Towards Sample Efficient Learners in Population based Referential Games through Action Advising

Extended Abstract

Shresth Verma
ABV-Indian Institute of Information Technology and Management
Gwalior, India
vermashresth@gmail.com

ABSTRACT

The ability of agents to learn to communicate through interaction has been studied through emergent communication tasks. Previous works in this domain have studied the linguistic properties of the emergent languages like compositionality, generalization, and as well as the environmental pressures that shape them. However, most of these experiments require a considerable amount of shared training time between agents to communicate successfully. Our work highlights the problem of sample inefficiency of agents in population-based referential games and proposes an Action Advising framework to counter it.

KEYWORDS

Emergent Languages; Action Advising; Multi-Agent Reinforcement Learning

ACM Reference Format:
Shresth Verma. 2021. Towards Sample Efficient Learners in Population based Referential Games through Action Advising: Extended Abstract. In *Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), Online, May 3–7, 2021*, IFAAMAS, 3 pages.

1 INTRODUCTION

One of the long-standing challenges of AI is developing agents capable of coordinating with one another through natural language communication. The current prevalent paradigm in language learning has been through capturing statistical patterns in language structure from large amounts of data. This approach has shown tremendous success in natural language tasks such as machine translation, sentiment analysis, image captioning [19, 26, 27] etc. An alternative approach to language learning is through its functional nature, that is, language must emerge out of the need to communicate. This learning paradigm has had a long history [2], but only recently it has been studied through neural agents.

Referential games, which are a form of Lewis' Signalling games [10], are widely used for facilitating such emergence of natural language in multi-agent setting. This includes referential game settings with real-world visual input [9], sequence of message tokens for communication [8], multi-step communication with different modalities [5], emergence of writing system through brush strokes [22], emergence of non-verbal communication [16] etc. Another line of work studies emergence of language in a population of agents. This learning setting, while being more realistic, has been shown to regularize emergent language [6], counter language drift [13] and promote languages that are easier to teach [11] and compositional [17]. Studying language emergence in a population of agents also allows us to study complex language dynamics and collective behaviour of agents in language games [7, 12, 14, 21, 25]. However, one limitation that is often under-addressed in neural emergent communication literature is that of sample efficiency. As the agents in a language game are usually optimized through Reinforcement Learning, a large amount of shared training time is required. Sample efficiency is primarily hampered because of sparse rewards in a language game and extreme non-stationarity. The problem is further aggravated when we consider larger population sizes. An area in multiagent-learning that has been useful for faster and sample efficient learning is that of learning from demonstration [1, 20]. Further, Action Advising has been proposed, both in the presence of an expert and among simultaneously learning agents [3]. While Action Advising has previously been used in convention emergence [23], it has largely remained unexplored in a referential game setting. Thus, in our work, we propose an Action Advising framework that mitigates the effects of non-stationarity and sparse rewards in referential games.

Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), U. Endriss, A. Nowé, F. Dignum, A. Lomuscio (eds.), May 3–7, 2021, Online.

2 GAMES AND TERMINOLOGY

Paired Referential Game

We adapt the referential game formulation from [9]. The game consists of two players, a Speaker and a Listener. From a given set of entities E, we sample a target entity $t \in E$ and $K-1$ distracting entities $D = \{d_1, d_2, ..., d_{k-1}\}$ s.t. $\forall j,\ t \neq d_j,\ d_j \in E$. The candidate set $C = t\ \cup D$ contains both the target and distracting entities. The speaker is shown ordered set C and must come up with a message token m chosen from a fixed vocabulary V of size $|V|$. The listener is then shown message token m and U, which is a random permutation of C and it must point to an entity t'. Communication success is defined when $t = t'$, i.e., listener can correctly identify the target, in which case a payoff of 1 is given to both the players. In all other cases, payoff is 0.

