データを用いたフィルタ分布の直接推定と状態推定への応用

金 田 泰 昌*・入 月 康 晴*

Direct Estimation of Filter Distribution from Data and Its Application to State Estimation

Yasuaki KANEDA* and Yasuharu IRIZUKI*

We propose direct estimations of filter distributions in Kalman filter from data using density ratio estimation and KDE. Moreover, we apply the direct estimations to state estimations and propose new state estimations without information on measurement equations and noise.

Key Words: Kalman filter, machine learning, density ratio estimation, kernel density estimation

1. はじめに

状態推定は制御工学において古くから研究されている分野の一つであり、現在も精力的に研究が進められている。最も有名な状態推定手法としてカルマンフィルタ (KF)¹⁾がある。KFは線形ガウス型システムを対象としており、この場合、状態の予測分布やフィルタ分布が解析的に求まることが知られている。非線形システムに対してはKFを拡張した拡張カルマンフィルタ (EKF) やヤコビアンを必要としない無香料カルマンフィルタ (UKF)²⁾などが、また非ガウス系に対する手法として粒子フィルタ (PF)³⁾など、システムやノイズの条件に応じてさまざまな手法が提案されている。

このように、カルマンフィルタ型の状態推定手法が数多く提案されているが、これらはすべてノイズの分布情報を既知として状態の予測分布およびフィルタ分布を逐次的に解く手法と解釈できる⁴⁾。しかしながら、現実的にはノイズの事前情報が既知であることは稀であり、データを用いてノイズ分布やそれに付随するパラメータを推定する手法が提案されている。たとえば、ノイズ分布のパラメータを状態と同時に推定する手法として、文献5)ではノイズ分布をガウス分布と仮定し、そのパラメータを変分ベイズ法で推定している。しかしながら、この手法は特定の分布形状に対してそのパラメータを推定しており、分布形状は既知としている。また、ガウスとフィルタ (GSF)と呼ばれる任意の分布を扱うことのできる手法に対して、そのパラメータを適応的に決定する手法⁶⁾が提案されているが、観測ノイズは非ガウス分布を仮定しているもののシステムに加わるノイズは既知のガウス分布を仮定している。

機械学習の分野ではデータから状態に関する確率密度関数を直接推定する手法⁷⁾が提案されている。これを状態の予測分布およびフィルタ分布の推定に応用できれば、ノイズも含

めシステムの事前情報が未知の場合でもデータから直接予測分布やフィルタ分布が推定できると期待される。そこで、本稿では、システムの事前情報が未知の場合でも状態推定が可能な手法を開発することを目指し、その第一段階として観測値の事前情報が未知の場合においてデータから直接フィルタ分布を推定する手法を提案する。

2. 問題設定と戦略

2.1 システム

次式の離散時間非線形システムを考える。

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k) + \mathbf{w}_k, \quad \mathbf{y}_k = \mathbf{g}(\mathbf{x}_k) + \mathbf{v}_k. \quad (1)$$

ここで、 $\mathbf{x}_k \in \mathbb{R}^n$ は時刻 k における状態、 $\mathbf{y}_k \in \mathbb{R}^m$ は観測値、 $\mathbf{w}_k \in \mathbb{R}^n$ はシステムノイズ、 $\mathbf{v}_k \in \mathbb{R}^m$ は観測ノイズをそれぞれ表わす。また、 $\mathbf{f}$ および $\mathbf{g}$ はそれぞれ状態方程式および観測方程式である。本稿では、 $\mathbf{w}_k$ の確率密度関数および $\mathbf{f}$ は既知とし、観測値の事前情報である $\mathbf{v}_k$ の確率密度関数および $\mathbf{g}$ は未知とする。

2.2 カルマンフィルタと条件付き確率密度関数

カルマンフィルタは観測更新および時間更新式の二つのステップより構成されており、それぞれ予測分布およびフィルタ分布と呼ばれる条件付き確率密度関数を逐次的に求めるアルゴリズムである⁴⁾。今、 $\mathbf{Y}_k = \{\mathbf{y}_1, \dots, \mathbf{y}_k\}$ とすると、予測分布およびフィルタ分布はそれぞれ次式で与えられる。

$$p(\mathbf{x}_{k+1}|\mathbf{Y}_k) = \int p(\mathbf{x}_{k+1}|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{Y}_k) d\mathbf{x}_k, \quad (2)$$

$$p(\mathbf{x}_k|\mathbf{Y}_k) = \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{Y}_{k-1})}{p(\mathbf{y}_k|\mathbf{Y}_{k-1})}, \quad (3)$$

