MAI4CAREU

Master programmes in Artificial Intelligence 4 Careers in Europe

Πανεπιστήμιο Κύπρου - Τεχνητή Νοημοσύνη

MAI612 - ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Διάλεξη 16: Εισαγωγή στην Ενισχυτική Μάθηση

Βασίλης Βασιλειάδης, PhD

Χειμερινό Εξάμηνο 2022/23

CYENS CENTRE OF EXCELLENCE

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU

Master programmes in Artificial Intelligence 4 Careers in Europe

Επανάληψη

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ανίχνευση ανωμαλίας

- Η ανίχνευση ανωμαλίας είναι το πρόβλημα της μοντελοποίησης ενός συνόλου δεδομένων φυσιολογικών συμβάντων και της ενεργοποίησης ενός συναγερμού όταν συμβαίνει ένα ασυνήθιστο συμβάν στο μέλλον.
  - Τα κανονικά γεγονότα θεωρείται ότι είναι συγκεντρωμένα
  - Outlier detection: υπάρχουν outliers στο σύνολο της εκπαίδευσης
  - Novelty detection: δεν υπάρχουν outliers στο σετ εκπαίδευσης
- Προσέγγιση:
  - Εκτίμηση πυκνότητας
  - Ταξινόμηση μιας κατηγορίας με χρήση διακριτικών μοντέλων
  - Αυτόματοι κωδικοποιητές
- Εκτίμηση πυκνότητας: Παραμετρική και μη παραμετρική
  - Δημιουργήστε ένα μοντέλο της πιθανότητας των σημείων
  - Χρησιμοποιήστε ένα όριο (ε) σχετικά με την πιθανότητα ταξινόμησης μιας ανωμαλίας (απίστευτο σημείο) από ένα κανονικό σημείο
  - Παραμετρική: τοποθετήστε ένα Gaussian (Παράμετροι: μέση και διακύμανση)
  - Μη παραμετρική: εκτίμηση πυκνότητας πυρήνα

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Ανίχνευση ανωμαλίας

- Ταξινόμηση μιας κατηγορίας με χρήση διακριτικών μοντέλων:
  - Δημιουργήστε ένα συντηρητικό όριο απόφασης
  - SVM μίας κατηγορίας: προσπαθήστε να συμπεριλάβετε όλα τα (κανονικά) δεδομένα κατάρτισης χρησιμοποιώντας τη μικρότερη υπερσφαίρα
  - Δάση απομόνωσης: οι ανωμαλίες είναι σημεία δεδομένων που έχουν μικρά μήκη διαδρομής σε ένα δέντρο
- Αυτόματοι κωδικοποιητές:
  - Τα κανονικά δεδομένα έχουν χαμηλό σφάλμα ανακατασκευής
  - Τα ανώμαλα δεδομένα έχουν υψηλότερο σφάλμα ανακατασκευής
  - Χρησιμοποιήστε ένα ιστόγραμμα των σφαλμάτων και αποφασίστε ένα όριο ανωμαλίας
- Η μηχανική χαρακτηριστικών μπορεί να είναι πολύ σημαντική στα συστήματα ανίχνευσης ανωμαλιών
- Εποπτευόμενη ανίχνευση ανωμαλίας: τρόπος αξιολόγησης των συστημάτων ανίχνευσης ανωμαλιών
  - Να έχετε μικρή ποσότητα δεδομένων με ετικέτα
  - Σύνολο εκμάθησης: κανονικά δεδομένα, χωρίς ετικέτες
  - Βιογραφικό σημείωμα και σύνολα δοκιμών: κανονική + ανώμαλη επισημασμένα δεδομένα
  - Μετρήσεις αξιολόγησης όπως στη δυαδική ταξινόμηση (ρυθμός TP, ακρίβεια, βαθμολογία AUC,...)

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Συστήματα συστάσεων

- Τα συστήματα συστάσεων είναι συστήματα που παρέχουν προτάσεις για στοιχεία που σχετίζονται περισσότερο με ένα συγκεκριμένο χρήστη
- Πρόβλημα ολοκλήρωσης μήτρας:
  - αραιά μήτρα με πολλές ελλείπουσες τιμές
  - πρόβλεψη των τιμών που λείπουν από τους άλλους
- Για παράδειγμα: πρόβλεψη αξιολογήσεων ταινιών (0-5)
  - Σειρά σειρών: ταινίες
  - Οι στήλες: χρήστες
  - Χρησιμοποιήστε διαφορετικό μοντέλο γραμμικής παλινδρόμησης για κάθε χρήστη (π.χ., εάν οι χρήστες είναι 1M, έχουμε 1M μοντέλα)
  - Όταν έχουμε τα χαρακτηριστικά για κάθε ταινία μπορούμε να διαμορφώσουμε μια συνάρτηση κόστους για την εκμάθηση των παραμέτρων όλων των χρηστών χρησιμοποιώντας το gradient descent
    - Εποπτευόμενο πρόβλημα παλινδρόμησης χρησιμοποιώντας την τετραγωνική απώλεια σφάλματος

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Συστήματα συστάσεων

- Collaborative filtering:
  - Προτείνετε στοιχεία με βάση τις βαθμολογίες των χρηστών που έδωσαν παρόμοιες αξιολογήσεις
  - Μη εποπτευόμενο επειδή δεν υποθέτει τη γνώση των χαρακτηριστικών
  - Διαμορφώνει μια συνάρτηση κόστους που μπορεί να χρησιμοποιηθεί για να μάθει ταυτόχρονα τόσο τα χαρακτηριστικά όσο και τις παραμέτρους για κάθε χρήστη χρησιμοποιώντας το gradient descent
  - Για τις δυαδικές ετικέτες μπορούμε να χρησιμοποιήσουμε ένα μοντέλο πρόβλεψης logistic regression και μια δυαδική απώλεια διασταυρούμενης εντροπίας
  - Όταν φτάσει ένας νέος χρήστης, μπορούμε να χρησιμοποιήσουμε τη μέση κανονικοποίηση, έτσι ώστε οι προβλεπόμενες βαθμολογίες του νέου χρήστη να ισούνται με τον μέσο όρο όλων των αξιολογήσεων για κάθε ταινία.
- Φιλτράρισμα με βάση το περιεχόμενο:
  - Σύσταση με βάση τα χαρακτηριστικά του χρήστη και του στοιχείου για την εύρεση καλής αντιστοίχισης
  - Υπολογισμοί ενσωμάτωσης (π.χ. με χρήση NN) από χαρακτηριστικά που πρέπει να είναι του ίδιου μεγέθους
  - Προβλεπόμενη βαθμολόγηση: εσωτετικό γινόμενο των ενσωματώσεων
- Η εύρεση σχετικών αντικειμένων (π.χ. ταινίες που σχετίζονται με την ταινία i) μπορεί να γίνει χρησιμοποιώντας μια k-κοντινότερη αναζήτηση γείτονα (σε λειτουργία ή ενσωματώνοντας χώρο)

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Διάλεξη 16: Εισαγωγή στην Ενισχυτική μάθηση

Μαθησιακά αποτελέσματα

Θα μάθετε για:

1. τι είναι η Ενισχυτική μάθηση (ΕΜ) και πώς διαφέρει από άλλους τύπους μηχανικής μάθησης
2. πώς να επισημοποιήσετε το πρόβλημα ΕΜ
3. τα διαφορετικά συστατικά και κατηγορίες πρακτόρων
4. οι διάφορες εφαρμογές της ΕΜ

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Σύντομο ιστορικό αυτοματισμού

- Πρώτον, η αυτοματοποίηση των επαναλαμβανόμενων φυσικών λύσεων
  - Βιομηχανική επανάσταση (1750-1850) και Εποχή της Μηχανής (1870-1940)
- Δεύτερον, η αυτοματοποίηση των επαναλαμβανόμενων διανοητικών λύσεων
  - Ψηφιακή επανάσταση (1950 — τώρα) και Εποχή της Πληροφορίας
- Επόμενο βήμα: επιτρέπουν στις μηχανές να βρίσκουν οι ίδιες τις λύσεις
  - Τεχνητή νοημοσύνη
  - Πρέπει μόνο να καθορίσετε ένα πρόβλημα ή/και έναν στόχο
  - Απαιτεί να μαθαίνεις αυτόνομα πώς να παίρνεις αποφάσεις

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Τι είναι η τεχνητή νοημοσύνη;

- Θα χρησιμοποιήσουμε τον ακόλουθο ορισμό της ευφυΐας:

Να είναι σε θέση να μάθουν να λαμβάνουν αποφάσεις για την επίτευξη των στόχων

- **Η μάθηση**, **οι αποφάσεις** και **οι στόχοι** είναι όλα κεντρικά

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Τι είναι η Ενισχυτική Μάθηση;

- Οι άνθρωποι και τα ζώα μαθαίνουν αλληλεπιδρώντας με το περιβάλλον μας
- Αυτό διαφέρει από ορισμένα άλλα είδη μάθησης
  - Είναι ενεργητικός και όχι παθητικός
  - Οι αλληλεπιδράσεις είναι συχνά διαδοχικές - οι μελλοντικές αλληλεπιδράσεις μπορούν να εξαρτώνται από προηγούμενες
- Είμαστε στόχοι-κατευθυνόμενοι
- Μπορούμε να μάθουμε χωρίς παραδείγματα βέλτιστης συμπε
- Αντ' αυτού, βελτιστοποιούμε κάποιο σήμα ανταμοιβής
  - Δοκιμή-και-λάθος μάθηση

