Neuroscience Research 104 (2016) 52–55

Neuroscience Research
journal homepage: www.elsevier.com/locate/neures

Review article

Slow dynamics perspectives on the Embodied-Brain Systems Science

Shiro Yano[a,*], Takaki Maeda[b], Toshiyuki Kondo[a]

[a] Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology, 2-24-16 Nakachou, Koganei-shi, Tokyo 184-8588, Japan
[b] Department of Neuropsychiatry, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan

ARTICLE INFO

Article history:
Received 2 September 2015
Received in revised form 5 November 2015
Accepted 10 November 2015
Available online 28 November 2015

Keywords:
Dynamical system
Fast-slow systems
Bayesian statistics
Dimension reduction
Statistical learning theory
Embodied-Brain Systems Science

ABSTRACT

Recent researches point out the importance of the fast-slow cognitive process and learning process of self-body. Bayesian perspectives on the cognitive system also attract research attentions. The view of fast-slow dynamical system has long attracted wide range of attentions from physics to the neurobiology. In many research fields, there is a vast well-organized and coherent behavior in the multi degrees-of-freedom. This behavior matches the mathematical fact that fast-slow system is essentially described with a few variables. In this paper, we review the mathematical basis for understanding the fast-slow dynamical systems. Additionally, we review the basis of Bayesian statistics and provide a fast-slow perspective on the Bayesian inference.

Contents

1. Introduction

The fast-slow dynamical system is a set of interacting objects such that at least one object varies much slower than the other objects. This viewpoint has attracted wide range of research attentions including physics, chemistry, sociology and neurobiology (Haken, 2004; Scheffer et al., 2012). One of the reasons for this is the mathematical fact that such system is essentially described only by the slow variables. Fast variables are enslaved to these slow variables. A number of experiences that the multi degrees-of-freedom (DOF) systems often show the DOF reduction encouraged the researchers to model them with a fast-slow system.

In the cognitive sciences, Bayesian perspectives on the cognitive systems attract lots of research attentions (Griffiths et al., 2008). Additionally, researchers point out the existence and importance of the fast-slow cognitive process about the self-body (Hagura and Haggard, 2015).

In this paper, we review the mathematical basis and related results of the fast-slow dynamical systems. Moreover we review the basis of Bayesian statistics and propose a fast-slow perspective on the Bayesian statistics. These review and perspectives would be helpful for advancing the Bayesian perspectives in cognitive science.

* Corresponding author. Tel.: +81 09047846062.
E-mail addresses: syano@cc.tuat.ac.jp (S. Yano), takaki@xa3.so-net.ne.jp (T. Maeda), t_kondo@cc.tuat.ac.jp (T. Kondo).

http://dx.doi.org/10.1016/j.neures.2015.11.002

2. Dynamical systems

2.1. Fast-slow dynamical system

We consider the following ordinary differential equation (ODE)

Definition 1 *(Fast-Slow ODE).*

$$\frac{dx}{dt} = f(x, y, \varepsilon) \tag{1}$$

$$\frac{dy}{dt} = \varepsilon g(x, y, \varepsilon), \tag{2}$$

with $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, $\varepsilon \in \mathbb{R}$, f: $\mathbb{R}^{n+m} \to \mathbb{R}^n$ and g: $\mathbb{R}^{n+m} \to \mathbb{R}^m$. A parameter ε is subjected to $0 \le \varepsilon \ll 1$.

The variable y is called slow variable due to $dy/dt \simeq 0$. Contrastingly the variable x is called fast variable (Jones, 1995).

We assume the functions f, g are the C^∞-differentiable functions and that f is hyperbolic at the equilibrium of Eq. (1). In other words, all of the eigenvalues of the Jacobian $\partial f/\partial x(x^*, y^*, \varepsilon)$ at any points $(x^*, y^*) \in \{(x, y)|f(x, y, \varepsilon) = 0\}$ have non-zero real parts. This assumption is important for applying the implicit function theorem on $f(x, y, \varepsilon) = 0$. As an especially important case $\varepsilon = 0$, we put down the function $x = h_0(y)$ as a solution of $0 = f(x, y, 0)$.

