3F8-3

自己増殖型ニューラルネットワークに基づく持続的半教師あり学習手法

Life-long Semi-supervised Learning Algorithm
Based on Self-organizing Incremental Neural Networks

神谷 祐樹[*1] 石井 利明[*1] 長谷川 修[*2]
Youki Kamiya Toshiaki Ishii Osamu Hasegawa

[*1]東京工業大学大学院 総合理工学研究科 知能システム科学専攻
Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology

[*2]東京工業大学大学院 理工学研究科 像情報工学研究施設
Imaging Science and Engineering Laboratory, Tokyo Institute of Technology

This paper presents an incremental network for online semi-supervised classification, which is based on a self-organizing incremental neural network (SOINN). Using labeled data and a large amount of unlabeled data, the proposed semi-supervised SOINN (ssSOINN) can automatically learn the topology structure of input data distribution without any priori knowledge such as number of nodes or a good network structure; It can separate the structure into sub-structures as need arises. Experimental results we obtained for artificial and real-world data sets show that ssSOINN has superior performance for separate the data distributions with high-density overlap and that ssSOINN Classifier (S3C) is an efficient classifier.

1. はじめに

持続的に新たな学習データを追加し，膨大なデータを効果的かつ頑健に扱う学習手法の構築を行う場合，膨大なデータのオンライン学習に起因する二つの問題が考えられる．一つは学習済みの識別器を壊すことなく新たに得られたデータをいかに学習するかである．これは安定性 - 可塑性ジレンマ [1] と呼ばれ，オンライン学習手法における主要な課題の一つとされている．もう一つは膨大なデータを扱う場合に大きな問題となる，学習データへの人手によるラベル付与の負担である．この問題は，オンライン学習手法では議論が少なく，主に半教師あり学習手法の間で議論され，従来から様々な手法が提案されている [2]．

しかし，これら二つの問題を共に考慮したオンライン半教師あり学習手法はほとんど提案されていない．Incremental Growing Neural Gas (IGNG)[3] では，オンライン半教師ありクラスタリングの方法が検討されている．しかし，可視化された学習結果を一定入力毎にユーザが全て確認・修正する必要がある．また，Semi-supervised Fuzzy ARTMAP (ssFAM) と Semi-supervised Ellipsoidal ARTMAP (ssEAM) は，誤認識許容パラメータを導入し，教示ラベルを選択的に利用することで，過学習を抑制し高い汎化能力を示すオンライン学習手法である [4]．しかし，学習データの全てに教示ラベルが必要であり，ラベル付与コストについては検討されていない．

そこで本稿では，自己増殖型ニューラルネットワークに基づくオンライン半教師ありクラスタリング手法を提案する．提案手法は，事前知識（クラス数など）を必要とせず，入力パターン情報を追加的に学習し続ける．さらに，半教師あり学習の性質を持ち，ラベル付与コストが問題となる状況に対処する．

2. Semi-supervised SOINN (ssSOINN)

semi-supervised SOINN (ssSOINN) は，自己増殖型ニューラルネットワーク (SOINN)[5] の性質を継承している．そのため，ノイズへの耐性があり，オンライン学習によって入力分布の近似が可能である．しかし，SOINN は半教師あり学習を考慮していない．また，第二層のネットワークに第一層の学習結果を利用するため，追加的データに対しては二層目の再学習が必要である．さらに，学習に必要なパラメータ数が多いことが問題とされる．そこで SOINN を，(1) より少ないパラメータ数で入力の分布をオンライン学習により近似する，(2) ラベル付きデータに基づいてオーバーラップを含むクラスタを分割する，ように改良する．

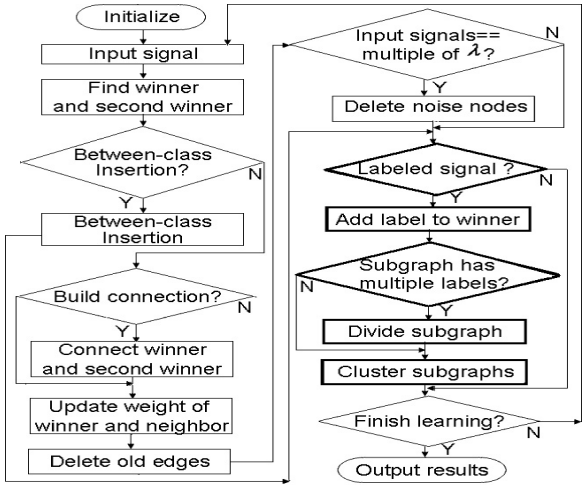

図 1: semi-supervised SOINN の学習過程

ssSOINN の学習過程を図 1 に示す．ssSOINN の学習過程は，adjusted SOINN の過程（図 1 細枠部分）とクラスタ分割過程（図 1 太枠部分）で構成される．

