Study on Importance Function for Particle Filter

Liu Lu[1], Meng Yang[2], Shu Geng[1], Shu-fen Wang[1] and Yong-hui Wang[1]

[1]Harbin Institute of Petroleum,150027,Harbin, China
[2]Haerbin Engineering University, 150001,Haerbin, China
{Liu Lu, Meng Yang, Shu Geng , Shu-fen Wang, Yong-hui Wang,
luliu_hsy }@163.com

Abstract

As an important nonlinear filter theory, particle filter is a heated issue in domestic and foreign researches. The option of importance density is one of the key steps of particle filter algorithm. The proper option of importance density can minish the negative influence of filter algorithm caused by degeneracy problem. This paper introduces several widely-used options of importance density systemically, and analyzes their features and applied perspectives respectively. The paper also advances a comprehensive method of importance density, analyzes its technical features, explores the adjudgement and improvement of this method based on various performance, and finally puts forward the necessary further study according to the engineer requirements.

Keywords: *particle filter, importance function, SIS, proposal distribution*

Particle filter is a widely used nonlinear filter algorithm recently. The main idea of particle filter is describing the posterior probability density of the random variable using weighted random sampling points. These sampling points are called particles. The major problem of the particle filter is the particle degeneration, *i.e.,* most particles' weights become tremendously less than before after several iteration steps with only a few particles have relatively high weights. So that a lot of calculation will be wasted on these low-weighted particles [1]. Plenty of research results show that, the best way to solve this problem is choosing a proper importance density and add the resampling step into the algorithm. In order to choose a good importance density, one have to consider several factors: first, the definition domain of probability density should cover all of the posterior probability distribution, i.e. the importance function should have a wide distribution, second, it should be sampled easily, furthermore, it should consider both prior probability density of the status and the newest observation data so as to get the smallest variance and make it close to the true posterior probability density.

Particle filte has been in signal processing, automatic control, target tracking, image processing, and many fields of application. For example, using particle filter for the pretreatment and image analysis of the underwater image target recognition, by lowering the complexity of the algorithm and improving algorithm importance in sampling and resampling step, we can gain good real-time and accuracy in image processing. In practical application, there is no common way to design the importance function. The importance function is usually designed by choosing a method to meet the performance requirement based on the specific case. This paper describes several mostly used methods that are used to design the importance function. The advantages and disadvantages of these methods have been analyzed.

1. The Optimum Importance Density Function

Theoretically, after choosing the importance density function $q\left(x_k \mid x_{k-1}^i, y_k\right) = p\left(x_k \mid x_{k-1}^i, y_k\right)$,when the reference distribution equal to the actual distribution, the importance weight w_k^i is minimized. w_k^i can be iteratively calculated to $w_k^i = w_{k-1}^i \int p\left(y_k \mid x_k\right) p\left(x_k \mid x_{k-1}^i\right) dx_k$. But there are two obvious disadvantages. First, the actual distribution $p\left(x_k \mid x_{k-1}^i, y_k\right)$ usually cannot be calculated. Second, the integration of w_k^i cannot be solved [2].

2. Commonly used Methods for Importance Function Generation

2.1. Use Prior Probability Density as Importance Function

In SIS or SIR algorithm, the prior probability density is directly chosen as importance function [3].

i.e., $q\left(x_k \mid x_{k-1}^i, y_k\right) = p\left(x_k \mid x_{k-1}^i\right)$.

