Title	A Distributed Asynchronous Deep Reinforcement Learning Framework for Recommender Systems
Authors(s)	Shi, Bichen, Tragos, Elias, Ozsoy, Makbule Gulcin, Dong, Ruihai, Smyth, Barry, Hurley, Neil J., Lawlor, Aonghus
Publication date	2020-09-26
Publication information	Shi, Bichen, Elias Tragos, Makbule Gulcin Ozsoy, Ruihai Dong, Barry Smyth, Neil J. Hurley, and Aonghus Lawlor. “A Distributed Asynchronous Deep Reinforcement Learning Framework for Recommender Systems.” ACM, September 26, 2020. https://doi.org/10.1145/3383313.
Conference details	The 14th ACM Conference on Recommender Systems (RecSys 2020), Online, 22-26 September 2020
Publisher	ACM
Item record/more information	http://hdl.handle.net/10197/12220
Publisher's statement	© ACM, 2020. This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in RecSys '20: Fourteenth ACM Conference on Recommender Systems http://doi.acm.org/10.1145/3383313
Publisher's version (DOI)	10.1145/3383313

A Distributed Asynchronous Deep Reinforcement Learning Framework for Recommender Systems

BICHEN SHI*, ELIAS Z. TRAGOS*, MAKBULE GULCIN OZSOY*, RUIHAI DONG*,BARRY SMYTH*, NEIL HURLEY*, AONGHUS LAWLOR*, *Insight Centre for Data Analytics, University College Dublin, Ireland

In this paper we propose DADRL, a distributed, asynchronous reinforcement learning recommender system based on the asynchronous advantage actor-critic model (A3C), which combines ideas from A3C and federated learning (FL). The proposed algorithm keeps the user preferences or interactions on local devices and uses a combination of on-device, local recommendation models and a complementary global model. The global model is trained only by the loss gradients of the local models, rather than directly using user preferences or interactions data. We demonstrate, using well-known datasets and benchmark algorithms, how this approach can deliver performance that is comparable with the current state-of-the-art while enhancing user privacy.

CCS Concepts: • **Computer systems organization** → **Distributed architectures**; • **Theory of computation** → **Reinforcement learning**; • **Information systems** → **Recommender systems**.

Additional Key Words and Phrases: reinforcement learning, recommender systems, distributed learning

ACM Reference Format:
Bichen Shi*, Elias Z. Tragos*, Makbule Gulcin Ozsoy*, Ruihai Dong*,Barry Smyth*, Neil Hurley*, Aonghus Lawlor*. 2018. A Distributed Asynchronous Deep Reinforcement Learning Framework for Recommender Systems. In *RecSys '20: ACM Conference on Recommender Systems, September 22–26, 2020, Online.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Traditional Recommender Systems (RS) use central servers that gather and store data from user interactions to compute user profiles and train global recommendation models. RS models which are trained centrally can achieve great performance because the full user profiles and other information is available to them during training. However, centralised RS require users to share their whole interaction history with the server, leading to problems with scalability as the number of users and interactions increase, and have a central point of attack with respect to user privacy, as all user profiles are stored centrally.

Motivated by the scalability and privacy issues related to a central server, we propose an asynchronous, distributed, deep reinforcement learning (RL) based recommendation algorithm (DADRL) based on the ideas from asynchronous advantage actor-critic model (A3C) [3] and from federated learning (FL) [1].

As shown in Figure 1, the DADRL framework comprised a global deep neural network (DNN) that exists on a central server and a local DNN on each user device, which is a copy of a version of the global model. On the device, the local DNN uses the locally stored user-item interactions in order to compute the gradient losses, which are sent to the central server for training the global DNN and updating its weights. A new copy of the global DNN is then sent back to the user device to be used as the new local DNN for calculating the losses of the next interactions (in the next session). DADRL

Manuscript submitted to ACM

addressed user privacy, as the global model is trained with only the gradients of the loss that it receives asynchronously from the local devices. No other private user information is shared. In addition, due to the asynchronous nature of DADRL, each user gets a local copy of the global DNN at different times, and it is rare that two users have exactly the same version of the model, which further mitigates privacy attacks and differentiates our model from FL [4].

