

コネクショニストアプローチによる数法則の発見

斉藤 和巳[†] 中野 良平[†]

科学的発見を支援するシステムにおいて、データからの数法則の発見は中心的な課題である。本論文では、コネクショニストアプローチに基づき、数法則を発見する新方法 RF5 を提案する。RF5 では、数法則の発見問題をニューラルネットの学習問題として定式化し、そのニューラルネットの学習において、つねに効率良く望ましい結果（法則候補）を得るため、2 次の新学習アルゴリズム BPQ を採用する。BPQ は準ニュートン法を基本枠組みとし、最適ステップ幅を 2 次近似の最小点として求めることを特徴とする。また、RF5 では、MDL 基準を用いて、学習結果の法則候補から適切な法則を選択する。記号処理、または、コネクショニストアプローチに基づく既存法と比較して、RF5 の最大の特長は、指数の値が整数に制限されない法則を効率良く発見できることである。実験では、ある程度のノイズを含むデータからでも、このような法則を効率良く発見できることを示す。また、既存のニューラルネットの学習アルゴリズムと比較して、BPQ の探索効率と収束能力は大幅に優れることを示す。最後に、現実問題での RF5 の有効性も評価する。

A Connectionist Approach to Numeric Law Discovery

KAZUMI SAITO[†] and RYOHEI NAKANO[†]

The discovery of an underlying law from a set of numeric data is the central part of scientific discovery systems. This paper proposes a new connectionist approach to numeric law discovery. In order to efficiently and constantly obtain near-optimal results (law-candidates), we introduce a new second-order learning algorithm; by adopting a quasi-Newton method as a basic framework, the optimal step-lengths are calculated as the minimal points of second-order approximations. The Minimum Description Length criterion selects the most suitable from law-candidates. The main advantage of our method over previous work of symbolic or connectionist approach is that it can efficiently discover numeric laws whose power values are not restricted to integers. Experiments showed that the proposed method works well in discovering such laws even from data containing a small amount of noise.

1. はじめに

科学的発見を支援するシステムにおいて、データからの数法則の発見は中心的な課題である。このようなシステムでは、たとえば、惑星と太陽の距離 r と惑星の公転周期 T から、ケプラーの第 3 法則 $T = kr^{3/2}$ を発見できる (k は定数)。

BACON システム^{5),6)}での先駆的な研究の後、いくつかの方法^{3),7),9),13),15)}が提案されている。ただし、これらの基本的な探索戦略はほとんど同じである。つまり、乗算、除算、または、あらかじめ定義した関数を用いて、2 つの既存変数を組み合わせ、新たな変数を再帰的に生成する。BACON と FAHRENHEIT⁷⁾ は trend detector と呼ばれる尺度で変数を組み合わせ、

ヒューリスティクスを加えた深さ優先探索を行う。ABACUS³⁾は proportional graph を生成し、ビーム探索を行う。IDS⁹⁾では、相関分析が適用され、ビーム探索を行う。E*アルゴリズム¹³⁾は 2 変数の法則のみを対象とする。また、Sutton-Matheus アルゴリズム¹⁵⁾は回帰を行い、自乗誤差と変数値の自乗についての相関¹²⁾を用いて変数を組み合わせる。

これらの既存法に対して、以下の課題が指摘できる。第 1 に、2 つの変数を順番に組み合わせて新たな変数を作るので、多くの変数からなるデータにおいて複雑な法則を探索すれば、容易に組合せ爆発が起き、また、探索パラメータが適切でなければ、望ましい法則を発見できないことが予想される。第 2 に、法則に現れる指数の値が整数ではないとき、適当な関数 (e.g., $r^{3/2}$) をあらかじめ定義しなければ、法則の発見は困難になる。しかし、多くの場合、事前知識はない。第 3 に、現実の観測データは確実にノイズを含むが、既存

[†] NTT コミュニケーション科学研究所

NTT Communication Science Laboratories

法は比較的ノイズに弱いことが指摘されている^{9),13)}.

コネクショニストアプローチは上記の課題解決に有望であると考えられる。指数の値が整数に制限されない一般化した多項式の各項を直接学習するには、入力値の重み付け和の代わりに、入力値を結合重みで累乗した値の積を計算する product unit と呼ばれる計算ユニットが提案されている²⁾。しかし、標準的な BP (Back Propagation) アルゴリズム¹¹⁾では、このタイプのユニットを含むネットワークの学習はきわめて困難であることが報告されている⁸⁾。これに対して、複数の学習アルゴリズムを組み合わせる方法⁸⁾などが提案されているが、BP と比較して、それらの有効性はあまり顕著ではない。また、これらの初期の研究では、2 値データのみを扱い、数法則の発見問題を対象としていない。

2 次非線形最適化法である準ニュートン法は、BP のような 1 次の学習法では困難な問題でも、効率良く学習できることが期待される⁴⁾。しかし、既存の多くのアプローチでは、各反復において最適ステップ幅を求めるとき、計算コストが高い直線探索法⁴⁾を採用しているので、全体として多くの計算量が必要となる。しかるに、最適ステップ幅を妥当な精度で効率良く求めることができれば、準ニュートン法は収束性と効率性の両面で優れた方法になる。