The method has simple calculation and is easily implemented. So the filter works well when the observation accuracy is not high. The disadvantage is the lack of consideration for the correction of importance weights by the newest observation values y_k , results in the consequent samples have a big difference compared to the samples generated by the real posterior distribution density. So this is somehow aimless, especially when the likelihood function is in the peak position and the forecasting status is close to the likelihood function. This often happens in the situation of high accuracy observation. Figure 1 shows the basic steps for the SIR algorithm.
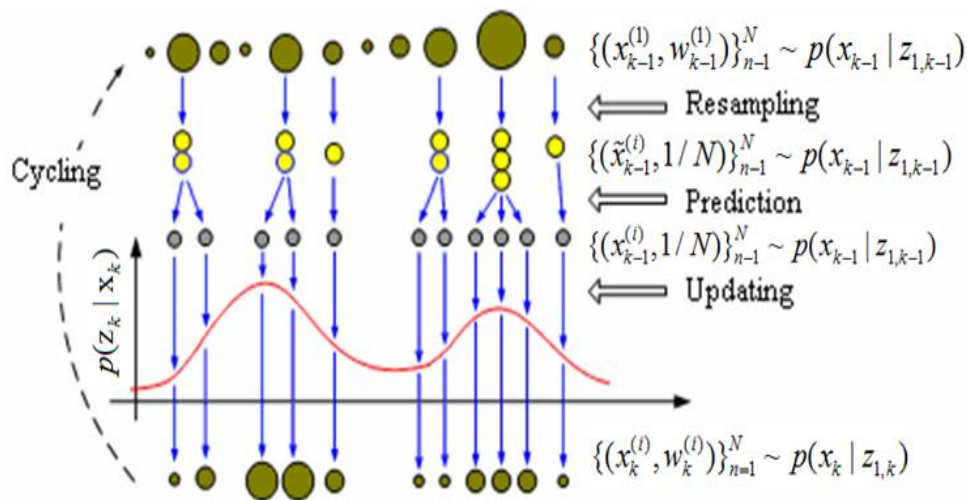

Figure 1. The Basic Steps for the SIR Algorithm

2.2. Mixed Importance Function Constituted of both Prior and Posterior Probability Density

With the consideration of the advantage and disadvantage of both prior and posterior probability density function as importance function, if the status parameters can be decomposed into two independent part, *i.e.,* $x_t = \{x_1, x_2\}$, while the samples that follow the distribution of $p\left(x_{2,t} \middle| x_{2,t-1}^i\right)$ and $p\left(x_{1,t} \middle| x_{2,t}^i, x_{t-1}^i, y_t\right)$ can be generated conveniently, the importance function that has the following form can be adopted:

$$q\left(x_t \middle| x_{t-1}^i, y_t\right) = p\left(x_{1,t} \middle| x_{2,t}^i, x_{t-1}^i, y_t\right) p\left(x_{2,t} \middle| x_{2,t-1}^i\right) \ ,$$ in which $x_{2,t}^i$ is the sample comes from $p\left(x_{2,t} \middle| x_{2,t-1}^i\right)$. Obviously, the above equation is a mixed equation constituted of prior and posterior probability density. The importance weights can be deduced to the follow equation based on the above equation:

$$w_k^i = w_{k-1}^i p\left(y_t \middle| x_{2,t}^i, x_{t-1}^i, y_{t-1}\right)$$

Compared to the posterior probability density as importance function, the mixed importance function has the advantage of simple calculation and easily renew of the weights. Moreover, because the mixed importance function uses the newest observation, the importance weights have smaller variance compared to that of prior probability density.

2.3. The Annealingprior Probability Density as the Importance Function

The main reason that SIS algorithms use the prior probability density as the importance function is that it can be achieved easily. However, without the consideration of the impact of the status noise and observation noise on the performance of particle filter, if the sample that extracted from prior distribution density cannot cover the whole likelihood area, the performance of particle filter will be lower due to the small contribution of most particles. So, people suggested the introduction of annealing factor β [4], in order to make the annealing prior probability density as importance function to overcome the above issue.

Set β as a relation factor defined by status noise covariance matrix and observation noise covariance matrix.

Assume $q\left(x_k \middle| x_{k-1}^i, y_k\right) = p\left(x_k \middle| x_{k-1}^i\right)^\beta$, then

$$w_k^i = w_{k-1}^i \frac{p\left(y_k \mid x_k\right) p\left(x_k \mid x_{k-1}\right)}{q\left(x_k \mid x_{k-1}, y_k\right)}$$

$$= w_{k-1}^i \frac{p\left(y_k \mid x_k\right) p\left(x_k \mid x_{k-1}\right)}{p\left(x_k \middle| x_{k-1}^i\right)^\beta}$$

$$= w_{k-1}^i p\left(y_k \mid x_k\right) p\left(x_k \mid x_{k-1}\right)^\alpha,$$ in which, $\alpha = 1 - \beta$ and $0 \le \alpha \le 1$.

