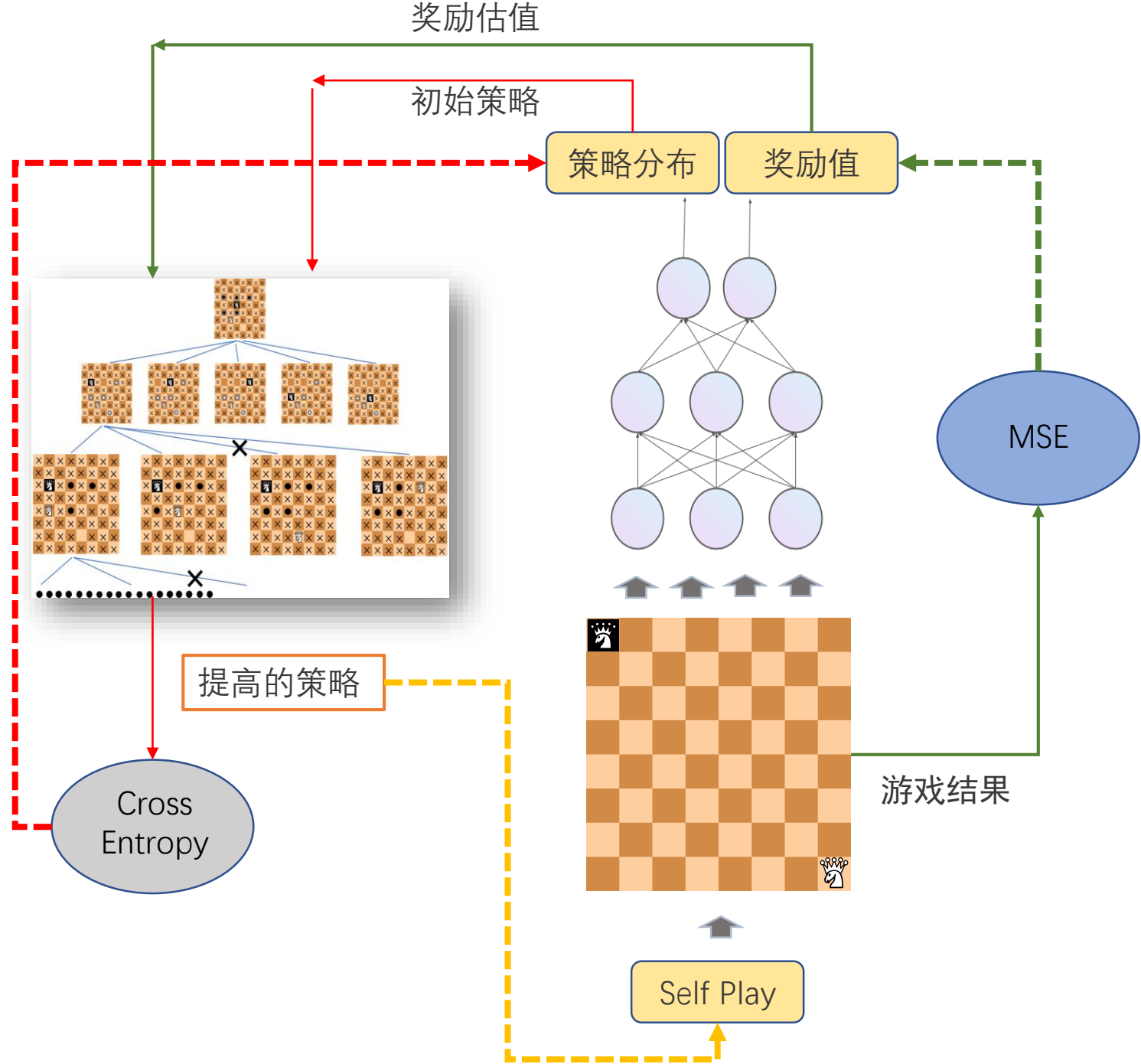
奖励估值
初始策略
策略分布
奖励值
MSE
提高的策略
Cross Entropy
游戏结果
Self Play