Surprise-modulated belief update: how to learn within changing environments?

Mohammad Javad Faraji[1], Kerstin Preuschoff[2] and Wulfram Gerstner[1]

1. Brain Mind Institute & School of Computer and Communication Sciences, EPFL, CH.
2. Geneva Finance Research Institute & Center for Affective Sciences, Univ. of Geneva, CH.

Abstract

- We propose a new framework for surprise-driven learning that can be used for modeling how humans and animals learn in changing environments. It approximates optimal Bayesian learner, but with significantly reduced computational complexity.
- This framework consists of two components: (i) a confidence-adjusted surprise measure to capture environmental statistics as well as subjective beliefs, (ii) a surprise-minimization learning rule, or SMiLe-rule, which dynamically adjusts the balance between new and old information for belief update.

Confidence-Corrected Surprise

- Assume that the world is governed by a set of (unknown) true parameters θ^*. We model our knowledge of the world by a joint distribution $\mathbf{p(X,\theta) = p(X|\theta)\pi_0(\theta)}$, where $\mathbf{p(X|\theta)}$ determines how likely data sample $\mathbf{X}$ is to be generated, if the model parameter is θ. $\pi_0(\theta)$ denotes our current belief about the model parameters.
- We define the confidence-corrected surprise $\mathbf{S_{corr}(X;\pi_0)}$ of a data sample $\mathbf{X}$, for a subject with the current belief $\pi_0(\theta)$, as a KL divergence between the prior belief $\pi_0(\theta)$ and the scaled likelihood $\hat{\mathbf{p}}_\mathbf{X}(\theta) = \frac{\mathbf{p(X|\theta)}}{\int \mathbf{p(X|\theta)} d\theta}$, i.e.,

$$\mathbf{S_{corr}(X;\pi_0) = D_{KL}[\pi_0(\theta)||\hat{p}_X(\theta)]} = \int_\Theta \pi_0(\theta) \ln \frac{\pi_0(\theta)}{\hat{\mathbf{p}}_\mathbf{X}(\theta)} \, d\theta. \quad (1)$$

- Eq (1) is a *subjective*, model-dependent, measure of surprise. It is calculated before inference or learning occurs. Moreover, it decreases with the confidence of subject in her belief, represented by the entropy $\mathcal{H}(\pi_0)$.

Surprise Minimization: the SMiLe Rule

- After receiving a data sample $\mathbf{X}$, we update our belief from the prior $\pi_0(\theta)$ to the posterior $\mathbf{q}(\theta)$. We refer to such a mapping as a *learning rule* $\mathbf{L}$, i.e., $\mathbf{q = L(X,\pi_0)}$.
- *Surprise minimization* as a learning strategy: the internal model of the world is modified such that a data sample $\mathbf{X}$ is perceived as less surprising under the posterior than under the prior, i.e., $\mathbf{S_{corr}(X;q) \leq S_{corr}(X,\pi_0)}$.
- The impact function $\mathbf{\Delta S_{corr}(X;L) = S_{corr}(X;\pi_0) - S_{corr}(X;q)}$ quantifies how much the data sample $\mathbf{X}$ has impacted the internal model.
- For a given data sample $\mathbf{X}$, the impact function is maximized by the learning rule $\mathbf{L}$ that maps the prior $\pi_0(\theta)$ to the posterior $\mathbf{q}(\theta) = \hat{\mathbf{p}}_\mathbf{X}(\theta)$.
- As this posterior $\mathbf{q} = \hat{\mathbf{p}}_\mathbf{X}$ discards all previously learned information, it amounts to a valid though meaningless solution.
- To avoid overfitting to the last data sample we limit our search to posteriors $\mathbf{q}$ that are not too different from the prior, i.e.,

