ARTICLE IN PRESS

Mechanical Systems and Signal Processing xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Probabilistic learning of nonlinear dynamical systems using sequential Monte Carlo

Thomas B. Schön *, Andreas Svensson, Lawrence Murray, Fredrik Lindsten

Department of Information Technology, Uppsala University, Sweden

A R T I C L E I N F O

Article history:
Received 2 February 2017
Received in revised form 26 September 2017
Accepted 22 October 2017
Available online xxxx

Keywords:
Probabilistic modeling
Nonlinear dynamical systems
System identification
Parameter estimation
Bayesian methods
Metropolis-Hastings
Sequential Monte Carlo
Particle filter

A B S T R A C T

Probabilistic modeling provides the capability to represent and manipulate *uncertainty* in data, models, predictions and decisions. We are concerned with the problem of learning probabilistic models of dynamical systems from measured data. Specifically, we consider learning of probabilistic nonlinear state-space models. There is no closed-form solution available for this problem, implying that we are forced to use approximations. In this tutorial we will provide a self-contained introduction to one of the state-of-the-art methods—the particle Metropolis-Hastings algorithm—which has proven to offer a practical approximation. This is a Monte Carlo based method, where the particle filter is used to guide a Markov chain Monte Carlo method through the parameter space. One of the key merits of the particle Metropolis-Hastings algorithm is that it is guaranteed to converge to the "true solution" under mild assumptions, despite being based on a particle filter with only a finite number of particles. We will also provide a motivating numerical example illustrating the method using a modeling language tailored for sequential Monte Carlo methods. The intention of modeling languages of this kind is to open up the power of sophisticated Monte Carlo methods—including particle Metropolis-Hastings—to a large group of users without requiring them to know all the underlying mathematical details.

1. Introduction

The true value of measured data arises once it has been analyzed and some kind of knowledge has been extracted from this analysis. The analysis often relies on the combination of a mathematical model and the measured data. The model is a compact representation—set of assumptions—of some phenomenon of interest, and establishes a link between that phenomenon and the data, which is expected to provide some insight. The knowledge we seek is typically a function of some unknown variables or parameters in the model. However, any reasonable model will be *uncertain* when making statements about unobserved variables, and uncertainty therefore plays a fundamental role in modeling. Probabilistic modeling allows the representation and manipulation of uncertainty in data, models, decisions and predictions. The capability to mathematically represent and manipulate uncertainty—which is essential to the development throughout this tutorial—is provided by probability theory. A good introduction to the ideas underlying probabilistic modeling in contemporary machine learning is provided by [1] and from a system identification point of view we recommend [2].

* Corresponding author.
E-mail addresses: thomas.schon@it.uu.se (T.B. Schön), andreas.svensson@it.uu.se (A. Svensson), lawrence.murray@it.uu.se (L. Murray), fredrik.lindsten@it.uu.se (F. Lindsten).

https://doi.org/10.1016/j.ymssp.2017.10.033
0888-3270/

Throughout this tutorial we are concerned with the problem of learning probabilistic models of nonlinear dynamical systems from measured data, which is sometimes referred to as the nonlinear system identification problem. These models provide an interpretable representation from which it is possible to extract the knowledge we seek, compared to doing so directly from the data. The model is thus a natural middle ground between the measured data and that knowledge. More specifically we are concerned with the nonlinear state-space model (SSM)

$$x_t = f(x_{t-1}, u_t, v_t, \theta), \tag{1a}$$
$$y_t = g(x_t, u_t, \theta) + e_t, \tag{1b}$$
$$x_0 \sim p(x_0 | \theta), \tag{1c}$$
$$\theta \sim p(\theta), \tag{1d}$$

where $y_t \in \mathsf{Y}$ denotes the observed output and u_t denotes a known input signal. In the interest of concise notation we will, without loss of generality, suppress the input signal u_t. The unknown variables are the state $x_t \in \mathsf{X}$ describing the system's evolution over time and the static parameters θ. Furthermore, $p(\theta)$ denotes the prior assumptions about θ. The stochastic variables v_t and e_t encode noise, commonly referred to as process noise and measurement noise, respectively. Finally, the functions f and g encode the dynamics and the measurement equation, respectively.

