

知識処理向きニューラルネットワーク LISM の 学習・想起アルゴリズムとルール抽出法

阿部 一裕

三菱電機（株）中央研究所

1992年1月16日

本稿では、部分的な情報を取り扱うために、不定値を含む入力パターンを学習・想起する階層型ニューラルネットワーク LISM(Learning In Subspace Method Network) のネットワークアーキテクチャと、その学習・想起アルゴリズムである 部分空間学習法, 部分空間想起法 について述べる。

また部分空間学習法により必要な教師事例の数とネットワークの中間層ノードの数を減少できること、部分空間想起法をもちいることで LISM が獲得した知識を決定木の形で取り出せることをしめす。

A Learning and Recall Algorithm of LISM and Rule Extraction from LISM

Kazuhiro Abe

Mitsubishi Electric Corporation
Central Research Laboratory

8-1-1 Tsukaguchi-Honmachi, Amagasaki, Hyogo, 661, Japan
E-mail: abek@sys.crl.melco.co.jp

In this paper I present the architecture of Multiple Layer Neural Network LISM(Learning In Subspace Method Network). I also describe its learning and recall algorithm - Learning In Subspace Method and Recall In Subspace Method - . In order to process partial information, LISM deals with patterns containing don't care values.

Then I show that Learning in Subspace Method can reduce the number of training set and the number of hidden layer nodes, and Recall in Subspace Method can extract knowledge from LISM in the form of decision trees.

1 はじめに

エキスパートシステム構築時における知識獲得の困難さを緩和するために、ルールインダクション、事例ベース推論、ニューラルネットワークなど領域専門家の判断事例を利用した推論方式が注目されている。

筆者らは、ニューラルネットワークの持つ、柔軟で高速な連想検索能力、開発・変更の容易性、ノイズへの頑健性に着目し、推論機構としてニューラルネットワークをもちいたエキスパートシステムの構築をおこなった [1]。また [2] では、ルールから生成した教師事例を学習することによりルールの形式で記述された知識をネットワークに獲得させる方法を提案した。

[1] のシステムでは、線形加算型ノード¹から構成された階層型ネットワークにバックプロパゲーション法 [3] をもちいて学習をおこなった。

しかし

- a. 学習に長時間を要する。
- b. 追加学習時にすべての教師事例を再学習する必要がある。
- c. 学習に必要な中間層ノード数をあらかじめ決める基準がない。
- d. 部分的な情報から学習・想起をおこなうことができない。
- e. ネットワークが獲得した知識をルールなどの明示的な形でとりだすことが困難である。

などニューラルネットワークを推論機構としてもちいる上で問題点があった。

学習速度が速く、追加学習時には新しい教師事例のみを学習するだけでよく、中間層ノードの数をあらかじめ決めておく必要がないニューラルネットワークとして RCE ネットワーク [4] がある。RCE ネットワークは、距離型ノード²から構成された階層型ネットワークであり、中間層ノ

ードの数を動的に増やしていく学習アルゴリズムをもちいている。

しかし RCE ネットワークにおいても上記の問題点 d,e は解決されていない。

従来提案されてきた階層型ニューラルネットワークでは、学習・想起時においてネットワークに提示する入力パターンの値を全て指定する必要がある、パターンの一部の値が不定である部分的な情報を学習・想起することができない。問題点 d,e は、この制約に起因する部分が多いと考えられる。

本研究は上記問題点 a～e を解決することを目的としておこなったものであり、距離型ノードから構成された階層型ネットワークに中間層ノードの数を動的に増やしていく学習アルゴリズムを採用することで問題点 a,b,c に対処し、さらに不定値を含む入力パターンを学習・想起する機構を設けることで問題点 d,e を解決することを試みた。

本稿では、不定値を含む入力パターンを学習・想起するために、入力層ノードがつくる状態空間の部分空間での距離の関数を出力値とするノードを中間層にもちいた階層型ニューラルネットワーク LISM (Learning In Subspace Method Network) のネットワークアーキテクチャについてのべる。また状態空間の部分空間上で学習・想起をおこなうアルゴリズムである部分空間学習法と部分空間想起法を説明する。そして部分空間学習をおこなうことにより必要な教師事例の数とネットワークの中間層ノードの数を減少できること、部分空間想起をもちいることで LISM が獲得した知識を決定木の形で取り出せることをしめす。

