**Assignment Part II - Dueling Double DeepQ Network**

**Dueling DeepQ Network**

Afbeelding met diagram, schets, ontwerp, lijn

Automatisch gegenereerde beschrijvingIn the next assignment a Dueling Double DeepQ Network is implemented. This method was first proposed by Wang et al. (2016) as an extension to the already existing DeepQ Network. They used already existing algorithms for model-free Reinforcement Learning and only changed the neural network used for training. The idea of the Dueling DeepQ network is that not one network is trained for the Q-values (see figure 1, top), but two separate neural networks are trained inside the general network for both the State-values (V(s)) and the advantage function (A(s,a)) (see figure 1, bottom). This advantage function is designed of the V and Q value based on the following formula: A(s,a) = Q(s,a) – V(s) and can be seen as the relative added value of taking a specific action in the determined state. This formula can be transformed to Q(s,a) = V(s) + A(s,a), so in the end both separate neural networks will be combined to get the Q(s,a) values. This returns a Q-values for every possible action (a) in the inputted state (s).

The authors of the article mention, however, that if the output estimator Q(s,a) is correct, this does no mean that the same holds for the V(s) and A(s,a). There are more (incorrect) combinations of these last two that can lead to a correct Q(s,a) value. We do know, however, that with a deterministic policy and the optimal action: Q(s,a) = V(s) indicating A(s,a) = 0. Therefore we have to control the algorithm to produce an A(s,a) of at maximum 0, the other values should be negative. It does, however, make the optimization more stable to not subtract the maxaA(s,a) but the mean of the A(s,a) resulting in the following finale formula to extract the Q-value: Q(s,a) = V(s) + A(s,a) – mean(Q(s,a))

The added value of training an individual part of the network for the v-value is that the value of a specific state can be seen easily without having to check the effect of the actions. This can make the model learn to not focus on too much on which action to take in case there is almost no added value of a specific action, while it does focus on which action to take in the case it is highly relevant for the outcome.

Figure 1

**Dueling DeepQ Network in code**

Wang et al. (2016) explains that difference in code with the original DeepQ Network is only in the neural network part. In figure 2 the predefined network layers of the DQN are shown, and in figure 3 the layers of the Dueling DQN. The first convolutional layer stays the same: a layer with

The next layer, however, does not consist of a single layer, but of two separate layers that we called v-values and adv-values. Both get the input from the first convolutional layer and both use the ReLU to make it non-linear. The output of the v-value part of the network is a single V-value and the output of the adv-value part is a different value for each possible action in that state.

Afbeelding met tekst, schermopname, Lettertype

Automatisch gegenereerde beschrijvingAfbeelding met tekst, schermopname, Lettertype

Automatisch gegenereerde beschrijving

Figure 3

Figure 2

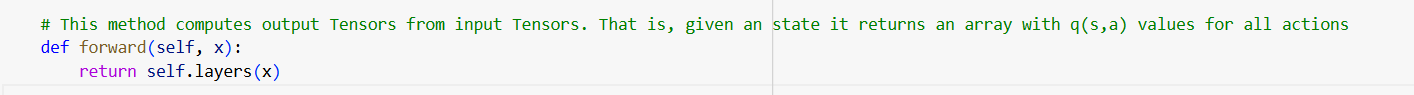
After that the adv-values and v-value should be combined and this is done in the forward function that pushes the given input state via all the layers through the network (figure 5). The output of the adv-values and v-values based on state x are collected and the mean of the adv-value is calculated. Then the formula described above is implemented which returns a q-value for all the different possible actions in the inputted state x. As the output of the neural network is still a Q value so no changes are done to the algorithm part in the rest of the code (figure 4).

Figure 4

Afbeelding met tekst, Lettertype, schermopname

Automatisch gegenereerde beschrijving

Figure 5

**Results of Dueling DeepQ Network**

**Source**

Wang, Z., Schaul, T., Hessel, M., Van Hasselt, H., Lanctot, M., & De Freitas, N. (2016). Dueling network architectures for deep reinforcement learning.