経路選択を考慮したGANによる適応型信号制御パラメータの最適化

吉澤尚輝 （東北大学） 井料隆雅（東北大学）

3-B-10

はじめに

- 強化学習を使用した適応型信号制御が近年注目されている.
- 適応型信号制御と利用者の経路選択の相互作用を考慮した適応型信号制御は少ない.
- 強化学習を使用しない，信号制御と経路選択の相互作用を考慮した制御にPolicy P_0がある.
  - Policy P_0には，飽和交通流率の小さな枝に交通量とスプリットが集中し，ネットワークの交通容量がネットワークへ流入する総需要より少なくなることを防ぐcapacity maximizationという性質がある.
  - Policy P_0以外の制御にこの性質があるか不明.
- 強化学習を使用した，信号制御と経路選択の相互作用を考慮した制御は知る限り存在しない.
- 適応型信号制御に使用する強化学習の学習時，および学習済みの適応型信号制御の性能評価時の両方で，利用者の経路選択を考慮する.
- この時，強化学習を使用した適応型信号制御がcapacity maximizationを達成するか調べる.

シミュレーション設定

- 飽和交通流率S_{we}は小さいが旅行時間が短い経路と，飽和交通流率S_{ns}は大きいが旅行時間が長い経路，の2経路が交差点に流入するネットワーク.
- スプリットの決定には，Deep Q network (DQN)を使用し，このネットワークをDQN_sと呼ぶ.
- DQN_sの学習時に流入する需要は経路選択を行う.
- 学習時の経路選択方法は，ロジットモデルと，経路選択用のDQN（DQN_r）の2種類.
  - DQN_rの報酬はDQN_sの報酬に-1を掛けたものになっており，DQN_rとDQN_sは互いに敵対的な行動を選択する.
    - これはGenerative Adversarial Networks (GAN)の考え方を応用している.
- DQN_sの性能評価に使用する需要はロジットモデルに従って経路変更する.

シミュレーション結果

ロジットモデルに従う需要で学習

- 表1に，学習時に流入する需要の分散パラメータθが各値のDQNsに対して，流入流率の期待値Iと分散パラメータθが各値の需要が流入した時のRを示す.
  - Rは適応型信号制御が実現する交通容量に対するネットワークへの流入流率の割合で，capacity maximizationが達成されている場合，R≤1.0となる.
  - 表中の色は，R=0.1に近いほど緑が，R=2.0に近いほど赤が濃くなる．R=1.0の時は白である.
  - 学習時のθ=0.25の行内にあるnanの表記は，DQNsの性能を評価するシミュレーションが終了しなかったため，各枝の平均スプリットが算出できなかったことを示す．この現象は，車両が存在する枝に対してスプリットが割り振られず，車両が交差点から流出できなかったため発生した.
- 流入する需要が各θの時に実現する制御のRは，学習時に使用した需要のθによって，大きく異なった.
  - 多くの条件でR>1.0となり，Iが大きい条件ではR=1.8程度まで達した.

GANを応用した手法で学習

- 表2は学習時の流入流率の期待値が各値のDQNsに対して，Iとθが各値の需要が流入した時のRで，表中の色は表1と同じ意味である.
- 多くの条件で，R>1.0となっていた.
  - 学習時の需要の流入流率の期待値が1440台/時，2160台/時，3600台/時の場合，R=2.0に近い値となっている場合も存在した.
  - 一方，2880台/時では，I=2880，3060の時，どのθが流入してもRは比較的1.0に近い値で安定している.

まとめ

- ロジットモデルに従う需要で学習した場合，各θの需要が流入した時に実現するスプリットは，学習時の需要のθや評価時の流入流率の期待値によって，大きく異なった.
  - 学習時のθの値によってRが1.2程度に収まるケースから1.8程度に達するものまで存在した.
- GANを応用した学習の場合，学習時の需要の流入流率の期待値によっては，どのようなθや流入流率の期待値の需要が流入しても，1.2程度の相対的に小さなRが実現した.