2.1 Adjusted SOINN

adjusted SOINN の過程では，ラベルの有無に関わらず全ての入力パターンの分布を学習する．ssSOINN では，SOINN

連絡先: 神谷 祐樹，横浜市緑区長津田町 4289-R2-52，045-924-5180，kamiya.y.ab@m.titech.ac.jp

の第一層のみを手法の基盤とすることで，オンラインクラスタリングを可能にした．また，クラスタ内のノード挿入の機能 [5] を除くことで，手法を平易にし SOINN のもつ 5 つのユーザ定義パラメータを除いた．この修正によって，ssSOINN ではクラスタ外のノード挿入の機能のみで分布の近似を行うが，その近似性能は保証される．

以下に adjusted SOINN の詳細を示す．adjusted SOINN の学習フローは，図 1 の太枠部分を除いたものである．

各記号の定義

W_i	ノード i の n 次元荷重ベクトル
A	ノード集合
N_i	ノード i と隣接するノード集合
C	エッジ集合
$age_{(i,j)}$	ノード i と j をつなぐエッジの年齢
$cl(i)$	ノード i の属するクラスタ番号

アルゴリズム 2.1: Adjusted SOINN

step0. ノード集合 A を，ランダムに選択した荷重ベクトル W を持つ二つのノード（c_1, c_2）に初期設定する．ノードの隣接関係を表すエッジ集合 C（$C \subset A \times A$）は空集合とする．

step1. 入力パターン $\xi \in R^n$ を取得する．

step2. ξ に対する勝者ノード s_1 と第二勝者ノード s_2 を探索する．I.e. $s_1 = \arg\min_{c \in A} \parallel \xi - W_c \parallel$, $s_2 = \arg\min_{c \in A \setminus \{s_1\}} \parallel \xi - W_c \parallel$. ξ とノード（s_1 または s_2）との距離が類似閾値（T_{s_1} または T_{s_2}）より大きい場合，ξ を新ノードとして A に追加する．その後，**step8.** に進む．ただし，類似閾値 T_i は以下の式で算出される．

$$T_j = \begin{cases} \max_{c \in N_j} \parallel W_j - W_c \parallel & (|N_j| \neq 0) \\ \min_{c \in A_i \setminus \{j\}} \parallel W_j - W_c \parallel & (|N_j| = 0) \end{cases} \quad (1)$$

step3. s_1 と s_2 との間にエッジが存在しなければ，新たにエッジを作成する．存在する場合はそのエッジの年齢を 0 にリセットする．

step4. s_1 に繋がる全てのエッジの年齢をインクリメントする．

step5. 勝者ノードとその勝者ノードに隣接するノードの荷重ベクトルを更新する．ただし，係数 ϵ_1 および ϵ_2 を，$\epsilon_1(t) = 1/t$, $\epsilon_2(t) = 1/100t$, また，t を該当ノードが勝者ノードに選択された回数と定義する．I.e. $\Delta W_{s_1} = \epsilon_1(t)(\xi - W_{s_1})$, $\Delta W_i = \epsilon_2(t)(\xi - W_i) \quad (\forall i \in N_{s_1})$.

