中国科学院大学
University of Chinese Academy of Sciences

Multi-Agent Uncertainty Sharing for Cooperative MARL

Guangkai Yang[1,2], Hao Chen[1,2], Junge Zhang[1,2†], Qiyue Yin[1,2], Kaiqi Huang[1,2,3]

[1]*CRISE, Institute of Automation, Chinese Academy of Sciences*

[2]*School of Artificial Intelligence, University of Chinese Academy of Sciences*

[3]*CAS Center for Excellence in Brain Science and Intelligence Technology*

Beijing 100190, P.R.China

{yangguangkai2019, chenhao2019}@ia.ac.cn, {jgzhang, qyyin, kqhuang}@nlpr.ia.ac.cn

University of Chinese Academy of Sciences

Contents

1 Motivation

- Core Problem
  - What will the partial observability cause?
    - Uncertainty of agents for action value estimations
  - What's worse in multi-agent reinforcement learning?
    - The uncertainty of each agent for a same action value can be different, which harms the cooperative exploration for the joint action space.
- The existing works with CTDE neglect the uncertainty of agents in MARL.

2 Method

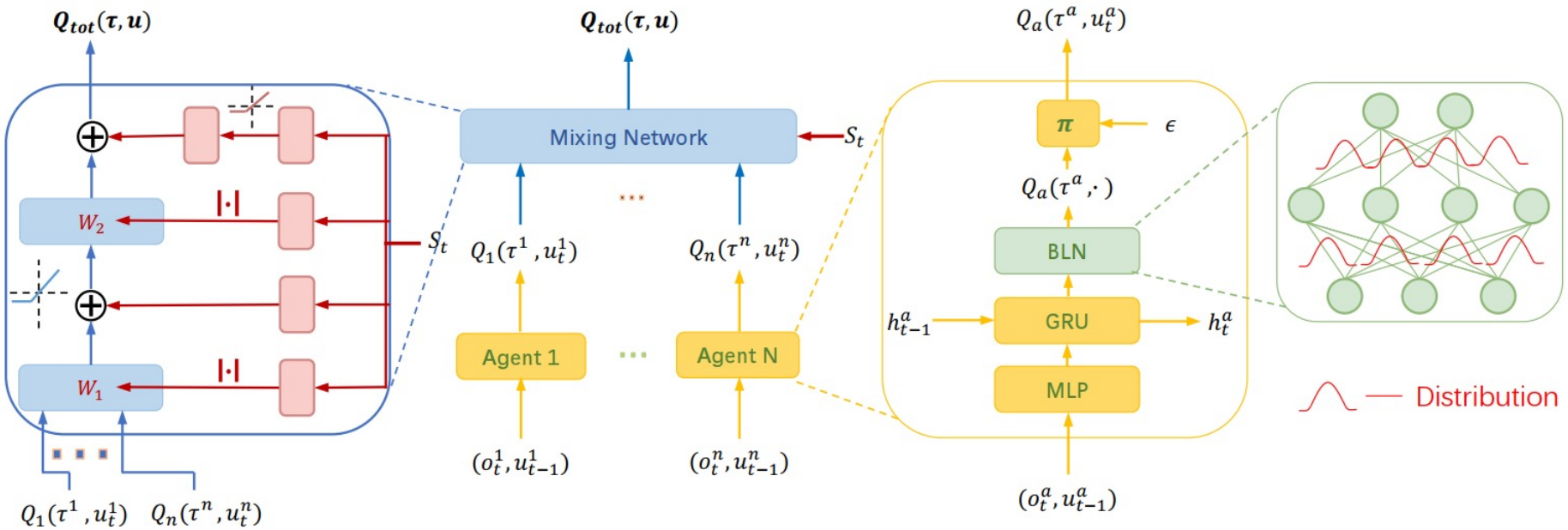
- How can we do to deal with the “uncertainty” in MARL?
  - Bayesian neural network to quantify the uncertainty.
    - Combine with the Thompson Sampling to select actions for each agent.
  - Enough?
  - No!
    - Besides, we impose the uncertainty sharing mechanism to align all the agents’ uncertainties to improve cooperative exploration and stabilize the training.