表1　各θの流入需要で学習したDQNsに，各θの需要が流入した時に実現するR

		評価時に流入する需要の分散パラメータ θ=0.125	評価時に流入する需要の分散パラメータ θ=0.25	評価時に流入する需要の分散パラメータ θ=0.5	評価時に流入する需要の分散パラメータ θ=0.75	評価時に流入する需要の分散パラメータ θ=1.0	評価時に流入する需要の分散パラメータ θ=∞
学習時に流入する需要の分散パラメータ θ=0.0	I=720 台/時	0.35	0.35	0.35	0.35	0.35	0.35
	I=2160 台/時	0.92	0.93	0.96	0.98	0.91	0.92
	I=2880 台/時	1.08	1.05	1.05	1.08	1.02	1.07
	I=3060 台/時	1.14	1.11	1.13	1.09	1.09	1.06
	I=3600 台/時	1.23	1.21	1.20	1.21	1.19	1.21
学習時に流入する需要の分散パラメータ θ=0.125	I=720 台/時	0.35	0.35	0.35	0.35	0.35	0.35
	I=2160 台/時	1.08	1.11	1.10	0.92	1.09	1.23
	I=2880 台/時	1.04	1.46	1.62	1.43	1.65	1.13
	I=3060 台/時	1.79	1.16	1.79	1.17	1.15	1.13
	I=3600 台/時	1.23	1.28	1.30	1.33	1.44	1.25
学習時に流入する需要の分散パラメータ θ=0.25	I=720 台/時	0.35	0.35	0.35	0.35	0.35	0.35
	I=2160 台/時	nan	nan	nan	nan	nan	nan
	I=2880 台/時	nan	nan	nan	nan	nan	nan
	I=3060 台/時	nan	nan	nan	nan	nan	nan
	I=3600 台/時	nan	nan	nan	nan	nan	nan
学習時に流入する需要の分散パラメータ θ=0.5	I=720 台/時	0.35	0.35	0.35	0.35	0.35	0.35
	I=2160 台/時	0.92	0.95	0.94	0.96	0.95	0.95
	I=2880 台/時	1.09	1.05	1.08	1.06	1.07	1.06
	I=3060 台/時	1.11	1.19	1.12	1.11	1.11	1.07
	I=3600 台/時	1.19	1.19	1.19	1.27	1.20	1.24
学習時に流入する需要の分散パラメータ θ=0.75	I=720 台/時	0.35	0.35	0.35	0.35	0.35	0.35
	I=2160 台/時	0.93	0.93	0.94	0.87	0.95	0.96
	I=2880 台/時	1.07	1.02	1.06	1.06	1.11	1.07
	I=3060 台/時	1.04	1.07	1.12	1.05	1.06	1.07
	I=3600 台/時	1.20	1.21	1.22	1.18	1.20	1.16
学習時に流入する需要の分散パラメータ θ=1.0	I=720 台/時	0.35	0.35	0.35	0.35	0.35	0.35
	I=2160 台/時	1.17	1.16	1.13	1.21	1.28	1.16
	I=2880 台/時	1.56	1.59	1.59	1.56	1.57	1.56
	I=3060 台/時	1.68	1.66	1.69	1.62	1.71	1.65
	I=3600 台/時	1.88	1.91	1.82	1.92	1.94	1.90
学習時に流入する需要の分散パラメータ θ=∞	I=720 台/時	0.35	0.35	0.35	0.35	0.35	0.35
	I=2160 台/時	0.98	1.01	0.95	0.97	0.96	0.94
	I=2880 台/時	1.10	1.13	1.10	1.05	1.07	1.12
	I=3060 台/時	1.13	1.18	1.15	1.13	1.11	1.17
	I=3600 台/時	1.32	1.28	1.30	1.30	1.29	1.28

表2　GANを応用した手法において，各流入流率の期待値を持つ流入需要で学習したDQNsに，各θの需要が流入した時に実現するR

		評価時に流入する需要の分散パラメータ θ=0.125	評価時に流入する需要の分散パラメータ θ=0.25	評価時に流入する需要の分散パラメータ θ=0.5	評価時に流入する需要の分散パラメータ θ=0.75	評価時に流入する需要の分散パラメータ θ=1.0	評価時に流入する需要の分散パラメータ θ=∞
学習時に流入する需要の流入流率期待値1440台/時	I=720 台/時	0.35	0.35	0.35	0.35	0.35	0.35
	I=2160 台/時	0.99	0.96	0.97	0.96	0.98	0.95
	I=2880 台/時	1.28	1.62	1.42	1.17	1.43	1.28
	I=3060 台/時	1.69	1.57	1.40	1.45	1.24	1.22
	I=3600 台/時	1.44	1.83	1.84	1.39	1.80	1.82
学習時に流入する需要の流入流率期待値2160台/時	I=720 台/時	0.35	0.35	0.35	0.35	0.35	0.35
	I=2160 台/時	0.98	0.95	1.00	0.98	0.96	0.95
	I=2880 台/時	1.21	1.64	1.16	1.68	1.57	1.59
	I=3060 台/時	1.22	1.27	1.26	1.72	1.21	1.74
	I=3600 台/時	1.31	1.47	1.29	1.30	1.31	1.30
学習時に流入する需要の流入流率期待値2880台/時	I=720 台/時	0.35	0.35	0.35	0.35	0.35	0.35
	I=2160 台/時	0.94	0.92	0.94	0.91	0.94	0.96
	I=2880 台/時	1.04	1.04	1.02	1.04	1.06	1.02
	I=3060 台/時	1.09	1.09	1.12	1.08	1.05	1.07
	I=3600 台/時	1.21	1.21	1.19	1.18	1.19	1.19
学習時に流入する需要の流入流率期待値3600台/時	I=720 台/時	0.35	0.35	0.35	0.35	0.35	0.35
	I=2160 台/時	0.96	0.97	0.97	0.97	0.98	0.96
	I=2880 台/時	1.15	1.10	1.12	1.05	1.10	1.09
	I=3060 台/時	1.25	1.12	1.15	1.12	1.15	1.10
	I=3600 台/時	1.41	1.25	1.45	1.29	1.31	1.38

今後の課題

- 本研究でDQN_rの選択する行動は使用する経路のみだが，流入間隔もDQN_rに選択させることで，学習時の需要の流入流率の期待値を設定する必要が無くなるかもしれない.

謝辞：本研究は，この研究は，科学研究費補助金（基盤A：20H00265）の支援によりなされた.