step6. 事前定義した閾値 age_{dead} を超える年齢のエッジを削除する．その結果，隣接関係を持たないノードが現れた場合は該当するノードを削除する．

step7. 入力パターン数が λ の整数倍なった場合，隣接ノード数が 1 以下のノードを削除する．

step8. A 内のノードを以下の手順でクラスタリングする．ただし，あるノード a からエッジを辿ることでノード b に着くとき，a は b との経路を持つと定義する．

i. 全ノードを未分類に初期化する．$k = 0$ とする．

ii. 未分類ノード $a \in A$ の中で，旧クラスタ番号 cl^{old} が $cl^{old}(a) = k$ であるノード u を一つ選択する．$cl_i^{old}(a) = k$ に該当するノード u がなければ，未分類ノードから $u = \arg\min_{a \in A} cl^{old}(a)$ に該当するノード u を選択する．ノード u の新クラスタ番号 cl^{new} を $cl^{new}(u) = k$ として分類する．

iii. ノード u との経路を持つ A 内の全てのノードを探索し，それらのノード v をノード u と同一クラスタ（$cl^{new}(v) = cl^{new}(u)$）に分類する．

iv. 未分類ノードが残っている場合は $k = k+1$ とし（ii）に戻る．そうでなければ **step9.** に進む．

step9. 学習が十分に行われるまで，**step1.** に戻り学習を繰り返す．

2.2 クラスタ分割過程

入力パターンがラベル付きである場合，クラスタ分割過程を adjusted SOINN に続いて実行する．この過程では，入力パターンに対する勝者ノードが含まれるクラスタに関して分割の必要性を判断する．ただし本手法では，互いにエッジを辿る経路が存在するノードは同一クラスタに属すると定義する．

クラスタ分割は，教示ラベル付きの入力パターンに対する勝者ノードの属するクラスタ内に，異なるラベルをもつ他のノードが存在する場合に実行される．クラスタ分割は以下の 2 つの仮定に基づいて実行する．

仮定 1： エッジで互いに直接連結しているノードはそれ以外のノードより類似性が高い傾向にある．

仮定 2： クラスタは単峰性の分布を形成する傾向にある．

具体的には，まず当該クラスタに含まれるラベル無しノードを全て未分類に初期化する．次に仮定 1 に基づいて，分割するクラスタのラベル付ノードから並列して，それらと連結する未分類の隣接ノードを探索し探索元ノードと同一クラスタに分類する．その後探索した隣接ノードを新たな探索ノード集合とし，それらと連結する未分類の隣接ノードの分類を繰り返す．

ただし，仮定 2 の定義から，分布の谷となるノードで探索を一時停止し，単峰性の分布を重視してクラスタの分割が実行される．分布の凹凸は各ノードの近似半径 R から推測する．adjusted SOINN で生成したノードは，入力分布の密度に合わせて分布する．そのため，分布が密の部分のノードは隣接ノードとの距離が近く R 値は小さくなる．したがって，R 値が隣接するノードより小さい場合を分布の谷と推測して探索を一時停止する．また，分割したクラスタについて，以後クラスタ分割の必要性が生じた場合は，以前の分割結果を破棄して再度クラスタ分割を行う．

アルゴリズム 2.2 に ssSOINN のアルゴリズムを示す．アルゴリズムの各ステップは図 1 内の数字に対応している．

各記号の定義

アルゴリズム 2.2 で初めて用いる各記号の定義を以下に示す．前述のアルゴリズムで用いた記号の定義は同義であるため省略する．

$\|N_i\|$	ノード i の隣接ノード数
R_i	ノード i の近似半径（平均エッジ距離）
L_i	ノード i に対応するラベル

アルゴリズム 3.2: semi-supervised SOINN (ssSOINN)

step0. アルゴリズム 3.1 と同様に，ノード集合 A およびエッジ集合 $C \subset A \times A$ を初期化する．

step1. 入力パターン $\xi \in R^n$ を取得する．

step2. ξ に対して adjusted SOINN 過程をアルゴリズム 3.1 に従って実行する．一入力分の処理が終了した後 **step3.** に進む．

step3. ξ にラベルが付与されていない場合，**step7.** に進む．

step4. ξ のラベル L_ξ を s_1 に付与する： $L_{s_1} = L_\xi$.

step5. s_1 とエッジを辿る経路が存在する全ノードについて，ラベルが付与されていないかまたは，s_1 と同一ラベルが付与されている場合，**step7.** に進む.

