広告効果予測のための制約付きニューラルネットワーク学習方式

正 員 谷 口 洋 司 (日立製作所)

正 員 水 野 浩 孝 (日立製作所)

正 員 矢 島 敬 士 (日立製作所)

Constrained Learning Algorithm of Neural Networks for an Estimate of Advertising effect
Yoji Taniguchi, Member, Hirotaka Mizuno, Member, Hiroshi Yajima, Member (Hitachi Ltd.)

This paper proposes an algorithm of neural networks for an estimate of advertising effect. The basic premise under this algorithm is that the relationship between input and output is monotonic increase. The algorithm is useful for the domain like an estimate of advertising effect which only deficient data are available.

The experimental simulation shows that the algorithm builds a reasonable model for an estimate of advertising effect.

キーワード：広告代理店，ニューラルネットワーク，単調増加性，先験的知識

1. はじめに

広告や流通などの分野では，効果的な商品企画や販売促進のために，顧客データベースを利用する需要予測などのデータベース・マーケティングが注目されている。広告代理店のマーケティング部門では，新商品の知名度向上のための4媒体（テレビ，ラジオ，新聞，雑誌）への広告出稿計画を立案し，予測される広告効果（知名度）を示すことにより，広告主を説得している。広告効果の予測には，過去の類似した商品の広告実績データが利用される。ここでは，「類似性の判定の曖昧さ」と「データの新鮮さが求められること」により，それぞれの新商品の広告効果予測に利用可能なデータは少数でばらつきが大きいものとなる。一方，新商品の広告に関しては，広告量を増やせば知名度は下がるという常識がある。

例えば広告出稿計画を作成する目的で，重回帰分析手法を用いて，広告4媒体への広告出稿量から一般顧客の商品の知名度を予測することが試みられてきた。しかし，データとの乖離を縮めることは困難で，さらに広告量を増やすほど知名度が下がるという常識に合わない結果が現れ，実用レベルに達していない⁽¹⁾⁽²⁾。

近年，データベースマーケティングの一モデル化手法として階層型ニューラルネットワークが利用されつつある⁽³⁾。何らかの因果関係は認められるがその構造が複雑で重回帰モデルなどの統計手法では捕らえにくいような対象に対して，ニューラルネットワークがその因果データ（教師データ）から構造をモデル化できるという特徴を持つ⁽⁴⁾⁽⁵⁾からである。

我々は，バックプロパゲーション法（以後BP法と呼ぶ）を用いて，広告4媒体への広告出稿量から一般顧客の商品の知名度を予測することを試みた。しかし，データとの乖離を縮めることには成功したものの，重回帰分析手法と同様に，広告量を増やすほど知名度が下がるという常識に合わない結果が現れた。

一方，例えば入力パターンのある種の変換に対する出力の不変性や入力パターンの変化に対する出力変化の滑らかさなどの先験的知識をニューラルネットワークの学習に反映させて，学習負荷を軽減するとともに，汎化能力を高める手法がいくつか提案されている⁽⁶⁾⁽⁷⁾。

しかし，「データが少数でばらつきが大きい。」という状況下で，「入力パターンの増加に対する出力パターンの単調増加性。」を実現する手法は提案されていない。

本論文では，データの分布に偏りがあり，少量データしか利用できない分野である広告効果予測問題を対象とし，ニューラルネットワークにデータの因果関係の単調増加性などの専門家の先験的知識をモデルに反映させる方式を提案する。本方式はニューラルネットワーク学習時に各ニューロンのウェイトの有効範囲に制限を設けることにより，入力層の特定のニューロンと出力層の特定のニューロンとの間の単調増加性を実現する。

広告業で実際に利用されているデータを用いたシミュレーションの結果，本方式を利用してニューロモデルを構築することにより，BP法に比べ，予測精度が高くなることが確認できた。

さらに，広告業における「広告効果モデル」に求められる「単調増加性」及び「飽和性」を特性として備えた模擬

データを利用したシミュレーションの結果、本方式を利用してニューロモデルを構築した場合に有効なデータ規模の範囲が確認できた。

2. 広告効果予測問題

2. 1 広告効果予測モデルとその要件

広告代理店は、広告依頼主（以後クライアントと呼ぶ）から、ある新商品の広告の制作・出稿を依頼される。マーケティング部門では4媒体（テレビ、ラジオ、新聞、雑誌）等への広告出稿計画案と予算の見積もりをクライアントに提示し、広告の受注決定後、制作部門では具体的な広告の内容（ストーリー、起用タレントなど）を決定し制作する。ここで、我々がシステム化の対象としているのはマーケティング部門で行われている新商品の知名度向上のための広告出稿計画業務である。