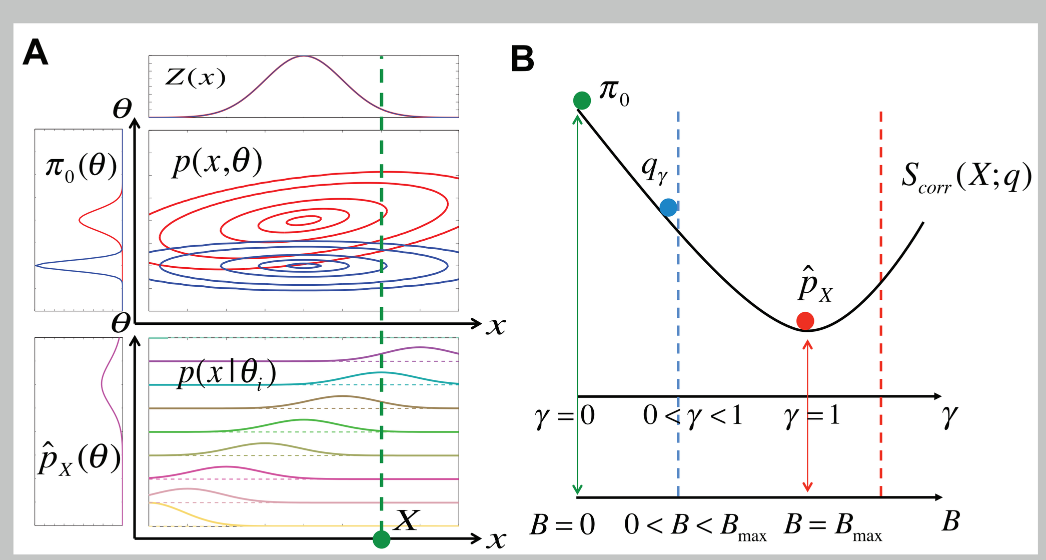
$$\min_{\mathbf{q: D_{KL}[q||\pi_0] \leq B}} \mathbf{S_{corr}(X;q)} \;\Rightarrow\; \mathbf{q}_\gamma(\theta) = \frac{\mathbf{p(X|\theta)^\gamma \pi_0(\theta)^{1-\gamma}}}{\mathbf{Z(X;\gamma)}} \quad (2)$$

- We call Eq (2) **S**urprise **Mi**nimization **Le**arning (**SMiLe**). It is reminiscent of Bayes rule except for γ that modulates the relative contribution of the likelihood and the prior to the posterior.
- Highly surprising data may signal a fundamental change in the context and should result in larger belief shifts. As such the bound $\mathbf{B}$ should increase with the level of surprise $\mathbf{S_{corr}}$:

$$\mathbf{B(X)} = \frac{\mathbf{mS_{corr}(X;\pi_0)}}{\mathbf{1 + mS_{corr}(X;\pi_0)}} \mathbf{B_{max}(X)}, \quad (3)$$

- Here, $\mathbf{m}$ is a subject-specific parameter that describes an organism's propensity toward changing its belief and $\mathbf{B_{max}(X) = D_{KL}[\hat{p}_X||\pi_0]}$.
- Note that there is a unique relationship between $\mathbf{B}$ and γ where

$$\mathbf{D_{KL}[q_\gamma||\pi_0] = B}, \quad (\mathbf{B > B'} \Rightarrow \gamma > \gamma') \quad (4)$$



