1J－6 ニューラルネットワークサーバとクライアント間でのデータ非開示計算

中島俊哉
富士通（株）

1．はじめに

階層型ニューラルネットワーク（ＮＮ）の応用として，学習済ＮＮをサーバとし，クライアントからの情報を入力してサーバの出力結果をクライアントに提供する形態がある（例えば金融機関の格付けにおいて格付け機関にＮＮサーバを置き，金融機関をクライアントとする場合など）．

このような場合クライアント・サーバ間の通信が暗号化されていてもクライアントの重要な情報がＮＮへの入力としてサーバに伝わることに変わりはない．クライアントとしては自己の持つ情報をサーバに知られずに結果のみ得られればその方が望ましいが，一方でサーバもＮＮの構造パラメタとして実現されているノウハウを外部に公開することはできない．

本稿では，以上のような状況において既存のＮＮの構造を変更せずに，クライアントの情報をサーバに開示せず，さらにサーバのＮＮ構造データをクライアントに開示することなく，正しい結果を得るためのクライアント・サーバ間の情報交換手順を考察する．

2．問題の定式化

前節の手順に対する要件は以下の各項で表わされる．

(1) サーバはＮＮへの入力データからクライアントデータを復元できない．
(2) クライアントはサーバからのデータによってＮＮ構造パラメタを復元できない．
(3) ＮＮの計算結果は手順の導入前と同一．
(4) 手順導入後でもＮＮ構造パラメタは変更しない（再学習不要）．
(5) クライアントは自己の情報をサーバが復元できないことを確認できる．

これらの要件を実現するための方針を検討する．まず，クライアント側のデータを n 次元ベクトル $\boldsymbol{x}$ とし，これを入力とするサーバ側の３階層ＮＮの入出力関係式を

$$z = g(Bf(Ax+t)) \equiv h(Ax+t) \qquad (2.1)$$

とする．ここで A は m×n 行列，B は k×m 行列，f, g は各々m 次元ベクトル間，k 次元ベクトル間の非線形変換，h は f, B, g の合成，z は k 次元ベクトル，t は閾値用 m 次元ベクトル．多くの場合中間層ユニット数は入力層ユニット数より少ないので m<n とする．

目的は，任意の入力 $\boldsymbol{x}$ に対して同一の出力 z を与える「代替入力」y を求めることであるが，このとき y から $\boldsymbol{x}$ が計算できないこと（ベクトルとして．また，先験的な知識では y から $\boldsymbol{x}$ は推定できないものとする），および $\boldsymbol{x}$ から y を計算するとき A, h, t を復元できないことが必要である．すなわち $\boldsymbol{x}$ から y への変換 T を

$$y = T(x, A, h, t) \qquad (2.2)$$

とするとき T に必要な条件は次の２点になる．

(1) $h(Ay+t) = h(Ax+t)$
(2) x, A, h, t について解けない

このような T が存在すれば，サーバからクライアントに T を送ったときクライアントにはＮＮの構造が伝わらず，クライアントが T で計算した y をサーバに送ったときサーバは $\boldsymbol{x}$ を知ることなく正しい出力を得られることになる．また $\boldsymbol{x}$ を直接入力する場合のＮＮの構造を変更する必要もない．

Computation by Camouflaged Data between Client and Neural Network Server

Toshiya NAKAJIMA
Fujitsu Ltd.
tossi@strad.se.fujitsu.co.jp

3．Tの構成

Tを $y=T(A)\,x$ として構成する（$T(A)$ はn×n行列）．任意の x について $Ax=AT(A)x$ により，$A=AT(A)$ に対する $T(A)$ は

$$T(A) = A^{-}A + (I_n - A^{-}A)\,C_1 \qquad (3.1)$$

となる．ここで A^{-}は A の一般逆行列[1]，I_n はn次単位行列であり，$m < n$ のとき $A^{-}A \neq I_n$．また C_i (i=1,2,..) は任意の（サイズは適切な）行列とする．

このとき(3.1)において C_1x は任意のベクトル w に置き換えられるから

$$y = A^{-}Ax + (I_n - A^{-}A)\,w \qquad (3.2)$$

とできる．したがってサーバが $A^{-}A$ をクライアントに送ればクライアントは y を計算しサーバに送ることになる．ここで w はクライアントが選ぶ（例えば $A^{-}Ax = x$ の場合には $(I_n - A^{-}A)w \neq 0$ となるようにする）．サーバは y を受け取り，x の場合と同一の計算により同一の結果（$Ay=Ax$）を得る．

以上の手順においてサーバは x を求められず，クライアントは A を求められないことが次のようにしてわかる．サーバについては，$A^{-}A$ が単位行列以外の射影行列であるため $A^{-}Ax$ に対して x は一意ではないので x を求めることはできない．一方，クライアントが A を求められないことについては以下のようになる．まず rank $A \equiv r$ (>0) として A を次式の形に分解する．

$$A = Q \begin{bmatrix} R & O \\ O & O \end{bmatrix} S^{-1} \qquad (3.3)$$

ここで Q, R, S は各々m次，r次，n次の正則行列．(3.3)は A の特異値分解を包含しているので，このような Q, R, S は必ず存在する．このとき A^{-}は

$$A^{-} = S \begin{bmatrix} R^{-1} & C_2 \\ C_3 & C_4 \end{bmatrix} Q^{-1} \qquad (3.4)$$

であるから $A^{-}A$ は

$$A^{-}A = S \begin{bmatrix} I_r & O \\ C_5 & O \end{bmatrix} S^{-1} \qquad (3.5)$$

と表わされる（C_3R は任意の行列 C_5 に置き換える）．これにより $A^{-}A$ には Q, R についての情報が含まれないため，$A^{-}A$ から A を求めることはできない．また，$A^{-}A$ が単位行列と異なる射影行列であることは容易に検証できるため，クライアントはサーバが x を求められないことを確認できる（あるいはより直接的に，サーバが S, r を送り(3.5)によってクライアントが $A^{-}A$ を計算するようにしてもよい）．

以上をまとめると，クライアント・サーバ間の情報交換手順は次のようになる．また全体構成を図1に示す．

(1) サーバ： A を Q, R, S^{-1} に分解する．
(2) サーバ： S と適当な C_5 から $A^{-}A$ を計算しクライアントに送る．
(3) クライアント： $x, A^{-}A$ と適当な w から y を計算しサーバに送る．
(4) サーバ： $h(Ay+t)$ を計算しクライアントに送る．

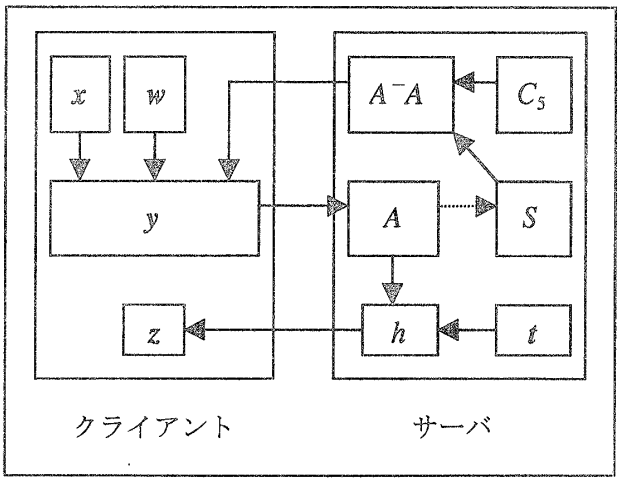

図1 クライアントとサーバの全体構成

参考文献：
1) 柳井・竹内，射影行列・一般逆行列・特異値分解，東京大学出版会，1983.