Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 15 (2011) 489 – 495

Procedia Engineering

www.elsevier.com/locate/procedia

Advanced in Control Engineering and Information Science

Overview of Nonlinear Bayesian Filtering Algorithm

ZHANG Feng[a,a*], XUE Wen-fang[a], LIU Xi[a]

[a]Integrated Information System Research Center, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China

Abstract

For many nonlinear dynamic systems, the choice of nonlinear Bayesian filtering algorithms is important. In this paper, we review the application of both optimal and suboptimal Bayesian algorithms for state estimation, and the latter is divided into four types, that is function approximation, numerical approximation, Gaussian sum approximation and sampling approximation. Then four typical suboptimal Bayesian algorithms are mainly discussed, that is EKF, STF, UKF and PF. Finally, characters of nonlinear Bayesian filtering algorithms are summarized, and further development is forecasted.

Selection and/or peer-review under responsibility of [CEIS 2011]

Keywords: *Nonlinear system; Bayesian filtering; Particle filter; Unscented Kalman filter;Strong tracking filter*

1. Introduction

System behaviors are reflected by the acquisition, keeping and change of state. Studies on a system mainly refer to research on the system state, the change of state and the transition between different states. The essence of tracking problems is estimating the time-varying states by measurements with noise. For multiple-dimension and nonlinear/non-Gaussian process, the state-space approach is very effective. Thus, this paper will mainly concentrate on the state-space approach to modeling dynamic system, and the focus will be on the discrete-time formulation. In order to analyze and make inference about a dynamic system, at least two models are required: First, a model describing the evolution of the state with time (the system

a* Corresponding author. Tel.: +86-010-82683298; fax: +86-010-82683230-
E-mail address: zf20087@139.com.

1877-7058 © 2011 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
doi:10.1016/j.proeng.2011.08.093

mode) and second, a model relating the noisy measurements to the state (the measurement model). State vector contains everything needed to describe the dynamic system, including target position, velocity and acceleration, so the key point of state-space approach is the determination of state vector. Measurement vector contains measurement information related to state vector, and is generally of lower dimension than state vector [1]. Bayesian approach provides a strict theory frame for state estimation. In Bayesian theory, one attempts to construct the posterior probability density function (PDF) of the system state based on all available information, including last measurements. Since this PDF contains all statistic information, it can be said to be the complete requirement to state estimation. Bayesian filtering theory is achieved by the mechanism of updating the target state based on new measurements.

2. Bayesian filtering theory

Bayesian theory was proposed by the British researcher Thomas Bayes in a posthumous publication in 1763. The well-know Bayes theorem reveals the fundamental probability law governing the process of logical inference. However, Bayesian theory hasn't gained its deserved attention in the early days until its modern form was rediscovered by the French mathematician PierreSimon de Laplace [2]. Bayesian inference, devoted to applying Bayesian statistics to statistical inference, has become one of the important branches in statistics, and has been applied successfully in statistical decision, detection and estimation, pattern recognition, and machine learning. Bayesian filtering provides a solution to nonlinear state estimation problems based on probability distribution. Bayesian filtering views state estimation as a probability inference process, and convert the problems of states estimation to the problems of solving the PDF $p(X_k \mid Y_k)$ to get optimal states estimation. To define the problem of tracking, consider the evolution of the state sequence $\{x_k, k \in N\}$ of a dynamic target given by

$$x_k = f_k(x_{k-1}, w_{k-1}) \tag{1}$$

Where f_k is a possibly nonlinear function of state x_{k-1}, and w_{k-1} is an process noise sequence. The objective of tracking is to recursively estimate x_k from measurements.

