Dynamic Paired Comparison Prediction Using Modified Elo Ratings

Grijalva, Julien (School: Northern Utah Academy for Math, Engineering and Science)

The performance of a modified Elo update rule was tested against the original Elo update rule. The Elo update rule can be interpreted as an application of gradient descent using a certain loss function. Research suggests that different optimizers and loss functions may improve performance. I tested the binary cross-entropy loss function and the Adam optimizer (Kingma and Ba, 2014), with hyperparameters estimated using a meta-optimization. This optimization was parallelized on 150,000 GPU threads using Python and CUDA. Rating update rules were evaluated using data generated from a simulator also running on the GPU with one thread for each simulation (750 simulations, 75,000,000 total games in the simulations). Although the accuracy measures were quantitatively similar, I found that the best performing Elo update rule used cross-entropy loss with gradient descent, having 67.03% accuracy. Hypothesis tests were conducted by treating each game as an individual of the population; in all treatments, the null hypothesis was rejected ($p<0.0002$). The results show that Elo ratings can be improved by using either cross-entropy or Adam.

Awards Won: