

道路ネットワークの変化に対する経路選択の学習

内田 英明^{1,a)} 藤井 秀樹^{1,b)} 吉村 忍^{1,c)} 荒井 幸代^{2,d)}

受付日 2012年1月27日, 採録日 2012年7月2日

概要: 現在, 多くの交通流シミュレーションでは定常状態の交通流を評価するモデルが採用されており, 過渡的な現象を再現することは困難である. しかし, 道路ネットワークの変更をとまなう交通施策の評価においてこの過渡的な現象を無視することはできない. そこで本研究では, 道路ネットワークが変化する状況下でどちらの現象も扱うことのできる経路選択モデルを新たに提案し, マルチエージェント型交通流シミュレータに実装する. はじめに, 基本として用いた強化学習の枠組みに基づく Q-routing アルゴリズムの説明を行い, 交通流シミュレーションに適用するため加えたいくつかの改良について説明する. 不規則な格子状のネットワークにおいて, 渋滞や信号制御に対する本経路選択モデルの基本的な振舞いを検証し, 最後に岡山市内の現実の LRT 延伸計画を対象としてシミュレーションを行い, 延伸実行後に小規模な渋滞が発生する可能性を示す. また, この渋滞現象は運転者の保持する延伸前の走行経験がバイアスとして影響した結果, 一時的に生じるもので, 十分な時間経過の後定常状態に至ることを示す.

キーワード: マルチエージェントシステム, 交通流シミュレーション, 強化学習, Q-routing

Learning Routing Policy for Changes in Road Network

HIDEAKI UCHIDA^{1,a)} HIDEKI FUJII^{1,b)} SHINOBU YOSHIMURA^{1,c)} SACHIYO ARAI^{2,d)}

Received: January 27, 2012, Accepted: July 2, 2012

Abstract: This paper describes the impact of changes in road network on driver's behavior. In agent-based traffic simulations, agents typically choose the shortest route, while drivers in a real world choose their routes through their own past experience. Since traffic simulations are used extensively in the evaluation and verification of traffic policy, accurate simulation reproducing the routing behavior of the real world is strongly demanded. Thus, we newly develop a reinforcement learning based routing algorithm, and implement it in Traffic Simulator. Then we explain some improvements in the Q-routing for traffic simulation. Firstly, we perform a preliminary experiment using an irregular grid network with various loads or signal control, and get good performance, i.e. robustness of the improved Q-routing in the cases of heavy traffic with signal control. Second, the simulator with/without the improved Q-routing is applied to simulate the LRT expansion project in Okayama city, and we evaluate transient traffic behaviors after the implementation of the plan. We observe transient congestion phenomenon only in the simulation with the improved Q-routing. Though the detailed analysis of the results, we find that such transient traffic congestions are caused due to the bias of the drivers' past experience.

Keywords: multi-agent system, traffic simulation, reinforcement learning, Q-routing

¹ 東京大学大学院工学系研究科
School of Engineering, The University of Tokyo, Bunkyo,
Tokyo 113-8656, Japan

² 千葉大学大学院工学研究科
Graduate School of Engineering, Chiba University, Inage,
Chiba 263-8522, Japan

a) uchida@save.sys.t.u-tokyo.ac.jp

b) fujii@sys.t.u-tokyo.ac.jp

c) yoshi@sys.t.u-tokyo.ac.jp

d) arai@tu.chiba-u.ac.jp

1. はじめに

道路交通は現代社会の基盤システムであると同時に大気汚染や交通渋滞などの問題を生んできた. これらの問題を解決するため様々な施策が提案されており, 近年ではその評価に交通流シミュレーションが採用されることが多い. これは交通施策の実証実験には多大なコストが必要となり,

代替手段としてのシミュレーションの需要が大きくなっているためである。しかしシミュレーションにおいて再現される交通流にはいくつかの注意すべき性質が存在する。その1つが本研究で扱う運転者の経路選択行動である。

多くのシミュレータにおいて運転者の経路選択はネットワークのコスト情報に基づいて行われる。一方、現実の運転者は限定された周囲の状況と経路に対するこれまでの知識に基づいてルートを選択する。この差異はシミュレーションによって再現される交通現象に大きく影響する。たとえば、交通施策の実施によりある時点から道路ネットワークが変化するという条件のもとでシミュレーションを行うとすると、ネットワークのコスト情報を瞬時に利用できるモデルではつねに各運転者が最適な経路選択を行うため、交通状況は初めから定常的な状態に落ち着く。しかし、過去の走行経験に基づいて経路選択が行われるとすれば、知識を修正していく過程で過渡的な交通状況が生じると考えられる。

過去の多くの研究成果は定常状態の予測にフォーカスしたものであるが、自治体などは交通施策の実行に際し、短期的にどのような現象が発生する可能性があるのかを把握することが必要である。たとえ一時的なものであっても過度の渋滞や交通事故などの現象が発生することは道路交通の特性上避けるべき事柄であるが、それがどの程度の期間継続し住民や利用者に負担を強いることになるかを事前評価することは、十分な時間経過の後収束する定常状態を予測することと同様に重要である。過渡期にどのような現象が発生するか知ることにより、その対策を立てることも可能になるためである。

そこで本研究では、マルチエージェント型交通流シミュレータに強化学習に基づいたルーティングアルゴリズムである Q-routing を導入し、過去の経験に基づいた経路選択の実現を目指す。同時に、岡山市で検討されている路面電車の軌道延伸計画に関し、実施直後の過渡的な交通状況から定常状態に収束するまでをシミュレーションする。

2. 交通流シミュレータにおける経路選択

2.1 従来のアルゴリズムの問題点

多くのマルチエージェント型交通流シミュレータでは、ダイクストラ法 [1] あるいは A* [2] による経路選択を採用している。これらはともに代表的なオフライン探索アルゴリズムであり、交通流シミュレータに限らず経路探索の分野で広く用いられている [3]、任意の 2 地点間の最短経路を求めるアルゴリズムである。

しかしダイクストラ法や A* などによる経路探索はネットワーク全体にわたって各リンクのコストを取得できることが前提であり、ある地点での道路ネットワークの変化を全エージェントが瞬時に認識するモデルとなっている。結果として、エージェントはネットワークの変化が起こった

直後から安定した経路選択を実現できてしまい、本来当初は判断できないような車線数の減少に起因する迂回行動など、現実と乖離した挙動を引き起こす。このように、定常状態の交通流をうまく表現できるモデルが、過渡状態の交通流をうまく表現できないという問題は、オフライン探索アルゴリズムを採用する既存の多くの交通流シミュレータに共通する。次節以降、過渡状態を表現するうえで満たすべきエージェントおよびシミュレータの要件をまとめ、過渡状態・定常状態の双方を表現する本研究の経路選択モデルの位置づけを説明する。