$$y_k = h_k(x_k, v_k) \tag{2}$$

Where h_k is a possibly nonlinear function, and $\{v_k, k \in N\}$ is an measurement noise sequence. In particular, we seek filtered estimates of x_k based on the set of all available measurements. From a Bayesian perspective, the tracking problem is to recursively calculate PDF, given the measurement up to time k [3]. And finally, state estimation and covariance of its error is calculated respectively,

$$\hat{x}_k = E(x_k \mid Y_k) = \int x_k p(x_k \mid Y_k) dx_k \tag{3}$$

$$P_k = \int (x_k - \hat{x}_k)(x_k - \hat{x}_k)^T p(x_k \mid Y_k) dx_k \tag{4}$$

Because PDF contains all information of state vectors, it is of importance in filtering theory. If PDF at k-1 time, $p(x_{k-1} \mid Y_{k-1})$ and latest measurement, y_k is given, PDF at k time is easily calculated by Bayesian theory. Before calculating PDF, two hypotheses are given: 1) state vector is Markov process; 2) measure is independent of state vector.

$$p(x_k \mid Y_k) = \frac{p(y_k \mid x_k) p(x_k \mid Y_{k-1})}{p(y_k \mid Y_{k-1})} \tag{5}$$

As equation (5) shows, PDF consists of five parts [2].

1) Prior: The prior $p(x_k \mid Y_{k-1})$ defines the knowledge of the model

$$p(x_k \mid Y_{k-1}) = \int p(x_k \mid x_{k-1}) p(x_{k-1} \mid Y_{k-1}) dx_{k-1} \tag{6}$$

2) Likelihood: the likelihood $p(y_k \mid Y_{k-1})$ essentially determines the measurement noise model in the equation (2)
3) Evidence: The denominator involves an integral

$$p(y_k \mid Y_{k-1}) = \int p(y_k \mid x_k) p(x_k \mid Y_{k-1}) dx_k \quad (7)$$

Calculation or approximation of these three terms are the essences of the Bayesian filtering and inference.

3. Bayesian optimal filtering

Dynamic system is divided into linear dynamic system and nonlinear dynamic system. When the system is linear or it satisfies some specific constraint of Bayesian optimal filtering, Bayesian optimal filtering is the best choice. But optimal estimator is just optimal in the sense of MMSE, MAP, ML, and so on. MMSE is widely used in practice, and Kalman filter is a typical filtering algorithm in the sense of MMSE.

Before Kalman filter, K.Gauss proposed least square method in 1795, but least square method didn't perform well since it didn't contain statistics characteristic for measurements. In 1942, Weiner proposed Weiner filter. It fully utilized input variable and statistics characteristic for measurements, and it is a linear least variance filtering. Kalman filter is proposed in 1960, and founds the modern filtering theory. Kalman filtering learns state-space idea from modern control theory, and describes dynamic system model by state equation, and measurement model by measurement equation. It adopts recursive filtering approach in time domain, which is easily realized by computer. It supposes PDF is Gaussian distribution at every moment, thus we can use statistics such as mean and variance to describe the dynamic system. By Bayesian theory, if $p(x_{k-1}/z_{1:k-1})$ is Gaussian distribution, $p(x_k/z_{1:k})$ is Gaussian distribution, too. Other hypotheses includes: 1) w_k, v_k are both Gaussian white noise; 2) $f_k(x_{k-1}, w_{k-1})$ is a linear function of x_{k-1}, w_{k-1}, and its expression is known; 3) $h_k(x_k, v_k)$ is a linear function of x_{k-1}, w_{k-1}, and its expression is known, too.

4. Bayesian suboptimal filtering

State estimation problems of nonlinear dynamic system is of great value in scientific research and engineering application area, such as process control, maneuvering target tracking, environment supervision, fault diagnosis, communication and machine learning; hence, its theory and approach has been researched for a long time at home and abroad. In nonlinear dynamic systems, it is very complicated to calculate high-dimension integral. To relieve calculation stress of nonlinear Bayesian estimation, exiting nonlinear Bayesian filters usually adopt approximate approach in dealing with integral problems to get the suboptimal solution. Numerous approximation approaches for state estimation of nonlinear systems have been proposed. According to the approximation techniques, most of them can be classified into four types: function approximation, numerical approximation, Gaussian sum approximation and sampling approximation.

