QMIX:
Monotonic Value Function Factorisation for Deep Multi-Agent Reinforcement Learning

Yanjie Ze @SJTU CS

Apr 6, 2021

Motivation: Improve VDN

In Value-Decomposition Network:

1. $$Q((h^1, h^2, ..., h^d), (a^1, a^2, ..., a^d)) \approx \sum_{i=1}^{d} \tilde{Q}_i(h^i, a^i)$$

2. $$Q^\pi(\mathbf{s}, \mathbf{a}) =: \bar{Q}_1^\pi(\mathbf{s}, \mathbf{a}) + \bar{Q}_2^\pi(\mathbf{s}, \mathbf{a}) \approx \tilde{Q}_1^\pi(h^1, a^1) + \tilde{Q}_2^\pi(h^2, a^2)$$

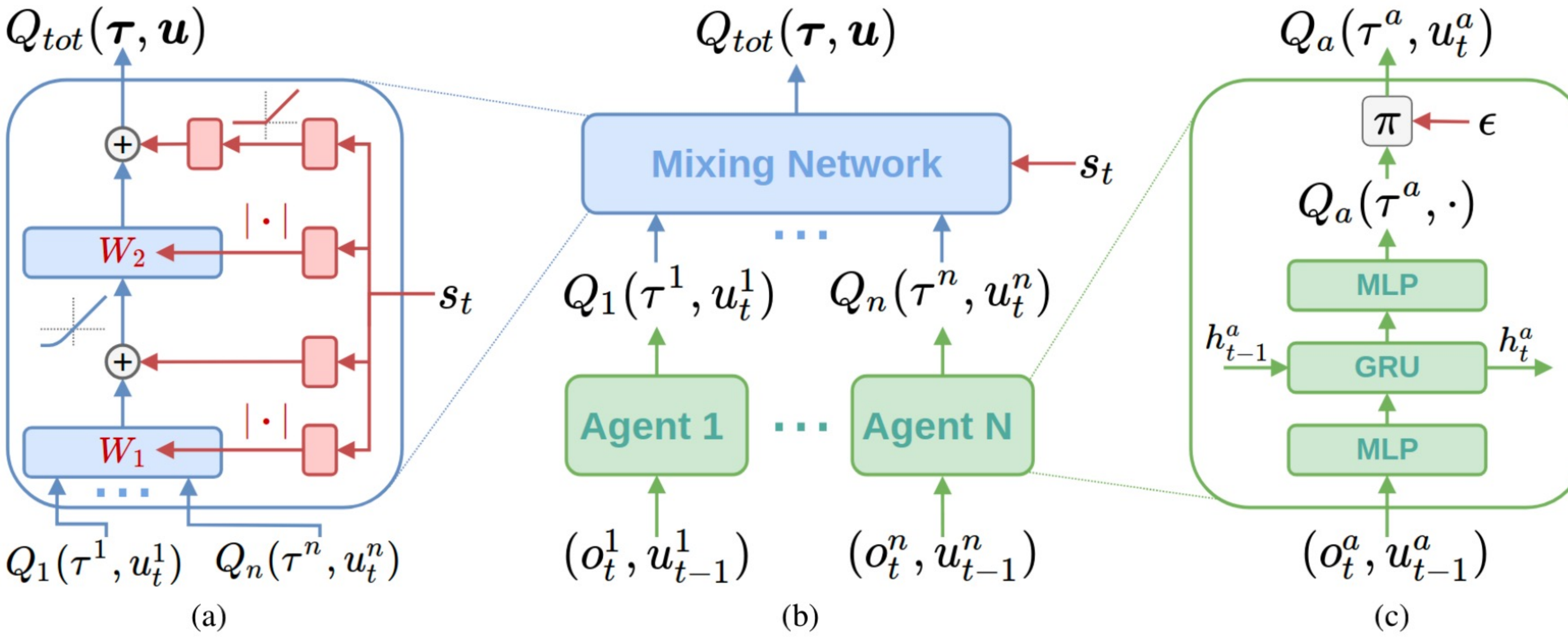
Motivation: Improve VDN

In QMIX：

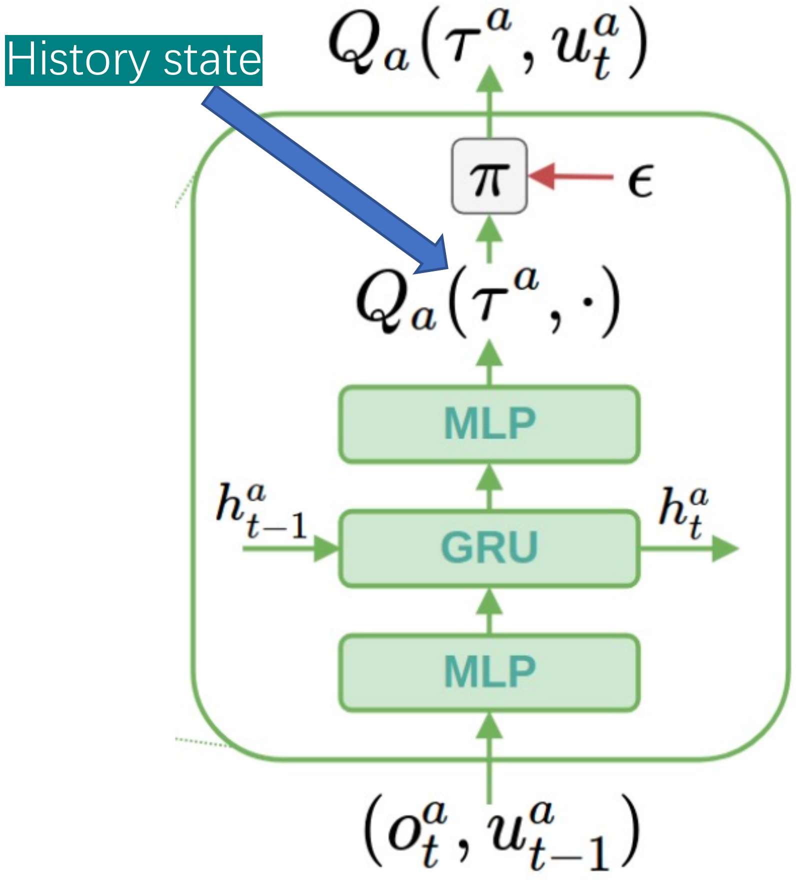
1.
$$\operatorname*{argmax}_{\mathbf{u}} Q_{tot}(\boldsymbol{\tau}, \mathbf{u}) = \begin{pmatrix} \operatorname{argmax}_{u^1} Q_1(\tau^1, u^1) \\ \vdots \\ \operatorname{argmax}_{u^n} Q_n(\tau^n, u^n) \end{pmatrix}$$

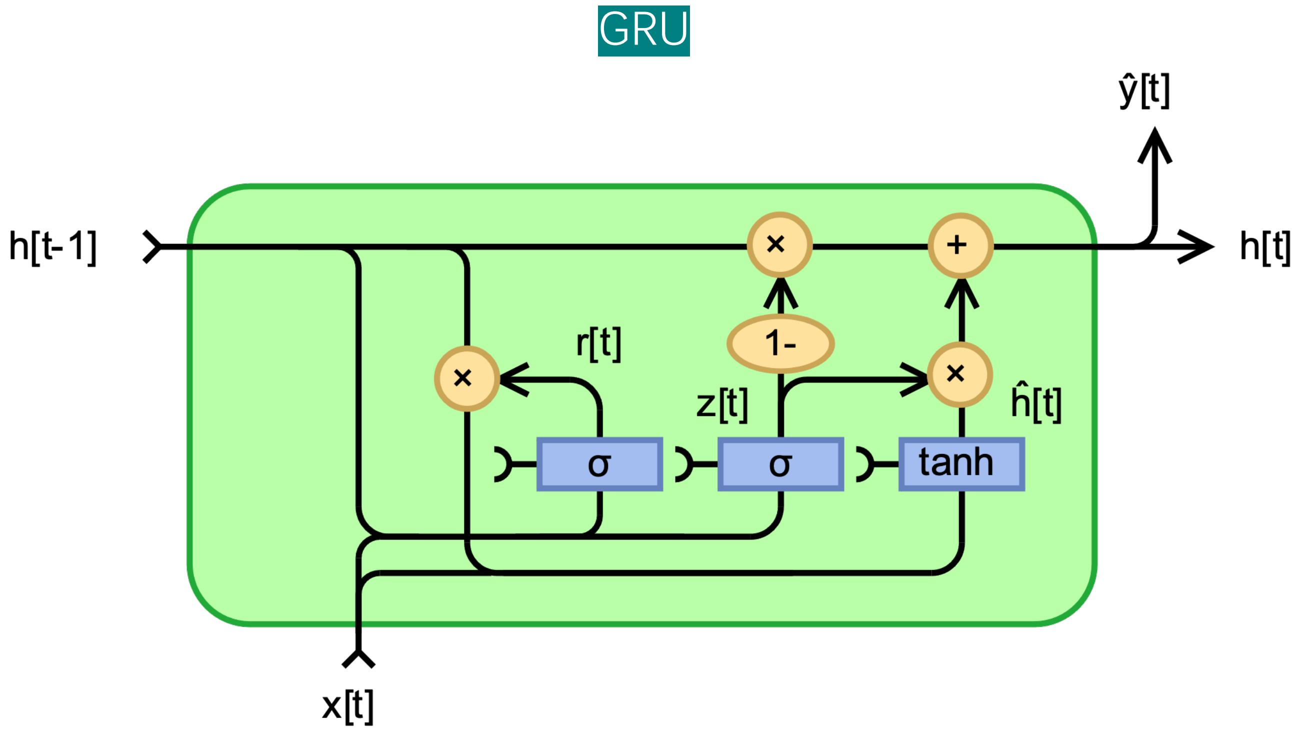
2.
$$\frac{\partial Q_{tot}}{\partial Q_a} \geq 0, \ \forall a \in A.$$

