認識システムの内部構造獲得のための Matchability指向の特徴選択

山川宏

新情報処理開発機構　つくば研究センタ

〒305茨城県つくば市竹園1-6-1つくば三井ビル16F

Tel: 0298-53-1664　　e-mail: yamakawa@trc.rwcp.or.jp

あらまし　内部に複数の処理モジュールをもつ大規模認識システムのモジュール間では外界入力の情報構造に適合した情報交換を行う必要がある。そこでシステムが柔軟な適応能力を持つには、処理モジュール毎だけでなくその間の接続構造もダイナミックに変化させなければならない。接続構造の学習を行う手掛かりとしては外界入力の部分独立性を利用する方法が有望であり、これは局所毎には関連の強い特徴量の選択となる。しかし従来このような視野に立つ特徴選択の研究はあまり進められていない。

そこで本報告書では、現入力状態と過去の入力状態にマッチングできる機会に対応するMatchabilityという尺度を特徴量毎に導入し、これを最大化する方向で関連の強い特徴量を選択する手法を提案する。さらに計算機実験によりアルゴリズムの動作を確認したことを報告する。

キーワード　内部表現の自動獲得, パターン認識, ニューラルネットワーク, マルチエージェント, 教師無し学習, 事例ベースド推論

Feature Selection for Acquiring Internal Structure of Recognition System based on Matchability

Hiroshi Yamakawa

Tsukuba Research Center, Real World Computing Partnership

Tsukuba Mitsui Building 16F, 1-6-1, Takezono Tsukuba, Ibaraki 305, JAPAN

Tel: +81-298-53-1664　e-mail: yamakawa@trc.rwcp.or.jp

Abstract　For effective recognition, a recognition structure that controls the information flow among the specialized processing modules should reflect the implicit correlation structure of the environmental input. Autonomous construction of a recognition structure will lead to extensive improve in the flexibility of the adaptive recognition system. For this purpose we propose a matchability-oriented feature selection that can collect highly correlated features at each local module. Matchability is a concept derived from the recognition functions of an adaptive intelligent agent (useful for action generation) and corresponds to the probability of input data items matching stored data items in the recognition system. We check this algorithm in simple artificial environment.

Key words　Learning internal representation, Pattern recognition, Neural networks, Multi-agent system, Unsupervised learning, Case based reasoning

1. まえがき

認識システム、行動生成部、評価部などをもつ知的エージェントを考えると、認識システムは環境から得た情報を行動に役立て易い表現に変換する役割を持つ(図 1 参照)。そこで知的エージェント全体が柔軟な適応能力をもつためには、認識システムにおいても、処理を実行している段階での処理能力と、それを獲得するまでの学習機能の両方が必要となる。

実世界からの膨大な入力情報全体を一気に取り扱うことは現実的な速度では不可能である。よって、認識システムは部分処理空間に専門化した多数の処理モジュールを用意し、それらが入力の情報構造を反映して接続した内部構造が必要であると考えられる。このような認識システム適応能力は、処理モジュール毎と処理モジュール間の接続構造の両方での適応能力に依存する。しかし、これまでは処理モジュール毎の学習に関する研究が多い。それ故、処理モジュール間の接続の学習が可能となれば、これまで以上に柔軟性の高い適応能力を実現できる可能性が開けるだろう。

そこで、本報告書では処理モジュール間の接続構造の学習を目的とした特徴選択の提案を行う。なぜなら、接続構造の決定は各処理モジュールから見れば特徴量の選択だからである。特徴量の評価選択に関しては、これまでにもパターン認識[1,2,3,4,5,6,7,8] ニューラルネットワーク[9]、事例ベースド推論[10,11]などの分野で研究が行われている。

