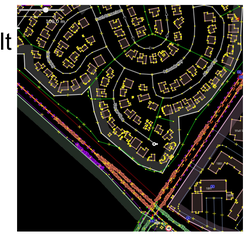
Traffic-Signal Control in Real-World Environments

Alex Mallery, Ben Cohen-Wang, Edward Lee
Stanford University

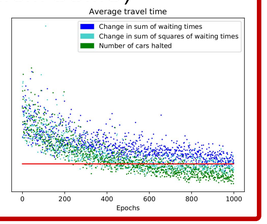
Motivation

- Traffic is a common problem
- Changing road infrastructure is difficult
- Changing traffic light timings is more practical
- Existing research generally optimizes traffic in symmetric grids or single intersections
- Traffic simulators available (SUMO) that allow gathering of realistic data



Primary Model

- Formulated as RL problem with state-action-reward tuple
- System **States** approximated with primary features Φ(s):
  - Number of vehicles present, number halting, average flow rate of vehicles, traffic light state (from SUMO)
- **Rewards** are number of cars halted in total
- **Actions** are every legal state (red, green combination) of every traffic light not currently in a yellow-light transition state
- Core features apply to each light



Q-learning with feature-based light dependencies

Linear function approximation

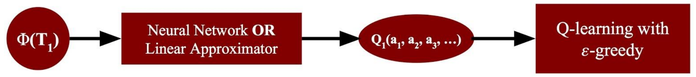
- Large state-space makes learning Q-value for every state-action pair impractical
- Approximating Q-value for each action as a linear function of state features allows generalization

Neural net function approximation

- Replacing raw features with neural net features of features allows increased expressivity
- Neural net with single hidden layer with (num features * num actions) neurons and σ fn used

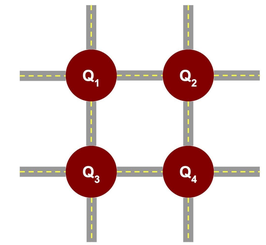
Feature-based light dependencies

- Primary features result in optimizing each light's actions independently
- Adding in states, halting information from adjacent lights better models inter-light dependencies

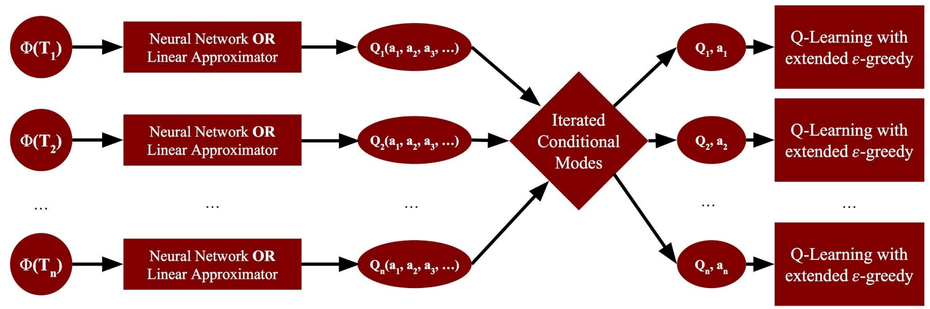


Multi-agent Q-learning with ICM

- Extend features used to approximate each Q_k to include actions of all neighboring intersections, forming a factor graph whose solution can be approximated using iterated conditional modes (ICM)
  - Using ICM instead of solving saves computing time
- Extended ε-greedy allows for exploration:
  - Modify ICM solution with probability ε_1. If modifying ICM solution, modify each component Q with probability ε_2.
- Each traffic light (agent) then takes the action given by the solution

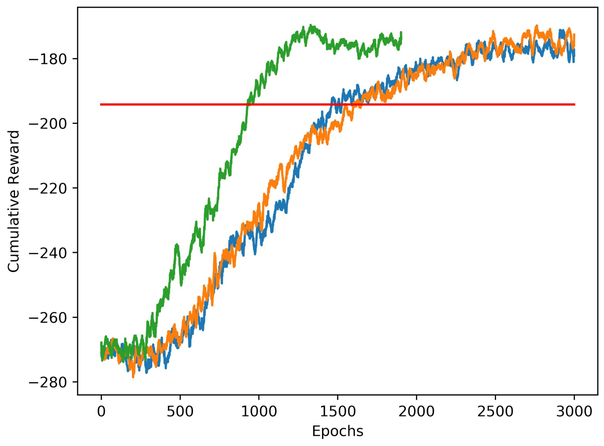


ICM finds a_1, a_2, a_3, a_4 that maximizes
$Q_{tot}=Q_1(a_1, a_2, a_3)+Q_2(a_2, a_1, a_4)$
$+Q_3(a_3, a_1, a_4)+Q_4(a_4, a_2, a_3)$

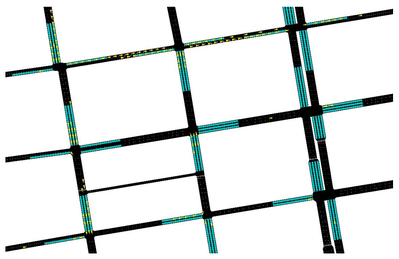


Results and Error Analysis

- All methods graphed (ICM MARL with lin. predictors, Ind. features with lin. predictors, Ind. features with NN predictors) give rewards above baseline default timings
- MARL converges faster but to same asymptote as others



- Adding in light dependencies to features gives similar performance but with slower training times
- ICM MARL with NN predictors fails to converge
- Initial tests conducted on a stretch of Page Mill Road between Foothill and El Camino
  - Lights too far apart for inter-light dependencies to be noticeable
- Current tests (graphed on left) conducted on a stretch of San Francisco grid containing many closely spaced one-way streets



Post-Training Statistics

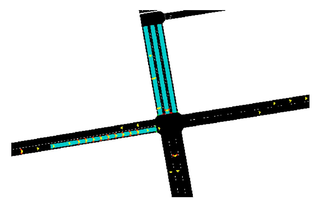
	Average Trip Duration	Average Wait Steps	Average Timeloss
Baseline	326.13 sec	115.34 steps	125.64 sec
Linear, Ind. feat.	307.73 sec	91.38 steps	110.15 sec
NN, Ind. features	**300.34 sec**	**88.42 steps**	**103.92 sec**
ICM MARL	In progress	In progress	In progress

Notes and Caveats

- Each epoch consists of 1000 seconds
- Yellow-light transitions take 3 seconds
- Flows defined based on per-second probabilities of pre-selected routes
- Optimizing for number of halting cars leads to policies that give very short durations for each light

Conclusions

- All predictors converge to about the same cumulative rewards
- Successful implementation of computationally feasible, better-than-baseline algorithm in urban (San Francisco) and suburban (Palo Alto) settings and under different traffic conditions
- Main limiting factor is lack of granularity in reward function
  - Halting car numbers have hard boundaries, limits precision
- Next steps include improving reward function, experimenting with different, simpler methods of modeling inter-light dependencies and coordinating joint traffic signal actions



References

W. Genders and S. Razavi, "Using a Deep Reinforcement Learning Agent for Traffic Signal Control," ARXIV, Nov. 2016. URL: https://arxiv.org/abs/1611.01142

Z. Zhang and D. Zhao, "Cooperative multiagent reinforcement learning using factor graphs," 2013 Fourth International Conference on Intelligent Control and Information Processing (ICICIP), Beijing, 2013, pp. 797-802. doi: 10.1109/ICICIP.2013.6568181 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6568181&isnumber=6568023