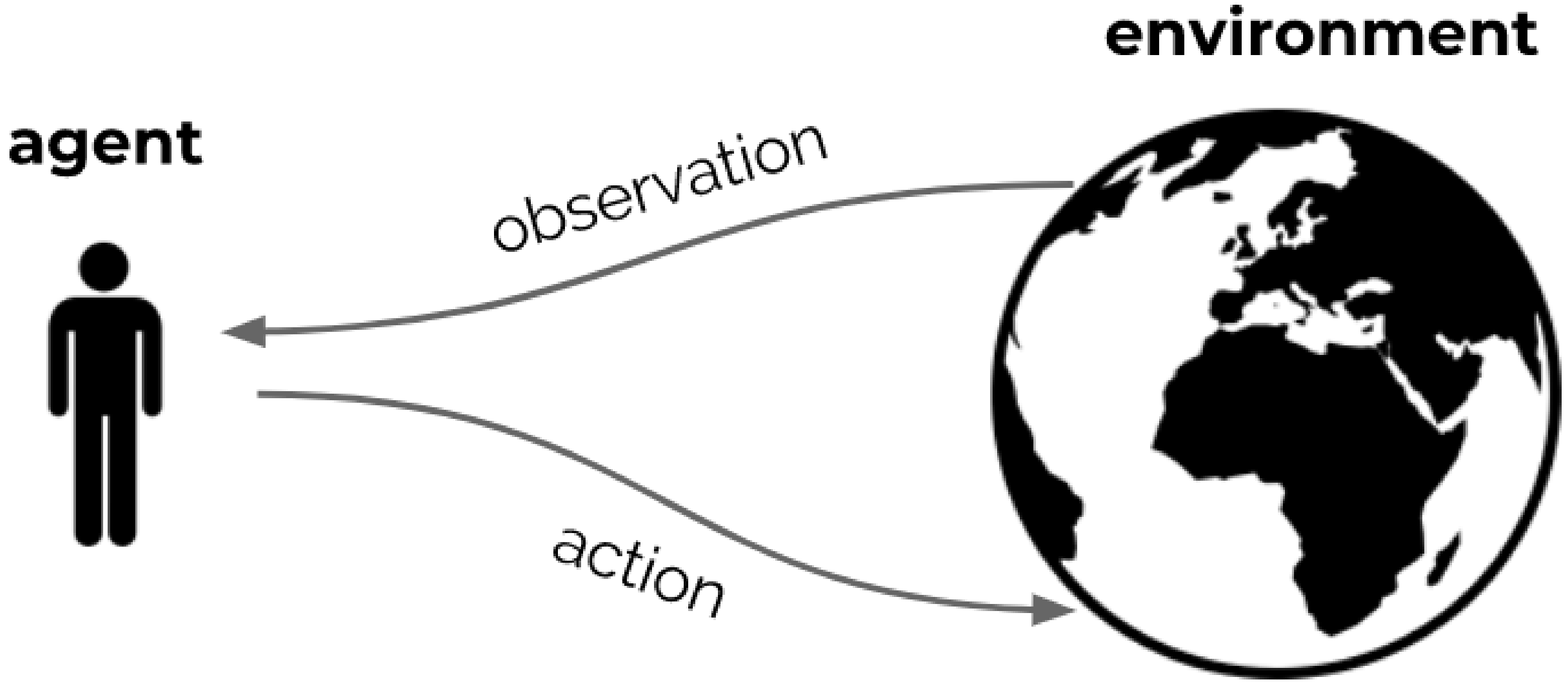
Πηγή εικόνας: ***UC Berkeley CS188 – Εισαγωγή στο μάθημα AI***

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ο βρόγχος αλληλεπίδρασης



Ο στόχος μας: βελτιστοποιήστε το άθροισμα των ανταμοιβών, μέσω επαναλαμβανόμενων αλληλεπιδράσεων

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Η υπόθεση της ανταμοιβής

Η Ενισχυτική Μάθηση βασίζεται στην υπόθεση ανταμοιβής

Οποιοσδήποτε στόχος μπορεί να επισημοποιηθεί ως αποτέλεσμα της μεγιστοποίησης μιας σωρευτικής ανταμοιβής.

Μπορείτε να βρείτε ένα αντιπαράδειγμα;

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Παραδείγματα προβλημάτων ΕΜ

- Πετάξτε ένα ελικόπτερο
- Διαχειριστείτε ένα επενδυτικό χαρτοφυλάκιο
- Έλεγχος ενός σταθμού παραγωγής ενέργειας
- Κάντε ένα ρομπότ να περπατήσει
- Παίξτε βιντεοπαιχνίδια ή επιτραπέζια παιχνίδια

Η επιβράβευση: χρόνος αέρα, αντίστροφη απόσταση,...

Η επιβράβευση: κέρδη, κέρδη μείον κίνδυνο,...

Η επιβράβευση: η αποτελεσματικότητα,...

Η επιβράβευση: απόσταση, ταχύτητα,...

Η επιβράβευση: κερδίστε, μεγιστοποιήστε το σκορ,...

Αν ο στόχος είναι να μάθουμε μέσω της αλληλεπίδρασης, όλα αυτά είναι ενισχυτικά μαθησιακά προβλήματα.

(Ανεξάρτητα από το ποια λύση χρησιμοποιείτε)

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Τι είναι η Ενισχυτική Μάθηση;

Υπάρχουν διακριτοί λόγοι για να μάθετε:

1. Βρείτε λύσεις
   - Ένα πρόγραμμα που παίζει πολύ καλά το σκάκι
   - Ένα ρομπότ κατασκευής με συγκεκριμένο σκοπό
2. Προσαρμογή στο διαδίκτυο, αντιμετώπιση απρόβλεπτων περιστάσεων
   - Ένα πρόγραμμα σκακιού που μπορεί να μάθει να προσαρμόζεται σε σας
   - Ένα ρομπότ που μπορεί να μάθει να πλοηγείται σε άγνωστα εδάφη

- Η ενισχυτική μάθηση μπορεί να παρέχει αλγορίθμους και για τις δύο περιπτώσεις
- Σημειώστε ότι το δεύτερο σημείο δεν είναι (απλά) σχετικά με τη γενίκευση - πρόκειται για τη συνέχιση της αποτελεσματικής μάθησης στο διαδίκτυο, κατά τη διάρκεια της λειτουργίας

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Τι είναι η Ενισχυτική Μάθηση;

- Επιστήμη και πλαίσιο μάθησης για τη λήψη αποφάσεων από την αλληλεπίδραση
- Αυτό απαιτεί από εμάς να σκεφτούμε
  - Χρόνος
  - (μακροπρόθεσμες) συνέπειες των ενεργειών
  - συλλέγοντας ενεργά την εμπειρία
  - προβλέποντας το μέλλον
  - αντιμετώπιση της αβεβαιότητας

- Τεράστιο δυναμικό πεδίο
- Επισημοποίηση του προβλήματος της τεχνητής νοημοσύνης

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Παράδειγμα: Μαθαίνοντας να παίζετε Atari

Πηγή βίντεο

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU

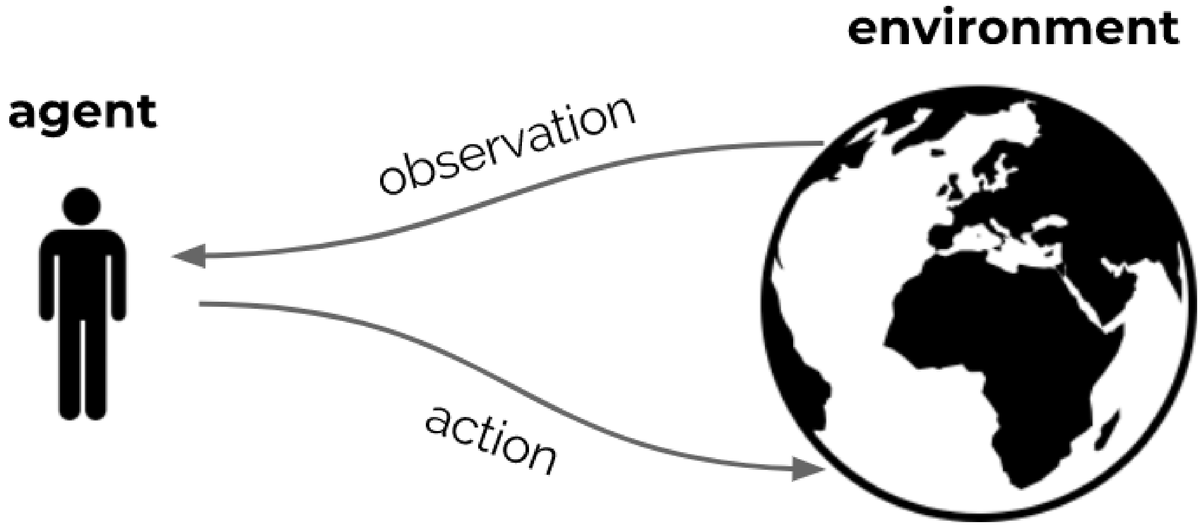
Master programmes in Artificial Intelligence 4 Careers in Europe

Επισημοποίηση του προβλήματος της ΕΜ

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Πράκτορας και Περιβάλλον



- Σε κάθε βήμα *t* ο πράκτορας:
  - Λαμβάνει παρατήρηση O_t (και ανταμοιβή R_t)
  - Εκτελεί τη δράση A_t
- Το περιβάλλον:
  - Λαμβάνει δράση A_t
  - Εκπέμπει παρατήρηση O_{t+1} (και ανταμοιβή R_{t+1})

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ανταμοιβές

- Μια **ανταμοιβή** R_t είναι ένα βαθμωτό σήμα ανατροφοδότησης
- Δείχνει πόσο καλά κάνει ο πράκτορας στο βήμα t — καθορίζει τον στόχο
- Η δουλειά του πράκτορα είναι να μεγιστοποιήσει τη σωρευτική ανταμοιβή

$$G_t = R_{t+1} + R_{t+2} + R_{t+3} + \ldots$$

- Το αποκαλούμε « **απόδοση**»

Η ενίσχυση της μάθησης βασίζεται στην **υπόθεση ανταμοιβής**:

Οποιοσδήποτε στόχος μπορεί να επισημοποιηθεί ως αποτέλεσμα της μεγιστοποίησης μιας σωρευτικής ανταμοιβής.

