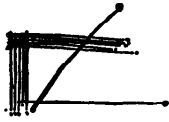


展望



コネクショニズムの展望

(IV) パターン処理の観点からの期待†

河原 英 紀††

1. はじめに

単純な機能をもつユニットの相互作用の結果として、人間の高度な認知機能を説明することを狙うコネクショニズム¹⁾は、現実の世界と記号の世界の橋渡しの役割を担うという点において、パターン認識・処理と深く関わっている。最近では、コネクショニストモデルを自然な形で実現することのできるニューラルネットの研究が急速に進展し、ハードウェア化による高速処理も期待できる状況となっている²⁾。このような状況を踏まえ、本稿では、音声を中心とするパターン処理の観点から、コネクショニズムに期待する事柄をまとめた。なお、急速に進展しつつある分野の常として、コネクショニズムやニューラルネットという言葉の意味についての共通の合意は形成されておらず、さまざまに異なったニュアンスで使われている。ここでは、コネクショニズムを思想や方法論を表すものとし、計算機構を表すニューラルネットと区別して用いることとする。

2. 音声情報処理と計算機関連科学との関わり

音声情報処理は、これまで、電子回路、デジタル計算機、時系列分析、マイクロエレクトロニクス、知識工学などの関連分野の技術の進歩を取り入れながら、能力と適用範囲を拡大してきた。その結果、今日では、日用品にまで入り込むほどの広範な応用の時代を迎えるに至っている。しかし、各ステップで大きな技術的進歩があったにもかかわらず、話者や話題を特定しない連続音声の認識という目標の達成には、まだかなりの距離が残されていることも事実である。これらの技術の導入のされかたを振り返ってみることは、コ

ネクショニズムが、今後、どのような形で取り入れられていくかを考える上で参考になろう。

単独に発声された母音の韻質と物理的特性との対応関係については、すでに前記の末から知られており、電子回路技術やデジタル計算機の進歩は、これらの知見を精密化する上で強力な道具を提供した。統計的な時系列解析に基礎を置く線形予測法(LPC)^{3),4)}は、音声の物理的特徴を効率よく表現する強力な計算手段を与え、広範な応用を可能にした。しかし、連続的に発声された音声の中の母音や、子音の場合には、音声学的な単位と物理的な特徴は直接には対応しておらず、これらの技術の利用による効果は、限られたものであった。実用的な音声認識技術の進歩は、話者と話題を限定し、認識の単位をより大きな単語程度のものとするでもたらされた⁵⁾。対応させるべきものについての視点の変更が、問題を解決したといえよう。

1971年から開始された音声理解プロジェクトは、当時、興隆しつつあった言語理解技術を利用して、意図の理解を目標に据えることによって、視点の変更を試みたものとみなすことができる。最終的には、音声の理解が能動的な過程であるとする枠組みの有効性を示し、所期の目標性能を達成することのできるシステムを生み出すことができたが、ここでも、物理的な波形のレベルから音声学的な記号にいたるギャップが、予想以上に深いものであることを印象づけた⁶⁾。余談になるが、このプロジェクトで開発された Hearsay-II は、さまざまな知識をもった独立のプロセスが、共通の情報にアクセスすることにより協調して問題を解決する構造をもっており、コネクショニズムの創設者の一人に強い印象を与えたことは、興味深い⁷⁾。

このように多くの問題を抱えているにもかかわらず、人間の有するような、柔軟で限定されない認識能力を実現するためには、音素あるいは音節のような要素の認識が不可欠であるとの考えから、1980年代に入ると再び音声学的な要素と物理的特徴の関連についての精力的な研究が進められるようになった⁸⁾。この時

† Perspectives on Connectionism (IV) Connectionism as a new Paradigm in Pattern Processing by Hideki KAWAHARA (Kakehi Research Group, Information Science Division, NTT Basic Research Laboratories).