Population based Referential Game (PopRG)

Consider two sets representing populations, both of size N, one consisting of speaker agents, $A_s = \{A_s^i\}_{i=1}^{i=N}$ and another for listener agents $\{A_l^i\}_{i=1}^{i=N}$. We define an undirected population interaction graph $G = (V_G, E_G)$ where $V_G = A_s \cup A_l$ and $E_G = \{(s, l)\ \forall\ s \in A_s, l \in A_l\}$. This graph thus represents the connection from speaker population to listener population. At every turn of gameplay, we randomly sample a speaker agent $A_s^i \in A_s$. Then, a listener $A_l^i \in A_l$ is sampled such that $(A_s^i, A_l^i) \in E_G$. This ensures that only the connected pair of agents are selected. At this point, a paired referential game described in the previous section is played between A_s^i and A_l^i.

Action Advising Framework

While the population interaction graph captures the interaction strictly between a speaker and listener, we introduce an undirected Advising Graph that allows interaction within the speaker population and listener population separately. We define Advising Graph $D = (V_D, E_D)$ where $V_D = A_s \cup A_l$ and $E_D = \{(s_1, s_2) \ \forall \ s_1 \in A_s, s_2 \in A_s, s_1 \neq s_2\} \cup \{(l_1, l_2) \ \forall \ l_1 \in A_l, l_2 \in A_l, l_1 \neq l_2\}$. During gameplay, we use teacher-induced advising whenever a successful episode is encountered. Any agent can assume the role of teacher and send advice to all the agents connected to it in the Advising Graph. Thus, this advice can be seen as the broadcasting of episode and action information to fellow agents of the same kind. Formally, for any agent A^i, advice is given to agents in A^i's set of advisees given by $Q(A^i) = \{A^j \in V_D \ s.t. \ (A^i, A^j) \in E_D\}$.

3 IMPLEMENTATION DETAILS

Agents and Learning

All the agents in the population, both speakers and listeners, are modelled as reinforcement learning policies. These policies are parameterized through neural networks with different parameters for each agent. We refer the reader to [9] for details on architecture. In all our experiments we take K as 3 and V as 5. The learning goal in Population based Referential Game is the maximization of the sum of rewards for all the agents. At any given game turn, it can be seen as optimizing expected reward for the pair of agents A_s^i and A_l^i in the paired referential game. Thus, $J(\theta_s^i, \theta_l^i) = E_{\Pi_s^i, \Pi_l^i}[R(t', t)]$ where θ_s^i, θ_l^i are the parameters for agents A_s^i and A_l^i respectively and Π_s^i, Π_l^i are their policies. Since message from speaker is sampled, the game is no longer end-to-end differentiable. As is a common practice in emergent language literature, we use REINFORCE update rule [24], to compute gradient of the objective and directly optimize the policies using the communication success as reward signal. Additionally, we use entropy regularization term [15], to encourage exploration. The gradient of cost functions can then be written as

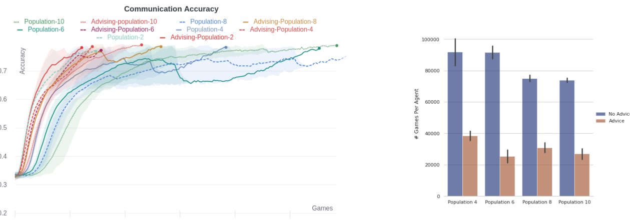
$$\nabla_{\theta_s^i} J = E_{\Pi_s^i, \Pi_l^i}[R(t', t).\nabla_{\theta_s^i} log \Pi_s^i(m|C)] + \lambda_s.\nabla_{\theta_s^i} H[\Pi_s^i(m|t)]$$

$$\nabla_{\theta_l^i} J = E_{\Pi_s^i, \Pi_l^i}[R(t', t).\nabla_{\theta_l^i} log \Pi_l^i(t'|m, U)] + \lambda_l.\nabla_{\theta_l^i} H[\Pi_l^i(t'|m, U)]$$

where H is the entropy function and λ_s and λ_l are hyperparameters controlling entropy regularization (both taken as 0.001).

Learning through Advice

Whenever a paired referential game between speaker S and listener L results in communication success, we allow S to advise all the speakers in advisee set $Q(S)$. Similarly, L is allowed to advise all fellow listeners in $Q(L)$ regarding the communication success. Specifically, this advice is implemented as a broadcast of experience trajectory containing state, action (that resulted in communication success) and reward (always 1) to fellow (similar) agents in the population. To learn from the advice, we first let the advisee generate probabilities over actions in a forward pass, from the state contained in the advice. However, instead of sampling from this action probability distribution, we use the action that is advised. After this, the optimization procedure described in previous section is followed. Advising in our setup can thus be seen as an exploration mechanism wherein, an agent explores actions that resulted in success for fellow agents.