2 Method

1.Thompson Sampling

$$P(u_t^a = u^a | \tau_t^a, \theta_t) = \int_{\mathbf{w}} \mathbf{1}\left\{Q_a(\tau_t^a, u^a; \mathbf{w}) > Q_a\left(\tau_t^a, u^{a'}; \mathbf{w}\right), \forall u^{a'} \neq u^a\right\} \cdot dq\left(\mathbf{w}|\theta_t\right)$$

3.Loss function $\mathcal{L}_{MAUS} = \mathcal{L}_{TD}(\phi) + \lambda_{kl} \cdot \mathrm{KL}[q(\mathbf{w}|\theta)\|p(\mathbf{w})]$



2.Target value generation

Multi-Agent Monte Carl，MAC：$y: y = r + \gamma \max_{\mathbf{u}'} Q_{tot}(s', \mathbf{u}'; \{\tilde{\mathbf{w}}_i\}_{i=1}^k)$

Maximum A Posterior，MAP：$y: y = r + \gamma \max_{\mathbf{u}'} Q_{tot}(s', \mathbf{u}'; \tilde{\mu})$

3 Experiments

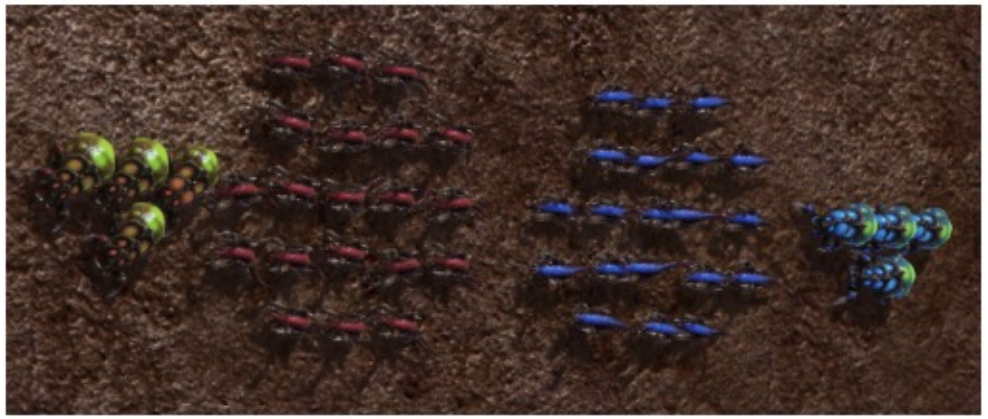
MAUS

Testing environment：the StarCraft Multi-Agent Challenge (SMAC)

Metric：**test winning rate**

Enemy：Built in heuristic rules vs Ally：The learned policy network

3s5z



bane_vs_bane

.....

3 Experiments

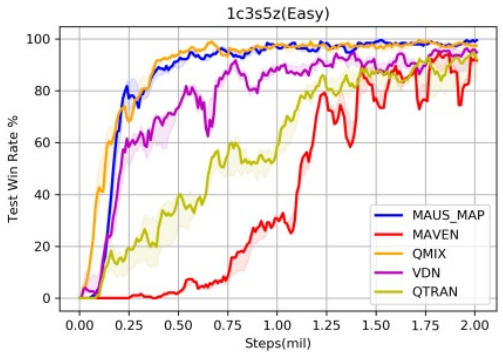
Selected Maps

SELECTED MAPS IN SMAC

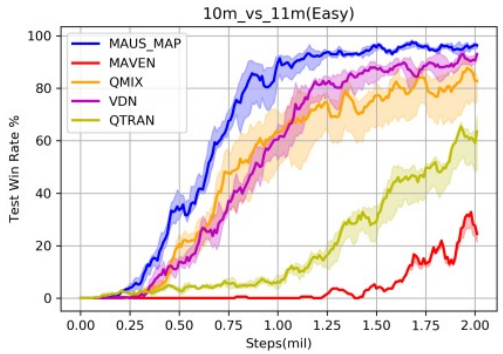
Name	Ally Units	Enemy Units	Type	Difficulty
1c3s5z	1 Colossus 3 Stalkers 5 Zealots	1 Colossus 3 Stalkers 5 Zealots	Heterogeneous, Symmetric Focus Firing, Macro Tactics	Easy
10m_vs_11m	10 Marines	11 Marines	Homogeneous, Asymmetric Focus Firing, Macro Tactics	Easy
bane_vs_bane	20 Zerglings 4 Banelings	20 Zerglings 4 Banelings	Heterogeneous, Symmetric Large action space, Macro Tactics	Hard
3s_vs_5z	3 Stalkers	5 Zealots	Homogeneous, Asymmetric Macro Tactics	Hard
2c_vs_64zg	2 Colossi	64 Zerglings	Homogeneous, Asymmetric Large action space	Hard
MMM2	1 Medivac 2 Marauders 7 Marines	1 Medivac 3 Marauders 8 Marines	Heterogeneous, Asymmetric Macro Tactics	Super Hard

