Few-Shot Learning with Weak Supervision
Ali Ghadirzadeh*, Petra Poklukar*2, Xi Chen2, Huaxiu Yao1, Hossein Azizpour2, Mårten Björkman2, Chelsea Finn1 and Danica Kragic2
1Stanford University, 2KTH Royal Institute of Technology

INTRODUCTION

Abstract We propose a Bayesian gradient-based meta-learning algorithm that can incorporate weak labels to reduce task ambiguity and improve performance. Our approach is cast in the framework of amortized variational inference and trained by optimizing a variational lower bound. The proposed method is competitive to state-of-the-art methods and achieves significant performance gains in settings where weak labels are available.

Challenges
• Lack of sufficient information contained in the few-shot training examples to uniquely determine a task.
• Existing methods fail to cover all possible solutions because of the complexity and multi-modality of the task distribution.

Contribution
• Introducing a probabilistic meta-learning framework that enables efficient integration of weak labels to mitigate task ambiguity and to improve the coverage.

VMAML FORWARD PROCESS

VMAML consists of three neural networks:
(1) Task Encoder
conditioned on the support set \( D^S \) outputs a Gaussian distribution over the latent variable \( z \)
\[
\mu_z, \sigma_z = h(D^S; \psi_c)
\]
\[
z' \sim q(z|D^S) = \mathcal{N}(z; \mu_z, \sigma_z^2)
\]
Approximate posterior dist.
(2) Classifier
conditioned on a latent sample \( z' \), the classifier outputs softmax probabilities over the weak labels.
• Accept the sample, if classifier’s output matches the provided weak label, else draw a new sample
(3) Generative Model
conditioned on the accepted latent sample, the generative model generates a task-specific neural network initialization that is further optimized by one gradient descent following MAML [1].
\[
\theta' \sim p(\theta|z', \psi_g)
\]
Task-spec. MAML Initialization
\[
\phi = \theta' - \alpha \nabla \mathcal{L}(\theta', D^S)
\]
Adapted task-specific parameters

Fig 1: VMAML extends MAML [1] into a conditional and probabilistic meta-learning algorithm.

Training

\[
\min_{\psi_c, \psi_g, \psi_d} \sum_t \mathbb{E}_{z_t \sim q(z|D^S_t)} \left[ \mathcal{L}(\phi_t, D^Q_t) \right]
- \beta \mathbb{D}_{KL}(q(z|D^S) \| p(z))
- l_t \ast \log C(z_t; \psi_d),
\]

Fig 2: 5-shot regression task. Predictions made by VMAML (top left) and VMAML (with weak labels) (bottom left) on the same support samples (red dots).

EXPERIMENTS

Fig 3: CelebA 5-shot attribute classification task. Visualization of classification made by three different classifiers sampled from a trained VMAML.

Table 1: CelebA 5-shot attribute classification results. wo: without weak labels info*: baseline to use weak labels info: proposed method

Reference