本論文では、コネクショニストアプローチにより数法則を発見する方法 RF5 を提案する。2 章では、数法則の発見問題をニューラルネットの学習問題として定式化する。3 章では、そのニューラルネットを学習するためのアルゴリズム BPQ について述べる。4 章では、複数の学習結果から適切なものを選択する評価尺度について説明する。5 章では、人工問題と現実問題による実験を行い、RF5 の有効性を評価する。

2. ニューラルネットを用いた数法則の発見

本章では、数法則の発見をコネクショニストモデル²⁾を用いて定式化する。 $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_m, y_m)\}$ を事例集合、 \mathbf{x}_t を n 次元入力ベクトル、 y_t を \mathbf{x}_t に対する目標出力値とする。本稿では、

$$y_t = c_0 + \sum_{i=1}^h c_i x_{t1}^{w_{i1}} \cdots x_{tn}^{w_{in}} \quad (1)$$

で表される数法則のクラスについて考える^{*}。ここで、

各パラメータ c_i , w_{ij} は未知の実数、 h は未知の整数である。以下では、 $(c_0, \dots, c_h)^T$, $(w_{i1}, \dots, w_{in})^T$ をそれぞれ \mathbf{c} , \mathbf{w}_i と表記する。ただし、 \mathbf{a}^T は \mathbf{a} の転置を意味する。また、すべてのパラメータからなる 1 つのベクトル $(\mathbf{c}^T, \mathbf{w}_1^T, \dots, \mathbf{w}_h^T)^T$ を Φ で表し、 $N (= nh + h + 1)$ を Φ の次元 (パラメータ数) とする。

必要ならば適当な値を各入力ベクトルの要素に加えることにより、 $x_{ti} > 0$ を仮定できるので、式 (1) は

$$y_t = c_0 + \sum_{i=1}^h c_i \exp \left(\sum_{j=1}^n w_{ij} \ln(x_{tj}) \right) \quad (2)$$

と等価である。式 (2) は各中間ユニットの活性化関数が $\exp(s) = e^s$ である 3 層ニューラルネットと見なすことができる。つまり、 h , \mathbf{w}_i , および、 \mathbf{c} はそれぞれ中間ユニット数、全入力ユニットと中間ユニット i との結合重み、および、全中間ユニットと出力ユニットとの結合重みである。以下では、中間ユニット i の出力値を $v_{it} = v_i(\mathbf{x}_t; \mathbf{w}_i) = \exp \left(\sum_{j=1}^n w_{ij} \ln(x_{tj}) \right)$ 、出力ユニットの出力値を $z_t = z(\mathbf{x}_t; \Phi) = c_0 + \sum_{i=1}^h c_i v_{it}(\mathbf{x}_t; \mathbf{w}_i)$ で表す。なお、このタイプの中間ユニットは product unit²⁾ と呼ばれる。よって、式 (1) を対象とする数法則の発見問題は

$$f(\Phi) = \frac{1}{2} \sum_{t=1}^m (y_t - z(\mathbf{x}_t; \Phi))^2 \quad (3)$$

を最小化する Φ を求めるニューラルネットの学習問題として定式化できる。

3. BPQ アルゴリズム

初期の研究⁸⁾で報告されたように、我々の実験でも、標準的な BP による式 (3) の最小化はきわめて困難であった。そこで、つねに効率良く望ましい結果を得るため、本稿では、新たな 2 次の学習アルゴリズム BPQ (Back Propagation based on Quasi-Newton) を提案する。BPQ は準ニュートン法を基本枠組みとし、最適ステップ幅を 2 次近似の最小点として計算する。

3.1 準ニュートン法

準ニュートン法⁴⁾は、多変数関数の極値を求める方法であり、探索の過程で反復により 2 次微分の逆行列 $(\nabla^2 f(\Phi))^{-1}$ の近似値 (\mathbf{H}) を各ステップで求めることを特徴とする。基本アルゴリズムは以下である：

^{*} 対象とする法則が周期関数や不連続関数からなる場合、式 (1) では厳密には対処できない。しかし、このような関数でも、入力ベクトル \mathbf{x} のレンジが限られている場合には、有限項数の多項式を用いて、ある程度の精度で近似可能であり、さらに、実数指

数の多項式ならば、各項の表現能力が向上するので、少ない項数での近似が期待できる。

一方、式 (1) で c_0 を除いた定数項なしモデルも考えられるが、本稿の議論を簡潔にするため式 (1) のみを対象とする。

- step 1:** Φ_1 を初期化し, $\mathbf{H}_1 = \mathbf{I}$, $k = 1$ とする.
step 2: 探索方向を求める: $\Delta\Phi_k = -\mathbf{H}_k \nabla f(\Phi_k)$.
step 3: 停止条件を満たせば, 反復を終了させる.
step 4: $f(\Phi_k + \lambda\Delta\Phi_k)$ を最小にする λ_k を求める.
step 5: 結合重みを修正する: $\Phi_{k+1} = \Phi_k + \lambda_k \Delta\Phi_k$.
step 6: $k \equiv 0 \pmod{N}$ ならば, $\mathbf{H}_{k+1} = \mathbf{I}$ とし,
 さもないければ, \mathbf{H}_{k+1} を更新する.
step 7: $k = k + 1$ とし, **step 2** に戻る.

