Lent Update: Week 5

Alex Darch
Supervisor: Dr Glenn Vinnicombe

February 12, 2019

1 Neural Network Adaptations

Neural network architecture. The input to the neural network is a 19 x 19 x 17 image stack comprising 17 binary feature planes. Eight feature planes, Xt, consist of binary values indicating the presence of the current player's stones (= X 1 t i if intersection i contains a stone of the player's colour at time-step t; 0 if the intersection is empty, contains an opponent stone, or if t ¡ 0). A further 8 feature planes, Yt, represent the corresponding features for the opponent's stones. The final feature plane, C, represents the colour to play, and has a constant value of either 1 if black is to play or 0 if white is to play. These planes are concatenated together to give input features st = [Xt, Yt, Xt1, Yt1,..., Xt7, Yt7, C]. History features Xt, Yt are necessary, because Go is not fully observable solely from the current stones, as repetitions are forbidden; similarly, the colour feature C is necessary, because the komi is not observable. The input features st are processed by a residual tower that consists of a single convolutional block followed by either 19 or 39 residual blocks4. The convolutional block applies the following modules: (1) A convolution of 256 filters of kernel size 3 3 with stride 1 (2) Batch normalization18 (3) A rectifier nonlinearity Each residual block applies the following modules sequentially to its input: (1) A convolution of 256 filters of kernel size 3 3 with stride 1 (2) Batch normalization (3) A rectifier nonlinearity (4) A convolution of 256 filters of kernel size 3 3 with stride 1 (5) Batch normalization (6) A skip connection that adds the input to the block (7) A rectifier nonlinearity The output of the residual tower is passed into two separate 'heads' for computing the policy and value. The policy head applies the following modules: (1) A convolution of 2 filters of kernel size 1 1 with stride 1 (2) Batch normalization (3) A rectifier nonlinearity (4) A fully connected linear layer that outputs a vector of size 192 + 1 = 362, corresponding to logit probabilities for all intersections and the pass move The value head applies the following modules: (1) A convolution of 1 filter of kernel size 1 1 with stride 1 (2) Batch normalization (3) A rectifier nonlinearity (4) A fully connected linear layer to a hidden layer of size 256 (5) A rectifier nonlinearity (6) A fully connected linear layer to a scalar (7) A tanh nonlinearity outputting a scalar in the range [1, 1] The overall network depth, in the 20- or 40-block network, is 39 or 79 parameterized layers, respectively, for the residual tower, plus an additional 2 layers for the policy head and 3 layers for the value head. We note that a different variant of residual networks was simultaneously applied to computer Go33 and achieved an amateur dan-level performance; however, this was restricted to a single-headed policy network trained solely by supervised learning.

2 Motion History Images

Temporal Convolutional Neural Networks https://www.cs.jhu.edu/ areiter/JHU/Publications$_f iles/ColinLea_T CN_C amer$