CS294-112 Assignment 5: Meta-Reinforcement Learning

Fan Zhang

Problem 1: Context as Task ID

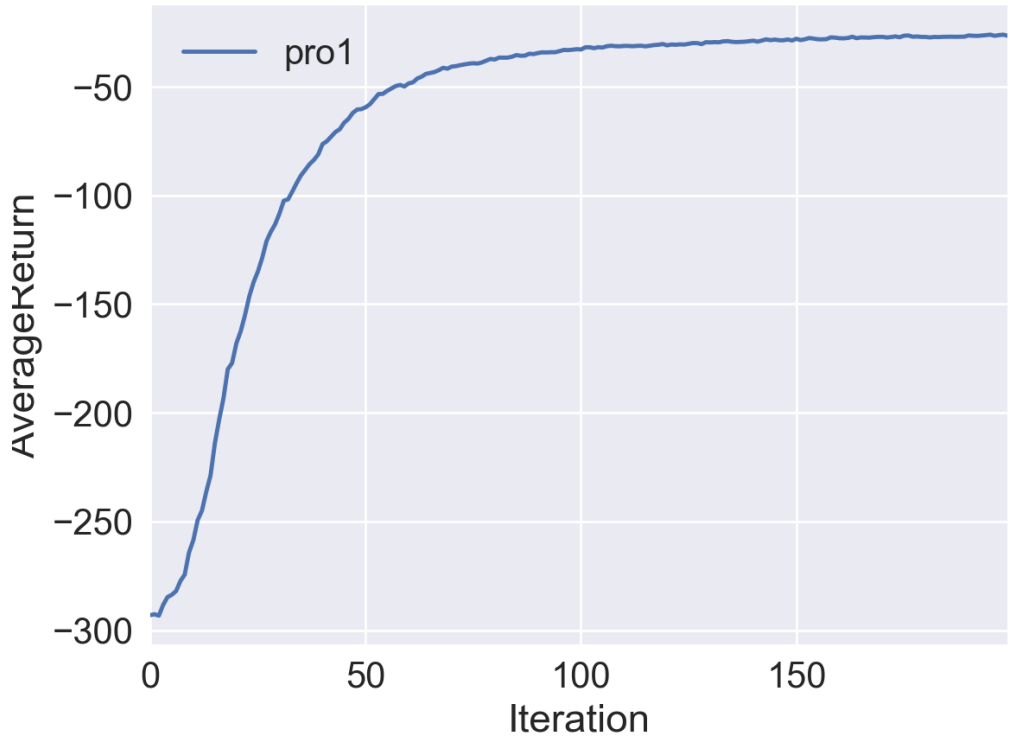


Figure 1: The average return versus training iteration in point mass environment with task ID as context. After 200 iterations, the average return reached above -50.