3 Experiments

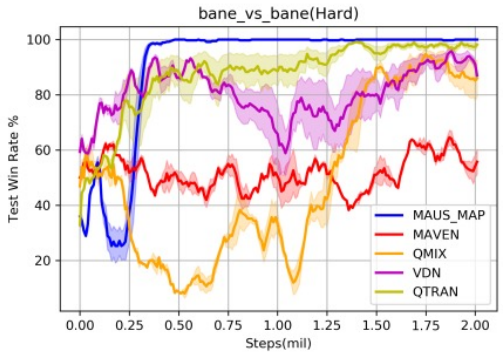
Performance



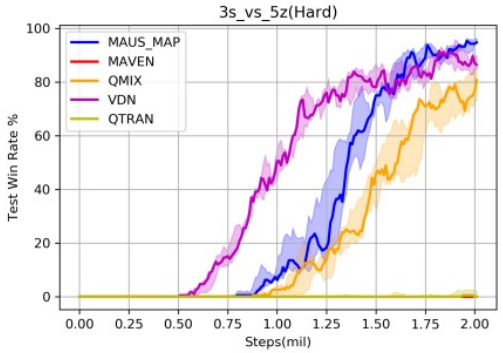
(a) 1c3s5z



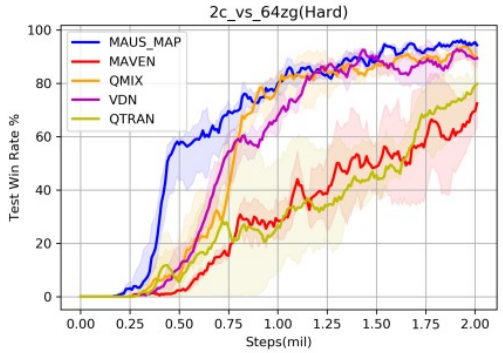
(b) 10m_vs_11m



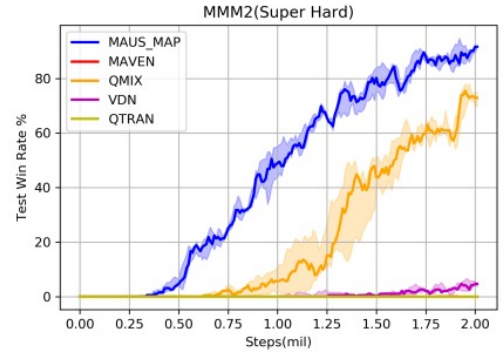
(c) bane_vs_bane



(d) 3s_vs_5z



(e) 2c_vs_64zg



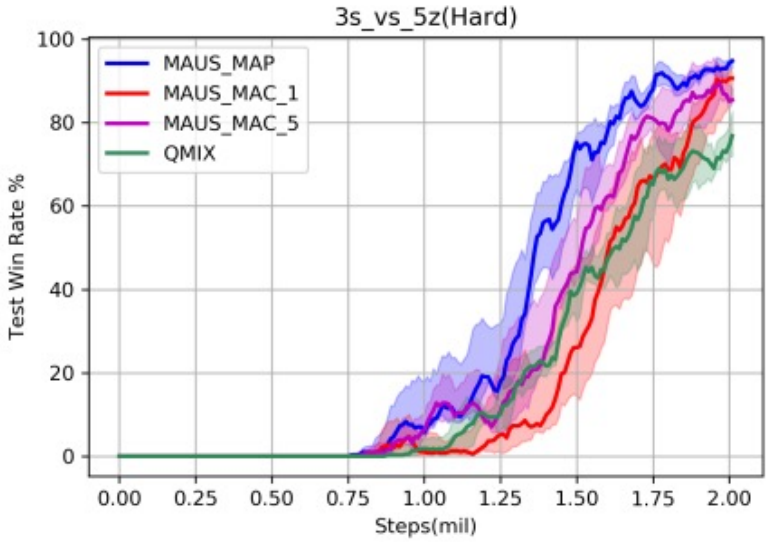
(f) MMM2

MEDIAN PERFORMANCE OF THE TEST WIN PERCENTAGE (%) IN DIFFERENT SCENARIOS.