step6. s_1 との経路が存在する全ノード（クラスタ D）について互いに異なるラベルを持つノードが存在する場合，クラスタ D を以下の手順で分割する.
(a) クラスタ D 内の全ノードを未分類に，探索回数を $k = 0$ に，終端ノード集合 B_{end} を空集合に初期化する．また探索ノード集合 B_k をクラスタ D 内の全ラベル付ノードに初期化する．I.e., $B_k = B_0 = \{a|\forall a(a \in D, L_a \neq \emptyset)\}$. また，クラスタ D 内の全ノードについて，近似半径 R を算出する．I.e., $R_i = \frac{1}{|N_i|}\sum_{c \in N_i} \parallel W_i - W_c \parallel$.
(b) ラベル付ノードのクラスタ番号を各々異なる番号に変更する．I.e. $cl(u_i) = cl^{new}$, $cl(u_i) \neq cl(u_j)$ $[cl^{new} \neq cl(a)$ $(\forall a \in A \setminus B_0)$, $1 \leq i \leq |B_0|$, $1 \leq j \leq |B_0|$, $i \neq j]$.
(c) 近似半径 R_a が隣接する全ての未分類ノードの R 値より小さいノード $a \in B_k$ を B_{end} に移動する.
(d) B_k 内のノードに隣接する未分類ノードを全て探索し B_{k+1} とする．I.e., $B_{k+1} = \{a|\forall a \in D, a \in N_u \ \ (\forall u \in B_k)\}$.
B_{k+1} 内の全ノードについて，各々の探索元ノードの持つクラスタ番号を付与し分類する．ただし二つ以上のノードに探索されたノードは，全てのクラスタ番号を付与して分類済とする.
(e) $B_{k+1} \neq \emptyset$ であれば $k = k + 1$ として (c) に戻る．そうでなければ $B_k = B_{end}$ とする.
(f) B_k 内のノードに隣り合う未分類ノードを重複を許して全て探索し，各々の探索元ノードの持つクラスタ番号を付与し分類する．探索した全ノードを B_{k+1} とする.
(g) D 内に未分類ノードが存在する場合，$k = k + 1$ として (f) に戻る.

step7. 新たな入力が得られる限り **step1.** に戻り，オンライン半教師ありクラスタリングを行う.

3. 汎化能力における他手法との比較実験

2 次元人工データセットと UCI レポジトリ (http://www.ics.uci.edu/~mlearn/MLRepository.html) のデータセットを用いて，他のオンライン学習手法および半教師あり学習手法と提案手法の汎化性能について比較実験を行った.

3.1 データセットと評価方法

提案手法はニューラルネットワークに基づくオンライン学習を行うため，Le らの行ったオンライン学習手法（GSC，GNG，ssEAM，ssFAM）の比較実験 [7] と同様の実験を行い，汎化性能を比較する．本実験では Le らの実験で用いられたデータセット 7 種類のうち 5 種類を用いた．Cancer データセットは Le らの比較実験において全ての手法の識別率が非常に高く，有意差が得難いため採用していない．また，Abalone データセットは，参照した実験において第一特徴の性別を実数値に変換した値が明記されていなかったため用いていない.

G4 データセット ：4 クラスの分布．各クラスの分布は単一の正規分布であり，同一の生起確率である．オーバーラップを含み，ベイズ誤り確率の程度によって G4LO(5%), G4ME(15%), G4HI(40%) と呼ぶ 3 通りのデータセットを定義する．以上に基づいて，学習データ 500，バリデーションデータ 5,000，テストデータ 5,000 を生成する．全学習サンプルに対するラベル付与率は 10%と 20%の 2 通りに設定した．図 2 上段に生成した学習データの例を示す．提案手法はこれら 3 種のデータセットについて，各クラスの分布をオンライン半教師あり学習によって分割することができる（図 2 下段）.

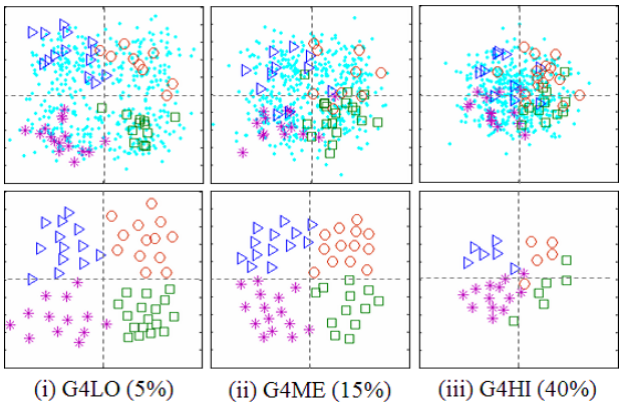

図 2: 人工データセット（G4）を用いた場合の提案手法の学習結果の一例：（上段）学習データセット．10%のラベル付サンプル（○，□など）と 90%のラベル無しサンプルを含む.（下段）上段のデータセットに対する提案手法の学習結果.