When $\beta = 1$, w_k^i can be simplified to a standard sequential importance resampling filter (SIR); when $\beta = 0$, the uniform distribution is used as the reference distribution. The choice of annealing factor β depends on the relationship between status noise statistical property Σ_d and observation noise statistical property Σ_v: when $\Sigma_d < \Sigma_v$, the major part of the prior probability distribution become out of the flat area of the likelihood function (see Figure 2). In this case, $0 < \beta < 1$ will make the prior probability function more flat, which equals to that

artificially add some noise to make the distribution of sampling much wider. When $\Sigma_d \approx \Sigma_v$, most part of the prior probability function overlaps with the likelihood function (see Figure 2B). In this situation, the prior probability function works well, so we can make $\beta = 1$. When $\Sigma_d > \Sigma_v$, the prior probability function is more flat compared to the peak of likelihood function (see Figure 2C). In this situation, only change β is not useful.
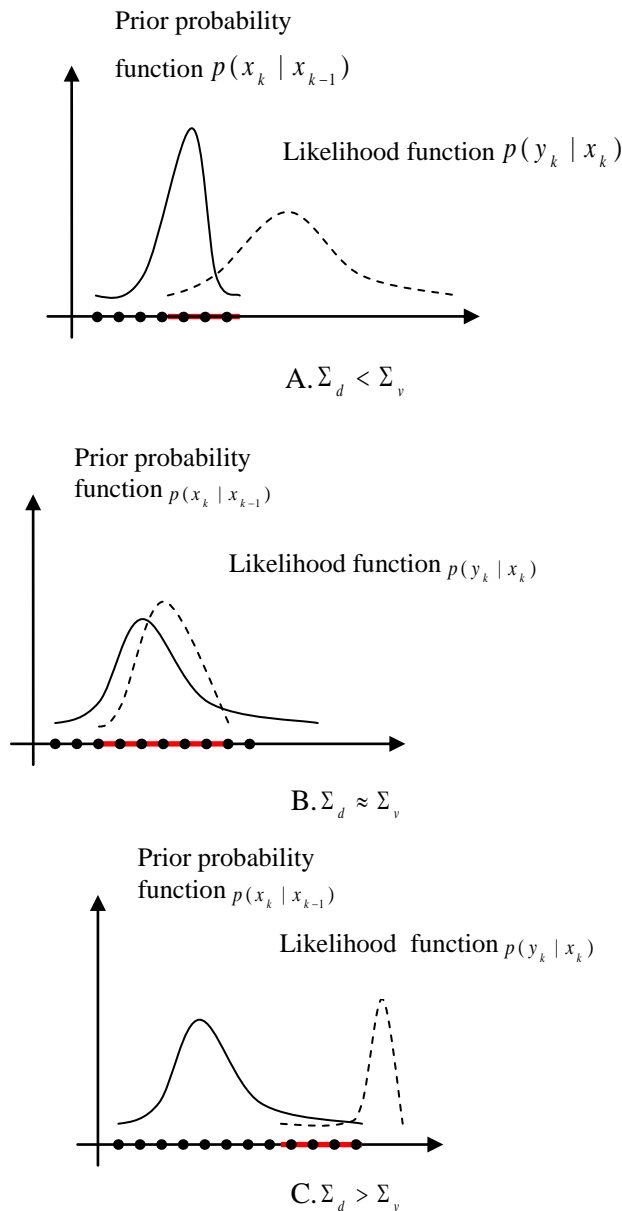

Figure 2. The Influence of Relationship between Status Noise Statistic Property and Observation Noise Statistic Property on the Performance of Particle Filter

3. Use Filter Algorithm to Design the Importance Function

3.1. Use EKF to Design the Importance Function

If the optimized reference distribution is linear, it can be approximated by EKF [5], *i.e.,* use extended Kalman filter to make a Gaussian importance density for each particle.

