Guía docente
270431 - ARA - Aprendizaje por Refuerzo Avanzado

Última modificación: 04/02/2026

Unidad responsable: Facultad de Informática de Barcelona
Unidad que imparte: 723 - CS - Departamento de Ciencias de la Computación.

Titulación: GRADO EN INTELIGENCIA ARTIFICIAL (Plan 2021). (Asignatura optativa).

Curso: 2025 **Créditos ECTS:** 6.0 **Idiomas:** Catalán

PROFESORADO

Profesorado responsable:

Otros:

CAPACIDADES PREVIAS

Conocimientos básicos de Deep Learning y de Aprendizaje por Refuerzo (haber cursado APRNS)

COMPETENCIAS DE LA TITULACIÓN A LAS QUE CONTRIBUYE LA ASIGNATURA

Específicas:
CE18. Adquirir y desarrollar técnicas de aprendizaje computacional y diseñar e implementar aplicaciones y sistemas que las utilicen, incluyendo las dedicadas a extracción automática de información y conocimiento a partir de grandes volúmenes de datos.
CE19. Utilizar los sistemas de computación actuales, incluidos sistemas de alto rendimiento, para el proceso de grandes volúmenes de datos desde el conocimiento de su estructura, funcionamiento y particularidades.
CE22. Representar, diseñar y analizar sistemas dinámicos. Adquirir conceptos como su observabilidad, estabilidad y controlabilidad.

Genéricas:
CG2. Utilizar los conocimientos fundamentales y metodologías de trabajo sólidas adquiridos durante los estudios para adaptarse a los nuevos escenarios tecnológicos del futuro.
CG4. Razonar, analizando la realidad y diseñando algoritmos y formulaciones que la modelen. Identificar problemas y construir soluciones algorítmicas o matemáticas válidas, eventualmente nuevas, integrando el conocimiento multidisciplinar necesario, valorando distintas alternativas con espíritu crítico, justificando las decisiones tomadas, interpretando y sintetizando los resultados en el contexto del dominio de aplicación y estableciendo generalizaciones metodológicas a partir de aplicaciones concretas.
CG9. Afrontar nuevos retos con una visión amplia de las posibilidades de la carrera profesional en el ámbito de la Inteligencia Artificial. Desarrollar la actividad aplicando criterios de calidad y mejora continua, y actuar con rigor en el desarrollo profesional. Adaptarse a los cambios organizativos o tecnológicos. Trabajar en situaciones de carencia de información y/o con restricciones temporales y/o de recursos.

Transversales:
CT6. Aprendizaje autónomo. Detectar deficiencias en el propio conocimiento y superarlas mediante la reflexión crítica y la elección de la mejor actuación para ampliar dicho conocimiento.

Básicas:
CB5. Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

METODOLOGÍAS DOCENTES

Las clases están divididas en sesiones de teoría, problemas y laboratorio.

En las sesiones de teoría se desarrollarán los conocimientos de la asignatura, intercalando la exposición de nuevo material teórico con ejemplos y la interacción con los alumnos para discutir los conceptos.

En las clases de laboratorio se desarrollarán pequeñas prácticas utilizando herramientas y utilizando librerías específicas que permitirán practicar y reforzar los conocimientos de las clases de teoría.

OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

1.Conocer los problemas al diseñar funciones de refuerzo y cómo resolverlos
2.Aprender técnicas para acelerar el aprendizaje por reforzamiento de forma que sea viable en aplicaciones reales.
3.Entender la problemática del aprendizaje simultáneo en sistemas multiagentes y las técnicas que permiten este aprendizaje
4.Aprendrez cómo incorporar el aprendizaje a partir de ejemplos para obtener políticas mejores que las que generan los ejemplos y para obtener la función de refuerzo oculta y detrás de estos ejemplos.

HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTADO

Tipo	Horas	Porcentaje
Horas aprendizaje autónomo	90,0	60.00
Horas grupo grande	30,0	20.00
Horas grupo pequeño	30,0	20.00

Dedicación total: 150 h

CONTENIDOS

Diseño de la función de refuerzo: Aprendizaje por reforzamiento inverso (IRL)

Descripción:
La función de refuerzo es crucial en RL, pero no siempre es fácil de definir. En este tema se ve cómo obtener una función de refuerzo a partir de ejemplos de comportamiento.

Aprendiendo la función de refuerzo con un humano en el Loop (RLHF)

Descripción:
En la definición de funciones de refuerzo complejas poder no tenemos ejemplos de comportamientos para aplicar IRL. En este caso veremos cómo crear la función de refuerzo con el feedback de humanos (RLHF). Este mecanismo es la base del entrenamiento para alinear modelos del lenguaje como ChatGPT y otros.

