DIVERSIFYING MESSAGE AGGREGATION IN MULTI-AGENT COMMUNICATION VIA NORMALIZED TENSOR NUCLEAR NORM REGULARIZATION

Yuanzhao Zhai¹,², Kele Xu¹,², Bo Ding¹,², Dawei Feng¹,²*, Zijian Gao¹,², Huaimin Wang¹,²

¹National University of Defense Technology, Changsha, China
²Key Laboratory of Software Engineering for Complex Systems, Changsha, China

ABSTRACT

The use of graph attention networks (GAT) in communication-enhanced multi-agent reinforcement learning (Comm-MARL) has become prevalent. While successful, GAT can lead to homogeneity in the strategies of message aggregation, which can severely limit multi-agent coordination. To address this challenge, we study the adjacency tensor of the communication graph. Then we define a new nuclear tensor rank and its convex surrogate, the normalized tensor nuclear norm to measure the homogeneity of message aggregation. Leveraging the norm, we further propose a plug-and-play regularizer on the adjacency tensor, named Normalized Tensor Nuclear Norm Regularization (NTNNR), to actively enrich the diversity of message aggregation during the training stage. NTNNR is agnostic to specific Comm-MARL algorithms and can be flexibly integrated with different graph-attention methods. Empirical results demonstrate that aggregating messages using NTNNR-enhanced GAT can improve the efficiency of the training and achieve higher asymptotic performance than existing message aggregation methods.

Index Terms— multi-agent reinforcement learning, graph neural networks, normalized tensor nuclear norm

1. INTRODUCTION

Multi-Agent Reinforcement Learning (MARL) has achieved remarkable success in a range of challenging sequential decision-making tasks, such as wireless edge caching [1], multi-player strategy games [2], and so on. As an under-explored issue in MARL, communication is a key component for multi-agent coordination. Agents can exchange their local observations via communication messages. These messages are aggregated and further utilized to augment individual local observations for selecting actions.

To model the interactions between agents, MARL has widely utilized graph neural networks (GNNs) [3] to allow for a graph-based representation. The multi-agent system is usually modeled as a complete graph, and each agent corresponds to a node. As one of the most popular GNNs variants, GAT has shown great potential in Comm-MARL [4]. With GAT, message aggregation can be achieved via attention-weighted message passing in the communication graph.

Despite the success of the GAT in Comm-MARL, we show that a lack of diversity still persists in the obtained message aggregation strategy. In essence, many agents in the graph may pay undue attention to a few “key” agents and are often excessively influenced. As shown in Fig. 1, the behavior obtained by Comm-MARL methods with homogeneous message aggregation strategies can be suboptimal, highlighting the urgent need for diverse message aggregation strategies. The homogeneity issue of GAT are identified in various tasks [5], whilst multiple agents with the parameter-sharing scheme exacerbate the problem severely.

In this paper, we aim to enable agents to explore diverse message aggregation strategies. Firstly, we study the adjacency tensor of the multi-agent communication graph, which consists of adjacency matrices generated by the multi-head attention mechanism of GAT. We present that the homogeneity of message aggregation could be measured by the normalized tensor rank. Since the rank optimization problem is known to be NP-hard, we define a new nuclear norm, which is a convex surrogate of normalized tensor rank, to replace the rank.

Fig. 1. With homogeneous message aggregation strategies obtained by GAT, predators 2, 3, and 4 are excessively influenced by the message of predator 1 who found a prey. Then all predators tend to pursue prey 1 while ignoring prey 2.
Accordingly, we propose a novel Normalized Tensor Nuclear Norm (NTNN) regularizer, which regularize adjacency tensors to actively enrich the diversity of the message aggregation strategies in Comm-MARL. In this way, agents could discover diverse behaviors and tend to find better coordination. In brief, our main contribution is threefold:

- We firstly propose to measure the diversity (or the homogeneity) of the message aggregation via the normalized tensor rank of the adjacency tensor.
- We define a novel normalized tensor nuclear norm to replace the rank. The norm can be further utilized as the regularizer to discover diverse message aggregation strategies for multi-agent communication.
- Experiments show that aggregating messages using GAT with NTNNR can improve training efficiency and asymptotic performance. Our regularizer also brings significant performance improvements for existing graph-attention Comm-MARL methods, using the plug-and-play manner.

2. RELATED WORK

In Comm-MARL, message aggregation strategies for agents determine how to aggregate received messages and partial observation to select the next actions. Some works aggregate messages with no preference, such as averaging [6, 7] and summing up [8]. Since messages encode the senders’ personal understanding of their observations, some may be more important than others.