$\mathbf{w}_k$ の確率密度関数および $\mathbf{f}$ は既知のため、予測分布 (2) 式は PF などのアルゴリズムを用いて計算することができる。一方、 $\mathbf{v}_k$ の確率密度関数および $\mathbf{g}$ は未知であり、その際のフィルタ分布 (3) 式をあらかじめ取得した学習データを用いて直接推定する。そこで、本稿ではつぎの条件を仮定する。
[仮定 1] 学習時のみ状態が観測できる。

2.3 基本的な戦略

(3) 式は条件付き確率密度 $p(\mathbf{y}_k|\mathbf{x}_k)$ が求まれば、PF な

* 東京都立産業技術研究センター 東京都江東区青海 2-4-10

* Tokyo Metropolitan Industrial Technology Research Institute, 2-4-10 Aomi, Koutou-ku, Tokyo

(Received December 20, 2016)

(Revised January 31, 2017)

どのアルゴリズムを用いて計算することができる。つまり、 $\mathbf{v}_k$ の確率密度関数および $\mathbf{g}$ が未知のときに条件付き確率密度 $p(\mathbf{y}_k|\mathbf{x}_k)$ をどのように求めるかが問題となる。一方、

$$p(\mathbf{y}_k|\mathbf{x}_k) = \frac{p(\mathbf{x}_k, \mathbf{y}_k|\mathbf{Y}_{k-1})}{p(\mathbf{x}_k|\mathbf{Y}_{k-1})} \quad (4)$$

より、条件付き確率密度 $p(\mathbf{y}_k|\mathbf{x}_k)$ は密度 $p(\mathbf{x}_k, \mathbf{y}_k|\mathbf{Y}_{k-1})$ と $p(\mathbf{x}_k|\mathbf{Y}_{k-1})$ の比で表わされる。そこで、本稿では機械学習の分野で提案されている密度比推定手法ならびに密度推定手法を応用し、データから条件付き確率密度関数 $p(\mathbf{y}_k|\mathbf{x}_k)$ を直接推定する。次章以降では二種類の条件付き密度推定手法を説明し、フィルタ分布推定への応用を説明する。

3. データを用いた条件付き密度の直接推定

3.1 密度比推定を用いた条件付き密度推定⁷⁾

今、(4) 式が次式で近似されるとする。

$$\hat{p}(\mathbf{y}_k|\mathbf{x}_k) = \boldsymbol{\theta}^T \boldsymbol{\phi}(\mathbf{x}_k, \mathbf{y}_k), \quad (5)$$

ここで、 $\boldsymbol{\phi}(\mathbf{x}_k, \mathbf{y}_k) \in \mathbb{R}^N$ は N 個のカーネル関数からなる基底関数ベクトルで、 $\boldsymbol{\theta} \in \mathbb{R}^N$ はその係数ベクトルである。このとき、次式の評価関数の最小化問題を解くことで近似分布 (5) 式の係数を決定する。

$$J_0 = \frac{1}{2} \iint (\hat{p}(\mathbf{y}_k|\mathbf{x}_k) - p(\mathbf{y}_k|\mathbf{x}_k))^2 p(\mathbf{x}_k) d\mathbf{x}_k d\mathbf{y}_k. \quad (6)$$

今、(4) 式の関係を利用して (6) 式を展開する。この際、定数項を無視し、また期待値計算を標本平均で置き換えると、(6) 式から次式の評価関数が得られる。

$$J = \frac{1}{2} \boldsymbol{\theta}^T \bar{\mathbf{H}} \boldsymbol{\theta} - \boldsymbol{\theta}^T \bar{\mathbf{h}}, \quad (7)$$

ここで、

$$\begin{aligned} \bar{\mathbf{H}} &= \frac{1}{N} \sum_{i=1}^N \boldsymbol{\Phi}_i(\mathbf{x}_i), \quad \bar{\mathbf{h}} = \frac{1}{N} \sum_{i=1}^N \boldsymbol{\phi}(\mathbf{x}_i, \mathbf{y}_i), \\ \boldsymbol{\Phi}_i(\mathbf{x}_i) &= \int \boldsymbol{\phi}(\mathbf{x}_i, \mathbf{y}_i) \boldsymbol{\phi}^T(\mathbf{x}_i, \mathbf{y}_i) d\mathbf{y}_i. \end{aligned} \quad (8)$$