ただし, \mathbf{I} は単位行列を表す. **step 6** での \mathbf{H}_{k+1} の計算法にはいくつかの提案があるが, 制約のない非線形最適化では, 最も優れた方法であると指摘される BFGS (Broydon-Fletcher-Goldfarb-Shanno) 公式⁴⁾を採用する. すなわち, $\mathbf{p} = \lambda_k \Delta\Phi_k$, $\mathbf{q} = \nabla f(\Phi_{k+1}) - \nabla f(\Phi_k)$ とおけば, BFGS 公式は以下で与えられる:

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \left(1 + \frac{\mathbf{q}^T \mathbf{H}_k \mathbf{q}}{\mathbf{p}^T \mathbf{q}}\right) \frac{\mathbf{p} \mathbf{p}^T}{\mathbf{p}^T \mathbf{q}} - \frac{\mathbf{p} \mathbf{q}^T \mathbf{H}_k + \mathbf{H}_k \mathbf{q} \mathbf{p}^T}{\mathbf{p}^T \mathbf{q}}. \quad (4)$$

3.2 最適ステップ幅の新計算法

準ニュートン法の効率は **step 4** で採用する最適ステップ幅 λ の計算法に大きく依存する. 以下では, 効率の良い新計算法を導出する.

3.2.1 基本計算法

step 4 では, 変数は λ だけなので, $f(\Phi + \lambda\Delta\Phi)$ を単に $g(\lambda)$ で表す. $g(\lambda)$ の 2 次近似式は

$$g(\lambda) \approx g(0) + g'(0)\lambda + \frac{1}{2}g''(0)\lambda^2 \quad (5)$$

であり, $g'(0) < 0$ かつ $g''(0) > 0$ のとき, 式 (5) の右辺の最小値は

$$\lambda = -\frac{g'(0)}{g''(0)} \quad (6)$$

で与えられる. ほかのケースについては, 次の節で扱う.

式 (3) で定義したニューラルネットでは, $g'(0)$ と $g''(0)$ を効率良く計算できる. $g(\lambda)$ を微分し, λ に 0 を代入すれば,

$$g'(0) = -\sum_{t=1}^m (y_t - z_t) z'_t,$$

$$g''(0) = \sum_{t=1}^m ((z'_t)^2 - (y_t - z_t) z''_t)$$

となる. ここで, $z_t = z(\mathbf{x}_t; \Phi)$ の微分は $(\frac{d}{d\lambda} z(\mathbf{x}_t; \Phi + \lambda\Delta\Phi))_{(\lambda=0)}$ で定義され,

$$z'_t = \Delta c_0 + \sum_{i=1}^h (\Delta c_i v_{ti} + c_i v'_{ti}),$$

$$z''_t = \sum_{i=1}^h (2\Delta c_i v'_{ti} + c_i v''_{ti})$$

となる. ここで, $v'_{ti} = v_{ti} \times \sum_{j=1}^n \Delta w_{ij} \ln(x_{tj})$, $v''_{ti} = v'_{ti} \times \sum_{j=1}^n \Delta w_{ij} \ln(x_{tj})$ であり, Δc_i , Δw_{ij} は **step 2** で計算される c_i , w_{ij} の変化量である.

3.2.2 望ましくないケースへの対処

上述のケースでは, $g'(0) < 0$ を仮定した. $g'(0) > 0$ のとき, その探索方向で目的関数の値を減少できないので, $\Delta\Phi_k = -\nabla f(\Phi_k)$, $\mathbf{H}_k = \mathbf{I}$ に設定する. このとき, $g'(0) = (\nabla f(\Phi_k))^T \Delta\Phi_k = -\|\nabla f(\Phi_k)\|^2 < 0$ より, $g'(0) < 0$ が保証される.

$g'(0) < 0$ かつ $g''(0) \leq 0$ のとき, 式 (6) の値は負または無限大となるので, ガウス-ニュートン法を用いて, このケースに対処する. $z(\mathbf{x}_t; \Phi + \lambda\Delta\Phi)$ の 1 次近似は $z_t + z'_t \lambda$ となるので, $g(\lambda)$ の近似は

$$g(\lambda) \approx \frac{1}{2} \sum_{t=1}^m (y_t - (z_t + z'_t \lambda))^2$$

$$= g(0) + g'(0)\lambda + \frac{1}{2} \sum_{t=1}^m (z'_t)^2 \lambda^2 \quad (7)$$

であり, 式 (7) の右辺の最小値は

$$\lambda = -\frac{g'(0)}{\sum_{t=1}^m (z'_t)^2}. \quad (8)$$

で与えられる. 明らかに, $g'(0) < 0$ のとき, 式 (8) の値は正となる. このケースでは, 探索は鞍点の近くかもしれないので, $k = N$ とし, \mathbf{H} を単位行列に戻す.

多くの場合, 各反復で Φ の修正量の上限を設定することは有効である⁴⁾. よって, $\|\lambda\Delta\Phi\| > 1.0$ ならば, λ を $1.0/\|\Delta\Phi\|$ とする. このケースでは, 目的関数の形状がかなり変化するかもしれないので, $k = N$ とし, \mathbf{H} を単位行列に戻す.