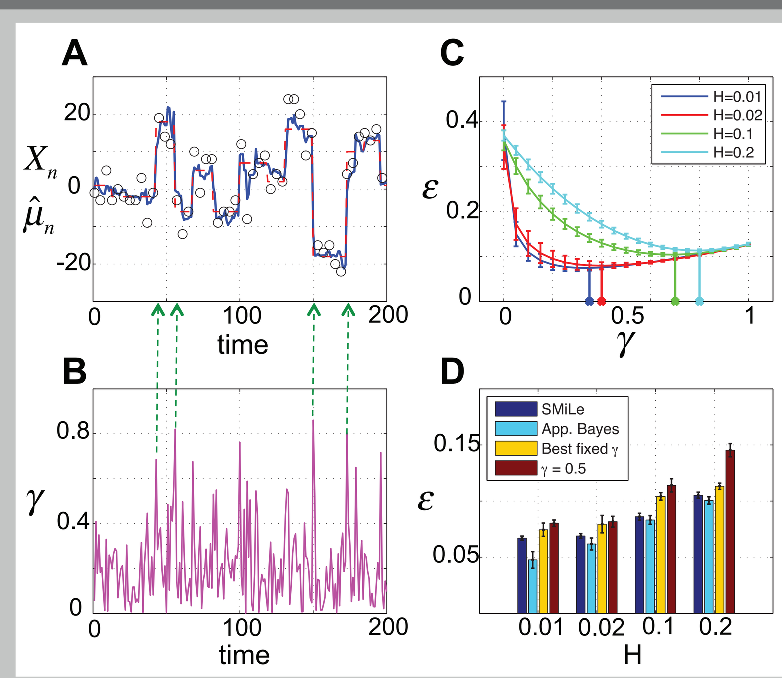
(A) A given data sample $\mathbf{X}$ (green dot) results in higher confidence-corrected surprise for the blue as compared to the red model, because π_0 in red is wider than in the blue model. (B) Solutions to the (constraint) optimization.

The objective function, i.e., the posterior surprise (black) is a parabolic landscape over γ. The boundary $\mathbf{B}$ constrains the range of γ and thus the set of admissible posteriors.

Surprise-Modulated Belief Update

After receiving a new data sample $\mathbf{X}$, we evaluate the surprise $\mathbf{S_{corr}(X;\pi_0)}$, Eq (1), which sets the bound $\mathbf{B}$, Eq (3), and allows us to solve for γ, Eq (4). We then update our belief using SMiLe, Eq (2).

Gaussian Mean Estimation Task



The SMiLe rule for this task becomes a delta-rule with a weight factor γ that increases with surprise.

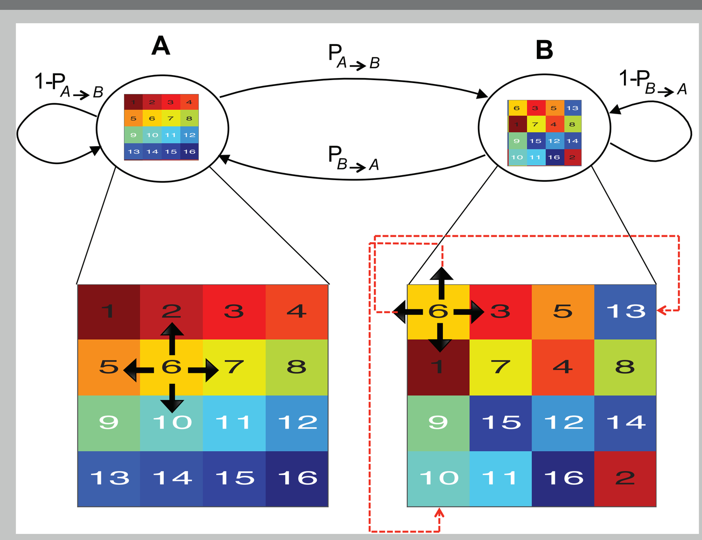
$$\hat{\mu}_\mathbf{n} = \gamma \mathbf{X_n} + (1-\gamma)\hat{\mu}_\mathbf{n-1} \quad (5)$$

$$\gamma = \sqrt{\frac{\mathbf{mS_{corr}(X_n)}}{\mathbf{1 + mS_{corr}(X_n)}}} \quad (6)$$

$$\mathbf{S_{corr}(X_n)} = \frac{(\mathbf{X_n} - \hat{\mu}_\mathbf{n-1})^2}{2\sigma_\mathbf{x}^2} \quad (7)$$