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Αξία

- Ονομάζουμε την αναμενόμενη αθροιστική ανταμοιβή, από ένα κράτος, την **αξία**

$$v(s) = \mathbb{E}\left[G_t \mid S_t = s\right]$$
$$= \mathbb{E}\left[R_{t+1} + R_{t+2} + R_{t+3} + \ldots \mid S_t = s\right]$$

- Η αξία εξαρτάται από τις ενέργειες που κάνει ο πράκτορας
- Στόχος είναι να **μεγιστοποιηθεί η αξία**, επιλέγοντας κατάλληλες ενέργειες
- Οι ανταμοιβές και οι αξίες καθορίζουν τη **χρησιμότητα** των κρατών και της δράσης (δεν υπάρχει εποπτευόμενη ανατροφοδότηση)
- Οι αποδόσεις και οι τιμές μπορούν να οριστούν αναδρομικά

$$G_t = R_{t+1} + G_{t+1}$$
$$v(s) = \mathbb{E}\left[R_{t+1} + v(S_{t+1}) \mid S_t = s\right]$$

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Μεγιστοποίηση της αξίας με τη λήψη μέτρων

- Ο στόχος μας: **επιλέξτε ενέργειες για να μεγιστοποιήσετε την αξία**
- Οι ενέργειες μπορεί να έχουν μακροπρόθεσμες συνέπειες
- Η ανταμοιβή μπορεί να καθυστερήσει
- Μπορεί να είναι καλύτερο να θυσιάσετε την άμεση ανταμοιβή για να κερδίσετε περισσότερη μακροπρόθεσμη ανταμοιβή
- Για παράδειγμα:
  - Ανεφοδιασμός ελικοπτέρου (μπορεί να αποτρέψει τη συντριβή σε αρκετές ώρες)
  - Αμυντικές κινήσεις σε ένα παιχνίδι (μπορεί να βοηθήσει τις πιθανότητες να κερδίσει αργότερα)
  - Εκμάθηση μιας νέας δεξιότητας (μπορεί να είναι δαπανηρή και χρονοβόρα στην αρχή)
- Μια χαρτογράφηση από τα κράτη στις δράσεις ονομάζεται **πολιτική**

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Τιμές δράσης

- Είναι επίσης δυνατό να καθοριστεί η αξία των **ενεργειών**:

$$q(s,a) = \mathbb{E}\left[G_t \mid S_t = s, A_t = a\right]$$
$$= \mathbb{E}\left[R_{t+1} + R_{t+2} + R_{t+3} + \ldots \mid S_t = s, A_t = a\right]$$

- Θα μιλήσουμε σε βάθος για τις αξίες της κατάστασης και της δράσης αργότερα

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Βασικές έννοιες

Η ενίσχυση του φορμαλισμού μάθησης περιλαμβάνει

- **Περιβάλλον** (δυναμική του προβλήματος)
- **Σήμα ανταμοιβής** (προσδιορίζει το στόχο)
- **Πράκτορας**, που περιέχει:
  - Πολιτεία πρακτόρων
  - Πολιτική
  - Εκτίμηση συνάρτησης τιμής;
  - Το μοντέλο;
- Τώρα θα μπούμε στον πράκτορα

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU

Master programmes in Artificial Intelligence 4 Careers in Europe

Μέσα στον Πράκτορα

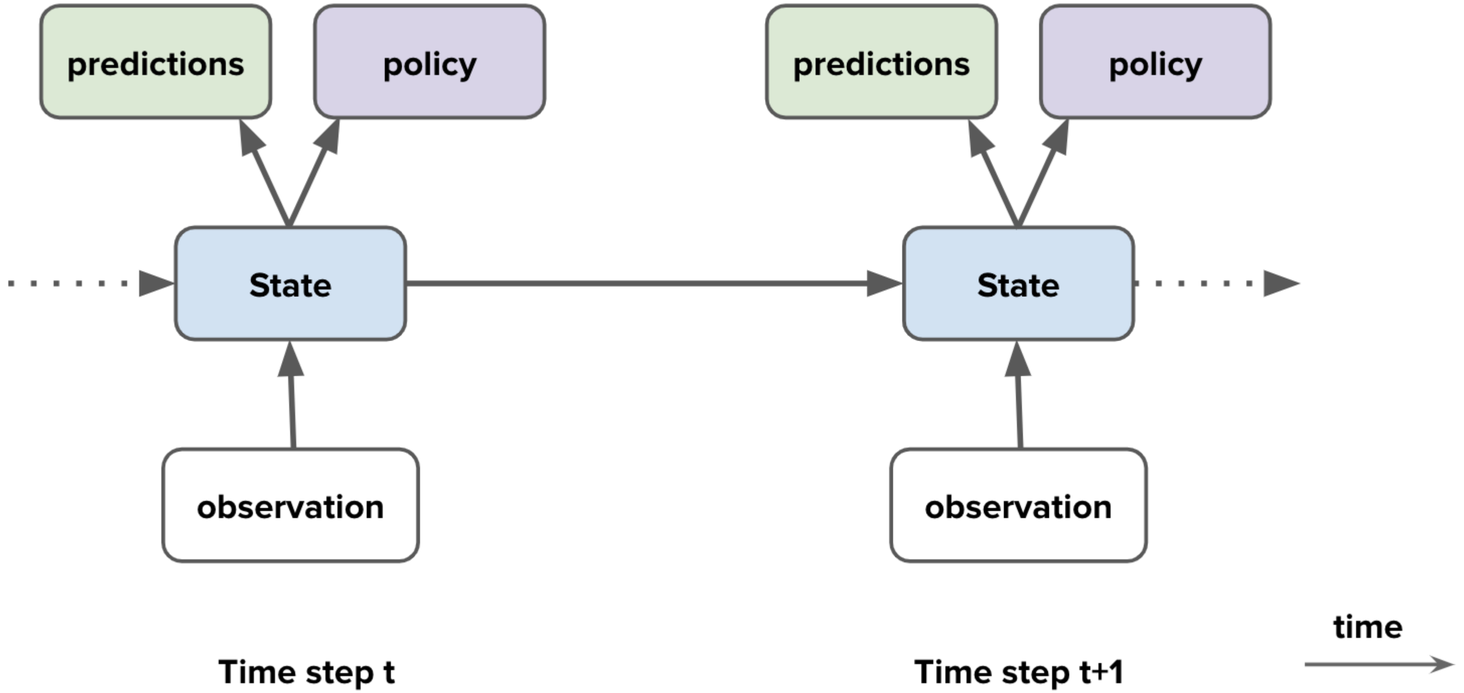
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Μέρη πρακτόρων

Μέρη πρακτόρων

- **Κατάσταση πρακτόρων**
- Πολιτική
- Συνάρτηση αξίας
- Μοντέλο

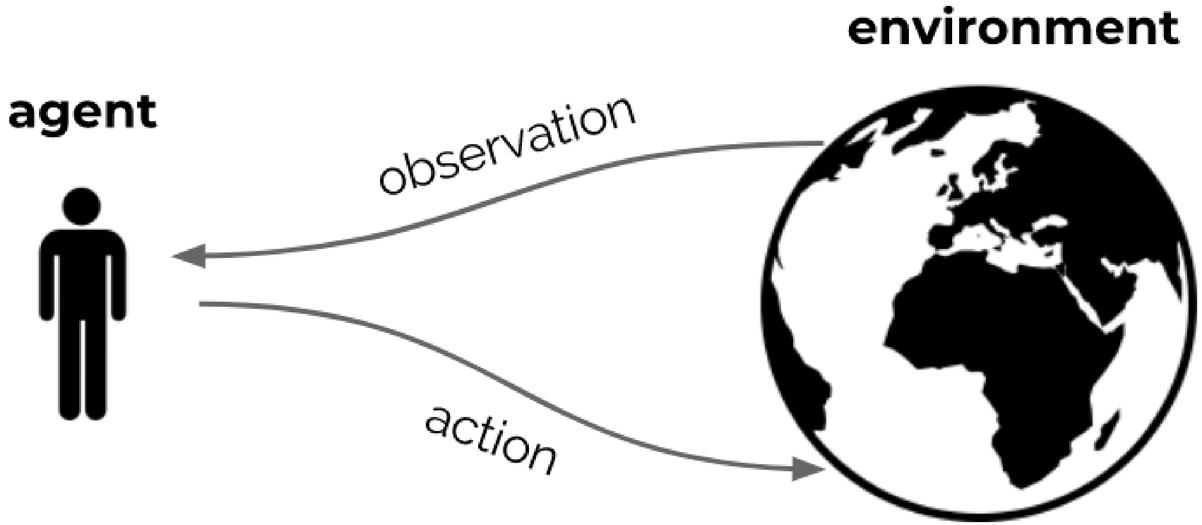


Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Κατάσταση Περιβάλλοντος



- Η **κατάσταση του περιβάλλοντος** είναι η εσωτερική κατάσταση του περιβάλλοντος
- Είναι συνήθως αόρατη για τον πράκτορα
- Ακόμη και αν είναι ορατό, μπορεί να περιέχει πολλές άσχετες πληροφορίες