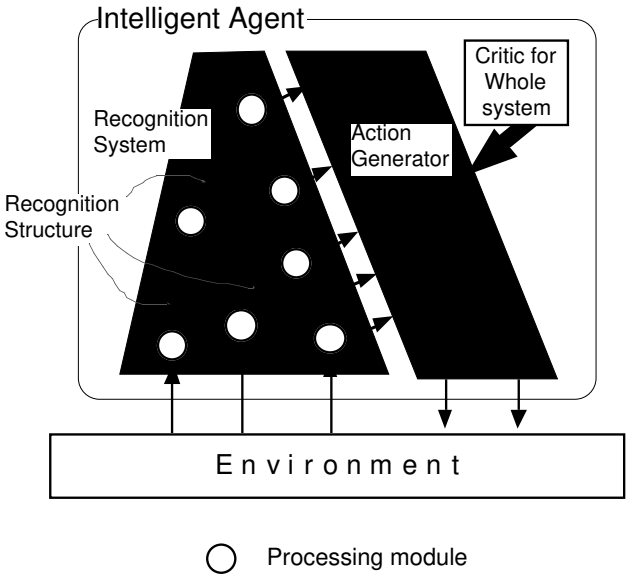

図 1 複数のモジュールを持つ認識システム

特徴を選択する尺度としては、タスクを限定する教師有り学習の枠組みの中で認識率や誤り確率などを利用するのが最も直接的でわかりやすい。事例ベースド推論の研究においては、この枠組みの中で有効性を表現するような重み係数を導入して特徴量の評価を行なっている。

しかしここではむしろ、タスクを限定せずに一般的な目的に用いる認識システムを考えいているため、入力情報だけを利用して自己組織的に構造を獲得する教師無し学習の枠組みに注目する。パターン認識におけるこの種の研究では、情報量基準などを用いて小さな次元の中に有用な情報を表現する方法が多数提案されている。つまり、出力の Compactnessを追及するアプローチである。さらに上記の二つの手法はしばしば組み合わされ、パターン認識の分野を初めとし、特にニューラルネットワークの分野における枝狩などはこれに類するケースが多い。また、多変量解析などの特徴抽出技術ににおいても、教師無し学習では出力の Compactness を追及することが多い。

しかし、Compactnessを追及する特徴量選択の手法は、認識システムの接続構造の学習の目的には、その性質上そぐわないことを説明する。

環境からの入力情報内に含まれる以下の二つの性質は、接続構造の学習に役立つと思われる。

1) 相関の空間的な局所性

特徴量空間内の高い相関は、多くの場合その部分特徴量に現れる。

2) 相関の時間的な局所性

部分特徴量空間内の高い相関は、定常的にではなく、特定の状況でのみ現れる。

接続構造の学習では、各処理モジュールは上記の局在性を利用し、特定の状況下で強い相関を持つ部分特徴量を選択すべきである。しかし、Compactnessを追及する特徴選択は、独立性が高く相互に相関の弱い特徴量を選択し、時間的にも異なった状況から情報を取り込む性質を持つ。よって、Compactnessの追及による特徴選択は接続構造の学習には不適切である。

そこで、次節以降ではCompactnessを指向する特徴選択の機能を補い得る、複数の処理モジュールの専門化を前提としたMatchabilityを指向する特徴選択の方法を提案する。

2. Matchabilityとその最大化の意義

本節では認識の目的の観点から、Matchabilityとその最大化の必要性について説明する。

ここでは認識の定義として計算機科学において比較的一般的である「一度認めたものを、もう一度認めること」と考える。つまり認識とは、ある入力を外部の観測者が解釈するのではなく、既に経験した入力と同一または、類似、または、それらの組み合わせとして解釈することである。

図 1 にも示したように認識システムは行動生成部を含む知的エージェントの一部であるから、認識の目的は行動の決定に役立つ情報の抽出だと考えられる。行動決定の過程では、まず現在の入力状態にマッチングする過去の入力状態を探索し、それに関連づけられた行動や、その行動の帰結に対する評価に基づいて行動を選択する。そこで、認識の定義と行動決定過程を考慮すると、「行動に有用な認識とは、出力表現がコンパクトであると同時に多くの機会にマッチングできることが望まれる。」(図 2 参照)。よって認識システムの接続構造の獲得においても、以下の二つの観点を考慮すべきであろう。