Aprendizaje por reforzamiento ayudado con el aprendizaje de modelos del mundo.

Descripción:
El aprendizaje por reforzamiento es lento. Para reducir el número de interacciones con el entorno, una posibilidad es aprender un modelo predictivo del entorno a partir de las interacciones con él y así poder generar experiencias simuladas de las que se podría aprender sin interaccionar tanto con el mundo real. En este tema se ve esa aproximación y sus limitaciones.

Exploración básica y avanzada en RL: implementando la curiosidad

Descripción:
Un elemento básico en RL es la exploración que permite encontrar mejores políticas. Los métodos básicos de exploración consisten en tomar acciones aleatorias, lo que conduce a ineficiencia y ralentización en el aprendizaje. Hay mejores formas de explorar nuevas opciones y en este tema se describen, pasando por determinación de incertidumbre en el conocimiento aprendido hasta la implementación de método de curiosidad para mejorar la exploración.

Aprendizaje en sistemas Multiagentes utilizando RL

Descripción:
En RL se presupone que el entorno es markoviano y que, por tanto, los cambios en el entorno sólo se producen por acciones del agente que aprende. Cuando el agente aprende en un entorno donde otros agentes también actúan y aprenden, esta condición ya no se cumple y los algoritmos de RL deben adaptarse. En este tema se ven los métodos más avanzados de aprendizaje por reforzamiento en sistemas multiagentes, poniendo especial énfasis en los problemas cooperativos.

Competición en sistemas multiagente utilizando RL: AlfaGo y familia

Descripción:
Un caso especial de interacción en sistemas multiagentes es la competición y, en especial, los juegos de suma cero. En este escenario, el aprendizaje por reforzamiento ha llevado al desarrollo de habilidades sobrehumanas en algunos casos, remarcablemente el caso del juego del Go. En este tema veremos las técnicas de self-play y MonteCarlo Tree Search que permiten desarrollar estas habilidades.

RL en funciones esparsas de refuerzo: Políticas condicionadas y hindsight

Descripción:
A menudo en RL la función de refuerzo es dispersa (poco informativa). Esto tiene como ventaja que las políticas obtenidas no son sesgadas, pero ralentiza el aprendizaje. En este tema se estudian las políticas condicionadas al objetivo y la técnica de hindsight que se han demostrado muy efectivas para acelerar el aprendizaje en estos casos.

Aprendizaje por refuerzo off-line

Descripción:
En algunas aplicaciones tenemos ejemplos de comportamientos generados por humanos o por otras políticas. Una posibilidad para aprovechar estos datos es realizar aprendizaje por imitación o aplicar IRL para aprender de ejemplos. Ahora bien, la política obtenida será como muy tan buena como la que genera los ejemplos. ¿Podemos obtener políticas mejores que las que generan los ejemplos utilizando RL? Off-line RL aprovecha la calidad de los métodos Off-policy para obtener políticas buenas no con los datos que ella misma genera sino con datos posiblemente subóptimos generados por otras políticas (los ejemplos)

Aprendizaje curricular y jerárquico

Descripción:
En RL es a menudo complicado aprender tareas complejas desde cero. Una aproximación, alineada con cómo los humanos aprendemos, consiste en definir un currículum o jerarquía de tareas para aprender de forma inicial antes de intentar aprender la tarea compleja por la que el agente no está preparado. En este tema se verá cómo realizar curriculum learning y aprendizaje jerárquico en estos casos.

Transfer learning, Meta learning, Lifelong learning y AGI

Descripción:
RL es una aproximación interesante al aprendizaje autónomo por agentes inteligentes. Sin embargo, por su naturaleza es enfocado a tareas concretas cuando es sabido que un agente inteligente debe resolver diferentes tareas. En este tema se plantea la interacción entre diferentes tareas que deben aprenderse respecto a la transferencia de conocimiento de una a otra (Transfer learning), el aprendizaje de tareas para mejorar el aprendizaje en tareas posteriores (Meta- Learning) y, por último, respecto al mantenimiento del conocimiento aprendido durante la vida del agente (Life-long learning). Veremos cómo todas estas técnicas podrían empoderar al agente y permitir una auténtica Inteligencia Artificial General (AGI).