By changing the timescales t to $\tau = \varepsilon t$, Eqs. (1) and (2) become

$$\varepsilon \frac{dx}{d\tau} = f(x, y, \varepsilon) \tag{3}$$

$$\frac{dy}{d\tau} = g(x, y, \varepsilon). \tag{4}$$

The time scale τ is a slower unit of measurement than t. It is for this reason that system (1) and (2) is called the fast system and system (3) and (4) is called the slow system.

2.2. Dimensionality reduction and Synergetics

Roughly speaking, in the limit of $\varepsilon \to 0$, there exists a $\Delta > 0$ such that the trajectory of (3) and (4) starts from $(x_0, y_0) \in \mathbb{R}^{n+m}$, gets closer to the trajectory of Eqs. (5) and (6) during $-\Delta < \tau < \Delta$.

$$x = h_0(y) \tag{5}$$

$$\frac{dy}{d\tau} = g(h_0(y), y, 0) \tag{6}$$

It means that the variable x is enslaved to satisfy $x = h_0(y)$. For a following precise explanation about this reduction, we denote a manifold $\mathcal{M}_0 = \{(x, y)|x = h_0(y), y \in K\}$, where K is a compact domain in $\mathbb{R}^m$.

Before explaining the theorem about the above mentioned dimensionality reduction, we introduce a term:

Definition 2 *(Locally invariant manifold (Chow et al., 2000)).* A submanifold $\mathcal{M} \subset \mathbb{R}^{n+m}$ with boundary $\partial\mathcal{M}$ is called locally invariant under (1) and (2), if, for any point $p \in \mathcal{M}/\partial\mathcal{M}$, there exists a $\Delta > 0$ such that $(x, y)_{t,p} \in \mathcal{M}$ for $t \in (-\Delta, \Delta)$, where $(x, y)_{t,p}$ is the solution of (1) and (2) with $(x, y)_{0,p} = p$.

Following theorem holds under a few appropriate assumptions (Jones, 1995)

Theorem 3 (Fenichel's theorem). *If $\varepsilon > 0$ is sufficiently small, there exists the locally invariant manifold under Eqs. (1) and (2) that $\mathcal{M}_\varepsilon = \{(x, y)|x = h_\varepsilon(y), y \in K\}$. Moreover h_ε is C^r for any $r < +\infty$ jointly in y and ε. $\mathcal{M}_\varepsilon$ is diffeomorphic to $\mathcal{M}_0$.*

Manifold M_ε is called slow manifold. Fenichel's theorem is known as the generalization of Tikhonov–Levinson theory (O'Malley, 2014). Tikhonov–Levinson theory assumes the stability of Eq. (1). Fenichel generalized this theory to be applicable for hyperbolic f at the equilibrium. Further historical review and extensions are reviewed in O'Malley (2014). Thus we get the reduced system Eq. (6).

We rewrite Eq. (6) to Eq. (7) without loss of generality.

$$\frac{dy}{d\tau} = g_0(y, \eta). \tag{7}$$

The vector field is parametrized by η. As noted before, the variable x is enslaved to this dynamics of the slow variable y.

Hermann Haken has investigated the mechanisms of the spontaneous emergence of new quantities and structures in the large degree of freedom system (Haken, 2004). He named this research field Synergetics. The fast-slow system is enslaved to the reduced system (7). Moreover once bifurcation occurs in this reduced system, the whole system spontaneously changes. A bifurcation of a dynamical system is a qualitative change on the system which is caused by parameters such as η (Crawford, 1991). A review of bifurcation theory is outside the scope of this paper. Readers are recommended to refer Crawford (1991) and Kuznetsov (2004). For this reason, the fast-slow phenomena have been one of the research subjects in Synergetics.

3. Bayesian statistics

3.1. Basis of Bayesian inference

At first, we introduce notations in this section. We represent a set of observed n samples as $X^n = (X_1, X_2, \ldots, X_n)$ which are independently taken from the true distribution $q(x)$, $x \in \mathbb{R}^n$. In general, true distribution $q(x)$ is unknown. Bayesian inference is a kind of the statistical inference which aims to construct a model of $q(x)$. It is based on the Bayes' theorem:

$$\varphi(w|X^n) = \frac{\varphi_0(w)\prod_{i=1}^{n} p(X_i|w)}{\int \varphi_0(w)\prod_{i=1}^{n} p(X_i|w)dw} \tag{8}$$

which consists of a conditional probability distribution $p(x|w)$, given a parameter $w \in \mathbb{R}^d$, prior distribution $\varphi_0(w)$ and samples X^n. The Bayes' theorem (8) is recursively derived from another form of the Bayes' theorem:

$$\varphi(w|X^n) = \frac{p(X_n|w)\varphi(w|X^{n-1})}{\int p(X_n|w)\varphi(w|X^{n-1})dw}, \tag{9}$$

where $\varphi_0(w) = \varphi(w|X^0)$. Bayesian inference is the updating process of prior distribution to the posterior distribution based on the Bayes' theorem. The denominator $Z = \int \varphi_0(w)\prod_{i=1}^{n} p(X_i|w)dw$ is called the marginal likelihood and the negative logarithm $-\ln Z$ is called the Bayes free energy (Watanabe, 2001a).

We are interested in the asymptotic agreement between true distribution $q(x)$ and predictive distribution $p(x|X^n)$ in the limit of $n \to \infty$, where predictive distribution is defined as

$$p(x|X^n) = \int p(x|w)\varphi(w|X^n)dw. \tag{10}$$

Watanabe (Watanabe, 2001b) showed that the generalization error $G(n) = E_{X^n}[d(q(\cdot), p(\cdot|X^n))]$ behaves $G(n) \to 0$ with $n \to \infty$ when $d(q(\cdot), p(\cdot|X^n))$ is Kull-back Leibler divergence (KL divergence):

$$d(q(\cdot), p(\cdot|X^n)) = \int q(x)\ln\frac{q(x)}{p(x|X^n)}dx. \tag{11}$$

KL divergence is a type of divergence function.

Definition 4 *(Divergence function (Gneiting and Raftery, 2007)).*

$$d(P, Q) = S(Q, Q) - S(P, Q) \quad P, Q \in \mathcal{P} \tag{12}$$

is called divergence function, where $\mathcal{P}$ is a convex class of probability measures on $(\Omega, \mathcal{F})$. Ω is a sample space and $\mathcal{F}$ is a σ-algebra of subset of *Omega*. $S(P, Q)$ is known as scoring rule.

It becomes KL divergence under the logarithmic scoring rule Eq. (13):

$$S(P, Q) = \int Q(x) \ln P(x) dx. \tag{13}$$

Logarithmic rule is widely used scoring rule in statistical decision problem and inference problem because of its strict propriety as it is explained in (Gneiting and Raftery, 2007). In short, the generalization error is automatically minimized by the Bayes' theorem.

Bayesian inference requires us to compute the marginalization at the denominator of Eq. (8). Because of the difficulty of marginalization, a number of approximation method have been proposed: Markov Chain Monte Carlo, Laplace approximation, variational approximation, expectation propagation, and so on (Ghahramani, 2004). Friston et al. focus on the variational approximation as a leading approximation in the biological systems (Friston et al., 2009). Litvak employs the belief revision algorithm (Litvak and Ullman, 2009). Further discussions would be required as to what kind of approximate inference algorithm drives the Bayesian brain.

In the field of cognitive neuroscience, the research viewpoint of fast-slow learning process about the body representation attracts research attentions (Hagura and Haggard, 2015). Hagura et al. define the term body representation as "organized maps of signal routing", which means the relationships between given signals (Hagura and Haggard, 2015). They advocate that the body representation would become an alternative concept to the body schema and body image, where "thought to hold bodily information required for the online control of action" is called body schema and a "relatively enduring representation of the physical structure of the body, which takes into account previous experiences and knowledge" is called body image (Kammers et al., 2009). As reviewed in this section, Bayesian inference enables us to construct the predictive distribution $p(x|X^n)$ of the true distribution $q(x)$ in the sense $G(n) \to 0$ in the limit of $n \to \infty$. They emphasize the multiple time scale of updating process of the prediction distribution.

In the next section, we present a problematic perspective to merge the slow-fast dynamical systems and Bayesian inference.

3.2. Perspectives on fast-slow inference

We consider a specific Bayesian inference problem with a statistical model $p(x|\theta, \mu)$ and priors $p(\theta|X^{n-1})$, $p(\mu|X^{n-1})$:

$$p(\theta, \mu|X^n) = \frac{p(X_n|\theta, \mu)p(\theta|X^{n-1})p(\mu|X^{n-1})}{\int p(X_n|\theta, \mu)p(\theta|X^{n-1})p(\mu|X^{n-1})d\theta d\mu}, \tag{14}$$

such that

$$0 \leq d[p(\theta|X^n), p(\theta|X^{n-1})] \ll d[p(\mu|X^n), p(\mu|X^{n-1})] < 1, \tag{15}$$

where $d[\cdot, \cdot]$ is KL divergence here. $p(\theta|X^n)$ which appears in (15) is obtained by the marginalization of $p(\theta, \mu|X^n)$ by μ. The same holds for $p(\mu|X^n)$. We call $p(\theta|X^n)$ as slow prior/posterior and $p(\mu|X^n)$ as fast prior/posterior.

