

平成 21 年度 情報工学コース卒業研究報告要旨

高木 研究室	氏 名	安 永 卓 矢
卒業研究題目	混合一括並列処理による 省メモリなモチーフ探索回路	
<p>隠れマルコフモデル (Hidden Markov Model, HMM) は、その時間的な処理に対する柔軟性から音声認識はもとより、今日では分子生物学の分野でも広く使われている。その応用例の 1 つにプロファイル HMM と呼ばれるタンパク質ファミリーの確率的モデルを用いたタンパク質の機能特定がある。機能が未知のタンパク質配列 (シーケンス) が、機能の知られたタンパク質配列パターン (モチーフ) のプロファイル HMM と比較され、シーケンスがそのモチーフを含んでいるかどうか探索される。この探索はモチーフ探索と呼ばれ、疾患遺伝子の特定や創薬などに応用されている。汎用計算機を用いたモチーフ探索には数時間から数日かかることもあり、その高速化が必要とされている。</p> <p>プロファイル HMM を用いたモチーフ探索において、高速化が必要とされる処理の 1 つに、ダイナミックプログラミングである Viterbi アルゴリズムを用いてあるプロファイル HMM とシーケンスの比較を複数のシーケンスについて行う処理がある。この処理は代表的なモチーフ探索プログラムの 1 つである HMMER においては <code>hmmsearch</code> と呼ばれており、FPGA を用いたハードウェア化による高速化の研究がなされている。FPGA を用いる場合、処理に必要なデータを回路内のメモリに保持し、複数の PE (Processing Element) を使い並列に計算を行うことで高速化を実現する。</p> <p>本研究では、モチーフ探索に適した混合一括並列処理とそれを用いた省メモリなモチーフ探索回路を提案する。提案する並列処理は 1 つのシーケンスに対する処理に必要なデータ (スコア) がそろってから計算を始め、シーケンスの各シンボルは順次ロードする方式 (ストリーム型一括並列処理) と、現在の計算を行いながら次のスコアを順次ダウンロードし、処理するシーケンスはあらかじめ回路内に保持しておく方式 (ストア型一括並列処理) を混合したものである。提案する手法は、現在の計算を行いながら次の計算に必要なスコアを順次ダウンロードすると同時に、シーケンスの各シンボルを順次ロードしていく方式である。また、提案するモチーフ探索回路は、計算の実行に必要な全てのスコアを回路内メモリに格納するのではなく、現在行っている計算と次の計算に必要なスコアのみを格納する回路内メモリを有する構成をとる。</p> <p>この提案するモチーフ探索回路とストリーム型一括並列処理によるモチーフ探索回路の比較を、回路内のメモリサイズと 1 つのシーケンスの処理に必要なクロックサイクル数について行った。提案回路における PE 数をストリーム型回路と同数の 100 として比較を行うと、クロックサイクル数は約 1% 増加するがメモリサイズは約 1/13 に削減できることが分かった。また、提案回路における PE 数を 25 とすると、メモリサイズはストリーム型回路の約 1/27 であった。なお、ストリーム型一括並列処理と混合一括並列処理で用いた PE は同じものである。</p> <p>提案手法を用いることによってモチーフ探索回路のメモリ量の削減が実現できる。また、省メモリ化により、モチーフ探索回路の小面積化が実現可能で、さらに、提案回路を複数個用いた回路構成が可能になり、それによりモチーフ探索の高速化も期待できる。例えば、先の PE 数 25 の提案回路を 27 個使った回路の性能はメモリサイズが PE 数 100 のストリーム型回路と同程度で、総 PE 数は 675 個になり、複数のシーケンスの処理に必要なクロックサイクル数は約 4/27 になる事が予想される。</p>		