文献 7) では、計算の簡易さを重視し、評価関数 (7) に対して L_2 正則化付き最小二乗法を解き、得た係数 $\boldsymbol{\theta}$ のうち負になるものを 0 とすることで密度関数の非負性を確保している。また、得られた密度関数は $\int p(\mathbf{y}_k|\mathbf{x}_k) d\mathbf{y}_k = 1$ を満たす必要がある。そこで、次式のように、最終的には正規化された値を条件付き密度の推定値として採用している。

$$\hat{p}(\mathbf{y}_k|\mathbf{x}_k) = \frac{\boldsymbol{\theta}^T \boldsymbol{\phi}(\mathbf{x}_k, \mathbf{y}_k)}{\alpha(\mathbf{x}_k)}, \quad (9)$$

$$\alpha(\mathbf{x}_k) = \int \boldsymbol{\theta}^T \boldsymbol{\phi}(\mathbf{x}_k, \mathbf{y}_k) d\mathbf{y}_k. \quad (10)$$

さらに、カーネル関数としてガウスカーネルを用いることで、(8) および (10) 式が解析的に計算できることが示されている。なお、(9) 式の推定精度は正則化パラメータやバンド幅に依存するが、これらは交差検定により決定している。

3.2 KDE を用いた条件付き密度推定

3.2.1 Kernel Density Estimation (KDE)

KDE は確率密度関数を推定するためのノンパラメトリッ

ク手法の一つである。今、標本 $\mathbf{x}_i, i = 1, \dots, N$ が与えられ、かつガウスカーネルを用いた場合、KDE による確率密度関数 $p(\mathbf{x}_k|\mathbf{Y}_{k-1})$ は次式で推定される。

$$p(\mathbf{x}_k|\mathbf{Y}_{k-1}) = \frac{1}{N (\sqrt{2\pi}\sigma)^n} \sum_{i=1}^N G_\sigma(\mathbf{x}_k, \mathbf{x}_i),$$

ただし、 $G_\sigma(\mathbf{x}_k, \mathbf{x}_i) = \exp\left(-\frac{\|\mathbf{x}_k - \mathbf{x}_i\|^2}{2\sigma^2}\right)$ はバンド幅 σ のガウスカーネルである。

同時分布も同様に考えることができ、標本 $(\mathbf{x}_i, \mathbf{y}_i)$ が与えられたときの同時分布 $p(\mathbf{x}_k, \mathbf{y}_k|\mathbf{Y}_{k-1})$ は次式で推定される。

$$p(\mathbf{x}_k, \mathbf{y}_k|\mathbf{Y}_{k-1}) = \frac{\sum_{i=1}^N \phi_i(\mathbf{x}_k, \mathbf{y}_k)}{N (\sqrt{2\pi}\sigma_x)^n (\sqrt{2\pi}\sigma_y)^m}. \quad (11)$$

ただし、

$$\phi_i(\mathbf{x}_k, \mathbf{y}_k) = G_{\sigma_x}(\mathbf{x}_k, \mathbf{x}_i) G_{\sigma_y}(\mathbf{y}_k, \mathbf{y}_i). \quad (12)$$

3.2.2 KDE を用いたプロパーな条件付き密度推定

(4) 式のとおり条件付き密度は密度比で表わされる。そのため、従来 KDE を用いた条件付き密度推定では、 $p(\mathbf{x}_k, \mathbf{y}_k|\mathbf{Y}_{k-1})$ と $p(\mathbf{x}_k|\mathbf{Y}_{k-1})$ をそれぞれ KDE で推定しその比を計算していたが、推定誤差が増幅されインプロパーな推定値となることが多い。この問題点を解決したものが密度比を直接推定する手法であり、3.1 節の手法である。一方、 $p(\mathbf{x}_k|\mathbf{Y}_{k-1})$ は同時分布を周辺化することで得られる。通常この周辺化の計算は困難であるが、KDE で同時分布を推定した場合、カーネルの与え方によってはその周辺分布を解析的に計算することができる。そして、このとき得られる条件付き密度は 3.1 節の結果のサブクラスであることを以下で示す。

今、同時分布が (11) 式で与えられたとする。このとき、周辺分布 $p(\mathbf{x}_k|\mathbf{Y}_{k-1})$ は次式で与えられる。

$$p(\mathbf{x}_k|\mathbf{Y}_{k-1}) = \int p(\mathbf{x}_k, \mathbf{y}_k|\mathbf{Y}_{k-1}) d\mathbf{y}_k = \frac{\sum_{i=1}^N G_{\sigma_x}(\mathbf{x}_k, \mathbf{x}_i)}{N (\sqrt{2\pi}\sigma_x)^n}.$$