ACTIVIDADES

Repaso rápido de los fundamentos, teoría y algoritmo de aprendizaje por reforzamiento

Dedicación: 10h
Grupo grande/Teoría: 2h
Grupo pequeño/Laboratorio: 2h
Aprendizaje autónomo: 6h

Diseño de la función de refuerzo: Aprendizaje por reforzamiento inverso (IRL)

Objetivos específicos:
1

Competencias relacionadas:
CG4. Razonar, analizando la realidad y diseñando algoritmos y formulaciones que la modelen. Identificar problemas y construir soluciones algorítmicas o matemáticas válidas, eventualmente nuevas, integrando el conocimiento multidisciplinar necesario, valorando distintas alternativas con espíritu crítico, justificando las decisiones tomadas, interpretando y sintetizando los resultados en el contexto del dominio de aplicación y estableciendo generalizaciones metodológicas a partir de aplicaciones concretas.
CE18. Adquirir y desarrollar técnicas de aprendizaje computacional y diseñar e implementar aplicaciones y sistemas que las utilicen, incluyendo las dedicadas a extracción automática de información y conocimiento a partir de grandes volúmenes de datos.
CB5. Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

Dedicación: 10h
Grupo grande/Teoría: 2h
Grupo pequeño/Laboratorio: 2h
Aprendizaje autónomo: 6h

Aprendiendo la función de refuerzo con un humano en el Loop

Objetivos específicos:
1

Competencias relacionadas:
CG4. Razonar, analizando la realidad y diseñando algoritmos y formulaciones que la modelen. Identificar problemas y construir soluciones algorítmicas o matemáticas válidas, eventualmente nuevas, integrando el conocimiento multidisciplinar necesario, valorando distintas alternativas con espíritu crítico, justificando las decisiones tomadas, interpretando y sintetizando los resultados en el contexto del dominio de aplicación y estableciendo generalizaciones metodológicas a partir de aplicaciones concretas.
CE18. Adquirir y desarrollar técnicas de aprendizaje computacional y diseñar e implementar aplicaciones y sistemas que las utilicen, incluyendo las dedicadas a extracción automática de información y conocimiento a partir de grandes volúmenes de datos.
CB5. Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

Dedicación: 10h
Grupo grande/Teoría: 2h
Grupo pequeño/Laboratorio: 2h
Aprendizaje autónomo: 6h

Aprendizaje por reforzamiento off-line

Objetivos específicos:
4

Competencias relacionadas:
CG4. Razonar, analizando la realidad y diseñando algoritmos y formulaciones que la modelen. Identificar problemas y construir soluciones algorítmicas o matemáticas válidas, eventualmente nuevas, integrando el conocimiento multidisciplinar necesario, valorando distintas alternativas con espíritu crítico, justificando las decisiones tomadas, interpretando y sintetizando los resultados en el contexto del dominio de aplicación y estableciendo generalizaciones metodológicas a partir de aplicaciones concretas.
CG2. Utilizar los conocimientos fundamentales y metodologías de trabajo sólidas adquiridos durante los estudios para adaptarse a los nuevos escenarios tecnológicos del futuro.
CE19. Utilizar los sistemas de computación actuales, incluidos sistemas de alto rendimiento, para el proceso de grandes volúmenes de datos desde el conocimiento de su estructura, funcionamiento y particularidades.

Dedicación: 10h
Grupo grande/Teoría: 2h
Grupo pequeño/Laboratorio: 2h
Aprendizaje autónomo: 6h

RL en funciones esparsas de refuerzo: Conditioned policies and hindsight

Objetivos específicos:
2

Competencias relacionadas:
CG9. Afrontar nuevos retos con una visión amplia de las posibilidades de la carrera profesional en el ámbito de la Inteligencia Artificial. Desarrollar la actividad aplicando criterios de calidad y mejora continua, y actuar con rigor en el desarrollo profesional. Adaptarse a los cambios organizativos o tecnológicos. Trabajar en situaciones de carencia de información y/o con restricciones temporales y/o de recursos.
CG2. Utilizar los conocimientos fundamentales y metodologías de trabajo sólidas adquiridos durante los estudios para adaptarse a los nuevos escenarios tecnológicos del futuro.
CT6. Aprendizaje autónomo. Detectar deficiencias en el propio conocimiento y superarlas mediante la reflexión crítica y la elección de la mejor actuación para ampliar dicho conocimiento.
CB5. Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

Dedicación: 10h
Grupo grande/Teoría: 2h
Grupo pequeño/Laboratorio: 2h
Aprendizaje autónomo: 6h

