交通流制御のための情報提供戦略の学習

内田英明† 荒井幸代‡

†‡ 千葉大学工学部

1 はじめに

交通渋滞が与える経済的損失は国土交通省の試算では年間 12 兆円にのぼる．この数値には渋滞の及ぼす環境負荷は含んでおらず，低炭素社会をめざす意味においても交通渋滞を緩和するための方策が望まれる．近年の高度交通システム (Intelligent Transport Systems) 導入の下では，事故や渋滞情報をリアルタイムで収集し運転者に配信することが可能になった．しかし高速道路では一般に，「ある区間における現在所要時間」のみが渋滞情報として提供される．

一方，運転者が，提供された情報に基づいて，選択可能な各経路の所要時間を推測し，経路を選択する状況では，交通の最適配分が行われるとは限らないことが知られている [1]．また，このような現象を回避するため，情報が交通情況に与える影響の評価について研究がなされている [2]．

そこで本研究では，交通流を制御する主体として渋滞情報センター（以下，センター）と運転者の 2 つをモデル化し，強化学習を適用する．センターはネットワーク全体の状態入力から経路選択の指針をトップダウンに提供し，運転者はその情報を基に行動することで，動的に変化する交通流に適応してルーティング政策を変化させる制御モデルを提案する．

2 問題設定と接近法

2.1 強化学習

センターは式 (1) に従って 1step ごとに Q 値を更新する．このとき，状態，行動は次のように離散化して表現する．

状態 $s \in \mathcal{S}$ は交通ネットワークにおけるリンクの密度 $d_l(l:$ リンクのラベル) を離散化した値で表現する．$0 \leq d_n \leq 1$ であり，0.1 刻みで 10 段階の評価とする．行動 $a \in \mathcal{A}$ は運転者が分岐ノードに到着した際，センターが情報提供する配分率を変動させることと表現する．時刻 t において，ある分岐ノード i における総流入交通量を x_i^t，対応する制御リンク $j(j = 1, 2, \cdots, n)$ の交通量を $x_{i,j}^t$ とする．ただし，n はノード i の出次数である．このとき，行動集合 $\mathcal{A} =$ { 減少, 一定, 増加 } とし，それぞれの行動を図 1 に記述する．また，旅行時間の最小化が目的であるため，負の報酬として平均旅行時間を与える．

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left\{ r + \gamma \max_{a' \in \mathrm{A}(s')} Q(s',a') - Q(s,a) \right\} \quad (1)$$

2.2 運転者の意思決定

均衡配分理論の仮定 [1] によれば，運転者が各リンクの旅行時間について情報を取得した場合，交通ネットワークは利用者均衡配分 (以下 UE) 状態に陥る．これに対し，センターの学習モデルによって獲得された情報提供戦略は，運転者に対し必ずしも最短経路の提示を行うとは限らない．そこで，センターの情報提供率 $p(0 \leq p \leq 1)$ を導入する．これによって，確率 p でセンターの指示した経路に従い，確率 $(1-p)$ でその時点での最短経路を選択する，という運転者の意思決定を反映する．

これは運転者の中に，センターの指示に従って経路選択をする経路浮動層が p の割合で存在することを表す．本来選択するはずであった経路とは別の経路に指示された場合，経路浮動層は通行料の割引など，何らかのインセンティブによって，指示を受け入れると考えることができる．

- 減少：$u^{t+1} = u^t - 0.01$
- 一定：$u^{t+1} = u^t$
- 増加：$u^{t+1} = u^t + 0.01$
  （ただし，配分率 $u^t = \frac{x_{i,1}^t}{x_i^t}$ とする）

図 1: 行動集合

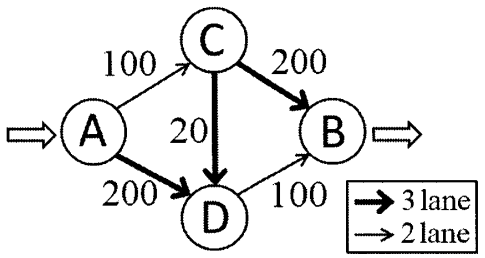

図 2: Braess のパラドクスを持つネットワーク

3 計算機実験と考察

3.1 対象ネットワーク

実験対象のネットワークとして図 2 のような有向グラフを考える．この交通ネットワークでは Braess のパラドクス [3] が発生することが知られている．このパラドクスは，ショートカットリンク CD の存在により，UE の平均旅行時間がネットワークの交通時間が最小化されたシステム最適配分 (以下 SO) での旅行時間に比べ大きくなってしまうというものである．つまり，運転者の自由度が大きくなった状態は，必ずしもネットワーク全体のパフォーマンス向上にはつながらないため，センターによる制御を必要とすると考えられる．