Problem 2: Meta-Learned Context

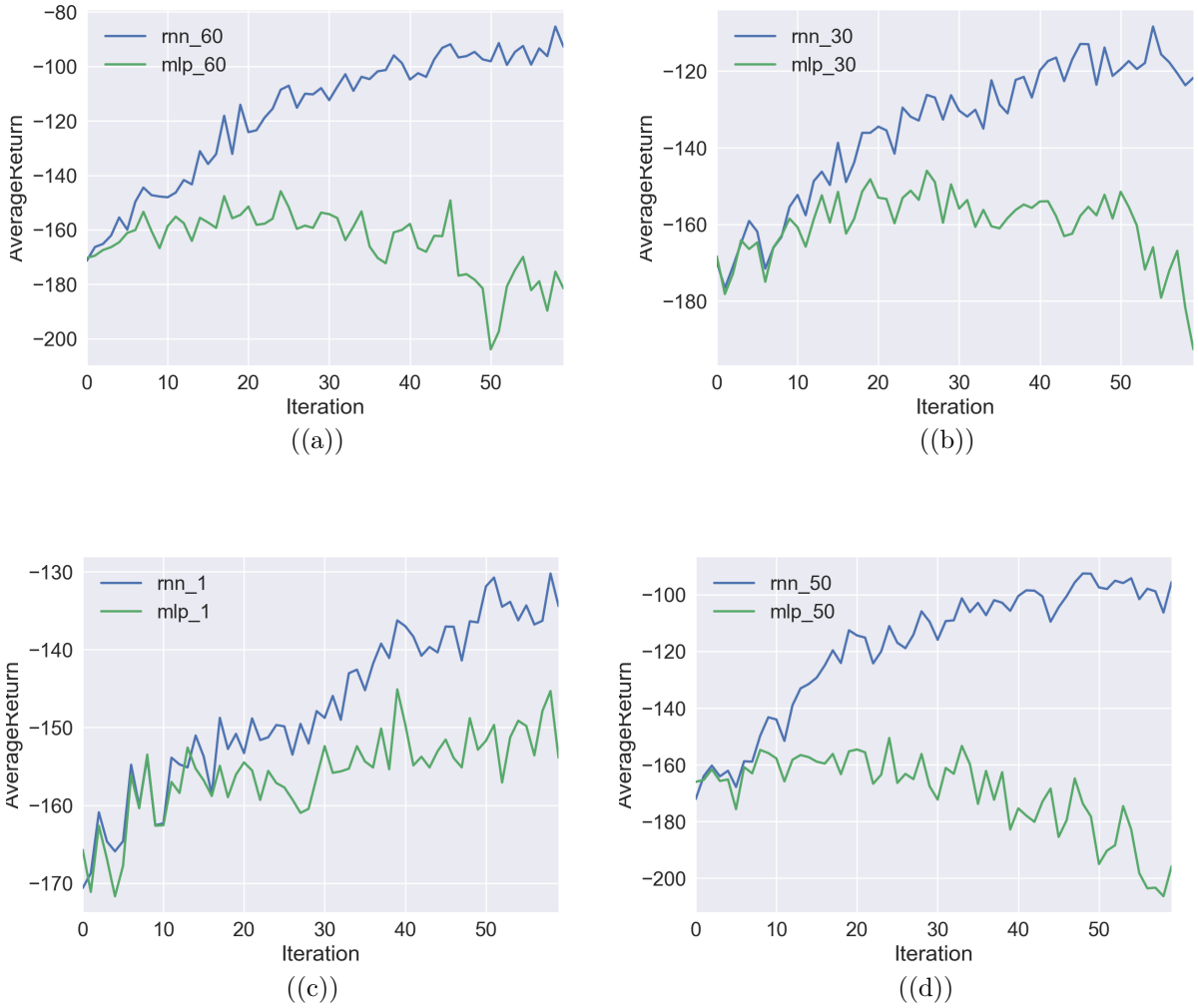
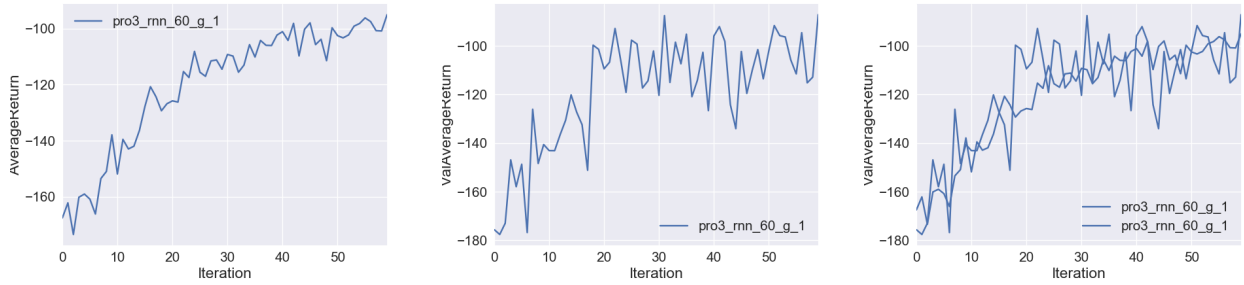


Figure 2: The performance of RNN model versus MLP model. (a) Hisory = 60 (b) History = 30 (c) History = 1 (d) History = 50.The long history performs better than short history and the RNN model performs better than MLP model.

