A Marginalized Particle Gaussian Process Regression

Yali Wang and Brahim Chaib-draa {wang;chaib}@damas.ift.ulaval.ca

Department of Computer Science, Laval University, Quebec, Canada

Introduction

- Gaussian process (GP) is an elegant nonparametric regression method
- Computation load $O(n^3)$ of GP training restricts its applicability in practice
- We propose an accurate and computationally efficient marginalized particle GP (MPGP) regression

Data Construction

- The large training set in practice is usually constructed by gathering small subsets several times
- The t_{th} collection: $(X_t, \mathbf{y}_t) = \{(\mathbf{x}_t^1, y_t^1), \cdots (\mathbf{x}_t^{n_t}, y_t^{n_t})\}$ where each output is generated by $y_t^i = f(\mathbf{x}_t^i) + N(0, a_0^2)$
- Goal: inferring $f(X_\star)$ at test inputs $X_\star = \{\mathbf{x}_\star^1, \cdots \mathbf{x}_\star^m\}$ by the training set $(X_{1:t}, \mathbf{y}_{1:t})$

Gaussian Process Regression

- **Definition**:

  A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution
- **Denotation**: $f(\mathbf{x}) \sim GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$
  - The mean function $m(\mathbf{x}) = E[f(\mathbf{x})]$ is usually set to zero
  - The covariance function $k(\mathbf{x}, \mathbf{x}') = E[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]$
    - ◇ $k_{SE} = a_1^2 exp[-0.5 a_2^{-2}(\mathbf{x} - \mathbf{x}')^T(\mathbf{x} - \mathbf{x}')]$
    - ◇ $k_{NN} = a_3^2 sin^{-1}[a_4^{-2}\tilde{\mathbf{x}}^T\tilde{\mathbf{x}}'((1 + a_4^{-2}\tilde{\mathbf{x}}^T\tilde{\mathbf{x}})(1 + a_4^{-2}\tilde{\mathbf{x}}'^T\tilde{\mathbf{x}}'))^{-0.5}]$ where $\tilde{\mathbf{x}} = [1 \;\; \mathbf{x}^T]^T$
- **Offline Inference**:

1. **Learning** $\theta = [a_0, a_1, a_2, a_3, a_4]^T$ given $(X_{1:t}, \mathbf{y}_{1:t})$
   - ◇ Method 1: drawing samples from $p(\theta|X_{1:t}, \mathbf{y}_{1:t})$ by MCMC
   - ◇ Method 2: maximizing $p(\mathbf{y}_{1:T}|X_{1:T}, \theta)$ by gradient based optimization
2. **Estimating** $p(f(X_\star)|X_{1:T}, \mathbf{y}_{1:T}, X_\star, \theta)$
   - ◇ Based on GP definition and noise property, it is jointly Gaussian distributed

     $\bar{f}(X_\star) = K_\theta(X_\star, X_{1:t})[K_\theta(X_{1:t}, X_{1:t}) + a_0^2 I]^{-1}\mathbf{y}_{1:t}$

     $P(X_\star, X_\star) = K_\theta(X_\star, X_\star) - K_\theta(X_\star, X_{1:t})[K_\theta(X_{1:t}, X_{1:t}) + a_0^2 I]^{-1}K_\theta(X_\star, X_{1:t})^T$
   - ◇ The $K_\theta(X_\star, X_{1:T})$, $K_\theta(X_{1:T}, X_{1:T})$ and $K_\theta(X_\star, X_\star)$ is calculated by $k(\mathbf{x}, \mathbf{x}')$

Marginalized Particle Gaussian Process Regression

- **State Space Model (SSM)**:

1. **State Equation**:
   - ◇ Artificial dynamics for θ by kernel smoothing: $\theta_t = b\theta_{t-1} + (1-b)\bar{\theta}_{t-1} + s_{t-1}$
     - ○ $b = (3\delta - 1)/(2\delta)$ where δ is typically around 0.95-0.99
     - ○ $\bar{\theta}_{t-1}$ and Σ_{t-1}: Monte Carlo mean and variance
     - ○ $s_{t-1} \sim N(0, r^2\Sigma_{t-1})$ where $r^2 = 1 - b^2$
   - ◇ Markovian transition between f_t^c and f_{t-1}^c by GP: $f_t^c = G(\theta_t)f_{t-1}^c + N(\mathbf{0}, Q(\theta_t))$
     - ○ $f_t^c = f(X_t^c)$ where $X_t^c = X_t \cup X_\star$ and $G(\theta_t) = K_{\theta_t}(X_t^c, X_{t-1}^c)K_{\theta_t}^{-1}(X_{t-1}^c, X_{t-1}^c)$
     - ○ $Q(\theta_t) = K_{\theta_t}(X_t^c, X_t^c) - K_{\theta_t}(X_t^c, X_{t-1}^c)K_{\theta_t}^{-1}(X_{t-1}^c, X_{t-1}^c)K_{\theta_t}(X_t^c, X_{t-1}^c)^T$
2. **Observation Equation**:

   $$\mathbf{y}_t = H_t f_t^c + N(0, R(\theta_t)) = f(X_t) + N(0, a_{0,t}^2 I)$$

- **Bayesian Online Inference by Marginalized Particle Filter**

$$p(f_t^c, \theta_{1:t}|X_{1:t}, X_\star, \mathbf{y}_{1:t}) = p(\theta_{1:t}|X_{1:t}, X_\star, \mathbf{y}_{1:t})p(f_t^c|\theta_{1:t}, X_{1:t}, X_\star, \mathbf{y}_{1:t})$$

