

Guia docent

270431 - ARA - Aprenentatge per Reforçament Avançat

Última modificació: 03/02/2025

Unitat responsable: Facultat d'Informàtica de Barcelona
Unitat que imparteix: 723 - CS - Departament de Ciències de la Computació.
Titulació: GRAU EN INTEL·LIGÈNCIA ARTIFICIAL (Pla 2021). (Assignatura optativa).
Curs: 2024 **Crèdits ECTS:** 6.0 **Idiomes:** Català

PROFESSORAT

Professorat responsable: MARIO MARTÍN MUÑOZ
Altres: Segon quadrimestre:
MARIO MARTÍN MUÑOZ - 10

CAPACITATS PRÈVIES

Coneixements bàsics de Deep Learning i d'Aprenentatge per Reforçament (haver cursat APRNS)

COMPETÈNCIES DE LA TITULACIÓ A LES QUALS CONTRIBUEIX L'ASSIGNATURA

Específiques:

CE18. Adquirir i desenvolupar tècniques d'aprenentatge computacional i dissenyar i implementar aplicacions i sistemes que les utilitzin, incloent les dedicades a extracció automàtica d'informació i coneixement a partir de grans volums de dades.
CE19. Utilitzar els sistemes de computació actuals, inclosos sistemes d'alt rendiment, per al procés de grans volums de dades des del coneixement de la seva estructura, funcionament i particularitats.
CE22. Representar, dissenyar i analitzar sistemes dinàmics. Adquirir conceptes com el seu observabilitat, estabilitat i controlabilitat.

Genèriques:

CG2. Utilitzar els coneixements fonamentals i metodologies de treball sòlides adquirits durant els estudis per adaptar-se als nous escenaris tecnològics de el futur.
CG4. Raonar, analitzant la realitat i dissenyant algoritmes i formulacions que la modelin. Identificar problemes i construir solucions algorísmiques o matemàtiques vàlides, eventualment noves, integrant el coneixement multidisciplinari necessari, valorant diferents alternatives amb esperit crític, justificant les decisions preses, interpretant i sintetitzant els resultats en el context de l'domini d'aplicació i establint generalitzacions metodològiques a partir de aplicacions concretes.
CG9. Afrontar nous reptes amb una visió àmplia de les possibilitats de la carrera professional en l'àmbit de la Intel·ligència Artificial. Desenvolupar l'activitat aplicant criteris de qualitat i millora contínua, i actuar amb rigor en el desenvolupament professional. Adaptar-se als canvis organitzatius o tecnològics. Treballar en situacions de carència d'informació i/o amb restriccions temporals i/o de recursos.

Transversals:

CT6. Aprenentatge autònom. Detectar deficiències en el propi coneixement i superar-les mitjançant la reflexió crítica i l'elecció de la millor actuació per ampliar aquest coneixement.

Bàsiques:

CB5. Que els estudiants hagin desenvolupat aquelles habilitats d'aprenentatge necessàries per emprendre estudis posteriors amb un alt grau d'autonomia

METODOLOGIES DOCENTS

Les classes estan dividides en sessions de teoria, problemes i laboratori.

A les sessions de teoria es desenvoluparan els coneixements de l'assignatura, intercalant l'exposició de nou material teòric amb exemples i la interacció amb els alumnes per tal de discutir els conceptes.

A les classes de laboratori es desenvoluparan petites pràctiques utilitzant eines i fent servir llibreries específiques que permetran practicar i reforçar els coneixements de les classes de teoria.

OBJECTIUS D'APRENTATGE DE L'ASSIGNATURA

1. Ser conscient dels problemes en el disseny de funcions de reforçament i com solucionar-los
2. Aprendre tècniques per accelerar l'aprenentatge per reforçament de manera que sigui viable en aplicacions reals.
3. Entendre la problemàtica de l'aprenentatge simultani en sistemes multiagents i les tècniques que permeten aquest aprenentatge
4. Aprendre com incorporar l'aprenentatge a partir d'exemples per obtenir polítiques més bones que les que generen els exemples i per obtenir la funció de reforç oculta i darrera d'aquests exemples.

HORES TOTALES DE DEDICACIÓ DE L'ESTUDIANTAT

Tipus	Hores	Percentatge
Hores grup gran	30,0	20.00
Hores aprenentatge autònom	90,0	60.00
Hores grup petit	30,0	20.00

Dedicació total: 150 h

CONTINGUTS

Disseny de la funció de reforç: Aprenentatge per reforçament invers (IRL)

Descripció:

La funció de reforç és crucial en RL però no sempre és fàcil de definir. En aquest tema es veu com obtenir una funció de reforç a partir d'exemples de comportament.

