルという言葉

モデルという言葉はいろいろなレベルで使われる：最適化モデル、統計モデル、線形計画モデル、ARモデル、ナースケジューリングモデル、授業割当てモデル等々。これらの用語はどのように使い分けられるのだろうか。

物理学は数理科学の雛形であるし数理科学の元祖である。そして、守備範囲を定めることなく、いろいろな分野に果敢に進出する（そのことをやっかみ半分に「物理帝国主義」と揶揄することもしばしばあるが:~:)。これは次のように考えられるのではないだろうか。物理学は「物理モデル」というメタモデル（方法論といってもよいかもしれない）に依拠して他の分野のモデリングに果敢に進出していくのだと。そのように考えると「最適化モデル」や「統計モデル」もここで書いた物理モデルと同じようなメタモデルのことなのだと考えれば、それなりに納得がいく。最適化というメタモデルに依拠したモデルが最適化モデルであるし、確率分布というメタモデルに依拠したモデルが統計モデルなのではないだろうか。モデリングという観点からは「幾何学モデル」という言葉もこのレベルで気になる言葉である。

一方、ARモデルや線形計画モデルはいろいろな現象の説明に共通の構造を集約した最大公約数的なモデルであるということが出来る。このレベルのモデルは一般性があり、数理的なモデルというに相應しいものなのかもしれない。モデリングの主要な道具であり、純粋数学や計算機科学との橋渡しの役割を果たす部分である。

ナースケジューリングモデルや授業割当てモデルはそれらに比べるとずっと個別的である。通常「モデリング」という時には、このレベルで現象やシステムをモデリングすることをいうように思う。これらのモ

つちや たかし
統計数理研究所
〒106-8569 港区南麻布4-6-7

デルは個別的ではあるが、現象やシステムの本質を捉えており、予測や理解に役立つ情報を提供できるのであれば、モデルと呼ぶに値するということになる。

4. モデルとは？

先日統数研を退職された田邊先生は「帰納は人間の原罪である」ということを退職講演の中でおっしゃっておられたが、これは、人間が適切に環境に適応し状況に対処するためには、どうしても有限の経験を一般化する必要がある、帰納的な推論が人間にとって不可欠であることを意味されたのだと思う。そして有限個の例から無限の場合を敷衍するのに必要なのがモデルである。

一方、「モデルの本質は模倣ではなく捨象である」、そして「客観性を標榜しながら、きわめて主観的なものである」ということもしばしば言われることであるし、大いに共感できることである（私はこのことを伊理先生が何かの折に書かれておられたのを読んで、な～るほど！ と思い随分感心した覚えがある：-)）。

人間は睡眠中に夢を見て忘却しているという。これも一種の捨象と見られないこともない。すべての物事を軽重の区別なく忘れることなく覚えていたら大変なことになるであろう。また、機械学習などのモデルで忘却係数を導入することは高い汎化能力を実現する上では不可欠であるとされているようである。捨象することにより、構造が固定され、有限個の例から適切な予測ができる。これがまさに本質を掴んでいるということなのではないかと思う。

さて、モデルを分類する軸はいくつか考えられる。例えば演繹的モデルと帰納的モデル、パラメトリックモデルとノンパラメトリックモデル、予測するためのモデルと構造を理解するためのモデルなどである。

演繹的モデルはモデルの背後にあるメカニズムを適当に想定してモデルが構築されるのに対し、帰納的モデルは入出力関係のみを適切に記述するところに力点を置いて構築される。

パラメトリックモデルはモデルが持つ自由度を少数のパラメータに完全に集約する。これに対し、ノンパラメトリックモデルは、パラメータは十分に多く用意しておき、それらの間に緩やかな制約を置いて縛る。

モデルは現実を適切に記述してはいけない。そのような立場からは、現象を予測できるモデルが一番理想的な立派な(?)モデルである。一方、現象をより人間が分かりやすく捉えるためのモデルも存在する。

その場合、モデルは眼の役割を果たす、視るための装置である。

筆者がモデルについて考える時には、これらの視点が微妙に絡まり合い交差する。

5. モデリングとアルゴリズム

モデリングを行う以上は「解けるモデル」を構築すべきである（ここで「モデルが解ける」とは何を意味するのかというのも難しい問題だが、とりあえずは「モデルを数学的、数値的に解析してそこから有用な情報を取り出せる」という程度の意味で考えてみたい）。現実を模倣すると称して複雑なモデルを考えることはいくらでもできるかもしれないが、解いて実際と比較することができないようなものが本当にモデルと呼べるのだろうか？

世の中の数学的問題で解析的・計算的に解けるものはごくわずかである。そのごくわずかなクラスの問題をうまく組み合わせ、解析しようとする対象のモデルを、必ずしもきれいでなくても何とかうまく解ける形に定式化するのがモデリングであると思う。そして「解けるモデルのクラス」は計算機の性能やアルゴリズムに依存する相対的なものである。そのような意味でモデリングと計算機の性能やアルゴリズムは不可分の関係にある（さらにMATLABやS、SIMPLE、AMPL等のモデリング環境も重要な要素であるが）。

