フォーラム

計算機による法則式発見への挑戦

鷲尾 隆 元田 浩

1 はじめに

法則発見は，人間の歴史においてもっとも知的な営みの1つである．経験的に得られるデータに潜み，かつ我々にとって何らかの意味を有する規則性の中で，更に幾つかの厳しい条件を満たすものを法則という．およそ過去20年に亘り科学的知識発見の研究分野において，法則発見の営みを計算機に実装して自動化，ないしは半自動化する試みがなされてきた．これらの多くが実験環境において得られる定量的データから数量間の法則式を発見するものである．法則式発見で重要な点は，対象分野やその観測過程に関する先験的知識に基づいて，対象を支配する原理が観測数量間に課す制約と同型ないしは準同型である関係式を探索することである．これは，従来のシステム同定や統計的手法が観測数量間の漸近式を見いだそうとすることと異なる点である．本論では数理的かつ人工知能的原理により，実験室のように条件変更操作を行いつつ能動的に種々の状態を測定可能である場合を対象とする法則式発見手法について述べる．これ以外のアプローチや条件に関しても多くの研究がなされているが，紙面の都合上割愛する[6],[10].

2 連立方程式の構造発見

対象が単一の法則式で表されることが予め知られていることは少ない．従って多くの場合には各法則式発見に先立ち，まず対象が法則式の集合である連立方程式で表されると仮定し，何本の方程式で表され，かつ何れの数量が何れの式に現れるかを決定しなければならない．この際，同じ対象挙動を表す連立方程式表現は一般にユニークではなく，多数の同値表現が存在することを考慮する必要がある．そこで，表現に依存しない連立方程式の不変構造を導出するために，“構造形”と“完全部分集合”という概念を用いることが必要となる[12]．ここでは，**図1**に示される簡単な電子回路の例を用いて説明を行う．式(1)はこの電子回路の1つの連立方程式記述である．一方，同回路は式(2)によっても記述される．

$$\begin{aligned} &V_1 = I_1R_1[1], \quad V_2 = I_2R_2[2], \\ &V_e = V_1[3] \text{ and } V_e = V_2[4], \end{aligned} \tag{1}$$

$$\begin{aligned} &I_1R_1 = I_2R_2[1], \quad V_2 = I_2R_2[2], \\ &V_e = V_1[3] \text{ and } V_e = V_2[4], \end{aligned} \tag{2}$$

両者は定量的には等価であるにもかかわらず，連立方程式の接続行列表現を考察すると明らかになる構造上の大きな違いがある．接続行列 T は何れの方程式に何れの数量が現れているかを0, 1の要素で示した行列である．式(1), (2)に対応する接続行列は，それぞれ以下の通りである．

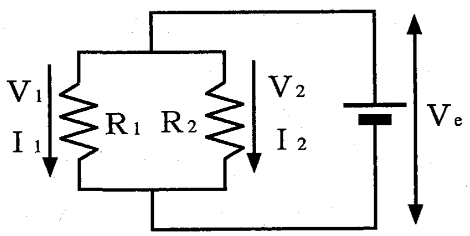

図1 並列抵抗回路

わしお たかし．大阪大学産業科学研究所．
もとだ ひろし．大阪大学産業科学研究所．

$$
T_1 = \begin{array}{c} \begin{array}{ccccccc} V_e & V_1 & V_2 & I_1 & I_2 & R_1 & R_2 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array} \quad (3)
$$

