

Variational Weakly Supervised Gaussian Processes

Melih Kandemir¹

melih.kandemir@iwr.uni-heidelberg.de

Manuel Haußmann¹

manuel.haussmann@iwr.uni-heidelberg.de

Ferran Diego¹

ferran.diego@iwr.uni-heidelberg.de

Kumar Rajamani²

KumarThirunellai.Rajamani@in.bosch.com

Jeroen van der Laak³

Jeroen.vanderLaak@radboudumc.nl

Fred A. Hamprecht¹

fred.hamprecht@iwr.uni-heidelberg.de

¹ Heidelberg University, HCI
Heidelberg, Germany

² Robert Bosch Engineering
Bangalore, India

³ Radboud University Medical Center
Nijmegen, Netherlands

We introduce the first model to perform weakly supervised learning with Gaussian processes (GPs) on up to millions of instances. The key ingredient to achieve this scalability is to replace the standard assumption of MIL that the bag-level prediction is the maximum of instance-level estimates with the accumulated evidence of instances within a bag. Given data set of N instances $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] = \{\mathbf{X}_1 \cup \mathbf{X}_1 \cup \dots \cup \mathbf{X}_B\}$ composed of B disjoint partitions, called bags, and supervised by bag labels $\mathbf{T} = [t_1, \dots, t_B]$, we propose the following model to infer a Bayesian weakly supervised predictor

$$p(\mathbf{u}|\mathbf{Z}) = \mathcal{N}(\mathbf{u}|\mathbf{0}, \mathbf{K}_{\mathbf{ZZ}}), \quad (1)$$

$$p(\mathbf{f}|\mathbf{u}, \mathbf{X}, \mathbf{Z}) = \mathcal{N}(\mathbf{f}|\mathbf{K}_{\mathbf{XZ}}\mathbf{K}_{\mathbf{ZZ}}^{-1}\mathbf{u}, \quad (2)$$

$$\text{diag}(\mathbf{K}_{\mathbf{XX}} - \mathbf{K}_{\mathbf{XZ}}\mathbf{K}_{\mathbf{ZZ}}^{-1}\mathbf{K}_{\mathbf{ZX}})),$$

$$p(\mathbf{T}|\mathbf{f}) = \prod_{b=1}^B \text{Bernoulli}(T_b | \sigma(\mathbf{f}_b^T \mathbf{1})), \quad (3)$$

where $\mathbf{Z} = [\mathbf{z}_1; \dots; \mathbf{z}_P]$ with $P \ll N$ is a tiny pseudo data set called the *inducing point set* and \mathbf{u} the corresponding outputs. We refer to this model as the *Variational Weakly Supervised Gaussian Process (VWSGP)*¹. Here, Equations 1 and 2 constitute the sparse GP prior and Equation 3 is a Bernoulli likelihood ensuring the model to predict binary outputs. Thanks to the sum term $\mathbf{f}_b^T \mathbf{1}$ in the likelihood, this model can be trained by closed-form variational inference updates. Hence, keeping all parameters but one fixed, the remaining parameter can be analytically fit to the global optimum. This virtue leads

to charmingly fast convergence, fitting perfectly to large-scale learning setups.

We evaluate our VWSGP on the Pascal VOC '07 benchmark and two medical image analysis applications: i) Diabetic Retinopathy screening (DR), and ii) metastatic tumor detection from histopathology images of lymph node tissues (Lymph). While VOC '07 consists of 19M instances (2000 region proposals per image), DR and Lymph have 361K and 1M instances, respectively. The results are summarized in Table 1. Our model proves to outperform various scalable MIL algorithms, as well as state-of-the-art adaptations of deep learning to weakly supervised learning.

Table 1: Bag-level average precision scores on two medical data sets.

	VOC'07	DR	Lymph
VWSGP (Ours)	83.7	0.98	0.68
VGG-S [1]	82.4	-	-
DMIL [4]	75.5	-	-
mi-FV [3]	-	0.92	0.48
e-MIL [2]	-	0.93	0.61

- [1] A. Vedaldi et al. Return of the devil in the details: Delving deep into CNNs. In *BMVC*, 2014.
- [2] G. Krümmermacher et al. Ellipsoidal multiple instance learning. In *ICML*, 2013.
- [3] X.-S. Wei et al. Scalable multi-instance learning. In *ICDM*, 2014.
- [4] J. Wu et al. Deep multiple instance learning for image classification and auto-annotation. In *CVPR*, 2015.

¹The source code of our model is publicly available under <https://github.com/melihkandemir/vwsgp>