高並列度計算機を用いた状態推定

予測発見戦略研究センター　データ同化グループ

助教　　中野　慎也

1 Introduction

Recently, sequential Bayesian estimation comes to be applied to high-dimensional problems with thousands or millions of variables. One important example are found in numerical weather prediction in meteorology. In numerical weather prediction, a numerical simulation model with millions of variables is employed to describe the temporal evolution of a complex atmospheric system. In order to obtain a credible prediction, it is necessary to incorporate meteorological measurement data into the numerical simulation model. For the purpose of this integration, meteorologists employ a variety of methods including a class of the methods based on sequential Bayesian estimation. In particular, since the assumption of Gaussianity is often invalid in atmospheric systems due to the high nonlinearity, the PF is expected as one of hopeful methods for incorporating meteorological measurement data into the numerical simulation model.

Assuming that both the prior and posterior PDFs are Gaussian, the posterior PDF at each time step can be obtained using the Kalman filter algorithm. However, the assumption of Gaussianity is invalid in many cases. A most common algorithm for such non-Gaussian problems is the particle filter (PF), which is sometimes referred to as the Monte Carlo filter or sampling importance resampling (SIR) filter.. The PF is a population-based algorithm which approximates a PDF by a population consisting of a large number of samples drawn from the PDF. Each sample is called a `particle' in the PF. An approximation of the posterior PDF is obtained using a procedure similar to the evolutionary algorithms. First, fitness of each particle to the current measurements is evaluated. Then, according to the fitness, some of the particles are abandoned and some of the particles are selected and duplicated. The number of the duplicates is determined in a Bayesian manner so that the population after the selection offers an approximation of the posterior PDF. In the PF, this selection procedure is called `resampling'. Now the PF is widely used for various sequential Bayesian problems in various fields.

In sequential Bayesian estimation for high-dimensional models, most of computational cost is due to the calculation of the temporal evolution of a state according to a system model. In the population-based algorithms such as the PF and the MPF, this temporal evolution for each particle can be calculated separately. Therefore, the population-based algorithms are basically suitable for parallel computing. However, the PF could require frequent communication between processing elements (PEs) every step if they are naively implemented on parallel computing systems. When these algorithms are applied to high-dimensional models which tend to require a large population size, the `inter-PE' communication can seriously impair the computational efficiency.

In this paper, we propose a hierarchical method similar to island models in the genetic algorithm for the purpose of reducing the inter-PE communication. The population is divided into multiple sub-populations and each sub-population is assigned to each PE. The PF procedure is applied to each sub-population locally at each time step. Meanwhile, fitness for each sub-population is calculated as an averaged fitness of the particles in the sub-population. If the fitness for some of the sub-populations becomes too low, the selection of the sub-populations is done in addition to the local procedure for each sub-population. The selection of the sub-populations can be done in a Bayesian manner so that the whole population which includes all the sub-population under consideration offers a Bayesian estimate of the posterior PDF. If fitness shows similar values among the sub-populations, the selection of the sub-populations can be skipped to avoid the communication between PEs.

2 Hierarchical method

In the PF, communication occurs in the filtering step. Some of particles with low fitness are abandoned and replaced by copies of particles with high fitness. Whenever the copy substituted for the abandoned particle is taken from a different PE, the two PEs must communicate with each other. Such inter-PE communication occurs many times every step, and it can seriously impair the computational efficiency.

We then propose the following strategy in order to reduce inter-PE communication. First, we divide the population into multiple sub-populations each of which is assigned to a different PE. Then, the PF procedure is applied to each sub-population; that is, the resampling or merging is performed locally within each PE at each time step. On the other hand, a weight of each sub-population is calculated as an averaged weight of the particles in the sub-population. If one or some of the weights for the sub-populations becomes too large in comparison with the other sub-populations, the selection of the sub-populations is carried out. If the weights take similar values among the sub-populations, the selection of the sub-populations can be skipped in order to avoid the network traffic between processing elements.

Since resampling is performed only occasionally, traffic among PEs can remarkably be reduced. The computational speed is remarkably improved using the proposed algorithm especially for the case with a large number of processing elements. The proposed algorithm can also achieve comparable or slightly better accuracy in comparison with the naïve PF. This hierarchical algorithm affords an opportunity to efficiently estimate temporal evolution of state for high dimensional complex systems.