1) Compactivity

Compactivityを指向する研究では入力特徴量空間は予め与えられたものとして、出力における Compactnessを追及するのが普通である。ここでは、情報圧縮や、目立った特徴を強調が技術的テーマとなっている。一方、コンパクトにする処理の視点に立つと、逆に入力部分では冗長な情報を持つ方が都合がよい。ここではこの性質を Redundancy of Input として示しておく。

2) Matchability

Matchability指向では「認識システムが過去において外界から得られた有限の経験を、将来の入力にマッチできる可能性を大きくすること」を目的とする。つまり、Matchabilityとはシステムが入力に対してどれだけ多くのマッチングを生み出すことができるかという指標である。Matchabilityを増大させるには具体的には二つの方向がある。局所的かつ時間的には、相関の高い部分特徴量を選択することにより Matching Chanceを増大させることができる。大局的かつ空間的には多様な反応性を持つ処理モジュールを並列に動作させることにより Matching Varietyを増大させることができる。

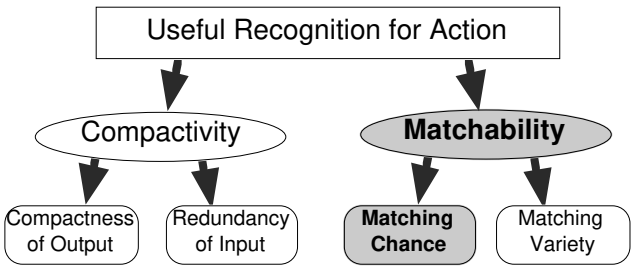

図 2 行動に有用な認識(Compativity & Matchability)

認識システムの接続構造の学習には、相補的な上記 2 つの尺度が重要であるが、現状では Compactness of Outputを評価の指標とする特徴選択 / 抽出などの研究が主流である。そこで、今後は Matchability指向の研究 / 開発も必要となると思われるので、次節において、この Matchability指向の特徴選択のアルゴリズムの一例を提案する。

3. 特徴選択のアルゴリズム

提案する特徴選択のアルゴリズムは各処理ユニットにおける Matching Chance(=局所的な Matchability)を増加させる学習則である。ここでは特徴量毎に与えられる重み係数 $\boldsymbol{W}(t)$ を変化させることで選択を行う。図 3 に示す様に処理ユニットの基本動作はマッチングで、保持している n個の参照データ $\boldsymbol{R}^P$ の中から、時刻 t における入力データ $\boldsymbol{I}(t)$ にもっとも近い参照データ $P_{match}(t)$ を選択する。この際に各特徴量毎の重み係数 $\boldsymbol{W}(t)$ を利用する。

t : 時刻

$\boldsymbol{W}(t) = \{W_i(t)\}$: 特徴量毎の重み係数

$\boldsymbol{R}^P = \{R^P_i\}$: 参照データ($P = 1 \sim n: n \le n_{max}$)

$\boldsymbol{I}(t) = \{I_i(t)\}$: 入力データ

さらに、図 3 に示すように重み係数の学習、入力データの参照データへの追加、不必要な参照データの削除を行なう。

3.1 距離の計算とマッチする参照データの選択

入力データ $\boldsymbol{I}(t)$ が与えられると。以下の二つの距離を計算する。

$$D^P(t) = \Sigma_i\{W_i(t)\ \Delta^P_i(t)\} \quad \text{荷重距離} \quad (1)$$

$$d^P(t) = \Sigma_i\{\Delta^P_i(t)\} \quad \text{非荷重距離} \quad (2)$$

ここで、 $\Delta^P_i(t) = |I_i(t) - R^P_i|$

マッチングでは、荷重距離 $D^P(t)$ が最も小さい参データが選択される。

$$P_{match}(t) = P(\min D^P(t)) \quad (3)$$

3.2 特徴量毎の重み係数の学習

Matching Chanceを増加させるための、重み係数 $\boldsymbol{W}(t)$ の学習則は、局所距離 $\Delta^{P_{match}}{}_i(t)$ が小さいほど重み係数が大きくなる学習規則で、

$$W_i^{(1)}(t) = W_i(t)\,\frac{d^{P_{match}}}{k\,\Delta^{P_{match}}{}_i(t) + d^{P_{match}}} \qquad (4)$$