Aprenent la funció de reforç amb un humà en el Loop (RLHF)

Descripció:

En la definició de funcions de reforç complexes poder no tenir exemples de comportaments per aplicar-hi IRL. En aquest cas veurem com crear la funció de reforç amb el feedback d'humans (RLHF). Aquest mecanisme és la base de l'entrenament per alinear models del llenguatge com ChatGPT i altres.

Aprenentatge per reforçament ajudat amb l'aprenentatge de models del món.

Descripció:

L'aprenentatge per reforçament és lent. Per reduir el nombre d'interaccions amb l'entorn, una possibilitat és aprendre un model predictiu de l'entorn a partir de les interaccions amb ell i així poder generar experiències simulades de les quals es podria aprendre sense interaccionar tant amb el món real. En aquest tema es veu aquesta aproximació i les seves limitacions.

Exploració bàsica i avançada en RL: implementant la curiositat

Descripció:

Un element bàsic en RL és l'exploració que permet trobar millors polítiques. Els mètodes bàsics d'exploració consisteixen en prendre accions aleatòries, la qual cosa condueix a ineficiència i alentiment en l'aprenentatge. Hi ha maneres millors d'explorar noves opcions i en aquest tema es descriuen, passant per determinació d'incertesa en el coneixement après fins a la implementació de mètode de curiositat per millorar l'exploració.

Aprenentatge en sistemes Multiagents fent servir RL

Descripció:

En RL es presuposa que l'entorn es markovià i que, per tant, els canvis en l'entorn només es produeixen per accions de l'agent que aprèn. Quan l'agent aprèn en un entorn on altres agents també actuen i aprenen, aquesta condició ja no es compleix i els algorismes de RL han d'adaptar-se. En aquest tema es veuen els mètodes més avançats d'aprenentatge per reforçament en sistemes multiagents, posant especial èmfasi en els problemes cooperatius.

Competició en sistemes multiagent fent servir RL: AlfaGo i família

Descripció:

Un cas especial d'interacció en sistemes multiagents és la competició i, en especial, els jocs de suma zero. En aquest escenari, l'aprenentatge per reforçament ha portat al desenvolupament d'habilitats sobre-humanes en alguns casos, remarcablement el cas del joc del Go. En aquest tema veurem les tècniques de self-play i de MonteCarlo Tree Search que permeten desenvolupar aquestes habilitats.

RL en funcions esparses de reforç: Polítiques condicionades i hindsight

Descripció:

Sovint en RL la funció de reforç és esparsa (poc informativa). Això té com avantatge que les polítiques obtingudes no són esbiaixades, però alenteix l'aprenentatge. En aquest tema s'estudien les polítiques condicionades a l'objectiu i la tècnica de hindsight que s'han demostrat molt efectives per accelerar l'aprenentatge en aquests casos.

Aprenentatge per reforçament off-line

Descripció:

En algunes aplicacions tenim exemples de comportaments generats per humans o per altres polítiques. Una possibilitat per aprofitar aquestes dades és fer aprenentatge per imitació o aplicar IRL per aprendre d'exemples. Ara bé, la política obtinguda serà com a molt tant bona com la que genera els exemples. Podem obtenir polítiques millors que les que generen els exemples fent servir RL? Off-line RL aprofita la qualitat dels mètodes Off-policy per obtenir polítiques bones no amb les dades que ella mateixa genera sinó amb dades possiblement subòptimes generades per altres polítiques (els exemples)

Aprenentatge curricular i jeràrquic

Descripció:

En RL és sovint complicat aprendre tasques complexes des de zero. Una aproximació, alineada amb com els humans aprenem, consisteix en definir un currículum o jerarquia de tasques per aprendre de forma inicial abans d'intentar aprendre la tasca complexa per la qual l'agent no està preparat. En aquest tema es veurà com fer curriculum learning i aprenentatge jeràrquic en aquests casos.