4.1. Function approximation

Function approximation techniques approximate a deterministic nonlinear function encountered in nonlinear filtering, which is most often the integrand of an expectation or part of a system model [4]. These techniques usually do not approximate an integral directly. Their best representations are those based on the Taylor series expansion (TSE). We denote the n_{th}-order TSE approximation of a function

$f(x,t)$ at $\hat{x}$ by

$$f(x,t) \approx TSE(x,t;n,\hat{x}) \\ = f(\tilde{x},t) + f'(\tilde{x},t)x + \frac{1}{2!}f''(\tilde{x},t)x^2 + \frac{1}{3!}f^{(3)}(\tilde{x},t)x^3 + \cdots + \frac{1}{n!}f^{(n)}(\tilde{x},t)x^n \tag{8}$$

Where $\tilde{x} = x - \hat{x}$, $f'(\tilde{x},t) = \frac{\delta f}{\delta x}|_{x=\tilde{x}}$, $\cdots$, $f^{(n)}(\tilde{x},t) = \frac{\delta^n f}{\delta x^n}|_{x=\tilde{x}}$

4.1.1. Extended Kalman filter

Bucy et al proposed a suboptimal filter applicable to nonlinear systems, called extended Kalman filter (EKF), which expands Kalman filter to nonlinear filtering area. EKF is based on TSE, and linearizes nonlinear characteristics of system model by parametric function approach. Its basic idea is linearizing nonlinear system partly by expanding 1_{st}-order TSE of state equation and measurement equation and then filtering using Kalman filter.

4.1.2. Strong tracking filter

Strong tracking filter (STF) chooses a proper time-varying gain, and makes

$$E\{[x(k+1) - \hat{x}(k+1)] \cdot [x(k+1) - \hat{x}(k+1)]^T\} = \min \tag{9}$$

$$E[\upsilon(k+1+j) \cdot \upsilon^T(k+1)] = 0, \qquad k = 0,1,2,\cdots, j = 1,2,\cdots \tag{10}$$

Where $\mathbf{\upsilon}(k)$ is the innovation sequence. We denote $\Phi(k, \hat{x}(k \mid k))$ as state transfer matrix, $H(k+1, \hat{x}(k+1 \mid k))$ as measurement matrix, $\Gamma(k)$ as state noise transferring matrix. Covariance matrix of state noise and measurement noise are denoted as $Q(k)$ and $R(k)$, respectively.

STF improves the performance of EKF mainly on adjusting covariance of state predictive error and filter gain automatically.

$$\mathbf{P}(k/k-1) = \lambda(k)\mathbf{\Phi}(k,k-1)\mathbf{P}(k-1/k-1)\mathbf{\Phi}^{\mathrm{T}}(k,k-1) + \\ \Gamma(k-1)\mathbf{Q}(k-1)\Gamma^{\mathrm{T}}(k-1) \tag{11}$$

Where $\lambda(k)$ is the fading factor of STF.

4.2. Numerical approximation

The typical numerical approximation is grid-based methods. It divides continuous state space into several grid units, and replaces the integral with sum of discrete variables to approximate PDF. If state space is continuous and can be divided into N units, that is, $\{x_k^i : i = 1, \cdots, N_s\}$, grid-based methods can be used to approximate PDF. We denote $p(x_{n-1} = x^i \mid y_{0:n-1}) = w_{n-1|n-1}^i$, and suppose PDF at k-1 time is known,

$$p(x_{k-1} \mid y_{1:k-1}) \approx \sum_{i=1}^{N_s} w_{k-1|k-1}^{i} \delta(x_{k-1} - x_{k-1}^{i}) \tag{12}$$

Then, prediction and update is as follows, respectively,

$$p(x_k \mid y_{1:k-1}) \approx \sum_{i=1}^{N_s} w_{k|k-1}^{i} \delta(x_k - x_k^{i}) \tag{13}$$

$$p(x_k \mid y_{1:k}) \approx \sum_{i=1}^{N_s} w_{k|k}^{i} \delta(x_k - x_k^{i}) \tag{14}$$

To get good approximation results, the density of grid must be high. So one of the drawbacks of grid-based methods is that calculation cost will rise rapidly as the increase of dimension of state space; another drawback is, as state space must be predefined, state space can't be divided evenly according to probability density area unless prior knowledge is used.

