Distillation of Deep Reinforcement Learning Models using Fuzzy Inference Systems

Arne Gevaert, Jonathan Peck, and Yvan Saeys

Ghent University

Recently, significant progress has been made in the field of Deep Reinforcement Learning, with advances in a wide variety of application domains, such as arcade game playing, continuous control, and the game of Go. Many of these advances have been made using deep neural networks, which are widely regarded as *black boxes*. This means that the inner workings of a deep neural network are hard to understand, making interpretation of the learned policy by humans a difficult task. However, interpretability is a critical property of machine learning models in many application domains, including legal and medical applications. In this work[1], we use policy distillation [1] to distill the learned policy from a deep Q-network to an ANFIS controller [2]. The advantage of neuro-fuzzy controllers such as ANFIS is that they can be trained much like a neural network, but can be much more interpretable. This interpretability is not intrinsic to the model however, and specific precautions must be taken to ensure that the result is as interpretable as possible. For this reason, we extend the original policy distillation algorithm with a pre-processing and a post-processing step to maximize model interpretability [3].

Policy distillation [1] is a method to transfer knowledge from a *teacher* agent to a *student* agent. This is done by using the output Q-values of the teacher to train the student in a supervised manner. In fuzzy control, of principles from fuzzy logic are applied to create control systems (*fuzzy controllers*). These fuzzy controllers consist of a set of fuzzy IF-THEN rules $\{R_i\}$. This is a rule of the form IF x IS A_i THEN y IS B_i, where x is the input and y is the output. A_i is a fuzzy set, and B_i can be a fuzzy set or a crisp value. The output of the system is a weighted combination of the outputs $\{B_i\}$, where weights are derived from the membership degrees of x in the fuzzy sets $\{A_i\}$. In this work, we use an ANFIS controller [2]. ANFIS encodes a number of fuzzy IF-THEN rules into a neural network architecture, allowing parameters to be found using gradient descent. In our experiments, we train a DQN with a single hidden layer of 128 nodes on the well-known cart pole environment. Next, we use this network as the teacher model in a distillation setup, with ANFIS as the student model. The setup consists of three steps: in a first pre-processing step, subtractive clustering is used to initialize the ANFIS weights [3]. Next, a distillation setup as in [1] is used. Finally, the post-processing step merges similar fuzzy sets and output values by averaging out their parameters, thus simplifying the model.

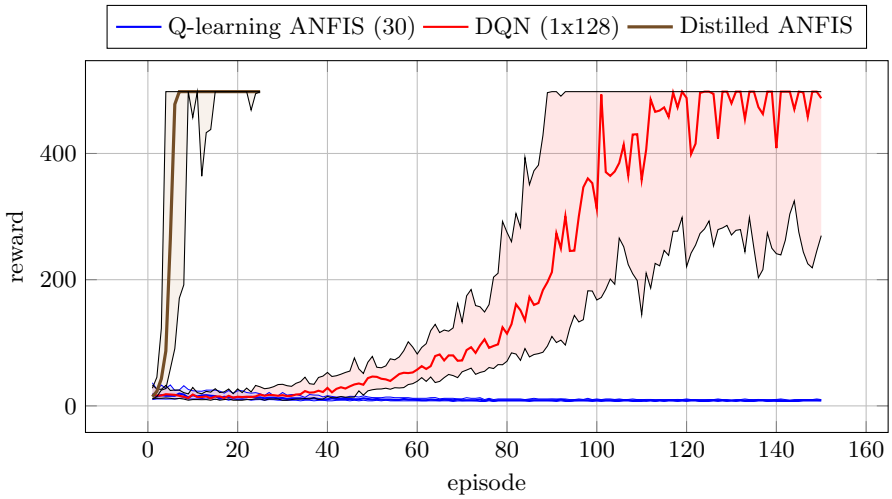

Fig. 1. DQN vs. ANFIS Q-learning vs. distillation of DQN to ANFIS on the cartpole environment. The experiment was repeated 50 times, the curves show the 20th, 50th and 80th percentiles of the reward for each episode.

Results of the experiments are shown in Figure 1. We first train ANFIS with 30 fuzzy rules using classical Q-learning (the blue curve). This agent is unable to solve the problem. Next, we train DQN (the red curve) and use our approach to distill it into ANFIS (brown curve). This distilled agent is able to solve the task in only 20 episodes every time, while using between only 3 and 6 fuzzy rules. In every run of the experiment, we were able to reduce the amount of distinct output values to 2 by iteratively merging the most similar ones, without a noticeable influence on performance. This is the minimal reasonable amount of output values, as there are 2 possible actions in the cartpole environment. We conclude that, although ANFIS is unable to learn a satisfying policy using a relatively large number of fuzzy rules, it is able to solve the problem through distillation. It is also possible to simplify the resulting rule base to a certain degree without significantly affecting performance. This opens new directions for interpretability research, including the application of more heuristic approaches to simplify the rule base after distillation, and regularization techniques to keep the rule base interpretable during distillation, or the exploration of new fuzzy architectures.

References

1. A. A. Rusu, S. G. Colmenarejo, C. Gulcehre, G. Desjardins, J. Kirkpatrick, R. Pascanu, V. Mnih, K. Kavukcuoglu, and R. Hadsell. Policy Distillation. *arXiv:1511.06295 [cs]*, November 2015.
2. J.-S.R. Jang. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(3):665–685, May-June/1993.
3. R. P. Paiva and A. Dourado. Interpretability and learning in neuro-fuzzy systems. *Fuzzy Sets and Systems*, 147:17–38, 2004.