Transfer learning, Meta learning, Lifelong learning i AGI

Descripció:

RL és una aproximació interessant a l'aprenentatge autònom per agents intel·ligents. Malgrat tot, per la seva naturalesa és enfocat a tasques concretes quan és sabut que un agent intel·ligent ha de resoldre diferents tasques. En aquest tema es planteja la interacció entre diferents tasques que s'han d'aprendre respecte a la transferència de coneixement d'una a una altra (Transfer learning), l'aprenentatge de tasques per millorar l'aprenentatge en tasques posteriors (Meta-Learning) i, finalment, respecte al manteniment del coneixement après durant la vida de l'agent (Life-long learning). Veurem com totes aquestes tècniques podrien empoderar l'agent i permetre una autèntica Intel·ligència Artificial General (AGI).

ACTIVITATS

Repàs ràpid dels fonaments, teoria i algorismes d'aprenentatge per reforçament

Dedicació: 10h

Aprenentatge autònom: 6h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Disseny de la funció de reforç: Aprenentatge per reforçament invers (IRL)

Objectius específics:

1

Competències relacionades:

CB5. Que els estudiants hagin desenvolupat aquelles habilitats d'aprenentatge necessàries per emprendre estudis posteriors amb un alt grau d'autonomia

CE18. Adquirir i desenvolupar tècniques d'aprenentatge computacional i dissenyar i implementar aplicacions i sistemes que les utilitzin, incloent les dedicades a extracció automàtica d'informació i coneixement a partir de grans volums de dades.

CG4. Raonar, analitzant la realitat i dissenyant algorismes i formulacions que la modelin. Identificar problemes i construir solucions algorísmiques o matemàtiques vàlides, eventualment noves, integrant el coneixement multidisciplinari necessari, valorant diferents alternatives amb esperit crític, justificant les decisions preses, interpretant i sintetitzant els resultats en el context de l'àmbit d'aplicació i establint generalitzacions metodològiques a partir de aplicacions concretes.

Dedicació: 10h

Aprenentatge autònom: 6h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Aprenent la funció de reforç amb un humà en el Loop

Objectius específics:

1

Competències relacionades:

CB5. Que els estudiants hagin desenvolupat aquelles habilitats d'aprenentatge necessàries per emprendre estudis posteriors amb un alt grau d'autonomia

CE18. Adquirir i desenvolupar tècniques d'aprenentatge computacional i dissenyar i implementar aplicacions i sistemes que les utilitzin, incloent les dedicades a extracció automàtica d'informació i coneixement a partir de grans volums de dades.

CG4. Raonar, analitzant la realitat i dissenyant algoritmes i formulacions que la modelin. Identificar problemes i construir solucions algorísmiques o matemàtiques vàlides, eventualment noves, integrant el coneixement multidisciplinari necessari, valorant diferents alternatives amb esperit crític, justificant les decisions preses, interpretant i sintetitzant els resultats en el context de l'àmbit d'aplicació i establint generalitzacions metodològiques a partir de aplicacions concretes.

Dedicació: 10h

Aprenentatge autònom: 6h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Aprenentatge per reforçament off-line

Objectius específics:

4

Competències relacionades:

CE19. Utilitzar els sistemes de computació actuals, inclosos sistemes d'alt rendiment, per al procés de grans volums de dades des del coneixement de la seva estructura, funcionament i particularitats.

CG2. Utilitzar els coneixements fonamentals i metodologies de treball sòlides adquirits durant els estudis per adaptar-se als nous escenaris tecnològics de el futur.

CG4. Raonar, analitzant la realitat i dissenyant algoritmes i formulacions que la modelin. Identificar problemes i construir solucions algorísmiques o matemàtiques vàlides, eventualment noves, integrant el coneixement multidisciplinari necessari, valorant diferents alternatives amb esperit crític, justificant les decisions preses, interpretant i sintetitzant els resultats en el context de l'àmbit d'aplicació i establint generalitzacions metodològiques a partir de aplicacions concretes.

Dedicació: 10h

Aprenentatge autònom: 6h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

RL en funcions esparses de reforç: Conditioned policies and hindsight

Objectius específics:

2

Competències relacionades:

CB5. Que els estudiants hagin desenvolupat aquelles habilitats d'aprenentatge necessàries per emprendre estudis posteriors amb un alt grau d'autonomia

CG2. Utilitzar els coneixements fonamentals i metodologies de treball sòlides adquirits durant els estudis per adaptar-se als nous escenaris tecnològics de el futur.

CG9. Afrontar nous reptes amb una visió àmplia de les possibilitats de la carrera professional en l'àmbit de la Intel·ligència Artificial. Desenvolupar l'activitat aplicant criteris de qualitat i millora contínua, i actuar amb rigor en el desenvolupament professional. Adaptar-se als canvis organitzatius o tecnològics. Treballar en situacions de carència d'informació i/o amb restriccions temporals i/o de recursos.