2 状態空間上での学習・想起

本章では、従来提案されてきた階層型ニューラルネットワークが状態空間上でおこなう学習・想起（全空間学習・想起）と、その問題点を説明する。そして、これらの問題点を解決するために LISM がおこなう学習・想起（部分空間学習・想起）をしめす。

¹入力信号を $\vec{I}$ 、信号 I_i の荷重を W_i とした時、出力値が $\sum W_i I_i$ の関数となるノード

²中心ベクトルを $\vec{X}$ とした時、出力値が $\sum (X_i - I_i)^n$ の関数となるノード

階層型ニューラルネットワークをエキスパートシステムの推論機構として使用する場合、操作対象の状態を属性値の組で表現し、これを符号化したものを入力パターンとし、対象に施す操作を符号化したものを出力パターンとする。

学習時には、教師事例（操作対象の状態を表す属性値の組と、それに対応した操作の対）を符号化した入出力パターンを学習することで知識を獲得する。実行時には、操作対象の状態を符号化した入力パターンから出力パターンを想起し、これを復号することで対象に施す操作を決定する。

対象の状態を状態空間上の1点、対象に加える操作をひとつのクラスと考えると、操作対象の状態におうじて施す操作を決定する動作は、状態空間上の1点が属するクラスを分類する問題を解くことに相当する。

操作対象を特徴付ける属性として A_1, A_2 のふたつの属性が、対象に施す操作を表すクラスとして $\square, \triangle$ のふたつのクラスがあり、教師事例を図1に示すようにあたえた場合を考える。各教師事例における操作対象の状態は、 A_1, A_2 がつくる状態空間上の1点として、操作は $\square, \triangle$ により表現する。

2.1 全空間学習・想起

従来提案されてきた階層型ニューラルネットワークでは、学習・想起時に入力パターンの値をすべて指定する必要がある。すなわち教師事例はすべての入力層ノードがつくる全状態空間上の点とその点が属するクラスの対の形式で与えられ、全状態空間上で学習・想起をおこなう。

図1に示した教師事例を階層型ニューラルネットワークに学習させた場合を考える。パーセプトロンのような線形加算型のノードを使用するネットワークでは、各中間層ノードが状態空間上の1つの超平面に相当し、状態空間内で超平面で囲むように各クラスの領域を形成していく。またRCEネットワーク[4]、RBFネットワーク[5]などの距離型ノードをもちいるネットワークでは、各中間層ノードが状態空間上の1つの超球に相当し、図2に示したように超球（円）の集合

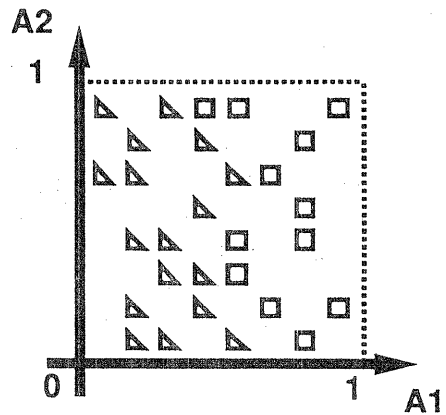


図 1: 教師事例

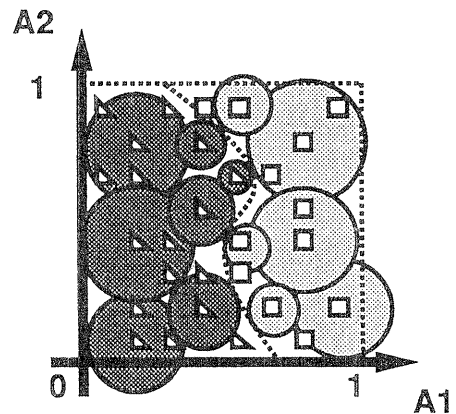


図 2: 全空間学習

で領域を形成していく。

エキスパートシステムが対象とする問題では、操作対象のあらゆる状態において施す操作を決定するために、すべての属性の値が必要となるわけではない。また、ある状態においては一部の属性のみから操作が決定できることが知識として与えられる場合も少なくない。