λ は近似に基づき計算されるので, 目的関数 $g(\lambda)$ の値が つねに減少するとは限らない. $g(\lambda) \geq g(0)$ のときは, λ を縮める処理を導入する. λ_0 を式 (6) または (8) で計算した値とすれば, 条件 $h(0) = g(0)$, $h(\lambda_0) = g(\lambda_0)$, $h'(0) = g'(0)$ を満たす 2 次近似式 $h(\lambda)$ は

$$g(\lambda) \approx h(\lambda) = g(0) + g'(0)\lambda + \frac{g(\lambda_0) - g(0) - g'(0)\lambda_0}{\lambda_0^2} \lambda^2 \quad (9)$$

となる. $g(\lambda_0) > g(0)$ かつ $g'(0) < 0$ より, $h(\lambda)$ の最小点は

$$\lambda = \frac{-g'(0)\lambda_0^2}{2(g(\lambda_0) - g(0) - g'(0)\lambda_0)} \quad (10)$$

で与えられる。ここで、式(10)では $0 < \lambda < \lambda_0$ が保証される。よって、この処理を $g(\lambda) < g(0)$ となるまで繰り返せば、 $g(\lambda) < g(0)$ となる λ をつねに求めることができる。

3.3 計算量の考察

すべての訓練事例をそれぞれ1回利用する1反復の計算量について考察する。BPでの計算量(乗算の回数)は、目的関数の値を求めるのに $nhm + O(hm)$ 、勾配ベクトルでは $nhm + O(hm)$ が必要となる。よって、 $N = nh + h + 1$ より、BPの全体の計算量は $2Nm + O(hm)$ である。

上述に加え、BPQでは、**step 2**で探索方向を求めるのに N^2 、**step 6**で \mathbf{H} を更新するのに $O(N^2)$ 、**step 4**で最適ステップ幅を計算するのに $Nm + O(hm)$ (3.2.1項を参照)の計算量が必要である。しかるに、BPQで目的関数の値を求める計算量は $Nm + O(hm)$ から $O(hm)$ に減少する。なぜなら、各中間ユニットの出力値は $\exp(\sum_{j=1}^n w_{ij} \ln(x_{tj}) + \lambda \sum_{j=1}^n \Delta w_{ij} \ln(x_{tj}))$ であるが、直前の反復での最適ステップ幅計算において、 $\sum_{j=1}^n \Delta w_{ij} \ln(x_{tj})$ はすでに計算されているからである。よって、BPQの全体の計算量は $2Nm + O(N^2) + O(hm)$ である。

未知の事例に対する汎化誤差を小さくするには、一般に、 m は N と比較して、ある程度大きくなければならない。PAC-学習理論¹⁶⁾に基づけば、汎化誤差の上限を ϵ 以下とするのに必要な事例数は、概算で $\epsilon^{-1}N$ 以上である¹⁾。よって、このケースでは、 $2Nm$ と比較して、計算量 $O(N^2)$ はかなり小さくなるので、BPQの計算量のオーダーはBPとほぼ等価になる。

4. 結果選択のための評価尺度

一般に、与えられたデータ集合に対して、最適な中間ユニット数あらかじめ知ることはできない。また、データは普通ノイズを含むので、式(3)を最小にする法則候補がベストとは限らない。よって、中間ユニット数を変えて発見した法則候補を適切に評価するための尺度が必要である。ここでは、ノイズを想定して、目標出力値とニューラルネットの出力値の差 $y - z$ が平均0、標準偏差 σ の正規分布に従うと仮定する。すなわち、その対数尤度は

$$\log p(\mathbf{x}, y; \Phi, \sigma) = -\log \sigma - \frac{(y - z(\mathbf{x}; \Phi))^2}{2\sigma^2} + r \quad (11)$$

で定義する。ただし、正規化定数 r は $-(1/2)\log(2\pi)$ である。よって、最適な中間ユニット数を求めることは、最尤推定におけるモデル選択問題となるので、その評価尺度としてMDL (Minimum Description Length) 尺度¹⁰⁾を採用できる。

事例集合に対する負の対数尤度は

$$\begin{aligned} & -\sum_{t=1}^m \log p(\mathbf{x}_t, y_t; \Phi, \sigma) \\ & = m \log \sigma + \sum_{t=1}^m \frac{(y_t - z(\mathbf{x}_t; \Phi))^2}{2\sigma^2} - mr \end{aligned} \quad (12)$$

である。ここで、 m は事例数である。式(12)の最小化を考えれば、 Φ については、式(3)の最小化と等価であり、ニューラルネットの学習で最尤推定量 $\hat{\Phi}$ を得ることができる。一方、 σ については、式(12)を σ で微分して0とおけば、

$$\frac{m}{\sigma} - \frac{1}{\sigma^3} \sum_{t=1}^m (y_t - z(\mathbf{x}_t; \hat{\Phi}))^2 = 0$$

であり、よって、分散の最尤推定量 $\hat{\sigma}^2$ は

$$\hat{\sigma}^2 = \frac{1}{m} \sum_{t=1}^m (y_t - z(\mathbf{x}_t; \hat{\Phi}))^2 \quad (13)$$

となり、この $\hat{\sigma}^2$ の値は平均自乗誤差 (MSE: Mean Squared Error) にほかならない。式(13)を式(12)に代入すれば、式(12)の第2項は事例数にのみ依存する値 ($m/2$) となり、第2項と第3項は各モデルで等しいので、本稿では、次のMDL値を評価尺度として採用する。

$$\text{MDL} = 0.5m \log(\text{MSE}) + 0.5N \log(m) \quad (14)$$

ここで、 N は Φ の総パラメータ数である。

以下では、式(3)を目的関数とし、BPQアルゴリズムとMDL尺度を用いた数法則発見法を *RF5 (Rule extraction from Facts version 5)* と呼ぶ。

5. 実験

5.1 人工データによる評価

Sutton-Matheusの問題¹⁵⁾とその修正問題を用いて、数法則発見法RF5を評価した。オリジナル問題は

$$y = 2 + 3x_1x_2 + 4x_3x_4x_5 \quad (15)$$

を復元する問題である。事例については、 x_1, \dots, x_5 の各変数に $[0,1]$ の範囲でランダムな値を与え、対応する y の値を式(15)より計算する。ただし、変数の総数は9 ($n=9$) であり、不要変数 x_6, \dots, x_9 にも $[0,1]$ の範囲でランダムな値を与え、合計200 ($m=200$)

表1 オリジナル問題
Table 1 Original problem.