筆者が大学院生であった1980年代中頃はモデリングは計算機センタにあるような大型計算機でないとできないと考えられていたが、今は90年代初頭の世界のトップクラスのスーパーコンピュータと同等の性能のパソコンが当然のごとく皆の机上におかれている。このような環境の違いの下では人々のモデリングに対する感覚や態度も自ずと異なるであろう。このことは、計算機性能の向上がメタヒューリスティックスなどの手法の実用化に結びつき、それが最適化モデリングに与えている影響などに端的に見ることができる。

次にアルゴリズムの進歩について見てみよう。例えば凸2次計画モデルは90年代初頭には標準的に使えるモデルではなかったが、内点法の登場などにより、現在では諸方面で必須のモデルとしていろいろなところで活躍している。サポートベクターマシンの登場も凸2次計画モデルの実用化がその背景にあるとも捉えることができる。そして線形計画モデルや凸2次計画モデルなどに加え、現在研究が進行中の半正定値計画問題や2次錐計画問題などが同様に標準的なモデリン

グの道具として使えるようになれば、最適化モデリングの世界もまた一段と広がるであろう。

逆に、難しい問題に対するアルゴリズムにおいてモデリングが果たす役割も見逃せない。マルコフ連鎖モンテカルロ法やインポートランスサンプリングなどの高次元積分法、組み合わせ最適化問題を解く上でのメタヒューリスティックス等において個別のモデルを解くためのパラメータの調整、非線形計画問題の適切な初期値の選択などは、まさに問題や現象を熟知したモデラーの腕の見せ所であろう。このような部分はアルゴリズムの一部であるともみなせるが、陰的な意味では対象の適切なモデリングを行っているといえなくはない。このような意味でも、モデリングとアルゴリズムは不可分な関係にある。

6. モデルと感度解析

モデルの当てはめは、モデルのパラメータを調整して、適当な当てはまりの度合いを示す目的関数を最適化することにより行われる。その際に気になるのは、パラメータに対する当てはまりの感度である。少しパラメータの値がずれたからといって、モデルの性質がまるで異なってしまうのは困る。しかし、パラメータの変動に対する目的関数値やモデルの安定性がきちんと検討されることは少ない（なかなかそこまで手が回らないということもあるだろうが）。

モデルは部品となる要素モデルを結合したものとして得られることも多い。その場合、要素ごとにパラメータを合わせて全体を合わせることはよく行われる。そのような場合でも要素モデルのどの部分をどの程度変化させるとモデルの性質がどの位異なるかを検証することはほとんど行われていないのではないだろうか。

パラメータの変動に対するモデルの安定性を確認することはモデルの性質を理解する上でも、そのモデルの妥当性について納得する上でも大事なことだと思う。特に統計モデルなどのように、「尤度や情報量基準などの目的関数が最適化されているという事実」を重要な情報としてモデルの良さを検討せざるを得ないことが多い場合などは、安定性に対する検討が理想的には不可欠なのではないだろうか。

計算機の能力が格段に向上した現在、何千、何万回とパラメータを摂動して問題を解いて安定性を確認することが可能な場合もあると思うし（その場合も人手を介入せずに問題を繰り返し解けるような安定した最適化アルゴリズムがあることが必須ではあるが）、高

速自動微分法やロバスト最適化などの手法がモデルの安定性を検討したり保証したりする上での有効な手段として今後大いに役に立っていくのではないだろうか。

7. モデルの高次元化

計算機の性能が向上したこともあり、時系列解析や時空間モデルにおいては、数千、数万次元空間上での高次元積分などを計算する必要がある。そのためにマルコフ連鎖モンテカルロ法や粒子フィルタなどの手法が盛んに研究されている。また、サポートベクターマシンに代表されるカーネル法では、非線形変換によって数千次元から数万次元の空間にデータを放り投げて、その上で種々の統計手法を適用する。このように、現在では、相当な高次元のモデルを何らかの形で取り扱うことはそれほど珍しいことではなくなりつつある。

高次元を扱う時には、単純な確率計算からは理解が難しい不思議な、あるいは、非直観的なことがしばしば起こる。そして次元に対し非常に少ないサンプル数でしばしばうまく所与の目的を達成するようなことができることもある。「次元の呪い」ということがしばしばいわれるが、いわばその逆である。今後、モデリングにおいて、我々はこのような高次元の不可思議さにより本格的に向き合うことになってくるのではないかと思う。