例えば、図1の例で、 A_1 が0または1の近傍の領域では、 A_2 に関わらずクラスが決定できることがあらかじめわかっているような場合である。

従来提案されてきた階層型ニューラルネットワークでは、入力パターンとして値がすべて定まっているパターンしかとれない。そのためクラスを

分類するのに不必要な属性があった場合でも、不必要な属性の値を指定する必要がある。ある属性が分類に影響しないことを表わすためには、その属性の値を変化させてもクラスが変わらないような教師事例を複数個用意する不自然な形で表現しなければならない。また教師事例が増加すると学習に要する時間も長くなる。

さらに距離型ノードを用いたネットワークでは不必要な属性がつくる軸方向にも領域を超球で被わなければならないので、中間層ノードが多数必要となる。

クラス分類に不必要な属性の数が増えるにつれ学習に必要な教師事例、中間層ノードの数も増加していく。

図1の例では、 A_1 が0または1の近傍でも A_1 のひとつの値について A_2 の値を変化させた複数個の教師事例を用意する必要がある。また図2に示めたように A_2 軸方向に変化させた領域を超球で被うために多くの中間層ノードが必要となる。

また想起時においても入力パターンの全ての値を確定する必要があるので、一部の属性だけしか値がわからない場合に値がわかっている属性だけから想起するような、部分的な情報の取り扱いができない。

2.2 部分空間学習・想起

すべての入力層ノードがつくる全状態空間上の局部において一部の属性の値のみから分類できることが知識として与えられている場合を考える。

このような局部において分類をおこなうのに必要な属性がつくる部分空間内で学習可能ならば、分類に必要な属性のみをもちいた自然な形式で教師事例を表現できる。また教師事例の数、中間層ノードの数も減少できる。

例えば、図3は図1に示される教師事例を属性 A_1 がつくる部分空間に射影したものである。 A_1 が0または1の近傍の領域では、部分空間上で教師事例を表すことにより、分類に必要な属性 (A_1) のみをもちいて、もとの全空間上でよりも少ない

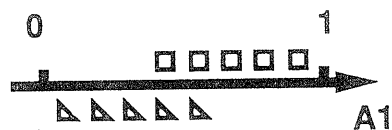


図 3: 部分空間に射影した教師事例



図 4: 部分空間学習

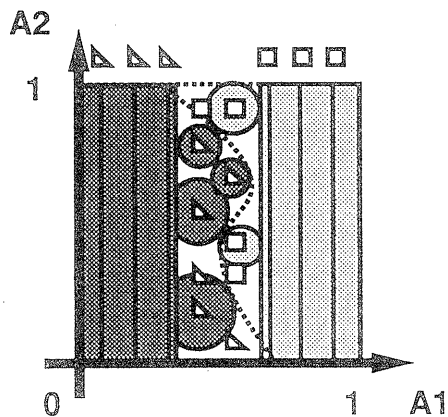


図 5: 全状態空間上でみた部分空間学習

数の教師事例で表現できる。

また属性 A_1 がつくる部分空間上で、距離型ノードで構成したネットワークをもちいて、超球がクラスの領域を形成していく学習をおこなえば、図4に示すように A_1 が0または1の近傍を部分空間内の超球（線分）で各クラスの領域を形成することができる。

部分空間内の超球は属性 A_2 方向に関しては不定であるので、全状態空間上で考えると、図5に示すように超円柱（長方形）で領域を形成することになる。

超円柱はクラス分類に不必要な属性方向に関しては任意の値をとるので、同じ領域を超球より

も少ない数で形成できる。そのため超球、超円柱に相当する中間層ノードの数を減少できる。

特定の属性の値のみ指定し、他の属性の値は不定とした教師事例は不定値を含む入力パターンとなる。これは全空間上において超立方体として表現される。

図5の例では、 A_1 が0または1の近傍にある A_2 方向に不定の教師事例は、全状態空間上の超立方体（直線）となる。入力パターンとしては $(a1, \#)$ （ $\#$ は不定値を表す記号）で表現する。不定値（ $\#$ ）を含む事例、パターンのことをそれぞれ汎化事例、汎化パターンと呼ぶ。

A_1 の値が0.5付近では、属するクラスを決定するために A_1, A_2 のふたつの属性が必要となるので全状態空間上の超球で領域を形成する。

知識処理をおこなう場合、各軸方向は属性の値に相当するので状態空間は等方的でない。したがってLISMでは超球、超円柱のかわりに、超楕円、超楕円柱をもちいて領域を形成する。

