RACA: Relation-Aware Credit Assignment for Ad-Hoc Cooperation in Multi-Agent Deep Reinforcement Learning

Hao Chen[1,2], Guangkai Yang[1,2], Junge Zhang[1,2†], Qiyue Yin[1,2], Kaiqi Huang[1,2,3]
[1] *School of Artificial Intelligence, University of Chinese Academy of Sciences*
[2] *CRISE, Institute of Automation, Chinese Academy of Sciences*
[3] *CAS Center for Excellence in Brain Science and Intelligence Technology*
Beijing 100049, P.R.China
{chenhao2019, yangguangkai2019}@ia.ac.cn, {jgzhang, qyyin, kqhuang}@nlpr.ia.ac.cn

CONTENTS

CONTENTS

PART 01 **Introduction**

PART 02 **Methodology**

PART 03 **Experiments**

PART 04 **Conclusion**

1. Introduction

One key Problem in MARL

- **Ad-Hoc Cooperation:** Adapt to new configurations of teammates at test time

Current Solutions

- Do not explicitly utilize the relationship between agents
- Can not adapt to different sizes of inputs

Utilizing the relationship between agents is vital to multiagent cooperation.

2. Methodology

Graph-Based Relation Encoder

$$e_{uv} = \begin{cases} 1, & \text{if } u \in \mathcal{N}(v) \text{ or } v \in \mathcal{N}(u) \\ 0, & \text{otherwise} \end{cases}$$

$$H^{l+1} = \sigma\left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^l W^l\right)$$

$$H_{input}^{l+1} = CONCAT\left[H_{output}^{l}; X\right]$$

Observation Abstraction via Attention Mechanism

$$o_{own}^{i},\ o_{var}^{i},\ o_{var}^{i}$$

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V$$

$$Q_i\left(\tau^i, u^i\right) = f_i\left(g_i\left(\text{Attention}(Q, K, V)\right), o_{inv}^{i}\right)$$

Loss Function

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

$$y^{tot} = r + \gamma \max_{\boldsymbol{u}'} Q_{tot}(\boldsymbol{\tau}', \boldsymbol{u}' | \theta^-)$$

2. Methodology

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

RACA: Relation-Aware Credit Assignment for Ad-Hoc Cooperation in Multi-Agent Deep Reinforcement Learning