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Κατάσταση πρακτόρων

- Το **ιστορικό** είναι η πλήρης ακολουθία παρατηρήσεων, δράσεων, ανταμοιβών

$$\mathcal{H}_t = O_0, A_0, R_1, O_1, \ldots, O_{t-1}, A_{t-1}, R_t, O_t$$

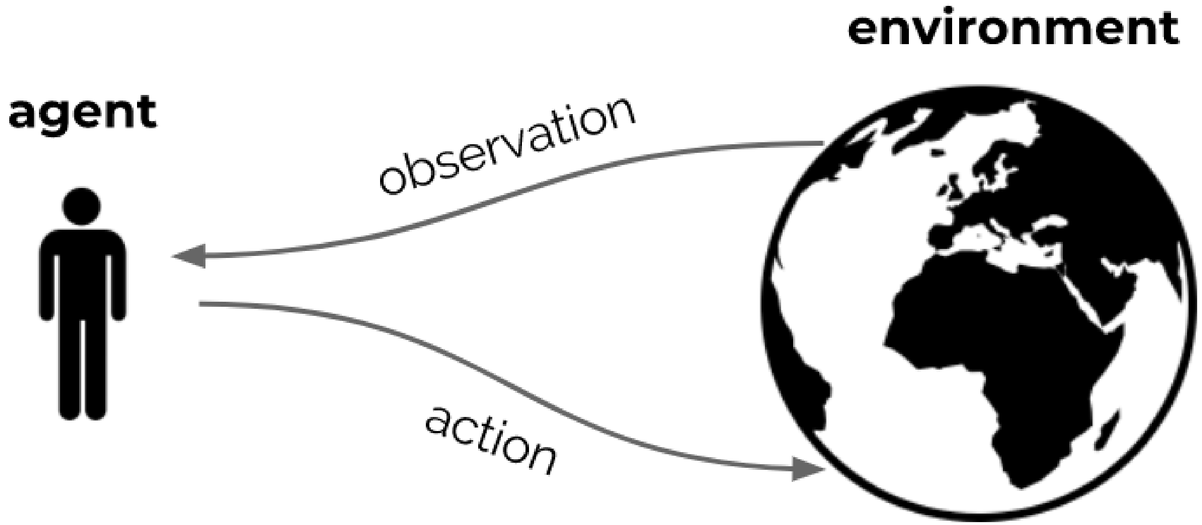
- Για παράδειγμα, η ροή αισθητήρων ενός ρομπότ
- Αυτό το ιστορικό χρησιμοποιείται για την κατασκευή της **κατάστασης πράκτορα** S_t

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Πλήρως παρατηρήσιμα περιβάλλοντα



Πλήρης παρατηρησιμότητα

Ας υποθέσουμε ότι ο πράκτορας βλέπει την πλήρη κατάσταση περιβάλλοντος

- παρατήρηση = κατάσταση περιβάλλοντος
- Το κράτος πράκτορα θα μπορούσε απλά να είναι αυτή η παρατήρηση:

$$S_t = O_t = \text{κατάσταση περιβάλλοντος}$$

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Διαδικασίες λήψης αποφάσεων Markov

Διαδικασίες αποφάσεων Markov (ΔΑΜ) είναι ένα χρήσιμο μαθηματικό πλαίσιο

Ορισμός: Μια διαδικασία λήψης αποφάσεων : $p(r, s \mid S_t, A_t) = p(r, s \mid \mathcal{H}_t, A_t)$

- Αυτό σημαίνει ότι η κατάσταση περιέχει όλα όσα πρέπει να γνωρίζουμε από το ιστορικό.
- Αυτό δεν σημαίνει ότι περιέχει τα πάντα, απλά ότι η προσθήκη περισσότερου ιστορικού δεν βοηθά
  - Μόλις γίνει γνωστή η κατάσταση, το ιστορικό μπορεί να πεταχτεί

- Τυπικά, η κατάσταση παράγοντα S_t είναι κάποια συμπίεση του H_t
- Σημείωση: χρησιμοποιούμε το S_t για να υποδηλώσουμε την **κατάσταση πράκτορα**, όχι την **κατάσταση περιβάλλοντος**

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Μερικώς παρατηρήσιμα περιβάλλοντα

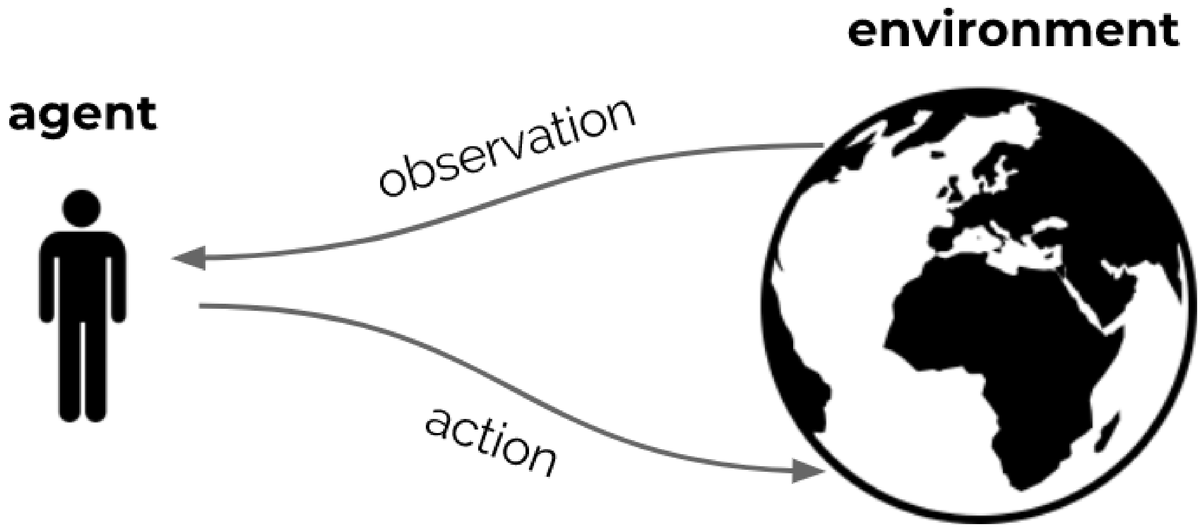
- **Μερική παρατηρησιμότητα:** Οι παρατηρήσεις δεν είναι Μαρκοβιανοί
  - Ένα ρομπότ με όραση κάμερας δεν λέει την απόλυτη τοποθεσία του
  - Ένας πράκτορας παιχνιδιού πόκερ παρατηρεί μόνο δημόσιες κάρτες
- Τώρα, χρησιμοποιώντας την παρατήρηση ως κράτος δεν θα ήταν Μαρκοβιανή.
- Αυτό ονομάζεται **μερικώς παρατηρήσιμη διαδικασία λήψης αποφάσεων Markov** (ΜΠΔΑΜ)
- Η κατάσταση του **περιβάλλοντος** μπορεί ακόμα να είναι ο Markov, αλλά ο πράκτορας δεν το ξέρει.
- Μπορεί να είμαστε ακόμα σε θέση να κατασκευάσουμε ένα Markov Agent State

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Κατάσταση πρακτόρων



- Οι ενέργειες του πράκτορα εξαρτώνται από το κράτος της
- Η **κατάσταση των πρακτόρων** είναι μια λειτουργία του ιστορικού
- Για παράδειγμα, $S_t = O_t$
- Γενικότερα:

$$S_{t+1} = u(S_t, A_t, R_{t+1}, O_{t+1})$$

όπου *u* είναι «συνάρτηση ενημέρωσης κατάστασης»