上記のように状態空間上の局部、局部により考慮する属性を変えることができ、考慮する属性のみをもちいて学習する方法を部分空間学習法と呼ぶ。部分空間学習法を全ての属性がつくる全状態空間上で見ると、クラスに属する超平面、点を含み、他のクラスに属する超平面、点を含まないように超楕円、超楕円柱を配置し各クラスの領域を形成していく。

また想起時に特定の属性にのみ着目した分類をおこなう場合は、学習時に配置された各クラスの領域を形成する超楕円、超楕円柱を考慮する属性がつくる部分空間上に射影し、その空間内でどのクラスに属するかを分類する。

3 LISMのネットワークアーキテクチャと学習・想起アルゴリズム

LISM(Learning In Subspace Method Network) は、3層の階層型ニューラルネットワークであり、部分空間内での距離の関数を出力値とするノードを中間層にもつ。本章では、LISMのネットワークアーキテクチャと、LISMをもちいた部分空間学習、部分空間想起のアルゴリズム

についてのべる。

3.1 ネットワークアーキテクチャ

LISMは3層の順伝搬階層型ネットワークであり、入力層、中間層、出力層の3層から構成される（図6）。

入力層は操作対象を特徴付ける属性の値を符号化したパターンを受け付ける。各ノードは入力値として0～1の実数、または $\#$ （不定値をあらわす記号）をとる。出力値は入力値と同じ値である。

中間層はクラスの状態空間上での特徴を記憶する。中間層のひとつのノードは図5などにしめた状態空間上の超楕円、超楕円柱のひとつに相当する。

中間層の各ノードは考慮する属性をあらわす入力層ノードとのみ結線している。また考慮する属性がつくる部分空間上で中心ベクトル $\vec{X}$ 、規格化ベクトル $\vec{\lambda}$ を記憶している。出力値は(1)式の値である。³

$$F_j = \exp\left(-\sum_{i \in \text{subspace}} \frac{(A_i - X_i)^2}{\lambda_i^2}\right) \quad (1)$$

和は結線している入力層ノードで、値が不定値でないノードの値から取る。

幾何学的考察から明らかなように、各中間層ノードの出力値は状態空間上の超楕円、超楕円柱の表面上で値が等しくなる。図5にしめた超楕円（円）、超楕円柱（長方形）は各中間層ノードの等出力値面である。

図6の例では、中間層ノード F_5 は、入力層ノード A_2, A_3, A_4 がつくる部分空間上で、中心ベクトル $\vec{X}_5 = (X_{52}, X_{53}, X_{54})$ 、規格化ベクトル $\vec{\lambda}_5 = (\lambda_{52}, \lambda_{53}, \lambda_{54})$ を記憶している。また入力パターンが $(\#, \#, a_3, a_4)$ の時の F_5 の出力値は

$$F_5 = \exp\left(-\left\{\frac{(a_3 - X_{53})^2}{\lambda_{53}^2} + \frac{(a_4 - X_{54})^2}{\lambda_{54}^2}\right\}\right) \quad (2)$$

となる。

³ここでは中間層ノードの出力値を計算する関数としてexpを採用したが、 $f(0)=1, f(-\infty) \rightarrow 0$ となる単調減少関数ならばどのようなものでよい。

出力層のひとつのノードは、入力パターンのひとつのクラスに相当する。出力層ノードは、クラスの領域を構成する超楕円、超楕円柱に相当する中間層ノードと接続している。出力値は(3)式の値である。

$$C_k = \max(F_j) \quad (3)$$

中間層ノードの出力値は、入力ベクトル $\vec{A}$ が中心ベクトル $\vec{X}_j$ から離れるにつれ減衰する。図7に示めすように異なるクラスに属する中間層ノード間で出力値が等しくなる場所でクラスの境界面を形成する。