2.2 道路ネットワークの変化をともなう環境下での経路選択

交通流の過渡状態の発生には、道路ネットワークの変化、道路の交通容量の変化（工事や事故による車線閉塞など）といった環境的な要因のほかに、発生交通量の時間的な変化、信号現示の時間的な変化などの要因が考えられるが、いずれも「現在の交通状態とエージェントの推測する交通状態の不一致」に起因し、その不一致について時間をかけて修正する過程が過渡状態であると見なせる。そのため、以上の問題を解消するためには、エージェントの経路選択モデルが次の要件を満たす必要があると考える。

- (1) 周囲の限定的な環境から状態観測を行う。
- (2) 過去の走行経験により事前情報を持つ。
- (3) 経路選択について説明可能なログを出力する。

本研究ではこの要件を満たすものとして強化学習法に基づく経路選択モデルを採用する。

はじめに、状態観測に関し、道路ネットワークのコスト情報が瞬時に取得可能な条件下で経路選択を行うことの問題点はすでに指摘した。強化学習法ではエージェントの状態入力を限定することで容易に (1) の条件を満たすことができる。次に事前情報について、状態観測によって取得できない情報は、過去の走行経験を利用して補完する。強化学習においては価値関数で行動の期待報酬値を保存しているため、過去の走行経験を利用でき、(2) の条件を満たせる。岡山市の路面電車延伸計画のような交通施策についてシミュレーションを行う場合、経路選択を決定する過程は説明可能なものでなくてはならない。確率的な要素の大きい発見的手法に比べ強化学習の行動と報酬は因果関係が明確であり、得られる価値関数はスカラー値のログをシミュレーションのステップごとに出力することができ、(3) の条件を満たすことができる。

2.3 既存の交通流シミュレータ

本研究で提案するアルゴリズムを実装するシミュレータの要件として、以下の項目をあげることができる。

- (a) ミクロシミュレータである。
 (b) 経路選択モデルを内包する。

交通流シミュレータはその詳細度や適用範囲の観点からマクロシミュレータとミクロシミュレータの2つに分類される。前者は流体モデルによって交通流を再現するため数学的な性質が良く、道路ネットワークにおける均衡解を導出するなど、解析的な議論が容易である。後者はより詳細な車両挙動を再現することが可能である。本研究ではエージェントの経路選択の変化を観察するため、個々のエージェントを離散的に扱えるミクロシミュレータが必要不可欠である。

一方、交通流シミュレータに実装されている経路選択行動は、経路選択モデルを外生するもの経路選択モデルを内包するものの2つに大別することができる。前者のモデルは発生交通量などを与える際、同時に走行経路を設定するものである。このような設計ではシミュレーションのリアルタイム性は向上するものの、エージェントの経路選択が限定されてしまい、動的な交通状況の変化を反映することができない。後者のモデルではエージェントの行動モデルに経路選択が組み込まれており、経路の一般化費用を算出することでシミュレーションの状況に応じた経路選択を行うことができる。なお、本論文で提案するアルゴリズムは多くのミクロシミュレータに実装可能である。

エージェントの経路選択モデルに学習モジュールを組み込んだ例としては MATSim (Multi-Agent Transport Simulation) [7] がある。MATSim は、各エージェントの1日の行動と学習をシミュレートすることで交通状況の変化を再現する。MATSim において、エージェントは OD (Origin-Destination) や経由地、出発・到着時刻や滞在時間を1日ごとの行動計画プランとして保持し、日々のシミュレーションを繰り返す。そしてその結果を基に各エージェントが学習を行い、1日のプランとシミュレーション結果との誤差から評価値を算出して、その値が最良のプランを実行する。同時に、観測に基づく各経路の予想通過時間などから新たなプランを作成することもできる。

また、学習ではないがファジィ推論による経路選択を組み込んだ例として MITRAM (Microscopic model for analyzing Traffic jams in the city area) [8] がある。MITRAM ではドライビングシミュレーションによる被験者実験やアンケート調査からファジィ推論則を生成し、AHP および多項ロジットモデルと組み合わせた経路選択を行う。得られた出力は現実の交通データによる検証によって調整される。これによりエージェントの多様性をあいまいさによって表現し、学習に比べ普遍的な（走行経験によって更新されない）経験則に基づいた経路選択を実現する。

このように経路選択1つをとっても様々なモデルが提案

されている。その中で、今回本研究で基本モデルとして採用する Q-routing には走行しながらリアルタイムに学習と経路選択を行えるという特徴がある。MATSim のように出発の時点で走行プランを作成する必要がないため、たとえば交通事故による突発的な渋滞に対しても状態入力を基にエージェントが自律的な走行を継続する。また、強化学習ベースであることから、なぜその経路が選択されたかを定量的に解析することが可能である。MITRAM ではファジィ推論則などを用いている運転者の経験則も、Q-routing ではスカラーの価値関数として表現されるため扱いが容易である。つまりこのどちらのシミュレータも、前節のエージェントの要件に照らすとはじめの2項目は満たすものの、「経路選択について説明可能なログを出力する」の項目が欠けていると考えられるため、新たにすべての要件を満たすモデルとして Q-routing を採用する。

3. Q-routing とその改良

3.1 Q-routing の概要

Q-routing とは強化学習の1つである Q-learning の枠組みに基づいた自律分散型のルーティングアルゴリズムである [4]。この手法はバケットルーティングの分野において提案され、多様なネットワーク環境への適応力が高いことで知られている [4], [5], [6]。Q-routing では、各ノード x が隣接ノード y とデスティネーションノード d の組に対してそれぞれ価値関数 $Q_x(y, d)$ を持ち、これを基にルーティングを決定する。ここで $Q_x(y, d)$ とは、 x から隣接ノード y を経由してデスティネーションノード d に到達するまでにかかる送信時間の推定値を表す。送信時間はノードにおける待ち時間とリンク上をバケットが移動する時間で占められるが、その大部分はノードにおける待ち時間である。

3.2 価値関数の更新

各ノード x において、バケットが移動する隣接ノード y は価値関数 $Q_x(y, d)$ が最小となるものを選択する。バケットは y に移動するとただちに $Q_x(y, d)$ の更新に必要となる q_y をローカルに情報交換する。 q_y は y からデスティネーションノード d までの推定配送時間の最小値であり、式 (1) に従って与えられる。

