

# 粒子フィルタを用いたクラウド計算サービス

予測発見戦略研究センター

特任研究員 長尾 大道

## 1 序

High Performance Computing (HPC)分野においては、ここ数年間の潮流であった地球シミュレータや次世代スーパーコンピュータに代表されるような理論性能が数PFlopsないし数十PFlopsに及ぶ単体システムの開発は非常に高価であるため、比較的安価に高い理論性能が得られるGPUを科学計算に積極的に利用したり、ネットワークを介して各研究機関のスーパーコンピュータを繋ぎ、超多並列計算を実現するためのクラウドコンピューティングシステム(CCS)を構築しようとする機運が高まり始めている。一方で、ベイズ的アプローチによる時系列データ解析の需要はますます高まっており、膨大な計算量を捌くためにこれらのHPC環境を積極的に利用することが当然考えられる。しかしながら、計算の精度、コーディングの手間、および計算時間の3つのバランスをいかに取るかは、ベイズ的時系列解析においては常に付き纏う問題であり、特に非ガウス分布を採り入れた多変量解析では非常に大きな計算コストを要求されるため、手間のかかるシステムやコーディングは敬遠されがちで、ベイズ的アプローチ自体を諦める研究者も少なからずいるのが実情である。そのため、並列計算機上で時系列解析をなるべく手軽に実施することができるプラットフォームを提供することは、非常に重要であると考えられる。

このような最近の研究動向を受け、本研究では将来的にCCSにインストール可能な、粒子フィルタを基盤とする時系列解析ソフトウェア “CloCK-TiME” (Cloud Computing Kernel for Time-series Modeling Engine) を開発している。本ソフトウェアは、ベイズ的アプローチが常に考えるべき上記3つのポイントに対して、解決の方向性を与えるものであると考えられる。まず計算の精度を得るためには単純に粒子数を増やせば良く、並列数を必要なだけ増やすことを簡単に可能にするというCCSの理念に沿っている。また粒子フィルタは並列計算機に実装しやすいアルゴリズムとなっているため、CCSの各拠点にインストールする際には特別なチューニングは必要としない。さらにはCCSを利用する際の最大の技術的障壁としては、ノード間通信コストが大きくなり、結果として実行時間が長くなることが考えられるが、粒子フィルタでモデルパラメータの探索や分布を計算する際には、ノード間通信の多発を抑えることが可能である。

本稿では、間もなくプロトタイプを公開開始予定である CloCK-TiME について、その概要を紹介する。

## 2 ソフトウェアの概要

図1は、CloCK-TiMEのシステム概念図を示したものである。CloCK-TiMEはFlash Playerを用いて開発されたユーザインターフェース(I/F、図2)を持っており、ユーザはI/Fを介して時系列データのアップロード、解析パラメータの設定、並列計算機による解析の実行、さらには解析結果の表示やダウンロードが可能である。CloCK-TiMEの解析エンジンでは、ユーザから与えられた多変量時系列データ  $y_t^l$  を、トレンド成分  $u_t^l$ 、季節成分  $s_t^l$ 、自己回帰(AR)成分  $p_t^l$ 、および観測ノイズ成分  $w_t^l$  に分解する。即ち、観測モデルは、

$$(1) \quad y_t^l = u_t^l + s_t^l + p_t^l + w_t^l$$

ただし、 $t$ は時刻、 $l$ は変数を表す。また状態ベクトル  $x_t$  は

$$(2) \quad x_t = \left( u_t^l \quad \cdots \quad u_{t-k+1}^l \quad s_t^l \quad \cdots \quad s_{t-j+2}^l \quad p_t^l \quad \cdots \quad p_{t-m+1}^l \right)'$$

と定義する。ただし、 $k$ はトレンド次数、 $j$ は季節成分の周期、 $m$ はAR次数であり、 $'$ は転置

を表す。最適化するモデルパラメータは、

$$(3) \quad \theta = (A_1 \ \cdots \ A_m \ \tau_u^l \ \tau_s^l \ \tau_p^l \ \sigma^l \ x_0)$$

