A Survey of Sample Size Adaptation Techniques for Particle Filters

Ondřej Straka and Miroslav Šimandl

Research Center Data, Algorithms, and Decision Making,
Department of Cybernetics,
University of West Bohemia, Pilsen,
Czech Republic

SYSID 2009, July 6-8

Outline

System specification

Dynamic stochastic nonlinear non-Gaussian system:

The system is specified by the state and measurement equations

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{f}_k(\mathbf{x}_k) + \mathbf{e}_k, \quad k = 0, 1, 2, \ldots \\ \mathbf{z}_k &= \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k, \quad k = 0, 1, 2, \ldots,\end{aligned}$$

and probability density functions $p(\mathbf{e}_k)$, $p(\mathbf{v}_k)$ and $p(\mathbf{x}_0)$,

- The sequences $\{\mathbf{e}_k\}$ and $\{\mathbf{v}_k\}$ are white
- $\mathbf{e}_k$, $\mathbf{v}_k$ and $\mathbf{x}_0$ are mutually independent
- $\mathbf{f}_k(\cdot)$ and $\mathbf{h}_k(\cdot)$ are known

Alternative specification of the system:

- by the transition probability density function (pdf)

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}),$$

  and the measurement pdf

$$p(\mathbf{z}_k|\mathbf{x}_k).$$

- used in particle filtering

State estimation

The aim of state estimation is to find the conditional pdf $p(\mathbf{x}_k|\mathbf{z}^l)$, where $\mathbf{z}^l \triangleq [\mathbf{z}_0^{\mathrm{T}}, \ldots, \mathbf{z}_l^{\mathrm{T}}]^{\mathrm{T}}$.

Particle Filtering - basic idea

- Particle filtering (PF) is currently a rapidly evolving method used for state estimation of discrete-time nonlinear non-Gaussian systems.
- The idea of PF is to approximate the filtering pdf by the empirical filtering pdf r_N, which is given by random samples $\{\mathbf{x}_k^{(i)}\}_{i=1}^N$ of the state and associated weights $\{w_k^{(i)}\}_{i=1}^N$

$$r_N(\mathbf{x}_k|\mathbf{z}^k) = \sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{x}_k - \mathbf{x}_k^{(i)}).$$

Particle filtering - algorithm

- The general PF algorithm can be decomposed into two principal steps
  1. drawing samples from an importance function specified by the user
  $$\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)},\mathbf{z}_k)$$
  2. computing the corresponding weights utilizing the transition and measurement pdf's as
  $$w_k^{(i)} = \frac{p(\mathbf{z}_k|\mathbf{x}_k^{(i)})p(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)})}{\pi(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)},\mathbf{z}_k)} w_{k-1}^{(i)}$$
- Obviously, the choice of the sample size N and the importance function π significantly affect quality of the state estimate.

Sample Size Specification

Only a few papers address sample size

- Non-adaptive sample size specification
  - constant sample size, i.e. $N_k = N$
  - calculating N off-line according to a criterion evaluating estimate quality
  - no increase of computational costs of the actual PF algorithm
- Adaptive techniques for sample size specification (SSA)
  - sample size N computed on-line
  - always increase computational costs

Motivation for Sample Size Adaptation

- Usually a fixed sample size is considered $\Longrightarrow$ quality varies due to stochastic nature of PF
- Motivation 1: To reduce computational costs which may be too high at some time instants
- Motivation 2: To guarantee a certain *quality*
- Q: Quality of what? - A: Samples or the estimate.
- Criteria:
  - empirical (usually consider samples quality)
  - evaluating *point estimate quality*
  - evaluating *pdf estimate quality*

Kullback-Leibler Distance (KLD) sampling

Fox D. 2003

Adapting the Sample Size in Particle Filters through KLD-sampling

International Journal of Robotics Research

IDEA: To bound the error between the true pdf p and the empirical pdf r_N by ε with probability $1-\delta$.