CT6. Aprenentatge autònom. Detectar deficiències en el propi coneixement i superar-les mitjançant la reflexió crítica i l'elecció de la millor actuació per ampliar aquest coneixement.

Dedicació: 10h

Aprenentatge autònom: 6h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Aprenentatge per reforçament ajudat amb l'aprenentatge de models

Objectius específics:

2

Competències relacionades:

CB5. Que els estudiants hagin desenvolupat aquelles habilitats d'aprenentatge necessàries per emprendre estudis posteriors amb un alt grau d'autonomia

CG2. Utilitzar els coneixements fonamentals i metodologies de treball sòlides adquirits durant els estudis per adaptar-se als nous escenaris tecnològics de el futur.

CG9. Afrontar nous reptes amb una visió àmplia de les possibilitats de la carrera professional en l'àmbit de la Intel·ligència Artificial. Desenvolupar l'activitat aplicant criteris de qualitat i millora contínua, i actuar amb rigor en el desenvolupament professional. Adaptar-se als canvis organitzatius o tecnològics. Treballar en situacions de carència d'informació i/o amb restriccions temporals i/o de recursos.

CT6. Aprenentatge autònom. Detectar deficiències en el propi coneixement i superar-les mitjançant la reflexió crítica i l'elecció de la millor actuació per ampliar aquest coneixement.

Dedicació: 10h

Aprenentatge autònom: 6h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Armed bandits

Objectius específics:

2

Competències relacionades:

CB5. Que els estudiants hagin desenvolupat aquelles habilitats d'aprenentatge necessàries per emprendre estudis posteriors amb un alt grau d'autonomia

CG2. Utilitzar els coneixements fonamentals i metodologies de treball sòlides adquirits durant els estudis per adaptar-se als nous escenaris tecnològics de el futur.

CG9. Afrontar nous reptes amb una visió àmplia de les possibilitats de la carrera professional en l'àmbit de la Intel·ligència Artificial. Desenvolupar l'activitat aplicant criteris de qualitat i millora contínua, i actuar amb rigor en el desenvolupament professional. Adaptar-se als canvis organitzatius o tecnològics. Treballar en situacions de carència d'informació i/o amb restriccions temporals i/o de recursos.

CT6. Aprenentatge autònom. Detectar deficiències en el propi coneixement i superar-les mitjançant la reflexió crítica i l'elecció de la millor actuació per ampliar aquest coneixement.

Dedicació: 10h

Aprenentatge autònom: 6h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Control primera part del curs

Dedicació: 2h

Activitats dirigides: 2h

Exploració avançada en RL: implementant la curiositat

Objectius específics:

2

Competències relacionades:

CB5. Que els estudiants hagin desenvolupat aquelles habilitats d'aprenentatge necessàries per emprendre estudis posteriors amb un alt grau d'autonomia

CG2. Utilitzar els coneixements fonamentals i metodologies de treball sòlides adquirits durant els estudis per adaptar-se als nous escenaris tecnològics de el futur.

CG9. Afrontar nous reptes amb una visió àmplia de les possibilitats de la carrera professional en l'àmbit de la Intel·ligència Artificial. Desenvolupar l'activitat aplicant criteris de qualitat i millora contínua, i actuar amb rigor en el desenvolupament professional. Adaptar-se als canvis organitzatius o tecnològics. Treballar en situacions de carència d'informació i/o amb restriccions temporals i/o de recursos.

CT6. Aprenentatge autònom. Detectar deficiències en el propi coneixement i superar-les mitjançant la reflexió crítica i l'elecció de la millor actuació per ampliar aquest coneixement.

Dedicació: 16h

Aprenentatge autònom: 10h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 4h

Aprenentatge en sistemes Multiagents fent servir RL

Objectius específics:

3

Competències relacionades:

CE22. Representar, dissenyar i analitzar sistemes dinàmics. Adquirir conceptes com el seu observabilitat, estabilitat i controlabilitat.

CG4. Raonar, analitzant la realitat i dissenyant algoritmes i formulacions que la modelin. Identificar problemes i construir solucions algorísmiques o matemàtiques vàlides, eventualment noves, integrant el coneixement multidisciplinari necessari, valorant diferents alternatives amb esperit crític, justificant les decisions preses, interpretant i sintetitzant els resultats en el context de l' domini d'aplicació i establint generalitzacions metodològiques a partir de aplicacions concretes.