(k: 学習定数)

$$W_i^{(2)}(t) = \frac{W_i^{(1)}(t)}{\sum_i W_i^{(1)}(t)} \qquad (5)$$

である。ここでは(4)式に示すようにマッチングしたデータとの非荷重距離を利用して重み係数を変更した後に(5)式で規格化を行なう。(4)式の効果により、マッチング精度が良く、誤差が少ない特徴量の重み係数が相対的に大きくなる。

さらに、ある程度大きな重み係数 $W_i(t)$ 同士を平均化させ、小さな重み係数 $W_i(t)$ に復活の機会を与えるために、W^{th} を重み係数の境界値として、

$$W_i^{(3)}(t) = \begin{cases} W_i^{(2)}(t) - \alpha \left\{ W_i^{(2)}(t) \right\}^2 & (W_i^{(2)}(t) > W^{th}) \\ \beta\, W_i^{(2)}(t) & (else) \end{cases}$$

$$(\alpha, \beta, W^{th}\text{: 定数},\ 0 < \alpha < 1,\ \beta > 1,\ 0 < W^{th} < 1) \qquad (6)$$

$$W_i(t+1) = \begin{cases} \dfrac{W_i^{(3)}(t)}{\sum_i W_i^{(3)}(t)} & (W_i^{(2)}(t) > W^{th}) \\ W_i^{(3)}(t) & (else) \end{cases} \qquad (7)$$

という学習則を追加し、最終的に更新された重み係数 $\boldsymbol{W}(t+1)$ が得られる。

3.3 入力データの参照データへの追加

新たに得られた入力データ $\boldsymbol{I}(t)$ が、保持している参照データ $\boldsymbol{R}^P$ の何れともかなり異なり、かつ参照データのスロットに余裕がある場合に、$\boldsymbol{I}(t)$ が新たな参照データとして追加される。つまり、マッチしたデータと入力データとの荷重距離があるしきい値 θ ($\theta > 0$) よりも大きければ追加が実行される。

$$\text{if } (D^{Pmatch}(t) > \theta \ \text{ and } \ P \neq P_{max})$$
$$\text{Adding } \boldsymbol{I}(t) \text{ to } \boldsymbol{R}^P \qquad (8)$$

3.4 参照データの削除

以下の二つの場合、参照データを削除する。

1) ほとんど利用されない古い参照データの削除

追加されてから 10 時刻過ぎても一度も利用されない参照データと、追加されてから 30 時刻経過しても一度しか利用されない参照データを削除する。

2) 類似した参照データの削除

重み係数の変化に伴って、類似であるとみなされた参照データの一方が削除される。削除される参照データ間の距離のしきい値はデータ追加時と同じ値 θ を利用する。

$$D^{PP'} = \Sigma\{ W_i(t)\, \Delta^{PP'}{}_i \} \qquad (9)$$

ここで、 $\Delta^{PP'}{}_i = | R^P{}_i - R^{P'}{}_i |$

$$\text{if}(D^{PP'} < \theta) \quad \text{Deleting } \boldsymbol{R}^{P'} \qquad (10)$$

4. シミュレーションとその結果

シミュレーションでは10通りの入力系列を、並列に動作する 8 つの処理モジュールに与えて、特徴量毎の重み係数 $W_i(t)$ の変化を調べた。以下、入力系列と処理モジュールの設定を説明し、シミュレーション結果を報告する。