マーケティング部門では、立案した広告出稿計画案が知名度向上に繋がることを示す必要がある。そのために、過去の広告実績データを利用して広告効果予測モデルを構築し、出稿計画案の広告効果をシミュレーションすることが試みられている。ここで利用される広告実績データの例を図1に示す。

商品名	テレビ	ラジオ	新聞	雑誌	知名度
α コーラ	1225	223	455	33	12.5
β サイダー	854	111	636	10	25.1
⋮	⋮	⋮	⋮	⋮	⋮
γ ドリンク	332	874	102	105	32.9

図1 広告実績データ例

Fig1. Sample data of an advertising effect.

例えば、 α コーラに関するテレビ広告量が1225で、ラジオ広告量が223で、新聞広告量が455で、雑誌広告量が33であった時に、知名度が12.5%であったことを示す。広告量とは、 α コーラに関する広告が行われた期間（通常、テレビ・ラジオは数週間から数カ月、新聞・雑誌は一日から数カ月）に一般顧客に到達した α コーラに関する情報の量を表す。テレビであれば、視聴率と広告回数から算出される。知名度とは、上記広告が終了した時点で、調査会社が行う一般顧客に対する知名度アンケート調査により獲得される。広告効果予測モデルとは、このような過去の広告実績データの広告量と知名度の関係をモデリングしたものであり、このモデルに出稿計画案を入力することにより、広告効果（知名度）を予測する。

このモデルに求められる要件としては、過去の実績データをうまく表現していることは勿論のことであるが、「広告量を増やせば知名度は下がる」「知名度は100%を越えない（飽和性がある）」という広告と知名度の関係の常識に合致したモデルであることが考えられる。

2. 2 予測モデルを利用した広告計画業務の流れ

上記要件を満足した広告効果予測モデルが構築できればマーケティング分析専門家による広告効果予測モデル構築業務とエンドユーザ（営業職員）による広告出稿計画立案を連携させることができ、以下のような広告出稿計画業務の流れ（図2）が成立するものと考えられる。

図2 広告出稿計画業務の流れ

Fig2. A flow of an advertising planning.

(1) 分析専門家による広告効果予測モデル構築

- (a) データ収集
広告出稿と効果に関する過去のデータ（4媒体への出稿量とそのときの一般顧客の知名度から成る）を収集する。
- (b) 商品分類
上記データを商品分類（飲料、食品、衣料品など）・ターゲット（20代の男性、40代の女性などの購買層を表す）別に分類する。
- (c) 予測モデル構築
上記グループのデータを分析し、4媒体への広告出稿量から知名度を予測する広告効果モデルを構築する。

(2) エンドユーザ（営業職員）による広告出稿計画立案

- (a) 広告出稿計画案の作成
4媒体への広告出稿計画案を作成する。計画案は、予算案の他、どの新聞の何面に広告を出稿するかといった出稿先と、何月に広告を出稿するかといった出稿時期から成る。
- (b) 広告効果シミュレーション
広告効果予測モデルを利用して広告効果をシミュレーションする。
- (c) 広告計画案の決定
クライアントを説得できる予測と知名度が得られるまで上記(a)(b)を繰り返す。

2. 3 重回帰分析による広告効果予測

上述のような広告出稿計画業務を実現するために、広告

実績データを利用して、重回帰分析手法による知名度予測モデルの構築が試みられている。しかし、データとの乖離を縮めることは困難で、さらに「新聞や雑誌の広告をすればするほど知名度が下がる」、「知名度が100%を越えてしまう」といった結果となり、常識に合わず、実用レベルに達していない。そのため、多くのエンドユーザは個人的に保存している過去の広告出稿計画や経験を元にして、新たな広告出稿計画を立案している。