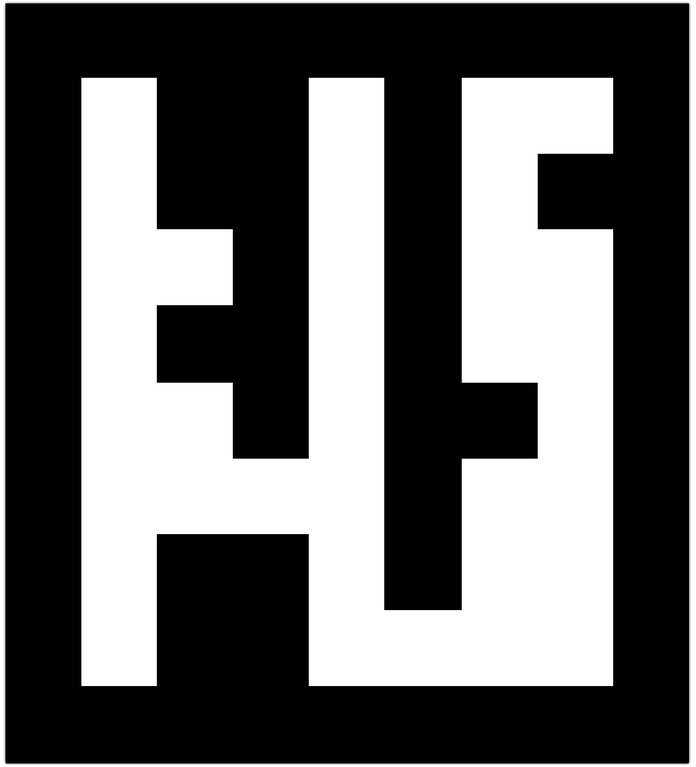
- Η κατάσταση του παράγοντα είναι συχνά **πολύ** μικρότερη από την κατάσταση του περιβάλλοντος

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Κατάσταση πρακτόρων



Η πλήρης κατάσταση του περιβάλλοντος ενός λαβύρινθου

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Κατάσταση πρακτόρων

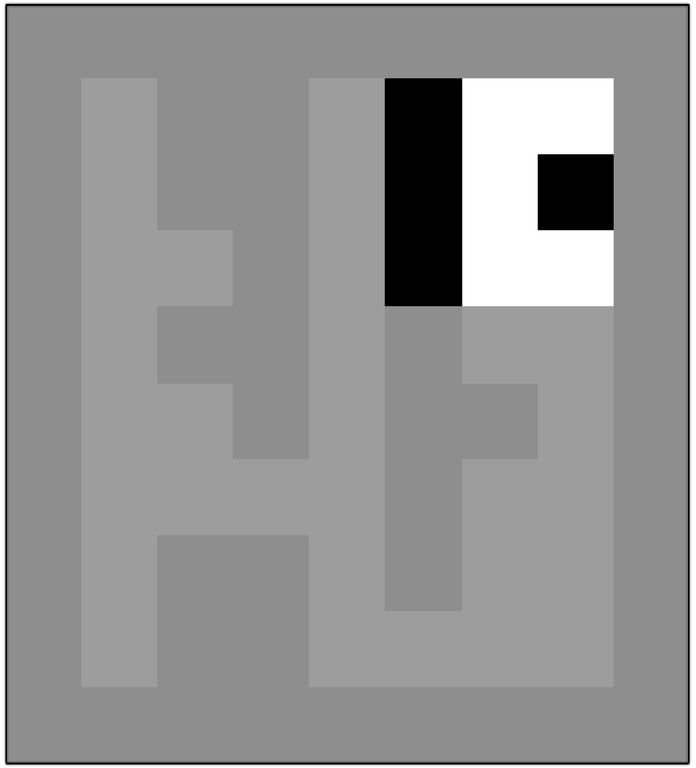
Μια πιθανή παρατήρηση

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Κατάσταση πρακτόρων



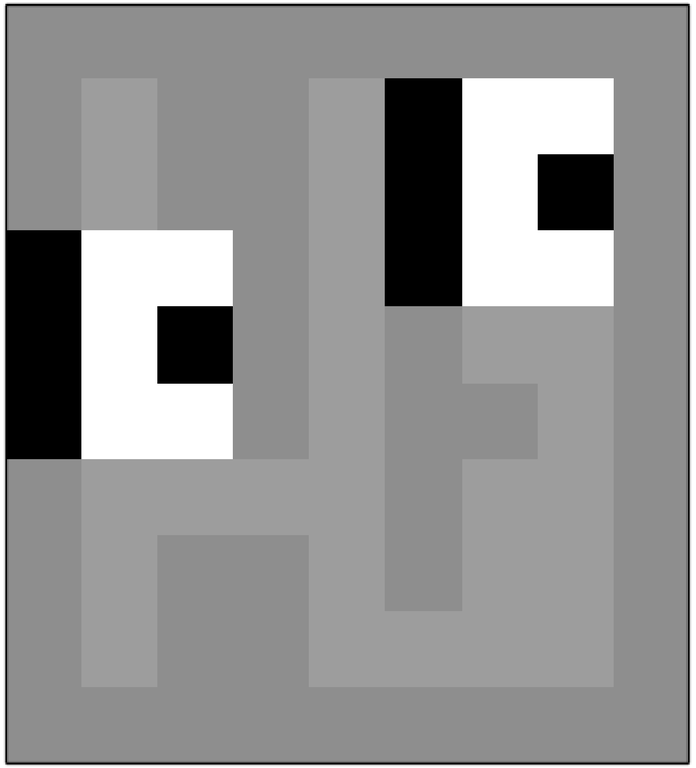
Μια παρατήρηση σε διαφορετική τοποθεσία

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Κατάσταση πρακτόρων



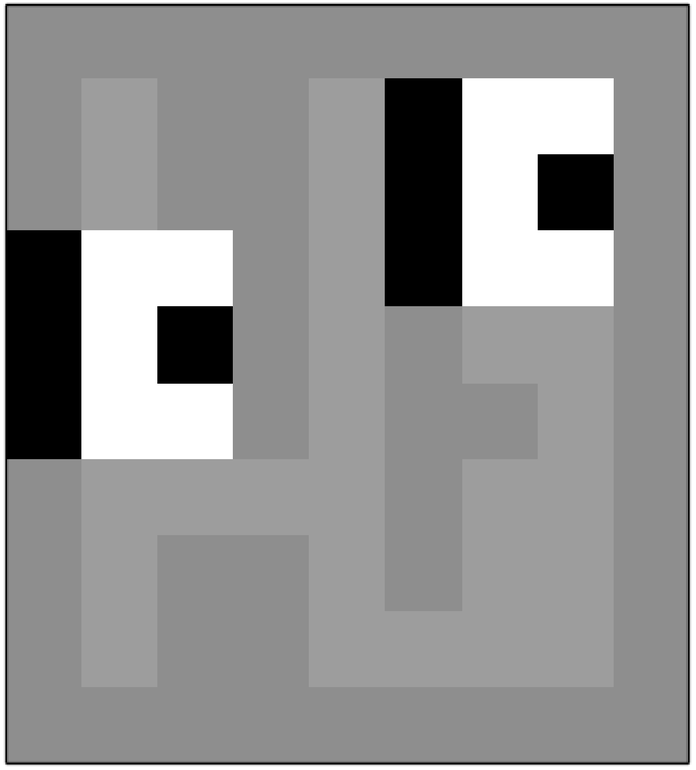
Οι δύο παρατηρήσεις είναι δυσδιάκριτες

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Κατάσταση πρακτόρων



Αυτές οι δύο καταστάσεις δεν είναι Markov.

Πώς θα μπορούσατε να κατασκευάσετε μια κατάσταση πράκτορα Markov σε αυτό το λαβύρινθο (για οποιοδήποτε σήμα ανταμοιβής);

Τι γίνεται αν λάβουμε υπόψη την προηγούμενη παρατήρηση;

Εξαρτάται από τις πολιτικές και τις μεταβάσεις κατάστασης (αν είναι πιθανοτικές).

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Μερικώς παρατηρήσιμα περιβάλλοντα

- Για την αντιμετώπιση της μερικής παρατηρησιμότητας, ο πράκτορας μπορεί να κατασκευάσει κατάλληλες κρατικές αναπαραστάσεις
- Παραδείγματα πρακτόρων αναφέρει:
  - Τελευταία παρατήρηση: $S_t = O_t$ (ίσως να μην είναι αρκετό)
  - Πλήρης ιστορία: $S_t = H_t$ (μπορεί να είναι πολύ μεγάλο)
  - Γενική ενημέρωση: $S_t = u(S_{t-1}, A_{t-1}, R_t, O_t)$ (αλλά πώς να επιλέξετε/μάθετε το u;)
- Η κατασκευή μιας πλήρως Μαρκοβιανής κατάστασης πρακτόρων συχνά δεν είναι εφικτή
- Το πιο σημαντικό είναι ότι η κατάσταση θα πρέπει να επιτρέπει καλές πολιτικές και προβλέψεις αξίας

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Μέρη πρακτόρων

Μέρη πρακτόρων

- Πολιτεία πρακτόρων
- **Πολιτική**
- Συνάρτηση αξίας
- Μοντέλο

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Πολιτική

- Μια **πολιτική** καθορίζει τη συμπεριφορά του πράκτορα
- Είναι ένας χάρτης από κατάσταση πράκτορα στη δράση
- Ντετερμινιστική πολιτική: $A = \pi(S)$
- Στοχαστική πολιτική: $\pi(A \mid S) = p(A \mid S)$

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Μέρη πρακτόρων

Μέρη πρακτόρων

- Πολιτεία πρακτόρων
- Πολιτική
- **Συνάρτηση αξίας**
- Μοντέλο

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Λειτουργία αξίας

- Η πραγματική συνάρτηση τιμής είναι η αναμενόμενη απόδοση

$$v_\pi(s) = \mathbb{E}\left[G_t \mid S_t = s, \pi\right]$$
$$= \mathbb{E}\left[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \ldots \mid S_t = s, \pi\right]$$