3.2 部分空間想起法

入力ベクトル $\vec{A}$ が与えられた場合、入力層から出力層にむかって信号を1回流し、(1)式、(3)式に従い出力層の各ノードの値を決める。

3.3 部分空間学習法

LISM の学習アルゴリズムは教師事例として、“入力パターン $(A_1, A_2, \dots, A_m)$ が、クラス C_n に属する” が与えられたときに、

$$C_n \geq TH_{out} \quad (4)$$

$$C_k \leq C_n - TH_{dif} \quad (k \neq n) \quad (5)$$

を満たす。

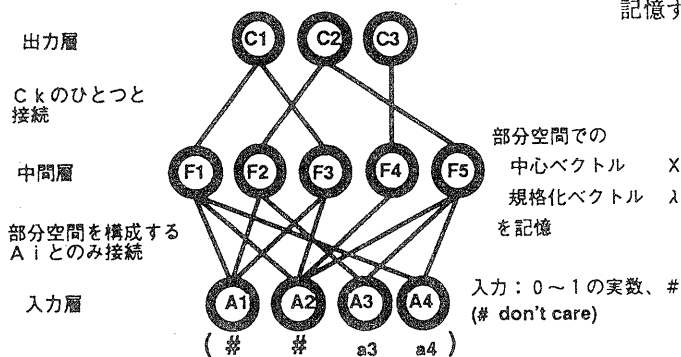


図 6: LISM のネットワークアーキテクチャ

(4) 式、(5) 式の全状態空間上での幾何学的意味についてのべる。

- (4) 式: クラス C_n に属する超平面・点は、クラス C_n の領域を構成する中間層の出力値 TH_{out} の等出力値面が囲む領域 (超楕円・超楕円柱) の少なくともひとつの内部に含まれる。
- (5) 式: クラス C_n に属する超平面・点は、クラス C_k ($k \neq n$) の領域を構成する中間層の出力値 $C_n - TH_{dif}$ の等出力値面が囲む領域 (超楕円、超楕円柱) のすべての外部にある。

学習過程は、ひとつの教師事例について中間層ノード組み込み過程、規格化ベクトル $\vec{\lambda}$ の変更過程の順におこなう。LISM の初期状態では中間層にはノードがなく、学習が進むにつれ中間層にノードが組み込まれていく。また学習中に入力層ノード、出力層ノードの数を増やしていくこともできる。

◦ 中間層ノード組み込み過程

(4) 式を保証する動作である。

$C_n \leq TH_{out}$ だった場合、中間層にノードを組み込む。組み込んだ中間層ノードは、入力層とは値が不定値でないノードとのみ結線し、出力層とはノード C_n とのみ結線する。そして値が不定値でない入力層ノードがつくる部分空間内で、中心ベクトル $\vec{X} = \vec{A}_i$ (入力ベクトルの部分ベクトル) と、規格化ベクトル $\vec{\lambda}$ (定ベクトル) を記憶する。

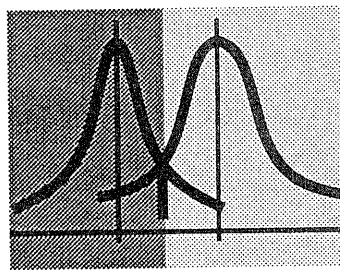


図 7: クラスの境界面

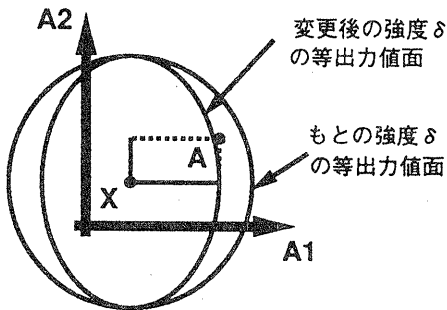


図 8: 規格化ベクトル $\vec{\lambda}$ の変更法

● 規格化ベクトル変更過程

(5) 式を保証する動作である。

$C_k (k \neq n)$ に接続されている中間層ノード F_j について、中間層ノード組み込み過程で

a. ノードを組み込んだ場合

$$F_j \leq 1 - TH_{dif} \quad (6)$$

b. 組み込まなかった場合

$$F_j \leq Cn - TH_{dif} \quad (7)$$

となるように (6) 式, (7) 式を満たさない F_j の規格化ベクトル $\vec{\lambda}$ を調節する。

F_j の値を δ 以下にする操作は、中間層のノードと結合しており、かつ値が不定値でない入力層ノード A_i に対して、 $|A_i - X_i|$ が最大となる i を求める。(図 8 の場合 $|A_1 - X_1|$ が最大)