また、ノード x の待ち行列における待ち時間と x から y までのリンクにおける遅延時間の和を t_x としたとき、 $Q_x(y, d)$ の更新式は式 (2) のようになる。またこの更新のモデルを図 1 に示した。

$$q_y = \min_{z \in \text{neighbors of } y} Q_y(z, d) \quad (1)$$

$$Q_x(y, d) \leftarrow Q_x(y, d) + \alpha(t_x + q_y - Q_x(y, d)) \quad (2)$$

ここで α は学習率 ($0 < \alpha \leq 1$) を表す。この式は観測された最新の推定配送時間に $Q_x(y, d)$ を漸近させるため、輻輳が発生するようなノードの $Q_x(y, d)$ は増大しホップ先

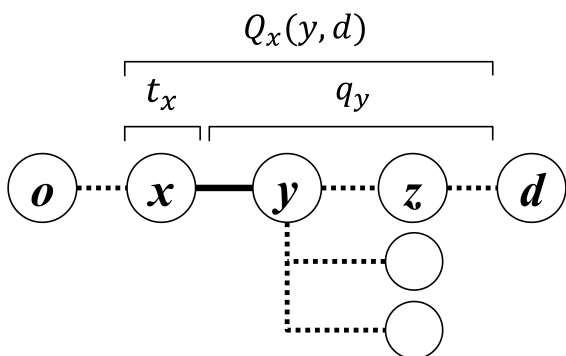


図 1 Q-routing の概略

Fig. 1 The framework of Q-routing.

として選択されにくくなる。

3.3 交通流シミュレータへの実装

本研究では、上述した Q-routing の交通流シミュレーションにおける経路選択行動への適用を提案する。パケットルーティングと交通流シミュレーションにおける経路選択の違いは、後者ではエージェントが物理的なボリュームを持つことと、ネットワークの性質が異なることである。通信ネットワークではノードでの待ち時間が送信時間を規定したが、ノードを交差点、リンクを単路と考えると、道路ネットワークでは単路における移動時間が旅行時間に相当する。待ち行列の取扱いと自律的に走行するエージェントの経路選択は異なるため、Q-routing を交通流シミュレーションに応用するために新たな改良を行う必要がある。そこで以下の事項に留意しながら交通流シミュレータへの実装を行った。これらは前章で述べたとおり多くのマイクロシミュレータに適用可能であるが、本研究では知的マルチエージェント型交通流シミュレータ MATES (Multi-Agent based Traffic and Environment Simulator) [9], [10], [11], [12] を対象に、MATES がこれまで持っていた経路選択アルゴリズムを修正する形で実装した。

3.3.1 意思決定のタイミングの変更

パケットと異なり車両には物理的な制約が存在する。たとえば次の交差点を右折するような経路を選択した場合、交差点に進入する以前に右折可能なレーンに車線変更しておく必要がある。そのため図 1 に示した従来の Q-routing のように直近の隣接ノード y を選択するのではなく、さらに 1 つ先の交差点 z に関して意思決定を行う必要がある。

価値関数 $Q_x(y, d)$ を、あるエージェントが交差点 x から隣接交差点 y を経由して目的地 d に到着するまでの推定旅行時間と再定義した新たなモデルを図 2 に示す。旅行時間 t_x が交差点ではなく単路にかかる点や、意思決定を行うための価値関数の最小値を隣接交差点 y ではなく、その先の z から取得する点が図 1 のモデルとは異なっている。

3.3.2 滞留発生時の予測更新

道路交通において右折待ちや信号待ちによる滞留は一

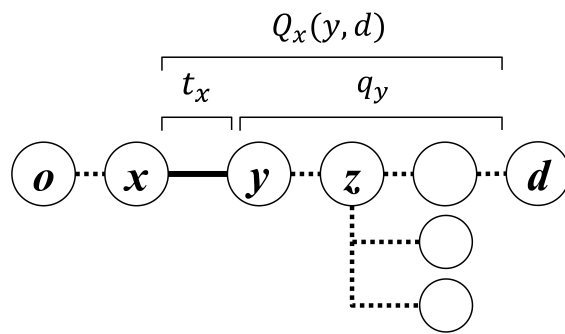


図 2 改良後の Q-routing の概略

Fig. 2 The framework of improved Q-routing.

般的に見られる現象である。しかし Q-routing における $Q_x(y, d)$ の更新は交差点に到達した時点で行われるため、単路内での滞留はこの更新のタイミングを遅らせることにつながる。後述する理由から、本研究では共通の目的地を持つエージェント群は価値関数を共有するものとしているため、このような現象は本来渋滞している経路へのルーティングを誘発し、最悪の場合はグリッドロックを引き起こす可能性がある。さらに、信号制御によるリンク（単路）内の滞留は通信ネットワークで想定されていないため、新たに考慮すべき点である。本研究では混雑の状況を素早く反映することを目的に、滞留の先頭にいるエージェントの価値関数の更新タイミングを前倒しすることを考える。先頭のエージェントは青信号になると同時に交差点へ侵入できるため、残りの赤信号の点灯時間をあらかじめ見積りに含んだ予測旅行時間を価値関数の更新に使うこととする。その後、青信号になり真の旅行時間が確定した時点で、実測値を用いた価値関数の補正も行う。更新式を式 (3)–(5) に示す。

予測旅行時間 \tilde{t}_x は、滞留の先頭に存在するエージェントが赤信号により停車し滞留が開始した時点での移動時間 t'_x と、信号待ち時間の期待値の和によって見積もる。信号待ち時間の期待値は赤信号点灯時間である $RedSignal$ の $1/2$ とした。また、最終的に確定する単路の旅行時間 t_x による補正は式 (5) のとおりである。式 (5) を式 (3) に代入することで本来の更新式である式 (2) に一致する。

滞留の先頭で赤信号により停車

$$Q_x(y, d) \leftarrow Q_x(y, d) + \alpha(\tilde{t}_x + q_y - Q_x(y, d)) \quad (3)$$

$$\tilde{t}_x = t'_x + \frac{RedSignal}{2} \quad (4)$$

青信号になり交差点に進入

$$Q_x(y, d) \leftarrow Q_x(y, d) - \alpha(\tilde{t}_x - t_x) \quad (5)$$

3.3.3 道路ネットワーク変化時の価値関数補完

交通施策の評価を行う場合、道路ネットワークの変化について扱う必要がある。ネットワークの変化は表 1 に示した 4 つが考えられる。ここでトランジットモールとは、公

表 1 交通施策の事例
Table 1 Various traffic policy.

