ARL-TN-1053 ● MAR 2021

DEVCOM ARMY RESEARCH LABORATORY

Statistical Evaluation of Multi-Agent Reinforcement Learning Models under Different Versions of TensorFlow

by Erin Zaroukian, Rolando Fernandez, and Derrik Asher

NOTICES

Disclaimers

ARL-TN-1053 ● MAR 2021

Statistical Evaluation of Multi-Agent Reinforcement Learning Models under Different Versions of TensorFlow

Erin Zaroukian, Rolando Fernandez, and Derrik Asher
Computational and Information Sciences Directorate,
DEVCOM Army Research Laboratory

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED (From - To)**
March 2021	Technical Note	November 2020–February 2021

4. TITLE AND SUBTITLE

Statistical Evaluation of Multi-Agent Reinforcement Learning Models under Different Versions of TensorFlow

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Erin Zaroukian, Rolando Fernandez, and Derrik Asher

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

DEVCOM Army Research Laboratory
ATTN: FCDD-RLC-IT
Aberdeen Proving Ground, MD 21005

8. PERFORMING ORGANIZATION REPORT NUMBER

ARL-TN-1053

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES

ORCID ID(s): Erin Zaroukian, 0000-0002-1381-085X; Rolando Fernandez, 0000-0001-6950-6251; Derrik Asher, 0000-0002-1217-8581

14. ABSTRACT

The Agents Leveraging Learning for Intelligent Engagement with Soldiers (ALLIES) team has been using TensorFlow, a machine learning software library, to train and evaluate agents in tasks such as a continuous 2D version of the Predator-Prey Pursuit game. These tasks have provided a practical, dynamic research environment for studying cooperation and competition in agent-agent and human-agent teams, but staying compatible with research partners requires that we update to a new release of TensorFlow. To maintain continuity in our research, we evaluated predator-prey data created under the original and the new release of TensorFlow and confirmed that this transition does not affect training and evaluation.

15. SUBJECT TERMS

machine learning, multi-agent reinforcement learning (MARL), adversarial games, coordinated behavior in cooperative games, statistical modeling

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					Erin Zaroukian
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER (Include area code)**
Unclassified	Unclassified	Unclassified	UU	14	(410) 278-3203

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures iv

List of Tables iv

1. Introduction 1

2. Methods 1

3. Results 2

4. Conclusion 5

5. References 6

List of Symbols, Abbreviations, and Acronyms 7

Distribution List 8

List of Figures

Fig. 1 Example of the bounded continuous 2D Predator-Prey Pursuit environment that was utilized for this analysis with a small, faster prey agent (green) being pursued by three larger, slower predators (red). In this environment, the black box represents an impassible barrier. 1

Fig. 2 Total collisions per episode by model and TensorFlow version........... 3

Fig. 3 Significance of Kolmogorov-Smirnov tests between models. Comparisons within the same version of TensorFlow are shown in light red and light blue. 4

List of Tables

Table 1 Agent properties.. 2

Table 2 Mean and median total collisions from the 10,000 episodes generated with each version of TensorFlow... 2

1. Introduction

The Agents Leveraging Learning for Intelligent Engagement with Soldiers (ALLIES) team has been using TensorFlow,[1] a machine learning software library, to train and evaluate agents in tasks such as a continuous 2D version of the Predator-Prey Pursuit game (Fig. 1), which has provided a practical, dynamic research environment for studying cooperation and competition in agent-agent and human-agent teams.[2] In September of 2019, TensorFlow 2.0 was officially released, and updating our workflow to use TensorFlow 2.x (2.3 at the time of writing) will 1) allow us to quickly and easily "swap" (replace jointly trained partners with novel partners from independent training) agents to assess their behavior with respect to emergent coordination, 2) efficiently analyze the neural network parameters for a more complete understanding of emergent coordinated policies, and 3) facilitate collaboration with research groups that rely on TensorFlow.[3]

These updates to TensorFlow should not affect the underlying performance of its algorithms, but out of an abundance of caution we seek to ensure that the data produced using TensorFlow 2.3 does not differ meaningfully from our existing data generated under the previous major version, thus allowing for continuity in our research. To accomplish this, we performed a comparison of these two versions in a simulated Predator-Prey Pursuit task.

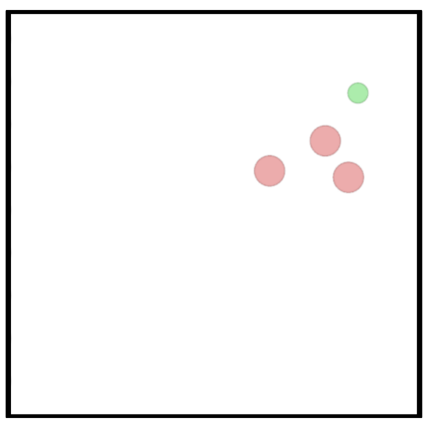

Fig. 1 Example of the bounded continuous 2D Predator-Prey Pursuit environment that was utilized for this analysis with a small, faster prey agent (green) being pursued by three larger, slower predators (red). In this environment, the black box represents an impassible barrier.