3.2 予備実験

予備実験として，流入交通量一定の環境（定常状態）について，Q 学習の適用による最適値への収束性を確かめる．交通流は CA モデルにより再現し，500split のシミュレーションを学習 1step とする．ここでは分岐ノード A, C において経路選択に関する情報提供を行うため，リンクの密度を離散化した値の組み合わせによって状態表現を行う．状態集合 $\mathcal{S} = \{s_{\mathrm{AC}}, s_{\mathrm{CD}}\}$ の状態数は 10^2，行動集合 $\mathcal{A} = \{a_{\mathrm{C}}, a_{\mathrm{D}}\}$ の行動数は 3^2 と

Learning Strategy of Effective Information Services for Indirect Traffic Control

Hideak UCHIDA†, Sachiyo ARAI‡
†‡Faculty of Engineering, Chiba University
263-8522, Chiba, Japan
†z6t0006@students.chiba-u.jp, ‡arai@tu.chiba-u.ac.jp

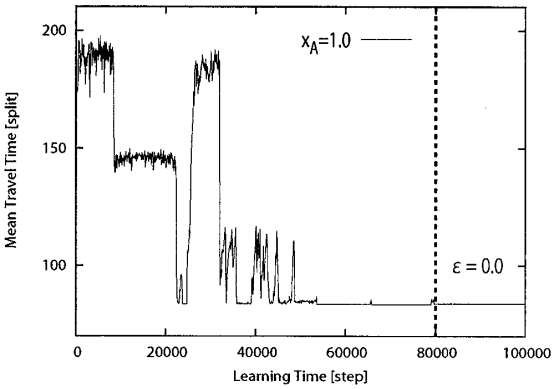

図 3: 学習のプロセス

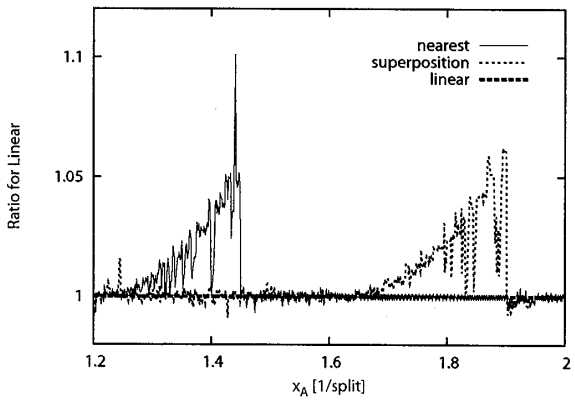

図 4: 補間手法による効果

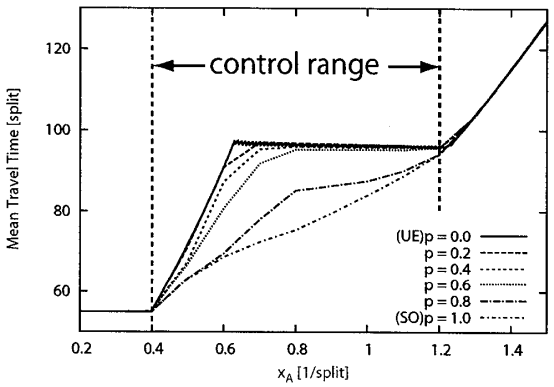

図 5: 情報提供率による変化

なる．Q 値は初期値を 0，学習のパラメータは学習率 $\alpha = 0.03$，割引率 $\gamma = 0.9$ とし，行動選択には ϵ-greedy 選択 ($\epsilon = 0.3$) を採用する．この設定は断りのない限り以降の実験でも同様とする．また，運転者に対する情報提供率 $p = 1$ とする．これはセンターの情報提供に必ず従うという状態であるが，学習の最適性を評価する必要があるためこの設定を用いる．

