1 Summary

1. The goal is to learn multi-state vision-based tasks from a single video of a human performing the task.
2. They argue that for task which can be decomposed into primitive skills, it is better to learn the primitive skills and how to compose them than learning how to perform the task directly.
3. Given a test-time video of human performing the task with compositional structure, they first segment it into primitive segments where each primitive segment requires the execution of one primitive skill. Each segment is then used to obtain the policy to perform the primitive skill through an adaptation procedures.

2 Background on Domain Adaptive Meta Learning

1. The MAML objective is

$$\begin{aligned} &\min_\theta \textstyle\sum_{\mathcal{T}\sim p(\mathcal{T})} \mathcal{L}\left(\theta - \alpha\nabla_\theta \mathcal{L}\left(\theta, \mathcal{D}_{\mathcal{T}}^{\mathrm{tr}}\right), \mathcal{D}_{\mathcal{T}}^{\mathrm{val}}\right) \\ &= \min_\theta \textstyle\sum_{\mathcal{T}\sim p(\mathcal{T})} \mathcal{L}\left(\phi_{\mathcal{T}}, \mathcal{D}_{\mathcal{T}}^{\mathrm{val}}\right) \end{aligned}$$

2. The inner loss fnc is referred to as the adaptation obj. The outer loss fnc is referred to as the meta-objective.
3. At test time, the adapted parameter is obtained by running:

$$\phi_{\mathcal{T}_{\text{test}}} = \theta - \alpha\nabla_\theta \mathcal{L}\left(\theta, \mathcal{D}_{\mathcal{T}_{\text{test}}}^{\mathrm{tr}}\right)$$

4. DAML aims to learn how to learn from a video of a human demo, using teleoperated demonstrations for evaluating meta-objective.
5. Since the human demonstration video does not come with ground truth label, DAML additionally learns the adaptation objective denoted as $\mathcal{L}_\psi$ along with the initial parameters θ.
6. The meta objective is a behavior cloning loss $\mathcal{L}_{\mathrm{BC}}$.
7. The DAML optimization problem is:

$$\min_{\theta,\psi} \sum_{\mathcal{T}\sim p(\mathcal{T})} \sum_{\mathbf{d}^h \in \mathcal{D}_{\mathcal{T}}^h} \sum_{\mathbf{d}^r \in \mathcal{D}_{\mathcal{T}}^r} \mathcal{L}_{\mathrm{BC}}\left(\theta - \alpha\nabla_\theta \mathcal{L}_\psi\left(\theta, \mathbf{d}^h\right), \mathbf{d}^r\right)$$

   where $\mathbf{d}^r, \mathbf{d}^h$ is the robot and human demonstration respectively.

3 Segmentation

1. Segmentation is done using a phase predictor model.
2. Given each segment, the meta-learned initial parameters θ is adapted to perform the skill demonstrated in the segment using the meta-learned adaptation objective $\mathcal{L}_u$.
3. There are two separate phase predictors, one for human data and one for robot data.
4. The phase predictor is trained to predict the temporal progress towards the completion of a primitive skill.
5. The supervision to train the phase predictor is t/T where t denotes the current timestep and T denotes the length of one demonstration video.

4 Application of DAML

1. During meta-train, DAML is used to find the initial parameters and the adaptation objective that leads to successful execution of the primitive skills given the human demonstration video.
2. This requires access to proprioceptive robot demonstration collected using teleoperation

5 Test-time algorithm

1. Given a video of human demo, it is first segmented into primitive segments using the human phase predictor by thresholding the output of the predictor.
2. Each primitive segment is used to derive the post-adaptation policy.
3. The post-adaptation policy is executed until the robot phase predictor predicts the end of the execution of the primitive skill, again by thresholding.

6 Experiments

1. The success rate is still quite low.
2. They attribute it to deficiencies in the one-shot imitation learning algorithm.

7 Question

1. Where is the data to train the robot phase predictor coming from? Is it coming from captured video of the robot executing each primitive skill after successful meta-train?