ネットワークの変化	交通施策
車線数の増加	道路の拡張, 右折レーンの新設
車線数の減少	事故・工事による車線閉塞
単路の増加	バイパスの新設
単路の削除	トランジットモールの導入

公共交通機関のみ進入・運行が許可されている歩車共存道路のことである。

本研究では、施策の適用前の学習結果である価値関数を、適用後のネットワークのシミュレーションで初期値として用いる。このとき、車線数の増減や単路の削除は問題なく価値関数を読み込むが、単路が増加するケースでは価値関数に新たな要素が発生する。たとえば2つの交差点 a と b の間に単路が新設されたとき、交差点 a から交差点 b を経由して目的地 d へ向かう際の価値関数 $Q_a(b, d)$ は未知である。施策適用後のシミュレーション時にこの要素の初期値を極端な値に設定してしまうと学習に悪影響を及ぼす懸念があるため、ここでは交差点 b における推定旅行時間の最小値 q_b を代入する。式 (1) によれば、この値は真の $Q_a(b, d)$ よりも ab 間の旅行時間 t_a だけ小さく、この経路が本来よりもわずかに選択されやすくなっている。 t_a についてはどのエージェントも走行経験を持たないため、学習初期に選択確率を高めておくことで早期に正確な推定値を得ることができる。

また、実験における行動選択には ϵ -greedy 戦略を導入した。これはエージェントが高い確率 $1-\epsilon$ で価値関数を最小化するルートを取り、低い確率 ϵ でその他のルートをランダムに選択するものである。エージェントにランダムな行動選択の機会が与えられることによって選択可能なルートの価値関数を万遍なく更新するため、局所解への収束を回避でき、同時に現実起こりうるランダム性をモデル化することにもつながる。また、学習が収束した後に道路状況が変化するようなケースでも、新たな経路を容易に学習しなおすことが可能となる。ここまでの内容を、Q-routing のシミュレーションフローとして図 3 に示す。

次章以降で行う実験に向け、現在 MATES で採用している A^* に従うエージェントと Q-routing に従うエージェントが共存する形での実装を行った。同時に、共通の目的地を持つエージェント群は価値関数を共有するものとした。この設定によりエージェントは自らの走行経験以外の要素(たとえばエージェントどうしの情報交換による影響や、その街・道路ネットワークにおける暗黙知の形成)を考慮でき、明示的にエージェントの協調行動を誘発させる必要なく学習を収束させることが可能である。これは、異なる時間軸に存在する車を同時に走らせるという仮想的なシナリオととらえることができる。たとえば、あるエージェント

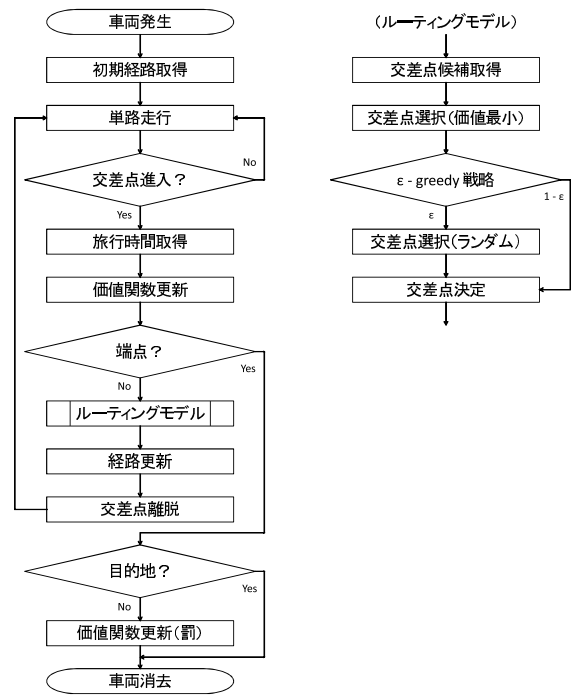


図 3 Q-routing のフローチャート

Fig. 3 Flowchart.

の任意の t 日目の走行経験は自身の $t+1$ 日目(あるいは近い将来)の経路選択に影響を及ぼすため、価値関数を共有していても結果的に問題がない。また、シミュレータ上に存在する多数のエージェント間での同時学習の問題も解決することができる。

4. 予備実験

4.1 実験設定

前章で述べた Q-routing による経路選択モデルの性能検証のために、先行研究 [4] を参考に不規則格子の道路ネットワークを用いた予備実験を行った。図 4 に道路ネットワークのトポロジを示す。端点となる各ノードからエージェントが流入し、ランダムな目的地を持った車両がそれぞれ等しい任意の交通量で発生する。エージェントは目的地に到着した時点でネットワークから消去され、発生から到着までの時間を旅行時間として記録する。このネットワークの特徴として左右のネットワークをつなぐ中継リンクが 10-11 と 26-27 の 2 本に制限されている点があげられる。このため、すべてのエージェントが最短経路を選択すれば中央部の中継リンク 26-27 において渋滞が発生する。そこで、このリンクを含む推定旅行時間が長い場合には上部の中継リンク 10-11 へ迂回する経路選択を行う必要がある。

実験では $Q_x(y, d)$ の初期値を 0, 学習率 $\alpha = 0.3$, $\epsilon = 0.05$ とし、発生交通量が各端点で 50, 300 [台/h] とした。また、Q-routing の提案された通信ネットワークでは存在しなかった信号制御による滞留の影響への適応可能性を検証するため、各交差点において信号制御をあり、なしとした 2

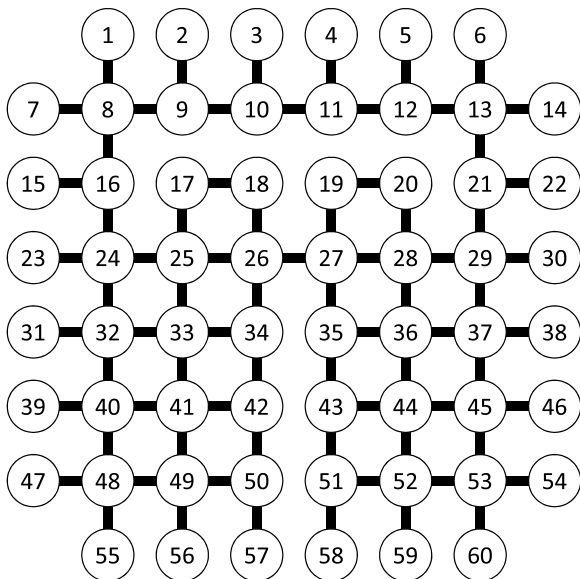


図 4 実験に使用した道路ネットワーク
 Fig. 4 The irregular grid topology.