Aprendizaje por reforzamiento ayudado con el aprendizaje de modelos

Objetivos específicos:
2

Competencias relacionadas:
CG9. Afrontar nuevos retos con una visión amplia de las posibilidades de la carrera profesional en el ámbito de la Inteligencia Artificial. Desarrollar la actividad aplicando criterios de calidad y mejora continua, y actuar con rigor en el desarrollo profesional. Adaptarse a los cambios organizativos o tecnológicos. Trabajar en situaciones de carencia de información y/o con restricciones temporales y/o de recursos.
CG2. Utilizar los conocimientos fundamentales y metodologías de trabajo sólidas adquiridos durante los estudios para adaptarse a los nuevos escenarios tecnológicos del futuro.
CT6. Aprendizaje autónomo. Detectar deficiencias en el propio conocimiento y superarlas mediante la reflexión crítica y la elección de la mejor actuación para ampliar dicho conocimiento.
CB5. Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

Dedicación: 10h
Grupo grande/Teoría: 2h
Grupo pequeño/Laboratorio: 2h
Aprendizaje autónomo: 6h

Armed bandits

Objetivos específicos:
2

Competencias relacionadas:
CG9. Afrontar nuevos retos con una visión amplia de las posibilidades de la carrera profesional en el ámbito de la Inteligencia Artificial. Desarrollar la actividad aplicando criterios de calidad y mejora continua, y actuar con rigor en el desarrollo profesional. Adaptarse a los cambios organizativos o tecnológicos. Trabajar en situaciones de carencia de información y/o con restricciones temporales y/o de recursos.
CG2. Utilizar los conocimientos fundamentales y metodologías de trabajo sólidas adquiridos durante los estudios para adaptarse a los nuevos escenarios tecnológicos del futuro.
CT6. Aprendizaje autónomo. Detectar deficiencias en el propio conocimiento y superarlas mediante la reflexión crítica y la elección de la mejor actuación para ampliar dicho conocimiento.
CB5. Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

Dedicación: 10h
Grupo grande/Teoría: 2h
Grupo pequeño/Laboratorio: 2h
Aprendizaje autónomo: 6h

Control primera parte del curso

Dedicación: 2h
Actividades dirigidas: 2h

Exploración avanzada en RL: implementando la curiosidad

Objetivos específicos:
2

Competencias relacionadas:
CG9. Afrontar nuevos retos con una visión amplia de las posibilidades de la carrera profesional en el ámbito de la Inteligencia Artificial. Desarrollar la actividad aplicando criterios de calidad y mejora continua, y actuar con rigor en el desarrollo profesional. Adaptarse a los cambios organizativos o tecnológicos. Trabajar en situaciones de carencia de información y/o con restricciones temporales y/o de recursos.
CG2. Utilizar los conocimientos fundamentales y metodologías de trabajo sólidas adquiridos durante los estudios para adaptarse a los nuevos escenarios tecnológicos del futuro.
CT6. Aprendizaje autónomo. Detectar deficiencias en el propio conocimiento y superarlas mediante la reflexión crítica y la elección de la mejor actuación para ampliar dicho conocimiento.
CB5. Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

Dedicación: 16h
Grupo grande/Teoría: 2h
Grupo pequeño/Laboratorio: 4h
Aprendizaje autónomo: 10h

Aprendizaje en sistemas Multiagentes utilizando RL

Objetivos específicos:
3

Competencias relacionadas:
CG4. Razonar, analizando la realidad y diseñando algoritmos y formulaciones que la modelen. Identificar problemas y construir soluciones algorítmicas o matemáticas válidas, eventualmente nuevas, integrando el conocimiento multidisciplinar necesario, valorando distintas alternativas con espíritu crítico, justificando las decisiones tomadas, interpretando y sintetizando los resultados en el contexto del dominio de aplicación y estableciendo generalizaciones metodológicas a partir de aplicaciones concretas.
CE22. Representar, diseñar y analizar sistemas dinámicos. Adquirir conceptos como su observabilidad, estabilidad y controlabilidad.

Dedicación: 18h
Grupo grande/Teoría: 4h
Grupo pequeño/Laboratorio: 4h
Aprendizaje autónomo: 10h

Competición en sistemas multiagente utilizando RL: AlfaGo y familia

Objetivos específicos:
3

Competencias relacionadas:
CG4. Razonar, analizando la realidad y diseñando algoritmos y formulaciones que la modelen. Identificar problemas y construir soluciones algorítmicas o matemáticas válidas, eventualmente nuevas, integrando el conocimiento multidisciplinar necesario, valorando distintas alternativas con espíritu crítico, justificando las decisiones tomadas, interpretando y sintetizando los resultados en el contexto del dominio de aplicación y estableciendo generalizaciones metodológicas a partir de aplicaciones concretas.
CE22. Representar, diseñar y analizar sistemas dinámicos. Adquirir conceptos como su observabilidad, estabilidad y controlabilidad.

