

7 E-4 疎な結合を持つ相互結合型神経回路網の動的データ駆動型並列実現法

大東 敏信 水野 正之 Ali M. Alhaj 岩田 誠 寺田 浩詔

大阪大学 工学部

1 はじめに

本研究の目的は、大規模な相互結合型神経回路網の高速模擬手法の確立にある。既に、我々は動的データ駆動型プロセッサを用いた相互結合型神経回路網の高速模擬手法を提案している [1] が、ここでは、神経回路網の結合が疎であることを利用してさらに高速な模擬を行なう手法を提案し、評価する。

動的データ駆動型処理方式は、自然な形で複数の入力データ組をパイプライン的に処理できること、また、文脈の切り換えが頻繁に起こるような処理でもパイプラインの乱れがないなどの特徴を持つ。従って、マルチプロセッサ化すれば、プロセッサ内での演算とプロセッサ間の通信をオーバーラップさせることによって、プロセッサ間の通信がボトルネックとならないため、神経回路網が本来持っているような粒度の細かな並列処理を効率的に実現できる。

2 完全相互結合型神経回路網の実現法

相互結合型神経回路網モデルでは各ニューロンが非同期的に動作しなければならず、複数のニューロンを並列に処理できない。したがって、従来の模擬手法では1つのニューロンの積和演算を並列に処理する水準での粗粒度の並列性しか活用できず、並列性があまりないものとして捉えられてきた。

これに対して、我々は“複数のニューロンが同時に発火できない”ことを利用し、従来に比べて、細粒度の並列性が得られる方式を既に提案した [1]。

この相互結合型神経回路網模擬手法では、まず初期設定として、

$$\forall i; u_i = \sum_k w_{ik} x_k + h_i \quad (1)$$

を実行する。以後、しきい値処理、

$$x'_i = \begin{cases} 1 & \text{Prob.}[f(u_i)] \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

全てのニューロンの膜電位補正処理、

$$\forall k; u'_k \leftarrow u_k + w_{ik}(x'_i - x_i) \quad (3)$$

を繰り返し行うことで神経回路網を模擬する。ここで、 x_i 、 x'_i 、 u_i 、 u'_i 、 w_{ik} 、 h_i はそれぞれ発火前の出力値、発火後の出力値、発火前の膜電位、発火後の膜電位、結合荷重、しきい値を示す。

すなわち、あるニューロンが発火した場合、他のニューロンは、改めて膜電位を計算するのではなく、発火したニューロンによって変化する量のみを計算し、膜電位を補正する。次に発火するニューロンは、この膜電位の補正のみを行なうことによって発火可能な状態となるため、連続発火を行なうためには、この膜電位の補正を高速に行なえば良い。一方、式 (3) に示すように、各ニューロンの膜電位補正処理は独立しているため、全く並列に演算を行なうことができ、マルチプロセッサ上での実現に適している。

さらに、この模擬手法を、動的データ駆動型プロセッサ上で実現すると、(1) プロセッサ内で膜電位補正処理を並列に処理できる (この並列度を内部多重度と呼ぶ)。さらに、(2) 全く独立した複数の神経回路網の模擬を同時並列に同じプロセッサ網上で処理できる (この並列度を外部多重度と呼ぶ)。

実際に、リング状に接続した動的データ駆動型マルチプロセッサ (最大 20MFLOPS) 上に実現し、プロセッサ・シミュレーションで評価したところ、ニューロン数が 100 万個、プロセッサ数が 1024 台の時、約 400MCPS の性能が得られている。

また、この実現法では、リング上のプロセッサ数を増加させることにより、スケーラブルな性能向上が望めることがわかっている。

3 疎な相互結合型神経回路網の実現法

巡回セールスマン問題などに見られるように、一般に、相互結合型神経回路網のアプリケーションでは、ニューロンの結合が疎であることが多い。本研究では、このことを積極的に利用して、先に述べた模擬手法を改良し、疎な結合を持つ相互結合型神経回路網向けの模擬手法を提案する。

式 (3) において、(1) w_{ik} が零である場合、そのニューロンの膜電位補正処理を実行する必要は無い。さらには、(2) 粗な結合を考慮すると、結合荷重行列はスパースであるため、荷重行列の記憶域を圧縮できる。

この場合、(a) w_{ik} が零であるか否かを判定するための遅延、および (b) 行列の非零要素を指すポインタアクセスを介する結合荷重行列アクセスに対する遅延、が問題になる。

一方、(c) 先に提案した完全相互結合型の場合は、膜電位補正処理が常に一定の演算量であるため、個々のプロセッサの負荷は常に一定であるが、粗な相互結合型では、一時的ではあるが、負荷の不均衡が生ずる。ここでは、プロセッサへ上で実現する時に、負荷の不均衡がボトルネックにならないよ

うに、各プロセッサでの発火制御部に FIFO を置いている。
 図1に本模擬手法を動的データ駆動型プロセッサ上に実現した構成例を示す。リング型トポロジに接続したプロセッサには、それぞれ全ニューロンをできるだけ等しい数に分割し、配置する。発火したあるニューロンは、(次に発火するニューロン番号、発火したニューロン番号、発火出力値の差分)を、リング上の各プロセッサに送る。各プロセッサは、割り当てられたニューロンの膜電位補正を行なうと同時に、次に発火するニューロンがそのプロセッサに割り当てられている場合、発火処理を行なう。

