アブダクションによるデータ解析 ——計算予測結果を読み解く力——

日本原子力研究開発機構 中島 憲宏

シミュレーションを設計過程において活用するためには、それを使いこなす技術の他に、シミュレーションが紡ぎ出す膨大なデータの読解力が必要である。設計案に対するシミュレーション解の分析や評価過程に、人工工学が提唱する「どの視点も取り入れた仮説・法則や行為を導出するためのアブダクション基盤」を取り入れ、想定外や見落としなどを最小化できるように人工知能などを活用したデータ解析技術の取り組みなどを概観する。

KEYWORDS: *Data analysis, Structural analysis, Assembly, Neural network, Design engineering, Computer aided design, Design Science, Artifacts, High Performance Computing, Artificial Intelligence*

I. はじめに

設計は、多数の人間がかかわっている問題である。設計問題では、設計条件が定められるものの必ずしも厳格に規定されていない場合や、設計作業を行うために必要となるデータが不足している場合もある。また、設計の結果となる設計解と設計条件の関係は複雑であり、一般的にはよくわからない問題とされている。このような設計過程を科学する学術としての設計科学や、設計過程の効率化や設計対象の高品質化などを目指す設計工学という分野がある。それらの総合化体系を目指す学術として人工工学が存在する。

設計過程において、設計解が出たとしても、その評価は必ずしも定まるわけではない。どういう結果が最も好ましいかを客観的に選択するために、複数生まれうる設計案を比較する必要がある。当然のこと、設計案の比較は多元的となり、異なる解の比較は論理的に判断することが難しい。このような設計空間にシミュレーションという設計計算技術はどうかかわるべきであろうか。また、シミュレーションが出力した計算結果をどのように分析し、評価するべきであろうか。

全く初めての設計対象と長年にわたり改修設計されてきた設計対象とでは、設計条件数やデータ量も異なる

が、ここでは、個別の設計案件にとらわれることなく設計過程にシミュレーションという道具を導入した場合、シミュレーションの紡ぎ出す結果をどのように読み解けるか、いわゆる人工知能技術を活用したデータ解析技法とその関連技術について解説する。

シミュレーションという一般の設計計算手段よりも幾分か高度化された道具を用いると、どうかすると与えられた計算モデル、あるいは選択したモデルやシミュレーション手段の準備といった計算結果を得る過程に多くの時間をかけ、結果が出たことに満足してしまう傾向を否定できない。シミュレーションという道具を使いこなすのに疲れて、その結果を吟味しないのでは、まさに「もったいない」。加えて、せっかくシミュレーションという道具を使い、それが紡ぎ出した多くの情報を全て明察しないのでは、何のためにシミュレーションという道具を用いたのかということになる。一方で、厳密な意味での至大な設計解を求めることは相当困難であると述べた通り、その設計解がどれだけ求めるものに近づいたのかを知ることは、大きな価値がある。

設計およびシミュレーションの実施では、ある種の先入観を持って取り組まざるを得ないが、その先入観を躊躇なく改めるべきときに改められることは肝要である。そのためにも、先入観を改めた「より良い」設計解、例えば想定外な発現を縮小できるような設計解を創出する方法論の確立が重要である。完全なる要求仕様が設計対象に与えられたとしても、一般的にそれらの仕様を満たす設計解を解析的に求めることはできないのだから、なん

Data analysis based upon abduction ; For better understanding the result discussion in computational science and engineering : Norihiro Nakajima.

(2017年5月31日 受理)

らかの初案を作って要求仕様を満たすまで改案するための分析過程は必須である。先入観は必要悪であるがゆえに、初案を肯定あるいは否定できる手段として、シミュレーション結果を様々な視点から分析しうるアブダクションを基盤としたシミュレーション結果の分析ツールを、筆者は提案している。

1. シミュレーションと結果の分析・評価

(1) 設計と解析

目的を明示されている設計計算は、理論計算など単機能な計算手段で十分である。設計過程においてシミュレーションを活用する意義は、目的とする設計計算解を求めるだけでなく、どの視点も取り入れて計算結果を吟味することにある。フロント・ローディング設計に代表されるように、設計案ができたら、まずはシミュレーションをする。その結果、当該設計案に対して、想定外や思いも知れない問題や過剰設計箇所が見つかるかもしれない。これは設計案をシミュレーションにより創出しようというものではない。シミュレーションはあくまで分析ツールであって、統合化ツールにはなりえないからである。重要なことは、どの視点も取り入れて設計案を分析してみるという行為である。