For each particle, make the importance density as below:

$$q\left(x_k \middle| x_{k-1}, y_k\right) \approx N\left(\overline{x}_k^i, p_k^i\right)$$

The mean value $\overline{x}_k^i$ and variance p_k^i can be obtained by EKF calculation. Then, sample from Gaussian importance density:

$$x_k^i \sim N\left(\overline{x}_k^i, p_k^i\right)$$

In which, $N\left(\overline{x}_k^i, p_k^i\right)$ is the Gaussian noise which has a mean value $\overline{x}_k'$ and variance p_k^i.

So, we can use the EKF algorithm and new observation data, calculate the mean value and variance of Gaussian distribution of every particle, and then obtain new particle by sampling in this distribution. The introduction of EKF can get relatively good importance function and make the prior probability distribution move toward high likelihood area. However, it is based on the linear approximation of the nonlinear system. For weakly nonlinear system, the EKUF can certainly overcome the problem that exists when using classical particle filter. But for many other cases, linear approximate will lead to big model error, which will cause the filter performance of EKF going lower quickly or even cause the filtering diverging.

3.2. Design the Importance Function through UKF

EKF is a partial linear method, so it is suboptimum for estimation of mean value and variance of importance distribution. Similar to EKF, Unscented Kalman Filter (UKF) can also be used to approximate the proposal distribution of particle filter [6]. UKF directly use the nonlinear system model and observation model, via several determined Sigma point to get the statistic property after nonlinear transformation. It can make posterior probability distribution's mean value and variance be exacted to second order or even higher. So UKF is better than EKF for its consideration of the newest observed importance density. This is so-called Unscented Particle Filter (UPF).

3.3. Design the Importance Function through Gaussian-Hermite Filter (GHF)

GHF is a recursive Bayes filter based on Gaussian-Hermitenumerical integration. The main ideas that use GHF to design the importance function are using each particle to forecast the next particle. Because GHF makes the newest observation value into the transfer process of system [7], it can get the sample closer to actual posterior probability function compared to EKPF. Due to the higher estimation accuracy of GHF compared to EKF, and the even higher tail of the status variance, GHPF has a better filter accuracy than EKPF.

4. Mixing of Prior Probability Density and UKF to Generate Importance Function

UPF first use the last particle and its variance to determine a set of sigma point. The position and weights of this set of point can only be determined by the expectation and

variance of particle. It can get the exactly character of particle probability distribution, then substitute it into status equation to get a new set of point. Use the weighted sum of this set of point as expectation, and use the weighted sum of the variance as variance, then use measurement equation to correct the obtained expectation and variance. The corrected value is used as the expectation and variance of the Gaussian distribution to generate a particle of currently point [8]. Because the sufficiently consideration of the influence of the current observation value on the posterior probability function, this algorithm improves the efficiency of particle. However, the calculation cost of each particle's generation is high. A report [9] gave the improvement of the UPF algorithm: use UKF that based on the estimation of the former status to get the importance function and generate a portion of particles. The remaining particles can be generated by prior probability. In this way, the influence of both the currently observation value and the prior probability on the posterior probability is under consideration. Also, the calculation cost is reduced while keeping the accuracy of filtering.

The improved UPF algorithm is shown below:

(1) Initialization $k = 0$

Get the sampling points $x_0^i \sim p(x_0)$, set $w_0^i = 1/N$, $i = 1,..., N$

(2) Importance sampling

For each sampling point x_{k-1}^i, using UKF to get the mean value $\bar{x}_k^i$ and variance P_k^i of the set of particles.

(3) Generation of particles

Generate 0.5N particles from the result of UKF $N(\bar{x}_k^i, P_k^i)$.

Generate 0.5N particles from the prior probability distribution $p(x_k | x_{k-1}^i)$.

(4) Update of weight value

$$w_k^i = w_{k-1}^i \frac{p(y_k | x_k^i) p(x_k^i | x_{k-1}^i)}{\pi(x_k^i | x_{k-1}^i, y_{1:k})}$$

The importance probability density function is:

$\pi(x_k^i | x_{k-1}^i, y_{1:k}) = N(\bar{x}_k^i, P_k^i)$. It introduces the newest observation value, so the performance of the filter is improved.