4.3. Gaussian sum approximation

Different from the linearized EKF that only concentrate on the vicinity of the mean estimate, Gaussian sum approximation uses a weighted sum of Gaussian densities to approximate the posterior density because single Gaussian distribution can hardly approximate posterior density effectively. It is based on hat any non-Gaussian distribution can be approximated by sum of as many Gaussian distribution as possible, so Gaussian sum approximation runs a bank of KF or EKF in parallel, and filters by updating mean, covariance and weighs of each Gaussian distribution.

We calculate the prior density and posteriori density of the nonlinear system described in equation (1) and (2) by Gaussian sum approximation approach, so the prediction and update is respectively,

$$\begin{aligned} p(x_k \mid y_{1:k-1}) &= \int p(x_k \mid x_{k-1}) p(x_{k-1} \mid y_{1:k-1}) dx_{k-1} \\ &= \sum_{i=1}^{N_s} \omega_{i,k-1} N(x_k, x_{i,k|k-1}, p_{i,k|k-1}) \end{aligned} \tag{15}$$

$$\begin{aligned} p(x_k \mid y_{1:k}) &= \frac{p(y_k \mid x_k) p(x_k \mid y_{1:k-1})}{\int p(y_k \mid x_k) p(x_k \mid y_{1:k-1}) dx_k} \\ &= \sum_{i=1}^{N_s} \omega_{i,k} N(x_k, x_{i,k|k}, p_{i,k|k}) \end{aligned} \tag{16}$$

Although Gaussian sum approximation approximates nonlinear functions better than EKF, it scatters easily as EKF for strong nonlinear system. Besides, its main drawback results from the choice of number of Gaussian function. If number of Gaussian function is too big, calculation cost rises; on the other hand, if number is too small, the performance deteriorates.

4.4. Sampling approximation

Sampling approximation approach approximates the posterior distribution of system state by a bank of samples; sampling approximation approximates integrated characteristics (mean and variance) as mentioned. According to sampling approach, sampling approximation approach is divided into deterministic sampling and stochastic sampling, which means its sampling points are deterministic or stochastic, and their typical representation is unscented Kalman filter and particle filter, respectively.

4.4.1. Unscented Kalman filter

Unscented Kalman filter (UKF) is proposed by Julier and Uhlmann et al of the University of Oxford in 1996. It represents prior and posterior distribution of stochastic states by Sigma points (a bank of deterministic samples) and transformed Sigma points. UKF still approximates posterior distribution of system state by Gaussian distribution, taking the mean and variance of some deterministic sampling points as that of Gaussian distribution. The approach that statistic characteristics of stochastic state are calculated to make nonlinear function of system state spread after nonlinear transformation is called Unscented Transform (UT). UT is based on the fact that PDF is easily approximated than nonlinear function. The key point of UT is the sampling strategy of Sigma points, including the amount, each position and weight of sampling points. Common sampling strategy includes symmetrical sampling, simplex sampling, 3-moments skew sampling, Gaussian 4-moments symmetrical sampling, proportional sampling and so on.

UKF has the similar algorithm structure with EKF, but what makes difference is that UKF uses nonlinear system model directly, updating by UT strategy and making posterior probability estimation precise to the fourth moment, compared with EKF which lineralizes nonlinear system and needs to calculate Jacoby matrix. However, UKF and EKF are both based on Gaussian approximation, both represent posterior distribution of system state by mean and variance of Gaussian distribution. When there are strong nonlinear or non-Gaussian characteristics in practice, actual distribution can't be accurately expressed by Gaussian approximation. As the development of computing science and the reduction of computing cost, particle filter becomes a hot in nonlinear/non-Gaussian stochastic system researches.

4.4.2. Particle filter

In the early 1950s, Hammersley adopted sequential importance sampling (SIS) based Monte Carlo methods to solve statistic problems. Because of the poor computing ability and the weights degeneration problems, SIS wasn't highlighted until the early 1990s, when Smith and Gordon et al brought re-sampling step to particle filter, which solved the weights degeneration problems to a certain extent, and proposed the sampling importance re-sampling (SIR) particle filter.

