Deep Reinforcement Learning for NLP

William Yang Wang
UC Santa Barbara
william@cs.ucsb.edu

Jiwei Li
Shannon.ai
jiwei_li@shannonai.com

Xiaodong He
JD AI Research
xiaohe@ieee.org

Abstract

Many Natural Language Processing (NLP) tasks (including generation, language grounding, reasoning, information extraction, coreference resolution, and dialog) can be formulated as deep reinforcement learning (DRL) problems. However, since language is often discrete and the space for all sentences is infinite, there are many challenges for formulating reinforcement learning problems of NLP tasks. In this tutorial, we provide a gentle introduction to the foundation of deep reinforcement learning, as well as some practical DRL solutions in NLP. We describe recent advances in designing deep reinforcement learning for NLP, with a special focus on generation, dialogue, and information extraction. We discuss why they succeed, and when they may fail, aiming at providing some practical advice about deep reinforcement learning for solving real-world NLP problems.

1 Tutorial Description

Deep Reinforcement Learning (DRL) (Mnih et al., 2015) is an emerging research area that involves intelligent agents that learn to reason in Markov Decision Processes (MDP). Recently, DRL has achieved many stunning breakthroughs in Atari games (Mnih et al., 2013) and the Go game (Silver et al., 2016). In addition, DRL methods have gained significantly more attentions in NLP in recent years, because many NLP tasks can be formulated as DRL problems that involve incremental decision making. DRL methods could easily combine embedding based representation learning with reasoning, and optimize for a variety of non-differentiable rewards. However, a key challenge for applying deep reinforcement learning techniques to real-world sized NLP problems is the model design issue. This tutorial draws connections from theories of deep reinforcement learning to practical applications in NLP.

In particular, we start with the gentle introduction to the fundamentals of reinforcement learning (Sutton and Barto, 1998; Sutton et al., 2000). We further discuss

their modern deep learning extensions such as Deep Q-Networks (Mnih et al., 2015), Policy Networks (Silver et al., 2016), and Deep Hierarchical Reinforcement Learning (Kulkarni et al., 2016). We outline the applications of deep reinforcement learning in NLP, including dialog (Li et al., 2016), semi-supervised text classification (Wu et al., 2018), coreference (Clark and Manning, 2016; Yin et al., 2018), knowledge graph reasoning (Xiong et al., 2017), text games (Narasimhan et al., 2015; He et al., 2016a), social media (He et al., 2016b; Zhou and Wang, 2018), information extraction (Narasimhan et al., 2016; Qin et al., 2018), language and vision (Pasunuru and Bansal, 2017; Misra et al., 2017; Wang et al., 2018a,b,c; Xiong et al., 2018), etc.

We further discuss several critical issues in DRL solutions for NLP tasks, including (1) The efficient and practical design of the action space, state space, and reward functions; (2) The trade-off between exploration and exploitation; and (3) The goal of incorporating linguistic structures in DRL. To address the model design issue, we discuss several recent solutions (He et al., 2016b; Li et al., 2016; Xiong et al., 2017). We then focus on a new case study of hierarchical deep reinforcement learning for video captioning (Wang et al., 2018b), discussing the techniques of leveraging hierarchies in DRL for NLP generation problems. This tutorial aims at introducing deep reinforcement learning methods to researchers in the NLP community. We do not assume any particular prior knowledge in reinforcement learning. The intended length of the tutorial is 3 hours, including a coffee break.

2 Outline

Representation Learning, Reasoning (Learning to Search), and Scalability are three closely related research subjects in Natural Language Processing. In this tutorial, we touch the intersection of all the three research subjects, covering various aspects of the theories of modern deep reinforcement learning methods, and show their successful applications in NLP. This tutorial is organized in three parts:

- **Foundations of Deep Reinforcement Learning**. First, we will provide a brief overview

of reinforcement learning (RL), and discuss the classic settings in RL. We describe classic methods such as Markov Decision Processes, REINFORCE (Williams, 1992), and Q-learning (Watkins, 1989). We introduce model-free and model-based reinforcement learning approaches, and the widely used policy gradient methods. In this part, we also introduce the modern renovation of deep reinforcement learning (Mnih et al., 2015), with a focus on games (Mnih et al., 2013; Silver et al., 2016).

