ベイズの定理の一拡張

Masayuki Kageyama 影山正幸 (名古屋市立大学大学院)*

Hiroe Tsubaki 椿広計 (統計数理研究所)

1 背景

MCMC の登場以来, ベイズ推定において正確な事前分布を推定することの重要性は低くなったように思われる. しかし, 学習データが十分存在しない場合, 適切な事前分布を選ぶことは事後分布を決定する際に非常に重要になる. 本稿では学習データを十分に獲得することが困難なケースを想定した数理モデルの構築を念頭に議論を進める. [7] において著者らは, 不確実な事前分布を採用した場合のベイズの定理を利用し [10], 古典的な確率論がもつ分布関数 $F(x)$ の性質を自然な形で拡張できることを証明した.

定理 1.1（[7]） 任意の fuzzy 化された事前分布 $\tilde{\nu}$ に対して,

- $\lim_{x\to\infty} \tilde{F}(x|\tilde{\nu}) = 1$, $\lim_{x\to-\infty} \tilde{F}(x|\tilde{\nu}) = 0$.
- $\tilde{F}(x|\tilde{\nu}) \succeq \tilde{F}(y|\tilde{\nu})$ if $x \geq y$.
- $\lim_{x\to y+0} \tilde{F}(x|\tilde{\nu}) = F(y|\tilde{\nu})$.

$\succeq$ は fuzzy max order.

ここでは、Schnatter [3] により提案された確率密度関数のパラメータの分布の推定方法を紹介する. 他方, 統計科学, 数理工学, 数理ファイナンスの研究分野における確率的制御システム等のテーマでの数理モデルがマルコフ決定過程により定式化されることがある. 研究自体は, 1950 年代に始まり, 様々な理論や応用研究が生み出されてきた. 1980 年代までの MDPs の研究は主として, ベルマン方程式とその解をみつけるための Policy iteration, Value iteration に関する Dynamic Programming(以後, DP) とよばれるアルゴリズムの研究であった [2, 9]. その後, マルコフ決定過程の構造, 特に推移確率の推定に関する研究が数多くなされるようになった. Bertsekas と Tsitsklis により提案されたニューロ動的計画法は推移確率が未知のモデルを解析する際の強力な道具となった [1]. また, 推移確率を fuzzy 数として捉えた研究などがある [5, 6]. 本稿では, [4] で

* kageyama@sda.nagoya-cu.ac.jp

提案された事前分布を組み込んだマルコフ決定過程を紹介する.

2 準備

任意の集合 M に対して $\mathcal{F}(M)$ を全ての fuzzy 集合 $\tilde{m}: M \to [0,1]$ の集合とする. $\tilde{m}$ の $\alpha-$cut を $\tilde{m}_\alpha := \{m \in M | \tilde{m}(m) \geq \alpha\} (\alpha \in [0,1])$ とする. $\tilde{\mathbb{R}}$ を全ての fuzzy 数の集合とする. つまり, $\tilde{r} \in \tilde{\mathbb{R}}$ とは, $\tilde{r} \in \mathcal{F}(\mathbb{R})$, normal, upper semi-continuous, fuzzy convex で compact support をもつとする. $\tilde{m} \in \tilde{\mathbb{R}}$ に対して, $\tilde{m}_\alpha = [\tilde{m}_\alpha^-, \tilde{m}_\alpha^+]$ と記す. Θ をパラメータ空間, S を state space, A を action space, Y を disturbance space, P を transition probability, $r_Y(\theta, s, a, y)$ を one-stage reward function とする.

3 ベイズの定理の拡張 [3]

Schnatter [3] により提案されたアイデアを紹介する. $\tilde{\boldsymbol{x}} = (\tilde{x}_1, \ldots, \tilde{x}_n)$ を fuzzy sample とする. このとき, fuzzy 化された尤度関数 $\tilde{L}(\boldsymbol{\theta})$ の $\alpha-$cut は次式で与えられる. $\tilde{L}_\alpha^-(\boldsymbol{\theta}) = \min_{\boldsymbol{x} \in \tilde{\boldsymbol{x}}_\alpha} L(\boldsymbol{\theta}|\boldsymbol{x}) = \prod_{i=1}^n (\tilde{f}_i(\tilde{x}_i))_\alpha^-(\boldsymbol{\theta}), \tilde{L}_\alpha^+(\boldsymbol{\theta}) = \max_{\boldsymbol{x} \in \tilde{\boldsymbol{x}}_\alpha} L(\boldsymbol{\theta}|\boldsymbol{x}) = \prod_{i=1}^n (\tilde{f}_i(\tilde{x}_i))_\alpha^+(\boldsymbol{\theta})$, ただし, $(\tilde{f}_i(\tilde{x}_i))_\alpha^-(\boldsymbol{\theta}) = \min_{x \in (\tilde{x}_i)_\alpha} f(x|\boldsymbol{\theta})$, $(\tilde{f}_i(\tilde{x}_i))_\alpha^+(\boldsymbol{\theta}) = \max_{x \in (\tilde{x}_i)_\alpha} f(x|\boldsymbol{\theta})$. また, 事前分布が fuzzy 化 $\tilde{\nu}$ の場合の事後分布は次式で与えられる. $(\tilde{h}_n)_\alpha^-(\boldsymbol{\theta}) = L(\boldsymbol{\theta}|\boldsymbol{x}) \cdot (\tilde{\nu})_\alpha^-(\boldsymbol{\theta}), (\tilde{h}_n)_\alpha^+(\boldsymbol{\theta}) = L(\boldsymbol{\theta}|\boldsymbol{x}) \cdot (\tilde{\nu})_\alpha^+(\boldsymbol{\theta}), (\tilde{\nu})_\alpha^-(\boldsymbol{\theta}) = \min_{\boldsymbol{\xi} \in \tilde{\boldsymbol{\xi}}_\alpha} \nu(\boldsymbol{\theta}|\boldsymbol{\xi}), (\tilde{\nu})_\alpha^+(\boldsymbol{\theta}) = \max_{\boldsymbol{\xi} \in \tilde{\boldsymbol{\xi}}_\alpha} \nu(\boldsymbol{\theta}|\boldsymbol{\xi})$. 両者を fuzzy 化した場合についても議論がなされているので, 興味のある方は [3] を参照されたい.