よって、条件付き確率密度関数 $p(\mathbf{y}_k|\mathbf{x}_k)$ は次式で与えられる。

$$p(\mathbf{y}_k|\mathbf{x}_k) = \frac{\sum_{i=1}^N \phi_i(\mathbf{x}_k, \mathbf{y}_k)}{(\sqrt{2\pi}\sigma_y)^m \sum_{i=1}^N G_{\sigma_x}(\mathbf{x}_k, \mathbf{x}_i)}. \quad (13)$$

一方、基底関数が (12) 式で与えられたとき、(9) 式より次式が得られる。

$$p(\mathbf{y}_k|\mathbf{x}_k) = \frac{\sum_{i=1}^N \theta_i \phi_i(\mathbf{x}_k, \mathbf{y}_k)}{(\sqrt{2\pi}\sigma_y)^m \sum_{i=1}^N \theta_i G_{\sigma_x}(\mathbf{x}_k, \mathbf{x}_i)}.$$

よって、基底関数が (12) 式で与えられたとき、KDE による条件付き確率密度関数 (13) 式は、密度比推定手法による条件付き確率密度関数 (9) 式の係数を $1/N$ にしたものと解釈することができる。つまり、カーネルの重み付き平均で近似するのか、算術平均で近似するのかの違いとなる。

(注意) 密度比推定を用いることでカーネルの重みベクトルがスパースベクトルとして求まるため、必要なカーネル数が少なくすむ可能性がある。一方、KDE を用いた場合、学習時

Procedure 1 Direct Kalman filter using
conditional density estimation

- 0) 事前準備: 条件付き密度の学習
 - a) データ $(\mathbf{x}_i, \mathbf{y}_i)$, $i = 1, \dots, N$ を取得する.
 - b) データより条件付き密度を学習し, $k = 1$ とおく.
 - c) 初期粒子群を生成する:
 $\mathbf{x}_{1|0}^{(i)} \sim p(\mathbf{x}_0)$, $i = 1, \dots, M$.
- 1) 観測更新
 - a) 観測値 $\mathbf{y}_k$, 事前分布の粒子群 $\mathbf{x}_{k|k-1}^{(i)}$, そして (9) もしくは (13) 式から (16) 式を計算する.
 - b) フィルタ分布の粒子群 $\mathbf{x}_{k|k}^{(i)}$ をリサンプリングする.
 - c) $\hat{\mathbf{x}}_{k|k} = \frac{1}{M} \sum_{i=1}^M \mathbf{x}_{k|k}^{(i)}$.
- 2) 予測更新
 - a) システムノイズの粒子群を生成する: $\mathbf{w}_k^{(i)} \sim p(\mathbf{w}_k)$.
 - b) 予測分布の粒子群を生成する:
 $\mathbf{x}_{k+1|k}^{(i)} = \mathbf{f}(\mathbf{x}_{k|k}^{(i)}, \mathbf{w}_k^{(i)})$.
- 3) $k \leftarrow k + 1$ とし, ステップ 1) に戻る.

の最適化プロセスがなく高速であるというメリットがある.
(注意) 3.1 節同様, カーネルのバンド幅は交差検定などにて決定する必要がある.

4. フィルタ分布の直接推定と状態推定への応用

フィルタ分布 (3) 式は次式で書き直すことができる.

$$p(\mathbf{x}_k | \mathbf{Y}_k) = \frac{p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{Y}_{k-1})}{\int p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{Y}_{k-1}) d\mathbf{x}_k}. \quad (14)$$

よって, 事前分布と前節で得られた条件付き密度を用いることでフィルタ分布をデータから直接推定できる.

本稿では, PF と同様に複数の粒子で事前分布を近似する. 今, 事前分布が次式で与えられたとする.

$$p(\mathbf{x}_k | \mathbf{Y}_{k-1}) = \frac{1}{M} \sum_{i=1}^M \delta(\mathbf{x}_k - \mathbf{x}_{k|k-1}^{(i)}). \quad (15)$$

ここで, $\delta(\mathbf{x})$ はディラックのデルタ関数である. このとき, 各パーティクル $\mathbf{x}_{k|k-1}^{(i)}$ に対して次式が成り立つ.

$$p(\mathbf{x}_k = \mathbf{x}_{k|k-1}^{(i)} | \mathbf{Y}_k) = \frac{\hat{p}(\mathbf{y}_k | \mathbf{x}_k = \mathbf{x}_{k|k-1}^{(i)})}{\sum_{i=1}^M \hat{p}(\mathbf{y}_k | \mathbf{x}_k = \mathbf{x}_{k|k-1}^{(i)})}. \quad (16)$$

(16) 式は離散的に与えられた状態の条件付き確率でフィルタ分布を近似しており, 次ステップの時間更新式でこの分布を用いるためにはリサンプリングが必要となる. その際, 粒子の多様性が失われフィルタ分布の近似精度が下がる可能性がある点に注意する.