$$
T_2 = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (4)
$$

式(1), (2)のいずれにおいても，もし V_e と R_1 の値が決まれば V_1, V_2, I_1 の値も決定されるが，I_2, R_2 の値は未決定のままである．このことは実際の式を見るまでもなく，T_1 の 0, 1 構造を見るだけで明らかである．即ち T_1 の第 1 及び第 6 列に対応する V_e と R_1 が決定されれば，第 3 及び第 4 行(式に対応)に現れている未定数量 V_1, V_2 の値が 2×2 の制約によって決定され，更に第 1 行から I_1 が決まることが分かる．このように幾つか数量の値を外生的に決定してやると他の数量の値も決まるような数量部分集合を“完全部分集合”という．この例では $\{V_e, V_1, V_2, I_1, R_1\}$ がこの系の 1 つの完全部分集合となる．一方，T_2 では第 1 行の 0, 1 構造が異なり，この行列上では I_1 が決定されることが分からない．しかしながら，元の式を定量的に解けば I_1 が決定されることは明らかであり，T_2 は系における数量間の依存関係を正確に反映していないことが分かる．これに対して T_1 は数量の値をどのように外生的に決定しても，定量的に決まる数量の組を正確に導くことが可能である．このように接続行列で構造だけを抜き出しても，系内の数量の依存構造を正確に表す連立方程式を“構造形”と呼ぶ．詳細は省略するが，連立方程式における一連の完全部分集合は“モジュラ束”をなし，しかもそれによって系の不変な依存構造は一意な“正準系”として表すことができる．

このような連立方程式が有する依存構造の数学的性質を利用して，その構造を発見する実験手順を定めることができる．鷲尾等はこの原理を用いて，実験室のような能動的観測条件の下で連立方程式の構造同定を行う Simuletaneous Structure Finder (SSF) を提案した[12]．尚，連立方程式の構造同定に関しては，他にマトロイドによる研究が行われている[7]．マトロイドは一般に少なくとも局所的に制約充足ないしは制約過剰な連立方程式構造に関する理論である．一方，上述の原理は何れの部分も制約不足ないしは制約充足な連立方程式の構造同定を行う理論である点が異なる．

3 実験データと一般的先見知識を用いる法則式発見

連立方程式構造を発見後，各々 1 本ずつの完全方程式について法則式関係を同定する．能動的な観測条件下における代表的な方法は“2 変項当てはめ”である[2], [6]．1 つの完全方程式を構成すると考えられる数量集合の中で，少なくとも何れか一方が実験的に制御可能な数量組を選び，両数量間の関係を表すデータを，これら 2 数量以外の数量値を固定した実験から採取する．そして，2 数量関係を表す式をこの実験データに当てはめる回帰分析を行う．更に回帰が統計的に有意であると判定された 2 数量関係式のパラメータを中間項とし，各数量や中間項の関係が 1 本の完全方程式にまとめ上げられるまで，以上の 2 変項当てはめを繰り返す．BACON はこの 2 変項当てはめに基づく法則式探索を行う代表的システムである[5]．ただし，この方法は法則式として尤もらしい形式を持ち，かつ見かけ上はデータに良く一致する式を導出するが，それが対象系を支配する法則式を表す保証はない．

この問題点を軽減すべく，一般的な先験的知識に基づいて数学的に許容される形式の範疇で，法則式を探索する手法が提案されている．もっとも代表的な手法は“単位次元解析”の“次元均一性”の原理を用いるものである．これは単位の異なる数量を足し引きすることは物理的に無意味であり法則式と見なすことはできないという制約である．これを用いうる法則発見システムとしては ABACUS が代表的である[2]．更に単位次元解析では，

質量や距離など大きさの比が意味を持つ数量同士の関係を強く限定する Product Theorem が知られており[1]，これを用いて法則式を探索する法則発見システムに COPER がある[4]．この制約は非常に強いので，流体の Bernooulli の法則など ある程度複雑な法則式を発見できる．

更により一般的な場合にも適用可能な手法として，“尺度”に関する知識を用いる手法が提案されている[10]．大半の数量は“比例尺度”と“間隔尺度”に分類される．比例尺度は絶対的原点を持ち，その比に意味がある数量である．前述の Product Theorem が対象とする数量はすべて比例尺度である．一方，間隔尺度は温度の摂氏や華氏などのように，原点が人為的に決められ値の差の比にしか意味の無い数量である．これら 2 種類の尺度を含む法則式の数学的許容形式について，以下の 2 定理が存在する[10]．