つのケースを想定した。制御なしの場合は信号の現示はつねに青であるが、MATES ではエージェントがお互いに衝突を回避する行動をとる。また制御ありの場合の信号パラメータはサイクル長 140 秒，点灯時間は交差する両方向ともに赤 70 秒，青 55 秒，右矢印 10 秒，黄色 5 秒とし，オフセットは設定しなかった。

このとき，すべてのエージェントが Q-routing による経路選択を行うとしたモデルとの比較のために，A* に従って経路選択を行う場合のシミュレーションも行った。本論文では，距離コストを最小化する A* を A*-distance，旅行時間コストを最小化するものを A*-time と呼ぶこととする。このうち A*-time は旅行時間を最小化する経路探索を行うことが数理的に保証されており，定常状態における交通状況は十分に再現可能であるため，学習収束後の Q-routing との比較対象として活用する。

4.2 発生交通量に対する比較

図 5, 図 6, 図 7 に交通量別の各ケースの旅行時間の推移を示す。横軸はシミュレーション時間 [min]，縦軸は 10 分ごとの平均旅行時間 [sec] である。ここでは信号制御は行っていない。まず図 5 に示した交通量の少ない 50 [台/h] のケースでは，A*-time・A*-distance とともに平均旅行時間が一定であり渋滞が発生していないことが分かる。一方 Q-routing はシミュレーション開始直後に平均旅行時間が増大し，学習が進展するにつれ開始後 90 分程度までには一定の収束が見られる。学習収束後の Q-routing が A* と比較し依然として旅行時間が大きいのは ϵ -greedy による行動選択の影響であり，ここでは $\epsilon = 0.05$ の割合で $Q_x(y, d)$ が最小ではない交差点を選択するためである。この実験において推定された $Q_x(y, d)$ を用い，新たに $\epsilon = 0$ とした実

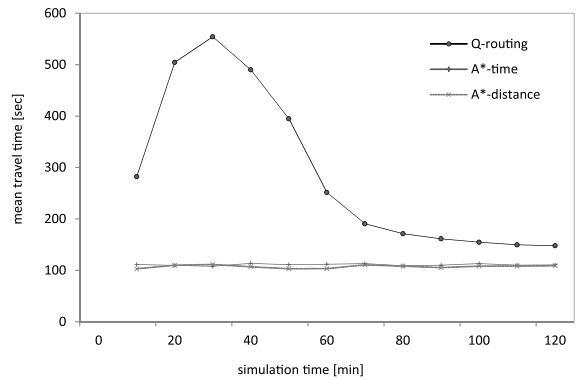


図 5 発生交通量 50 [台/h] のケースの旅行時間
 Fig. 5 Simulated result of mean travel time (Case: Network with a low load).

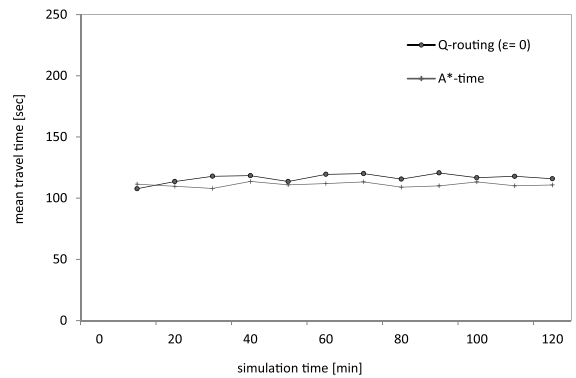


図 6 発生交通量 50 [台/h], $\epsilon = 0$ としたケースの旅行時間
 Fig. 6 Simulated result of mean travel time ($\epsilon = 0$).

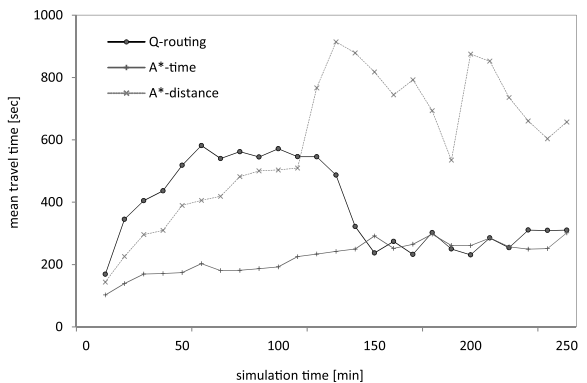


図 7 発生交通量 300 [台/h] のケースの旅行時間
 Fig. 7 Simulated result of mean travel time (Case: Network with a high load).

験を行った Q-routing の結果は，図 6 のように A* とほぼ同じになることを確認することができる。

また，交通量が多い 300 [台/h] のケースを図 7 に示す。ここでは A*-time が最も平均旅行時間が短く，A*-distance は時間が進展するにつれ旅行時間が増大していくことが分かる。単に最短距離を選択するだけでは，このネットワークの特徴である中央部の中継リンク 26-27 への交通集中により渋滞が発生するためである。一方 Q-routing は 50 [台/h] のケースに比べ収束に時間がかかるものの，A*-distance

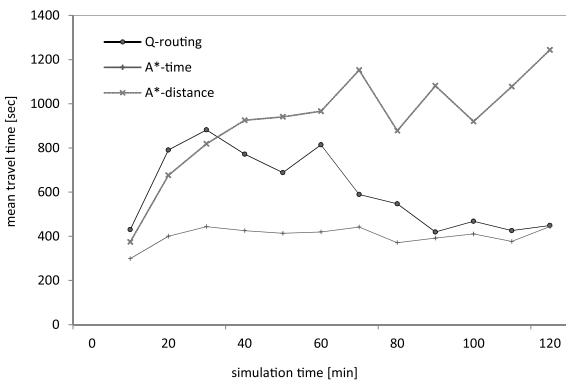


図 8 信号制御ありとしたケースの旅行時間

Fig. 8 Simulated result of mean travel time (Case: traffic light is available).