もしシミュレーションを単に設計計算の代替道具としてのみ使うのであれば、シミュレーション結果の分析は、定められた事項についてのみ分析すればよい。しかし、シミュレーションが出力する情報は多岐多様にわたり、従来の設計過程では必要としなかったデータまで出力できるため、これらの計算結果を吟味することにより、発見的に問題点や過剰設計箇所あるいは想定外の設計結果を見出せる可能性がある。余談だが、設計の目的事項だけを分析するのであれば、目的にあった単機能の計算を行えば十分であり、シミュレーションを実行したとしても単なる設計案の確認のためだけにというジレンマに陥る。この議論は、シミュレーションは設計過程に不要であるという計算機援用設計論という方法論への批判である。実際、シミュレーションを設計過程に導入するだけでは、この不要論問題は解決しないのである。

(2) 人工物工学

吉川弘之元東京大学総長は、1992年に人工物工学を次のように提唱した¹⁾。

「現在、我々は、環境、貧富、安全、健康など、多くの困難な問題(現代の邪悪なるもの)に直面している。これらに共通する点は、人類の安全と豊かさを求めてきた行為の結果として、全く予期せず生じた問題であるということである。人類はこれまで、知恵を駆使し、学問を構築することによって、多くのものを生み出してきた。しかし、現在の学問は領域性と視点の限定によって構築されたものであり、これらの問題の解決のために適用でき

ないのはおろか、むしろこれらの問題を生ぜしめた原因となっている。その解決のためには、人間が創出するものすべてを対象とし、領域を否定し、どの視点も取り入れることが可能な新たな学問を構築する必要がある。これは、従来の演繹を基盤とする学問ではなく、仮説・法則や行為を導出するためのアブダクションを基盤とした学問である。それを人工物工学と呼ぶ。」

ここでアブダクションとは、仮説形成や仮説的推論と和訳されることが多いが、推論という帰納法の一つであり、個別の事象を最も適切に説明しうる仮説を導出する論理的推論であり、個別的・特殊な事例から一般的・普遍的な規則・法則を見出そうとする論理的推論の方法のことである。帰納とは、演繹ではない推論をさし、具体的には枚挙的帰納法、類推、アブダクションの三種がある。ここで、数学的帰納法や構造的帰納法などの帰納法と称されるものは、名前上は帰納法とされているが、演繹手段であるものが多い。

演繹は、仮定と規則から結論を導く。仮定と規則に合致しているならば、その結論は真となる。妥当な演繹は、仮定が真であれば結論も真であることを保証している。帰納は、仮定したことが結論を伴ういくらかの事例を観察した結果、その規則はたぶんそうであろうと考えられるとし、ある程度確実であると推論する。帰納は、推論した規則が真であることを保証していない。

設計過程においては、先入観を持って取り組むことが強いられると前段で述べた。この先入観こそが仮説として立てられていく。しかし、人工物工学的には、あらゆる視点が肝要であるため、一つの先入観ではなく、複数の先入観から仮説を立てることになる。複数の先入観をいかにして立案すべきであろうか。逆説的にシミュレーションが紡ぎ出した膨大なデータのすべての組み合わせを考慮し、推論する考え方は、まさに視点限定を解除する方法となりえる。ここで、計算科学の出番がくる。なぜなら計算機は、ある領域をすきまなく一面網羅する作業や処理を得意としているからである。つまり、膨大なデータの組合せを、計算機を用いて情報分析し、その結果から出力される多様な仮説を設計者に開示し、様々な視点からの分析結果を設計者に知ってもらうのである。

2. データを活用するとは

人工物工学にもとづく設計では、データをどう活用するかが最大の課題である。シミュレーションが紡ぎ出した膨大なデータをどのように分析し、評価すべきか。演繹的な手段は容易に準備できる。体系化された経験則や技術者の勘、多くの設計対象物に対する寸法や性能、特性、動作条件等にかかわる推奨値や設計にかかわる報告書などが含まれるデータベース、工学的な例えば材料力学や統計学等を用いた定理や理論を駆使して、シミュレーションが紡ぎ出したデータを演繹的に分析あるいは

評価できる。これに加えて、仮説・法則や行為を導出するためのアブダクションを基盤とした分析あるいは評価を実現すれば、人工物工学的な考察にもとづく設計過程を一部実現しえる。