- the error is measured by the Kullback-Leibler distance
- the true pdf p is supposed to be *discrete, piecewise constant*
- $N_k = \frac{1}{2\varepsilon}\chi^2_{m-1,1-\delta}$
- m is the number of bins of the true pdf p with support
- *adaptation with respect to the complexity of the target pdf p*

Self adaptive particle filter

Soto A. 2005

Self adaptive particle filter

International Joint Conference on Artificial Intelligence Systems

- enhancement of KLD sampling

IDEA: To respect the fact that the samples are drawn from the importance function π (different from the target pdf p).

- the relative accuracy measured in terms of point estimates (MSE)

$$\frac{\mathrm{var}_p(\mathbf{x})}{N_k} = \frac{\sigma_{IS}^2}{N_{IS,k}} \quad \Rrightarrow \quad N_{IS,k} = \frac{\sigma_{IS}^2}{\mathrm{var}_p(\mathbf{x})} N_k$$

- $N_{IS,k} = \frac{\sigma_{IS}^2}{\mathrm{var}_p(\mathbf{x})} \frac{1}{2\varepsilon} \chi^2_{m-1,1-\delta}$

Self adaptive particle filter - asymptotic normal approximation

Soto A. 2005

Self adaptive particle filter

International Joint Conference on Artificial Intelligence Systems

IDEA: Instead of checking the accuracy of the *true pdf estimate*, monitor the accuracy of the particle filter in *estimation of a moment of the true pdf p*

- using strong law of large numbers – estimate of the mean is asymptotically unbiased
- if the variance of estimator is finite – the central limit theorem justifies asymptotic normal approximation for it
- $N_{IS,k} = \frac{\sigma_{IS}^2}{\mathrm{E}_p(x_k)^2} \frac{1}{\varepsilon^2} Z_{1-\alpha/2}^2$

Fixed efficient sample size adaptation

Straka O. and Šimandl M. 2006

Adaptive particle filter based on fixed efficient sample size

Proceedings of the 14th IFAC symposium on System Identification

IDEA: To preserve *Efficient sample size (ESS)* and to adapt sample size accordingly.

- *Efficient sample size (ESS)* - the number of samples drawn from the true pdf p achieving the same estimate quality as N_k samples drawn from the importance function π.
-
$$N_k = N_k^* \int \frac{[p(\mathbf{x}_k)]^2}{\pi(\mathbf{x}_k)} \mathrm{d}\mathbf{x}_k$$
- evaluating point estimate quality

Information theoretic rule adaptation

Lanz O. 2007

An information theoretic rule for sample size adaptation in particle filtering

14th International Conference on Image Analysis and Processing (ICIAP 2007)

- Consequence of the **Asymptotic equipartition property theorem :** *For a given density p and n large, the volume of the smallest n-sized sample set that contains most of the probability is approximately* $e^{nH(p)}$

IDEA: The number of i.i.d. samples needed to properly represent a density p with resolution ρ is

$$N_\rho(p) \sim \rho\, e^{H(p)}.$$

- Shannon differential entropy is approximated using kernel density estimation

Fixed Empirical Density Quality

Straka O. and Šimandl M. 2008

Adaptive particle filter with fixed empirical density quality

17th IFAC World Congress

IDEA: To guarantee the quality of the empirical pdf r_N.

- The quality is measured by inaccuracy (cross-information) between the empirical pdf r_N and the filtering pdf p
- $$N_k = z^2_{1-\delta/2}\frac{\sigma^2_W r^2_{1-\delta/2} - 2\mathrm{cov}(Y,W) r_{1-\delta/2} + \sigma^2_Y}{(\mu_W r_{1-\delta/2} - \mu_Y)^2}$$
- user-defined parameters: $r_{1-\delta/2}$, δ

Likelihood based adaptation

Koller D. and Fratkina R. 1998

Using learning for approximation in stochastic processes.

