Rapid Uncertainty Computation with Gaussian Processes and Histogram Intersection Kernels

Alexander Freytag¹, Erik Rodner², Paul Bodesheim¹, and Joachim Denzler¹
¹Computer Vision Group, Friedrich Schiller University Jena, Germany
²ICSI Vision Group, UC Berkeley, California
http://www.inf-cv.uni-jena.de/gp_hik.html

Motivation

- Is this decision reliable?
- Calculate the GP predictive variance!

Classification hides the underlying uncertainty of decisions

- Gaussian process (GP) inference is not available for n \(\gg 10,000\) \(O(n^3)\) for learning and \(O(n)\) for prediction
- Previous work: large-scale GP classification [4]

Goal: use the GP framework even for large-scale incremental and active learning by using approximated GP variance estimates

Our contributions

2. Several approximations and methods able to compute the predictive GP variance even for large-scale scenarios.
3. Evaluation of our methods with active learning scenarios.

Gaussian process inference and fast kernel calculations

- Flexible Bayesian approach for regression and classification.
- We use label regression for one-vs-all classification [3].
- Closed form solution of the posterior \(p(y_i|x^*) \sim N(\mu_x, \sigma^2_x)\)
- \(\mu_x = k^T(x_\star)(K + \sigma^2 \cdot I)^{-1}y\)
- \(\sigma^2_x = k_\star - k^T(x_\star)(K + \sigma^2 \cdot I)^{-1}k_\star + \sigma^2\)
- Fast calculation of \(K\) \& \(\alpha\) and \(K \cdot v\) with HIK [2, 4, 5]:
  \(K_{\text{HIK}}(x, x') = \sum_{i=1}^{D_{\text{HIK}}} \min(x_i, x'_i)\)
- Observation: \(K\) \& \(\alpha\) is piecewise linear
- Idea: sorting \(x_i\) and calculating look-up tables
- Learning: Calculation of \(\alpha\) with conjugate gradients and fast \(K \cdot v\) multiplications (kernel matrix is not stored explicitly)

Predictive variance estimates

1. Represents uncertainty of the classification estimate
2. Useful for active learning and one-class classification

Efficient approximations of the predictive variance

1. Precise Uncertainty Prediction (PUP): compute \(k_\star\) and apply a linear solver.
2. Fine Approximation of the Pred. Uncertainty (FAPU):
   - Based on the top \(k\) eigenvalues \(\mu_i\) and \(-v\)-vectors of the regularized kernel matrix:
   \[ k^T(\Sigma + \sigma^2 \cdot I)^{-1}k_\star \geq \left( \sum_{i=1}^{k} \frac{1}{\mu_i}v_i + \frac{1}{\mu_{k+1}} \left( \|k_\star\|^2 - k_n^2 \right) \right), \]
   where \(v_i\) is the projection on the \(i\)th eigenvector. The norm of \(k_\star\) can also be approximated efficiently.
3. Rough Approximation of the Pred. Uncertainty (RAPU):
   - Use the FAPU method with \(k = 1\). The quantization idea of [2] can also be applied here (q-RAPU).

Runtime evaluation

<table>
<thead>
<tr>
<th>Approach</th>
<th>Asymptotic runtime</th>
<th>Exact?</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>q-RAPU</td>
<td>(O(D))</td>
<td>no</td>
<td>34(\mu)s</td>
</tr>
<tr>
<td>RAPU</td>
<td>(O(D \log n))</td>
<td>no</td>
<td>267ms</td>
</tr>
<tr>
<td>FAPU</td>
<td>(O(D \log n + kn))</td>
<td>no</td>
<td>1.15s</td>
</tr>
<tr>
<td>PUP</td>
<td>(O(Dn))</td>
<td>yes</td>
<td>(&gt;1)min</td>
</tr>
<tr>
<td>GP-standard</td>
<td>(O(n^2 + nD))</td>
<td>yes</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Conclusions

- Active and incremental learning with tens of thousands of learning examples
- Estimation of the GP predictive variance with different degrees of approximation
- Visual recognition with large-scale data requires more than classification decisions!

References