人工物工学では、データ活用を強化するモデル論が議論されている。人工物には多くの IT 技術が含まれつつあるからでもあるが、人工物そのものや人工物システムの設計では、データをどう活用するかは、領域性と視点限定を排除していく上で重要な課題である。シミュレーションが紡ぎ出すデータの価値は、これらデータから抽出しうるデータの意味を判断して決まる。例えば、構造解析の場合、変位の持つ設計上の意味である。変位が許容量を超えると、その部品が機能しなくなる場合、変位量というデータの最大変位量すなわち許容量という仕様が変位量の評価において意味を与える。演繹的には、その許容量を超えたか否か、あるいはその値近傍にどのくらい近づいたかなどは分析・評価できる。このような分析は自動処理が容易である。モデル内では、許容量は変位量の意味ではなく、また変位量そのものでもないの、意味のメタファー表現となる。メタファーの表現は、「変位量は許容量である。」というように直接的に許容量の属性を変位量に移して記述するものである。こういったデータを活用していくためには、データベースすなわちデータ表現を記述するモデルが必要となる。シミュレーションが紡ぎ出すデータとそれらのメタデータ、そしてこの二つを連想結合させるデータがモデル内で記述される。このような、いわゆる知識ベースなどに類するデータモデルを構築することで、いわゆる人工知能技術を活用したデータ解析技法の確立が可能となる。

II. 人間が考えるようなデータ解析

1. 生物学的な情報処理

電子的に動作する計算機械をコンピュータと呼ぶが、現在普及しているコンピュータはノイマン型と呼ばれるものが主流である。このノイマン型コンピュータを発明したジョン・ルートヴィヒ・フォン・ノイマンは、原子力分野にも大きくかかわった数学者であり物理学者である。演算能力に長けたノイマン型コンピュータに対して、コンピュータの機能を人間の知的活動に近づけたいという「思い」は誰にでもあったし、その思いを今も抱くこの分野の研究者は少なからずいる。ノイマンも例外ではなく、自己増殖オートマトンの理論を構築し、生物学的な情報処理やコンピュータの発案を同時期にしている²⁾。生物学的なコンピュータ例えばニューロ型コンピュータの発案もノイマン型コンピュータと同時期の1940年代に行われている。その流れは1980年代に至って、神経回路網プロセッサとして発表されている^{3, 4)}。情報工学の分野では、マルチタスク処理やデーモンと呼ばれるバックグラウンドプロセスとして動作するプログ

ラム、エージェントとそのマルチ化などが生物学的な情報処理技術の促進に寄与した。設計工学の分野では、沖野により自律的に自己の諸元を決定する概念として、オブジェクト指向の前身的あるいは揺籃期的存在であるモデルロンが提唱され、1982年には具現化したものが現れ、1988年に生物型生産システムが提唱されている⁵⁾。人間社会はあいまいさがともなうということから、不正確さと不確定性を取り入れるためにファジー集合を情報工学に取り入れ始めたのは1960年代である。エキスパートシステムも1970年代に研究レベルではほぼやりつくされた。以上のように、人間の脳内処理を情報処理化しようとする研究は、それなりに長い歴史の中で行われてきた。中世の時代ではないがこの頃、人工的に知能を創るということは、それこそ神をも畏れぬ研究だという議論もあったほどである。

Million Connections Per Second という生態情報処理速度を計る単位がある。これは1秒間に何百万回情報伝送がされるかというもので、人間の場合 10^9 MCPS と推定されている。そのようなことから、かのパーセプトロンとの係りも深いミンスキーの Thinking Machine, いわゆる並列計算機の提案が始まる⁶⁾。心はどうはたらくのか？人間の知能についての新しい考え方を示し、心とは、「一つ一つは心を持たない小さなエージェントたちが集まってできた社会」と提示する。これが計算機内脳の一つの概念である。ちなみに彼が1954年に書いた学位論文は、Neural Nets and the Brain Model Problem である。

2. ニューロ

Deep Learning という言葉により、第4次か第5次かのニューラルネットワークや人工知能ブームが訪れている。人間の神経細胞を計算機上に具現化したものをニューロンと呼び、このニューロンを活用したシステムや仕組みをニューロコンピューティングであるとか、ニューラルネットワークなどと呼んでいる。研究の最終目標は、人間に近い能力を持ったコンピュータシステムの実現である。人間の脳が科学的に解明されているわけではないので、分かる範囲と推察可能な範囲で人間の脳をお手本にして、研究が進められている。

計算機上に具現化されたニューロンは、ユニットとかセルあるいは Processing Element などと呼ばれ、多入力1出力を基本とする。出力の特性関数として0と1が切り替わるようなステップ関数や0から1へ滑らかに変化するようなシグモイド関数など様々な応答特性が提案されている。このニューロンをたくさん集めて結合させることで脳のモデル化を進めており、ニューロンを計算機上でネットワーク化し、システムを形成している。ネットワークの形式ではそれぞれのニューロンが直接結合する相互結合型ネットワークや上下の階層間のニュー