- **Practical Deep Reinforcement Learning: Case Studies in NLP** Second, we will focus on the designing practical DRL models for NLP tasks. In particular, we will take the first deep reinforcement learning solution for dialogue (Li et al., 2016) as a case study. We describe the main contributions of this work: including its design of the reward functions, and why they are necessary for dialog. We then introduce the gigantic action space issue for deep Q-learning in NLP (He et al., 2016a,b), including several solutions. To conclude this part, we discuss interesting applications of DRL in NLP, including information extraction and reasoning.

- **Lessons Learned, Future Directions, and Practical Advices for DRL in NLP** Third, we switch from the theoretical presentations to an interactive demonstration and discussion session: we aim at providing an interactive session to transfer the theories of DRL into practical insights. More specifically, we will discuss three important issues, including problem formulation/model design, exploration vs. exploitation, and the integration of linguistic structures in DRL. We will show case a recent study (Wang et al., 2018b) that leverages hierarchical deep reinforcement learning for language and vision, and extend the discussion. Practical advice including programming advice will be provided as a part of the demonstration.

3 History

The full content of this tutorial has not yet been presented elsewhere, but some parts of this tutorial has also been presented at the following locations in recent years:

1. "*Deep Reinforcement Learning for Knowledge Graph Reasoning*", William Wang, presented at the IBM T. J. Watson Research Center, Yorktown Heights, NY, Bloomberg, New York, NY and Facebook, Menlo Park, CA, total attendance: 150.

2. "*Deep Learning and Continuous Representations for NLP*", Wen-tau Yih, Xiaodong He, and Jianfeng Gao. Tutorial at IJCAI 2016, New York City, total attendance: 100.

3. "*Teaching a Machine to Converse*", Jiwei Li, presented at OSU, UC Berkeley, UCSB, Harbin Inst. of Technology, total attendance: 500.

4 Information About the Presenters

William Wang is an Assistant Professor at the Department of Computer Science, University of California, Santa Barbara. He received his PhD from School of Computer Science, Carnegie Mellon University. He focuses on information extraction and he is the faculty author of DeepPath—the first deep reinforcement learning system for multi-hop reasoning. He has published more than 50 papers at leading conferences and journals including *ACL, EMNLP, NAACL, CVPR, COLING, IJCAI, CIKM, ICWSM, SIGDIAL, IJCNLP, INTERSPEECH, ICASSP, ASRU, SLT, Machine Learning, and Computer Speech & Language*, and he has received paper awards and honors from CIKM, ASRU, and EMNLP. Website: `http://www.cs.ucsb.edu/~william/`

Jiwei Li recently spent three years and received his PhD in Computer Science from Stanford University. His research interests are deep learning and dialogue. He is the most prolific NLP/ML first author during 2012-2016, and the lead author of the first study in deep reinforcement learning for dialogue generation. He is the recipient of a Facebook Fellowship in 2015. Website: `https://web.stanford.edu/~jiweil/`

Xiaodong He is the Deputy Managing Director of JD AI Research and Head of the Deep learning, NLP and Speech Lab, and a Technical Vice President of JD.com. He is also an Affiliate Professor at the University of Washington (Seattle), serves in doctoral supervisory committees. Before joining JD.com, He was with Microsoft for about 15 years, served as Principal Researcher and Research Manager of the DLTC at Microsoft Research, Redmond. His research interests are mainly in artificial intelligence areas including deep learning, natural language, computer vision, speech, information retrieval, and knowledge representation. He has published more than 100 papers in *ACL, EMNLP, NAACL, CVPR, SIGIR, WWW, CIKM, NIPS, ICLR, ICASSP, Proc. IEEE, IEEE TASLP, IEEE SPM*, and other venues. He received several awards including the Outstanding Paper Award at ACL 2015. Website: `http://air.jd.com/people2.html`

References

Kevin Clark and Christopher D Manning. 2016. Deep reinforcement learning for mention-ranking coreference models. *EMNLP*.

