

Neural Networks の数量化理論への適用

2M-1

鈴木 健嗣

橋本 周司

早稲田大学理工学部応用物理学科

1. はじめに

統計手法の一つである多変量解析は、データ解析によって多くのデータの相互関係を解析するものである。近年、コンピュータの進歩によりさまざまな分野での応用が容易になり、理学、工学を始め、製品開発やマーケティング調査まで、必要不可欠なものになってきている。しかしながらこれはデータの全体的傾向を視覚的に効率よく説明するためには有効であるが、数量化後の結果と、新たに加えられたデータの関係は明確には判らない。ここではデータの数量化を非線形写像と考え、ニューラルネットワークを用いることにより、従来の数量化法と比較することで、その有効性を検討した。

2. ネットワークモデル

ここでは、ネットワークモデルとしてDMLN (Distance Mapping Learning Network) を用いる。

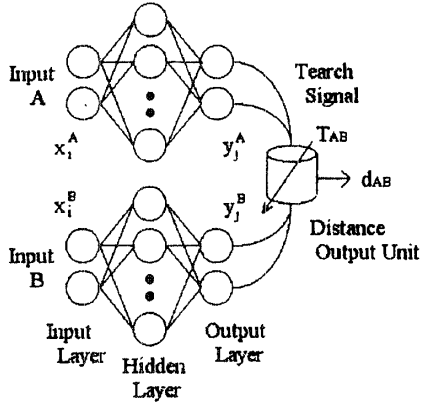


Fig 1. DMLN (Distance Mapping Learning Network)

DMLNは、階層型パーセプトロンの一変形である。学習則はBP (Back Propagation) 法を利用しており、ネットワークが並列構造をしているのが特徴である。Fig. 1 に示したように、A、Bの2列のネットワークが同じ写像を行うものであ

り、座標  $x_i^A(t)$ 、 $x_i^B(t)$  を時刻  $t$  における入力とすると、出力層から写像後の座標  $y_i^A$ 、 $y_i^B$  が出力される。これが距離計算ユニットに入り、ネットワークは写像後の2つの入力間の距離  $d_{AB}$  を出力し、学習はこの距離  $d_{AB}$  を教師信号に近づけるように行われる。

各ユニットの出力にはSigmoid関数  $f$  を使い、ユニット間は荷重  $w_{ij}$  でユニット  $j$  から  $i$  へ結合している。距離計算ユニットの出力  $d_{AB}$  は、単純にユークリッド距離を与えている。ここでネットワークの出力誤差  $E_p$  は、下式を利用し、これを最小にするように学習を行う。

$$E_p = \sum (T_{AB} - d_{AB})^2 / 2 \tag{1}$$

したがって、学習の目標は、次式を0にすることであり、荷重  $w_{ij}$  と誤差関数  $E_p$  の関係式は、となる。

$$\frac{\partial E_p}{\partial w_{ij}} = \frac{1}{2} \left\{ \frac{\partial E_p}{\partial y_i^A} \frac{\partial y_i^A}{\partial x_j^A} \frac{\partial x_j^A}{\partial w_{ij}} + \frac{\partial E_p}{\partial y_i^B} \frac{\partial y_i^B}{\partial x_j^B} \frac{\partial x_j^B}{\partial w_{ij}} \right\} \tag{2}$$

$$= -\frac{1}{2} (T_{AB} - d_{AB}) \frac{y_i^A - y_i^B}{d_{AB}} \{ f'(x_j^A) \cdot y_i^B - f'(x_j^B) \cdot y_i^A \}$$

さらに、結合荷重の更新には、下式のように前学習時の慣性項を付与して学習させている。

$$\Delta w_{ij}(t+1) = \alpha \frac{\partial E_p}{\partial w_{ij}(t)} + \eta \Delta w_{ij}(t) \tag{3}$$

この時、座標点の決定に際して、固定条件を与えていないため、回転に対しては自由である。

3. 数量化理論IV類

多変量解析の中で、データを目的に応じて適当な数量に変換する方法を数量化法 [1] という。ここでは、 $n$ 個の対象に対して、それら相互の親近性より各対象の相対的な位置づけを行う数量化理論IV類について考える。

これはサンプル  $ij$  間の親近性  $e_{ij}$  より次式を最小とするようにサンプルの座標  $\{x_i\}$  を決める方法である。

$$Q = \sum_{i=1}^n \sum_{j=1}^n e_{ij} (x_i - x_j) \tag{4}$$

これにより、親近性が大きい(似ている)ものは、近くに、そうでないものは遠くに配置されるのである。

### 4. 評価実験

ここでは、Ekmanが8種類の色彩の組み合わせの類似度を、31人の被験者に評価させたものについて、本提案手法と数量化理論IV類より数量化したものの結果とを比較する。学習は、UNIXワークステーション (Sun SPARC Station 20) 上で行い、すべて30秒程度で収束した。なお、学習係数は、 $\alpha = 0.55$ 、 $\eta = 0.10$  を用いた。

