.UBAexactas
FACULTAD DE CIENCIAS EXACTAS Y NATURALES

Licenciatura en Ciencias de Datos
Propuesta de tema de tesis

# Abordando el espacio combinatorio en recomendaciones de ranking mediante la descomposición de acciones en reinforcement learning

*El uso del aprendizaje por refuerzo, también conocido como Reinforcement Learning (RL), cuenta con la capacidad de optimizar métricas a largo plazo, a diferencia de técnicas tradicionales de Machine Learning. Sin embargo, en el caso de recomendaciones de contenido (e.g., recomendaciones de películas de Netflix), donde el espacio de posibles acciones es muy grande, se suelen hacer restricciones en qué se recomienda para reducir el tamaño de dicho espacio. Este estudio propone investigar una nueva representación de los agentes de RL donde, en lugar de limitar el espacio de acciones, se limita el espacio de distribución que puede ser representada sobre todas las acciones.*

Palabras clave: `Machine Learning, Reinforcement Learning, Recommender Systems, Learning to Rank`

## Conocimientos deseables

*Estadística, Reinforcement Learning, Machine Learning, Python, PyTorch*

## ¿Qué podría aprender quien realice esta tesis?

*El estudiante aprenderá técnicas de recomendaciones usadas actualmente en la industria y sus limitaciones. Habrá un énfasis en las técnicas basadas en reinforcement learning.*

---

## Dirección de la tesis

*Garcia, Francisco*
*Roku | Search and Recommendations*

Contacto: *fmaxgarcia@gmail.com*

Más información en el pdf a continuación.

# Abordando el espacio combinatorio en recomendaciones de ranking mediante la descomposición de acciones en reinforcement learning

September 25, 2023

## 1 Introducción

En los sistemas de recomendación es común encontrar situaciones en el que es necesario elegir un subset de elementos para mostrar a un usuario proveniente de un set significativamente más grande; por ejemplo, un sitio de streaming recomendando diez películas sobre un total de miles de películas disponibles.

En estos escenarios de recomendación, el uso de agentes de reinforcement learning (RL) ha demostrado ser capaz de optimizar metricas con horizontes más largos que simplemente buscar el contenido que llevaría al óptimo immediate. Considere la decisión de mostrarle un contenido nuevo a un usuario; si se le recomienda una nueva película podría llevar a dos horas de streaming hoy, sin embargo, si se le recomienda una nueva serie podría llevar a una hora de streaming hoy, pero diez horas de streaming en el transcurso de una semana [IJW+19].

Para lidiar con estos casos, los agentes de RL típicamente son modelados en una de las siguientes formas: 1) el agente elige un elemento a la vez de manera recursiva [RT22], 2) el numero de posibles elementos es limitado a un set pequeño y el agente elige entre los posibles rankings de esos elementos [DAEvH+16, WFS+20], o 3) los elementos son rankeados de manera independiente sin tener en consideración la posición de otros elementos [LCLS10, CBJ+23].

## 2 Propuesta

En este proyecto se propone el desarrollo de un algoritmo del estilo "actor-critic" que sea capaz de hacer un ranking de manera escalable dónde el número de elementos a rankear no sea una limitación. Para llegar a ese objetivo, se propone representar el set de elementos a recomendar de una manera invariante y descomponer las acciones del agente en una secuencia de comparaciones de elementos. De esta manera, el agente podrá representar todos los posibles rankings del set de elementos al limitar la clase de distribución que pueda ser representada. Más formalmente, dado un set $\mathcal{X} = \{x_1, \ldots, x_N\}$ de $N$ elmentos, buscamos obtener una secuencia ordenada $Y = (y_1, \ldots, y_N)$, dónde $y_i \in \mathcal{X}$. El número total de rankings del set en cuestión es $N!$, pero dado evaluaciones parciales de los rankings, podríamos determinar que ciertos secuencias son más similares que otras, lo cuál permitiría la evaluación de un ranking en particular que no se ha encontrado en los datos de entrenamiento en base a otros rankings similares.

En esta propuesta se espera que el estudiante contribuya con la implementación y ejecución de testeos, el desarrollo de la teoría necesaria para mostrar los pros y contras de la propuesta, y (dependiendo los resultados) ayude con la preparación de un paper para presentar en una conferencia o journal de índole internacional.

## References

[CBJ+23] Tianchi Cai, Shenliao Bao, Jiyan Jiang, Shiji Zhou, Wenpeng Zhang, Lihong Gu, Jinjie Gu, and Guannan Zhang. Model-free reinforcement learning with stochastic reward sta-

bilization for recommender systems. *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2023.

[DAEvH+16] Gabriel Dulac-Arnold, Richard Evans, Hado van Hasselt, Peter Sunehag, Timothy Lillicrap, Jonathan Hunt, Timothy Mann, Theophane Weber, Thomas Degris, and Ben Coppin. Deep reinforcement learning in large discrete action spaces, 2016.

[IJW+19] Eugene Ie, Vihan Jain, Jing Wang, Sanmit Narvekar, Ritesh Agarwal, Rui Wu, Heng-Tze Cheng, Tushar Chandra, and Craig Boutilier. Slateq: A tractable decomposition for reinforcement learning with recommendation sets. In *International Joint Conference on Artificial Intelligence*, 2019.

[LCLS10] Lihong Li, Wei Chu, John Langford, and Robert E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*. ACM, apr 2010.

[RT22] Mehrdad Rezaei and Nasseh Tabrizi. Recommender system using reinforcement learning: A survey. In *Delta*, 2022.

[WFS+20] Fengsheng Wei, Gang Feng, Yao Sun, Yatong Wang, Shuang Qin, and Ying-Chang Liang. Network slice reconfiguration by exploiting deep reinforcement learning with large action space. *IEEE Transactions on Network and Service Management*, 17:2197–2211, 2020.