Result Discussion: RNN model performs better than MLP model and the longer history we feed into RNN model, the better performance we will get. From my experiments, the minimum history that reaches an average return of -100 is around 50. Compared with MLP model, the input data (batch x history x observation) is processed by every timestep in RNN model while the MLP model processes a whole batch of input data. The hidden state of the GRU cell helps the model to memorize the former information that will learn to learn more with time accumulation so it will perform better with long term memory.

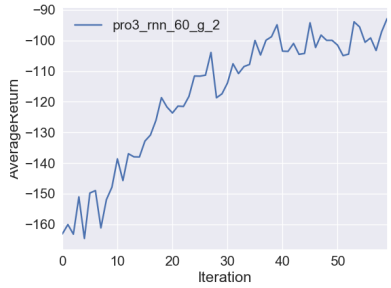
Problem 3: Generalization



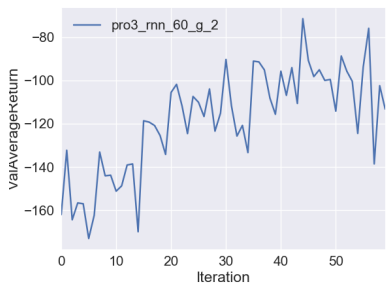
(a) Training(granularity=1)

(b) Evaluation(granularity=1)

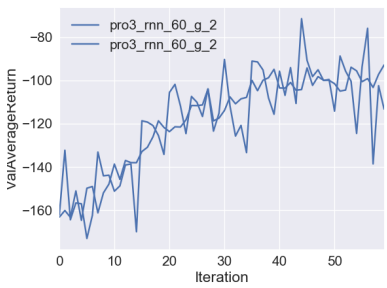
(c) Compare(granularity=1)



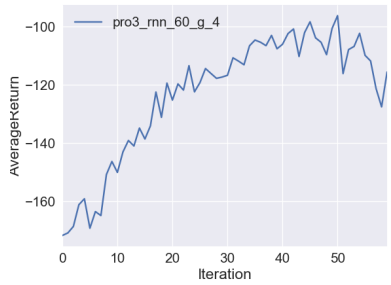
(d) Training(granularity=2)



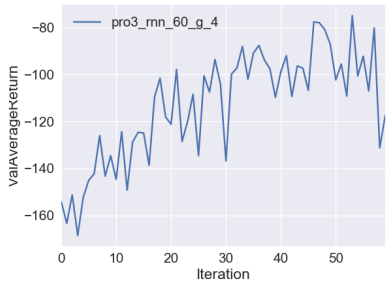
(e) Evaluation(granularity=2)



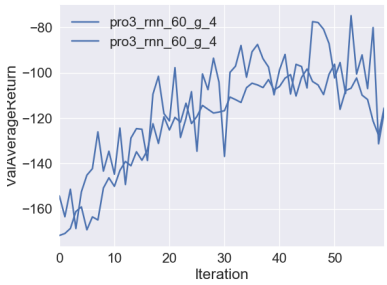
(f) Compare(granularity=2)



(g) Training(granularity=4)



(h) Evaluation(granularity=4)



(i) Compare(granularity=4)

Figure 3: The performance of the RNN policy on training goals and testing goals. The first row corresponds to granularity = 1, the second row corresponds to granularity = 2 and the third row corresponds to granularity = 4.

From the result that using RNN model as meta-reinforcement learning has a good performance on generalization. Since there is no overlap between training goals and testing goals and the model reached nearly the same result as training return during evaluation. The variance of evaluation results is larger because there are some specific testing goals that the

model didn't learning well from the training goals but overall the generalization ability is good for this model.