Dedicació: 18h

Aprenentatge autònom: 10h

Grup gran/Teoria: 4h

Grup petit/Laboratori: 4h

Competició en sistemes multiagent fent servir RL: AlfaGo i família

Objectius específics:

3

Competències relacionades:

CE22. Representar, dissenyar i analitzar sistemes dinàmics. Adquirir conceptes com el seu observabilitat, estabilitat i controlabilitat.

CG4. Raonar, analitzant la realitat i dissenyant algoritmes i formulacions que la modelin. Identificar problemes i construir solucions algorísmiques o matemàtiques vàlides, eventualment noves, integrant el coneixement multidisciplinari necessari, valorant diferents alternatives amb esperit crític, justificant les decisions preses, interpretant i sintetitzant els resultats en el context de l' domini d'aplicació i establint generalitzacions metodològiques a partir de aplicacions concretes.

Dedicació: 16h

Aprenentatge autònom: 10h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 4h

Aprenentatge curricular i jeràrquic

Dedicació: 10h

Aprenentatge autònom: 6h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Transfer learning, Meta learning, Lifelong learning

Dedicació: 10h

Aprenentatge autònom: 6h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

RL i AGI

Dedicació: 8h

Aprentatge autònom: 6h

Grup gran/Teoria: 2h

Control final

Objectius específics:

1, 2, 3, 4

Competències relacionades:

CB5. Que els estudiants hagin desenvolupat aquelles habilitats d'aprenentatge necessàries per emprendre estudis posteriors amb un alt grau d'autonomia

CE19. Utilitzar els sistemes de computació actuals, inclosos sistemes d'alt rendiment, per al procés de grans volums de dades des del coneixement de la seva estructura, funcionament i particularitats.

CE18. Adquirir i desenvolupar tècniques d'aprenentatge computacional i dissenyar i implementar aplicacions i sistemes que les utilitzin, incloent les dedicades a extracció automàtica d'informació i coneixement a partir de grans volums de dades.

CE22. Representar, dissenyar i analitzar sistemes dinàmics. Adquirir conceptes com el seu observabilitat, estabilitat i controlabilitat.

CG2. Utilitzar els coneixements fonamentals i metodologies de treball sòlides adquirits durant els estudis per adaptar-se als nous escenaris tecnològics de el futur.

CG9. Afrontar nous reptes amb una visió àmplia de les possibilitats de la carrera professional en l'àmbit de la Intel·ligència Artificial. Desenvolupar l'activitat aplicant criteris de qualitat i millora contínua, i actuar amb rigor en el desenvolupament professional. Adaptar-se als canvis organitzatius o tecnològics. Treballar en situacions de carència d'informació i/o amb restriccions temporals i/o de recursos.

CG4. Raonar, analitzant la realitat i dissenyant algorismes i formulacions que la modelin. Identificar problemes i construir solucions algorísmiques o matemàtiques vàlides, eventualment noves, integrant el coneixement multidisciplinari necessari, valorant diferents alternatives amb esperit crític, justificant les decisions preses, interpretant i sintetitzant els resultats en el context de l'àmbit d'aplicació i establint generalitzacions metodològiques a partir de aplicacions concretes.

CT6. Aprentatge autònom. Detectar deficiències en el propi coneixement i superar-les mitjançant la reflexió crítica i l'elecció de la millor actuació per ampliar aquest coneixement.

SISTEMA DE QUALIFICACIÓ

L'assignatura comprendrà els següents actes avaluatoris:

- Informes de les activitats de laboratori, que caldrà haver lliurat dins un termini indicat per a cada sessió (orientativament, 2 setmanes). A partir d'una mitjana ponderada de les notes d'aquests informes es calcularà una nota de laboratori, L.

- Un primer examen parcial, fet cap a meitat del curs, de la matèria vista fins llavors. Sigui P1 la nota obtinguda en aquest examen.

- En el dia designat dins del període d'exàmens, un segon examen parcial de la matèria no coberta pel primer parcial. Sigui P2 la nota obtinguda en aquest examen.

Les tres notes L, P1, P2 són entre 0 i 10. La nota final de l'assignatura serà:

$$0.4 \cdot L + 0.3 \cdot P1 + 0.3 \cdot P2$$



BIBLIOGRAFIA

Bàsica:

- 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 línia]. 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.

Complementària:

- 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

Enllaç web:

- <https://sites.google.com/upc.edu/ara>. Pàgina web on trobar les transparències i els materials del laboratori