†† NTT 基礎研究所情報科学研究部 研究グループ

期に盛んになった知識工学的手法が、物理特徴と音声学的な記号を結びつける知識を記述するための手段として利用されている。しかし、知識の抽出は困難な問題である。特に、音声の知覚のようにわれわれが無意識のうちに容易に行っているような行為の機構を、明示的な規則の集合として記述することへの直接的な方法の適用は、ほとんど不可能に近い。現在は、スペクトラムリーディングのように、音声を視覚的に表現したものを人間のエキスパートが読みとる過程を調べ、利用されている知識を抽出するというような、多くの労力を要する間接的な方法が用いられている⁹⁾。このような方法では、抽出された知識と人間の知覚過程で利用されている知識との対応が保証されず、また、抽出された知識が十分に強力で相互に矛盾していないかを検証することには、多くの困難をともなう。

コネクショニズムは、このような従来の技術の弱点であった、明示的に記述しにくい知識やスキルを抽出し利用するための方法論を与えるものとして、期待できる。また、コネクショニズムの考え方は、意識のレベルで記号的に行われている活動は人間の知的活動の微小な部分を占めるに過ぎず、ほとんどは前記号的なレベルで処理されていることを示唆するものであり、パターン認識研究者の共感を誘う魅力がある。

3. 認識手法としてのコネクショニズム

以上のようなコネクショニズムの位置づけを踏まえて、具体的に期待できることをまとめてみたい。

最も直接的に期待できることは、パターン認識やクラスタリングに必要とされる複雑な形状の識別曲面を形成する方法としての利用である。この観点は、Lippman による解説¹⁰⁾ に詳しいが、工学的な立場である。

パターン処理の分野では、これまで数多くのアドホックな工夫の利用が認識率を向上させるなどの理由により正当化されてきた。これらのあるものは、適当な変数変換により、複雑な形状の識別曲面をできるだけ超平面や2次曲面に近付けようとする工夫であるとみなすことができるし、またあるものは、複雑な識別曲面を何とか近似しようとする試みであると考えることができる。今やコネクショニズムと同義的にさえ扱われている誤差逆向き伝播法 (BP)¹¹⁾ などは、複雑な形状の識別曲面を組織的に形成できるアルゴリズムとしての側面をもっていることから、あまり手を加えることなく、従来のアドホックな工夫に勝る識別率の達成

が期待できる。音声認識への応用も、この点に期待するものが多い^{12), 13)}。

しかし、現実の問題は、このような一段階の力づくの方法で解決できるほど簡単なものではないであろう。コネクショニズムへの次の期待は、複雑な現実の過程を、理解可能なサブシステムに分解する上でのヒントを与えてくれることにある。

認知心理学的な知見によれば、われわれの認識や問題解決の能力は、経験に強く束縛されている。日常経験することと類似しているならば、われわれは非常に高い能力を示すことができるにもかかわらず、そうでないときには、同じ論理的構造をもつ簡単な問題すら解けない場合があることが知られている。パターン認識の過程に含まれるであろう多くの階層構造を見いだすことは、われわれにとっては、日常の経験の適用できない困難な課題である。この困難を克服する上で、PDP グループの主張する知識の内部表現の探索¹¹⁾ という観点が重要になる。具体例で説明したい。

BP や Boltzmann マシン¹⁴⁾ では、隠れユニットと呼ばれる外界と直接には接触しないユニットを用いることが多い。これらの隠れユニットの機能には、入力データと出力の関係を反映したなんらかの規則性が抽出されていることが予想できる。したがって、活性化パターンやユニット間の結合の重みを詳しく調べることで、外界に関する知識の内部表現についての手がかりが得られる可能性がある。実際、コネクショニストモデルによる音声合成規則の抽出の研究として有名な NETtalk の場合には、隠れユニットの活性化パターンなどの因子分析やクラスタ分析の結果、母音の符号化が声道の形状に基づいて行われていることが見いだされている¹⁵⁾。この例では、学習用に用いた資料の母音の出現状況に当初見落とされていた規則性があったことが、その後の分析で確認されている。

