

ΕΝΙΣΧΥΤΙΚΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ ΚΑΙ ΣΤΟΧΑΣΤΙΚΑ ΠΑΙΓΝΙΑ

(multi agent reinforcement learning)

Περιγραφή

Το μάθημα εξετάζει την ακόλουθη ερώτηση:

- Πως μπορεί μια μονάδα/πράκτορας (agent) να μάθει να ενεργεί για την επίτευξη κάποιου σκοπού, σε ένα δυναμικό αβέβαιο και μερικώς παρατηρήσιμο περιβάλλον, παρουσία και άλλων δρώντων μηχανισμών που επιδιώκουν τους δικούς τους σκοπούς?

Το ερώτημα και οι ειδικότερες εκδοχές του (ένας πράκτορας, κοινός σκοπός και συντονισμός ενεργειών των μονάδων) έχει κεντρικό ενδιαφέρον σε μεγάλο φάσμα εφαρμογών:

- Υπολογιστική όραση, ρομποτική και αυτόνομη οδήγηση
- Επεξεργασία φυσικής γλώσσας
- Παιχνίδια (επιτραπέζια, παιχνίδια με τράπουλες, βιντεο)
- Κυβερνοφυσικά συστήματα
- Ανίχνευση σπάνιων γεγονότων (ανίχνευση και διάγνωση στην ιατρική φροντίδα)
- Επικοινωνίες και ασφάλεια (αντιμετώπιση απειλών και επιθέσεων σε αισθητήρες, κανάλια και υπολογιστικούς κόμβους)
- Το μάθημα ασχολείται με εκείνες τις απαντήσεις στο παραπάνω ερώτημα οι οποίες στηρίζονται στη θεωρία τις μεθόδους και τους αλγόριθμους της ενισχυτικής μηχανικής μάθησης (reinforcement learning). Ετσι, η αλληλεπίδραση μεταξύ των μονάδων και του περιβάλλοντος στο οποίο ενεργούν είναι συνεχής. Οι μονάδες παρατηρούν την κατάσταση του περιβάλλοντος μέσω σχετικών μετρήσεων και έμμεσα αντλούν πληροφορίες για τις ενέργειες των άλλων μονάδων. Κάποια πληροφοριακά στοιχεία είναι κοινά σε όλες τις μονάδες, άλλα συνιστούν ιδιωτική πληροφόρηση. Με βάση τις παρατηρήσεις που έχει στη διάθεση της, κάθε μονάδα επιλέγει μια ενέργεια και την εκτελεί. Η κατάσταση του περιβάλλοντος μεταβάλλεται χρονικά και με αβέβαιο (πιθανοτικό) τρόπο, ανάλογα με την μέχρι τώρα εξέλιξη της και βασει των ενεργειών των μονάδων. Οι αποφάσεις κάθε μονάδας λαμβάνονται ακολουθιακά (στό χρόνο ή σε στάδια), στηρίζονται σε κανόνες, συνιστούν δηλαδή πολιτικές και κατά συνέπεια και η μάθηση είναι ακολουθιακή. Ο κύκλος αυτός επαναλαμβάνεται συνεχώς επιτρέποντας στις μονάδες να μαθαίνουν πως να ενεργούν ώστε να βελτιώνουν, όχι τη στιγμιαία ανταμειβή, αλλά το συνολικό μακροπρόθεσμο όφελος/απόδοση.

Ειδικότερα, θα μελετήσουμε:

1. τις θεμελιώδεις μεθόδους της ενισχυτικής μάθησης:

- Μαρκοβιανές διαδικασίες απόφαση (MDPs), μερικώς παρατηρήσιμες MDPs (POMDPs) και πίστη, Δυναμικός προγραμματισμός και εξίσωση Bellman, Q learning, στοχαστικά παίγνια, κοινή και ιδιωτική πληροφόρηση, πολλαπλοί πράκτορες, ισορροπίες
- Εξομείωση και αλγόριθμοι στοχαστικής προσέγγισης
- Βασικές προσεγγιστικές δομές, βαθειά νευρωνικά δίκτυα με και χωρίς μνήμη.

2. Αντιπρόσωπους αλγορίθμων ενισχυτικής μάθησης (vanilla policy gradient VPG, trust region policy optimization TRPO, proximal policy optimization PPO, Deep deterministic policy gradient DPPG, Twin delayed DDPG, TD3, Soft Actor Critic SAC, DQN, DDQN, και άλλους)

3. Βιβλιοθήκες αλγορίθμων ενισχυτικής μάθησης (Stable baselines3), βιβλιοθήκες με προεγκατεστημένα περιβάλλοντα (OpenAI Gym) και υποκείμενες βιβλιοθήκες νευρωνικών δικτύων (Pytorch).

Μαθησιακοί στόχοι

Απόκτηση γνώσεων και δεξιοτήτων σε μια σημαντική περιοχή της μηχανικής μάθησης που έχει ισχυρούς δεσμούς με τον βέλτιστο έλεγχο (Optimal control), την επιχειρησιακή έρευνα (Operations Research), τις προσεγγιστικές δομές (στοχαστική προσέγγιση, νευρωνικά δίκτυα), την εξομείωση, τη θεωρία πληροφορίας και τη θεωρία παιγνίων, με εξαιρετικές επιτυχίες τα τελευταία χρόνια

Αξιολόγηση:

Συμμετοχή στο μάθημα, Εργασία

Πιστωτικές μονάδες: ECTS 6