Eq. (15) means that the prior distribution of μ is much more variable than that of θ. In other words, if we consider the prior distributions are parametrized, i.e. $p(\theta|X^n) = p(\theta\,; \alpha_n)$ with $\alpha_n \in \mathbb{R}^l$ and $p(\mu|X^n) = p(\mu\,; \beta_n)$ with $\beta_n \in \mathbb{R}^m$, the assumption (15) means the rate of change of α_n is much slower than that of β_n.

As a brief example, we introduce a inference problem of the mean value $m \in \mathbb{R}$ and the variance value $\sigma^2 \in \mathbb{R}$ of normal distribution $N(x|m, \sigma^2)$. Conjugate prior is normal distribution $N(m|m_0, (\kappa_0\lambda)^{-1})$ and gamma distribution $Ga(\lambda|a_0, b_0)$ where $\lambda = \sigma^{-2}$ (Murphy, 2007). With n samples, parameters of the posterior distribution become

$$m_n = \frac{\kappa_0 m_0 + n\bar{x}}{\kappa_n} \tag{16}$$

$$\kappa_n = \kappa_0 + n \tag{17}$$

$$a_n = a_0 + \frac{n}{2} \tag{18}$$

$$b_n = b_0 + \frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2 + \frac{\kappa_0 n(\bar{x} - m_0)^2}{2\kappa_n} \tag{19}$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{20}$$

If $\kappa_0 \gg n$ and $m_0 \gg \frac{n\bar{x}}{\kappa_0}$, posterior distribution of m is approximately invariant, but λ is not. This invariance is approximately explained as a limiting case $N(m|m_0, (\kappa_0\lambda)^{-1}) \to \delta(m - m_0)$ in the limit of $\kappa_0 \to \infty$ because posterior distribution of m becomes $\delta(m - m_0)$ by the Bayes' theorem.

We present two research ideas which would be applicable for any kind of system that faces the fast-slow inference problem (14) and (15). First idea is about the hastening method of updating process of slow prior. It would be of help for modulating the persistent prior. Second idea is about the dimensionality reduction of inference problem. To the best of our knowledge, there is no established theory about the dimensionality reduction method shown below. Mathematical refinement would be an issue in the future.

Let us present the first research idea. Assume the agent that faces the fast-slow inference problem such as the person who challenges the inference problem with some rigid beliefs. It makes no difference to this idea by assuming artificial agent or human beings. As described in this section, their rigid beliefs are approximately represented as delta function. These strong beliefs are remained to be unchanged. Alternatively, they flexibly change other variables to establish the inference problem. Although the existence of rigid beliefs would be beneficial under static circumstances, it becomes a big disadvantage in case the agent should change his mind drastically. From the fast-slow inference perspective, it would be effective to increase the breadth of the prior distribution of rigid beliefs for this purpose. In other words, to enforce the agent to think of a number of possibilities on the rigid beliefs would be effective to achieve that.

Let us present the second research idea which is about the dimensionality reduction method. Key idea is that slow prior behaves as the likelihood function rather than the prior distribution from a standpoint of the fast prior. By marginalizing out the slow prior $p(\theta\,; \alpha)$, we get the fast inference process

$$p(\mu|D; \alpha, \beta_{n+1}) = \frac{p(D|\mu; \alpha)p(\mu; \beta_n)}{\int p(D|\mu; \alpha)p(\mu; \beta_n)d\mu}. \tag{21}$$