(a) Communication success

(b) Games per Agent

Figure 1: (a): Communication success over game steps for different population sizes, with and without Action Advising. (b) Games required per agent to reach high communication success, with and without Action Advising

Data

In all our experiments, we use different classes of images in the ImageNet dataset [4] as entities. 200 images are randomly sampled from each of the classes to form the training set, and 100 images are sampled for the test set. Each image is represented as output from second last layer of a pretrained VGG16 network [18], resulting in a vector of length 2048. Further, since population based referential games are computationally expensive to run, we restrict the entities to top 26 classes in the synset of ImageNet.

4 EXPERIMENTS AND RESULTS

We report the number of games required to reach a high level of communication success and repeat this experiment for population sizes 2 to 10. Additionally, we observe the sample efficiency achieved through Action Advising. Figure 1a shows the effect of Action Advising on sample efficiency (number of games required to reach a high communication success (>80 %)) for different population sizes. All results are averaged over 5 runs with different seed values. It can be seen that for all the population sizes, using Action Advising results in faster convergence towards high communication success. Another trend that can be noted is that, as population size increases, more samples are required to reach convergence. This is because every counted gameplay is between a pair of agents. Hence, in Figure 1b, we show games required per agent for different population sizes to reach convergence. This analysis also supports our argument on the benefit of Action Advising.

5 CONCLUSION

In this work, we highlight the issue of sample inefficiency in simulating language emergence through reinforcement learning in large populations. We show empirical results suggesting that Action Advising can help mitigate this issue. While our work uses a simple advising mechanism, taking into account advice budget and action uncertainty would be promising directions for future work.

REFERENCES

[1] Tim Brys, Anna Harutyunyan, Halit Bener Suay, Sonia Chernova, Matthew E. Taylor, and Ann Nowé. 2015. Reinforcement Learning from Demonstration through Shaping. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, Qiang Yang and Michael J. Wooldridge (Eds.). AAAI Press, 3352–3358. http://ijcai.org/Abstract/15/472

[2] Vincent Crawford. 1998. A Survey of Experiments on Communication via Cheap Talk. *Journal of Economic Theory* 78, 2 (Feb 1998), 286–298. https://doi.org/10.1006/jeth.1997.2359

[3] Felipe Leno da Silva, Ruben Glatt, and Anna Helena Reali Costa. 2017. Simultaneously Learning and Advising in Multiagent Reinforcement Learning. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems, AAMAS 2017, São Paulo, Brazil, May 8-12, 2017*, Kate Larson, Michael Winikoff, Sanmay Das, and Edmund H. Durfee (Eds.). ACM, 1100–1108. http://dl.acm.org/citation.cfm?id=3091280

[4] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Fei-Fei Li. 2009. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA*. IEEE Computer Society, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

[5] Katrina Evtimova, Andrew Drozdov, Douwe Kiela, and Kyunghyun Cho. 2018. Emergent Communication in a Multi-Modal, Multi-Step Referential Game. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net. https://openreview.net/forum?id=rJGZq6g0-

[6] Nicole Fitzgerald. 2019. To Populate is To Regulate. *CoRR* abs/1911.04362 (2019). arXiv:1911.04362 http://arxiv.org/abs/1911.04362

[7] Laura Graesser, Kyunghyun Cho, and Douwe Kiela. 2019. Emergent Linguistic Phenomena in Multi-Agent Communication Games. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan (Eds.). Association for Computational Linguistics, 3698–3708. https://doi.org/10.18653/v1/D19-1384

[8] Serhii Havrylov and Ivan Titov. 2017. Emergence of Language with Multi-agent Games: Learning to Communicate with Sequences of Symbols. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (Eds.). 2149–2159.

[9] Angeliki Lazaridou, Alexander Peysakhovich, and Marco Baroni. 2017. Multi-Agent Cooperation and the Emergence of (Natural) Language. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net. https://openreview.net/forum?id=Hk8N3Sclg

[10] David Lewis. 1969. Convention cambridge. *Mass.: Harvard UP* (1969).