この分野で扱うデータの特徴として以下の2点が考えられる。

- (1) 「データに新鮮さが求められること」と「ターゲットとしている商品との類似性の判定の曖昧さ」により、新商品の広告効果予測モデル構築に利用可能なデータは少数(30~40件)でかつばらつきが大きいものとなる。
- (2) 広告量と知名度の間の先験的知識として、単調増加性と飽和性(知名度が100%を越えない)がある。

これに対し、重回帰分析手法は、「各入力と出力との全体的な相関を線形的に捕らえる。」傾向がある。そのため上記特徴を持つデータに適用した場合、一部の特異なデータに引きずられた相関結果となってしまう可能性が高い。この結果が、先験的知識と合致しなかったと考えられる。

以上のように、広告効果の予測モデル構築に関しては、この分野で扱うデータの特徴である前記(1)(2)に適應できるモデル化手法が求められる。

3. 制約付きニューラルネットワーク学習方式

3.1 基本的な考え方

前述のモデルに求められる要件の中で、「過去の実績データをうまく表現すること」「知名度は100%を越えない(飽和性がある)こと」に関しては、ニューラルネットワークの学習能力⁽⁸⁾および表現能力⁽⁵⁾で対応できるものとする。

一方「広告量を増やせば知名度は下がる」ことに関しては、ニューラルネットワークの入出力に単調増加性を持たせるための学習方式が必要であると考えられる。次節では、ニューラルネットワークの入出力に単調増加性を持たせるための十分条件について述べる。

3.2 ニューラルネットワーク入出力の単調増加性

階層型ニューラルネットワークの入力層の特定のニューロンへの入力と出力層の特定のニューロンからの出力の間に単調増加性を持たせるための十分条件を以下で示す。

$$y_{jn} = h_j(z_{jn}) \quad \dots\dots\dots (1)$$

$$z_{jn} = \sum_i w_{ij} x_{in} \quad \dots\dots\dots (2)$$

まず、対象とする階層型ニューラルネットワークの第 n 入力パターンに対する第 $s-1$ 層第 i ニューロンの出力を x_{in} 、第 s 層第 j ニューロンの出力を y_{jn} とし、ニューロンの入出力関係が次の(1)式および(2)式で与えら

れているものとする。

但し、 $h_j(\cdot)$ は、微分可能な非減少関数で与えられるシグモイド関数であり、 w_{ij} は、第 $s-1$ 層第 i ニューロンと第 s 層第 j ニューロンを結ぶ結合の重みを意味する。

ここで、入力層(p 個の入力ノード)、中間層(q 個の中間ノード)、出力層(r 個の出力ノード)の3層から構成される階層型ニューラルネットワークを考える(図3)。

図3 3層ニューラルネットワーク
Fig3. 3 layers neural networks.

入力層第 i ニューロンと中間層第 j ニューロンを結ぶ結合の重みを w_{ij} 、中間層第 j ニューロンと出力層第 k ニューロンを結ぶ結合の重みを w_{jk} とすると、入力層のニューロンへの入力はそのままだ入力層のニューロンの出力となるので、(1)式および(2)式により、ニューラルネットワークの第 n 入力パターンに対する入力層の各ニューロンへの入力 x_{in} ($i=1,2,\dots,p$) と、出力層第 k ニューロンの出力 y_{kn} との関係は、(3)式および(4)式で表わされる。

$$y_{kn} = h_k \left(\sum_{j=1}^q w_{jk} y_{jn} \right) \quad \dots\dots\dots (3)$$

$$y_{jn} = h_j \left(\sum_{i=1}^p w_{ij} x_{in} \right) \quad \dots\dots\dots (4)$$

ここで、入力層の特定のニューロンへの入力と出力層の特定のニューロンからの出力の単調性を調べるために、出力層第 k ニューロンの出力 y_{kn} を入力層第 m ニューロンへの入力 x_{mn} で偏微分すると、

$$\begin{aligned} \frac{\partial y_{kn}}{\partial x_{mn}} &= \frac{\partial \{h_k(\sum_{j=1}^q w_{jk} y_{jn})\}}{\partial x_{mn}} \\ &= \frac{\partial \{h_k(\sum_{j=1}^q w_{jk} y_{jn})\}}{\partial (\sum_{j=1}^q w_{jk} y_{jn})} \cdot \frac{\partial (\sum_{j=1}^q w_{jk} y_{jn})}{\partial x_{mn}} \quad \dots\dots (5) \end{aligned}$$