4 事前分布を考慮したマルコフ決定過程

ここで, [4] により議論されている推移確率のパラメータの事前分布 ν が与えられた場合のマルコフ決定過程について考える. 確率分布と総合利得は次式で定義される.

定義 4.1（[4]）

$$\mathcal{Q}_T^\nu(E) \triangleq \int_\Theta \nu(d\theta) \int P(\theta, dy_1) \cdots \int P(\theta, dy_T) \mathbf{1}_E(\theta, y^T) \tag{1}$$

$$R_T^\pi(\theta, s_0, y^T) \triangleq r_Y(\theta, s_0, \pi_0(s_0), y_1) + \sum_{t=1}^{T-1} \beta^t \cdot r_Y(\theta, \zeta_t^\pi, \pi_t(y^t, \zeta_t^\pi), y_{t+1}) + \beta^T \cdot V_0(\theta, \zeta_T^\pi),$$

$0 < \beta < 1$ は discount factor.

定義 4.2（[4]） 事前分布 ν を用いた $T-$stage Bayes reward を次式で定義する.

$$\begin{aligned} v_T^\pi(\nu, s) &\triangleq \int \mathcal{Q}_T^\nu(d(\theta, y^T)) R_T^\pi(\theta, s, y^T) \\ &= \int \nu(d\theta) V_T^\pi(\theta, s). \end{aligned} \tag{2}$$

我々の目的は上式を最大化する policy π^* をみつけることである．通常のベイズ学習と同様に [4] においても Bayes operator とよばれる条件付き確率分布を利用してパラメータの更新が行なわれている．

5 課題と今後の展望

通常のベイズ学習では真の確率分布を求めることだけが目標になることが多いが, 多段階決定過程ではそれと同時に最適政策を見つけなければならず著者の知る限り, DP のようなアルゴリズムはみつかっていない.
また, よりロバストなモデルとして以下のような (1) を拡張した確率分布を考えることも可能であり, 現在この測度上でのモデルの構築に取り組んでいる [8].

$$
\begin{aligned}
\widetilde{\mathcal{Q}}_T^{\nu}(E) \triangleq (F)\int \tilde{\nu}(d\theta)\int P(\theta, dy_1)\cdots \\
\cdots\int P(\theta, dy_T)\mathbf{1}_E(\theta, y^T).
\end{aligned} \tag{3}
$$

ただし, $(F)\int$ は fuzzy 積分 [10].

謝辞

本研究は科研費 (課題番号:19K01735)(代表:影山正幸) の助成を受けたものである.

参考文献

[1] Dimitri P. Bertsekas, Jhon N. Tsitsklis, Neuro-Dynamic Programming, Athena Scientific, 1996.

[2] E. A. Feinberg, A. Shwartz, Handbook of Markov decision processes:Methods and Applications, Kluwer Academic Publishers, 2002.

[3] S. F. Schnatter, ”On fuzzy Bayesian inference”, Fuzzy sets and systems, 1993, 60, 41-58.

[4] Hinderer, Karl, Rieder, Ulrich, Stieglitz, Michael, Dynamic Optimization:Deterministic and Stochastic Models, Springer, 2016.

[5] M. Kurano, M. Yasuda, J. Nakagami, Y. Yoshida, ”Fuzzy Optimality Relation for Perceptive MDPs — The average case”, Fuzzy Sets and Systems, Volume 158, Issue 17, 1 September 2007, 1905-1912.

[6] M. Kurano, M. Yasuda, J. Nakagami, Y. Yoshida, ”A fuzzy approach to Markov decision processes with uncertain transition probabilities”, Fuzzy Sets and Systems, 157, 2006, 2674-2682.

[7] M. Kageyama, T. Fujii, K. Kanefuji and H. Tsubaki ”An extension of risk measures using non-precise a-priori densities”, Journal of Uncertainty Systems, 2011, 5(4) 314-320.

[8] 影山正幸, ベイジアンマルコフ決定過程, 日本数学会秋季総合分科会, 2019 年 9 月 18 日

[9] Martin L. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming, Wiley-Interscience, 2005.

[10] Viertl, R., and D. Hareter, Generalized Bayes' theorem for non-precise a-prior distribution, Metrika, 2004, 59, 263-273.