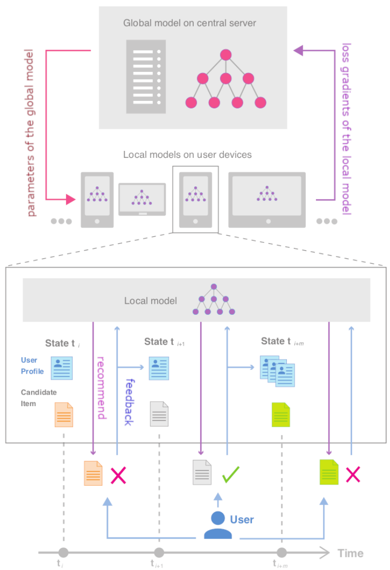

Figure 1. Overview of the privacy-preserving distributed deep reinforcement learning (DADRL) framework.

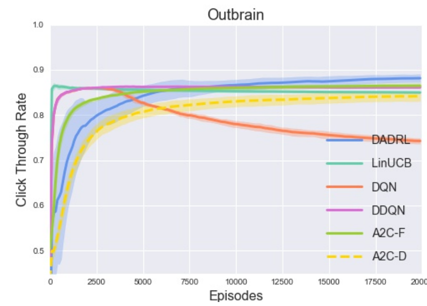

Figure 2. Global average CTR vs number of training episodes for Outbrain.

Outbrain	k=1	k=10	k=20	k=50	k=100
LinUCB	0.856	0.359	0.213	0.096	0.049
DQN	0.883	0.417	0.261	0.120	0.063
DDQN	0.881	0.403	0.246	0.113	0.059
A2C-F	0.888	0.428	0.261	0.122	0.064
A2C-D	0.869	0.381	0.240	0.109	0.057
DADRL	**0.914**	**0.567**	**0.393**	**0.201**	**0.111**

Table 1. The Click Through Rate (CTR) of the 6 methods in One Plus K Random Items test setting on the dataset.

In Fig. 2 we show the moving global average CTR of DADRL during the training process, compared with various state of the art baselines. As shown in the figure, DADRL achieves comparable results with the state-of-the-art algorithms, either centralised (LinUCB[2], DQN[5], DDQN) or synchronous (A2C-F, A2C-D). The fully trained models are evaluated on a realistic testing environment, employing the One-Plus-k-Random-Items evaluation protocol. The performance of all methods on the Outbrain datasets, with different values of k are presented in Table 1. The results reveal that DADRL outperforms the other models for all k. We point out that when there are significantly more negatively sampled items than user interactions ($k \geq 50$) the performance of DADRL is much better than the other baseline models.

ACKNOWLEDGMENTS

This work was supported by the Science Foundation Ireland through the Insight Centre for Data Analytics under grant number SFI/12/RC/2289_P2.

REFERENCES

[1] Keith Bonawitz, Hubert Eichner, Wolfgang Grieskamp, Dzmitry Huba, Alex Ingerman, Vladimir Ivanov, Chloe Kiddon, Jakub Konecny, Stefano Mazzocchi, H Brendan McMahan, et al. 2019. Towards federated learning at scale: System design. *arXiv preprint arXiv:1902.01046* (2019).

[2] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. 2010. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*. ACM, 661–670.

[3] 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.

[4] Milad Nasr, Reza Shokri, and Amir Houmansadr. 2019. Comprehensive Privacy Analysis of Deep Learning: Passive and Active White-box Inference Attacks against Centralized and Federated Learning. In *IEEE Symposium on Security and Privacy (SP)*.

[5] Guanjie Zheng, Fuzheng Zhang, Zihan Zheng, Yang Xiang, Nicholas Jing Yuan, Xing Xie, and Zhenhui Li. 2018. Drn: A deep reinforcement learning framework for news recommendation. In *Proc. of the 2018 Conference on World Wide Web*. International WWW Conferences Steering Committee.