As one of the most popular GNNs variants, GAT [9] generalizes the multi-head self-attention mechanism [10] from sequences to graphs, which allows the model to attend to information from different representation subspaces jointly. GAT has been proved an effective tool to aggregate messages with weighted message aggregation [11, 12, 13, 14]. TarMAC [11] utilize the signature-based soft-attention mechanism to enable targeting. DICG [13] introduces the deep implicit coordination graph with the self-attention mechanism.

GATv2 [5] finds by a theoretical analysis that the ranking of the attention scores generated by GAT may be unconditioned on the query node, which causes homogeneous message aggregation. To address this problem, GATv2 proposes to modify the order of weight calculation operation in GAT. Complementary with GATv2, our method regularize the adjacency tensor, encouraging diverse aggregation actively.

3. METHODOLOGY

We model the multi-agent tasks as a Decentralized Partially Observable Markov Decision Process (Dec-POMDP) augmented with communication, which can be described as a tuple $< N, S, U, P, R, O, M, G, \gamma >$. $N$ is the number of agents. $S$ represents the space of global states. $O$ denotes the space of observations of robots, and each agent receives a private observation $o_i \in O$ according to the observation function $\sigma(s_i) : S \rightarrow O$. $M$ represents the space of messages. Agents generate messages $m_{ij} \in M$ encoded by its observations and others’ messages at the last timestep, which could be modeled by the multi-agent communication graph $G = (< V, E >)$. Node $v_i \in V$ represent agents, and edges $e_{ij} \in E$ represent communication links. We denote $h_i$ as the feature of $v_i$. Combining with local observations, each agent aggregates communication messages and generates its own action $u_i = \pi_\theta(u_i, m_{ij \neq i})$, where $\pi_\theta$ is the policy with parameter $\theta$ shared across all agents. The transition probability of reaching state $s'$ from state $s$ by executing action $a$ is $P(s'|s, u)$. $R$ is the joint reward function. $\gamma \in [0, 1]$ denotes the discount factor. Each agent $i$ aims to maximize its discounted reward $E_{s, u \sim \pi} [r_t^i] = \sum_{t=0}^{\infty} \gamma^t r_t^i(s', u')$.

3.1. Measuring Message Aggregation’s Diversity with the Normalized Tensor Rank

We first consider the case that a single attention head is used. For each agent $i$, GAT computes the attention scores, as the elements of the adjacency matrix $A$, are normalized across all neighbors using the softmax function:

$$a_{ij} = \text{Softmax}_j(\text{LeakyReLU}(W'(W[h_i \| h_j])), \quad (1)$$

where $W$ and $W'$ are learnable, and $\|$ denotes vector concatenation. We denote vectors selected from the i-th and j-th rows of the matrix as $a_i$ and $a_j$, which represent the attention scores of agent $i$ and $j$ respectively for aggregating messages. If agents $i$ and $j$ have homogeneous message aggregation strategies, they focus on the same “key” agents. In this case, vectors $a_i$ and $a_j$ are less different. In other words, two vectors are linearly dependent. Therefore, we could measure the diversity (or the homogeneity) of the message aggregation with the rank of the adjacency matrix $A \in \mathbb{R}^{N \times N}$.

With the multi-head attention mechanism in GAT, independent $K$ attention mechanisms execute the attention function in parallel. Then we obtain a three-way adjacency tensor $A \in \mathbb{R}^{N \times N \times K}$. Its $K$ frontal slices $\{A^{(k)}\}_{k=1,\ldots,K}$ represent independent adjacency matrices. From another view, its mode-3 fiber $A(i, j, :)$. represents the attention scores of agent $j$ to agent $i$ using different attention heads.

From the above analysis, to enrich the message aggregation strategies, we should enlarge the rank of frontal slices, while maintaining the diversity of mode-3 fibers. However, most widely used tensor ranks, such as CP rank and Tucker rank [15] can not exactly measure the correlation from both frontal slices and mode-3 fibers views. This motivates us to define a new tensor rank to measure the homogeneity of message aggregation with multi-head attention GAT.

