ご購入はこちら

第1章 行列演算やニューラル・ネットワーク

並列にすると速くなる処理

鈴木 量三朗

GPUに注目の理由

● アクセラレータとして使いやすくなった

もともとGPU(Graphical Processing Unit)は2D/3Dグラフィックス用の処理エンジンでした．快適なゲーム環境などを提供するアクセラレータとして専用のドライバおよびライブラリとともに存在していました．

その主用途であるゲームが3Dの座標計算を要求するため，GPUは32ビット浮動小数点数の演算に強いものでした．近年，それらを科学技術計算[注1]を含めた一般用途でも利用するための手段が用意され始めました．

そのような用途でGPUを使うことに対して，当初はGPGPU(最初のGPはGeneral Purposeの意味)という単語が使われていましたが，今では単にGPUと総称されており，画像処理ばかりでなくAI関連の処理には欠かせない技術になっています．

● CPUの動作クロックは頭打ち

コンピュータの性能を示す指標のCPUクロックは，上限がないかのように向上していった時期がありました．しかし，もはやそういう時代は終わりました．昨今，5GHzで動作するCPUがあるものの，CPUの動作クロックの向上は頭打ちとなっており，10GHzで動作するCPUが現れるのはいつになるか分かりません．

抱えている性能問題はよいアルゴリズムや良いプログラムでソフトウェアを改良することで解決していかなければなりません．

● 高速化のトレンドは並列処理

CPUの動作クロックが倍になることは期待できない一方で，CPUに載るコア数や，GPUのコア数の増加には期待ができます．

そこで，既存のプログラムを並列で動かすことができれば処理時間を短縮できるかもしれません．並列処理と言えばGPUやFPGAといった特別なハードウェアの名称を思い浮かべる方も多いかと思います．

GPU以外にも同時に複数の処理ができるハードウェアはあります．最近のCPUはコアがたくさん搭載されていますし，FPGAも並列に処理ができます．これらを使うには，並列処理というものを理解して，演算対象に応じて適切な選択をする必要があります．

特集 第2部ではGPUを使った並列演算，とりわけGPUが得意とする32ビット浮動小数点演算の並列演算にGPUプログラムの代表の1つであるCUDAを使ってトライしていきます．

NVIDIAを使うならCUDAがお勧め

● GPUベンダが作った専用の開発環境

開発元であるNVIDIA社によると，CUDA(クーダ)(Compute Unified Device Architecture)は，GPU上の汎用コンピューティング向け並列コンピューティング・プラットフォームとプログラミング・モデル[注2]だそうです．

この開発環境では，GPUを一般の演算用途にも使える仕組みが用意されています．これは市販されているグラフィックス・ボードのうち，NVIDIAのGPUを搭載したものが対象となります，CUDAの開発環境は無償で入手できます．

● デメリット…NVIDIAのGPUでしか使えない

一方，CUDAを利用できるのはNVIDIAのボードだけというデメリットもあります．ベンダに依存しないGPU向けの並列コンピューティング環境としては，OpenCLが有名です．

注1：科学技術計算では64ビット浮動小数点数を使うことが多いでしょう．その場合，GPUによっては性能が十分ではないものもあるため注意が必要です．

注2：英文で次のように紹介されています．CUDA(R) is a parallel computing platform and programming model developed by NVIDIA for general computing on graphical processing units (GPUs).