Perpetua: Multi-Hypothesis Persistence Modeling for Semi-Static Environments - Supplementary Material

Miguel Saavedra-Ruiz, Samer B. Nashed, Charlie Gauthier and Liam Paull

March, 2025

Contents

1 Overview

This document contains supplementary material for the paper "Perpetua: Multi-Hypothesis Persistence Modeling for Semi-Static Environments". It provides the learning equations for a mixture of log-normal priors and highlights the changes needed to do training with the mixture of emergence filters.

2 Parameter Learning With a Mixture of Log-Normal Priors

We begin this section by revisiting the notation introduced in the learning section of the paper (§ IV-D). The goal is to estimate the parameters of the distributions over T^E/T and C from noisy data. To simplify notation, we represent C as a one-hot vector with K binary random variables $C = [C_1, C_2, \dots, C_K]^T$ instead of $C \in \{1, 2, \dots, K\}$. The element $C_k = 1$ if and only if the result belongs to class k, and zero otherwise. Without loss of generality, we demonstrate our derivations for the persistence model but the same steps can be followed to obtain the learning equations of the emergence model.

Mixture of Log-Normal Priors. Assume p_{T_k} is a log-normal distribution:

$$p_{T_k}(T \mid C_k = 1) \triangleq \frac{1}{T\sigma_k\sqrt{2\pi}} \exp\left(-\frac{(\ln T - \mu_k)^2}{2\sigma_k^2}\right) \tag{1}$$

since the mixture of persistence filters contains two hidden variables (C and T), we use the expectation-maximization algorithm Murphy [2022] to estimate the parameters $\Theta = \{(\mu_k, \sigma_k, \pi_k)\}_{k=1}^{K}$.

Given an observation sequence $\mathcal{Y}_{1:S}$ with $S >> N$, where a feature may transition multiple times between presence and absence, we assume that these transition times are identifiable. Therefore, we partition our data into M disjoint sets of the form $\mathcal{Y} \triangleq \{\mathcal{Y}_{N_j:N_j^{\text{Last}}}\}_{j=1}^{M}$, where N_j is the index of the first observation in the j^{th} set, N_j^{Last} the index of the last observation, and $N_1 = 1$. To ensure consistency in the estimator, we subtract the observation time $t_{N_{j-1}^{\text{Last}}}$ from all times $t \in \{t_{N_j}, \dots, t_{N_j}^{\text{Last}}\}$ in the j^{th} set, where we define $t_{N_0^{\text{Last}}} = 0$[1].

Denoting $\mathcal{T} = \{T_j\}_{j=1}^{M}$ and $\mathcal{C} = \{c_{jk}\}_{j=1,k=1}^{M,K}$, the *complete-data likelihood* can be written as

$$p(\mathcal{Y}, \mathcal{T}, \mathcal{C}; \Theta) = \prod_{j=1}^{M} p(\mathcal{Y}_{N_j:N_j^{\text{Last}}} \mid T_j) p(T_j \mid C_j) p(C_j), \tag{2}$$

where the distributions in (2) are

[1] This adjustment in the observation times is required irrespective of the choice of prior.

$$p(\mathcal{Y}_{N_j:N_j^{\text{Last}}} \mid T) = \prod_{t_i \le T} P_M^{1-y_{t_i}} (1-P_M)^{y_{t_i}} \prod_{t_i > T} P_F^{y_{t_i}} (1-P_F)^{1-y_{t_i}}, \tag{3}$$

$$p(T_j \mid C_j) = \prod_{k=1}^{K} \left[\frac{1}{T\sigma_k\sqrt{2\pi}} \exp\left(-\frac{(\ln T - \mu_k)^2}{2\sigma_k^2} \right) \right]^{c_{jk}}, \tag{4}$$

$$p(C_j) \triangleq \prod_{k=1}^{K} \pi_k^{c_{jk}}, \tag{5}$$

with (3) denoting the likelihood of the model, (4) a conditional prior over survival time T, and (5) a prior over mixture components.

