Simulation Modelling Practice and Theory 85 (2018) 1–14

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

journal homepage: www.elsevier.com/locate/simpat

On-demand data assimilation of large-scale spatial temporal systems using sequential Monte Carlo methods

Feng Gu[1]

Department of Computer Science, College of Staten Island, 2800 Victory Boulevard, Staten Island, NY 10314, USA

ARTICLE INFO

Article history:
Received 10 November 2017
Revised 24 March 2018
Accepted 26 March 2018

Keywords:
Data assimilation
Dynamic systems
Sequential Monte Carlo methods
Wildfire spread simulation
State estimation

ABSTRACT

The proliferation of sensors is generating rapidly increasing quantities of data like never before. These extensive amounts of data can provide useful information for more accurate state inference of large-scale spatial temporal systems. Sequential Monte Carlo methods are used to assimilate the observed data from sensors to improve the state estimation of large-scale spatial temporal systems, which highly rely on the available real time observation data. In many scenarios, the real time data are limited in space and time. Therefore, it is important to effectively obtain critical sensor data in real time and then dynamically feed them into the running model. In this paper, we propose the on-demand data assimilation method for large-scale spatial temporal systems, in which we quantify the spatial states using run-time state quantification methods and decide if we need to trigger data assimilation on demand and obtain more relevant real time data when the state uncertainty is high. The effectiveness of the developed framework is evaluated based on large-scale wildfire spread simulations.

1. Introduction

The proliferation of sensors is producing rapidly increasing data in quantity and variety like never before. The extensive amount of data on one hand provide useful information for spatial temporal systems, and on the other hand pose challenges on effectively and efficiently utilizing the information for better state inferences to support real time decision-making. Considering the example of a wildfire spread and firefighting scenario, when a wildfire occurs, various types of real time data, including data from satellite systems, multi-spectrum infrared airborne images collected by unmanned or remotely controlled aerial vehicles and the temperature data collected from ground fire sensors will be provided to wildfire managers. Therefore, how to utilize these data and integrate them into simulation models to accurately estimate the evolving fire fronts and the fireline intensity will help the wildfire managers make decisions in wildfire suppression. Another example is collaborative target inference and target tracking in a battlefield. When targets appear, streams of sensor data collected from disparate sensors are transferred to the command center. Effective methods to assimilate these sensor data in space and time for real time target inference are crucial for system operators and commanders to decide the actions to take.

Both of the above systems are spatial temporal dynamic systems, exhibiting dynamic behaviors in both space and time. Many applications fall in the above spatial temporal systems, such as target tracking, wildfire spread, nuclear radiation dispersion, and chemical/biological release spread. They are inherently difficult to study due to reasons such as involving large

E-mail address: Feng.Gu@csi.cuny.edu
[1] URL: http://www.cs.csi.cuny.edu/~gu.

https://doi.org/10.1016/j.simpat.2018.03.007
1569-190X/

spatial areas, the stochastic process of system dynamics, and complex interactions among many different factors, e.g., vegetation, terrain, and weather in a wildfire. Another common feature of the above systems is the need of accurate state inference (or state estimation) from spatial temporal sensor data. To support state inference, dynamic models that capture systems' behaviors are developed. For example, in collaborative target inference and target tracking, models of targets' motions are developed. Real time sensor data are then assimilated into these models to infer targets' states, e.g., their positions and velocities (see [1–3] for examples). Similarly, in the wildfire spread example, simulation models are developed to predict the fire spread behavior over time. Real time sensor data are assimilated into the simulation models to estimate the dynamically evolving fire fronts (see [4–7] for examples). By infusing spatial temporal information and assimilating real time sensor data into dynamic models, we can better estimate the current system state and thus provide more accurate behavior analysis and prediction. Furthermore, the real time sensor data also carry "feedback" information for us to calibrate the model parameters in order to better match the model outputs with observations. These capabilities are very useful for supporting real time system analysis and decision-making.

