ISSN 2186-5647

－日本大学生産工学部第46回学術講演会講演概要（2013-12-7）－

P-3

線形パーセプトロンを用いた相互学習に関する研究

日大生産工 (学部)　◯ 斎藤大輔　日大生産工　原一之

1　はじめに

ニューラルネットワークはニューロンモデル (脳の神経細胞をモデル化したもの) をネットワークとしたものであり, パーセプトロンはニューラルネットワークモデルの一つである. このモデルでは, ネットワークの出力信号と正解信号 (教師信号) を比較して, 間違っている場合は結合荷重を変化させて両者を一致させる. これを学習という. 本報告では, 正解を生成する教師ネットワークを生徒ネットワークが学習することを教師あり学習, 生徒ネットワーク同士で学習することを相互学習と呼ぶ. 本研究では, パーセプトロンの線形的なモデルを用いて教師あり学習を行ない, 途中で教師ネットワークを外して相互学習を行い, 教師あり学習をどの程度行えば相互学習の効果があるかを検討する.

2　学習方法

教師ネットワークの出力を v, 結合荷重を B, 生徒ネットワークの出力を u_1, u_2, 結合荷重を J, K とする (図 1).

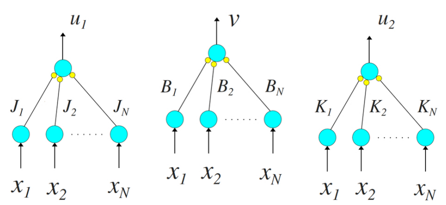

図 1 ネットワークの概要

はじめに, 生徒ネットワークは教師ネットワークの出力を見て, 自分自身の結合荷重を変化させる (教師あり学習). 誤差がある値に達したら生徒ネットワーク同士で学習を行う (相互学習). ここで, 相互学習とは, 見本となる教師ネットワークを学習対象から外し, 生徒ネットワーク同士でお互いの出力を教師と見なして, 学習を行うことをいう. 各ネットワークの出力値は, それぞれの入力と結合荷重の積の総和とする (式 (1)).

$$v = \sum_{i=1}^{N} x_i \cdot B_i \tag{1}$$

誤差は, 二乗平均誤差とする. 二乗平均誤差 ϵ_g は, 二乗誤差を次元数 N 回計算し, その平均により計算する (式 (2)).

$$\epsilon_g = \frac{1}{N}\{\sum_{i=1}^{N}\frac{1}{2}(v_i - u_i)^2\} \tag{2}$$

学習には最急降下法を用いる. 最急降下法を用いて, 教師あり学習を行うときの結合荷重の更新式を以下に示す.

$$J_i^{m+1} = J_i^{m} + \eta \cdot (v - u) \cdot x_i \tag{3}$$

ここで m は学習回数, η は学習係数, i は次元数 1〜N を指す. 教師あり学習においては, v は教師ネットワークの出力 v, u は生徒ネットワークの出力 u_1(または u_2) となる. また相互学習において, v は u_1, u は u_2(またはその逆) となる. この更新式を用いて, 教師あり学習, 相互学習それぞれを行った. 教師あり学習から相互学習へ切り替えるには, 式 (2) の二乗平均誤差を用いた. また, 教師ネットワークと生徒ネットワークを比較する一つの指標として, オーバーラップがある. オーバーラップ R は式 (4) で表され, 教師ベクトルと生徒ベクトルの

Research on the Mutual Study Using a Linear Perceptron
Daisuke SAITOH and Kazuyuki HARA

重複度合いを示すものである．ϵ_g と R, この二つの要素を比較して実験を行った．

$$R = \frac{\boldsymbol{B} \cdot \boldsymbol{J}}{|\boldsymbol{B}| \cdot |\boldsymbol{J}|} \tag{4}$$

