

- [18] Ross, S., Bagnell, J.A. (2014). Reinforcement and Imitation Learning via Interactive no-regret learning. arXiv:1406.5979 [cs, stat], Jun. 2014.
- [19] Daumé III, H., Langford, J., Marcu, D. (2009). Search-based structured prediction. arXiv:0907.0786 [cs], Jul. 2009.
- [20] Ross, S., Bagnell, J.A. (2010). Efficient reductions for imitation learning. AISTATS.
- [21] Mnih, V., Badia, A., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., Kavukcuoglu, K. (2016). Asynchronous Methods for Deep Reinforcement Learning. arXiv:1602.01783 [cs], Feb. 2016.
- [22] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O. (2017). Proximal policy optimization algorithms. arXiv:1707.06347 [cs], Jul. 2017.
- [23] Wu, Y., Mansimov, E., Liao, S., Grosse, R., Ba, J. (2017). Scalable trust-region method for deep reinforcement learning using Kronecker-factored approximation. arXiv:1708.05144 [cs], Aug. 2017.
- [24] Sun, W., Bagnell, J.A., Boots, B. (2018). Truncated horizon policy search: Combining reinforcement learning & imitation learning. arXiv:1805.11240 [cs, stat], May 2018.
- [25] Cheng, C.A., Yan, X., Wagener, N., Boots, B. (2018). Fast policy learning through imitation and reinforcement. arXiv:1805.10413 [cs, stat], May 2018.
- [26] Dropout: A simple way to prevent neural networks from overfitting. <http://jmlr.org/papers/v15/srivastava14a.html>, accessed on 04-May-2019.
- [27] Deep Learning. <https://srdas.github.io/DLBook/ImprovingModelGeneralization.html>, accessed on 04-May-2019.
- [28] OpenAI. (2019). Gym: A toolkit for developing and comparing reinforcement learning algorithms. <https://gym.openai.com>, accessed on 04-May-2019.
- [29] Chauvin, S. (2018). Hierarchical decision-making for autonomous driving. <https://doi.org/10.13140/RG.2.2.24352.43526>
- [30] Schorner, P., Tottel, L., Doll, J., Zöllner, J. (2019). Predictive trajectory planning in situations with hidden road users using partially observable Markov decision processes. 2019 IEEE Intelligent Vehicles Symposium (IV), pp. 2299-2306. <https://doi.org/10.1109/IVS.2019.8814022>
- [31] Sharifzadeh, S., Chiotellis, J., Triebel, R., Cremers, D. (2016). Learning to drive using inverse reinforcement learning and deep Q-networks.

NOMENCLATURE

| | |
|---------|--|
| RL | Reinforcement learning |
| DRL | Deep Reinforcement Learning |
| IL | Imitation learning |
| BC | Behavioral Cloning |
| PER | Prioritized Experience Replay |
| A2C | Advantage Actor-Critic algorithm |
| A2CfD | Advantage Actor-Critic algorithm from Demonstrations |
| A2CfDoC | Advantage Actor-Critic algorithm from Demonstrations optimally Constrained |

Greek symbols

| | |
|-------------------|---|
| β_{Expert} | Expert's trust margin threshold |
| ϵ^{Clip} | Expert Gradient clipping hyperparameter |
| $L_{A2CfDoC}$ | A2CfDoC loss function |
| $R_{Explorer}$ | The Explorer Network Reward |
| R_{Expert} | The Expert Network Reward |