2. Methods

In our Predator-Prey Pursuit simulations, all agents' behaviors were guided by a multi-agent deep deterministic policy gradient (MADDPG) algorithm.[4] The MADDPG algorithm is an extension of a deep deterministic policy gradient

(DDPG) algorithm[5] into the multi-agent domain. Like DDPG, MADDPG utilizes an actor-critic model with deep neural networks representing policy and Q-learners. Multi-agent criteria are met with a centralized training technique that passes information about an agent's state and actions to each agent's critic network (only during training). The learning agents are joint action learners as opposed to independent learners, which indicates that one agent's actions were selected with respect to the actions of all other agents.

In the version of the Predator-Prey Pursuit task used here, one prey agent is pitted against three larger, slower predators (see Table 1 for agent properties) in an open arena. In training a reinforcement learning (RL) model in this task, the pursuers all receive a positive reward when any one of them comes into contact with the prey, and the prey receives a comparable negative reward. A trained model is then implemented and evaluated by summing the number of collisions across a fixed-length episode.

Table 1 Agent properties

Properties	Predator	Prey
Size	0.075	0.05
Maximum velocity	1.0	1.3
Maximum acceleration	3.0	4.0

We trained 10 models under each version of TensorFlow for a total of 20 models. Each model trained for 100,000 episodes, with 25 time steps per episode. We then generated a data set from each model, where each data set was composed of 1,000 episodes with 1,000 time steps per episode. Aggregated data and code for analysis can be found at https://osf.io/tvncr/.

3. Results

The data generated from the 10 models under each version of TensorFlow (TF1, TensorFlow 1.15; TF2, TensorFlow 2.3) are summarized in Table 2 and are similar to baseline results reported previously.[6]

Table 2 Mean and median total collisions from the 10,000 episodes generated with each version of TensorFlow

TensorFlow version	Mean of total collisions per episode	Median total collisions per episode
TF1	103.92	102
TF2	104.16	103

The data sets are broken down by model in Fig. 2, showing boxplots of total collisions per episode for each of the 10 models in each of the two TensorFlow versions.

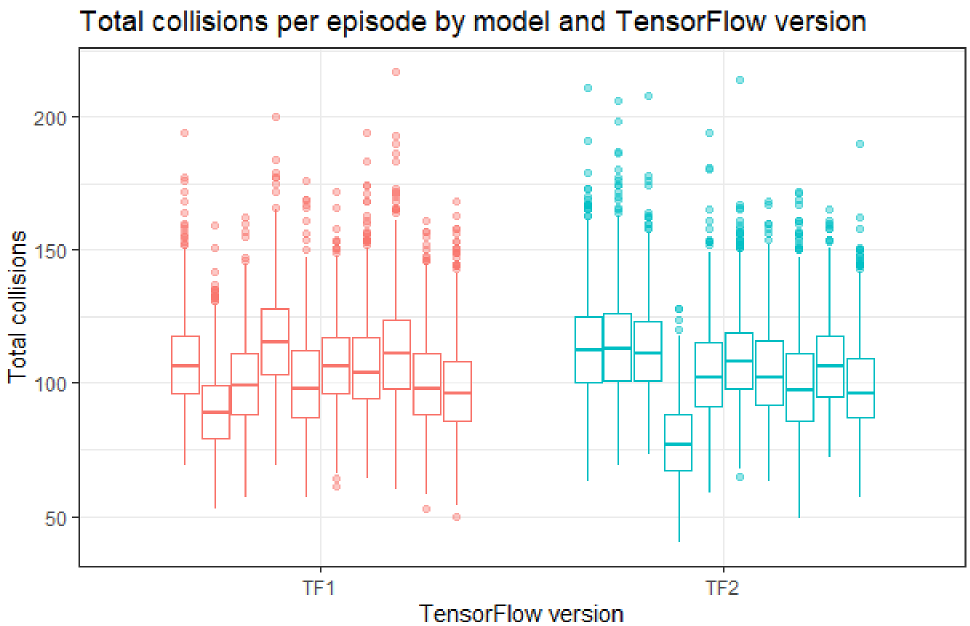

Fig. 2 Total collisions per episode by model and TensorFlow version

A mixed effects linear regression was performed, with TensorFlow version as fixed effect and random intercepts for Model. The results confirm that there is no meaningful difference between the TF1 and TF2 data sets (CI [7.95, 8.44]).*

In addition, Kolmogorov-Smirnov tests were performed pairwise between each model's empirical cumulative probability distributions of total collisions per episode. Bonferonni correction was used to compensate for multiple ($\binom{20}{2} = 190, \alpha = \frac{0.05}{190}$) comparisons. The result of these comparisons are shown in Fig. 3, where blue squares indicate significant comparisons, and red squares indicate non-significant comparisons.

