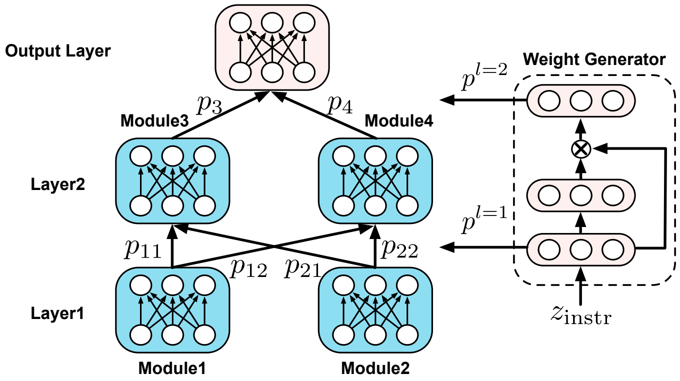

Fig. 1: Weights generation process. We visualize the weights generation process for two module layers with two modules each.

Appendix

A. Operational Details of the Weight Generation Network

We employ a pre-trained DistilBERT model to encode the natural language task instructions into fixed-length vector in $\mathbb{R}^{768}$. The encoded vector, denoted as $\mathbf{z}_{\text{instr}}$, is then used as input to a weight generation network, which outputs the weights $\alpha_1, \ldots, \alpha_k$ for the parameterized modules. This component is used throughout all Meta-World ML-10 and ML-45 training and evaluations.

We now illustrate how the weight generation network operates. For example, consider a setting with two module layers, each containing two modules. The weight generation network computes the weight probabilities $\mathbf{p}^{l=1}$ between the first and second module layers, and $\mathbf{p}^{l=2}$ between the second module layer and the output layer. A schematic diagram of this process is provided in Figure 1.

The probabilities $\mathbf{p}^{l=1} = (p_{11}, p_{12}, p_{21}, p_{22})$ are computed as follows:

$$\mathbf{p}^{l=1} = \text{softmax}\left(\mathbf{W}_d^{l=1} \cdot \text{ReLU}(\mathbf{z}_{\text{instr}})\right), \tag{1}$$

where $\mathbf{z}_{\text{instr}}$ is a D-dimensional task instruction vector, and $W_d^{l=1} \in \mathbb{R}^{4 \times D}$ is a fully connected layer that projects the instruction embedding into a probability vector over all possible connections between the modules in the first and second layers.

The probabilities $\mathbf{p}^{l=2} = (p_3, p_4)$ are computed as:

$$\mathbf{p}^{l=2} = \text{softmax}\left(\mathbf{W}_d^{l=3} \cdot \text{ReLU}\left(\mathbf{W}_d^{l=2}\mathbf{p}^{l=1} \cdot \mathbf{z}_{\text{instr}}\right)\right), \tag{2}$$

where $W_d^{l=2} \in \mathbb{R}^{D \times 4}$ maps the vectorized connection probabilities $p^{l=1}$ back into the task embedding space, and $W_d^{l=3} \in \mathbb{R}^{2 \times D}$ projects this embedding to produce a probability vector over all possible connections between the modules in the second and output layers.