よりも小さい値に収束し、交通量の増加に対しても問題なく収束することが確かめられた。

4.3 信号制御の有無に対する比較

次に各交差点において信号制御を導入したケースの実験を行った。ここでも比較対象として A*-distance と A*-time を用い、発生交通量は各端点で 50 [台/h] とした。

図 8 に旅行時間の推移を示す。これまでと同様、横軸はシミュレーション時間、縦軸は 10 分ごとの平均旅行時間である。A*-time はネットワーク全体の情報を活用できるため、信号制御下でも一定して良好な平均旅行時間を示した。また A*-distance はシミュレーションが進むにつれ旅行時間が増大した。これは発生交通量 50 [台/h] で信号機制御を行わないケース (5) では見られなかった現象であることから、信号待ちによるエージェントの滞留が進展し渋滞が起きたことが分かる。A*-distance は信号の有無にかかわらず最短距離の経路をとるためこのような現象が起こったと考えられる。

一方 Q-routing はシミュレーションが進展するにつれ旅行時間が減少し、開始 90 分ごろからは A*-time と同等の良好な旅行時間が得られた。これは学習によって信号のサイクル自体を価値関数に内包したことによる。また、収束後の平均旅行時間が Q-routing と A*-time で変わらないことから、信号によって動的に変化する交通流に対しても Q-routing が適応することが示された。

4.4 まとめ

予備実験では改良した Q-routing の基本的な性質を確かめ、経路選択モデルとしての性能検証を行った。その結果、十分な時間経過の後収束する定常状態は、数理的に最適性が保証されている A*-time と同等の旅行時間となった。このことから Q-routing は、交通施策の仮想社会実験を実施する際に、過渡期の現象が再現可能であり、かつ定常状態についても従来モデルと同等の妥当性を主張することがで

きる。

5. 路面電車の軌道延伸シミュレーション

5.1 岡山市における路面電車軌道延伸計画

2010 年 4 月、岡山商工会議所は市中心部の路面電車の軌道延伸・環状化を核に地域活性化を図る「人と緑の都心 1km スクエア構想」の推進・実現への協力を岡山市に要望した。また同 5 月には岡山電気軌道を傘下に持つ両備グループが同様の構想をまとめ、さらに、路面電車と都市の未来を考える会 (RACDA) は後楽園など観光施設へのアクセス向上を含めた複数の環状化案を提起している [13]。これらの概略を図 9 に示す。路面電車延伸は岡山市において以前から繰り返し提案されてきたものの、様々な問題から計画が難航していたものである。しかし、近年の富山ライトレールの成功などから路面電車の有効性が再認識されるようになった。一方で、このような大規模な交通施策については事前に行われるべき社会実験についてもコストが高いと予想されるため、計算機による現実的なシミュレーションの需要が存在すると考えられる。また、延伸が計画されている経路は主要な幹線道路であり多くの運転者に影響を与えることから、交通流が定常状態に収束するまでに一定以上の時間を要することが想定される。計画実行に先立つ社会実験の実施期間がこの過渡状態の期間より短ければ、本来の延伸計画による効果を評価できない可能性がある。同時に、軌道の敷設をとるため 1 度実行されると容易に復元できない不可逆的な計画であり、計画実行後に渋滞が発生したとすればその渋滞が永続的なものか過渡的なものか、また過渡的であればどの程度の期間で収まるのかをシミュレーションすることに大きな意義がある。

5.2 延伸による影響

本研究では、路面電車延伸の影響としては、(1) 路面電車の軌道敷設による主要道の車線数減少、(2) 路面電車への乗換による交通量の減少、の 2 点を想定した。(1) については、延伸後のシミュレーションで延伸経路の車線数を片側 1 車線ずつ減少させることを考える。また (2) については、近年路面電車が導入された富山市における富山ライトレールの事例を活用した。

富山市が行った LRT 化の整備効果調査 [14] よれば、富山ライトレール以前に運営されていた JR 富山港線の平日の利用者が 2,331 人、開業 1 年間で自動車から乗り換えた利用者が 572 人であった。岡山市の路面電車は平日 1 時間あたり約 600 人の利用者が存在するため自動車からの乗換えは 147 人とすることができる。道路交通センサス [15] によれば平日に走行する乗用車の平均乗員数は 1.33 [人/台] であることから、路面電車延伸による交通量の減少は 1 時間あたり 111 台となる。

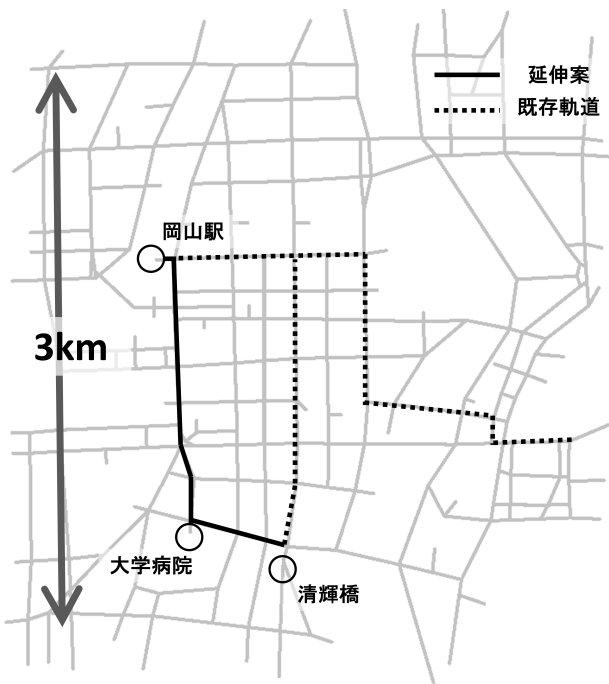


図 9 路面電車延伸案 (岡山市中心部)

Fig. 9 The LRT expansion project in Okayama city.

5.3 実験設定

路面電車軌道延伸計画を扱うため、図 9 に示した岡山市の中心部の道路ネットワークを実験環境として使用する。領域は約 3km 四方であり、262 ノード (交差点)・381 リンク (単路) である。エージェントが発生するのはグラフの端点であり、本実験における岡山市の道路データでは、主に対象地域の境界部に端点が存在する。また、対象地域内部の主要な集客施設 (大学病院や百貨店) の駐車場にあたる地点にも端点を設定している。信号の制御方法については予備実験と同様の設定とし、主要な信号 4 カ所については実測値を使用した。また、交通量は 2001 年に行われた社会実験のデータを基に設定し、平日の平均的な 2 時間を再現するため時間あたりの発生交通量は一定とした。シミュレーションは延伸前 1 ケース、延伸後 3 ケースの計 3 ケース行う。延伸前は価値関数の各要素を 0 で初期化した Q-routing、延伸後はこの学習済みの価値関数を初期値として利用する Q-routing (learned) と、再び価値関数の各要素を 0 で初期化した Q-routing、学習をまったく行わない従来の non learning (A*-time) である。

延伸による影響としては、延伸後のシミュレーションで図 9 に図示した延伸経路の車線数を片側 1 車線ずつ減少させた。また、前節で算出した乗換利用者数を考慮し、図 9 の岡山駅～大学病院の発生交通量を毎時 100 台減少させた。