同様な例は、曲面形状の識別¹⁶⁾ や、眼球運動と視覚の応答野との連動の学習¹⁷⁾ に関しても報告されており、既存の枠組みから脱して新しい機能や概念を見いだす上で、コネクショニストモデルが有用な道具となり得ることを予想させる。音声関連の分野では、この方向に沿った研究はまだ少ないが、Elman and Zipser の研究¹⁸⁾ が有名であり、我が国では入野、河原の一連の研究¹⁹⁻²¹⁾ がある。

音声認識などの場合には、高次認知過程との関連を無視するわけにはいかない。コネクショニズムに期待する次の点は、自然な形でトップダウン処理とボト

ムアップ処理を融合させる機構の解明である。

従来の方では、複雑なデータ構造と制御構造を用いることで、これらの処理の統合が試みられてきた。コネクショニズムは、この部分に、新しい枠組みを提供する。具体的な音声認識システムの構築を目指したものではないが、McClellandによるTraceモデルは、音響的特徴、音素、単語などのさまざまなレベルに存在する要素間の相互作用の結果として、外面的には単語辞書や単語の標準パターンが利用されているような機能が構成できることを示し、複雑な制御機構を設けなくとも、これらの処理が自然に融合される可能性を示した²²⁾。このような機構を単純に言語のレベルにまで拡張できるかについては、議論も多いが、柔軟で高度な人間のパターン認識機能を説明する魅力的なモデルの一つである。

一方、ニューラルネット研究が大変なブームになり、研究者の数が急速に増大したことにともない、コネクショニストモデルも提案者の意図とは異なったところで有用性を発揮する場面が生じ始めている。

たとえば、BPは、重回帰分析を非線形の領域に拡張した方法であるとみなすこともできる²³⁾。したがって、既存の認識システムの出力から望ましい出力へのファインチューニングをBPを用いて行い、性能を改良するといった応用も十分に期待できる方向である。いわば、対象についての知識の不足や理論モデルの不完全さを埋めるための利用である。

また、非線形の予測機構としての利用も考えられる。たとえば、声道形状から音声波形への変換については、ほとんど線形の過程であるため、たとえBPを利用したとしても、LPCなどの従来の方と大差ない結果しか得られないであろう。しかし、数10ms～数100msにわたるスペクトル形状の動的変化の予測のように、調音モデルまでを考慮しなければならない場合には、運動指令から調音器官の運動を介して声道形状に変換されるまでに多くの非線形の過程が含まれるため、BPが遙かに良好な成績をあげることが期待できる。

Hopfieldモデル²⁴⁾についても、最近、新しい展開があり、組み合わせ問題の最小値を確定的な方法で解くことのできる可能性が出てきている²⁵⁾。現在までのところ、どのような初期値が最小値に収束するかという肝心の部分は予想としてしか与えられていないが、もし、この予想が成立するのであれば、組み合わせ問題の強力な解法として、パターンの検索などを始めと

する多くの問題への応用が期待できる。

さらに、最近では、これらのモデルをハードウェア化する試みがいくつか行われており、ソフトウェアでのシミュレーションに比べると数十万倍以上の速度が得られるとの報告もある²⁶⁾。このような高度の処理能力は、シミュレーションできる範囲を拡大し、質的に異なった展開も期待できる。

パターン認識の研究者は、これまで慢性的に計算能力の不足に悩まされてきた。たとえば、前の章で触れた不特定の話者による連続音声を認識するためには、100,000 MIPSの計算能力が心要であるとの試算もある²⁷⁾。上記のハードウェア化による速度向上は、この要求条件を、達成可能なものに変える。