$$\exp\left(-\frac{(A_i - X_i)^2}{\lambda_i^2}\right) = \delta \quad (8)$$

を λ_i についてとき、この値を新しい λ_i として記憶する。

4 LISM からルールの抽出法

LISM は不定値を含む入力パターンから出力パターンを想起でき、部分的な情報が分類におよ

はす影響を検出することができる。本章では、この機能を利用し LISM の獲得した知識を、決定木の形で抽出する方法をしめす。

学習後の LISM に入力パターン $\vec{A} = (A_1, \dots, A_m)$ を与えた時の、出力層の各ノードの出力値を $C_1, \dots, C_n$ とする。このとき、入力パターン $\vec{A}$ がクラス C_n に属する確率を

$$Pr(C_n|\vec{A}) = \frac{C_n}{\sum C_k} \quad (9)$$

と考える。

決定木の生成には、(9) 式を使用し、ID3 でもちいられたトップダウン的な分割統治の方法 [6] を使用する。

属性 A_i は値として $\{A_{i1}, A_{i2}, \dots, A_{iL}\}$ を取るものとする。

決定木の根部においては、ひとつの属性のみ値を設定し他の属性の値は不定値とした入力パターンを LISM に提示する。出力パターンから (9) 式を計算し各属性の平均相互情報量を求める。そして平均相互情報量が最大の属性を根部における分類属性として選択する。

平均相互情報量は、(10) 式で定義される。

$$I(A_i) = -\sum_k Pr(C_k) \log(Pr(C_k)) + \sum_l Pr(A_{il}) \sum_k Pr(C_k|A_{il}) \log(Pr(C_k|A_{il})) \quad (10)$$

すべての属性について第 1 項は同じ値となるので、第 2 項が最大となる属性を選択すればよい。また属性 A_i が属性値として L 個の値をとるとすると、

$$Pr(A_{il}) = \frac{1}{L} \quad (11)$$

とし、(10) 式の第 2 項を計算する。

決定木の節では、根からその節にいたるまでに分類に使われていない属性の中から、分類属性をひとつ決定する。候補の中のひとつの属性の値と、その節にいたる前に使用された属性の値を設定し、他の属性の値は不定値とした入力パターンを LISM に提示し、根部と同様の手続きから、その節での分類属性を選択する。

分類属性を選択していく過程で、ある入力パターンに対して (4) 式、(5) 式が満たされた場合、クラスが決定されたと考え、それ以上の分類をおこなわない。

属性値が連続値をとる場合 (9) 式を確率密度関数と考え、しきい値に関しては、例えば (9) 式が極大、極小をとる値の中間値をしきい値とする方法などが考えられる。

5 実験結果

本章では、多重 Exclusive OR 問題と、ID3 に関する文献 [6] で例題として用いられている Saturday Mornings Classification 問題を LISM に適用した結果についてのべる。

5.1 多重 Exclusive OR 問題

Exclusive OR 関数を 3 重に用いた問題である。入力属性として、 $A_1, A_2, A_3, A_4, A_5, A_6$ の 6 つの属性があり、属性値として 0 または 1 をとる。分類するクラスは C_1, C_2 のふたつがある。入力パターンは図 9 の規則にしたがい分類される。

```

if      xor( $A_1, A_2$ ) then  $C_1$ 
else if xor( $A_3, A_4$ ) then  $C_2$ 
else if xor( $A_5, A_6$ ) then  $C_1$ 
else       $C_2$ 

```

図 9: 多重 Exclusive OR 問題

全ての属性の値を指定した $2^6 = 64$ 通りの入力パターンに対して図 9 の規則からクラスを決定した教師事例を LISM に学習させた (事例学習)。

図 9 の規則では、 $xor(A_1, A_2)$ が真の領域では A_1, A_2 のみから、 $xor(A_1, A_2)$ が偽、 $xor(A_3, A_4)$ が真の領域では A_1, A_2, A_3, A_4 のみから入力パターンは分類される。これらの知識を利用できるならば、クラス分類に不必要な属性を不定値とすることで、64 個の事例は 22 個の汎化事例で表現できる。これらの事例を LISM に学習させた (汎化事例学習)。

比較のため 64 個の事例を BP 法で学習させる

こともおこなった。

上記各場合の教師事例数、中間ノード数、学習回数を表 1 にしめす。なお BP 法では中間層ノードの数が 8 個より少ない場合、学習が収束しなかった。

また事例学習、汎化事例学習をおこなった LISM から、第 4 章でしめした方法をもちいて決定木を抽出した。多重 Exclusive OR 問題の場合、ひとつの属性の値を定めただけでは、(10) 式の値が全て同じになるので、ふたつの属性の組合せについて (10) 式をもとめた。