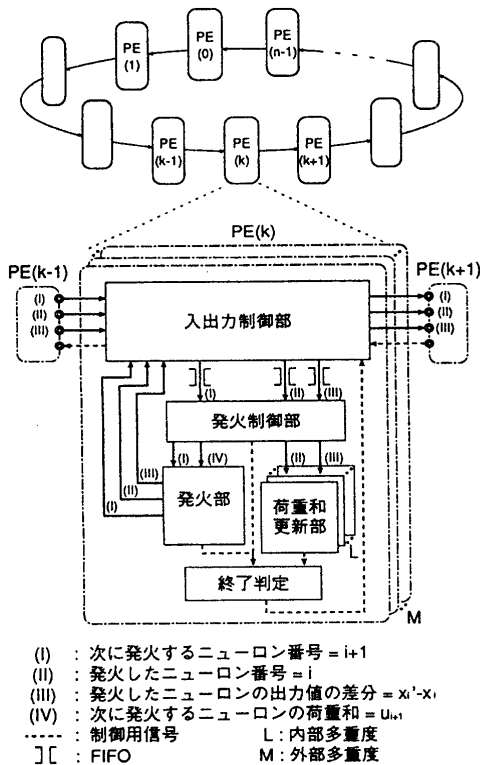


図1: プロセッサ・トポロジとプロセッサ内部での処理

動的データ駆動型プロセッサの特徴から、結合荷重の読みだしなどのEDS(外部記憶域)へのアクセスと、膜電位補正などの数値演算処理を、文脈の切り替え無しにオーバーラップさせることができるので、EDSアクセスの遅延(latency)が全体の実行時間に影響を与えない。したがって、非常に効率の良い処理が実現されている。

4 本方式の評価

本方式を、プロセッサ・シミュレーションで評価した結果を図2に示す。ここでは内部多重度は全て4としている。ここで注目したいのは、ある結合密度(今回の実験では67%)以下で、2節で述べた完全相互結合向き実現法に比べて今回提案した実現法が有効で、この境界値がニューロン数が増加しても一定なことである。すなわち、このことは大規模な網にも対応できることを意味する。例えば、10都市の巡回セー

ルスマン問題では、結合密度は36%であり、従来提案した実現法に比べて、実行速度では1.7倍の改善がみられた。

すなわち、結合密度が小さい場合には、(a),(b)に要する遅延に対して、各ニューロンが発火する際の処理数を減少させることができるため全体としての処理数は減少し、全体のシミュレーション時間が短縮される。

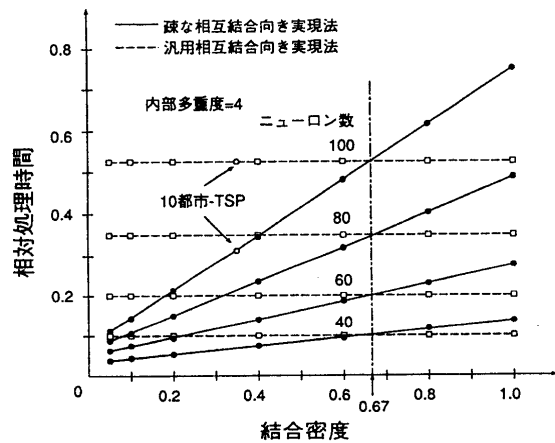


図2: 結合密度による処理時間の変化

5 おわりに

疎な相互結合型神経回路網の動的データ駆動型並列模擬手法を提案し、従来提案していた完全相互結合型神経回路網向き手法との比較を行なった。ある結合密度以下では本手法が有効であり、また、ニューロン数の増加に対するオーバーヘッドは見られず、大規模な網にも対応できることが分かった。

結合が非常に疎な場合には、ニューロンモデル自体の非同期動作という制約があっても、ニューロン同士の同時発火ができる可能性がある。これについては今後の課題である。

なお、相互結合型とは別に、階層型神経回路網の動的データ駆動型並列実現法についても検討を進めている[2]。

本研究で用いた動的データ駆動型プロセッサQ-v1[3]は、シャープ(株)ならびに三菱電機(株)から提供を受けた。両社の関係各位に厚く御礼申し上げる。

参考文献

[1] 水野正之, 青山和之, Ali M. Al-Haj, 寺田浩詔: “動的データ駆動型マルチプロセッサによる神経回路網処理”, 電気関係学会関西支部連合大会, S3-5, pp.S17 (1991-11).
 [2] Ali Mohammad Alhaj and Hiroaki TERADA: “A Data-driven Implementation of Back Propagation Learning Algorithm”, Proc. IJCNN(Int'l. Joint Conf. on Neural Networks) pp. II-588-II-593 (June 1992).
 [3] Shinji KOMORI, Kenji SHIMA, Souichi MIYATA, Toshiya OKAMOTO and Hiroaki TERADA: “The Data-Driven Micro-processor”, IEEE Micro, 9, 3, pp.45-49 (June 1989).