In the limit of $n \to \infty$, we solve the posterior parameter $\beta_{n+1} \to \beta$ as a function of α, i.e. $\beta = \beta(\alpha)$. By marginalizing out the fast prior $p(\mu\,; \beta(\alpha_m))$, we get the slow inference process based on Bayes' theorem

$$p(\theta|D; \alpha_{m+1}) = \frac{p(D|\theta; \beta(\alpha_m))p(\theta; \alpha_m)}{\int p(D|\theta; \beta(\alpha_m))p(\theta; \alpha_m)d\theta}, \tag{22}$$

or the maximum likelihood method

$$\alpha_{m+1} = \arg\ \max_{\alpha_m \in \mathbb{R}^l} \int p(D|\theta; \beta(\alpha_m))p(\theta; \alpha_m)d\theta. \tag{23}$$

By the equation $\beta = \beta(\alpha)$, we obtain the fast priors at the same time as updating slow prior. Some dimension reduction methods and bifurcation analysis methods for the fast-slow nonlinear stochastic differential equations are proposed (Kuehn, 2011). This dimensional reduction would correspond to the reduction of the statistical model rather than that of the priors.

4. Conclusion

In this paper, we reviewed the fast-slow concept in the ordinary differential equations. We also reviewed the mathematical basis of Bayesian statistics and proposed a fast-slow perspective on it.

In the fast-slow perspective of Bayesian statistics, we introduced the concept of fast/slow prior and provided two research ideas related to these concepts. First idea was the hastening method of updating process of slow prior. It would be of help for modulating the persistent prior. Second idea was about dimensionality reduction method. A drastic change in the set of beliefs would be caused by a small number of beliefs encoded as the slow priors.

Acknowledgement

This work was partially supported by JSPS KAKENHI, Grant-in-Aid for Scientific Research on Innovative Areas "Understanding brain plasticity on body representations to promote their adaptive functions" (Grant Numbers 26120002 and 26120005).

References

Chow, S.-N., Liu, W., Yi, Y., 2000. Center manifolds for smooth invariant manifolds. Trans. Am. Math. Soc. 352 (11), 5179–5211.

Crawford, J.D., 1991. Introduction to bifurcation theory. Rev. Mod. Phys. 63 (4), 991.

Friston, K.J., Daunizeau, J., Kiebel, S.J., 2009. Reinforcement learning or active inference? PLoS ONE 4 (7), e6421, 07.

Ghahramani, Z., 2004. Unsupervised learning. In: Advanced Lectures on Machine Learning. Springer, pp. 72–112.

Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. J. Am. Stat. Assoc. 102 (477), 359–378.

Griffiths, T.L., Kemp, C., B. Tenenbaum, J., 2008. Bayesian models of cognition. In: The Cambridge Handbook of Computational Psychology. Cambridge University Press, pp. 59–100 (Chapter 3).

Hagura, N., Haggard, P., 2015. Body representation and neuroprosthetics. In: Clinical Systems Neuroscience. Springer, pp. 169–188.

Haken, H., 2004. Synergetics: Introduction and Advanced Topics. Springer Science & Business Media.

Jones, C.K., 1995. Geometric singular perturbation theory. In: Dynamical Systems. Springer, pp. 44–118.

Kammers, M.P., Longo, M.R., Tsakiris, M., Dijkerman, H.C., Haggard, P., 2009. Specificity and coherence of body representations. Perception 38 (12), 1804–1820.

Kuehn, C., 2011. A mathematical framework for critical transitions: Bifurcations, fast-slow systems and stochastic dynamics. Phys. D: Nonlinear Phenom. 240 (12), 1020–1035.

Kuznetsov, Y.A., 2004. Elements of Applied Bifurcation Theory, vol. 112. Springer Science & Business Media.

Litvak, S., Ullman, S., 2009. Cortical circuitry implementing graphical models. Neural Comput. 21 (11), 3010–3056.

Murphy, K.P., 2007. Conjugate Bayesian Analysis of the Gaussian Distribution. Tech. Rep. $2\sigma 2$. University of British, Columbia.

O'Malley, R.E., 2014. Historical Developments in Singular Perturbations. Springer.

Scheffer, M., Carpenter, S.R., Lenton, T.M., Bascompte, J., Brock, W., Dakos, V., Van De Koppel, J., Van De Leemput, I.A., Levin, S.A., Van Nes, E.H., et al., 2012. Anticipating critical transitions. Science 338 (6105), 344–348.

Watanabe, S., 2001a. Algebraic analysis for nonidentifiable learning machines. Neural Comput. 13 (4), 899–933.

Watanabe, S., 2001b. Algebraic information geometry for learning machines with singularities. Adv. Neural Inf. Process. Syst., 329–335.