Proceedings of 15th International Conference on Machine Learning

- Dagum and Luby in “Optimal approximation algorithm for Bayesian inference” argued that actual number of effective samples is their total weight.

IDEA: To keep a fixed sum of likelihoods of the whole sample set instead of keeping a fixed sample size.

- likelihoods - unnormalized weights
- samples with low weights do not match the target pdf $\Longrightarrow$ more low weighted samples are necessary
- *empirical criterion*

Localization-basic adaptation

Straka O. and Šimandl M. 2004

Sample size adaptation for particle filters

Proceedings of the 16th IFAC symposium on Automatic Control in Aerospace

IDEA: To check position of samples according to measurement pdf to have a sufficient number of "good" samples.

- at least M_k of N_k generated samples are located in the significant part $\mathscr{A}_k$ of the measurement pdf support.
- M_k can be specified using the measurement pdf (idea of relative accuracy)
- $$N_k = \bar{N} \frac{\int_{\mathscr{A}_k} p(y_k|z_k)\mathrm{d}y_k}{\int_{\mathscr{A}_k} \pi(y_k|\mathbf{x}_{k-1}^{(1:\nu_{k-1})}, z_k)\mathrm{d}y_k}$$

$y_k = h_k(x_k)$

The overspill

Particle number controller: comparing a performance index of two PF's with different sample size

Rule-based algorithm: monitoring a set of criteria and consequently deciding to alter the sample size

Monitoring efficiency of sampling: measuring efficiency by closeness of p and π (KLD)

Performance estimate using neural networks: training a neural network by an expert to provide a decision about sample size

Computational aspects of the adaptation techniques

- Most of the relations for sample size adaptation depends on the true filtering pdf
- Naturally, the importance sampling technique is utilized for approximation of the filtering pdf

Convenient procedure:

1. Set admissible sample sizes $[N_{min}, N_{max}]$.
2. **Burn-in:** Draw N_{min} samples from the importance function and compute the corresponding weights, set N_{curr} to N_{min} and N_k to N_{max}
3. **Iterate:**
   - *while* $Ncurr < \min(N_k, N_{max})$
   - draw new N_Δ samples, compute the weights
   - set N_{curr} to $N_{curr} + N_\Delta$
   - update N_k using the new samples

Comparison - table

	provides N_k	#parameters	MC method	complexity
KLD	✓	2^*	✓	$\mathcal{O}(N_{MC} \cdot n \cdot n_B)$
KLD-I	✓	2^*	✓	$\mathcal{O}(N_{MC} \cdot n \cdot n_B)$
ANA	✓	2	✓	$\mathcal{O}(N_{MC} \cdot n)$
FESS	✓	1	✓	$\mathcal{O}(N_{MC} \cdot n)$
ITR	✓	2	✓	$\mathcal{O}(N_{MC}^2 \cdot n)$
FEDQ	✓	2	✓	$\mathcal{O}(N_{MC} \cdot n \cdot N_{k-1})$
FLA	✓	1		$\mathcal{O}(N_{k-1})$
PNC		3		$\mathcal{O}(N_{k-1} \cdot n)$
RBA		¶		$\mathcal{O}(N_{k-1} \cdot n)$
MES		2		$\mathcal{O}(N_{k-1} \cdot n)$
LBA	✓	‖	✓	$\mathcal{O}(N_{MC} \cdot n)$
PENN		**		$\mathcal{O}(N_{k-1} \cdot n)^{\dagger\dagger}$

Concluding notes and remarks

- The problem of sample size setting requires more attention
- The techniques should take into account
  - System specification (determines posterior pdf of the state)
  - Importance function (determines samples location)
- The presented design techniques – incomparable w.r.t. estimate quality (different design criteria)
- The techniques can be compared w.r.t. computational complexity, universality (tailoring to specific requirements) number of user-defined parameters, etc.

QUESTIONS ?