Glass, Pima Indian Diabetes データセット ：UCI レポジトリ内のデータセット．スペースの都合上，詳細は関連する文献を参照されたい．このサンプルを学習データ，バリデーションデータ，テストデータに分配する割合は Le らの実験と同様に関連文献 [8] の設定を用いた．学習データのラベル付与率は 10%と 20%の 2 通りに設定した.

以上のデータセットは，Le らの実験と同様の定義であるが，人工データの値やデータセットの分配などの理由で完全には一致しない．また，我々の実験ではラベル付サンプルの選択によってもデータセットが異なる．したがって，この定義に基づいたデータセットに対する実験を 10 回行い，それらの結果の平均値を算出して Le らの実験結果と比較した.

各データセットに対する提案手法の汎化性能の評価方法についても同様に，Le らの実験に基づいて設定した．まず，パラメータ選択による結果の偏りを防ぐため，多数のパラメータセットにおいて提案手法に学習データを学習させた．パラメータ (λ, age_{dead}) はそれぞれ，$[100 : 50 : 300]$, $[100 : 300 : 1000]$, の値を取るように設定した．さらに，オンライン学習では学習サンプルの入力順序によって結果が異なるため，100 通りの異なる入力順序を設定した．したがって，各データセットについて 2,000 通りの識別器が形成される．なお，学習サンプルの入力回数は合計 100,000 回に設定した．その後，2,000 個の識別器の中から，バリデーションデータに対する識別性能の高い 100 モデルを選択する．そしてそれら 100 モデルのテストデータに対する識別性能を算出し，その平均値を汎化性能とした.

前述の通り，ラベル付サンプルの選択などによって学習結果は変化すると予測される．そのため，以上の操作をデータセットを 10 回変化させて同様に行い，汎化性能の平均を算出した.

表 1: 5 種のデータを用いた汎化性能の比較（太字は各データセットに対する最小値）

		PIC Test for 100 classifiers		PIC Test for 100 best classifiers					
		1-NN		Proposed		GCS	GNG	ssEAM	ssFAM
		10%	20%	10%	20%	with 100% labeled data [7]			
G4LO	avg.	8.82	8.12	7.09	**6.76**	10.48	10.4	10.62	10.38
	std.	1.9167	1.1488	0.4958	0.3993	0.09211	0.07888	**0.06457**	0.10203
G4ME	avg.	22.44	21.87	19.09	**17.92**	25.36	24.74	26	25.2
	std.	3.1218	2.0340	0.9127	0.7471	0.14061	**0.12173**	0.26508	0.25875
G4HI	avg.	50.07	49.03	42.33	**40.42**	42.16	41.58	42.4	42.06
	std.	4.0464	2.6683	1.1799	0.9599	0.21283	**0.1378**	0.34975	0.25216
GLASS	avg.	49.76	42.45	47.33	42.11	35.8491	39.6226	39.6226	**33.9623**
	std.	10.6437	8.7528	6.3257	5.7478	**1.1024**	2.4243	1.8443	1.3564
PIMA	avg.	34.76	33.00	30.84	30.00	25.8333	26.3542	**25**	**25**
	std.	4.6356	4.0069	2.8882	2.8713	**1.0586**	1.0824	1.0232	1.6368

比較する手法には，Le らの実験で用いられたオンライン学習手法（GCS，GNG，ssEAM，ssFAM）に加えて，ラベル付サンプルのみを用いて構成した最近傍識別器（1-NN）[6] と比較し，提案手法の半教師あり学習の効果を確認した．最近傍識別器については各データセットを 100 通りに変化させてそれら全てのテストデータに対する識別性能を算出しその平均値を汎化性能とした．

3.2 結果と考察

表 1 に，各データセットに対する各手法を用いた場合の汎化性能を示す．表 1 の各値は，誤識別率（*PIC: percent incorrect classification*）について 100 モデルの平均値と標準偏差を表している．ラベル付サンプルのみを用いた最近傍識別器（1-NN）に比べて，提案手法の汎化性能は全ての場合において上回っている．この結果から，ラベル無しサンプルを利用する半教師あり学習の性質が提案手法に備わっているといえる．