ユニット数	MSE 値			MDL 値			反復回数		処理時間	
	best	avg.	s.d.	best	avg.	s.d.	avg.	s.d.	avg.	s.d.
1	0.126	0.13	0.0	-177.9	-178	0.0	71	1.0	0.32	0.01
2	0.000	0.0	0.0	-1786.4	-1786	0.0	81	8.4	0.68	0.07
3	0.000	0.0	0.0	-1759.9	-1717	64	130	29	1.65	0.36

$$y = 2.000 + 3.000x_1^{1.000}x_2^{1.000} + 4.000x_3^{1.000}x_4^{1.000}x_5^{1.000}$$

図1 オリジナル問題で発見した法則
Fig. 1 Discovered law for original problem.

表2 修正問題
Table 2 Modified problem.

ユニット数	MSE 値			MDL 値			反復回数		処理時間	
	best	avg.	s.d.	best	avg.	s.d.	avg.	s.d.	avg.	s.d.
1	1.317	1.32	0.00	56.7	57	0.0	70	5	0.32	0.02
2	0.000	0.03	0.17	-1786.4	-1731	314	116	27	0.97	0.23
3	0.000	0.00	0.00	-1760.0	-1727	47	240	104	3.02	1.31

$$y = 2.000 + 3.000x_1^{-1.000}x_2^{3.000} + 4.000x_3^{1.000}x_4^{0.500}x_5^{-0.333}$$

図2 修正問題で発見した法則
Fig. 2 Discovered law for modified problem.

事例を生成した。実験では、各結合重みの初期値は、平均0、標準偏差0.1の正規分布に基づいて独立に生成し、MSE値が十分小さいとき、

$$\frac{1}{m} \sum_{t=1}^m (y_t - z(\mathbf{x}_t; \Phi))^2 < 10^{-8},$$

勾配ベクトルの大きさが十分小さいとき、

$$\frac{1}{N} \|\nabla f(\Phi)\|^2 < 10^{-8},$$

または、CPU処理時間が100秒を越えたときに、アルゴリズムの反復を終了させた。

5.1.1 オリジナル問題

実験では、中間ユニット数を1から3まで変化させ ($h = 1, 2, 3$)、それぞれ100回の試行を行った。実験結果のMSE値、MDL値、反復回数、および、処理時間(秒)^{*}に関する基本統計量を表1に示す。表より、 $h = 2$ でMDL値が最小になり、正しい中間ユニット数を発見できたことが分かる。また、このとき100回すべての試行が最適解に収束した。発見した法則を図1に示す。ただし、各値は小数点第4位で四捨五入した。図よりRF5は元の法則を完全に復元できたことが分かる。ここで、各試行の平均反復回数は93.7回、

平均処理時間は0.878秒であり、この実験に要した全処理時間は4.4分であった。

5.1.2 修正問題

指数の値が実数の場合でのRF5の有効性を評価するため、式(15)の代わりに、

$$y = 2 + 3x_1^{-1}x_2^3 + 4x_3x_4^{1/2}x_5^{-1/3} \quad (16)$$

を用いて実験を行った。ただし、実験の条件はオリジナル問題のときとすべて同じとした。結果を表2に示す。この実験でも正しい中間ユニット数を発見できた。しかし、 $h = 2$ のとき、数回の試行は望ましくない局所最適解に収束した。発見した法則を図2に示す。これは式(16)と等価である。既存法では、適切な関数を用意しなければこのような法則を発見できないので、既存法と比べて、RF5には重要な長所があることが示された。

5.1.3 ノイズ許容性

RF5のノイズ許容性を評価するため、式(15)、または、式(16)で計算する各 y の値に、平均0、標準偏差0.1の正規分布に基づく独立なノイズを与えて実験を行った。ただし、これ以外の実験の条件は、以前のものとすべて同じとした。結果を表3に示す。最良のMSE値は $h = 3$ のときであるが、最良のMDL値は $h = 2$ のときであり、いずれの問題でも、正し

^{*} 実験にはHP/9000/735を用いた。

表3 ノイズ許容性
Table 3 Noise tolerance.