このように、実際の応用という観点からみた場合には、コネクショニズムの考え方自体よりも、道具として開発されたアルゴリズムや専用のハードウェアによるインパクトのほうが、当面は、大きいと予想される。しかし、長期的な観点に立てば、コネクショニズムの興隆により、記号化される以前の、経験によって得られる知識の重要性が改めて認識されたことは、人間と同様の、学習によりだんだん賢くなるシステムを本当に実現する上で、重要な契機となるであろう。

4. 今後の課題

しかし、コネクショニズムを本格的にパターン処理の問題に適用するためには、解決しなければならない多くの課題がある。それらは、規模依存性およびダイナミックレンジの問題、時間情報の表現・処理の問題、高次過程との整合の問題である。

現在、音声認識に用いられているコネクショニストモデルのノード数は、高々数1000個程度であり、出力としても全部の音韻をカバーするものは少ない。これは、モデルの規模の増大にともない、学習のために必要なサンプル数と、収束までに要する演算回数が急速に増加するためである。また、パリティ問題のように悪条件の問題の場合には、規模の増大にともない、枝の重みのダイナミックレンジが急速に増加する。これらは、すでに1960年代に「パーセプトロン」²⁷⁾で指摘された問題であるが、最近のモデルにおいても本質的には解決されていない。

大規模な並列処理マシンの存在証明である脳も、一様にランダムに配線されているわけではなく、さまざまに組織化されたユニットが規則的に配置・接続された複雑なシステムであることが明らかにされてきてい

ることを考慮すると、この問題は並列計算機構にとって本質的な制約条件である可能性がある。もし、そのような制約が存在するのであれば、コネクショニストモデルの能力についての理論的分析を踏まえて、与えられた問題を適切なサブシステムに分解して解く方法を明らかにすることが課題となる。いわば、ニューラルネットについての計算の複雑性の理論と回路網の設計理論が求められている。

従来のコネクショニストモデルでは、時間の問題を正面から取り上げた例はあまり多くない。これは、時間の概念の捉え方の難しさによるものであるとも考えられる。たとえば、別の次元と考えるか、系列により定義されるものとするか、事象が生起するためのパラメータと考えるか、単一のものとしてではなく、階層構造をもったものとして考えるかなど、時間についてのさまざまな捉え方が可能である。実際、人間の聴覚における時間知覚²⁹⁾については、表-1に示すようにさまざまなレベルがあり、上記の捉え方のすべてが実際に実現されている可能性もある。

従来の音声認識でも、フレームごとのマッチングから始まり、次第に複雑な時間構造が扱えるようさまざまな手法が開発されてきているが、BPに代表される手法は、非線形の写像を簡単な手続きで形成できるという点で、従来の方法に対する潜在的な優位性をもっている。さまざまな時間の概念に対応する適切なモデルを定式化できれば、音声のように時間的変化に本質的な情報が含まれるパターンの認識に利用できる強力な道具となることが期待できる。具体的には、ループを含む系のダイナミクスと学習則についての理論の整備が当面の課題となろう。

パターン処理も、最終的には高次の認知に結びつく。コネクショニズムが、言語などの高次認知過程を理解するための枠組みとして適切であるかについては、他の著者による詳しい議論が予定されているが、知能の本質についての哲学的な論争^{29,30)}ともかかわっており、簡単に決着のつく状況にはない。この枠組みの変更がパターン処理のレベルと高次認知過程のインタフェースにどのような影響を与えるかを検討しておくことが必要であろう。

以上、音声認識への応用例にとり、パターン処理の観点からのコネクショニズムに対する期待を中心に、説明してきた。現在は、パーセプトロン時代とは異なり、これらの期待が、実体をともなったものであるかについて、各自で試してみるための手段が与え

表-1 音声知覚における時間の階層構造

時間長	感覚・現象	対応する技術
10 μ s	●方向知覚	●信号処理
1ms	●音色の識別	
10ms	●母音韻質	●パターン処理
100ms	●逐次感	
	●時間順序の知覚	
	●子音の識別	
1s	●調音結合の補償	●記号処理
	●単語	
	●音韻の対比効果	
	●リズム感の上限	
10s	●文	●知能処理
	●文章 ●会話	