Theorem 3. 1 (Extended Buckingham II-theorem) $\phi(x_1, x_2, x_3, \cdots)=0$ が 1 つの完全方程式であり，かつその各引数が比例尺度ないしは間隔尺度の何れかであるならば，その式は以下の形式に書き換え可能である．

$$F(\Pi_1, \Pi_2, \cdots, \Pi_{n-w}) = 0, \tag{5}$$

ここで n は ϕ の引数の個数，w は $x_1, x_2, x_3, \cdots$ が有する測定基準自由度の総数である．測定基準とは単位目盛りの間隔や原点など，測定において人為的に設ける基準である．また，各中間項 Π_i は無次元量である．

Theorem 3. 2 (Extended Product Theorem) 幾つかの基礎単位を共有する比例尺度数量の集合 R，間隔尺度数量の集合 I について，中間項 Π を各 $x_i \in R \cup I$ と関係づける関数は以下のいずれかの形式を取る．

$$\Pi = \left(\prod_{x_i \in R} |x_i|^{a_i}\right)\left(\prod_{I_k \subseteq I}\left(\sum_{x_j \in I_k} b_{kj}|x_j| + c_k\right)^{a_k}\right) \tag{6}$$

$$\begin{aligned}\Pi = &\sum_{x_i \in R} a_i \log|x_i| \\ &+ \sum_{I_k \subseteq I} a_k \log\left(\sum_{x_j \in I_k} b_{kj}|x_j| + c_k\right) \\ &+ \sum_{x_l \in I_g \subseteq I} b_{gl}|x_l| + c_g,\end{aligned} \tag{7}$$

ここで，Π を除いてすべての係数は定数であり，$I_k \cap I_g = \phi$ である．

基礎単位とは[m], [sec], [kg]など，それ以上分解できない単位のことである．もしいずれかの数量が何らかの基礎単位を共有しているならば，それらはこれらの関係を満たす．例え基礎単位の共有が定かではなくとも，法則式中の数量の幾つかは必ず基礎単位を共有しているため，必ずこれらの関係が含まれる．従って，Extended Product Theorem に示される関係をデータ中で優先的に探索することで，探索範囲を絞り込みかつ解の健全性を向上できる．心理学や経済学を含む殆どの問題領域において数量の尺度がその測定過程の分析から明らかである場合が多いので，この原理は幅広い分野に適用可能である．

鷲尾等は，この原理に基づいて能動的観測条件の下で 1 本の完全方程式からなる法則式を発見するシステム Smart Discovery System (SDS) を構築した[10]．また，このシステムを第 2 節の最後に述べた SSF と組み合わせ，連立方程式で表される法則式を発見できるように拡張した．以下に固体物質平板に垂直に開いた 2 つの環状穴流路に関する流体の熱伝達及び凝縮現象に関する 17 数量を含む 8 本の連立方程式を示す．

$$\begin{aligned}&\omega = 0.9423\left(\frac{vsk^3}{L\mu}\right)^{1/4}, \quad \dot{H} = \dot{H}_1 + \dot{H}_2, \\ &\Delta T_1 = T_f - T_{w1}, \quad \Delta T_2 = T_f - T_{w2} \\ &h_1 = \Delta T_1^{-1/4}\omega, \quad h_2 = \Delta T_2^{-1/4}\omega, \\ &\dot{H}_1 = 2\pi\gamma L h_1 \Delta T_1, \quad \dot{H}_2 = 2\pi\gamma L h_1 \Delta T_2\end{aligned} \tag{8}$$

ここで v, s, k, μ は液相流体の単位体積当たりの潜熱，密度，熱伝導率，粘性である．また L は各流路の長さ，T_f は流体の温度，T_{w1}, T_{w2} は各