Particle filter uses plenty of stochastic samples, and achieves recursive Bayesian filtering process by Monte Carlo simulation. Particle filter expresses posterior distribution of system state by a couple of weighted stochastic samples. Its basic structure is choosing an importance probability density and sampling randomly, adjusting weights and particles' positions based on measurements after obtaining some weighted stochastic samples, then approximating posterior distribution of system state by these samples, and finally weighting these samples as the estimations of system states. Because particle filter isn't constrained by the assumption that system model is linear and Gaussian distribution, and describes PDF by sampling approximation approach instead of function approximation approach, it is applicable to any nonlinear non-Gaussian dynamic systems.

5. Conclusions

For a certain problem, if Bayesian optimal filtering satisfies the linearity and Gaussian distribution assumptions, then the best choice will be Kalman filter or grid-based filter. However, actual systems often don't satisfy linearity, and Bayesian suboptimal filtering is needed. EKF approximates system model to approximate PDF of system state by the form of Gaussian function; STF introduces the fading factor based on EKF, which makes the filter adjust divergence self-adaptively; grid-based approximation separates continuous state space into predefined discrete areas, but computing cost rises rapidly as the increase of the dimension of state space; Gaussian sum approximation approximates PDF by weighting

Gaussian distributions; UKF and PF approximates PDF by limited samples, and is very effective in dealing with parameter estimation and system filtering problems for nonlinear non-Gaussian time-varying system.

In recent years, researches on nonlinear Bayesian filtering have been carried out in the direction of engineering applications, so the accuracy, real time performance and stability of nonlinear Bayesian filtering is paid more and more attention. There are unsolved problems for nowadays nonlinear Bayesian filtering as follows.

1） Filtering accuracy can be improved. In recent years, to improve the filtering accuracy of nonlinear system, more and more combined filters appear, such as unscented particle filter, extended Kalman particle filter, and it is a hopeful direction worthy of exploring.

2） Real time performance can be strengthened. Nowadays, calculation is very complicated for most nonlinear Bayesian filtering. Take particle filter for example, we have to maintain substantial particles against particle degeneration and sample impoverishment, which results in calculation increase. Therefore, partly linearized particle filter is researched. It is an urgent problem to reduce calculation and strengthen real time performance of nonlinear Bayesian filtering on the premise of satisfactory estimation accuracy.

3） Stability can be strengthened. Self-adaptive ability to model error is not satisfactory enough for current nonlinear Bayesian filters, which mainly reflects in highly sensitive to system noise, measurement noise and statistic characteristics of initial values, low robustness for system state mutation and lack of the ability to restrain divergence when dealing with non-stationary systems. Self-adaptive approach is one of the effective methods in strengthening self-adaptive ability, and many professors have fruitfully progressed in self-adaptive filtering at home and abroad, but researches on nonlinear Bayesian filtering is poor now.

Modern filtering techniques are developed based on modern computing technique, and will advance as modern computing technique develops. We believe another spring will not be long for nonlinear Bayesian filtering as the rapid development of computer science and related areas.

References

[1] Arulampalam, M.S.,et al. A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking [C]. *IEEE Transactions on Signal Processing*, 2002, 50(2):174-188.

[2] Z. Chen. Bayesian Filtering From Kalman Filters to Particle Filters, and Beyond. Hamilton: McMaser University, 2003.

[3] Arulampalam, M.S.,Maskell, S.,Gordon, N.,Clapp, T.. A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking [C]. *IEEE Transactions on Signal Processing*, 2002, 50(2):174-188.

[4] X. Rong Li, Vesselin P. Jilkov. A Survey of Maneuvering Target Tracking: Approximation Techniques for Nonlinear Filtering [C]. *Proceedings of 2004 SPIE Conference on Signal and Data Processing of Small Targets*. San Diego, CA, USA. April 2004, 537-550.