ここで、 $h_k(\cdot)$ は、微分可能な非減少関数で与えられるシグモイド関数であるから、

$$\frac{\partial \{h_k(\sum_{j=1}^q w_{jk} y_{jn})\}}{\partial (\sum_{j=1}^q w_{jk} y_{jn})} \geq 0 \quad \dots\dots\dots (6)$$

である。従って、

$$\begin{aligned} \frac{\partial y_{kn}}{\partial x_{nm}} &= \frac{\partial \{h_k(\sum_{j=1}^q w_{jk} y_{jn})\}}{\partial x_{nm}} \cdot K_1 \quad (\text{但し, } K_1 = \frac{\partial \{h_k(\sum_{j=1}^q w_{jk} y_{jn})\}}{\partial (\sum_{j=1}^q w_{jk} y_{jn})}) \\ &= \frac{\partial \{h_k(\sum_{j=1}^q w_{jk} h_j(\sum_{i=1}^p w_{ij} x_{in}))\}}{\partial x_{nm}} \cdot K_1 \\ &= \sum_{j=1}^q \left\{ w_{jk} \cdot \frac{\partial h_j(\sum_{i=1}^p w_{ij} x_{in})}{\partial \sum_{i=1}^p w_{ij} x_{in}} \cdot \frac{\partial \sum_{i=1}^p w_{ij} x_{in}}{\partial x_{nm}} \right\} \cdot K_1 \quad \dots\dots\dots (7) \end{aligned}$$

ここで、 $h_j(\cdot)$ は、微分可能な非減少関数で与えられるシグモイド関数であるから、

$$\frac{\partial h_j(\sum_{i=1}^p w_{ij} x_{in})}{\partial \sum_{i=1}^p w_{ij} x_{in}} \geq 0 \quad \dots\dots\dots (8)$$

である。従って、

$$\begin{aligned} \frac{\partial y_{kn}}{\partial x_{nm}} &= \sum_{j=1}^q \left(w_{jk} \cdot \frac{\partial \sum_{i=1}^p w_{ij} x_{in}}{\partial x_{nm}} \right) \cdot K_1 \cdot K_2 \\ &\quad (\text{但し, } K_2 = \frac{\partial h_j(\sum_{i=1}^p w_{ij} x_{in})}{\partial \sum_{i=1}^p w_{ij} x_{in}}) \\ &= \sum_{j=1}^q \left\{ w_{jk} \cdot \frac{\partial (w_{1j} x_{1n} + w_{2j} x_{2n} + \dots + w_{nj} x_{nn} + \dots + w_{pj} x_{pn})}{\partial x_{nm}} \right\} \\ &\quad \cdot K_1 \cdot K_2 \\ &= K_1 \cdot K_2 \cdot \sum_{j=1}^q w_{nj} w_{jk} \quad \dots\dots\dots (9) \end{aligned}$$

ここで、(6)式および(8)式より、

$$K_1 \geq 0, \quad K_2 \geq 0$$

であるから、

$$\sum_{j=1}^q w_{nj} w_{jk} \geq 0 \quad \dots\dots\dots (10)$$

であれば、

$$\frac{\partial y_{kn}}{\partial x_{nm}} \geq 0 \quad \dots\dots\dots (11)$$

となるので、(10)式は、階層型ニューラルネットワークの入力層の特定のニューロン（入力層第 m ニューロン）への入力と出力層の特定のニューロン（出力層第 k ニューロン）からの出力の間に単調増加性を持たせるための十分条件である。

3. 3 制約付きニューラルネットワーク学習方式

我々は、ニューラルネットワークの入出力に単調増加性を持たせるために、BPアルゴリズムを改良した制約付きニューラルネットワーク学習方式を以下に提案する。ここでは、学習の容易化を目的として、上記(10)式の条件を強めた下記条件を用いる。

$$w_{nj} \geq 0 \text{ and } w_{jk} \geq 0 \text{ for } j=1,2,\dots,q \quad \dots\dots\dots (12)$$

すなわち、BPアルゴリズムにおいて、重み w を常に監視し、(12)式を犯す重み w の修正時は、 w を0に設定する。本学習方式を図4に示す。

図4 制約付き学習ニューラルネットワーク学習方式
Fig4. A constrained learning algorithm of neural networks.