られている。PDPの第3巻には、前2巻で紹介されたモデルのほとんどについてのソースコードのついたプログラムが添付されており、解説と練習問題をあわせて読めば、すぐにも自分の問題に適用してみることができるようになっている。また、このような試みを通じて生じた疑問点を、議論して解決するうえで効果の大きい、全国的な計算機ネットワークも整備されつつある。このような状況を生かし、コネクショニズムを、確実な基盤の上に立つ共有の技術/枠組みとして育てていきたいと考えている。

5. まとめ

コネクショニズムは、従来のパターン処理の弱点を補うための方法論として位置づけることができる。現在、ニューラルネットは、過熱気味とも言えるブームになっており、コネクショニズムもその一部であるかのように扱われている。このような先行した期待に攪乱されることなく、着実な理論的研究や実証的研究が積み上げられれば、コネクショニズムは、パターン処理の分野に多くの実りをもたらすであろう。また、並行して進められているニューラルネットの研究により生み出されるであろう高速のハードウェアは、かつてLPC理論とそのLSI化が音声情報処理の分野に与えたと同様のインパクトを、この分野に与える可能性がある。

謝辞 本稿をまとめるにあたり、議論していただいた当研究所読者を始めとする多くの方々に感謝します。コネクショニズムやニューラルネットワーク研究の動向を把握するうえで、計算機ネットワークは今や不可欠のものとなっている。これらを育て、維持されている方々に感謝します。

