Modification notes

JUST FOLLOW THIS FOR HOW TO HANDLE INDIVIDUAL TASKS

<https://github.com/rlworkgroup/metaworld#running-a-benchmark>

ml10 = metaworld.ML10() # Construct the benchmark, sampling tasks

training\_envs = []

for name, env\_cls in ml10.train\_classes.items():

env = env\_cls()

task = random.choice([task for task in ml10.train\_tasks

if task.env\_name == name])

env.set\_task(task)

training\_envs.append(env)

python -m garage.examples.torch.maml\_trpo\_metaworld\_ml10.py

<https://github.com/rlworkgroup/garage/blob/master/src/garage/examples/torch/maml_trpo_metaworld_ml10.py>

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/examples/torch/maml_trpo_metaworld_ml10.py#L84>

So this line is strange they are setting up a sampler that only has 1 domain randomization for each of the 10 envs but they INITIALLY send out the same env (first one) to every single worker

Only when they do sample on the task sampler in maml later does this get updated

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/examples/torch/maml_trpo_metaworld_ml10.py#L47>

Trainer runs train loop

<https://github.com/rlworkgroup/garage/blob/master/src/garage/trainer.py>

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/trainer.py#L63>

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/trainer.py#L355>

NOTICE that the trainer is passed to the train call on each of the algos run

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/trainer.py#L396>

SO THIS IS WHAT TRAIN METHOD GETS RUN AND I SHOULD BE ADJUSTING

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/torch/algos/maml.py#L79>

HERE IS HOW YOU CAN GET LAST RETURN

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/torch/algos/maml.py#L95>

Want to modify this obtain samples and merge it with train

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/torch/algos/maml.py#L201>

Maml trpo just sets up TRPO as subclass of MAML

<https://github.com/rlworkgroup/garage/blob/master/src/garage/torch/algos/maml_trpo.py>

VPG def (trpo)

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/torch/algos/vpg.py#L18>

TRPO actually extends VPG

For some reason on the trpo maml class they handle trust region stuff in maml

IMPORTANT STUFF

MAML class is responsible for most of action eg calling sample and optimization

Outer maml loop

OBTAIN SAMPLES

<https://github.com/rlworkgroup/garage/blob/c56513f42be9cba2ef5426425a8ad36097e679c2/src/garage/torch/algos/maml.py#L182>

ENV UPDATE

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/torch/algos/maml.py#L204>

OH OK GOOD TO KNOW HERE IS WHERE THEY JUST SAY THERE ARE 50 ENVS PER CLASS

<https://github.com/rlworkgroup/garage/blob/f056fb8f6226c83d340c869e0d5312d61acf07f0/src/garage/experiment/task_sampler.py#L234>

Meta world task sampler NOTE NUMBER SAMPLED IS ALWAYS MULTIPLE OF ENVS AND IS ALWAYS PASSED BACK IN ORDER <https://github.com/rlworkgroup/garage/blob/f056fb8f6226c83d340c869e0d5312d61acf07f0/src/garage/experiment/task_sampler.py#L237>

THIS IS HELPFUL FOR UNDERSTANDING HOW TASKS AND ENVS AND THE SAMPLE METHOD RELATE

<https://github.com/rlworkgroup/garage/blob/f056fb8f6226c83d340c869e0d5312d61acf07f0/src/garage/experiment/task_sampler.py#L312>

OK SO EACH ENV APPEARS TO BE LOCKED TO A SEPERATE WORKER AND THIS IS RANDOMIZING DOMAINS?

<https://github.com/rlworkgroup/garage/blob/f056fb8f6226c83d340c869e0d5312d61acf07f0/src/garage/experiment/task_sampler.py#L307>

THIS IS WHERE WHEN SAMPLE IS RUN IT SELECT TASKS PER CLASS OF AN ENV

<https://github.com/rlworkgroup/garage/blob/f056fb8f6226c83d340c869e0d5312d61acf07f0/src/garage/experiment/task_sampler.py#L364>

This is the actual function wrapper that does an env update

<https://github.com/rlworkgroup/garage/blob/f056fb8f6226c83d340c869e0d5312d61acf07f0/src/garage/experiment/task_sampler.py#L367>

This generates the env update to be used by obtain episodes so when it goes out to different workers they are using the new randomized domain

DEF OF SET\_ENV\_UPDATE

<https://github.com/rlworkgroup/garage/blob/f056fb8f6226c83d340c869e0d5312d61acf07f0/src/garage/sampler/env_update.py#L65>

BUT NOTE EVEN THOUGH MAML DOES SAMPLE TO GET A WHOLE LIST OF THESE IT GETS ONE PER CLASS AND DOES ITERATE ONE ENV AT A TIME

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/torch/algos/maml.py#L201>

This is where that env update is used on the sampler for obtaining episode results. This can sample multiple rollouts at a time

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/sampler/local_sampler.py#L112>

OK THIS MAKES A SINGLE ENV UPDATE INTO A LIST IF A LIST IS NOT GIVEN

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/sampler/worker_factory.py#L70>

Def of episode batch

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/_dtypes.py#L455>

For some reason maml filters raw rewards with this(isnt exactly normal discounted reward

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/np/_functions.py#L111>

WHEN IT DOES ADAPT THIS IS WHAT IT USES FOR COMPUTE LOSS

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/torch/algos/vpg.py#L270>

So just unwrapping it works

ACTUALLY SKIP ALL THE META STUFF JUST DO THIS FOR UPDATE

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/torch/algos/vpg.py#L137>

Env dict defines env\_names used in maketasks per each task

OK IT IS A PRESET LIST OF 10 TASKS for ML10

<https://github.com/rlworkgroup/metaworld/blob/29fe5d6d95cf9ad86f63eac38db8c0aef3837994/metaworld/envs/mujoco/env_dict.py#L384>

Actual def of ML10 class

<https://github.com/rlworkgroup/metaworld/blob/9e3863d3a3b7d3f0ab6f1b759bb156abbbb27a1a/metaworld/__init__.py#L162>

Random selection of tasks

<https://github.com/rlworkgroup/metaworld/blob/9e3863d3a3b7d3f0ab6f1b759bb156abbbb27a1a/metaworld/__init__.py#L78>

THIS SAMPLE IS FOR TASKS(DOMAIN RANDOMIZATION IM PRETTY SURE)

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/trainer.py#L218>

Maml makes call to obtain episodes ON THE TRAINER which call OBTAIN SAMPLES ON ITS DEF OF SAMPLER BUT NOTE ITS DEF OF SAMPLER IS NOT A TASK SAMPLER IT IS THE RAY SAMPLER

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/torch/algos/maml.py#L196>

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/examples/torch/maml_trpo_metaworld_ml10.py#L65>

Ray obtain samples

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/sampler/ray_sampler.py#L150>

Local obtain samples

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/sampler/local_sampler.py#L134>

Maml update

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/torch/algos/maml.py#L287>

Actual inner loop for maml grad descent per task and loss calculation

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/torch/algos/maml.py#L274>

HERE IS HOW TO GET LOSS

<https://github.com/rlworkgroup/garage/blob/6461a071f0155712add1b41316003e90c9c77899/src/garage/torch/algos/maml.py#L283>

loss = self.\_inner\_algo.\_compute\_loss(\*last\_update[1:])