本方式では、BP法と同様に、学習パターンに基づき階層型ニューラルネットワークの学習を行う。まず、学習パターンの入力値を階層型ニューラルネットワークの入力層に入力し、出力層からの出力値と学習パターンの出力値（教師データ）との比較し誤差を求める。次に誤差に基づきパラメータ（重みやシグモイド関数のしきい値など）の修正値を算出し、修正する。この修正において、重み w_{ij} を常に監視し、重み w_{ij} が負となる修正時は、 w_{ij} を0に設定する。

4. シミュレーション実験

4. 1 シミュレーション内容

提案方式の有効性を確認するために、以下の2つのシミュレーション実験を行った。

【実験1】実データによる予測精度検証

実際に広告効果予測に用いられているデータを2ケースに関して、それぞれ教師データと検証データに分割して予測モデルを構築し、他のモデル化手法（重回帰分析、BP法によるニューラルネットワーク）と予測精度を比較した。

図5 模擬データ

Fig5. Mock data of advertising effect.

【実験2】模擬データによる対応可能なデータ規模の検証

広告効果の予測モデルに求められる先験的知識として「広告量と知名率の間の単調増加性と飽和性（知名率が100%を越えない）」があり、これを満足するデータ（図5）にノイズを付加したものを教師データとして予測モデルを構築し、図5のデータを検証データとして、他のモデル化手法（同上）と予測精度を比較し、提案方式の対応可能なデータ規模を検証した。

4. 2 実験データ

【実験1】

飲料に関するデータ（18件）と化粧品／薬品に関するデータ（21件）の2ケースの実データであり、入力4媒体（テレビ、ラジオ、新聞、雑誌）への広告量、出力は知名率である4入力1出力のパターンからなる。それぞれランダムに2分割し、教師データと検証データを作成した。

【実験2】

模擬データは、指数関数を元に合成した単調増加性・飽和性のある関数にデータのばらつきを表すノイズを付加したもの（式11）により発生させた。ここでは、検証の容易化を図るために、広告効果予測に用いられている実データ（4入力1出力）よりも自由度の小さい2入力1出力とし、「広告量」を想定した2入力の定義域は同一で0以上の整

数とし、「知名率」を想定した1出力の値域を0以上100以下とした。

ノイズ（式12）は、実データのばらつき具合（ほぼ同一の広告量に対する知名率の最大偏差が50%程度であったこと）から判断して、乱数を用いノイズ幅が最大のケースで絶対値25となるようにした。模擬データ作成に使用した式を以下に示す。

$$z = 100 \{1 - \exp(-\frac{x+y}{3})\} + noise() \dots\dots\dots (11)$$

$$0 \leq x, 0 \leq y \text{ 但し, } x, y \text{ は整数}$$

$$noise() = 25 (rand()) \dots\dots\dots (12)$$

但し、rand()は-1.0~1.0の乱数

入力値（広告量を想定）をx、y、出力値（知名率を想定）をzとし、入力値をそれぞれ同一の定義域の0以上の整数とし、実際に広告効果予測モデル構築に利用可能なデータ規模が30~40件であることから、検証する模擬データの規模を小さなものに限定し、それぞれ9件、36件、81件、144件、225件の入出力パターンからなる模擬データ5ケースを作成した。

4. 3 シミュレーション結果

【実験1】

実データによる予測精度検証結果を表1、表2に示す。

表1 飲料データによる誤差の二乗和比較

Table 1. Error sum of squares in drink data.

モデル化手法	教師データの学習 誤差の二乗和	検証データの予測 誤差の二乗和
重回帰	2072.08	9992.26
BPニューロ	11.76	12036.70
提案方式	1207.97	1234.53

表2 化粧品／薬品データによる誤差の二乗和比較

Table 2. Error sum of squares in cosmetics and medicines data.