- Παρουσιάσαμε έναν **εκπτωτικό παράγοντα** $\gamma \in [0, 1]$
  - Ανταλλάσσει τη σημασία των άμεσων και μακροπρόθεσμων ανταμοιβών
- Η αξία εξαρτάται από μια πολιτική
- Μπορεί να χρησιμοποιηθεί για την αξιολόγηση της σκοπιμότητας των καταστάσεων
- Μπορεί να χρησιμοποιηθεί για την επιλογή μεταξύ ενεργειών

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Λειτουργία αξίας

- Η επιστροφή έχει αναδρομική μορφή $G_t = R_{t+1} + \gamma G_{t+1}$
- Ως εκ τούτου, η αξία έχει επίσης

$$v_\pi(s) = \mathbb{E}\left[R_{t+1} + \gamma G_{t+1} \mid S_t = s, A_t \sim \pi(s)\right]$$
$$= \mathbb{E}\left[R_{t+1} + \gamma v_\pi(S_{t+1}) \mid S_t = s, A_t \sim \pi(s)\right]$$

- Εδώ $A_t \sim \pi(s)$ σημαίνει A_t επιλέγεται από την πολιτική π σε κατάσταση s (ακόμη και αν το π είναι ντετερμινιστικό)
- Αυτή είναι γνωστή ως **εξίσωση Bellman** (Bellman 1957)
- Μια παρόμοια εξίσωση ισχύει για τη βέλτιστη (=υψηλότερη δυνατή) τιμή:

$$v_*(s) = \max_a \mathbb{E}\left[R_{t+1} + \gamma v_*(S_{t+1}) \mid S_t = s, A_t = a\right]$$

- Αυτό **δεν** εξαρτάται από μια πολιτική

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Προσεγγίσεις συνάρτησης τιμής

- Οι παράγοντες συχνά προσεγγίζουν τις συναρτήσεις αξίας
- Θα συζητήσουμε τους αλγορίθμους για να τους μάθουμε αποτελεσματικά
- Με μια ακριβή λειτουργία αξίας, μπορούμε να συμπεριφερθούμε βέλτιστα
- Με τις κατάλληλες προσεγγίσεις, μπορούμε να συμπεριφερόμαστε καλά, ακόμη και σε δυσεύρετα μεγάλους τομείς

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Μέρη πρακτόρων

Μέρη πρακτόρων

- Πολιτεία πρακτόρων
- Πολιτική
- Συνάρτηση αξίας
- **Μοντέλο**

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Μοντέλο

- Ένα **μοντέλο** προβλέπει τι θα κάνει το περιβάλλον στη συνέχεια
- Π.χ., $\mathcal{P}$ προβλέπει την επόμενη κατάσταση

$$\mathcal{P}(s, a, s') \approx p\left(S_{t+1} = s' \mid S_t = s, A_t = a\right)$$

- Π.χ., $\mathcal{R}$ προβλέπει την επόμενη (άμεση) ανταμοιβή

$$\mathcal{R}(s, a) \approx \mathbb{E}\left[R_{t+1} \mid S_t = s, A_t = a\right]$$

- Ένα μοντέλο δεν μας δίνει αμέσως μια καλή πολιτική - θα πρέπει σχεδιάσουμε κι'άλλο
- Θα μπορούσαμε επίσης να εξετάσουμε τα **στοχαστικά (generative)** μοντέλα

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU

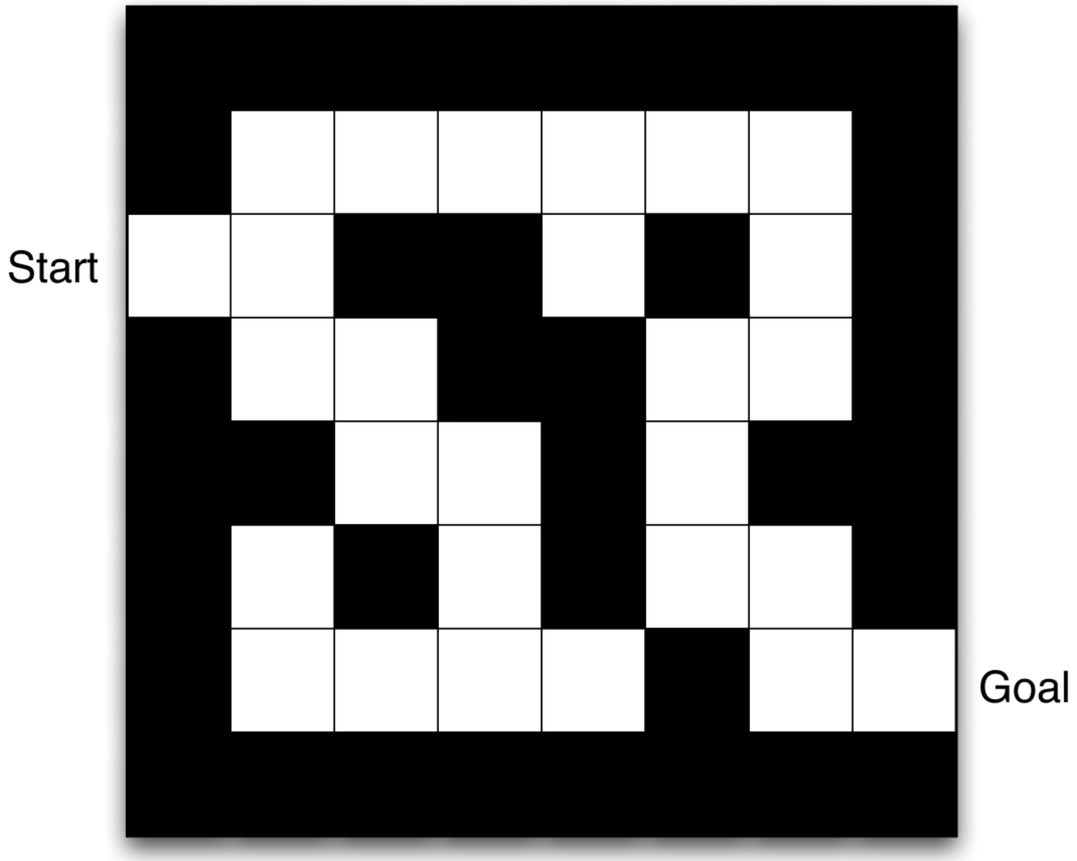
Master programmes in Artificial Intelligence 4 Careers in Europe

Ένα παράδειγμα

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Παράδειγμα λαβύρινθου



- Επιβραβεύσεις: -1 ανά χρονοβήμα
- Δραστηριότητες: Β, Α, Ν, Δ
- Οι καταστάσεις: Η τοποθεσία του πράκτορα

Συμβουλή: Όταν χρησιμοποιούμε αρνητική ανταμοιβή ανά βήμα και ανταμοιβή 0 στο στόχο, ορίζουμε ένα πρόβλημα συντομότερης διαδρομής.

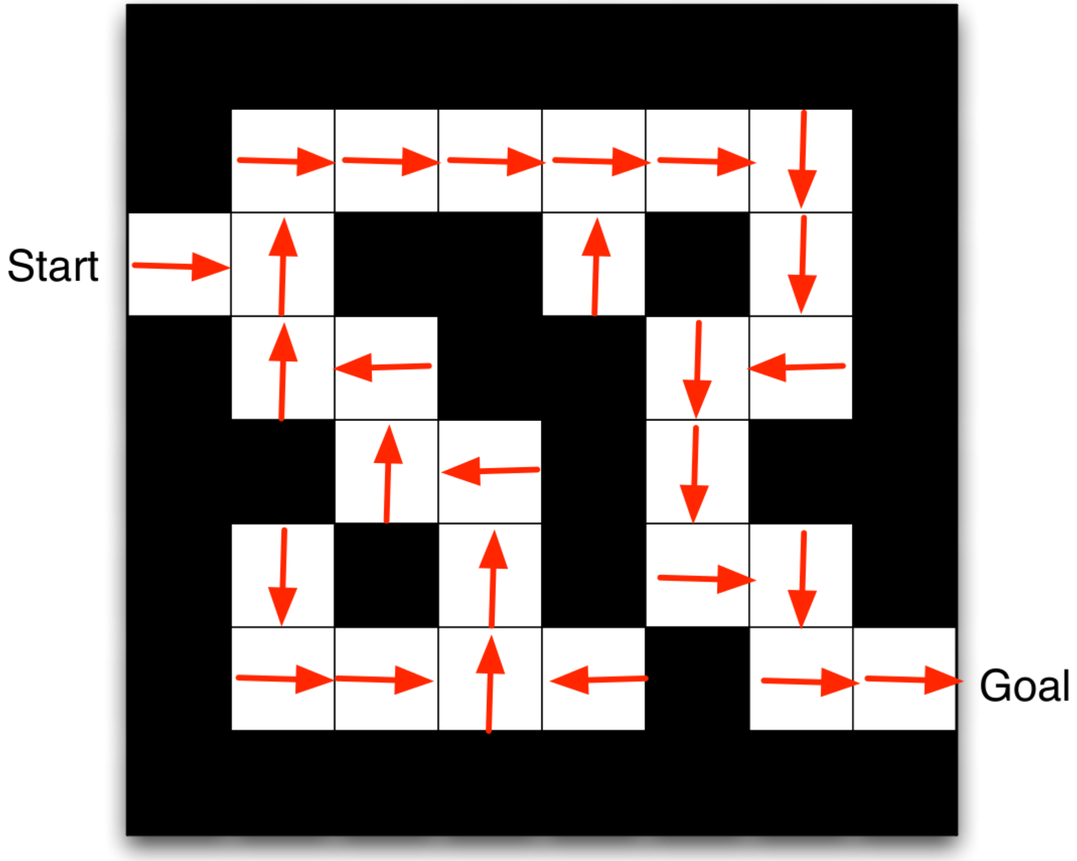
Δηλαδή, ενθαρρύνουμε τον πράκτορα να πάει στο στόχο όσο το δυνατόν γρηγορότερα, επειδή αυτό θα του έδινε το χαμηλότερο ποσό -1.

