Bayesian Models of Data Streams with Hierarchical Power Priors

AMiDST TOOLBOX

A Java Toolbox for Scalable Probabilistic Machine Learning

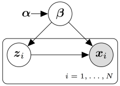
www.amidsttoolbox.com

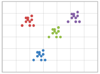
contact@amidsttoolbox.com

Andres R. Masegosa(1), Thomas D. Nielsen(2), Helge Langseth(3), Dario Ramos-Lopez(1), Antonio Salmeron(1), Anders L. Madsen(2,4)

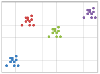
(1) University of Almeria [ES], (2) University of Aalborg [DK], (3) Norwegian University of Science and Technology [NO], (4) Hugin Expert A/S [DK]

The problem





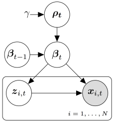
$\mathbf{x}_{t-1}$



$\mathbf{x}_t$

- **Variational Inference**
  - Latent Variable Models (LVMs).
  - Conjugate Exponential Family (CEF).
- **Learning from Data Streams**
  - Continuous Model Updating.
  - Bayesian posterior conditioned to non-finite data set.
  - Presence of Concept Drift (i.e. non i.i.d data).

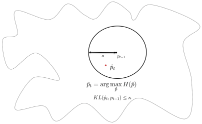
Our proposal



- **Out-of-the-box temporal extension.**
  - Global parameters β_t evolve over time.
  - Hierarchical prior modeling concept drift.
  - Closed-form Variational inference.
- $p(\beta_t|\beta_{t-1}, \rho_t)$ defined by a general implicit transition model.
  - Non-parametric form.
  - No need of expert knowledge modeling.
- $\rho_t \sim TrucantedExponential(\gamma)$, $\Omega(\rho_t) = [0, 1]$.
  - ρ_t close to 1 → No Drift at time t (i.e. $\beta_{t-1} \approx \beta_t$).
  - ρ_t close to 0 → Drift at time t (i.e. $\beta_{t-1} \not\approx \beta_t$).
  - $p(\rho_t|\mathbf{x}_{1:t})$ tracks concept drift.

Implicit Transition Models

$$\underbrace{p(\boldsymbol{\beta}_t|\boldsymbol{x}_{1:t-1})}_{\hat{p}_t} = \int \boxed{p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1})}\underbrace{p(\boldsymbol{\beta}_{t-1}|\boldsymbol{x}_{1:t-1})}_{p_{t-1}} d\boldsymbol{\beta}_{t-1}$$



$$\hat{\boldsymbol{\lambda}}_t = (1-\rho)\boldsymbol{\lambda}_u + \rho\boldsymbol{\lambda}_{t-1}$$

- **Closed-form solution for the Exponential Family**
  - λ natural parameter vector.
  - $\rho \in [0, 1]$ is defined by the user.
  - $\rho = 1$ equals $\kappa = 0$.
  - $\rho = 0$ equals $\kappa = \infty$.

Variational Inference

$q(\boldsymbol{\beta}, \boldsymbol{z}|\boldsymbol{x}, \boldsymbol{\alpha})$

- **Variational Inference in plain LVMs**
  - $(\boldsymbol{\lambda}^\star, \boldsymbol{\phi}^\star) = \arg\max_{\boldsymbol{\lambda},\boldsymbol{\phi}} \mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\phi}|\mathbf{x}, \boldsymbol{\alpha})$
  - Closed-form gradients for CEF models.

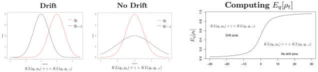
$q(\boldsymbol{\beta}_t, \boldsymbol{z}_t, \rho_t|\boldsymbol{\lambda}_t, \boldsymbol{\phi}_t, \omega_t)$

- **Variational Inference in temporal LVMs**
  - $(\boldsymbol{\lambda}_t^\star, \boldsymbol{\phi}_t^\star, \omega_t^\star) = \arg\max_{\boldsymbol{\lambda}_t,\boldsymbol{\phi}_t,\omega_t} \mathcal{L}_{HPP}(\boldsymbol{\lambda}_t, \boldsymbol{\phi}_t, \omega_t|\mathbf{x}, \boldsymbol{\lambda}_{t-1})$
  - **No closed-form gradients.**

Variational Inference with Hierarchical Porwer Priors

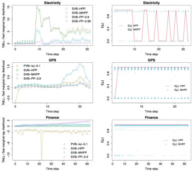
$$\mathcal{L}_{HPP} \geq \hat{\mathcal{L}}_{HPP}$$

- **A double-lower bound**
  - $\frac{\partial \hat{\mathcal{L}}_{HPP}}{\partial \boldsymbol{\beta}} = \frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}}$ (i.e. computed in closed-form).
  - $\frac{\partial \hat{\mathcal{L}}_{HPP}}{\partial \boldsymbol{\phi}} = \frac{\partial \mathcal{L}}{\partial \boldsymbol{\phi}}$ (i.e. computed in closed-form).



- **Closed-form gradient**
  - $\frac{\partial \hat{\mathcal{L}}_{HPP}}{\partial \omega_t} = KL(q_t, p_u) - KL(q_t, q_{t-1}) + \gamma - \omega_t$.
  - A measure of concept drift.
- **If only part of the data drifts:**
  - Multiple Hierarchical Power Priors (M-HPP).
  - Place independent $\rho_{k,t}$ for each parameter of the model.
  - Closed-form Variational inference.

Experimental Evaluation



(a) TMLL_t improvement over SVB (b) $\mathbb{E}_q[\rho_t]$ for SVB-HPP and SVB-MHPP

- **Summary of the evaluation:**
  - M-HPP is the most robust approach.
  - Adaptive forgetting mechanisms are usually needed.
  - Concept drift usually affects only a part of the model.

ICML @ Sydney
Thirty-fourth International Conference on Machine Learning

This work was partly carried out as part of the AMIDST project. AMIDST has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 619209.