



Σ Ε Μ Ι Ν Α Ρ Ι Ο Τ Μ Η Μ Α Τ Ο Σ

ΟΜΙΛΗΤΗΣ:



Κωνσταντίνος Μπλέκας

Αναπληρωτής Καθηγητής
Τμήμα Μηχ. Η/Υ & Πληροφορικής
Πανεπιστήμιο Ιωαννίνων

ΗΜΕΡΟΜΗΝΙΑ:

Πέμπτη, 22 Απριλίου 2021

ΩΡΑ:

12:00

ΑΙΘΟΥΣΑ:

Θα μεταδοθεί διαδικτυακά μέσω MS Teams

Link MsTeams:

<https://teams.microsoft.com/l/meetup-join/19%3a84545243afd4489da0e6c4d1a8aa3d0a%40thread.tacv2/1618824473966?context=%7b%22Tid%22%3a%2208bea52a-5ad3-4627-9549-5ff3a65676be%22%2c%22Oid%22%3a%229cf679bd-1cb1-4a8c-8613-d87951467e61%22%7d>

Θέμα

Multi-agent reinforcement learning for autonomous driving in traffic networks with unsignalized intersections

Περίληψη

Ένα σημαντικό πρόβλημα της αυτόνομης οδήγησης οχημάτων σε αστικά περιβάλλοντα για τη μείωση της κυκλοφοριακής συμφόρησης και την ασφαλή κίνηση χωρίς συγκρούσεις είναι η διαχείριση διασταυρώσεων που δεν περιέχουν φωτεινούς σηματοδότες. Στην ομιλία αυτή θα παρουσιαστεί μία μέθοδος πολλών πρακτόρων ενισχυτικής μάθησης (multi-agent reinforcement learning) για την επίλυση αυτού του προβλήματος. Η προτεινόμενη μέθοδος εισάγει την έννοια των route-agents θεωρώντας ως πράκτορες όλες τις δυνατές διαδρομές που μπορούν να ακολουθήσουν τα οχήματα στο οδικό δίκτυο. Οι route-agents έχουν σημαντικά πλεονεκτήματα αφενός στην οικονομία του πολυπρακτορικού συστήματος, και αφετέρου στη διαδικασία μάθησης αυξάνοντας την γενικευτική ικανότητα (generalization) των παραγόμενων πολιτικών και επιτρέποντας την επαναχρησιμοποίησή τους (knowledge reusing) σε άγνωστα σενάρια με μεγαλύτερο βαθμό πολυπλοκότητας. Έμφαση δίνεται στην κατασκευή του χώρου καταστάσεων (state space) των πρακτόρων όπου, παράλληλα με τις ατομικές πληροφορίες των οχημάτων χρησιμοποιούνται πληροφορίες πρόβλεψης που αποτυπώνουν την κατάσταση της ροής κυκλοφορίας που αναμένεται να συμβεί



πάνω στις διασταυρώσεις. Η χρήση αυτής της πληροφορίας προσφέρει τον συντονισμό (coordination) των οχημάτων που κατευθύνονται τοπικά σε κάθε κόμβο-διασταύρωση του οδικού δικτύου και την συνεργατική συμπεριφορά των πρακτόρων στο περιβάλλον ώστε να εξυπηρετήσουν βέλτιστα όλα τα οχήματα. Για την προσέγγιση της συνάρτησης αξίας (value function) των πρακτόρων χρησιμοποιούνται μορφές (τυπικών ή βαθιών) νευρωνικών δικτύων (neural networks), ενώ για τη μάθηση των πολιτικών τους και την εκτίμηση των βαρών των νευρωνικών δικτύων χρησιμοποιείται ο αλγόριθμος double Q-learning. Επίσης, προτείνεται μία κατάλληλη συνάρτηση ανταμοιβής (reward function) με βάση την οποία πραγματοποιείται η μάθηση. Τέλος, θα παρουσιαστούν πειραματικά αποτελέσματα της μεθοδολογίας σε ποικίλα οδικά δίκτυα (τεχνητά και πραγματικά) και σενάρια που παρήχθησαν χρησιμοποιώντας το γνωστό περιβάλλον προσομοίωσης SUMO (simulation for urban mobility) και παράλληλα θα γίνει σύγκριση με τις μεθόδους που υπάρχουν ενσωματωμένες στο πακέτο αυτό.