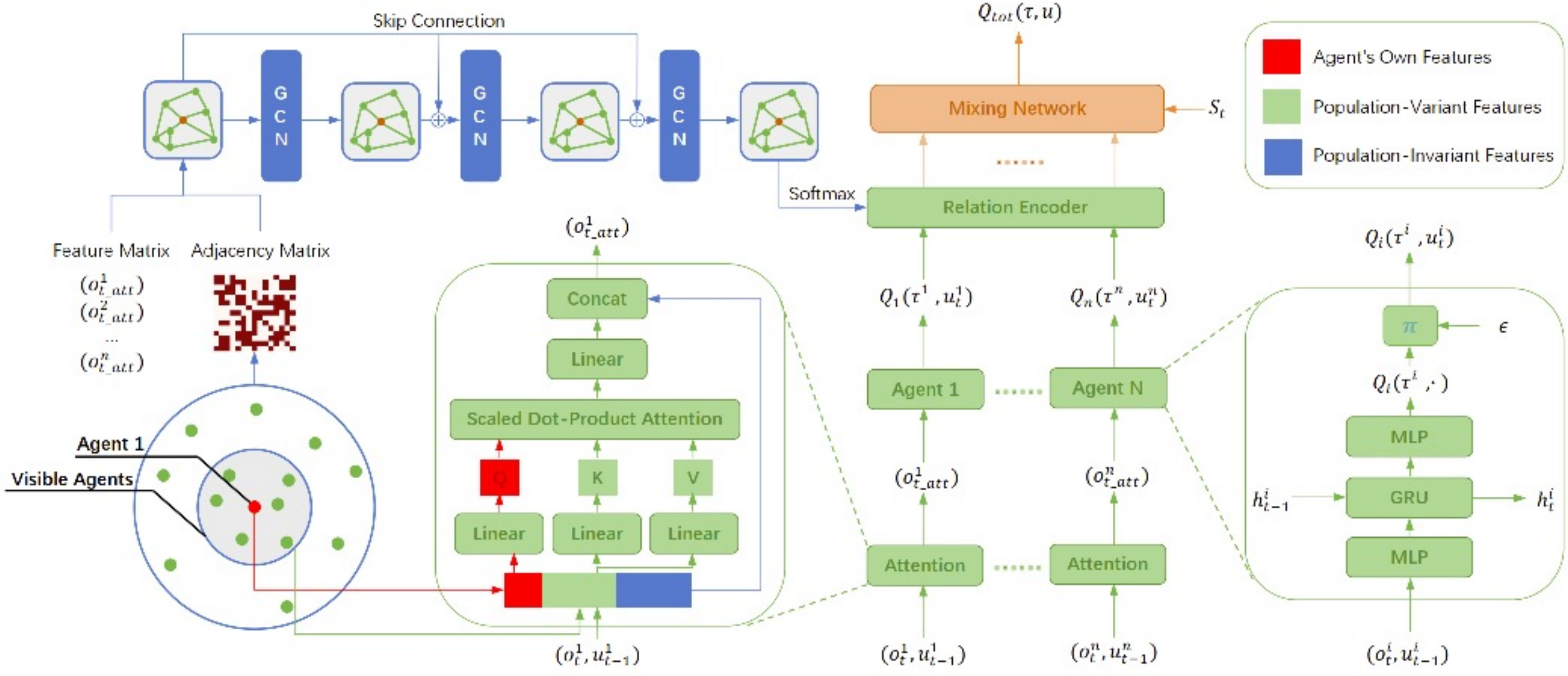


Fig. 1. The overall architecture of RACA. The right part is a deep recurrent Q-network, which takes the local action-observation history as input and outputs the individual action-values. The left part is the graph-based relation encoder and the attention-based observation abstraction module.

3 Experiments

Experiment Setup

TABLE I
MAPS IN DIFFERENT SCENARIOS.

Map Name	Ally Units	Enemy Units	Map Type	Critical Challenges	Map Difficulty
5m	5 Marines	5 Marines	Homogeneous, Symmetric	Focus Fire	Easy
5ma	5 Marauders	5 Marauders	Homogeneous, Symmetric	Focus Fire	Easy
1c3s5z	1 Colossus 3 Stalkers 5 Zealots	1 Colossus 3 Stalkers 5 Zealots	Heterogeneous, Symmetric	Focus Fire, Macro Tactics	Easy
XsYz	X Stalkers Y Zealots	X Stalkers Y Zealots	Heterogeneous, Symmetric	Focus Fire, Macro Tactics	Easy
Xm_vs_Ym	X Marines	Y Marines	Homogeneous, Asymmetric	Focus Fire	Easy(Y=X), Hard(Y=X+1)
MMM2	1 Medivac 2 Marauders 7 Marines	1 Medivac 3 Marauder 8 Marines	Heterogeneous, Asymmetric	Focus Fire, Macro Tactics	Super Hard

TABLE II
BASELINE AND ABLATION ALGORITHMS.

	Alg.	Description
Related Works	IQL VDN QMIX QTRAN	Independent Q-learning Additivity constraint Monotonicity constraint Constraint-free
Abla-tions	IQL_Attn VDN_Attn QMIX_Attn QMIX_Gcn QTRAN_Attn	IQL with attention VDN with attention QMIX with attention QMIX with gcn module QTRAN with attention

