Rapid Uncertainty Computation with Gaussian Processes and Histogram Intersection Kernels

Code Available

Alexander Freytag[1], Erik Rodner[1,2], Paul Bodesheim[1], and Joachim Denzler[1]

[1]Computer Vision Group, Friedrich Schiller University Jena, Germany
[2]ICSI Vision Group, UC Berkeley, California

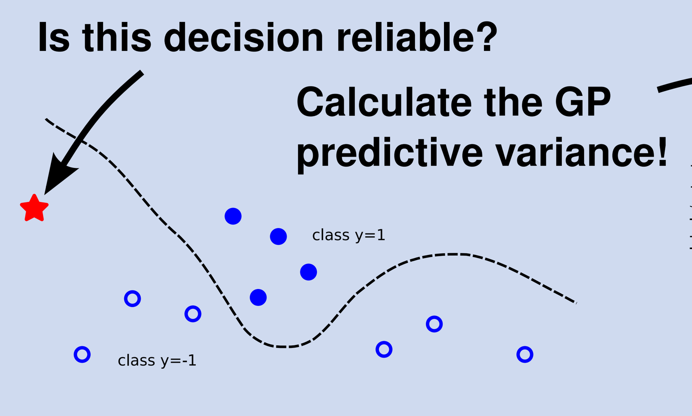
http://www.inf-cv.uni-jena.de/gp_hik.html

since 1558

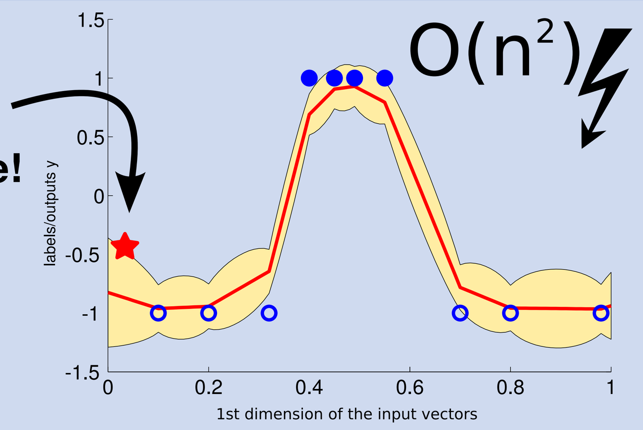
INTERNATIONAL COMPUTER SCIENCE INSTITUTE

Friedrich Schiller University Jena
Computer Vision Group

Motivation



Classification hides the underlying uncertainty of decisions



GP Regression with variance estimates

- Gaussian process (GP) inference is not available for $n \gg 10,000$ ($\mathcal{O}(n^3)$ for learning and $\mathcal{O}(n)$ 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

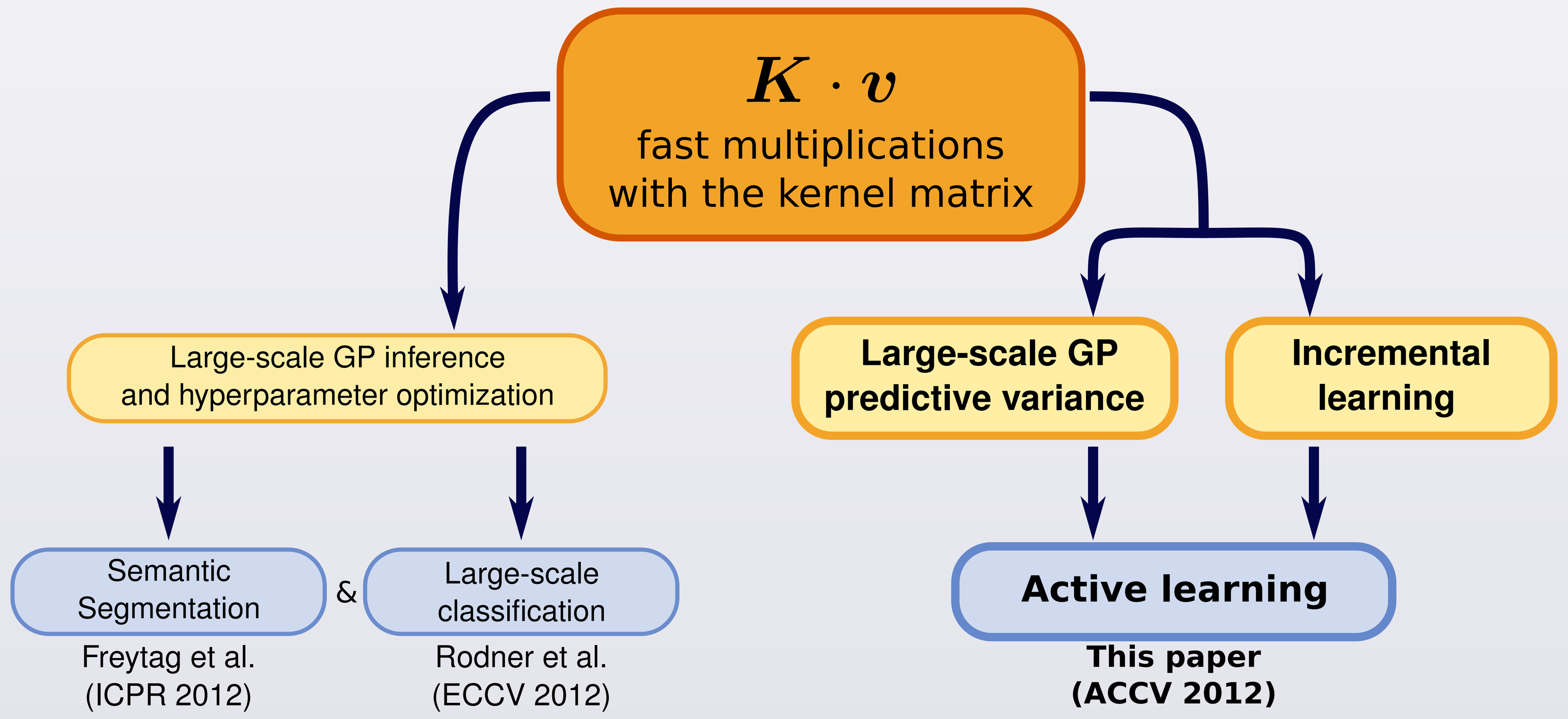
1. Extending [4] towards efficient **incremental learning with GP** (new categories, new examples).
2. Several approximations and methods able to compute the **predictive GP variance even for large-scale scenarios**.
3. Evaluation of our methods with active learning scenarios.

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 $p(y_*|\boldsymbol{x}^*) \sim \mathcal{N}(\mu_*, \sigma_*^2)$

$$\mu_* = \boldsymbol{k}_*^T(\mathbf{K} + \sigma^2 \cdot \mathbf{I})^{-1}\boldsymbol{y}$$
$$\sigma_*^2 = k_{**} - \boldsymbol{k}_*^T(\mathbf{K} + \sigma^2 \cdot \mathbf{I})^{-1}\boldsymbol{k}_* + \sigma^2 \ .$$

- Fast calculation of $\boldsymbol{k}_*^T\boldsymbol{\alpha}$ and $\mathbf{K}\boldsymbol{v}$ with HIK [2, 4, 5]: $K^{\text{hik}}(\boldsymbol{x}, \boldsymbol{x}') = \sum_{d=1}^{D} \min(x_d, x'_d)$
- **Observation**: $\boldsymbol{k}_*^T\boldsymbol{\alpha}$ is piecewise linear
- **Idea**: sorting x_d and calculating look-up tables
- **Learning**: Calculation of $\boldsymbol{\alpha}$ with conjugate gradients and fast $\mathbf{K} \cdot \boldsymbol{v}$ 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 $\boldsymbol{k}_*$ and apply a linear solver.
2. Fine Approximation of the Pred. Uncertainty (FAPU): based on the top k eigenvalues μ_i and -vectors of the regularized kernel matrix:

$$\boldsymbol{k}_*^T(\mathbf{K} + \sigma^2\mathbf{I})^{-1}\boldsymbol{k}_* \geq \left(\sum_{i=1}^{k}\frac{1}{\mu_i}\nu_i^2 + \frac{1}{\mu_{k+1}}\left(||\boldsymbol{k}_*||^2 - \sum_{i=1}^{k}\nu_i^2\right)\right) \ ,$$

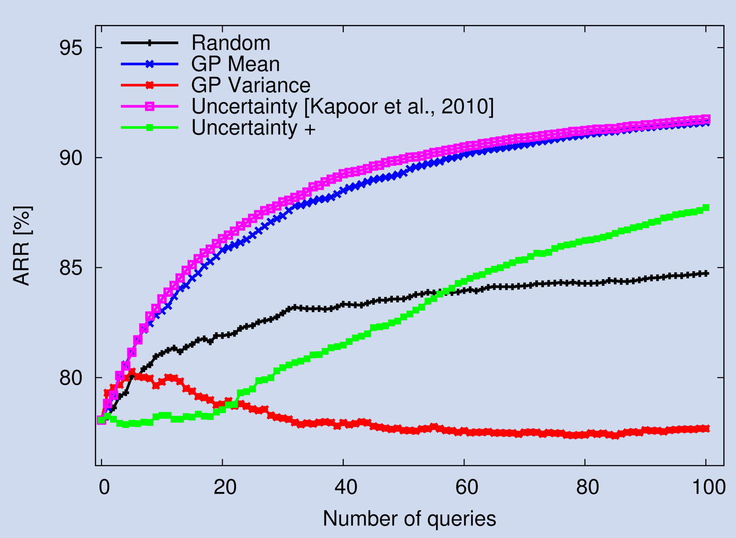
where ν_i is the projection on the ith eigenvector. The norm of $\boldsymbol{k}_*$ can also be approximated efficiently.
3. Rough Approximation of the Pred. Uncertainty (RAPU): use the FAPU method with $k = 0$. The quantization idea of [2] can also be applied here (q-RAPU).