8. 特異値分解や主成分分析、クラスタリング

モデリングにおいては、対象に対し、何らかのメカニズムや事前情報を想定し、モデルに組み込む。データから有用な情報を抽出するための枠組みは必ずしもこのようなレベルのモデルとは限らない。そのような枠組みとして代表的な手法に特異値分解がある。統計で言えば主成分分析である。特異値分解は画像データやテキストデータマイニングなどにも用いられているようである。このような手法が有効であることに異論はないのだが、モデリングという観点からどう位置づけられるかは興味のあるところである。線形代数そのものがモデルと考えるのが適当なのかも知れないが、よく分らない。同様なことは、クラスタリングについても言える。これらの手法が妙に説得力があるのは何とかして情報を要約したいという人間の本能に根ざすものだからなのかもしれない。

9. 数理科学とメタモデル

ここで少し脱線して、数理科学の話をしてみたい。昔から折に触れ思ってきたのは、世の中には「基準のある数理科学」と「基準のない数理科学」があるということである。これが一体モデリングとどんな関係にあるのかと思われるかもしれないが、筆者の頭の中では次のようにして繋がっている。「計算複雑度という一つの基準」を例にとって説明してみよう。アルゴリズムの世界が発展した大きな理由の一つは計算複雑度という、アルゴリズムの良し悪しを計る合理的な尺度の存在であると思う。これは「計算複雑度がアルゴリズムの性能を計るよい（メタ）モデルであった」ということを意味するのではないだろうか。同じようなことは、AICなどの情報量規準についてもいえる。

さて、世の中で研究されている最適化問題というところ、いろいろなものがあるが、「線形計画問題」と「巡回セールスマン問題」はその中でも特別のものであるように感じている人は多いと思う。それは、これらの問題が「最適化アルゴリズムの研究をする上での良いモデル」であったからではないだろうか。線形計画問題は50年間もの長い間、最適化の分野でいくつもの大きな研究の進展を導く突破口となってきた。巡回セールスマン問題についても同様である。そして、それらの進展を裏側から支えてきたのが、計算複雑度という「問題を解く手間の良いモデル」であったのではないだろうか。

ひと頃、筆者は数理科学の分野が発展していくには計算複雑度のような良い規準が存在することが重要であると考えていたが、最近はやはり面白いモデルを考えるということが重要なのだと思うようになった。しかし最終的には、モデルやアルゴリズムの有効性を主張するためには、皆がその良さを共有できるための「良い基準」が隠し味として必要不可欠なような気がするのである。Karmarkar法における多項式性やサポートベクターマシンにおけるVC次元など、説得力のある合理的な基準（しつこいようだがこれもまたモデルである！）に照らして良いといわれれば、とりあえず人は使う気になるものである。

10. モデリングは科学か工芸か芸術か？

本稿の最初の方で整理したように、モデルにはその抽象性に応じていろいろな段階があるが、やはり身近に感じるのは、現実問題を解決したり理解したりする

ためにされるより具体的なモデリングであり、モデルである。その種のモデリングはもちろん世間で山のよう日々営まれていると思うが、得られるモデルは工芸作品に性格に近いのではないかと思う（実はさらにモデリングは「工芸」か「芸術」かという話はあるのだが、ここでは工学的な立場に立って工芸作品と呼んでみたい）。人が絵や音楽を楽しむことができるように、モデルも鑑賞の対象となるのではないだろうか。鑑賞にもいろいろな段階があるとは思いますが、うまくいったモデリングの話を読んだり聞いたり学んだりするのは面白いし、わくわくする（心躍る）ものである。

赤池先生と中川先生の有名なセメントキルンの統計的制御の話や、深尾先生、田邊先生、尾形氏、井上氏の地震波による地球内部構造トモグラフィ、池上さんのナーススケジューリング、今野先生の授業割り当ての話、統数研の北川所長と共同研究者の松本氏による地下水位変動と巨大地震の関係の解析、同じく統数研の尾崎氏と共同研究者の彭氏、豊田氏、中野氏らの非線形モデルによる火力発電所の統計的制御の話などは印象に残ったモデルである。また、鉄道総研の笹川氏によるリニアモーターカーの磁気シールド最適設計のモデルも、簡単でありながらも巧妙で現実によく合うモデルとして、筆者には印象深いものであった。このモデルについては、内点法による求解やロバスト最適化について共同研究を行う機会に恵まれたが、モデルをいじりながらいろいろと学ぶ所が多かったと思っている。

それに対し、モデリングの科学とは何だろうか？それは工芸的なモデリングのための道具立て、方法論の研究だと思う。例えばAICに代表される情報量基準などは典型的なものであろうし、最適化や数値計算法などもそうであろう。田邊先生はソフトウェア、ハードウェアに対比されるべき重要な要素としてモデリングの方法論を「メタウェア」と呼んでおられる。筆者にとってはなじみやすい、この重要な概念を表すための良い言葉であると思う。

11. おわりに

さて、そろそろ紙面も尽きてきたようである。いろいろとまとまりのないことを書いてきてしまったが「雑感」ということでお許しいただきたい。もっと修行を積んでいつかは立派な「論」を書きたいものである。:-)