Fig. 2 に、数量化理論によって配置されたサンプルを示した [2]。これに対して、Fig. 3 にはサンプルの初期座標値を長方形の配置に与えたもの、同様に Fig. 4 には、平面状に一様に与えたものに対して、それぞれ本手法によりサンプル座標を非線形写像したものを示す。

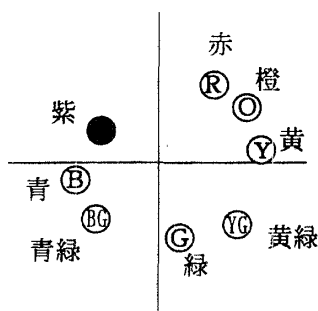


Fig. 2 数量化理論による配置 (Ekman, 1954)

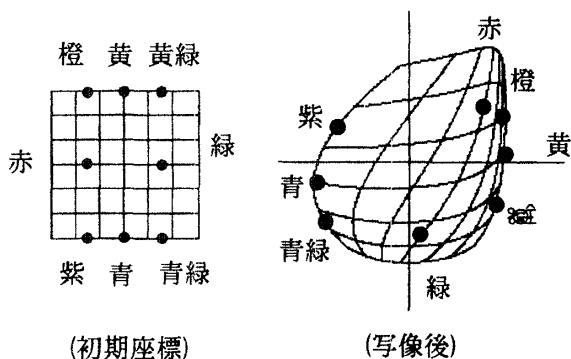


Fig. 3 長方形の配置からの写像

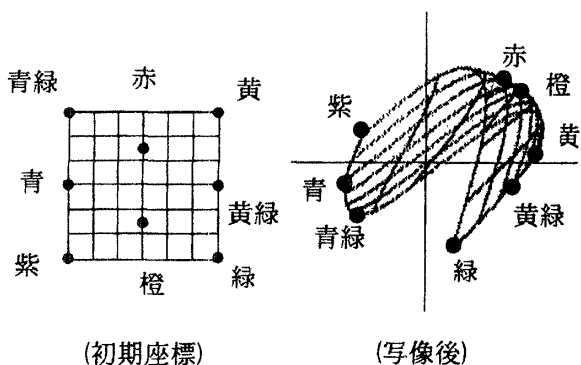


Fig. 4 一様配置からの写像

また Fig. 5-a は、8種類の色相を1次元配置で与えたものである。また、ここでは紫を入力せずに学習を行ったが、Fig. 5-b にあるように、写像後の結果は環状に配置され、未知の入力である紫は、数量化理論と同様の位置に配置されていることがわかる。

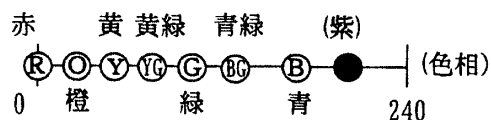


Fig. 5-a 8種類の色相順による配置

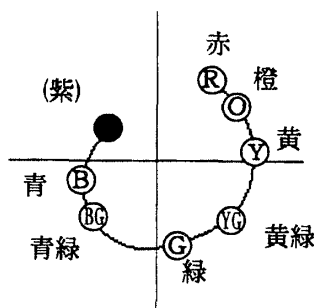


Fig. 5-b 未知入力に対する評価

### 5. おわりに

提案したDMLNにより数量化理論と同様の配置づけが、可能となるばかりでなく新たな入力においてもその親近性の評価が可能になった。また、サンプルの初期座標を利用目的に対して任意で与えられることから、数量化理論の利用用途が広がると考えられる。

またDMLNは多次元への拡張が容易である。さらに距離計算ユニットには、ユークリッド距離の他、ハミング距離やマハラノビスの汎距離などの適用も可能である。

今後は、様々な分野のサンプルデータを用いて感性的情報の対応づけについても考えていきたい。またデータベースの感性的検索方法への応用についても検討している。

### 参考文献

[1] 林 知己夫「数量化」朝倉書店 (1993)  
 [2] Ekman, G. "Dimensions of color vision", Journal of Psychology (1954)  
 [3] J. Hertz, A. Krogh, R. G. Palmer "Introduction to the theory of Neural Computation" Addison-Wesley Publishing (1991)