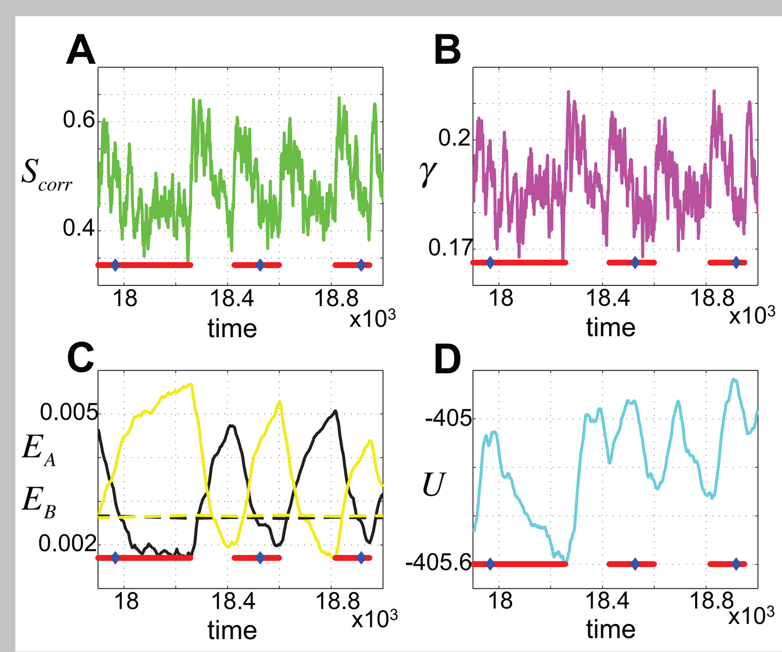
Our results show that the SMiLe rule outperforms delta-rule with *any* fixed learning rate.