Taking the logarithm of (2) and then the expectation with respect to $q^{[u+1]}(\mathcal{T}, \mathcal{C}) \triangleq p(\mathcal{T}, \mathcal{C} \mid \mathcal{Y}; \Theta^{[u]})$, where u is the current EM iteration, gives the maximization objective: *the expected complete-data log-likelihood*, defined as

$$\mathcal{L}(q, \Theta) \triangleq \mathbb{E}_q[\log p(\mathcal{Y}, \mathcal{T}, \mathcal{C}; \Theta)] \propto \sum_{j=1}^{M} \sum_{k=1}^{K} \left(\mathbb{E}_q[c_{jk}] (\log \pi_k - \log \sigma_k) - \mathbb{E}_q[c_{jk} \log T_j] - \mathbb{E}_q \left[\frac{c_{jk}(\ln T_j - \mu_k)^2}{2\sigma_k^2} \right] \right), \tag{6}$$

where we dropped the terms that do not depend on Θ.

E-Step: In this step, we fix the parameters $\Theta^{[u]}$ and compute the distribution $q^{[u+1]}$. Although q can be intractable, we show that it allows for a closed-form solution. From the definition of $q^{[u+1]}(T_j, c_{jk}) \triangleq p(T_j, c_{jk} \mid \mathcal{Y}_{N_j:N_j^{\text{Last}}})$, we have

$$\begin{aligned} q^{[u+1]}(T_j, c_{jk}) &= p(T_j, c_{jk} \mid \mathcal{Y}_{N_j:N_j^{\text{Last}}}) \\ &= \frac{p(\mathcal{Y}_{N_j:N_j^{\text{Last}}} \mid T_j) p(T_j \mid c_{jk}) p(c_{jk})}{\int_0^\infty \sum_{l=1}^{K} p(\mathcal{Y}_{N_j:N_j^{\text{Last}}} | \tau) p(\tau | c_{jl}) p(c_{jl}) d\tau}, \qquad (7) \\ q^{[u+1]}(c_{jk}) &= p(c_{jk} \mid \mathcal{Y}_{N_j:N_j^{\text{Last}}}). \qquad (8) \end{aligned}$$

Note that (8) correspond to the *posterior weights* derived in § IV-A. We define the expectation $\phi_{jk}^{[u+1]} \triangleq \mathbb{E}_q[c_{jk}] = q^{[u+1]}(c_{jk} = 1)$. Since this expectation is exactly the posterior weights, it can be computed by following the procedure outlined in § IV-A.

By using the same decomposition used to incrementally compute the evidence (see § III-B and (5) in the paper), the other two expectations can be computed as:

$$\nu_{jk} \triangleq \mathbb{E}_q[c_{jk} \ln T_j] = \frac{p(c_{jk}=1)}{p(\mathcal{Y}_{N_j:N_j^{\text{Last}}})} \sum_i p(Y_{N_{j-1}:N_j} \mid t_i) \int_{t_i}^{t_{i+1}} p(\tau \mid c_{jk}=1) \ln \tau \, d\tau, \tag{9}$$

$$\kappa_{jk} \triangleq \mathbb{E}_q[c_{jk}(\ln T_j - \mu_k^2)] = \frac{p(c_{jk}=1)}{p(\mathcal{Y}_{N_j:N_j^{\text{Last}}})} \sum_i p(Y_{N_{j-1}:N_j} \mid t_i) \int_{t_i}^{t_{i+1}} p(\tau \mid c_{jk}=1)(\ln \tau - \mu_k)^2 \, d\tau, \tag{10}$$

with $i \in \{0\} \cup \{N_j, \dots, N_j^{\text{Last}}\}$. The u superscript is omitted for clarity. Additionally, we set $t_0 = 0$ and $t_{N_j^{\text{Last}}+1} = \infty$ (see (5) in the paper). The evidence, $p(\mathcal{Y}_{N_j:N_j^{\text{Last}}})$, is computed alongside the posterior weights, as outlined in §IV-A. For a log-normal distribution, it can be shown that the integrals in (9) and (10) admit the following closed-form

$$\int_{t_i}^{t_{i+1}} p(\tau \mid c_{jk}=1) \ln \tau \, d\tau = \frac{1}{\sqrt{2\pi}} \left[\sigma_k^{[u]} \exp(-a_{ik}) - \sigma_k^{[u]} \exp(-b_{ik}) + \mu_k^{[u]} \sqrt{\frac{\pi}{2}} \left(\text{erf}(\sqrt{b_{ik}}) - \text{erf}(\sqrt{a_{ik}}) \right) \right],$$