To support data assimilation of spatial temporal data into dynamic models, novel approaches are needed to effectively and efficiently incorporate the real time data. Many of such approaches are designed based on probabilistic methods because spatial temporal systems are often stochastic and dynamic, such as Kalman filter [8] and sequential Monte Carlo (SMC) methods [9]. SMC methods are sample-based, utilizing Bayesian inference and stochastic sampling to recursively estimate the state inference of dynamic systems under study from the given observations. SMC methods use a large number of random samples and their associated weights to represent the sequence of probability distributions of interest. With the arrival of observation data, the weights of samples are updated. SMC methods are able to represent any probability densities and have no restrictions on the properties of the system model. Therefore, SMC methods are effective in supporting data assimilation of large-scale spatial temporal systems with sophisticated models and complex non-linear, non-Gaussian, and unsteady behaviors [5,6]. In addition, SMC methods are executed in a recursive manner that the state inference of systems is updated when new observations arrive. This pattern well matches the feature of data assimilation of dynamic systems that the availability of new real time data drives the updates of the states of the systems.

Over the past several decades, many research efforts of applying SMC methods to large-scale spatial temporal systems focus on their large state space and the required intensive computational costs [10–12]. With the proliferate types of sensor data available, they can be used to improve the accuracy of data assimilation using SMC methods. In the scenario of wildfire spread, the primary real time data are the temperatures from the deployed ground fire sensors. However, in some cases, the wildfires spread to the field without fire sensors, or the quality of the collected temperatures is low due to various reasons, such as a low number of sensors available and malfunctioned sensors. This decreases the accuracy of the data assimilation results. To improve the predicted results, additional types of available data can be used, such as satellite image data and multi-spectrum infrared airborne photographs from unmanned or remotely controlled aerial vehicles. The availability and heterogeneity of these real time data ask for methods that can integrate multiple types of spatial temporal data to support data assimilation using SMC methods. In this paper, we develop the on-demand data assimilation framework of large-scale spatial temporal systems using SMC methods, which obtains multiple types of real time data and dynamically feeds them into the simulation model to improve the state inference.

The rest of the paper is organized as follows. Section 2 introduces the related work in SMC methods and their applications. Section 3 describes the foundations of SMC methods. Section 4 presents the on-demand data assimilation framework using SMC methods. Section 5 illustrates data assimilation of wildfire spread simulation using SMC methods. Section 6 provides the experimental design and results. Section 7 concludes the paper and points out the future research directions.

2. Related work

Data assimilation is an analysis technique used in numerous application domains, such as meteorology and oceanography, weather forecasting, geosciences, and environmental systems. It aims to obtain the best estimates of states of dynamic systems by infusing the observational data into the system models. Many different methods have been developed to support data assimilation, among which the variational approach and the probabilistic approaches are widely used. The variational approach minimizes the cost functions containing observations, and it eliminates additional initialization steps. Three-dimensional variational analysis (3D-VAR) and four-dimensional variational assimilation (4D-VAR) are two of the popular variational approaches. 3D-VAR achieves data assimilation by iteratively minimizing a predefined cost function and penalizing the differences between the analysis and observations according to their estimated errors. 4D-VAR generalizes 3D-VAR by handling the observations distributed in time. It uses the same cost function, but includes a forecast model that compares the model state and the observations at different time steps. Their related applications can be found in many literatures, such as [13] and [14]. The probabilistic approach mainly includes filtering methods and smoothing methods, such as Kalman filter and particle filters. Kalman filter [15] estimates states of a system represented by a linear state space model with the given observations. For applications with non-linear behaviors, Kalman filter needs to be extended [16]. Its broad applications can be found in [17–20].

Another popular probabilistic approach of data assimilation is SMC methods, also called particle filters. SMC methods approximate the states of dynamic systems by a set of particles and their associated weights. They can be applied to dynamic systems with non-linear behaviors and have little restrictions on the system's properties. Therefore, SMC methods are popularly adopted in many applications, such as civil engineering, finance, oceanography, smart buildings, weather pre-