ロが互いに結合する階層型ネットワークなどがある。ニューロンは外部環境に合わせて自身を変化させて調整するという学習機能があり、これを自己組織化と呼んでいる。簡単に言えば、自動プログラミングに相当する。この学習機能は、多入力の入力データにそれぞれ重みをつけるというもので、100個のニューロンの相互結合型ネットワーク型の場合だと、99の入力に重みがそれぞれ割り振られることになる。この重みを例えば0から9までに整数で与えるとすれば、10段階の重みと100個のニューロンからなるシステムでの組合せ(10の100乗)について重みをどうつけるかという評価をしなければならなくなるため、高性能な計算機が必要となるわけでもある。この評価処理においては、教師付きと教師なし型の学習方式が提案されており、外部から正解を与えるものを教師付き学習と呼んでいる。

人間の脳の神経細胞は百億個以上あるともいわれており、それに近づくためのシミュレーション環境として、超並列かつ高度分散処理を可能とするコンピュータが必要である。京(毎秒1京回の浮動小数点演算ができることから、 10^7 の和式の単位から由来として命名)というコンピュータで705,024並列、現在、世界最高水準といわれる中華人民共和国の神威・太湖之光で10,649,600並列であり、Thinking Machineの具体化が見えつつある⁷⁾。

III. Cerebral Methodology in Data Analysis

設計計算においてシミュレーションを有効活用しようとするならば、シミュレーションが出力する全ての情報を活用すべきである。設計者は設計の目的のために何をすればよいかは理解しているが、シミュレーションが何を出力するか、その詳細までは必ずしも見えていない。そこで、CAEシステム等が具備するポスト・プロセッサや可視化ソフトウェアにより、シミュレーション結果の確認をするだけで終わらせないために、シミュレーションが紡ぎ出すデータをくまなく分析・評価するためのシステムとして、Cerebral Methodology in Data Analysis Systemを提案している^{8, 9)}。

これは、一つに当該設計案に対して、想定外や思いも知れない問題や過剰設計箇所を確認できるようにするための研究でもある。このシステムでは、演繹的データ解析と帰納的データ解析を図1に示すように分散並列計算で実現している。演繹的データ解析処理部では、規定、規程、科学則、工学則、技術則、設計則、経験則等からシミュレーションが紡ぎ出すデータを統計的技法により分析する。帰納的データ解析処理部では、枚举的帰納、類推(Analogy)、仮説形成(abduction)からシミュレーションが紡ぎ出すデータを分析する。それぞれに作図可視化解析と作画可視化解析処理部を設けているが、これはCAEシステム等が具備するポスト・プロセッサ機能と同等のものであり、シミュレーションが紡ぎ出した

図1 Cerebral Methodology for Data Analysis System

図2 四辺形有限要素の等高線模式図

データそのものを可視化し、シミュレーションが紡ぎ出したデータの分析結果を視覚的に提示し、ドキュメント化するものである。作図可視化解析処理部では、情報可視化(グラフ等の作画)を行う。作画可視化解析処理部では、画像・映像可視化、可視化画像の画像解析結果の可視化を行う。Synthesis処理部は、シミュレーション結果の設計データモデルを駆使して、知見の統合化をする。そして分析結果の帰結を求めるために、Cerebral Methodology for Data Analysisでは演繹的な分析と帰納的な分析の両方結果を統合(Synthesis)して、一つの結論を導く。

本システムの一つの特長であるニューラルネットワークを用いたアブダクション解析は、濃度特長処理による画像データのパターンマッチングにより実現している。有限要素法を用いた構造解析の結果得られた物理量、例えば応力値を等高線で描画した場合、一つの有限要素における等高線図は図2に示すような、濃淡分布に分類できる。濃度特長処理では、図2に示すように濃淡が一つの四辺形有限要素内で変化するマスクパターンを用意しておき、実際の四辺形要素の等高線図画像における標本点の色を観測し、四辺形要素内の全ての標本点の濃度の総和を濃度特長として出力する。このとき、四辺形要素内を $n \times n$ の格子状に分割した小領域をそれぞれ標本点とすれば、 n の数が多いほど精緻な特長抽出となる。

具体的には、四辺形要素内等高線パターンを $S(u, v)$ 、マスクパターンを $MP(i, j)$ とし、標本点の座標を (u_p, v_p) としたとき、その標本点での濃度 $C(u_p, v_p)$ は、
$$C = \sum \sum MP(i, j) \cdot S(u_p + i, v_p + j)$$
である。

このようなやり方で、時刻歴変化のある可視化した画像の類似性を比較し、四辺形有限要素の画像を個々にパターンマッチングしながら状態を推論する。これは数値でもできることだろうとの反論もあるが、数値で行う場合には同じとみなす閾値の設定に恣意的な操作が加わ