参考文献

- 1) Rumelhart, D. E. and McClelland, J. L. (Eds.): *Parallel Distributed Processing: Exploration in the Microstructure of Cognition*. Vol. 1, 2, MIT Press, Cambridge (1986).
- 2) 三宅 誠: ニューラルコンピューテーション, 電子情報通信学会誌, Vol. 70, No. 12, pp. 1255-1262 (1987年12月).
- 3) 板倉, 斎藤: 統計的手法による音声スペクトル密度とホルマント周波の推定, 電子通信学会論文誌, Vol. 53-A, No. 1, pp. 35-42 (1970).
- 4) Atal, B. S. and Schroeder, M. R.: *Adaptive Predictive Coding of Speech Signals*, Bell System Tech. J., Vol. 49, No. 8, pp. 1973-1986 (1970).
- 5) 迫江, 千葉: 動的計画法を利用した音声の時間正規化に基づく連続単語認識, 音響学会誌, Vol. 27, No. 9, pp. 483-500 (1971).
- 6) Klatt, D. H.: Review of the ARPA Speech Understanding Project, J. Acoust. Soc. Am., Vol. 62, No. 6, pp. 1345-1366 (1977).
- 7) Rumelhart, D. E.: *Introduction to Human Information Processing*, John Wiley (1977).
- 8) Reddy, R. and Zue, V. W.: Recognizing Continuous Speech Remains an Elusive Goal, *IEEE Spectrum*, pp. 84-87 (Nov. 1983).
- 9) Zue, V. W. and Lamel, L. F.: An Expert Spectrogram Reader: A Knowledge based Approach to Speech Recognition, *Proc. ICASSP 86*, pp. 1197-1200 (1986).
- 10) Lippmann, R. P.: An Introduction to Computing with Neural Nets, *IEEE ASSP Magazine*, pp. 4-22 (Apr. 1987).
- 11) Rumelhart, D. E., Hinton, G. E. and Williams, R. J.: Learning Internal Representations by Error Propagation, In D. E. Rumelhart and J. L. McClelland (Eds.), *Parallel Distributed Processing: Exploration in the Microstructure of Cognition*. Vol. 1., pp. 318-362, MIT Press, Cambridge (1986).
- 12) 稲葉則夫: ニューラルネットをパターン認識, 信号処理, 知識処理に使う: 学習アルゴリズムは“バックプロパゲーション”, *日経エレクトロニクス*, No. 427, pp. 115-124 (1987).
- 13) 住永, 稲葉: ニューラルネットで音声認識は変わる, *日経エレクトロニクス*, No. 441, pp. 175-179 (1988).
- 14) Hinton, G. E. and Sejnowsky, T. J.: Learning and Relearning in Boltzmann Machines, In D. E. Rumelhart and J. L. McClelland (Eds.), *Parallel Distributed Processing: Exploration in the Microstructure of Cognition*. Vol. 1., pp. 282-317, MIT Press, Cambridge (1986).
- 15) Rosenberg, C. R.: Revealing the Structure of NETtalk's Internal Representations, *Proc. of the Ninth Annual Conference of the Cognitive Science Society*, pp. 537-554 (1987).
- 16) Leaky, S. R. and Sejnowsky, T. J.: Neural Network Model for the Cortical Representation of Surface Curvature from Images of Shaded Surface, In Lund, J. S. (ed.) *Sensory Processing*, Oxford University Press, Oxford (1988).
- 17) Zipser, D. and Anderson, R. A.: A Back-Propagation Programmed Network That Simulates Response Properties of a Subset of Posterior Parietal Neurons, *Nature*, Vol. 331, No. 25, pp. 679-684 (1988).
- 18) Elman, J. L. and Zipser, D.: Learning the Hidden Structure of Speech, *ICS Report 8701*, UCSD (Feb. 1987).
- 19) 入野, 河原: 神経回路網による母音認識, 音響学会聴覚研究会資料, H-87-52 (1987年10月).
- 20) 入野, 河原: 基底膜振動を入力とした母音特徴抽出の検討, 音講論, 3-P-15 (1988年3月).
- 21) 河原, 入野: 状態縮約を用いた神経回路網による時空間パターン特徴抽出の検討, 音響学会聴覚研究会資料, H-88-15 (1988年3月).
- 22) McClelland, J. L. and Elman, J. L.: Interactive Processes in Speech Perception: The TRACE Model, In D. E. Rumelhart and J. L. McClelland (Eds.), *Parallel Distributed Processing: Exploration in the Microstructure of Cognition*, Vol. 2., pp. 55-121, MIT Press, Cambridge (1986).
- 23) 李, 板倉: 重回帰係数の適応化による中国語の声調認識, 音講論, 1-2-10 (1988年3月).
- 24) Hopfield, J. J.: Neurons with Graded Response Have Collective Computational Properties Like Those of Two-State Neurons, *Proc. Natl. Acad. Sci. USA*, Vol. 81, pp. 3088-3092 (1984).
- 25) 上坂吉則: ニューロン回路のエネルギータについて, *AVIRG 予稿* (1988年3月).
- 26) Alspector, J. and Allen, R. B.: A Neuromorphic VLSI Learning System, In *Advanced Research in VLSI: Proceedings of the 1987 Stanford Conference*, Paul Losleben (Ed.), pp. 313-349, MIT Press, Cambridge (1987).
- 27) Minsky, M. and Papert, S.: *Perceptrons: An Introduction to Computational Geometry—Expanded Edition*, MIT Press, Cambridge (1988).
- 28) 寺西: 聴覚の時間的側面, 難波編 聴覚ハンドブック第7章, pp. 276-319, ナカニシヤ出版 (1984).
- 29) Fodor, J. A. and Pylyshyn, Z. W.: Connectionism and Cognitive Architecture: A Critical Analysis, *Cognition*, 28, pp. 1-71 (1988).
- 30) Smolensky, P.: On the Proper Treatment of Connectionism, *CU-CS-377-87*, Colorado Univ. (1988). (To appear in *The Behavior and Brain Science* (1988)).

(昭和63年5月9日受付)