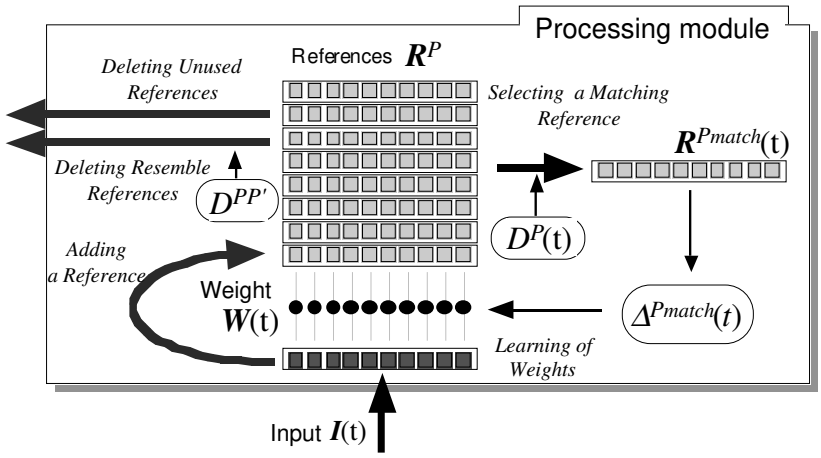

図 3 特徴選択アルゴリズムのダイアグラム

4.1 実験条件

入力系列は時刻 1 から 500まで継続する10次元の特徴量ベクトルである。2 クラス・フェーズにおいて、このベクトルはそれぞれ独立なスカラー変数に従属する二つの部分特徴量ベクトルに分割できる(図 4 参照)。 2 つの独立変数は、*rand* (min, max)を [min, max]間の一様乱数として以下のように定義する。

$$j_{\max}{}^A = rand(0.0, 4.0), \quad j_{\max}{}^B = rand(0.0, 6.0) \qquad (11)$$

Class Aは$j_{max}{}^{A}$によって制御される4次元特徴量で、Class Bは、$j_{max}{}^{B}$によって制御される6次元特徴量で、それぞれ0.1以下の雑音を含んでいる。

$$I_j(t) = \left| j_{\max}{}^{A} - j + 1 \right| + rand(0.0, 1.0) \qquad (j = 1,2,3,4)$$

$$I_j(t) = \left| j_{\max}{}^{B} - j + 5 \right| + rand(0.0, 1.0) \quad (j = 5,6,7,8,9.10) \tag{12}$$

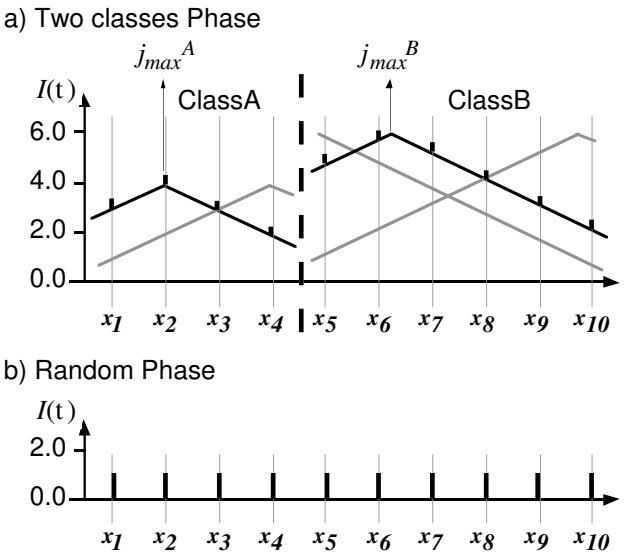

図4 シミュレーションに用いた入力系列

P_{max}	θ	k	W^{th}	α	β
20	0.1	1.0	0.05	0.2	1.04

表1 シミュレーション・パラメータの設定

一方、ランダム・フェーズでは各特徴量は各々独立な一様乱数に従い、次式で表される。

$$I_i(t) = rand(0.0, 1.0) \tag{13}$$

次に各処理モジュールの、パラメータは表1のように設定し、さらに初期状態では参照データを保持せず、重み係数は乱数により各次元毎に独立に決定(式14)した後に規格化される(式15)。

$$W_i(0) = 10^{-[1 + rand(0.0, 2.0)]} \tag{14}$$

$$W_i(0) = \frac{W_i(0)}{\sum_i W_i(0)} \tag{15}$$

4.2 実験結果

提案した特徴選択のアルゴリズムは相関の高い部分特徴量を選択するが、2節で述べた相関の時間的な局在性は考慮していない。従って、一つ目の定常的に2クラス・フェーズが継続する入力系列に対しては適切な動作が期待されるが、二つ目のランダム・フェーズを交えた入力系列に対しては時間的な局在性があるため重み係数の挙動は予想し難い。