ユニット数	MSE 値			MDL 値			反復回数		処理時間	
	best	avg.	s.d.	best	avg.	s.d.	avg.	s.d.	avg.	s.d.
ノイズありオリジナル問題										
1	0.160	0.16	0.0	-154.0	-154	0.0	68	4	0.31	0.02
2	0.009	0.01	0.0	-416.0	-416	0.0	93	9	0.77	0.07
3	0.008	0.01	0.0	-405.5	-405	0.8	784	81	9.80	1.02
ノイズあり修正問題										
1	2.326	2.33	0.0	113.6	114	0.0	90	12	0.41	0.05
2	0.010	0.03	0.21	-403.9	-399	53	228	134	1.90	1.11
3	0.009	0.01	0.0	-388.9	-385	1.2	753	127	9.41	1.58

$$\begin{aligned}
 y &= 1.968 + 3.028x_1^{1.000}x_2^{0.969}x_4^{-0.007}x_5^{-0.007}x_6^{0.004}x_7^{0.008}x_8^{-0.007}x_9^{0.001} \\
 &\quad + 3.880x_1^{-0.027}x_2^{-0.014}x_3^{1.025}x_4^{0.995}x_5^{1.048}x_6^{-0.008}x_7^{-0.020}x_8^{0.010}x_9^{-0.014} \\
 y &= 2.012 + 3.004x_1^{-1.000}x_2^{3.001}x_6^{-0.001}x_7^{0.001} \\
 &\quad + 3.983x_1^{0.002}x_2^{-0.003}x_3^{1.022}x_4^{0.500}x_5^{-0.333}x_6^{-0.005}x_7^{-0.002}x_8^{0.003}x_9^{-0.007}
 \end{aligned}$$

図3 ノイズを含むデータから発見した法則
Fig. 3 Discovered laws from data containing noise.

$$\begin{aligned}
 y &= 2.0 + 3.0x_1^{1.0}x_2^{1.0} + 3.9x_3^{1.0}x_4^{1.0}x_5^{1.0} \\
 y &= 2.0 + 3.0x_1^{-1.0}x_2^{3.0} + 4.0x_3^{1.0}x_4^{0.5}x_5^{-0.3}
 \end{aligned}$$

図4 小数点第2位で四捨五入した結果
Fig. 4 Results rounded off to the first decimal place.

い中間ユニット数を発見できた。オリジナルと修正問題で RF5 が発見した法則を図 3 に示す。図 3 の各値を小数点第 2 位で四捨五入した結果を図 4 に示す。いくつかの値はわずかに異なるが、元の法則とはほぼ等価な法則を発見できた。このことは、RF5 は頑健で、ある程度のノイズを許容できることを示している。

5.2 学習効率の評価

ここでは、RF5 で採用する学習アルゴリズム BPQ の効率を評価する。

5.2.1 BP との比較

学習効率をグラフィカルに評価するため、2 変数からなる人工問題を作成した。つまり、法則の一般形を $y = x^{w_1} + w_2$ (17) とし、 $(w_1, w_2) = (0.4, 0.2)$ で真の法則を与えるとする。入力事例 x_t は $\{0.1, 0.2, 0.3, 0.4, 0.5\}$ の各要素とし、目標出力値 y_t は式 (17) に真のパラメータを代入して各 x_t から計算した。すなわち、最小値は真のパラメータの値で与えられる。実験では、標準的な BP、慣性項 (momentum term) 付き BP、および、BPQ を比較した。誤差曲面上において、 $(w_1, w_2) = (0.0, 0.0)$ を初期値とし、最大で 100 反復させた学習軌跡を図 5 に示す。ただし、図の学習定数や慣性項の係数は試行錯誤で決定した。図より、標準的な BP と慣性項付き

BP では、初期値を比較的最小値の近くに設定したにもかかわらず、100 反復では最小値に到達できなかった。この理由は、谷底近くでは、連続する 2 つの勾配ベクトルの方向がほぼ逆向きになるためであり、これは 1 次の学習法の本質的な問題点である。一方、BPQ では、十数反復で効率良く最小値に到達できた。さらに、BPQ には試行錯誤で決定するパラメータがないので、一般の問題への適用が容易となる。

5.2.2 適応 BP との比較

ノイズありの Sutton-Matheus のオリジナルと修正問題を用いて、BPQ の効率を評価した。この実験では、標準的な BP では、すべての試行が収束しなかったため、Silva-Almeida の学習定数適応規則¹⁴⁾を用いたアルゴリズム (適応 BP) と比較した。適応 BP では、各結合重み Φ^i に対する学習定数 η^i は

$$\eta_k^i = \begin{cases} \eta_{k-1}^i \times u & \text{if } \frac{\partial f(\Phi_k)}{\partial \Phi^i} \times \frac{\partial f(\Phi_{k-1})}{\partial \Phi^i} \geq 0 \\ \eta_{k-1}^i \times u^{-1} & \text{otherwise} \end{cases}$$

で調整される。ここで、 k は反復回数を表し、パラメータ u は提案者が推奨するように 1.1 に設定した¹⁴⁾。ただし、目的関数の値が減少しないときには、すべての学習定数の値はその半分の値に設定される。