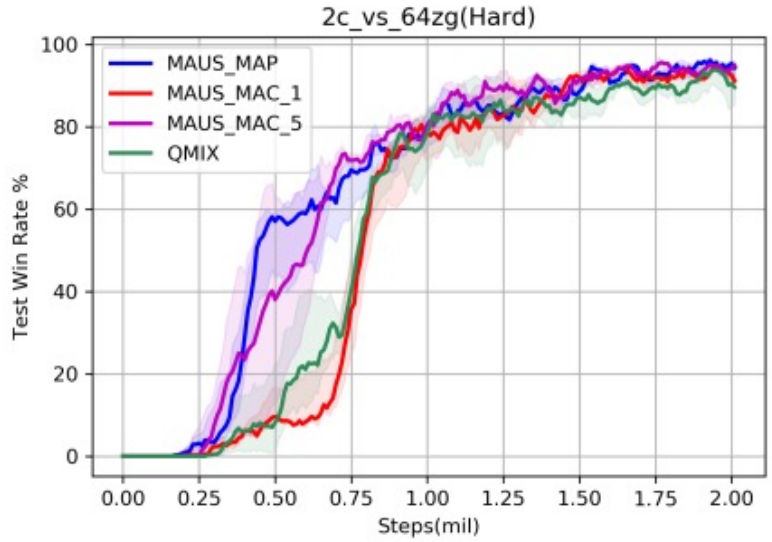
Scenario	MAUS_MAP	MAVEN	QMIX	VDN	QTRAN
1c3s5z	**100**	94	98	94	91
10m_vs_11m	**97**	22	84	92	70
bane_vs_bane	**100**	56	84	67	99
3s_vs_5z	**97**	0	84	93	0
2c_vs_64zg	**93**	81	89	90	80
MMM2	**90**	0	75	3	0

3 Experiments

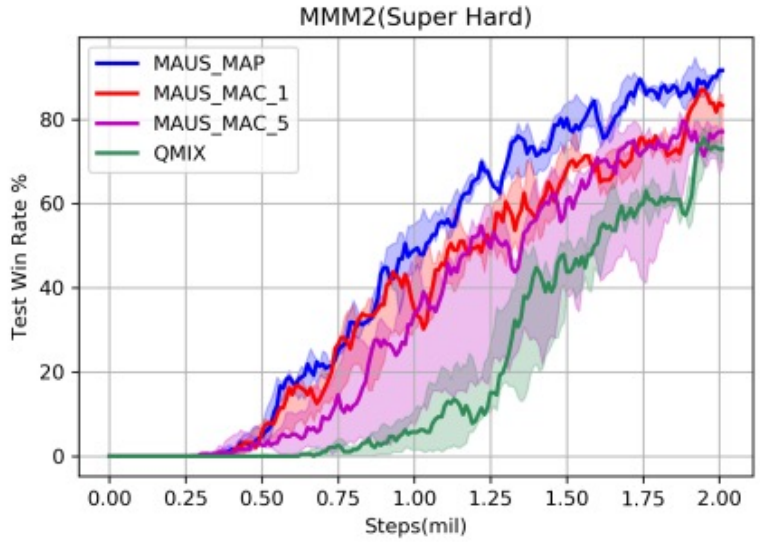
Ablation study: MAUS_MAP vs MAUS_MAC



(a) 3s_vs_5z



(b) 2c_vs_64zg



(c) MMM2

4 Conclusion

- What we do?
  - A simple yet effective method called MAUS
    - Bayesian neural network to quantify the uncertainty
    - Imposing the uncertainty sharing mechanism

- Future?
  - Uncertainty for credit assignment in MARL
  - Breaking the distribution type constraint by the Bayesian hypernet


中国科学院大学
University of Chinese Academy of Sciences
Thank you all for listening!
11/11