事例学習、汎化事例学習をおこなったどちらの場合からも図 10 にしめした決定木が抽出できた。

表 1 の結果より、汎化事例学習をおこなうことにより事例学習をおこなう場合と比較し教師事例数、中間層ノード数を減少できることがわかる。ただし XOR 関数は 1bit 値が異なると他のクラスとなるので多くの中間層ノードが必要となった。

また図 10 より LISM が学習した図 9 の知識を抽出できることがわかる。

表 1: 多重 Exclusive OR 問題実行結果

学習法	教師事例数	中間層ノード数	学習回数
汎化事例学習	22	21	3
事例学習	64	51	3
BP 学習	64	8	276
		20	222
		50	231

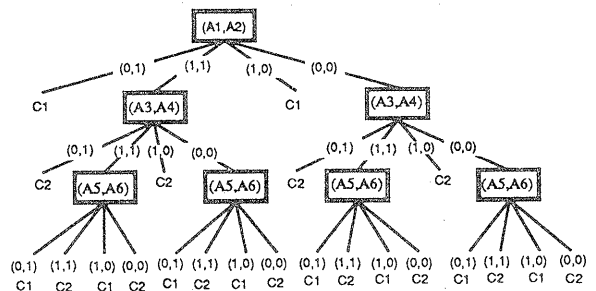


図 10: xor 問題を学習した LISM から抽出した決定木

5.2 Saturday Mornings Classification 問題

ID3 に関する文献 [6] で例題として用いられている問題である。Outlook, Temperature, Humidity, Windy の 4 つの属性から特徴付けられる対象が N, P ふたつのクラスに分類される。

表 2 にしめした 14 個の教師事例を図 11 の規則に従い符号化し, LISM に学習させた (事例学習)。

また表 2 の教師事例をみると, 属性 Outlook が overcast なら, 他の属性に関係なくクラスが P であることがわかる。このことが知識としてあたえられていると想定すると, No.3,7,12,13 の教師事例は, 汎化事例 (*overcast, #, #, #*) $\rightarrow P$ で表現できる。この汎化事例と No.1, 2, 4, 5, 6, 8, 9, 10, 11, 14 の 11 個の事例を LISM に学習させた (汎化事例学習)。

各場合の教師事例数, 中間ノード数, 学習回数を表 3 にしめす。

また事例学習, 汎化事例学習をおこなった LISM から抽出した決定木はどちらの場合も図 12 のようになった。

表 2: Saturday Mornings Classification Problem training set

No.	attributes				Class
	Outlook	Tem.	Humidity	Windy	
1.	sunny	hot	high	false	N
2.	sunny	hot	high	true	N
3.	overcast	hot	high	false	P
4.	rain	mild	high	false	P
5.	rain	cool	normal	false	P
6.	rain	cool	normal	true	N
7.	overcast	cool	normal	true	P
8.	sunny	mild	high	false	N
9.	sunny	cool	normal	false	P
10.	rain	mild	normal	false	P
11.	sunny	mild	normal	true	P
12.	overcast	mild	high	true	P
13.	overcast	hot	normal	false	P
14.	rain	mild	high	true	N

この問題の場合, 一部の属性のみからクラスが決定できる部分が少ないので, 汎化事例学習をおこなっても多重 XOR 問題ほどには, 教師事例数, 中間層ノード数は減少していない。

図 12 の決定木は文献 [6] 中でしめされている ID3 から生成した決定木と同一のものである。このことより LISM がおこなう部分空間想起は, 状態空間における教師事例の分布を反映したものになっていることがわかる。

Outlook: *sunny* $\rightarrow$ 100, *overcast* $\rightarrow$ 010,
rain $\rightarrow$ 001
Temperature: *hot* $\rightarrow$ 1, *mild* $\rightarrow$ 0.5, *cool* $\rightarrow$ 0
Humidity: *high* $\rightarrow$ 1, *normal* $\rightarrow$ 0
Windy: *true* $\rightarrow$ 1, *false* $\rightarrow$ 0
Class: *P* $\rightarrow$ 01, *N* $\rightarrow$ 10