(5) Get the normalized weight value $\tilde{w}_k^i$

$$\tilde{w}_k^i = \frac{w_k^i}{\sum_{j=1}^{N} w_k^j}$$

(6) Resampling

Define N_{eff} for evaluation of the number of effective particle

If $N_{eff} < N_{thr}$, resample for $\{\bar{x}_k^i, \quad i = 1,..., N\}$, generate new set $\{x_k^i, \quad i = 1,..., N\}$,

readjust the weights of particle as: $w_k^i = 1/N$.

(7) Update the status

$$x(k | k) = \sum_{i=1}^{N} x_k^i w_k^i$$

The number ratio of the particles that generated by UKF to the particles that generated by prior probability density is not always fixed. A parameter c ($0 < c < 1$) can be introduced to control the ratio. The evaluation of c can be based on the accuracy requirement and the speed requirement of the filtering. For smaller c, the calculation time is shorter while the filtering accuracy is lower; For bigger number c, the calculation time is longer and the filtering accuracy is higher.

Table 1. The Calculation Time with Different c

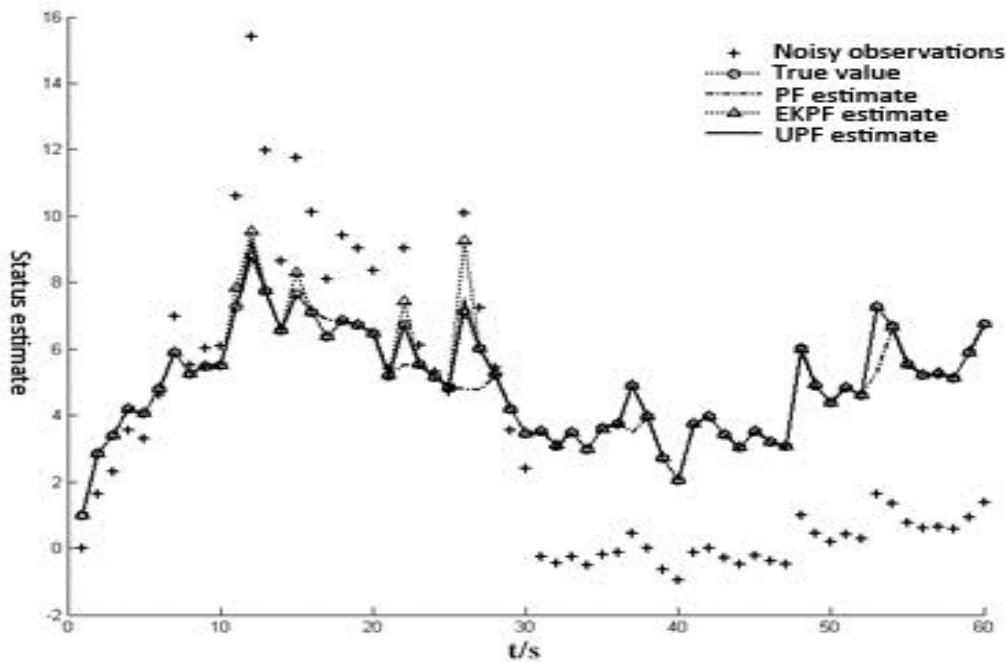| c	0.1	0.2	0.3	0.4	0.5
Calculation time(s)	0.6832	0.8455	0.9559	1.1522	1.3117
c	0.6	0.7	0.8	0.9	1
Calculation time(s)	1.4237	1.6365	1.8387	1.9771	2.1644

Figure 3. The Status Estimation for Different Particle Filtering Algorithm (in which, c is 0.35 in UPF Algorithm）

Performance analysis and comparison of several particle algorithm has been done using simulation. Example simulation uses the nonlinear non-Gaussian model. The status equations and observation equations are shown below:

$$\begin{cases} x_k = 1 + \sin((4e-2)\pi(k-1)) + 0.5 x_{k-1} + v_{k-1} \\ y_k = \begin{cases} 0.2 x_k^2 + n_k & k \le 30 \\ 0.5 x_k - 2 + n_k & k > 30 \end{cases} \end{cases}$$

During the simulation, the algorithms used are PF, EKPF and the UPF which is suggested in this report. The number of particles is set to 200 for each algorithm. The observation time is T=60. The simulation conducts 100 times independent experiments. From the simulation results, the filter accuracy that suggested in this report is obviously better than classical PF algorithm and EKPF algorithm.