実験結果を図 6 に示す。ただし、図の値は 100 回の

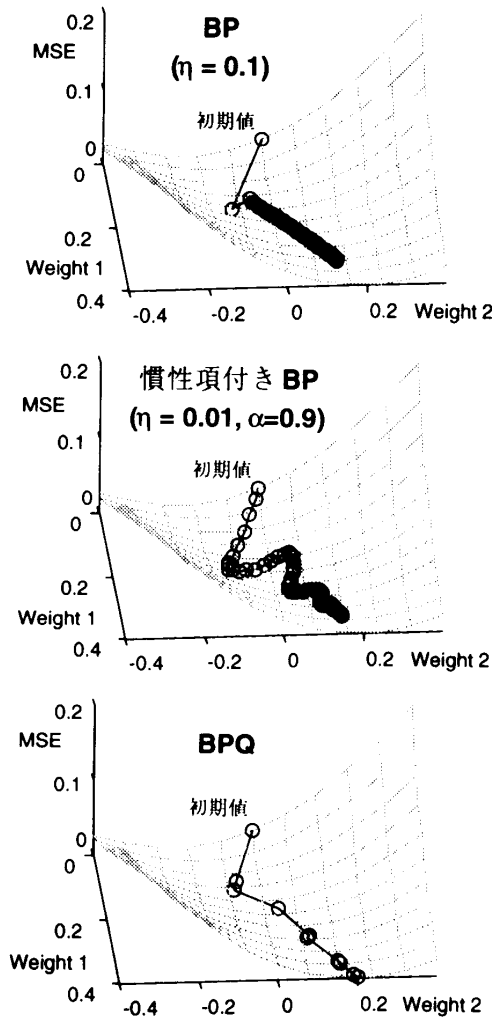
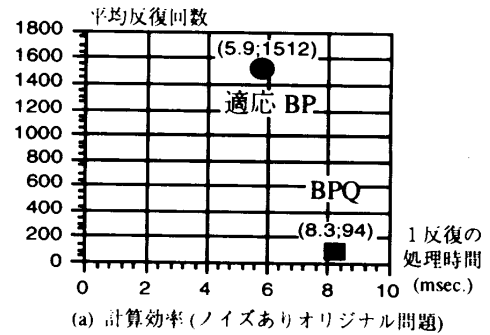


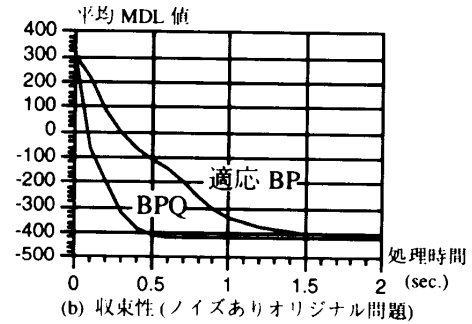
図5 学習軌跡
Fig. 5 Learning trajectories.

試行の平均である。図 6(a) では、ノイズありオリジナル問題において、1 反復の処理時間と収束までに要した反復回数との関係を示す。図より、適応 BP の 1 反復の処理時間は BPQ よりわずかに少ないが、適応 BP の反復回数は BPQ の 16.1 であり、全体では、BPQ は適応 BP より 11.4 倍速いことが分かる。図 6(b) では、ノイズありオリジナル問題において、MDL 値による収束性を比較する。明らかに、BPQ の収束性は適応 BP より優れている。図 6(c) では、ノイズあり修正問題において、MDL 値による収束性を比較する。この問題では、適応 BP のすべての試行が収束しなかった。

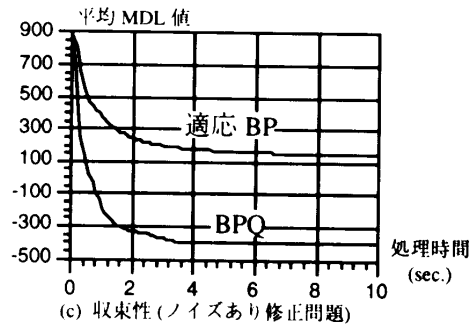
適応 BP がノイズあり修正問題をうまく学習できなかった理由については、目標値 y_t に関する基本統計量がヒントを与える。今回の実験では、ノイズありオリジナル問題の平均と標準偏差は 3.33 と 0.90 (値のレンジは [1.8, 6.3]) であったが、ノイズあり修正問



(a) 計算効率(ノイズありオリジナル問題)



(b) 収束性(ノイズありオリジナル問題)



(c) 収束性(ノイズあり修正問題)

図6 学習結果
Fig. 6 Learning results.

題では、19.40 と 60.95 (値のレンジは [2.1, 555.9]) であった。これは、結合重みベクトルが変化すると、オリジナル問題と比較して、修正問題での勾配ベクトルがより急激に変化することを意味する。これが適応 BP では学習できなかった理由であると考えられる。

5.3 現実データによる評価

現実データとして、Boyle の法則、Kepler の第三法則、および、Ohm の法則に従うデータ[☆]を用いた実験を行った。ここで、Boyle の法則は気体の体積と圧力の関係、Kepler の法則は太陽との距離と惑星の公転周期の関係、Ohm の法則は銅線とそこを流れる電流の関係である。この実験で、最尤推定における事例数とパラメータ数の割合を考慮すれば、中間ユニット数が 2 以上では事例数が十分でないので、ユニット数 1 のケースだけを考えた。なお、この場合でも、定数項 c_0 を考慮するので、この学習は単純な回帰問題に

[☆] これらのデータは UCI 機械学習データベースから得た。

表4 現実データによる評価
Table 4 Evaluation using real data.