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Παράδειγμα λαβύρινθου: Πολιτική



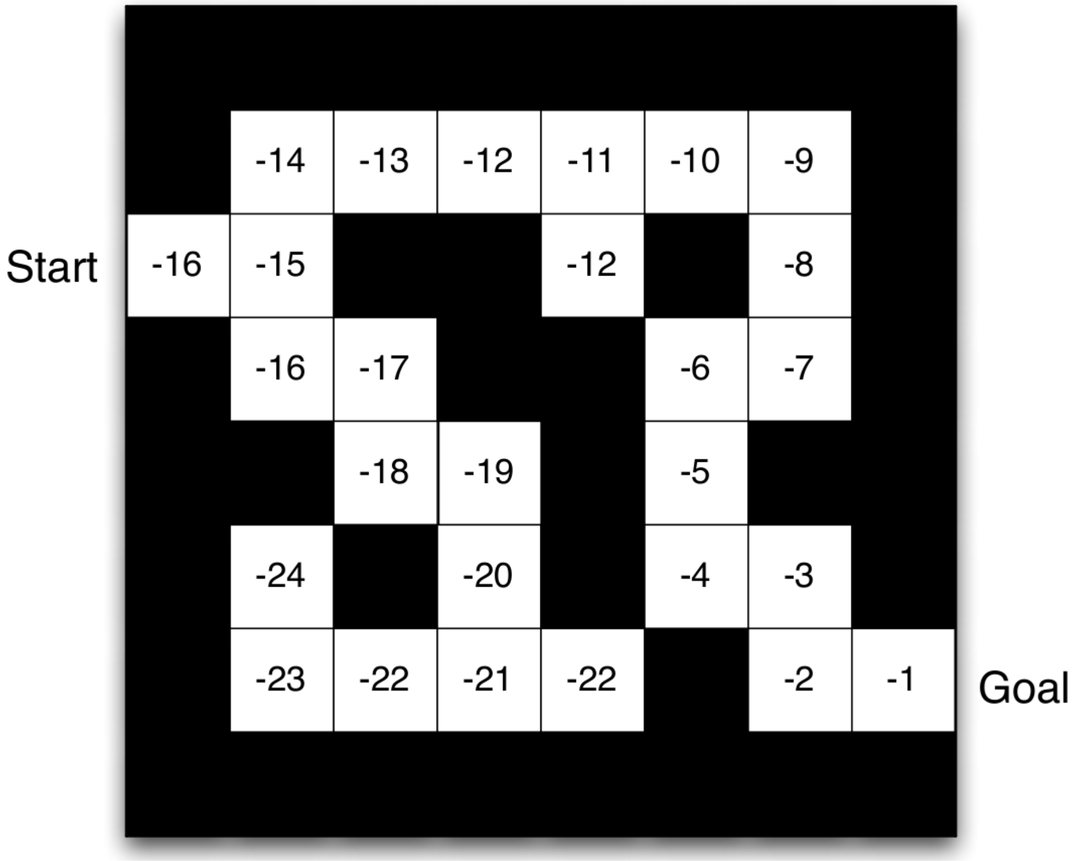
- Τα βέλη αντιπροσωπεύουν *την πολιτική π(s) για* κάθε κατάσταση *s*

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Παράδειγμα λαβύρινθου: Συνάρτηση αξίας



- Οι αριθμοί αντιπροσωπεύουν την τιμή $v_\pi(s)$ κάθε κατάστασης s

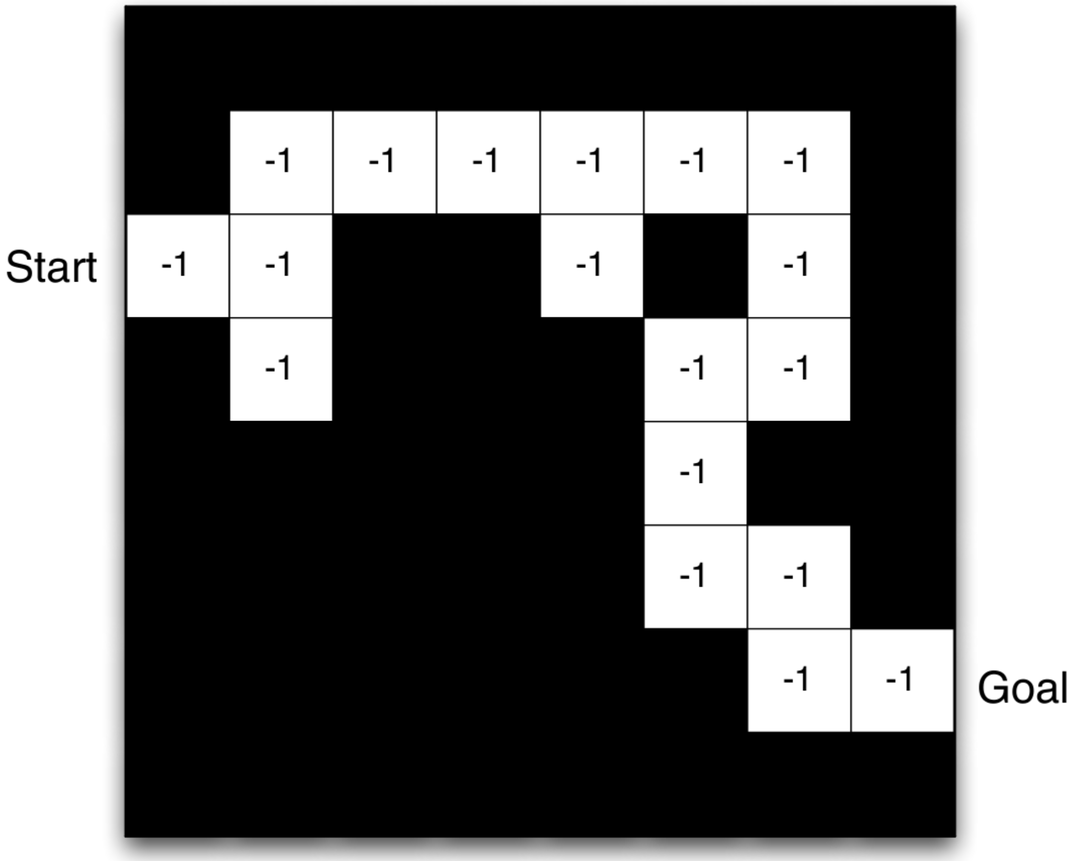
Εδώ, δεδομένου ότι χρησιμοποιήσαμε μια ανταμοιβή -1 ανά βήμα χρόνου, κάθε τιμή αντικατοπτρίζει τον αριθμό των βημάτων προς το στόχο.

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Παράδειγμα λαβύρινθου: Μοντέλο



- Η διάταξη πλέγματος αντιπροσωπεύει το μοντέλο μερικής μετάβασης $\mathcal{P}^{a}_{ss'}$,
- Οι αριθμοί αντιπροσωπεύουν άμεση ανταμοιβή $\mathcal{R}^{a}_{ss'}$, από κάθε κράτος *(* ίδιο για όλα τα *a* και *s σε* αυτή την περίπτωση)

Ο πράκτορας έμαθε αυτό το ανακριβές μοντέλο επειδή μπορεί να μην είχε εξερευνήσει το κάτω μέρος του λαβύρινθου.

Ο πράκτορας μπορεί να χρησιμοποιήσει αυτό το μοντέλο για να σχεδιάσει τις ενέργειές του και να φτάσει στη βέλτιστη λύση για τις καταστάσεις που έχει δει.