Dedicación: 16h
Grupo grande/Teoría: 2h
Grupo pequeño/Laboratorio: 4h
Aprendizaje autónomo: 10h

Aprendizaje curricular y jerarquico

Dedicación: 10h
Grupo grande/Teoría: 2h
Grupo pequeño/Laboratorio: 2h
Aprendizaje autónomo: 6h

Transfer learning, Meta learning, Lifelong learning

Dedicación: 10h
Grupo grande/Teoría: 2h
Grupo pequeño/Laboratorio: 2h
Aprendizaje autónomo: 6h

RL y AGI

Dedicación: 8h
Grupo grande/Teoría: 2h
Aprendizaje autónomo: 6h

Control final

Objetivos específicos:
1, 2, 3, 4

Competencias relacionadas:
CG9. Afrontar nuevos retos con una visión amplia de las posibilidades de la carrera profesional en el ámbito de la Inteligencia Artificial. Desarrollar la actividad aplicando criterios de calidad y mejora continua, y actuar con rigor en el desarrollo profesional. Adaptarse a los cambios organizativos o tecnológicos. Trabajar en situaciones de carencia de información y/o con restricciones temporales y/o de recursos.
CG4. Razonar, analizando la realidad y diseñando algoritmos y formulaciones que la modelen. Identificar problemas y construir soluciones algorítmicas o matemáticas válidas, eventualmente nuevas, integrando el conocimiento multidisciplinar necesario, valorando distintas alternativas con espíritu crítico, justificando las decisiones tomadas, interpretando y sintetizando los resultados en el contexto del dominio de aplicación y estableciendo generalizaciones metodológicas a partir de aplicaciones concretas.
CG2. Utilizar los conocimientos fundamentales y metodologías de trabajo sólidas adquiridos durante los estudios para adaptarse a los nuevos escenarios tecnológicos del futuro.
CE19. Utilizar los sistemas de computación actuales, incluidos sistemas de alto rendimiento, para el proceso de grandes volúmenes de datos desde el conocimiento de su estructura, funcionamiento y particularidades.
CE22. Representar, diseñar y analizar sistemas dinámicos. Adquirir conceptos como su observabilidad, estabilidad y controlabilidad.
CE18. Adquirir y desarrollar técnicas de aprendizaje computacional y diseñar e implementar aplicaciones y sistemas que las utilicen, incluyendo las dedicadas a extracción automática de información y conocimiento a partir de grandes volúmenes de datos.
CT6. Aprendizaje autónomo. Detectar deficiencias en el propio conocimiento y superarlas mediante la reflexión crítica y la elección de la mejor actuación para ampliar dicho conocimiento.
CB5. Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

SISTEMA DE CALIFICACIÓN

La asignatura comprenderá los siguientes actos evaluatorios:

- Informes de las actividades de laboratorio, que será necesario haber entregado dentro de un plazo indicado para cada sesión (orientativamente, 2 semanas). A partir de una media ponderada de las notas de estos informes se calculará una nota de laboratorio, L.

- Un primer examen parcial, realizado a mitad del curso, de la materia vista hasta entonces. Sea P1 la nota obtenida en este examen.

- En el día designado dentro del período de exámenes, un segundo examen parcial de la materia no cubierta por el primer parcial. Sea P2 la nota obtenida en este examen.

Las tres notas L, P1, P2 son entre 0 y 10.

La nota final de la asignatura será:0.4*L +0.3*P1+ 0.3*P2

BIBLIOGRAFÍA

Básica:
- Zai, Alexander; Brown, Brandon. Deep reinforcement learning in action. Shelter Island, NY: Manning Publications Co, 2020. ISBN 9781617295430.
- Lapam, Maxim. Deep Reinforcement Learning Hands-On: apply modern RL methods to practical problems of chatbots, robotics, discrete optimization, web automation, and more [en línea]. 2nd. ed. Birmingham ; Mumbai: Packt Publishing, 2020 [Consulta: 24/02/2025]. Disponible a: https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=6034344. ISBN 9781838820046.
- Albrecht, Stefano V.; Christianos, Filippos; Schäfer, Lukas. Multi-Agent Reinforcement Learning Foundations and Modern Approaches. Cambrige ; London: MIT Press, 2024. ISBN 9780262049375.

Complementaria:
- Bilgin, Enes. Mastering reinforcement learning with Python : build next-generation, self-learning models using reinforcement learning techniques and best practices. Birmingham, England: Packt Publishing, [2020]. ISBN 9781838644147.

RECURSOS

Enlace web:
- https://sites.google.com/upc.edu/ara