* In the future, it may be useful to apply equivalence testing, using two one-sided tests against a minimal meaningful effect size.[7]

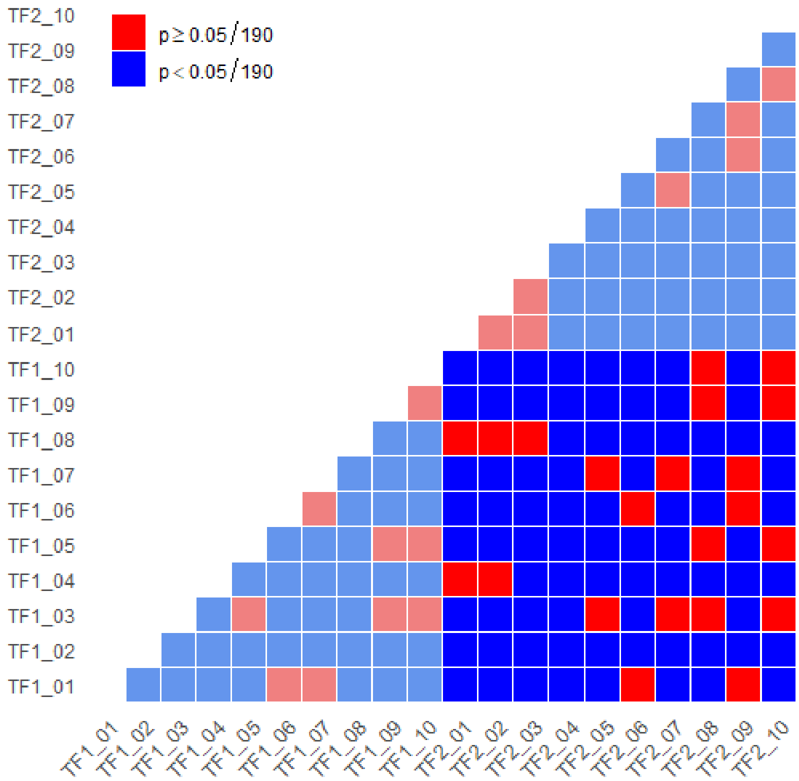

Fig. 3 Significance of Kolmogorov-Smirnov tests between models. Comparisons within the same version of TensorFlow are shown in light red and light blue.

While there is a notable amount of variation between versions of TensorFlow (78/100 model comparisons are significant), it is comparable to the amount of variation within versions of TensorFlow (74/90 model comparisons are significant).

A permutation test using the D statistics from these Kolmogorov-Smirnov tests (a measure of the maximum difference between the two distributions) confirms that there is no significant difference between the between-and within-version comparisons ($p = .477$). In this test, 2,000 permutations were performed to shuffle the between-version/within-version labels on the 190 D statistics, and for each permutation the difference between the median within-version statistic and the median between-versions statistic is computed. Compared to this distribution of 2,000 median D-statistic differences, the empirical difference of 0.0175 is within 1.96 standard deviations of the mean of this distribution.

4. Conclusion

By generating data sets and comparing their total collisions through mixed effects linear regression, as well as by comparing distributions of total collisions through Kolmogorov-Smirnov tests, we established that updating to TensorFlow 2.3 is unlikely to affect the results of our Predator-Prey Pursuit simulations. To further validate this transition, additional comparisons can be made using different performance metrics, difference agent parameters, and even different tasks.

5. References

1. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, et al. TensorFlow: large-scale machine learning on heterogeneous systems; 2015. Software available from tensorflow.org.

2. Asher DE, Zaroukian E, Barton SL. Adapting the predator-prey game theoretic environment to Army tactical edge scenarios with computational multiagent systems. Proceedings of the 23rd International Command and Control Research and Technology Symposium (ICCRTS); 2018.

3. TensorFlow Team. Effective TensorFlow 2.0: best practices and what's changed. TensorFlow Blog; 2019 Feb 13 [accessed 2021 Feb 5]. https://blog.tensorflow.org/2019/02/effective-tensorflow-20-best-practices.html.

4. Lowe R, Wu Y, Tamar A, Harb J, Abbeel OP, Mordatch I. Multi-agent actor-critic for mixed cooperative-competitive environments. Neural Information Processing Systems (NIPS); 2017.

5. Lillicrap TP, Hunt JJ, Pritzel A, Heess N, Erez T, Tassa Y, Silver D, Wierstra D. Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971; 2015.

6. Asher DE, Zaroukian E, Perelman B, Perret J, Fernandez R, Hoffman B, Rodriguez SS. Multi-agent collaboration with ergodic spatial distributions. In: Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications II (Vol. 11413, p. 114131N). International Society for Optics and Photonics; 2020.

7. Lakens D, Scheel AM, Isager PM. Equivalence testing for psychological research: a tutorial. Advances in Methods and Practices in Psychological Science. 2018;1.2:259–269.

List of Symbols, Abbreviations, and Acronyms

2D	two-dimensional
ALLIES	Agents Leveraging Learning for Intelligent Engagement with Soldiers
CI	Confidence Interval
DDPG	deep deterministic policy gradient
MADDPG	multi-agent deep deterministic policy gradient
TF1	TensorFlow 1.15
TF2	TensorFlow 2.0

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
1 (PDF)	DEVCOM ARL FCDD RLD DCI TECH LIB
3 (PDF)	DEVCOM ARL FCDD RLC IS R FERNANDEZ FCDD RLC IT E ZAROUKIAN D ASHER