[11] Fushan Li and Michael Bowling. 2019. Ease-of-Teaching and Language Structure from Emergent Communication. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada*, Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett (Eds.). 15825–15835. http://papers.nips.cc/paper/9714-ease-of-teaching-and-language-structure-from-emergent-communication

[12] Ryan Lowe*, Abhinav Gupta*, Jakob Foerster, Douwe Kiela, and Joelle Pineau. 2020. On the interaction between supervision and self-play in emergent communication. In *International Conference on Learning Representations*. https://openreview.net/forum?id=rJxGLlBtwH

[13] Yuchen Lu, Soumye Singhal, Florian Strub, Olivier Pietquin, and Aaron C. Courville. 2020. Countering Language Drift with Seeded Iterated Learning. *CoRR* abs/2003.12694 (2020). arXiv:2003.12694 https://arxiv.org/abs/2003.12694

[14] Gary Lupyan and Rick Dale. 2010. Language structure is partly determined by social structure. *PloS one* 5, 1 (2010).

[15] Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. 2016. Asynchronous methods for deep reinforcement learning. In *International conference on machine learning*. 1928–1937.

[16] Igor Mordatch and Pieter Abbeel. 2018. Emergence of Grounded Compositional Language in Multi-Agent Populations. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, Sheila A. McIlraith and Kilian Q. Weinberger (Eds.). AAAI Press, 1495–1502. https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/17007

[17] Yi Ren, Shangmin Guo, Matthieu Labeau, Shay B. Cohen, and Simon Kirby. 2020. Compositional languages emerge in a neural iterated learning model. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net. https://openreview.net/forum?id=HkePNpVKPB

[18] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[19] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to Sequence Learning with Neural Networks. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, Zoubin Ghahramani, Max Welling, Corinna Cortes, Neil D. Lawrence, and Kilian Q. Weinberger (Eds.). 3104–3112.

[20] Matthew E. Taylor and Peter Stone. 2009. Transfer Learning for Reinforcement Learning Domains: A Survey. *J. Mach. Learn. Res.* 10 (2009), 1633–1685. https://dl.acm.org/citation.cfm?id=1755839

[21] Olivier Tieleman, Angeliki Lazaridou, Shibl Mourad, Charles Blundell, and Doina Precup. 2019. Shaping representations through communication: community size effect in artificial learning systems. *CoRR* abs/1912.06208 (2019). arXiv:1912.06208 http://arxiv.org/abs/1912.06208

[22] Shresth Verma and Joydip Dhar. 2020. Emergence of Writing Systems through Multi-Agent Cooperation (Student Abstract). In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 13941–13942. https://aaai.org/ojs/index.php/AAAI/article/view/7243

[23] Yixi Wang, Wenhuan Lu, Jianye Hao, Jianguo Wei, and Ho-fung Leung. 2018. Efficient Convention Emergence through Decoupled Reinforcement Social Learning with Teacher-Student Mechanism. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS 2018, Stockholm, Sweden, July 10-15, 2018*, Elisabeth André, Sven Koenig, Mehdi Dastani, and Gita Sukthankar (Eds.). International Foundation for Autonomous Agents and Multiagent Systems Richland, SC, USA / ACM, 795–803. http://dl.acm.org/citation.cfm?id=3237501

[24] Ronald J Williams. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning* 8, 3-4 (1992), 229–256.

[25] Alison Wray and George W Grace. 2007. The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua* 117, 3 (2007), 543–578.

[26] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron C. Courville, Ruslan Salakhutdinov, Richard S. Zemel, and Yoshua Bengio. 2015. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. In *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015 (JMLR Workshop and Conference Proceedings, Vol. 37)*, Francis R. Bach and David M. Blei (Eds.). JMLR.org, 2048–2057. http://proceedings.mlr.press/v37/xuc15.html

[27] Lei Zhang, Shuai Wang, and Bing Liu. 2018. Deep learning for sentiment analysis: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* 8, 4 (2018). https://doi.org/10.1002/widm.1253