1. **Sequential Importance Sampling Mechanism for** $p(\theta_{1:t}|X_{1:t}, X_\star, \mathbf{y}_{1:t})$

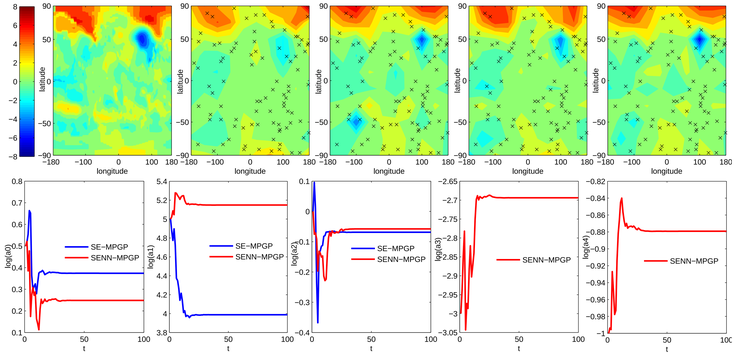
   $p(\theta_{1:t}|X_{1:t}, X_\star, \mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \theta_{1:t}, X_{1:t}, X_\star)p(\theta_t|\theta_{t-1})p(\theta_{1:t-1}|X_{1:t-1}, X_\star, \mathbf{y}_{1:t-1})$
   - ◇ Importance weight: $p(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \theta_{1:t}, X_{1:t}, X_\star) = N(H_t f_{t|t-1}^c, H_t P_{t|t-1}^c H_t^T + R(\theta_t))$
   - ◇ $f_{t|t-1}^c$ and $P_{t|t-1}^c$: predictive mean and covariance of Kalman filter
2. **Kalman Filtering Framework for** $p(f_t^c|\theta_{1:t}, X_{1:t}, X_\star, \mathbf{y}_{1:t})$
   - ◇ **Predict**: $f_{t|t-1}^c = G(\theta_t)f_{t-1|t-1}^c$, $P_{t|t-1}^c = G(\theta_t)P_{t-1|t-1}^c G(\theta_t)^T + Q(\theta_t)$
   - ◇ **Update**: $\Gamma_t = P_{t|t-1}^c H_t^T(H_t P_{t|t-1}^c H_t^T + R(\theta_t))^{-1}$

     $f_{t|t}^c = f_{t|t-1}^c + \Gamma_t(\mathbf{y}_t - H_t f_{t|t-1}^c)$, $P_{t|t}^c = P_{t|t-1}^c - \Gamma_t H_t P_{t|t-1}^c$

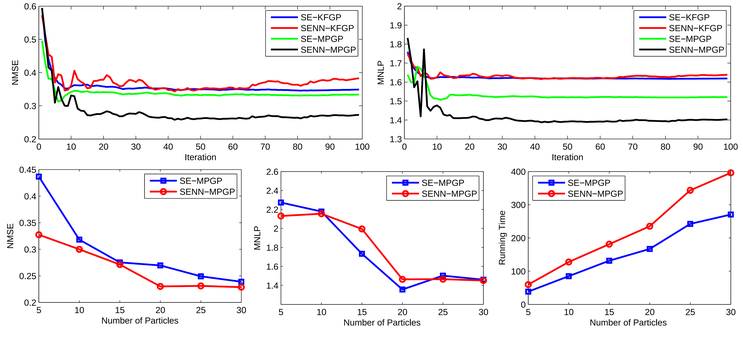
- **Marginalized Particle GP (MPGP) Algorithm**
  - ◇ For $i = 1, 2, ...N$
    - ○ Drawing $\theta_t^i \sim p(\theta_t|\tilde{\theta}_{t-1}^i)$ to construct $G(\theta_t^i)$, $Q(\theta_t^i)$, $R(\theta_t^i)$ in SSM
    - ○ One-step predict-update Kalman filter
    - ○ Using $f_{t|t-1}^{c,i}$, $P_{t|t-1}^{c,i}$, $R(\theta_t^i)$ to compute the normalized importance weight w_t^i
  - ◇ Estimating $\hat{\theta}_t$, $\hat{f}_{t|t}^c$, $\hat{P}_{t|t}^c$ by weighted particles
  - ◇ According to w_t^i, resampling θ_t^i, $f_{t|t}^{c,i}$, $P_{t|t}^{c,i}$ to obtain $\tilde{\theta}_t^i$, $\tilde{f}_{t|t}^{c,i}$, $\tilde{P}_{t|t}^{c,i}$ for the next step

Experiments: Global Surface Temperature Dataset

- **Predicted Temperature by SE, SENN-KFGP and our SE, SENN-MPGP**



- **Accuracy and Efficiency Evaluation**



Conclusion

- We have proposed a highly accurate marginalized particle GP (MPGP) regression using a recursive Bayesian filtering framework
- Our MPGP provides a new online learning approach for hidden function values and GP hyperparameters with computation efficiency $O(NTS^3)$