図3 応力分布の画像比較模式図

る。ニューラルネットワークを用いた解析では、閾値と言えそうではあるが、ある程度、こんな感じだという曖昧な推論をしていくことが可能となる。これにより、図3の(1)と(2)に示すような実線と破線で示した等高線図は、それらに違いはあるものの図3の(3)のように人間の目で見れば概ね一致しているという推定が可能となる。すなわち、シミュレーションの結果データから判断論理を導出し、仮説形成をしながらデータ解析を行うことになる。例えば、50秒間の時刻歴変化のあるシミュレーションが0.1秒刻みで計算結果を出力した500ステップ分の状態図を当該機能で分析した結果、図3の右手方向に応力負荷がかかっているということがミーゼス相当応力の分布からわかり、これを仮説として、剪断力の分布などを同様な手段で分析していくことにより、帰納的帰結として図3の右手方向の耐力要件を評価することになる。このように人工物工学的な手段により、本システムは帰納的分析や帰納的帰結、すなわちアブダクションに重きをおいた分析を実現している。

本システムのもう一つの特長であるデータを活用したどの視点も取り入れて計算結果を吟味する機能は、演繹的な分析と帰納的な分析の両方結果を統合(Synthesis)する機能により具現化している。ここでいう統合とは、演繹的な分析結果と帰納的な分析結果という知を比較分析した結果を、本システムが有するデータモデルから伝達される知により判断する過程である。図3を例にとれば、演繹分析では、過大応力は見当たらないと判定するも帰納分析では応力の分散には偏りがあり、しかも特定の部位にのみ集中する傾向がある、となる。この二つの分析結果を比較することにより、この設計対象が必ずしも均一の断面をもった構造である必要はないというような総合知を獲得するのである。

IV. おわりに

多くの設計工学や設計科学に従事された方々は、経験

則や技術者の勘をきちんと体系化すべきであると主張してきている。また企業においてもノウハウの蓄積は重要とされ、継承されてきている。こういった経験則や特性値等を活用して、分析や照合を行うことの必要性はよく認知されている。しかし設計上、絶対はない。すなわち、モレやヌケは人間が考える以上あり得るし、コストや目的の限定からくる仕様外の議論はついてまわる。その意味では、エキスパートシステムのような演繹的手段のみでは人工物工学的な分析は不十分である。一方で、PL法に代表されるように製造側の責任は、消費者側によって訴求されるものでもある。想定外を排除していくためには、利用できるもの、例えば人工知能や計算科学をどんどん利用して、人間の英知を設計の中で高めていくべきと考える。人工物工学が重要と考えるあらゆる視点からの行為、例えば観察という行為を考えるとアブダクションを用いた仮説形成による分析は必要不可欠である。必要なデータだけを分析しているとすれば、シミュレーション活用として「もったいない」のである。

— 参考資料 —

- 1) 藤田豊久, 太田順(編), 人工物工学入門, 東京大学出版会, (2015).
- 2) Neumann, J.V. (高橋秀俊監訳), 自己増殖オートマトンの理論, 岩波書店, (1975).
- 3) Farhat, N.H., Psaltis, D., Prata, A. and Paek, E., *Optical implementation of the Hopfield model*, *Applied Optics*, Vol.24, No.10, pp.1469-1475, (1985).
- 4) Takeda, M. and Goodman, J.W., *Neural networks for computation : number representations and programming complexity*, *Applied Optics*, Vol.25, No.18, pp.3033-3046, (1986).
- 5) 沖野教郎, 生物型生産システム—集中から分散へ, 朝倉書店, (1993).
- 6) Minsky, M. (安西 祐一郎訳), 心の社会, 産業図書, (1990).
- 7) <https://www.top500.org/lists/2016/11/>
- 8) 木野千晶, 鈴木喜雄, 西田明美, 櫛田慶幸, 林幸子, 中島憲宏, 認識能力を備えたデータ解析システムの概念設計, 日本計算工学会論文集, 2008(P200818), pp.1-8, (2008).
- 9) 木野千晶, 鈴木喜雄, 宮村浩子, 武宮博, 中島憲宏, 物理的意味の認識を可能とするデータ解析システムのための科学概念語彙モデルの提案, 日本計算工学会論文集, 2009(P200922), pp.1-8, (2009).

著者紹介

中島憲宏 (なかじま・のりひろ)

日本原子力研究開発機構 システム計算科学センター, 東京大学 人工物工学研究センター

(専門分野/関心分野) 設計工学, 構造解析, 計算科学, 可視化, 認知科学