3 Experiments

Main results

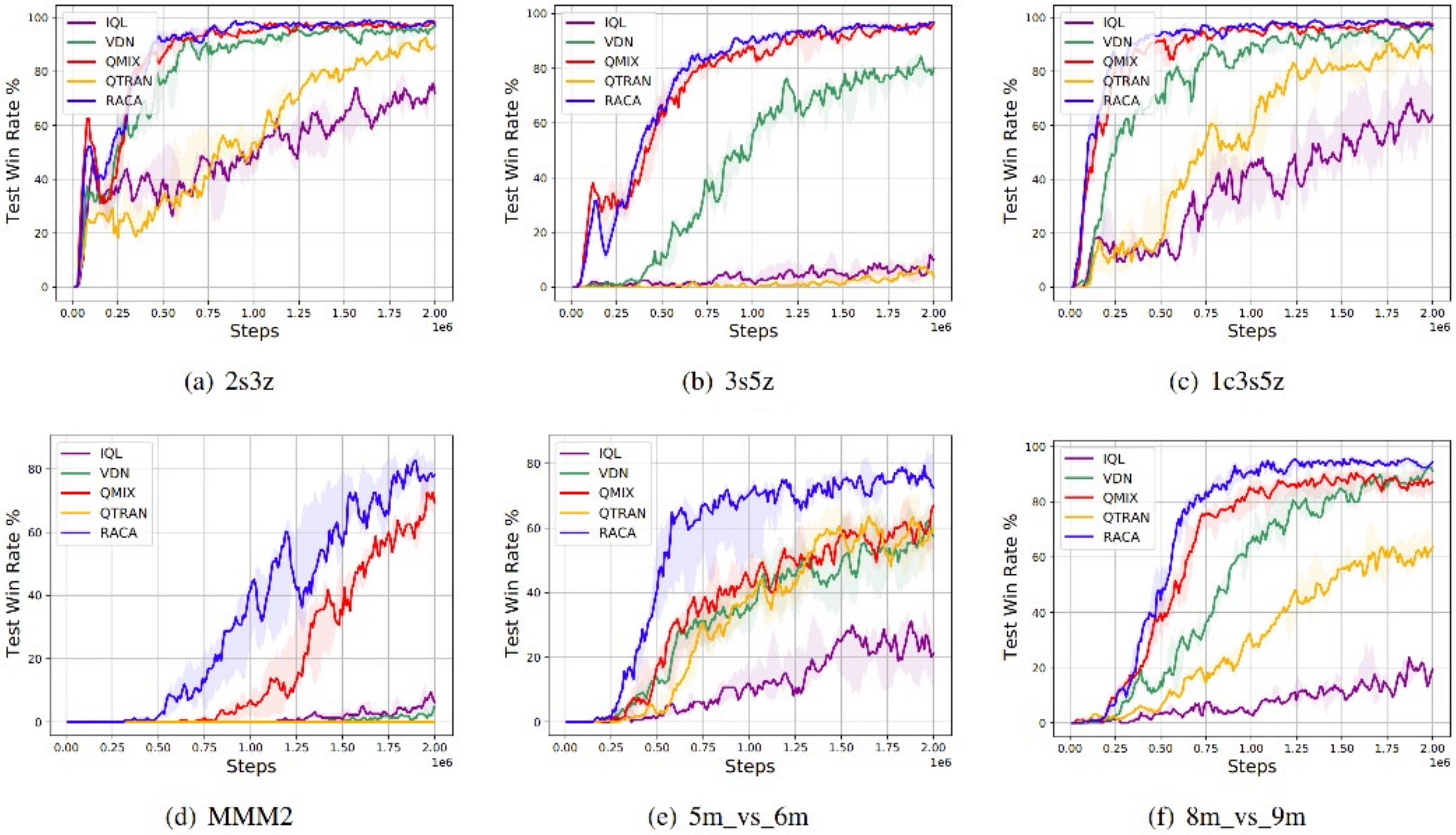
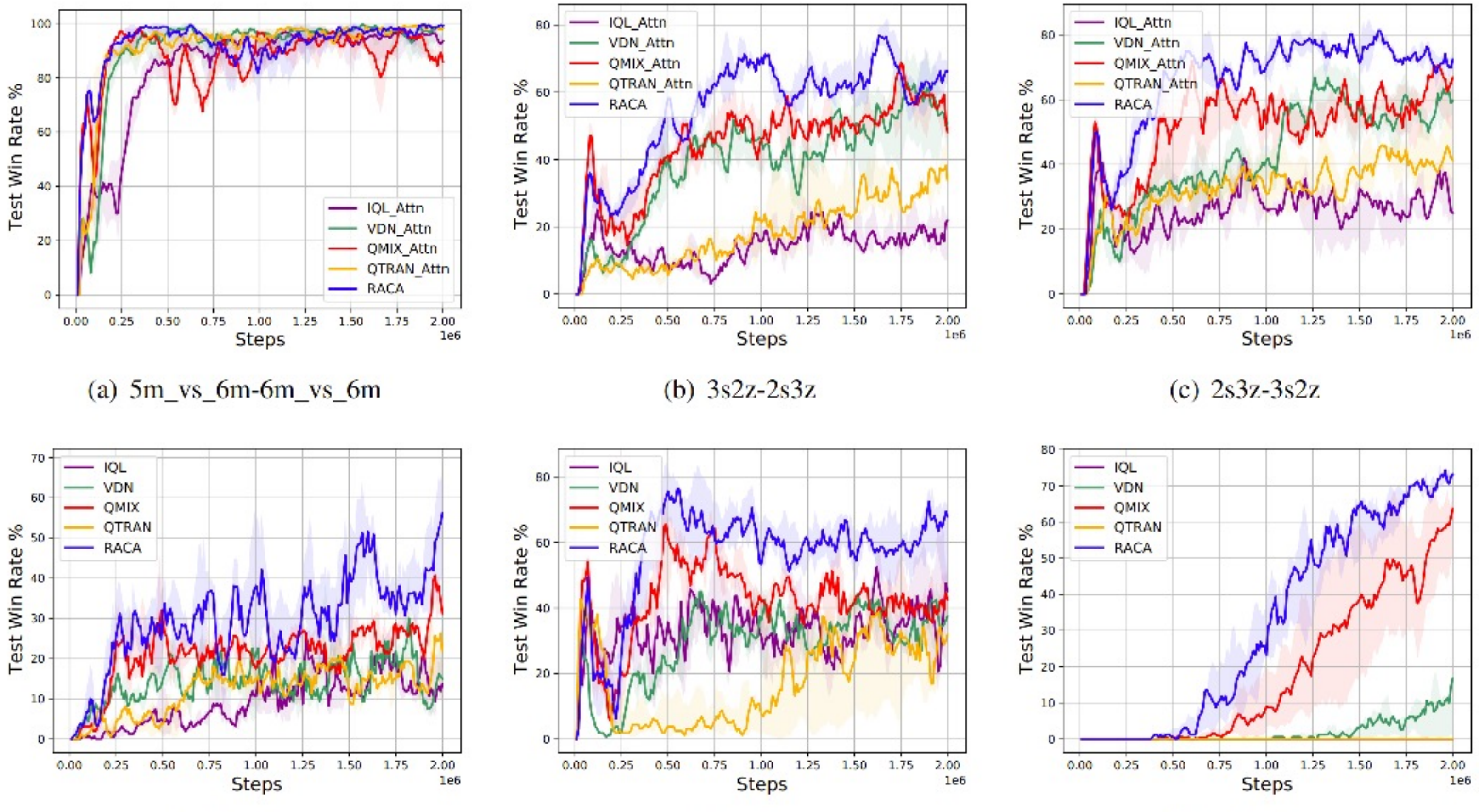