The first step towards extracting knowledge from a set of measured data $y_{1:T} = \{y_1, \ldots, y_T\}$ is to learn a probabilistic model of the form (1) by computing the conditional distribution of the unknown $x_{0:T}$ and θ conditioned on $y_{1:T}$, denoted $p(x_{0:T}, \theta | y_{1:T})$. This provides a useful representation which is typically much closer to the knowledge we seek than the measured data itself. Once we have a representation of this conditional distribution, it can be used to compute more specific quantities that typically constitute the end result of the analysis. To mention just a few examples of such quantities we have mean values, variances or estimates of some tail probability.

The key challenge is that, in general, there is no closed form expression available for $p(x_{0:T}, \theta | y_{1:T})$, so we must resort to approximations. We will focus on approximations based on Monte Carlo sampling which, while being costly to compute, have the appealing property of converging to the true solution as the amount of computations increases. Over the last decade, these approximations have evolved rapidly, so that we now have computationally feasible solutions available.

This naturally brings us to the aim of this tutorial, which is to provide a gentle introduction to probabilistic learning of nonlinear dynamical systems, and to introduce in some detail one of the current state-of-the-art methods to do so. This method relies on the systematic combination of two Monte Carlo algorithms, where a sequential Monte Carlo algorithm is used to compute a good proposal distribution for a Markov chain Monte Carlo (MCMC) algorithm. Hence, in terms of methods, this tutorial is focused on introducing one particular solution rather than surveying all available methods (see e.g. [3,4] for recent accounts of that sort). However, the key ideas discussed in this tutorial (in the context of the specific method) are in fact central to many other state-of-the-art Monte Carlo learning methods as well. For example the particle filter itself, and the fact that it is capable of producing an unbiased estimate of the likelihood, are also of more general interest. In the accompanying paper [5], we show how probabilistic modeling, implemented via a new algorithm of the type outlined in this tutorial, is used to solve one of the challenging benchmark problems in this special issue with promising results. We will also hint at the tailored software that is being developed to make these mathematical tools available to a much wider audience without a thorough knowledge of Monte Carlo methods. Such software allows the user to focus entirely on the modeling problem and leave the computational learning problem to the software.

In Section 2 we introduce probabilistic nonlinear state-space models in more detail. The model as it is stated in (1) clearly incorporates uncertainty due to the presence of the noise sources v_t and e_t, as well as uncertainty in the initial state x_0 and in the parameters θ. In probability theory, uncertainty is represented using random variables and in Section 2 the probabilistic nature of the model will be made even more explicit when we represent it as a joint distribution $p(x_{0:T}, \theta, y_{1:T})$ of all the random variables present in the model. The basic Monte Carlo idea is then introduced in Section 3 together with an explanation of how this idea can be used to learn the conditional distribution of the parameters given the measurements $p(\theta | y_{1:T})$. The idea is developed further in Section 4, where it becomes clear that we also need information about the unknown state variables, resulting in the introduction of the sequential Monte Carlo method (a.k.a. the particle filter) to estimate the state variables. In Section 5, the particle filter is used inside the MCMC method introduced in Section 3. The basic particle filter construction from Section 4 can be improved in several ways and in Section 6 we discuss some of the most important developments in this direction. The resulting method is then illustrated using a nonlinear spring-damper system in Section 7. Finally we conclude with a discussion in Section 8.

2. Probabilistic modeling of dynamical systems

We will refer to the joint distribution of all observed (here $y_{1:T}$) and unobserved (here $x_{0:T}$ and θ) variables as the *full probabilistic model*, which in our present setting amounts to $p(x_{0:T}, \theta, y_{1:T})$. The idea of using the mathematics of probability to represent and manipulate uncertainty is commonly referred to as Bayesian statistics ([6]). In order to write down the full probabilistic model for (1) let us start by noticing that for many models we can express the conditional distribution of y_t given x_t and θ as

$$p(y_t | x_t, \theta) = p_{e_t}(y_t - g(x_t, \theta), \theta), \tag{2}$$