学習率 $\alpha = 0.3$, $\epsilon = 0.03$ とした。また、規模の大きなネットワークで実験を行うにあたり学習を行うエージェントを限定した。図 9 に図示した岡山駅～大学病院や岡山駅～清輝橋といった OD を持つエージェントは路面電車延伸の影響を直接的に受けるため、Q-routing による学習を

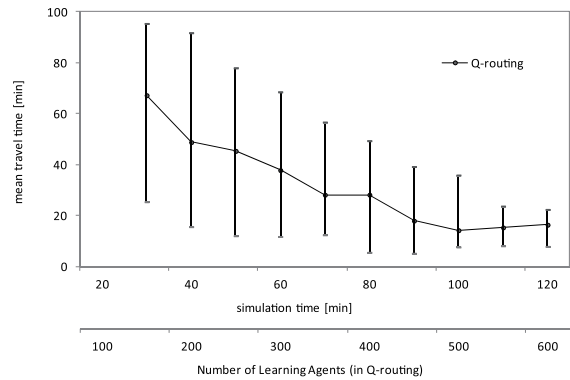


図 10 路面電車延伸前の平均旅行時間

Fig. 10 Simulated result of mean travel time before the implementation of the LRT expansion.

行い、その他のエージェントは A*-time によって経路選択を行うこととした。

5.4 実験結果

はじめに路面電車延伸前の道路ネットワークでシミュレーションを行った。図 10 は図 9 に示した清輝橋から岡山駅へ向かうエージェントの平均旅行時間を示す。縦軸は 10 分ごとの平均旅行時間 [min]、横軸はシミュレーション時間 [min] である。プロットは平均旅行時間で、各 10 分間の最大値と最小値をエラーバーで表記している。また、横軸には学習エージェントの発生数も併記した。これは、経路の学習を行うエージェントの発生数を観察することで現実の運転者の行動と関連付けるためである。

シミュレーション開始直後は学習がほとんど進んでいない状態のため旅行時間は長く、また分散も大きくなっているが、100 分経過後にはほぼ学習が収束している様子が確認できる。ネットワークにエージェントが飽和してから交通流が定常状態に遷移するまでに、約 400 台のエージェントが学習を行っていることが分かる。

次に延伸後の道路ネットワークでシミュレーションを行った。同様に清輝橋から岡山駅へ向かうエージェントの平均旅行時間を図 11 に示す。縦軸を 10 分ごとの平均旅行時間 [min]、横軸をシミュレーション時間 [min] および学習エージェントの発生数とした。

結果を比較すると non learning (A*-time)、Q-routing (learned)、Q-routing の順に旅行時間が小さくなっている。これまでどおり、non learning (A*-time) はシミュレーション開始直後から定常状態を示し、また、Q-routing は環境同定に大きなコストがかかるため過渡状態の旅行時間が非常に大きくなる。しかし、今回新たに導入した Q-routing (learned) は経路選択についてすでにある程度の経験を保持している状態であるため、学習中の旅行時間は最大でも収束後の 30%程度までしか増加しない。この Q-routing (learned) について、シミュレーションの各時間帯でどのよ

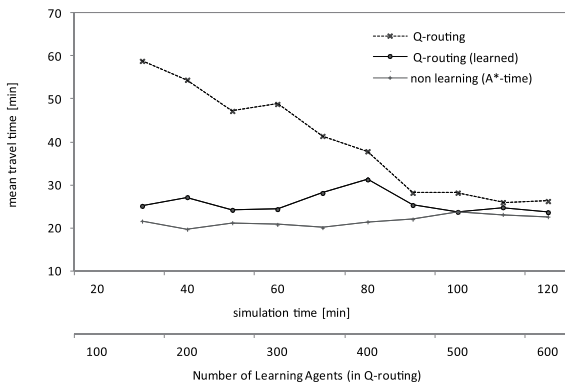


図 11 路面電車延伸後の平均旅行時間

Fig. 11 Simulated result of mean travel time after the implementation of the LRT expansion.

うな現象が発生していたかを詳しく見ていく。

5.4.1 非飽和期 (0~30分)

ネットワークが大規模であるため、シミュレーション開始直後はエージェントが道路に行きわたっていない状況が継続する。エージェントは各自自由走行している状態である。

5.4.2 再学習期 (30~100分)

エージェントが飽和し、路面電車延伸により狭くなった大学病院~岡山駅の経路が混雑するようになると、延伸前のネットワークで学習済み価値関数から推定される旅行時間と、実際の旅行時間が次第に乖離し始める。それに合わせて新たな迂回経路の学習が始まり、試行錯誤の過程でさらに渋滞の程度が増す。ただし再学習は局所的な経路にとどまり、延伸経路との相互作用の弱い道路についてはこれまでの価値関数を流用することが可能であるため、Q-routing (learned) の過渡状態における渋滞現象はQ-routingほど深刻なものには進展しない。

5.4.3 収束期 (100分~)

迂回経路の再学習が収束し、定常状態に遷移する。旅行時間は non learning (A*-time) がシミュレーション開始直後から達成している定常状態のそれと同程度である。

再学習期にあたる過渡的な渋滞はシミュレーション上で約70分間継続したが、この期間は実際のタイムスケールを定量的に示すものではない。ただし、継続期間を学習エージェントの発生台数に読み替えると約350台となる事が分かる。これは出発地点から到着地点までの移動、つまりQ-routingの価値関数を350回更新したことに相当する。また、延伸前後のシミュレーションを比較すると収束した平均旅行時間は延伸前のQ-routingが約16分、延伸後のQ-routing (learned) は約23分であった。ただしこの数値は延伸の影響を最も大きく受ける経路の1つで計測したものであり、すべてのエージェントの旅行時間が悪化するわけではない。一方、延伸後のQ-routing (learned) における旅行時間の最大値は約31分であり、エージェントが再学

習によって環境に適応していく過程では特定の経路に車両が集中し比較的大きな渋滞が発生してしまう可能性を示している。

今回のシミュレーションの結果、再学習期において過渡的な渋滞現象が生じた。これはQ-routing (learned) において延伸前の走行経路がバイアスとなり、新たな環境に適応するまで時間を要したことによる。しかし、学習中の旅行時間の増大はQ-routingに比べ小さく、延伸前の走行経路が多分に有用な知見を含むことも同時に示している。これまで多くの交通流シミュレータでは交通施策の計画に対し、長期的視点に立った定常状態の予測のみで意思決定を行っていた。しかし実際には本実験で示したような短期的で複雑な過渡状態が存在する可能性があるため、定常状態と過渡状態の2つの交通現象をどちらも考慮したうえでの実行判断が必要であるといえる。