また他のオンライン学習手法が教師あり学習（ラベル付データ 100%）であるのにも関わらず，G4 データセットに対しては提案手法が他手法に比べて最も低い誤識別率を示した．ただし，オーバーラップの少ない場合はラベル付サンプルのみ用いた最近傍識別器でも低い誤識別率を示しており，最近傍法が有効となるデータセットであるとも考えられる．しかし，G4HI においては，最近傍識別器が非常に高い誤識別率を示したのに対し，提案手法はラベル無しサンプルの情報を用いて 7.67–8.61%程度誤識別率を減少させることができている．以上から，クラスタ分割過程において定義した仮定が理想的に満たされる場合，本手法は十分な性能を示すと考えられる．

Glass および Pima データセットでは，他のオンライン教師あり学習手法（ラベル付与 100%）に比べて汎化性能は劣る結果となった．しかし，近年のオンライン学習でない半教師あり手法と比較した場合，Glass データセットに対する追加的半教師あり学習手法の識別率は，ラベル付与率 10%のときの PIC: 55.48–59.91，ラベル付与率 20%のときの PIC: 53.51–59.18，と報告されている [9]．また Pima データセットについては，Dimitriadou らが提案した手法において，PIC: 30.8 ± 2.6（ラベル付与率 10%）と報告されている [10]．提案手法は，これらの手法と同等もしくはそれ以上の汎化性能を示している．したがって，オンライン半教師あり学習が必要となる場合に提案手法は有効な手法であるといえる．

4. むすび

本稿では，オンライン半教師ありクラスタリング手法を提案した．提案手法は，オンライン学習の性質を持ち，膨大なパターン情報を追加的に学習可能である．また，半教師あり学習の性質を持ち，ラベル付与コストが問題となる状況にに対処する．5 種のデータセットを用いた比較実験により，提案手法が従来の半教師あり学習手法と同等かそれ以上の汎化性能をオンライン学習によって達成することを示した．

謝辞

本研究の実施にあたり NEDO 産業技術研究助成事業から支援を頂きました。記して感謝いたします。

参考文献

[1] Grossberg, S.: Nonlinear Neural Networks: Principles, Mechanisms, and Architectures, *Neural Networks*, Vol. 1, pp. 17–61 (1988).

[2] Zhu, X.: *Semi-Supervised Learning Literature Survey*, Tech. Rep., (2006) http://www.cs.wisc.edu/jerryzhu/research/ssl/semireview.html.

[3] Prudent, Y. and Ennaji, A.: An Incremental Growing Neural Gas Learns Topologies, *in Proc. IJCNN '05*, pp. 1211–1216, (2005).

[4] Anagnostopoulos, G.C., et. al.: Exemplar-Based Pattern Recognition via Semi-Supervised Learning, *in Proc. IJCNN '03*, Vol. 3, pp. 1350–1356, (2003).

[5] Shen, F. and Hasegawa, O.: An Incremental Network for On-line Unsupervised Classification and Topology Learning, *Neural Networks*, 19(1), pp. 90–106 (2006).

[6] Cover, T. and Hart, P.: Nearest Neighbor Pattern Classification, *IEEE Trans. on Information Theory*, Vol. IT-13, No. 1, pp. 21–27, (1967).

[7] Le, Q., et. al.: An Experimental Comparison of Semi-Supervised Artmap Architectures, GCS and GNG Classifiers, *in Proc. IJCNN '05*, pp. 3121–3126, (2005).

[8] Hamker, F. and Heinke, D.: *Implementation and Comparison of Growing Neural Gas, Growing Cell Structures and Fuzzy ARTMAP*, Schriftenreihe des FG Neuroinformatik der TU Ilmenau, Tech. Rep. 1/97, (1997).

[9] Zhang, R. and Rudnicky, A.: A New Data Selection Principle for Semi-Supervised Incremental Learning, *in Proc. ICPR '06*, Vol. 2, pp. 780–783, (2006).

[10] Dimitriadou, E., Weingessel, A. and Hornik, K.: A Mixed Ensemble Approach for the Semi-Supervised Problem, *in Proc. ICANN '02*, pp. 571–576, (2002).