3 結果

二乗平均誤差が 0.1, 0.3, 0.5, 0.7 の時に，教師あり学習から相互学習へ切り替えたところ，以下の結果となった．今回は生徒ネットワーク u_1 のみについて記す．ここで，相互学習に切り替える時の二乗平均誤差を ER, 相互学習開始前の ϵ_g を $\epsilon_g(0)$, 相互学習を十分に行った学習回数 $3 \times 10^4 \sim 4 \times 10^4$ における ϵ_g の平均値を $\epsilon_g(\infty)$, 相互学習開始前の R を $R(0)$, 学習回数 4×10^4 における R を $R(\infty)$ と呼ぶ．また，相互学習において学習方向 $u_1 \to u_2$, $u_2 \to u_1$ を 1 セットとする ($u_1 \to u_2$ はネットワーク u_1 がネットワーク u_2 を学習することを指す)．相互学習は 5×10^3(8 セット) 行った．

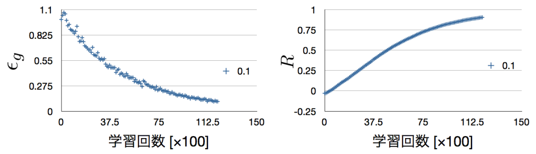

図 2 教師あり学習 (ER : 0.1)

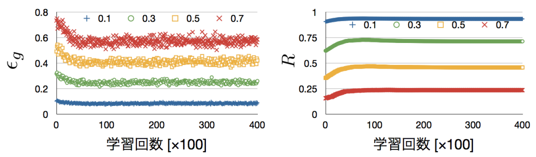

図 3 相互学習 (ER : 0.1, 0.3, 0.5, 0.7)

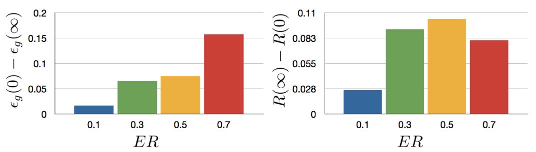

図 4 学習効果

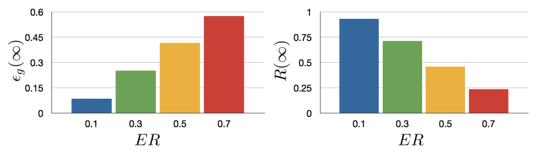

図 5 最終的な ϵ_g 及び R

表 1 ϵ_g 及び R の各値

教師あり学習前（初期値）		相互学習				
ϵ_g	R	ER	$\epsilon_g(0)$	$\epsilon_g(\infty)$	$R(0)$	$R(\infty)$
0.991877	-0.30682	0.1	0.100170	0.083389	0.905145	0.931016
		0.3	0.315018	0.250293	0.619530	0.711331
		0.5	0.489432	0.414756	0.352682	0.455715
		0.7	0.732092	0.575369	0.153492	0.233334

図 2 は教師あり学習，図 3 は相互学習の結果である．教師あり学習を行う前の ϵ_g と R, 相互学習における ϵ_g と R の初期値，収束値を表 1 に示す．全ての ER において，学習を行うことで ϵ_g が小さくなり，R は大きくなることを確認できた．このことから，教師あり学習から相互学習へ切り替えても，学習が行えることがわかった．図 3 の ϵ_g の減少，R の増加の様子を図 4 に示す．本図より，ϵ_g は ER 0.7 の時に最も減少しており，また R は ER 0.5 の時に最も増加していることがわかる．このことより，教師あり学習から相互学習に切り替える誤差によって，相互学習の動きも変化することがわかった．図 4 より，R が最も大きくなる ER 0.5 の時に，よく学習できていると考えられる．また，図 5 は最終的な $\epsilon_g(\infty)$ 及び $R(\infty)$ を比較したものである．$\epsilon_g(\infty)$ 及び $R(\infty)$ と ER は比例関係であることがわかる．

今回の結果では，教師あり学習を誤差 0.5 になるまで行えば，相互学習によって性能を向上できると考えられる．

4 今後の予定

今回の結果より，ER を 0.3〜0.5 とすると相互学習で性能が大きく向上することがわかった．今後は生徒数を増やした時の相互学習の効果を検討する予定である．

「参考文献」

1) W. H. Press , B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, 石井出 宏嘉, "*Numerical Recipes in C*", 技術評論社, (1993), p.214-217.