6. おわりに

本研究ではより現実的な交通施策の評価を目標とし、運転者の過去の走行経路を経路選択に反映させるため、交通流シミュレーションにエージェントの学習機能を実装した。実装に際し、通信ネットワークと交通ネットワークの相違点に着目しながらQ-routingに改良を加えた。

続いて不規則格子での予備実験を行い、Q-routingによる経路選択が過渡的な交通状況を再現するだけでなく、定常状態についてもA*と同等の性能を示すことを確認した。岡山市の路面電車軌道延伸計画を扱った実験では複雑なネットワーク・信号制御といった環境下でも学習が収束することを示すと同時に、従来のシミュレータでは再現できなかった、エージェントのバイアスによる延伸後の過渡的な渋滞現象を確認することができた。

今後の課題としては、過渡的な渋滞が収束し定常状態に至る過程で、経路選択がどのように遷移していったかをさらに詳細に解析することがあげられる。これにより、学習の収束を促進する有用な情報(価値関数)の抽出などを目指す。また、今回の実験ではQ-routingにおける行動選択にε-greedy戦略を用いた。しかしこの方法ではエージェントの選択がその時点での最良のものに偏ってしまうため、経路の分散を許容する行動選択の導入を行うことが考えられる。

謝辞 本研究は文部省科学研究費補助金(課題番号: 22500120)の助成を受けたものである。

参考文献

- [1] Edsger, W.D.: A note on two problems in connexion with graphs, *Numerische Mathematik*, Vol.1, pp.269-271 (1959).
- [2] Peter, E.H., Nils, J.N. and Bertram, R.: A Formal Basis for the Heuristic Determination of Minimum Cost Paths, *IEEE Trans. Systems Science and Cybernetics*, Vol.4,

- No.2, pp.100-107 (1968).
- [3] 社団法人交通工学研究会 (編著): やさしい交通シミュレーション, 交通工学研究会 (2000).
 - [4] Justin, A.B. and Michael, L.L.: Packetrouting in dynamically changing networks: A reinforcement learning approach, *Advances in Neural Information Processing Systems*, Vol.6, pp.671-678 (1994).
 - [5] Samuel, P.M.C. and Dit-Yan, Y.: Predictive Q-routing: A memory-based reinforcement learning approach to adaptive trac control, *Advances in Neural Information Processing Systems*, Vol.8, pp.945-951 (1996).
 - [6] Said, H., Abdelhamid, M. and Yacine, A.: K-Shortest Paths Q-Routing: A New QoS Routing Algorithm in Telecommunication Networks, *Lecture Notes in Computer Science*, Vol.3421, pp.164-172 (2005).
 - [7] Balmer, M., Meister, K., Rieser, M., Nagel, K. and Axhausen, K.W.: Agent-based simulation of travel demand: Structure and computational performance of MATSim-T, *Proc. 2nd TRB Conference on Innovations in Travel Modeling* (2008).
 - [8] 立本真治, 本多中二: 微視的道路交通シミュレータ MITRAM への経路選択モデルの導入と検証, *日本知能情報ファジィ学会誌*, Vol.18, No.4, pp.586-597 (2006).
 - [9] Shinobu, Y.: MATES: Multi-Agent Based Traffic and Environment Simulator-Theory, Implementation and Practical Application, *Computer Modeling in Engineering and Sciences*, Vol.11, No.1, pp.17-25 (2006).
 - [10] Hideki, F., Takafumi, S. and Shinobu, Y.: Virtual Social Experiment of Tram Railway Extension Using Multi-Agent-Based Traffic Simulator, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol.15, No.2, pp.226-232 (2011).
 - [11] 吉村 忍, 西川紘史, 守安 智: 知的マルチエージェント交通流シミュレータ MATES の開発, *日本シミュレーション学会論文誌*, Vol.23, No.3, pp.228-237 (2004).
 - [12] 藤井秀樹, 仲間 豊, 吉村 忍: 知的マルチエージェント交通流シミュレータ MATES の開発: 第二報: 歩行者エージェントの実装と歩車相互作用の理論・実測値との比較, *日本シミュレーション学会論文誌*, Vol.25, No.4, pp.274-280 (2006).
 - [13] NPO 法人公共の交通ラクダ (RACDA): RACDA かわら版第 74 号, 入手先 (http://image02.wiki.livedoor.jp/r/a/racda_okayama/b4db0ada0e026b9e.pdf) (参照 2012-01-26).
 - [14] 国土交通省, 富山市: 富山港線 LRT 化の整備効果調査結果, 富山市 (オンライン), 入手先 (<http://www.city.toyama.toyama.jp/toshiseibibu/kotsuseisakuka/seibikokachosakekka.html>) (参照 2012-01-26).
 - [15] 国土交通省: 平成 17 年度全国道路・街路交通情勢調査, 入手先 (<http://www.mlit.go.jp/road/ir/ir-data/data/203.xls>) (参照 2012-04-30).



内田 英明

平成 22 年千葉大学工学部都市環境システム学科卒業。平成 24 年東京大学工学系研究科システム創成学専攻修士課程修了。修士 (工学)。交通工学研究会会員。



藤井 秀樹 (正会員)

平成 15 年東京大学工学部システム創成学科卒業。平成 21 年同大学大学院新領域創成科学研究科人間環境学専攻博士後期課程修了。同年同大学人工物工学研究センター特任助教。平成 22 年より同大学大学院工学系研究科システム創成学専攻助教。博士 (環境学)。複雑システムと交通流シミュレーションに関する研究に従事。日本機械学会, 日本シミュレーション学会, 人工知能学会, 交通工学研究会各会員。



吉村 忍

昭和 62 年東京大学大学院工学系研究科原子力工学専攻博士課程修了。工学博士。同大学講師, 助教授を経て, 平成 17 年同大学大学院工学系研究科システム量子工学専攻 (現, システム創成学専攻) 教授。知的シミュレーションの研究開発と社会・環境分野, 工学分野への応用に関する研究・教育に従事。日本計算力学連合会長, 日本機械学会計算力学部門長, 国際計算力学連合評議員, アジア太平洋計算力学連合評議員, 日本機械学会フェロー, 日本シミュレーション学会等会員。



荒井 幸代

慶應義塾大学理工学部卒業後, ソニー (株) を経て, 平成 10 年東京工業大学大学院総合理工学研究科知能科学専攻博士課程修了。Carnegie Mellon University (米) 上級研究員, 平成 14~16 年京都大学大学院情報学研究科客員助教授, 平成 17 年より千葉大学大学院工学研究科准教授。マルチエージェント強化学習, 多目的最適化の研究に従事。電気学会, 計測自動制御学会, 人工知能学会, 電子情報通信学会, AAAI, ACM 各会員。