Departamento de Ingeniería Eléctrica y Electrónica

Reinforcement Learning

IELE4922

Fernando Lozano Martínez
e-mail: flozano@uniandes.edu.co
Horas de Atención: Miércoles y Viernes 2:30-3:30 p.m. ML 427

1. Descripción

Reinforcement Learning (RL) o Aprendizaje por Refuerzo, es un paradigma de aprendizaje de máquina agentes que aprenden autónomamente a realizar una tarea a partir de su interacción en el ambiente en el que están inmersos. La aplicación exitosa de RL a problemas reales en robótica, control, juegos por computador y múltiples otras áreas hacen que sea una de las áreas de estudio más promisorias en inteligencia artificial. En este curso se estudiará el paradigma general de RL en el contexto de procesos de decisión de Markov (MDP) y los algoritmos de solución de problemas de RL tanto en entornos discretos como continuos, incluyendo algoritmos modernos de RL profundo (Deep Reinforcement Learning). Se pretende que al finalizar el curso, el estudiante pueda identificar un problema de RL en un contexto real, seleccionar el algoritmo apropiado de RL para resolverlo y evaluar la solución obtenida.

2. Evaluación

Criterio	**Porcentaje**
Tareas	60 %
2 Exámenes	40 %

3. Reglas

- Las tareas realizarse en grupos de máximo dos personas. Aunque es válido discutir los problemas con sus compañeros de otros grupos, el trabajo debe ser de completa autoría de los estudiantes del grupo. Esto quiere decir que no es permitido por ejemplo usar el trabajo de otro grupo (u otra fuente similar) como "guía"para resolver su tarea. Está prohibido copiar cualquier material/código desarrollado por otro estudiante, u obtener soluciones del internet, soluciones de semestres pasados u otros medios, a no ser que se especifique en el enunciado. Cualquier transgresión a esta regla se considerará **FRAUDE** y se reportará sin excepciones. Me reservo el derecho de solicitar sustentación detallada de las tareas y de asignar la calificación de acuerdo a la sustentación.
- Cada tarea tendrá una fecha y hora de entrega predeterminada. Usted (o su grupo) dispone de un "presupuesto" de 8 días de retardo que puede utilizar libremente sin incurrir en ninguna penalización (por ejemplo usted puede entregar la primera tarea tres días tarde y la segunda cinco días tarde y las demás a tiempo). Una tarea entregada tarde cuando se haya agotado el presupuesto tendrá nota de cero.
- La calificación final se obtendrá mediante redondeo a las centésimas (por ejemplo $2,995$ corresponde a $3,00$, pero $2,994$ corresponde a $2,99$).
- Los reclamos en calificaciones se deben hacer de acuerdo a lo estipulado en el reglamento de estudiantes.

4. Contenido

1. Introducción al problema de RL ([Sutton and Barto, 2018], capítulo 1).
2. El dilema entre exploración y explotación, multi-armed bandits ([Sutton and Barto, 2018], capítulo 2).
3. Procesos de decisión de Markov ([Sutton and Barto, 2018], capítulo 3, [Puterman, 2014]).
4. Métodos de solución tabulares: Programación dinámica, métodos de Montecarlo ([Sutton and Barto, 2018], capítulos 3,4).
5. Métodos de solución tabulares: método de diferencia temporal (TD), Q-Learning, SARSA ([Sutton and Barto, 2018], capítulos 6,7).
6. Redes Neuronales Convolucionales[1] [Goodfellow et al., 2016].
7. RL con aproximación de funciones ([Sutton and Barto, 2018], capítulos 9,10).
8. Deep Reinforcement Learning [Mnih et al., 2013, Hasselt et al., 2016, Schaul et al., 2016, Wang et al., 2016].
9. Métodos de búsqueda de política (policy search) ([Sutton and Barto, 2018], capítulos 9,13), [Lillicrap et al., 2019],[Schul .

Referencias

[Goodfellow et al., 2016] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.

[Hasselt et al., 2016] Hasselt, H. v., Guez, A., and Silver, D. (2016). Deep reinforcement learning with double q-learning. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, pages 2094–2100. AAAI Press.

[Lillicrap et al., 2019] Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., and Wierstra, D. (2019). Continuous control with deep reinforcement learning.

[Mnih et al., 2013] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M. (2013). Playing atari with deep reinforcement learning. cite arxiv:1312.5602Comment: NIPS Deep Learning Workshop 2013.

[Puterman, 2014] Puterman, M. L. (2014). *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

[Schaul et al., 2016] Schaul, T., Quan, J., Antonoglou, I., and Silver, D. (2016). Prioritized experience replay. In *International Conference on Learning Representations*, Puerto Rico.

[Schulman et al., 2017] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. (2017). Proximal policy optimization algorithms.

[Sutton and Barto, 2018] Sutton, R. S. and Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. The MIT Press, second edition.

[Wang et al., 2016] Wang, Z., Schaul, T., Hessel, M., Hasselt, H., Lanctot, M., and Freitas, N. (2016). Dueling network architectures for deep reinforcement learning. volume 48 of *Proceedings of Machine Learning Research*, pages 1995–2003, New York, New York, USA. PMLR.

[1] Opcional, dependiendo de la audiencia.