We denote $\hat{A}$ as a result of applying normalization to $A$ along the 3-rd way. Specifically, we apply Softmax on every
tube fibers \( \mathcal{A}(i, j, :) \), i.e.,
\[
\hat{A}_{ijk} = \frac{\exp(A_{ijk})}{\sum_{t \in [0, K-1]} \exp(A_{ijl})}, \quad K \geq 2.
\] (2)

Then we can define the normalized tensor rank as:
\[
\text{rank}_n(\mathcal{A}) = \sum_k \text{rank}(\hat{A}^{(k)}). \tag{3}
\]

3.2. Normalized Tensor Nuclear Norm Regularization

The rank optimization problem is known to be NP-hard. An alternative is to utilize the nuclear norm, and the matrix nuclear norm is defined as:
\[
\|A\|_* = \sum_i \sigma_i(A),
\] (4)
where \( \sigma_i(\mathcal{A}) \) are singular values of \( \mathcal{A} \).

For normalized tensor \( \mathcal{A} \), we denote \( \hat{A} \in \mathbb{R}^{NK \times NK} \) as the block diagonal matrix with its \( i-th \) block on the diagonal as the \( i-th \) frontal slice, i.e.,
\[
\hat{A} = \text{bdia}(\hat{A}) = \begin{bmatrix} \hat{A}^{(0)} & \hat{A}^{(1)} & \cdots & \hat{A}^{(K-1)} \end{bmatrix}.
\] (5)

Based on the matrix nuclear norm, we define a novel tensor nuclear norm for the normalized tensor rank, which is called Normalized Tensor Nuclear Norm (NTNN):
\[
\|\mathcal{A}\|_* = \frac{1}{K} \|\hat{A}\|_*.
\] (6)

As a special case, if \( \mathcal{A} \) reduces to a matrix \( (K = 1) \), it is not necessary to normalize the third dimension. In this case, NTNN reduces to the matrix nuclear norm. Considering the nuclear norm is the convex relaxation of the matrix rank [16], \( \|\hat{A}\|_* \) is a tight convex surrogate of rank(\( \hat{A} \)). Combining Equation 3 and 5, \( \|\mathcal{A}\|_* \) is a tight convex surrogate of rank\(_n(\mathcal{A})\).

3.3. Overall Optimization Objective

Following most Comm-MARL methods, we implement our framework with the policy decentralization with shared parameters (PDSP) paradigm. Then the gradient of Comm-MARL’s original loss function can be formulated as:
\[
\nabla_\theta L_{RL}(\theta) = E_{t,i}[\nabla_\theta \log \pi_\theta (u_t^i | o_t^i, m_{j \neq i}^t)] \Psi_t^i,
\] (7)
where \( \Psi_t^i \) is related to the discounted reward \( r_t^i \) and has various forms depending on different algorithms [17], and \( \theta \) denotes all parameters of the policy network.

As shown in Fig. 2, we apply NTNNR to the adjacency tensor \( \mathcal{A} \) of GAT layers. The corresponding loss function of NTNNR in the \( l-th \) layer can be formulated as:
\[
L_{NTNNR}(\theta_l) = -\|\mathcal{A}\|_*,
\] (8)
where \( \theta_l \) is part of parameters \( \theta \) to obtain the adjacency tensor \( \mathcal{A} \) of the \( l-th \) GAT layer.

Overall, we update the model parameter \( \theta \) by minimizing the following loss function:
\[
L(\theta) = L_{RL}(\theta) + \sum_l \lambda_l L_{NTNNR}(\theta_l),
\] (9)
where \( \lambda_l \) is the regularization weights of NTNNR for layer \( l \).

To anneal \( \lambda_l \) during the training process, we introduce new scaling hyper-parameters \( \beta_l \) and obtain adaptive weight:
\[
\lambda_l = \frac{|L_{RL}(\theta)|}{\beta_l \times |L_{NTNNR}(\theta_l)|},
\] (10)

4. EXPERIMENTAL RESULTS

4.1. Predator-Prey and Traffic Junction

We utilize two-layer GATs. The first layer contains two attention heads in the predator-prey scenario and four in the traffic junction scenario, while the second always contains one.

In the Predator-Prey scenario [18], we set 8 predators pursuing four fixed preys. We compare GAT with NTNNR with two baselines: vanilla GAT and applying our defined tensor nuclear norm without normalization (TNNR) to GAT.

We set scaling hyper-parameters to \( \beta_1 = 0.2, \beta_2 = 0.005 \) for the two GAT layers respectively. Fig. 3(a) shows the average reward as the training epoch increases. Integrating NTNNR into GAT could boost training efficiency, and obtain the highest asymptotic performance. We also record the corresponding NTNN values in Fig. 3(b) and 3(c) respectively. We can observe that vanilla GAT keeps small NTNN values
in both layers, which suggests that all agents have homogeneous message aggregation strategies. Compared to TNNR, NTNNR can maintain high NTNN values in both layers with the same scaling hyper-parameters. This is due to additional diversity among different attention heads introduced by the normalization along the 3-rd way, which reflects the effectiveness of the normalization operation of NTNNR.