$$\begin{aligned} \int_{t_i}^{t_{i+1}} p(\tau \mid c_{jk}=1)(\ln \tau - \mu_k)^2 \, d\tau = -2\mu_k^{[u+1]} \nu_{jk} &+ \left(\frac{(\mu_k^{[u+1]})^2}{2} + \frac{(\mu_k^{[u]})^2}{2} + \frac{(\sigma_k^{[u]})^2}{2\sqrt{\pi}} \right) \left(\text{erf}(\sqrt{b_{ik}}) - \text{erf}\sqrt{(a_{ik})} \right) \\ &- \sqrt{\frac{2}{\pi}} \mu_k^{[u]} \sigma_k^{[u]} \left(\exp(-b_{ik}) - \exp(-a_{ik}) \right) \\ &+ \frac{(\sigma_k^{[u]})^2}{\pi} \left(\sqrt{a_{ik}} \exp(-a_{ik}) - \sqrt{b_{ik}} \exp(-b_{ik}) \right), \end{aligned}$$

where $\mathrm{erf}(z) \triangleq \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$ is the error function, and the auxiliary variables $a_{ik} = \frac{(\ln t_i - \mu_k^{[u]})^2}{2(\sigma_k^{[u]})^2}$ and $b_{ik} = \frac{(\ln t_{i+1} - \mu_k^{[u]})^2}{2(\sigma_k^{[u]})^2}$ are introduced to simplify the notation. Note that $\mu_k^{[u+1]}$ is the solution obtained in the M-step. Thus, computing the variance requires first optimizing the mean (this is analogous to the standard MLE for a standard normal distribution).

M-Step: In this step, we fix the variational distribution $q^{[u+1]}$ and optimize the parameters $\Theta^{[u]}$ by maximizing the objective in (6). This leads to the parameters updates

$$\mu_k^{[u+1]} = \frac{\sum_{j=1}^{M} \nu_{jk}}{\sum_{j=1}^{M} \phi_{jk}}, \qquad \sigma_k^{[u+1]} = \sqrt{\frac{\sum_{j=1}^{M} \kappa_{jk}}{\sum_{j=1}^{M} \phi_{jk}}}, \qquad \text{and} \qquad \pi_k^{[u+1]} = \frac{\sum_{j=1}^{M} \phi_{jk}}{\sum_{j=1}^{M} \sum_{l=1}^{K} \phi_{jl}}. \tag{11}$$

Therefore, to optimize $\Theta^{[u]}$, we first perform the E-step using (8), (9) and (10) while keeping the parameters fixed. Then, with $q^{[u+1]}$ updated, we optimize $\Theta^{[u+1]}$ using (11). This process is repeated for U iterations or until convergence[2].

3 Parameter Learning for a Mixture of Emergence Filters

The learning procedure for mixtures of exponential and log-normal distributions can be directly applied to estimate the parameters of the emergence filter. More generally, this holds for any prior distribution over T. If we want to estimate the parameters of the distribution over T^E and C, we need to do the two following modifications to the procedure presented in §2 and § IV-D in the paper:

1. The likelihood (3) has to be updated according to the mixture of emergence filters' model (see § IV-B)

$$p(\mathcal{Y}_{N_j:N_j^{\text{Last}}} \mid T^E) = \prod_{t_i \leq T^E} P_F^{y_{t_i}} (1 - P_F)^{1 - y_{t_i}} \prod_{t_i > T^E} P_M^{1 - y_{t_i}} (1 - P_M)^{y_{t_i}} \tag{12}$$

2. To compute the expectation $\phi_{jk}^{[u+1]}$ and the evidence $p(\mathcal{Y}_{N_j:N_j^{\text{Last}}})$, we must use the filtering equations for the mixture of emergence filters (§ IV-B) instead of those for the mixture of persistence filters (§ IV-A).

Therefore, with these two modifications, we can now compute the parameters for the mixture of emergence filters.

References

Kevin P. Murphy. *Probabilistic Machine Learning: An introduction.* MIT Press, 2022.

[2]The learning code for a mixture of exponential and log-normal distributions can be found here: https://montrealrobotics.ca/perpetua.