流路壁の温度，$\dot{H}$ は流体からの流路壁を通じた単位時間当たりの熱伝達量である．SSF+SDSはデータと各数量の尺度に関する情報のみから，これら8本からなる連立方程式モデルを正しく再発見した．

4 おわりに

法則式は各学問体系発展の歴史的経緯の中で発見・命名されてきており，慣習によって広く知られた経験的な関係であるという解釈が取られることも多い．例外なき公理化には無理があろうが，学問体系の最も基本的な道具である法則式の定義や成立条件を公理論的な立場から明確化しようという努力も大切である．Newtonは自然法則を発見する際の規範として，自然界の原因以外を排除する客観性，なるべく少数の原因のみを仮定する簡潔性，再試行に対する再現性，広範囲での成立を要請する普遍性，すべての試行結果に反しない健全性を要求した[8]．近年でも，Simonが法則記述の簡潔性の重要さを主張している[9]．更にFeynmanは法則形式の時間や空間に関する数学的許容性の重要性を述べている[3]．筆者等はこのような考察を踏まえ，法則式に要請される条件の体系化を試みた[11]．

能動的観測条件における法則式発見システムは，実験的に広範囲の対象状態に関するデータを追試も含めて採取して発見を試みるので，客観性はもちろんのこと，実験系に制約される面はあるものの普遍性，再現性，健全性に関して一定の条件の満たした法則式発見を行うことができる．また何れもがボトムアップに単純な式を優先的に探索するので，簡潔性の条件も満たされる．また，SDSのように数学的許容性の検証を行う方法も提案されている．しかしながら，法則式が満たすべき条件を考慮した法則式発見の研究はまだ端緒についたばかりであり，より完全かつ広範囲な条件を考慮する手法の探求が待たれるところである．

参考文献

[1] Bridgman, P. W., Dimensional Analysis, New Haven, Connecticut, Yale University Press, 1922.

[2] Falkenhainer, B. C., and Michalski R. S., Integrating Qualitative and Quantitative Discovery : The ABACUS System, Machine Learning, 1, Boston, Kluwer Academic Publishers, 1986, 367-401.

[3] Feynman, R. P., The Character of Physical Law, Charles E. Tuttle Co. Inc., 1965.

[4] Kokar, M. M., Determining Arguments of Invariant Functional Descriptions, Machine Learning, 1 (1986), 403-422.

[5] Langley, P. W., Bradshaw, G. L., and Simon, H. A., BACON. 5 : The Discovery of Conservation Laws, Proceedings of IJCAI-81 : Seventh International Conference on Artificial Intelligence, (1981), 121-126.

[6] Langley, P. W., Simon, H. A., Bradshaw G., and Zytkow, J. M., Scientific Discovery; Computational Explorations of the Creative Process, Cambridge, Massachusetts, MIT Press, 1985.

[7] Murota, K., Systems Analysis by Graphs and Matroids-Structural Solvability and Controllability, Algorithms and Combinatorics, Vol. 3, Berlin, Springer-Verlag, 1987.

[8] Newton, I., Principia Vol. II The System of the World, 1686. (Translated into English by Motte, A., London, England, University of California Press, Ltd, 1729. (Copyright 1962))

[9] Simon, H. A., Models of Discovery, Dordrecht, Holland, D. Reidel Publishing Company, 1977.

[10] Washio, T., and Motoda, H., Discovering Admissible Models of Complex Systems Based on Scale-Types and Identity Constraints, Proceedings of IJCAI-97 : the Fifteenth International Joint Conference on Artificial Intelligence, (1997), 810-817.

[11] 鷲尾隆，元田浩，属性変量の尺度認知に基づく構成的法則発見手法，認知科学，5, 2 (1998), 80-94.

[12] Washio, T., and Motoda, H., Discovering Admissible Simultaneous Equations of Large Scale Systems, Proceedings of AAAI-98 : Fifteenth National Conference on Artificial Intelligence, (1998), 189-196.