ただし、 $A_i (i=1, \dots, m)$  は AR 係数行列、 $\tau_u^l$ 、 $\tau_s^l$ 、 $\tau_p^l$  は各成分のシステムノイズの分散、 $\sigma^l$  は観測ノイズの分散、および  $x_0$  は初期状態である。すなわち状態ベクトルの次元は  $(k+j-1)l+m$ 、未知パラメータ数は  $(k+j+3)l+m(l^2+1)$  となる。

CloCK-TiME で設定可能な主な解析パラメータは、解析モード、トレンド次数、季節成分の周期、AR 次数、システムノイズの有無および分布形である。解析モードとしては、解析エンジンが自動的に設定した事前分布  $p(\theta)$  からサンプリングした  $\theta$  から最適なものを探索する「ランダムサーチモード」、およびそこからマルコフ連鎖モンテカルロ法(MCMC)の一種である Adaptive Direction Sampling アルゴリズム(Roberts and Gilks (1994))を用いて事後分布  $p(\theta|y_{1:T})$  を求める「事後分布推定モード」がある。特に AR 係数行列については、事前分布をどのように与えるかが問題となるが、ここでは入力された時系列から大まかにトレンド成分と周期成分を差し引いたものに対し、Yule-Walker 法で推定した多変量 AR モデルを事前分布として与え、Lehman-Schur 法を用いて、AR 特性根が常に定常条件を満たすように監視するようにしている。また AR 成分は周波数領域における特徴をよく表現できることを活かし (例えば、Higuchi (1991))、オプションとして AR 周期を指定することも可能である。

解析が終了すると、最適パラメータの一覧表、時系列、各パラメータの事前/事後分布、およびスペクトルやコヒーレンシーが、I/F 画面に表示される。ユーザは、これらのプロットおよび数値データファイルをダウンロードすることができる。

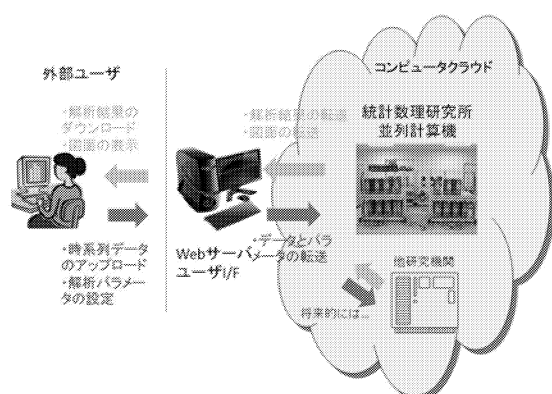


図 1: CloCK-TiME の概念図。

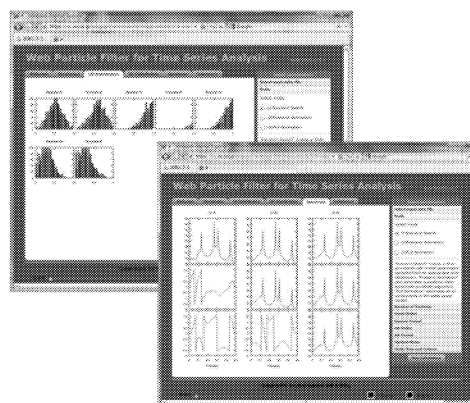


図 2: CloCK-TiME のユーザインターフェース。

### 3 今後の展開

CCS の構想は以前よりあったが、具体的な構築の動きを見せつつあるのはごく最近のことであり、CloCK-TiME が CCS 上で動くのはまだ当分先のことでありと考えられる。まずはプロトタイプの本研究所内での試用運転を開始し、所内外からの要望やアドバイスに基づいて改善しながら、将来 CCS 上で躍動するようなソフトウェアに仕立て上げたいと考えている。

### 参考文献

Higuchi, T. (1991), Frequency domain characteristics of linear operator to decompose a time series into the multi-components, *Ann. Inst. Statist. Math.*, **43**, 469-492.  
 Roberts, G. O. and W. R. Gilks (1994), Convergence of Adaptive Direction Sampling, *J. Multi. Anal.*, **49**, 287-298.