QMIX: Overall Structure

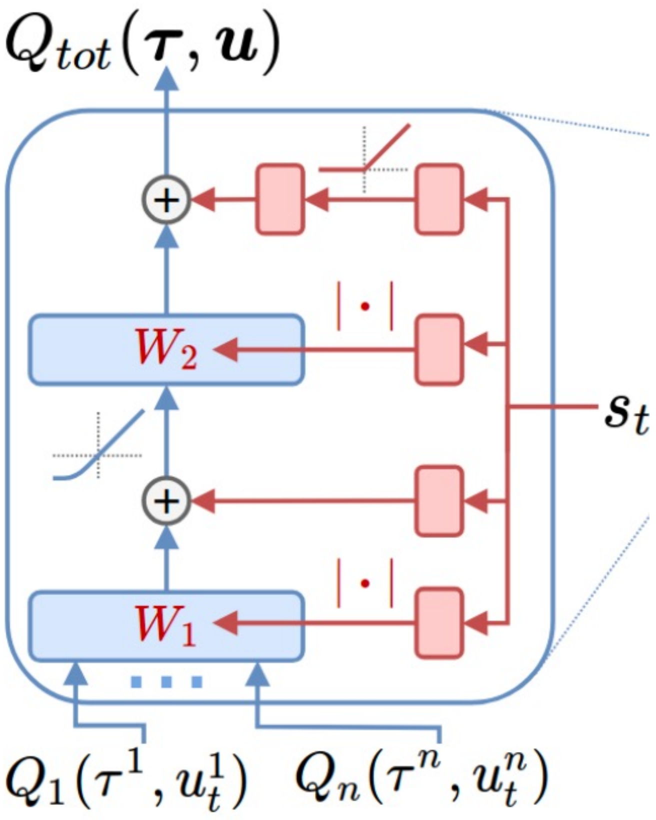


QMIX: Agent Network(DRQN)





QMIX: Mixing Network and Hypernetwork



Implementation Detail:

1. The weights of the mixing network are produced by separate hypernetworks.

2. Each hypernetwork consists of a single linear layer, followed by an absolute activation function.

3. The biases are produced in the same manner but are not restricted to being non-negative. The final bias is produced by a 2 layer hypernetwork with a ReLU non-linearity

4. s_t is full state.

QMIX: Loss Function

$$\mathcal{L}(\theta) = \sum_{i=1}^{b} \left[\left(y_i^{tot} - Q_{tot}(\boldsymbol{\tau}, \mathbf{u}, s; \theta) \right)^2 \right]$$

Where: $y^{tot} = r + \gamma \max_{\mathbf{u}'} Q_{tot}(\boldsymbol{\tau}', \mathbf{u}', s'; \theta^-)$