The second scenario we employ is cooperative hard-mode traffic junction scenario [6]. We set the maximum number of cars in the environment to 20 and the maximum time steps to 50. We compare our proposed message aggregation method, GAT with NTNNR, against a variety of widely used message aggregation methods, including averaging used in CommNet [6], signature-based attention mechanism used in TarMAC [11], GAT, and GATv2 [5].

We set $\beta_1 = 0.01, \beta_2 = 0.005$ for the two GAT layers respectively. Fig. 3(d) shows the success rate per epoch attained by various message aggregation methods. GAT with NTNNR is competitive when compared to other methods with a higher success rate and sample efficiency.

4.2. StarCraft II

StarCraft II Multi-Agent Challenge (SMAC) [19] is a benchmark to evaluate various reinforcement learning works. Using the plug-and-play manner, we integrate NTNNR with one of the state-of-the-art Comm-MARL methods, DICG [13].

|          | DICG | DICG with NTNNR |
|----------|------|-----------------|
| 2s3z     | 0.97 | 0.99            |
| 3s5z     | 0.81 | 0.88            |
| 1c3s5z   | 0.91 | 0.93            |

Table 1. Success rates in the SMAC scenario.

The evaluation win rates are shown in Table 1. DICG with NTNNR achieve outstanding performance compared with its vanilla version. We suppose the improvement is due to the emergent behaviors brought by the diverse message aggregation strategies. To better explain why our regularizer performs well, we further visualize the final trained strategies in Fig. 4. In this 3s5z map, three parameter-sharing zealots with similar observations can select diverse actions and finally surround the enemy stalkers to attack. The sophisticated coordination reflects the effectiveness of diverse message aggregation in Comm-MARL.

5. CONCLUSION

In this paper, we present that the diversity of message aggregation in graph-attention Comm-MARL methods could be measured by the normalized tensor rank, and further define the corresponding nuclear norm to quantify the diversity. Then we propose a plug-and-play regularizer named NTNNR, to actively enrich the diversity of message aggregation. Experiments show that GAT with NTNNR can provide superior performance and better training efficiency compared to existing message aggregation methods. Furthermore, NTNNR can be easily applied to existing graph-attention Comm-MARL methods and improve their performance.

6. ACKNOWLEDGEMENT

This work is partially supported by the major Science and Technology Innovation 2030 “New Generation Artificial Intelligence” project 2020AAA0104803.
7. REFERENCES

[1] Navneet Garg and Tharmalingam Ratnarajah, “Cooperative scenarios for multi-agent reinforcement learning in wireless edge caching,” in ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021, pp. 3435–3439.

[2] Siqi Shen, Yongquan Fu, Huayou Su, Hengyue Pan, Peng Qiao, Yong Dou, and Cheng Wang, “Graphcomm: A graph neural network based method for multi-agent reinforcement learning,” in ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021, pp. 3510–3514.

[3] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini, “The graph neural network model,” IEEE transactions on neural networks, vol. 20, no. 1, pp. 61–80, 2008.

[4] Changxi Zhu, Mehdi Dastani, and Shihan Wang, “How attentive are graph attention networks?,” International Conference on Learning Representations, 2022.

[5] Shaked Brody, Uri Alon, and Eran Yahav, “Individualized controlled continuous communication model for multiagent cooperative and competitive tasks,” in International conference on learning representations, 2019.

[6] Yali Du, Bo Liu, Vincent Moens, Ziqi Liu, Zhicheng Ren, Jun Wang, Xu Chen, and Haifeng Zhang, “Learning correlated communication topology in multi-agent reinforcement learning,” in Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems, 2021, pp. 456–464.

[7] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio, “Graph attention networks,” International Conference on Learning Representations, 2018.

[8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin, “Attention is all you need,” Advances in neural information processing systems, vol. 30, 2017.

[9] Abhishek Das, Théophile Gervet, Joshua Romoff, Dhruv Batra, Devi Parikh, Mike Rabbat, and Joelle Pineau, “Tarmac: Targeted multi-agent communication,” in International Conference on Machine Learning. PMLR, 2019, pp. 1538–1546.

[10] Yong Liu, Weixun Wang, Yujing Hu, Jianye Hao, Xingguo Chen, and Yang Gao, “Multi-agent game abstraction via graph attention neural network,” in Proceedings of the AAAI Conference on Artificial Intelligence, 2020, vol. 34, pp. 7211–7218.