Fig. 2. Learning curves of our method and baseline algorithms on the StarCraftII micromanagement benchmark.

3 Experiments

Main results



(a) 5m_vs_6m-6m_vs_6m (b) 3s2z-2s3z (c) 2s3z-3s2z

(d) 5m-5ma (e) 5ma-5m (f) MMM2_rnd

Fig. 3. Learning curves of our method and baseline algorithms on ad-hoc cooperation scenarios.

3 Experiments

Ablation studies

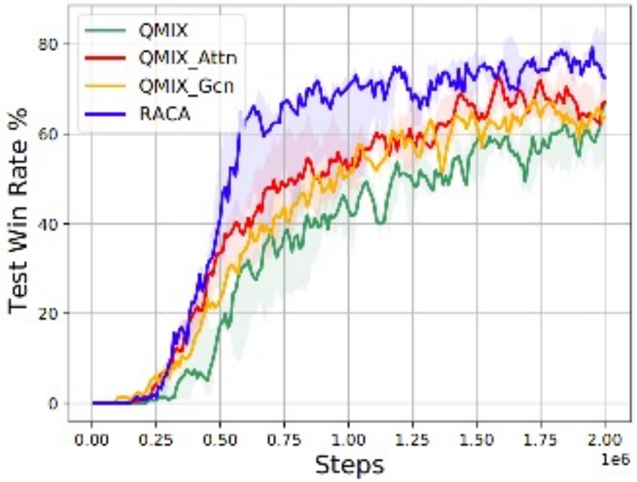


Fig. 4. Ablation studies on map 5m_vs_6m.

TABLE II
BASELINE AND ABLATION ALGORITHMS.

	Alg.	Description
Related Works	IQL	Independent Q-learning
	VDN	Additivity constraint
	QMIX	Monotonicity constraint
	QTRAN	Constraint-free
Ablations	IQL_Attn	IQL with attention
	VDN_Attn	VDN with attention
	QMIX_Attn	QMIX with attention
	QMIX_Gcn	QMIX with gcn module
	QTRAN_Attn	QTRAN with attention

4 Conclusion

A novel multi-agent credit assignment method RACA

- Graph-based relation encoder
- Attention-based observation abstraction mechanism

Future work

- Dynamic graph for ad-hoc cooperation

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

CASIA

Thank you.