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU

Master programmes in Artificial Intelligence 4 Careers in Europe

Κατηγορίες πρακτόρων

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Κατηγορίες πρακτόρων

- Με βάση την αξία
  - Καμία πολιτική (Implicit)
  - Λειτουργία αξίας
- Με βάση την πολιτική
  - Πολιτική
  - Καμία λειτουργία αξίας
- Ηθοποιός Κριτικός
  - Πολιτική
  - Λειτουργία αξίας

- Μοντέλο Δωρεάν
  - Λειτουργία Πολιτικής και/ή Αξίας
  - Κανένα μοντέλο
- Με βάση το μοντέλο
  - Προαιρετικά Λειτουργία Πολιτικής ή/και Αξίας
  - Μοντέλο

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU

Master programmes in Artificial Intelligence 4 Careers in Europe

Υποπροβλήματα του προβλήματος ΕΜ

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Πρόβλεψη και έλεγχος

- **Πρόγραμμα**: αξιολόγηση του μέλλοντος (για μια δεδομένη πολιτική), π.χ. εκμάθηση μιας συνάρτησης αξίας
- **Έλεγχος**: βελτιστοποίηση του μέλλοντος (βρείτε την καλύτερη πολιτική)
- Αυτά μπορεί να είναι στενά συνδεδεμένα:

$$\pi_*(s) = \underset{\pi}{\operatorname{argmax}}\ v_\pi(s)$$

- Αν μπορούσαμε να προβλέψουμε **τα πάντα** χρειαζόμαστε κάτι άλλο;

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Μάθηση και προγραμματισμός

Δύο βασικά προβλήματα στην ενίσχυση της μάθησης

- **Μάθηση**:
  - Το περιβάλλον είναι αρχικά άγνωστο
  - Ο παράγοντας αλληλεπιδρά με το περιβάλλον
- **Σχεδιασμός**:
  - Ένα μοντέλο του περιβάλλοντος δίνεται (ή μαθαίνεται)
  - Ο πράκτορας σχεδιάζει σε αυτό το μοντέλο (χωρίς εξωτερική αλληλεπίδραση)
  - αλλιώς συλλογισμός, σκέψη, αναζήτηση, σχεδιασμός

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Μέρη Εκπαιδευτικού Πράκτορα

- Όλα τα συστατικά είναι λειτουργίες
  - Πολιτικες: $\pi: S \rightarrow A$ (ή σε πιθανότητες πάνω από A)
  - Συναρτήσεις αξίας: $\beta: S \rightarrow R$
  - Πρότυπα: $\mu: S \rightarrow S$ και/ή $r: S \rightarrow R$
  - Ενημέρωση της κατάστασης: $u : S \times O \rightarrow S$
- Π.χ., μπορούμε να χρησιμοποιήσουμε νευρωνικά δίκτυα και να χρησιμοποιήσουμε τεχνικές βαθιάς μάθησης για να μάθουμε
- Προσοχή: συχνά παραβιάζουμε υποθέσεις από την εποπτευόμενη μάθηση (iid, stationarity)
- Η βαθιά μάθηση είναι ένα σημαντικό εργαλείο
- Η βαθιά ενισχυτική μάθηση είναι ένας πλούσιος και ενεργός ερευνητικός τομέας

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU

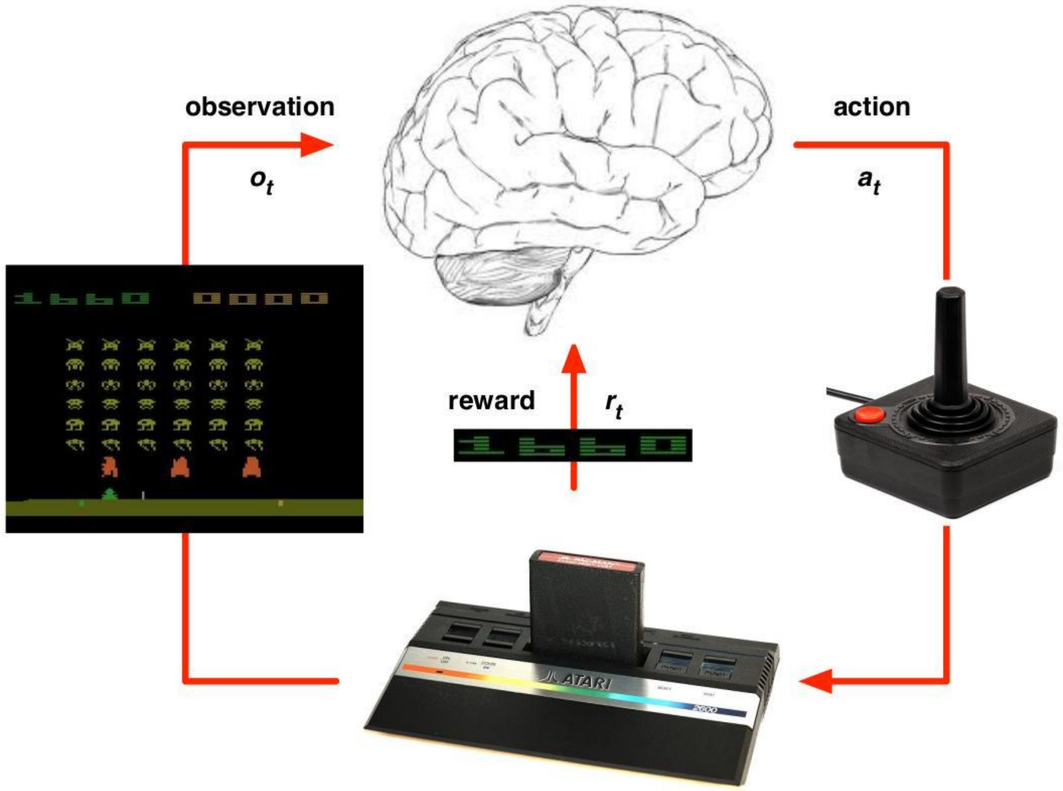
Master programmes in Artificial Intelligence 4 Careers in Europe

Παραδείγματα

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Παιχνιδια Atari



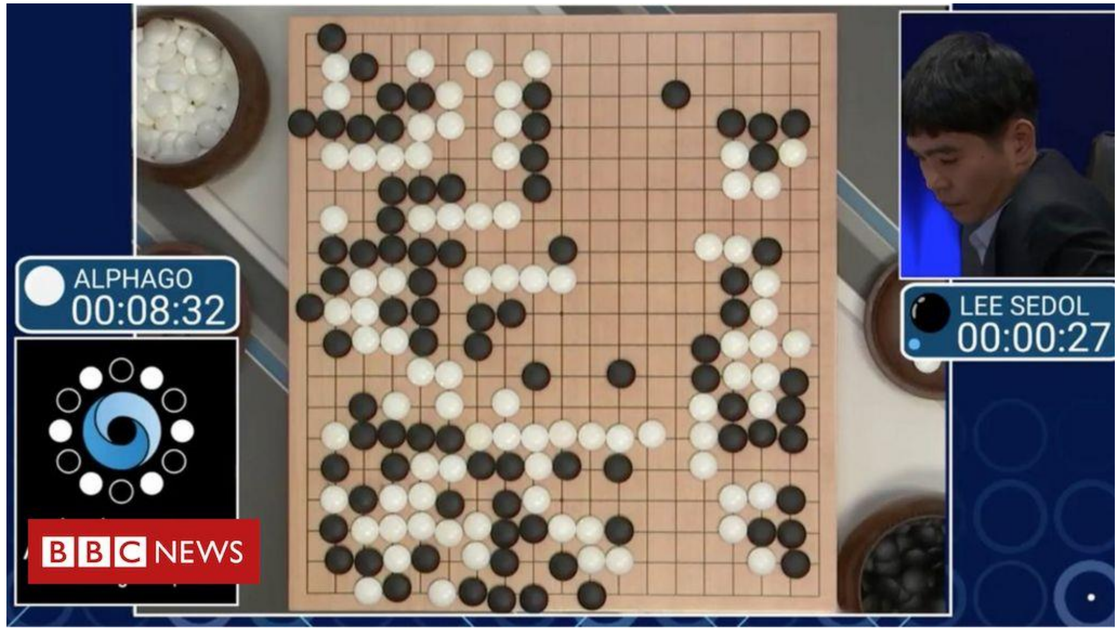
- Οι κανόνες του παιχνιδιού είναι άγνωστοι (χωρίς μοντέλο ΕΜ)
- Μάθηση απευθείας από το διαδραστικό παιχνίδι
- Επιλογή δράσης στο joystick, προβολή pixels και σκορ

Διαφάνειες βασισμένες στις Διαλέξεις Ενισχυτικής μάθησης της DeepMind

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Το AlphaGo



Πηγή εικόνας

- Go ~ 10^{170} αριθμός διαμορφώσεων πινάκων
  - Το σκάκι: ~ 10^{44}
  - Αριθμός ατόμων στο σύμπαν: ~ 10^{82}
- Οι κανόνες του παιχνιδιού είναι γνωστοί (με βάση το μοντέλο RL)
- Συνδυάζει το σχεδιασμό (αναζήτηση δέντρων) με βαθιά νευρωνικά δίκτυα
- Δίκτυο πολιτικής: επιλέγει την επόμενη κίνηση
- Δίκτυο Αξίας: προβλέποντας τον νικητή του παιχνιδιού
- Νίκησε παγκόσμιους πρωταθλητές

- Δείτε ντοκιμαντέρ στο Netflix ή στο YouTube

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU

Master programmes in Artificial Intelligence 4 Careers in Europe

Robotic Quadrupedal Locomotion

Βίντεο στο YouTube

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Άλλες επιτυχίες RL

- TD-Gammon computer backgammon program
- Flying helicopters for aerobatic maneuvers
- Flying stratospheric balloons
- AlphaZero: mastering the games of chess, shogi and go without expert knowledge
- MuZero: master games without knowing their rules
- AlphaStar: Mastering the real-time strategy game StarCraft 2
- Controlling nuclear fusion reactors
- Designing hardware chips
- Discovering more efficient matrix multiplication algorithms
- Solving Rubik’s cube with a robot hand
- …

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU

Master programmes in Artificial Intelligence 4 Careers in Europe

Σας ευχαριστούμε

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423