	Number of Input series										Total
	0	1	2	3	4	5	6	7	8	9	
A	0	5	0	1	3	1	1	1	1	1	14
B	8	3	8	7	5	7	7	7	7	7	66

A=Class A, B=Class B

表2 2クラス入力系列に対する専門化

初期値の異なる入力系列毎のそれぞれのクラスに専門化した処理モジュールの数、およびその総和。

図5 (a) 重み係数の減衰(2クラス入力系列)
ClassBに専門化した処理モジュールの場合

図5 (b) 重み係数の減衰(2クラス入力系列)
ClassAに専門化した処理モジュールの場合

１）２クラス・フェーズの入力系列

異なる初期化を行ったいずれの入力系列においても100から200時刻経過後には、図５に示す例のように、各処理モジュールの重み係数はいずれかのクラスの特徴量を選択した。表２に示すように10通りの何れのケースにおいてもClass Bに専門化するモジュールが多いことから、冗長性の高い特徴量に対して専門化が起こりやすい傾向が伺える。

２）２つのフェーズを混合した入力系列

初めの8時刻はランダム・フェーズとし、引き続く8時刻を２クラス・フェーズとし、以下このサイクルを500時刻まで繰り返す入力系列を与えて同様の実験を行った。ランダム・フェーズを含めたことで入力情報に時間的な局在性があるにも関わらず、今回の実験の範囲では、図６(a), (b)及び表３に示すように、多くの場合にクラス毎の部分特徴量が選択された。

しかしながら500時刻の時点ではまだ、専門化が不十分な場合もあり、とくにその一部のケースで見られた現象として、ほぼ選択されたクラスと相関の強い重み係数の値が、平均化のしきい値 W^{th} よりも小さい領域に落ち込むために、学習を500時刻以降まで行っても、その値がなかなか大きくならない例があった。また500時刻経過後にも、何れのクラスにも専門化しなかったケースもある(図６(c)参照)。そして前記の実験と同様に冗長性の高い特徴量に専門化しやすい傾向が認められた。

	Number of Input series										Total
	0	1	2	3	4	5	6	7	8	9	
A	2	4	1	0	1	2	1	1	0	2	14
B	5	3	7	8	3	5	6	6	7	6	56
O	1	1	0	0	4	1	1	1	1	0	10

A=Class A, B=Class, O=obscure

表３ 混合入力系列に対する専門化

初期値の異なる入力系列毎のそれぞれのクラスに専門化した処理モジュールの数、およびその総和。なおここでの集計では専門化が多少不十分でもどちらかのクラスに分類しうる場合はそれぞれのクラスに分類した。それでもなお、判断できなかったものはObscureに分類した。

図６(b) 重み係数の減衰(混合入力系列)
ClassA に専門化した処理モジュールの場合

図６(a) 重み係数の減衰(混合入力系列)
ClassB に専門化した処理モジュールの場合

図６(c) 重み係数の減衰(混合入力系列)
専門化しなかった処理モジュールの場合

5. 議論

5.1 Compactivity の最大化と重み係数の平均化

本文中ではMatchabilityの最大化を中心に議論を進めた。しかし、実際には提案したアルゴリズムはCompactivityと Matchabilityの両尺度のバランスの上に成り立っている(図 7 参照)。つまり、(6)式による重み係数の平均化は選択する特徴量の数をを増加させ、Compactivityの最大化に必要な入力空間の冗長性(Redundancy of Input)を増大させる（ 2 節参照）。これに対して、(4)式による特徴量毎の重み係数の選択的減衰は Matching Chanceを増加させる。

