DRON-MoE (Mixture of Experts)

$$Q(s,a) = \sum_o Prob(o|s)Q(s,a,o) \quad (s:\text{ state} \quad a:\text{ action} \quad o:\text{ opponent action})$$

State features ϕ_t^s

ϕ_t^o Opponent features

k: number of latent opponent types (hyperparameter)

h^s

h^o

$Q_1 \cdots Q_k$

w_i

Experts

Gating

$\otimes$

$Q^{\cdot|\pi^o}(s_t, a_t)$