Ji He, Jianshu Chen, Xiaodong He, Jianfeng Gao, Lihong Li, Li Deng, and Mari Ostendorf. 2016a. Deep reinforcement learning with a natural language action space. *ACL*.

Ji He, Mari Ostendorf, Xiaodong He, Jianshu Chen, Jianfeng Gao, Lihong Li, and Li Deng. 2016b. Deep reinforcement learning with a combinatorial action space for predicting popular reddit threads. *EMNLP*.

Tejas D Kulkarni, Karthik Narasimhan, Ardavan Saeedi, and Josh Tenenbaum. 2016. Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation. In *NIPS*.

Jiwei Li, Will Monroe, Alan Ritter, Michel Galley, Jianfeng Gao, and Dan Jurafsky. 2016. Deep reinforcement learning for dialogue generation. *EMNLP*.

Dipendra K Misra, John Langford, and Yoav Artzi. 2017. Mapping instructions and visual observations to actions with reinforcement learning. *arXiv preprint arXiv:1704.08795*.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. 2013. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. 2015. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533.

Karthik Narasimhan, Tejas Kulkarni, and Regina Barzilay. 2015. Language understanding for text-based games using deep reinforcement learning. *EMNLP*.

Karthik Narasimhan, Adam Yala, and Regina Barzilay. 2016. Improving information extraction by acquiring external evidence with reinforcement learning. *EMNLP*.

Ramakanth Pasunuru and Mohit Bansal. 2017. Reinforced video captioning with entailment rewards. *EMNLP*.

Pengda Qin, Weiran Xu, and William Yang Wang. 2018. Robust distant supervision relation extraction via deep reinforcement learning. In *ACL*.

David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. 2016. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489.

Richard S Sutton and Andrew G Barto. 1998. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge.

Richard S Sutton, David A McAllester, Satinder P Singh, and Yishay Mansour. 2000. Policy gradient methods for reinforcement learning with function approximation. In *NIPS*.

Xin Wang, Wenhu Chen, Yuan-Fang Wang, and William Yang Wang. 2018a. No metrics are perfect: Adversarial reward learning for visual storytelling. In *ACL*.

Xin Wang, Wenhu Chen, Jiawei Wu, Yuan-Fang Wang, and William Yang Wang. 2018b. Video captioning via hierarchical reinforcement learning. In *CVPR*.

Xin Wang, Wenhan Xiong, Hongmin Wang, and William Yang Wang. 2018c. Look before you leap: Bridging model-free and model-based reinforcement learning for planned-ahead vision-and-language navigation. *arXiv preprint arXiv:1803.07729*.

Christopher John Cornish Hellaby Watkins. 1989. *Learning from delayed rewards*. Ph.D. thesis, King's College, Cambridge.

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

Jiawei Wu, Lei Li, and William Yang Wang. 2018. Reinforced co-training. In *NAACL*.

Wenhan Xiong, Xiaoxiao Guo, Mo Yu, Shiyu Chang, and William Yang Wang. 2018. Scheduled policy optimization for natural language communication with intelligent agents. In *IJCAI-ECAI*.

Wenhan Xiong, Thien Hoang, and William Yang Wang. 2017. Deeppath: A reinforcement learning method for knowledge graph reasoning. *EMNLP*.

Qingyu Yin, Yu Zhang, Wei-Nan Zhang, Ting Liu, and William Yang Wang. 2018. Deep reinforcement learning for chinese zero pronoun resolution. In *ACL*.

Xianda Zhou and William Yang Wang. 2018. Mojitalk: Generating emotional responses at scale. In *ACL*.