二つの尺度のバランスから、できるだけ少ない参照データで記述可能でありかつ記述が増えなければノードを増やすという効果が生まれる。これは選択された特徴量によって張られる超空間中の参照データの存在領域を相対的に小さくすることである。よって選択された部分特徴空間に対して特徴抽出処理を行えば効果的な圧縮が可能となる。

もし二つの尺度を融合した適当な評価量を定義できれば、直感的には各処理モジュールの専門化は、選択する特徴量と保持する参照データなどを組み合わせた空間におけるその評価量に関するローカルミニマムへの引き込となるだろう。よって、シミュレーションにおいて特徴量の数が多いクラスに対して専門化が起こり易かったのは冗長度が高いほど、引き込み領域が大きい為であると考えられる。

5.2 処理モジュールの専門化のバランス

Matching Varietyの観点からは、多様な専門化が行われることが望ましい（ 2 節参照）。しかし、先の議論から推察できるように、専門化した状態はローカルな安定状態であり、しかもその発生確率には本質的に偏りがあるという問題がある。提案した、各モジュール毎に Matching Chanceを考慮した学習則ではこの問題は回避できない。そこで、今後はモジュール間で相互作用を行うグローバルな機構の導入が必要であり。これにより多様性に富んだ専門化が実現できるであろう。

5.3 認識システムの作成に向けた特徴抽出装置との結合

1 、 2 節に述べたように、提案した Matchability を指向する特徴選択手法は、Compactivity指向の特徴抽出手法に対して相補的な技術であるから、認識システムを構築するには両技術の融合が不可欠であると考えられる。

融合の具体的例を図 8 に示した。まず、Matchability指向の特徴選択器によって選ばれた相関の高い部分的な特徴量をCompactivity指向の特徴抽出器に入力することで圧縮した表現を生成する結合処理モジュールを作成する。さらにそれら接続してネットワーク構造を作成する。なお自律的に認識システムを構築するには、必要に応じて結合処理モジュールをインクリメンタルに追加したり、予め大きく作ったネットワークを枝狩りして小さくする方法などが考えられる。

5.4 入力情報構造に時間的な局在性への対応

混合入力系列に対するシミュレーションでは、入力情報に相関の時間的な局在性があるにも関わらず、特徴量の選択がある程度行えた。しかし、より複雑な時間的な局在性を持つ入力系列に対しては、重み係数の学習が不安定になると予測される。今後この問題に対応するには、参照データの評価や取捨選択についても検討する必要がある。

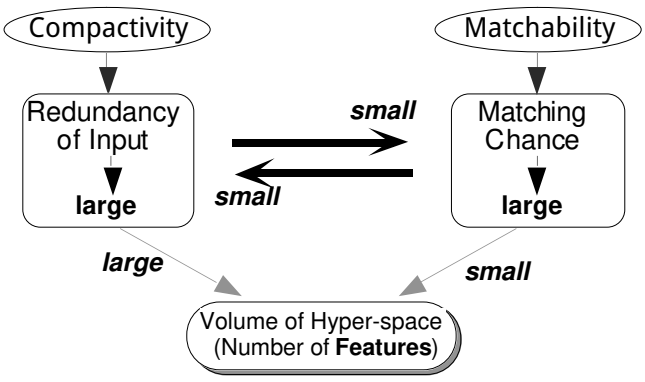

図 7 特徴選択におけるバランス

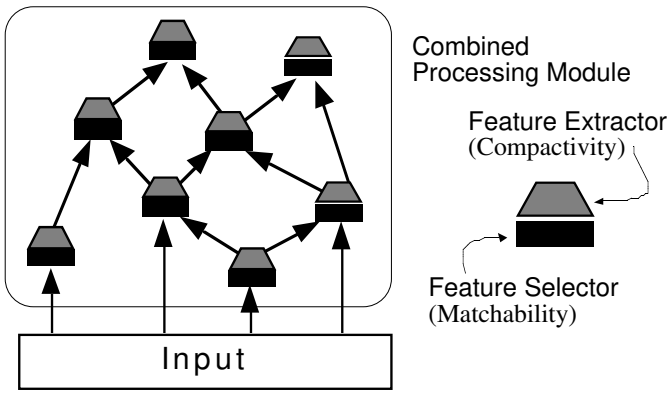

図 8 認識システムの構築

6. 結論

柔軟な適応能力を持つ知的エージェントの実現や、それらの設計負担を低減する為には、認識のための内部構造が自律的に獲得されると都合が良い。しかし従来から研究されたタスクに依存しない特徴抽出 / 選択技術では、多くの場合に出力情報をコンパクトにすることを最終目的とする。一方、認識システム内部の接続構造の学習では相互の相関が大きな特徴量を収集すべきなので、上記の技術を利用するのは不適切であると思われる。そこで、本報告書では従来技術と相補的な機能を実現しうるMatchabilityという尺度に基づく学習の必要性を主張した。Matchability指向では相互に相関の強い特徴量を収集するので、接続構造の学習に都合が良い。そこで、本報告書では特徴量毎の重み係数をMatchabilityの増大に基づいて更新する特徴選択アルゴリズムを提案した。シミュレーションでは、冗長性が大きく圧縮しやすい特徴量を選択し易い傾向が見られた。

今後は認識システムの構築に向けて、議論にも述べたように、特徴抽出技術との融合、処理モジュールの専門化におけるグローバルなバランスによる多様性の確保、相関の時間的な局在性に対応するための拡張などの研究を進めるべきであると考えている。

さらに、Matchability指向では局所的には冗長な特徴量を収集するので、一部の特徴量が隠蔽された場合に対応できるなど、情報統合技術への応用が期待できる。そして、入力特徴量に行動情報を含めることで、関連の強い認識行動関係の抽出に応用することも考えられる。また、脳やニューラルネットワークの研究に関してもMatchabilityの最大化を盛り込んだモデルを作ることで成果が得られる可能性がある。

参考文献

[1] Marill, T. and Green, D. M., (1963). "On the effectiveness of receptors in recognition system." *IEEE Trans. Inform. Theory*, vol. 9, pp. 11-17.

[2] Tou, J. T. and Heydoren, R. P., (1963). "Some approaches to feature extraction." *Computer and Information Sciences-II*, Academic Press, New York, pp. 57-89.

[3] Chien, Y. T. and Fu, K. S., (1968). "Selection and ordering of feature observations in pattern recognition systems." *Inform. Contr.*, vol. 12, pp. 394-414.

[4] Backer, E. and Scipper, J. A. De., (1977). "On the max-min approach for feature ordering and selection." *The Seminar on Pattern Recognition* , Liege University, Sart-Tilman, Belgium, 2.4.1.

[5] Narendra, P. M. and Fukunaga, K., (1977). "A branch and bound algorithm for feature subset selection." *IEEE Trans. Comput.*, vol. 26, no. 9, pp. 917-922.

[6] Pudil, P., Novovicova, J. and Kittler, J., (1994). "Floating search methods in feature selection." *Pattern Recognition Letter*, vol. 15, no. 11, pp. 1119-125.

[7] Malina, W., (1987). "Some multiclass fisher feature selection algorithms and their comparison with Karhunen-Loeve algorithms." *Pattern Recognition Letter*, vol. 6, no. 5, pp. 279-285.

[8] Siddiqui, K. J., Liu, Y.-H., Hay, D. R. and Suen, C. Y., (1990). "Optimal waveform feature selection using a pseudo similarity method." *J. Acoustic Emission*, vol. 9, no. 1 , pp. 9-16.

[9] Read, R., (1993). "Pruning Algorithms –A Survey." *IEEE Trans. Neural Networks*, vol. 4. no. 5, pp. 740-747.

[10] Aha, D. W. and Kibler, D., (1991). "Instance-based learning algorithms." *Machine Learning*, vol. 6, pp. 37-66.

[11] Wettschereck, D., Aha, D. W., and Mohri, T. (1995). "A review and comparative evaluation of feature weighting methods for lazy learning algorithms." *(Technical Report AIC-95-012)*. Navy Center for Applied Research in Artificial Intelligence.