Also, some scholars put forward the method of creating the importance function with the combination of the prior distribution and the EKF, and of EKF and UKF, with the purpose of getting the balance between the precision and the rapidity from improvement. The main purpose is to get the balance between accuracy and the speed. Furthermore, people also suggested that renew twice of a portion of particles using UKF and EKF, then mixed with the particles generated by the prior distribution. The procedure is more or less the same as that reported in this paper, so we won't describe it here.

5. Conclusion

The option of importance function is the vital factor for the particle filter algorithm. The algorithm existed have both advantages and disadvantages: SIR algorithm has a simple calculation procedure, but the directly usage of prior probability function as importance function will cause the particle degeneration problem. So it cannot get high accuracy of filtering; mixing the prior probability density and the posterior probability density to generate importance function and introducing the degeneration factor can somehow cover the shortage of performance of SIR filter. However, the requirement of status condition make the application of these two algorithm much restricted; EKPF has better performance compared to SIR, but with too much error introduced in linearity and Gaussian assumption, the improvement of filter's performance is very limited; UPF algorithm has the best usage of the newest observation value, so it has a big advantage compared to SIR. But with the introduction of UT transform, it increases the calculation cost and affects the speed of the filter; Mixing the UPF and SIR can have the consideration of both calculation efficiency and accuracy of the filter. It needs more verification in the solution of high-order nonlinear issues.

As a conclusion, in the design of filter algorithm, the first thing is analyzing the details of problem to be solved, while systematically considering the statistical property of noise and observing the influence of the relationship between the statistical properties on the particle filter's performance, making the best usage of prior distribution function, likelihood function and the newest observation; second, the relationship between the filter's performance and the calculation cost, complexity of calculation can also affect the filter's performance. Only if the above factors are considered, a good importance function can be designed when using proper method.

Acknowledgements

Reference

[1] J. Yao, "Study on Particle Filter Based Visual Tracking Method", Ph.D Dissertation of CAS, **(2004)**.
[2] S. Hu and Z. Jing, "Overview of particle filter algorithm", Control and Decision, vol. 20, No. 4, **(2005)**, pp. 361-365.
[3] Doucet, "Rao-Black wellised particle filtering for dynamic Bayesuan networks", Proc UAI2000, Cambridge U K, **(2000)**, pp. 253-259.

[4] Z. Chen, "Robust particle filters with applications in tracking and communications", Halnilton, Ontario, Cananda, Adaptive Systems Lab, **(2003)**.

[5] R. Chen, "Mixture Kalmanfilters", J Roy Statist Soc, vol. 62, **(2000)**, pp. 493-508.

[6] R. V. Merwe, A. Doucet and N. D. Freitas, "The Unscented Particle Filter", Technical Report CUED/F-INPENG/TR 380, Cambridge University Engineering Department, **(2000)**.

[7] Z. Yuan, N. Zheng and X. Jia, "The Gauss-Hermite Particle Filter", ACTA ELECTRONICA SINICA, vol. 31, no. 7, **(2000)**, pp. 970-973.

[8] Q. Pan, F. Yang, L. Ye, Y. Liang and Y. Cheng, "Survey of a kind of nonlinear filters-UKF", Control and Decision, vol. 20, no. 5, **(2005)**, pp. 481-489.

[9] S. J. Julier and J. K. Uhlmann, "Reduced sigma point filters for the propagation of means and covariances though nonlinear transformations", Proceeding of the American Control Conference. Anchorage AK, **(2002).**

Author

Liu Lu, Worked in Harbin Institute of Petroleum. Job title: Lecturer. Research Direction: Computer Technology and Application. As one of the Heilongjiang Provincial Department of Education Science and Technology Research Project.. In two studies of the Heilongjiang Provincial Department of Education Science and Technology Research Projects.