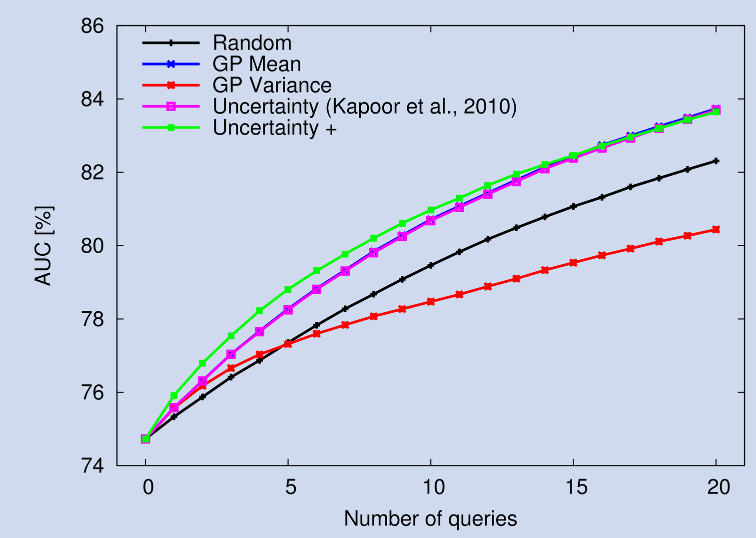
Runtime evaluation

Approach	Asymptotic runtime	Exact?	Time
q-RAPU	$\mathcal{O}(D)$	no	34 μs
RAPU	$\mathcal{O}(D \log n)$	no	267ms
FAPU	$\mathcal{O}(D \log n + kn)$	no	1.15s
PUP	$\mathcal{O}(Dn)$	yes	$>$ 1min
GP-standard	$\mathcal{O}(n^2 + nD)$	yes	n/a

Runtime evaluation done with $50,050$ training examples.

Active learning experiments





- Experiment with a (left fig.) synthetic checkerboard learning task and (right fig.) parts of the ImageNet database (ILSVRC'10)
- Different active learning strategies [1]:

$$\mathcal{Q}_{\mu_*} = \operatorname*{argmin}_i |\mu_*(\widehat{\boldsymbol{x}}_i)| \qquad \mathcal{Q}_{unc} = \operatorname*{argmin}_i \frac{|\mu_*(\widehat{\boldsymbol{x}}_i)|}{\sqrt{\sigma^2 + \sigma_*^2(\widehat{\boldsymbol{x}}_i)}}$$

$$\mathcal{Q}_{\sigma_*^2} = \operatorname*{argmax}_i \sigma_*^2(\widehat{\boldsymbol{x}}_i) \qquad \mathcal{Q}_{\text{Unc+}} = \operatorname*{argmin}_i \left(|\mu_*(\widehat{\boldsymbol{x}}_i)| + \sqrt{\sigma^2 + \sigma_*^2(\widehat{\boldsymbol{x}}_i)}\right)$$

- The heuristic approaches $\mathcal{Q}_{\text{Unc}}$ and $\mathcal{Q}_{\text{Unc+}}$ give consistent performance boosts when compared to random selection.
- **Further experiments in the paper: incremental learning and one-class classification**

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

[1] Asish Kapoor, Kristen Grauman, Raquel Urtasun, and Trevor Darrell. Gaussian processes for object categorization. *International Journal of Computer Vision*, 88:169–188, 2010.

[2] S. Maji, A.C. Berg, and J. Malik. Classification using intersection kernel support vector machines is efficient. In *CVPR*, pages 1 –8, 2008.

[3] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. The MIT Press, 01 2006.

[4] Erik Rodner, Alexander Freytag, Paul Bodesheim, and Joachim Denzler. Large-scale gaussian process classification with flexible adaptive histogram kernels. In *ECCV*, 2012.

[5] Jianxin Wu. A fast dual method for hik svm learning. In *ECCV*, pages 552–565, 2010.