流入交通量 $x_A = 1.0$[台/split] における学習結果を図 3 に示す．横軸は学習回数，縦軸は学習の更新 100[setp] 毎の平均旅行時間である．$\epsilon > 0$ であるため収束後も平均旅行時間にノイズが発生しているが，収束性を確かめるため実験では 80000[step] 以降は決定的な行動選択 $\epsilon = 0$ とした．このとき，学習終了時のそれぞれの平均旅行時間は最適値には収束しないが，準最適値であることが確認された．これは行動集合の要素数が少ないことが原因であると考えられる．

3.3 情報提供戦略の拡張

予備実験から，Q 学習によって獲得された情報提供戦略の有効性を示した．しかし現実の交通流は日々変化するため，流入交通量を観測してから学習を適用することは困難である．そこで，事前にオフラインで学習し，観測された流入交通量に応じた戦略を選択する方法を用いる．このとき，戦略とは学習によって関数近似された Q 値表を指す．また，流入交通量は連続量であることから，全ての状態に対して学習を行うことは不可能である．そこで，事前学習で獲得した離散的な Q 値表を補間することを考える．補間の方法として次の 3 手法を比較する．(1) 0 次補間：観測点から最も近い Q 値表を採用する．(2) 重ね合わせ：隣接する 2 値点の Q 値表の平均をとる．(3) 1 次補間：隣接する 2 値点の Q 値表を，観測点からの距離によって重み付けする．このとき，補間された Q 値表の再学習は行わず，決定的行動選択による解探索を 100[step] 行う．実験設定は流入交通量が $0 < x_A \leq 2.0$ の範囲において，0.1[台/split] 刻みで事前学習を行い，0.01[台/split] 刻みに補間を行った．

補間手法ごとの実験結果を図 4 に結果を示す．ここで縦軸は，1 次補間の結果に対する他手法の平均旅行時間の割合，横軸は流入交通量である．補間手法としては 1 次補間の精度が安定して最も高く，実際に学習した状態入力でなくとも，Q 値表を基に実時間の解探索を行うだけで準最適解が得られることがわかった．これは，各状態間が連続な位相構造であるためと考えられる．

3.4 情報提供率

提案手法では全ての運転者が車載器を搭載している状況を考えている．このとき，$p = 1$ で SO 状態が実現されることはこれまで述べてきたとおりであるが，$p = 0$ では UE 状態が実現される．これは全ての運転者が車載器を持たない状況下で仮定される均衡状態と等価であり，センターの情報提供効果を著しく低下させる．また，社会的に望ましい状態は SO であるため，p の値は大きいほど良い制御であるが，p は同時に運転者の負担に関するパラメータでもある．そこで，情報提供率を $0 < p \leq 1.0$ の範囲で 0.2 刻みに変化させた場合の学習結果を図 5 に示す．ここで，縦軸は平均旅行時間，横軸は流入交通量である．このとき，$x_A \leq 0.4$ 及び $1.2 \leq x_A$ の範囲では p による有意な差はなく，SO と UE が等価であった．また，p によって違いの確認された $0.4 < x_A < 1.2$ の範囲においても，$p > 0.4$ では UE 状態と変わらない平均旅行時間であり，系全体のパフォーマンスを向上させるには，情報提供率に関する一定の閾値が存在すると考えられる．

4 まとめと今後の課題

本研究では，はじめに，Braess のパラドクスを持つ交通ネットワークにおける，情報提供の制御方法を提案し，強化学習の特徴を利用した Q 値表の補間によって戦略を連続な状態に対しても拡張可能であることを示した．また，情報提供におけるコストを考慮したネットワークの最適化に向け，センターの情報提供率 p を変化させた場合の平均旅行時間の変化を考察した．

今後の課題として，交通量の増減や事故によるリンク容量の増減など，交通流の特徴である動的環境での制御を検討する必要がある．また，運転者の意思決定モデルを更に階層化し，環境の多様性を表現する必要があると考える．

参考文献

[1] 加藤晃, “交通量配分理論の系譜と展望”, 土木学会論文集, No.389, IV-8, pp.15-27, 1988.

[2] 吉井稔雄, 桑原雅夫, “リアルタイム交通情報の提供効果”, 土木学会論文集, No.653, IV-48, pp.39-48, 2000.

[3] D. Braess, A. Nagurney, Tine Wakolbinger, “On a Paradox of Traffic Planning”, Transportation Science, Vol. 39, No 4, pp.446-450, 2005.