法則名	参照法則	発見した法則
Boyle の法則	$y = 29.30/x$	$y = 29.05x^{-1.08} - 0.61$
Kepler の法則	$y = 0.41x^{1.5}$	$y = 0.41x^{1.50} + 0.19$
Ohm の法則	$y = 7268.74/x - 20.30$	$y = 7812.02x^{-1.02} - 19.14$

帰着されない。参照法則 \star と発見した法則を表4に示す。ただし、結果は10回の試行でMDL値を最小にしたものであり、各値は小数点第3位で四捨五入した。Boyleの法則とKeplerの第3法則では、望ましくない定数項が現れたが、参照法則と類似した法則を発見できた。また、これらの結果からは、定数項の値が比較的小さいので、それを0に固定したモデルでの試行が示唆される。一方、Ohmの法則には、参照法則からかなり外れたデータも含まれるが、参照法則と類似した法則を発見できた。

6. おわりに

数値データから未知の法則を発見するため、コネクショニストアプローチに基づく方法RF5を提案した。RF5では、数法則の発見問題がニューラルネットの学習問題として定式化され、最適ステップ幅を2次近似の最小点として求める2次の学習アルゴリズムBPQで学習し、学習結果の数法則候補から適切な法則をMDL基準を用いて選択する。実験では、ある程度のノイズを含むデータからでも、RF5は指数の値が整数に制限されない法則を効率良く発見することができた。今後は、重要でない結合重みを自動的に枝刈りできるようにアルゴリズムの改良を進めるとともに、幅広い問題へRF5を適用して評価する予定である。

参考文献

- 1) Baum, E. and Haussler, D.: What Size Net Gives Valid Generalization, *Neural Computation*, Vol.1, No.1, pp.151-160 (1989).
- 2) Durbin, R. and Rumelhart, D.: Product Units: a Computationally Powerful and Biologically Plausible Extension, *Neural Computation*, Vol.1, No.1, pp.133-142 (1989).
- 3) Falakenhainer, B. and Michalski, R.: Integrating Quantitative and Qualitative Discovery in the Abacus System, *Machine Learning: an Artificial Intelligence Approach, Volume III*, Kodratoff, Y. and Michalski, R. (Eds.), pp.153-190, Morgan Kaufmann, San Mateo, CA (1990).
- 4) Gill, P., Murray, W. and Wright, M.: *Practical Optimization*, Academic Press, London (1981).
- 5) Langley, P.: Bacon.1: A General Discovery System, *Proc. Second National Conference of the Canadian Society for Computational Studies of Intelligence*, pp.173-180 (1978).
- 6) Langley, P., Simon, H., Bradshaw, G. and Zytkow, J.: *Scientific Discovery: Computational Explorations of the Creative Process*, MIT Press, Cambridge, MA (1987).
- 7) Langley, P. and Zytkow, J.: Data-Driven Approaches to Empirical Discovery, *Artificial Intelligence*, Vol.40, pp.283-312 (1989).
- 8) Leerink, L., Giles, C., Horne, B. and Jabri, M.: Learning with Product Units, *Advances in Neural Information Proc. Systems 7*, Tesauro, G., Touretzky, D. and Lee, T. (Eds.), MIT Press, Cambridge, MA (1995).
- 9) Nordhausen, B. and Langley, P.: A Robust Approach to Numeric Discovery, *Proc. Seventh International Machine Learning Conference*, pp.411-418, Austin, Texas (1990).
- 10) Rissanen, J.: A Universal Prior for Integers and Estimation by Minimum Description Length, *Ann. of Statist.*, Vol.11, No.2, pp.416-431 (1983).
- 11) Rumelhart, D., Hinton, G. and Williams, R.: Learning Internal Representations by Error Propagation, *Parallel Distributed Processing*, Rumelhart, D. and McClelland, J. (Eds.), pp.318-362, MIT Press (1986).
- 12) Sanger, T.: Basis-Function Trees as a Generalization of Local Variable Selection Method for Function Approximation, *Neural Information Processing Systems 3*, Touretzky, D. (Ed.), pp.707-713, Morgan Kaufmann, San Mateo, CA (1991).
- 13) Schaffer, C.: Bivariate Scientific Function Finding in a Sampled, Real-Data Testbed, *Machine Learning*, Vol.12, No.1/2/3, pp.167-183 (1993).
- 14) Silva, F. and Almeida, L.: Speeding up Backpropagation, *Advanced Neural Computers*, Eckmiller, R. (Ed.), pp.151-160, North-Holland, Amsterdam (1990).
- 15) Sutton, R. and Matheus, C.: Learning Polynomial Functions by Feature Construction, *Proc.*

\star Ohmの法則に定数項が現れるのは不自然であるが、UCIデータベースでは定数項付きのものを参照法則としていた。

Eighth International Machine Learning Workshop, pp.208-212, Evanston, IL (1991).

- 16) Valiant, L.: A Theory of the Learnable, *CACM.*, Vol.27, No.11, pp.1134-1142 (1984).

(平成7年6月23日受付)

(平成8年6月6日採録)



斉藤 和巳 (正会員)

昭和38年生。昭和60年慶応義塾大学理工学部数理科学科卒業。同年NTT入社。平成3年より1年間オタワ大学客員研究員。神経回路網、機械学習の研究に従事。現在、NTTコミュニケーション科学研究所主任研究員。人工知能学会、神経回路学会、日本認知科学会各会員。



中野 良平 (正会員)

昭和22年生。昭和46年東京大学工学部計数工学科卒業。工学博士。同年、日本電信電話公社(現NTT)入社。以来、統計解析、分散処理、データベース、人工知能、学習ロボットの研究に従事。現在、NTTコミュニケーション科学研究所主幹研究員。人工知能学会、神経回路学会、日本応用数理学会各会員。