図 11: 符号化の規則

表 3: Saturday Mornings Classification 問題実行結果

学習法	教師事例数	中間層ノード数	学習回数
汎化事例学習	11	6	4
事例学習	14	7	4

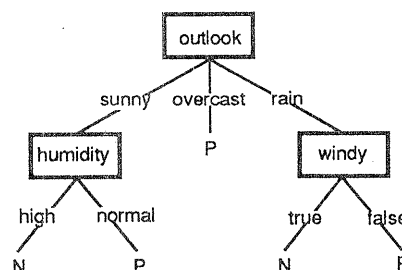


図 12: Saturday Mornings Classification 問題を学習した LISM から抽出した決定木

6 おわりに

本稿では、不定値を含む入力パターンを学習・想起可能で、部分的な情報を取り扱える階層型ニューラルネットワーク LISM と、その学習・想起アルゴリズムである部分空間学習法、部分空間想起法について述べた。また部分空間学習法により、必要な教師事例数、中間層ノード数を減らせること、部分空間想起法により LISM が獲得した知識を決定木の形式で取り出せることをしめた。

部分的な情報が取り扱え、説明能力をもつニューラルネットワークとしては Gallent のコネクショニストエキスパートシステムがある [7]。

このシステムでは、1 を true、-1 を false、0 を unknown と符号化している。すなわち unknown の幾何学的位置を true と false の中間点に置くことにより、ニューラルネットワークの汎化能力をもちいて true、false のいずれの場合にも同様な出力結果を得ようとしている。しかし汎化能力にも限界があるので unknown な属性の数が多い場合にもちいることはできない。

また Gallent のシステムがもつ説明機能は、ある状態を入力した時の出力結果に関する説明であり、本稿でしめたようにネットワークが獲得した知識の全体をしめすことはできない。

階層型ニューラルネットワークから知識を抽出しようとした研究としては、斉藤らの研究 [8] がある。

しかし、階層型ネットワークとして通常の入力ノードの値をすべて指定する必要があるものを使用しているため、属性数の多い場合に組合せ爆発をおこす問題がある。

本稿では、例題として属性値が離散的な値をとる場合しかあげなかったが、属性値が連続値をとる場合でも、何の変更もなくネットワークの学習・想起をおこなうことができる。ルール抽出法に関しては、しきい値の決定法などが解決すべき課題として残っている。

距離型ノードから構成された階層型ニューラルネットワークは、最近さかんに研究されており、中間層ノード数の決定法 [9] や、系統的な中心ベ

クトルの決定法 [10] などが提案されている。今後これらの成果を部分空間学習・想起法に取り入れ LISM の学習・想起機能を高めていくとともに、LISM を推論機構としてもちいた知識処理システムの検討を進めていく予定である。

参考文献

- [1] 小中裕喜, 阿部一裕, 瀬尾和尾: ニューロ・AI 統合型エキスパートシステム - システム構成 -, 情報処理学会第 42 年全国大会論文集, 3F-9 (1991).
- [2] 阿部一裕, 小中裕喜, 小船隆一: ニューロ・AI 統合型エキスパートシステム - SampleMaster の実装 -, 情報処理学会第 42 年全国大会論文集, 3F-10 (1991).
- [3] Rumelhart, D.E., McClelland, J.L. and The PDP Research Group: Parallel Distributed Processing, MIT Press, Cambridge, MA (1986).
- [4] Reilly, J.R., Cooper, L.N., and Elbaum, C.: A neural mode for category learning, Biological Cybernetics, vol.45, pp35-41 (1982).
- [5] Broomhead, D.S., and Lowe, D.: Radial basis functions, multi-variable functional interpolation and adaptive networks, Technical Report Memorandum No.4148, RSRE (1988).
- [6] Quinlan, J.R.: Induction of Decision Trees, Machine Learning, vol.1, pp81-106 (1986).
- [7] Galland, S.I.: Connectionist Expert System, CACM, vol.31, pp.152-169 (1988).
- [8] 斉藤和巳, 中野良平: 事例とニューラルネットからの分類ルールの抽出法, 情報処理学会研究会報告, 89-AI-67-3 (1989).
- [9] Nakano, R. and Saito, K.: Dream: A Heuristic Approach to hypersphere Minimum Covering, Artificial Neural Network, pp.427-432, North-Holland (1991).
- [10] Chen, S., Cowan, C.F.N., and Grant, P.M.: Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks: IEEE Trans. Neural Networks, vol.2, pp.302-309 (1991).