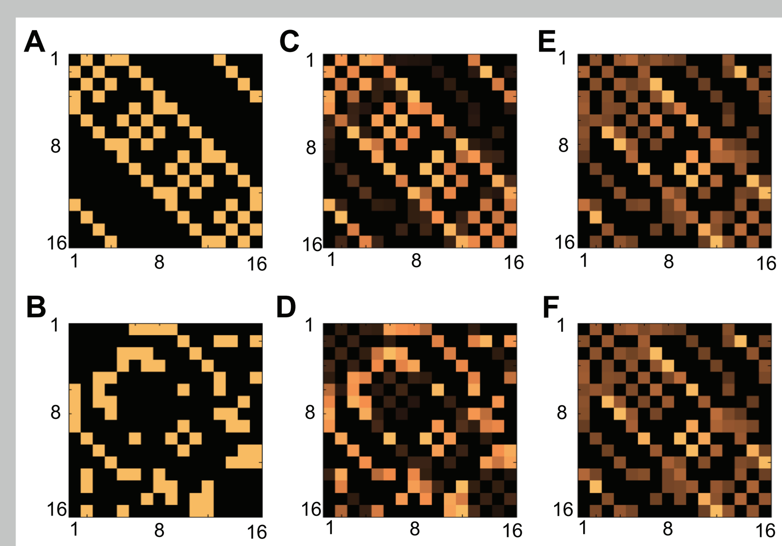
Maze-Exploration Task



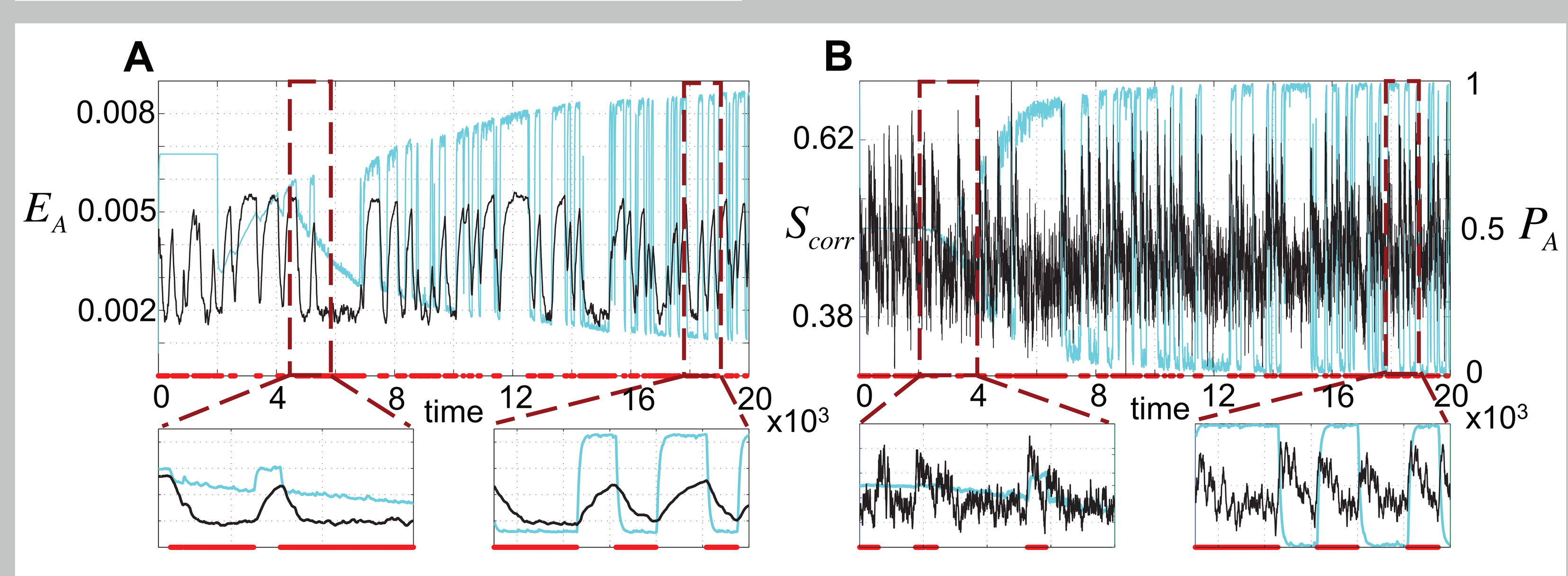
The task is to quickly adjust the model parameters (state transition probabilities) while exploring two environments $\mathcal{A}$ and $\mathcal{B}$, that are unexpectedly switched in a Markovian fashion. Both environments consist of **16** rooms but differ in topology. Neighboring rooms are connected and accessible through doors.

When the environment is switched, the sudden increase in surprise (green) causes the parameter γ (magenta) to increase. This is equivalent to discounting previously learned information and results in a quick adaptation to the new environment (see the estimation errors in yellow and black). Following a change point, model uncertainty (cyan) increases indicating the current model of topology is inaccurate.





The matrix of estimated transition probabilities using SMiLe (C&D), **100** time steps after a switch, resembles that of the present environment (A&B). For the naive Bayesian learner (which assumes there is only a single stable environment), the probabilities are estimated by averaging over the true parameters of both environment (E&F).



(A) Already after less than **1000** time steps the estimation error of SMiLe rule (black) drops below **0.002**. Only after **10000** time steps, the online EM algorithm (cyan), which benefits from knowing the true hidden Markov model of the task, achieves the same level of accuracy. While the solution of SMiLe in the long run is not as good, our algorithm benefits from a reduced computational complexity and simpler implementation. (B) The inferred probability of being in environment $\mathcal{A}$ (cyan) used in the online EM algorithm and the confidence-corrected surprise (black) used in the SMiLe.

References & Acknowledgment

[1] M.J. Faraji, K. Preuschoff, and W. Gerstner. "Balancing New Against Old Information: The Role of Surprise." arXiv:1606.05642

Funded by the European Research Council (grant no. 268 689, "MultiRules").

mohammadjavad.faraji@epfl.ch