Procedure 1 に提案する状態推定手法の処理の流れを示す. 提案手法は条件付き確率密度をデータセットから事前に近似し, 推定時にはそれを用いてフィルタ分布を求めるが, それ以外は既存の PF のアルゴリズムが適用できる点に注意する.

5. 数値シミュレーション

非線形非ガウス系に対する状態推定のベンチマーク問題としてしばしば取り上げられる次式の 1 次元非線形時変システ

Table 1 Root mean squared errors of estimates

PF	Proposed methods	
	density ratio estimate	KDE
3.998	3.878	3.915

ムを考える⁴⁾.

$$\mathbf{x}_{k+1} = \frac{1}{2} \mathbf{x}_k + \frac{25\mathbf{x}_k}{1+\mathbf{x}_k^2} + 8 \cos(1.2k) + \mathbf{w}_k, \quad \mathbf{y}_k = \frac{\mathbf{x}_k^2}{20} + \mathbf{v}_k.$$

ここで, $\mathbf{w}_k \sim \mathcal{N}(0, 1)$, $\mathbf{v}_k \sim \mathcal{N}(0, 1)$ であり, $\mathcal{N}(\mu, \sigma^2)$ は平均 μ , 分散 σ^2 の正規分布を表わす.

提案手法の比較として, 観測方程式ならびに観測ノイズの標準偏差が既知の場合の PF を考える. 提案手法の学習データ数は 200 個とし, 評価用データを別途 200 個用意する. また, カーネルの中心値は学習データから 100 個をランダムに選択し, 学習に必要なパラメータは平均対数尤度を基準とした 5 分割交差検定により決定する. PF および提案手法で用いる粒子数は 100 とし, 初期粒子の分布は $p(\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_{e0}, 2.0 \times 10^4)$ とする. ここで, $\mathbf{x}_{e0}$ は評価用データの初期値である.

Table 1 に推定値の二乗平均平方根誤差 (RMSE) を示す. なお, Table 1 は各手法ともに 1000 回のモンテカルロシミュレーションを行なった際の平均値を示している. これより提案手法は観測方程式ならびに観測ノイズが既知の場合の PF と同程度の性能が得られていることがわかる.

6. おわりに

本稿では, 密度比推定手法により条件付き確率密度関数をデータから直接推定し, それをフィルタ分布推定に応用することで観測値の事前情報が未知の場合の状態推定手法を提案した. また, 密度比を直接推定しない従来の手法において, ガウスカーネルを選択することで密度非推定手法による結果のサブクラスになることを示した. そして, 数値シミュレーションにより提案手法の有効性を確認した.

謝辞 本研究は JSPS 科研費 16K18128 の助成を受けたものです. ここに感謝の意を表します.

参考文献

- 1) R.E. Kalman: A New Approach to Linear Filtering and Prediction Problems, J. Basic Eng., **82**-1, 35/45 (1960)
- 2) S. Julier, J. Uhlmann and H.F. Durrant-Whyte: A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators, IEEE Trans. Automat. Contr., **45**-3, 477/482 (2000)
- 3) 樋口知之: 粒子フィルタ, 信学会誌, **88**-12, 989/994 (2005)
- 4) 片山 徹: 非線形カルマンフィルタ, 朝倉書店 (2011)
- 5) S. Sarkka and A. Nummenmaa: Recursive Noise Adaptive Kalman Filtering by Variational Bayesian Approximations, IEEE Trans. Automat. Contr., **54**-3, 596/600 (2009)
- 6) G. Terejanu, P. Singla, T. Singh and P.D. Scott: Adaptive Gaussian Sum Filter for Nonlinear Bayesian Estimation, IEEE Trans. Automat. Contr., **56**-9, 2151/2156 (2011)
- 7) M. Sugiyama, I. Takeuchi, T. Suzuki, T. Kanamori, H. Hachiya and D. Okanohara: Least-Squares Conditional Density Estimation, IEICE Trans. Inf. & Syst., **E93**-D-3, 583/594 (2010)