モデル化手法	教師データの学習 誤差の二乗和	検証データの予測 誤差の二乗和
重回帰	2030.85	7728.92
BPニューロ	0.33	10999.00
提案方式	495.36	7513.83

いずれの場合もBP法は学習誤差が極端に小さくなっているが、予測精度は提案方式によるニューロモデルが最も高くなった。これは、BP法によるニューロモデルの過学習が顕著に現れているものと思われる。

従って、広告効果予測に用いている実データを利用した検証実験において提案方式は有効であることが確認できた。

【実験2】

各種モデル化手法による予測モデルの推定値と、ノイズを付加する前の模擬データとの平均二乗誤差のデータ件数による変化を図6に示す。

図6 推定値の平均二乗誤差
Fig.6 Prediction Errors of 3 models

提案方式によるニューラルネットワークは、データ件数が81件までの場合には、BP法によるものに比べ、平均二乗誤差が小さく、144件以上の場合には逆転している。提案方式は、BP法に比べ過学習しにくいので、少数のデータに対してノイズの影響を受けにくい、データ数が増加してくると学習そのものが困難となり、予測精度も低下しているものと思われる。

また、実データに比べ自由度の小さい2入力1出力のデータにおいてデータ規模が少なくとも81件までの場合に、提案方式がBP法よりも有効であることが確認できた。

従って、広告効果予測に用いている実データの場合には少なくとも81件程度までは提案方式が有効に機能することが期待でき、利用可能なデータ数が多くて30~40程度しかない広告効果予測において、提案方式は有効であると考えられる。

5. おわりに

本論文では、データの分布に偏りがあり、少量データしか利用できない分野である広告効果予測問題を対象とし、ニューラルネットワークにデータの因果関係の単調増加性などの専門家の先見的知識をモデルに反映させる方式を提案する。本方式は、ニューラルネットワーク学習時に各ノードのウェイトの有効範囲に制限を設けることにより、任意の入力ノードと出力ノードの間の単調増加性を実現する。実データによる検証と、模擬データを利用したシミュレーションの結果、広告効果予測に実際に利用可能な規模のデータ（少なくとも80件程度）に関して、提案方式がBP法よりも有効であることが確認できた。

最後に、本研究の遂行にあたり有効なご助言を頂いている株式会社大広瀬戸口香情報科学研究所チーフディレクターに厚くお礼申し上げます。

(平成 8年 5月28日受付, 平成 8年11月29日再受付)

- (1) 西尾チヅル：ニューラルネットワークによるマーケティングモデル，オペレーションズリサーチ，Vol.39, No.4, pp.203-208
- (2) 青木：リサーチ革命下における消費者行動研究の課題，マーケティングジャーナル，Vol.43 (1992) pp.47-54
- (3) Benard W., David E.R., and Michael A.L. : Neural Networks : Applications in Industry, Business and Science, COMMUNICATIONS OF THE ACM March 1994 /Vol.37, No.3
- (4) D.E.Rumelhart, J.L.McClelland, and the PDP Research Group : Parallel Distributed Processing, MIT Press (1986)
- (5) K.Funahashi : On the Approximate Realization of Continuous Mappings by Neural Networks : Neural Networks, Vol.2, pp.183-192 (1989)
- (6) 福島：位置ずれに影響されないパターン認識機構の神経回路モデル—ネオコグニトロン—：電子通信学会論文誌，Vol.J62-A, No.10, pp.658-665 (1979)
- (7) T.Poggio : Learning, Regularization and Splines : Neural Networks, Vol.1-S1, p.211 (1988)
- (8) 喜多，高岡，西川：共役勾配法を用いた誤差逆伝搬学習について，第34回システム制御情報学会研究発表講演会講演論文集，pp.129-130 (1990)

谷口 洋司 (正員) 1963年6月29日生まれ。87年3月早稲田大学理工学部電気工学科卒業。89年同大学院理工学研究科修士課程終了。同年(株)日立製作所入社。システム開発研究所関西システムラボラトリ勤務。事例ベースシステム，意思決定支援システムなどの開発に従事。情報処理学会会員。

水野 浩孝 (正員) 1957年3月13日生まれ。79年3月大阪大学工学部電気工学科卒業。81年同大学院工学研究科修士課程終了。同年(株)日立製作所入社。システム開発研究所関西システムラボラトリ勤務。画像処理システム，意思決定支援システムなどの開発に従事。計測自動制御学会，情報処理学会会員。

矢島 敬士 (正員) 1950年10月5日生まれ。73年3月京都大学工学部機械工学科卒業。75年同大学院工学研究科修士課程終了。同年(株)日立製作所入社。システム開発研究所関西システムラボラトリ勤務。エキスパートシステム，意思決定支援システムなどの開発に従事。工学博士。IEEE，情報処理学会会員。