経営課題にAIを!ビジネスインフォマティクス研究会 (第18回)
JSAI Special Interest Group on Business Informatics (SIG-BI #18)

宅配中の交通安全を考慮した タスク割当手法・インセンティブ設計

Designing task allocation and incentives for safe delivery to prevent traffic accidents

西野　貴志[1*]　章　進[1]　山成　侑香[1]　宮尾　勝[1]　劉　鵬達[1]　林　久志[1*]

Takashi Nishino[1], Sin Syo[1], Yuka Yamanari[1], Masaru Miyao[1], Pengda Liu[1], and Hisashi Hayashi[1]

[1]東京都立産業技術大学院大学　産業技術研究科

[1]Advanced Institute of Industrial Technology School of Industrial Technology

In online food delivery, which is expanding all over the world, the increasing number of traffic accidents during delivery is becoming problematic. To ensure the safety of delivery workers and the residents, it is necessary to understand the incentives of workers for behavior choices. While many existing sharing platforms are pulling workers into the online labor process by such incentives as on-peak/off-peak surcharges, workers try to accomplish more than their goals within a limited time. In this paper, to prevent over-speeding during delivery, we compare and evaluate some task assignment methods and incentive schemes by multi-agent simulation. We model worker's rational choices of behaviors based on reinforcement learning considering the profit and over-speeding of delivery workers.

1. はじめに

世界の交通事故における死者数のおよそ 4 分の 1 は自転車と歩行者で占められ日本では 2011 年から自転車や歩行者の死亡・重症事故の予防を目的として生活道路区域での速度規制が進められている[1].

一方 2000 年代から世界中で拡大している, オンラインフードデリバリーでは，簡単な手段で募集されたワーカーによる配達中の交通事故が社会問題化しつつある[2].

このため, 拡大が続くフードデリバリーにおける, ワーカーの行動選択に関するさまざまなインセンティブスキームを理解し，ワーカーや地域住民の安全を確保する必要がある.

オンラインフードデリバリーのサービス事業者は, ワーカーの配達中の交通ルール違反に対し，サービス利用停止によるペナルティを与えることや，格付け等の社会的な評価によって速度超過を抑制している.

本論文では，罰金によるペナルティと注文減少による機会損失とを組み合わせることで，ワーカーの配達中の速度超過を抑制することを目的とする.

状況に応じたワーカーの行動選択と，稼働時間あたりの収入・移動中の速度超過率を評価可能な，シミュレーションを構築し，タスク割当手法とインセンティブスキームを検討する.

強化学習によりワーカーエージェントの合理的な行動選択を推論し，シミュレーションの結果を学習データとしてより良い行動選択を学習させる.

本論文は，以下のように構成される．2 章で関連研究について議論する．3 章で提案の概要と目的を述べる． 4 章でシミュレーションモデル, 5 章で強化学習について, 6 章で提案手法を説明する．7 章で評価結果を示し，8 章で本論文をまとめる.

2. 関連研究と課題

配達を行うワーカーの主な収入は，店舗まで移動して商品を受け取り依頼主へ届ける基本収入と，移動距離に応じた収入がある.

サービス事業者は依頼主との仲介料と配達管理サービスの利用料を，手数料としてワーカーから受け

*連絡先：産業技術大学院大学産業技術研究科
〒140 - 0011 東京都品川区東大井1-10-40
E-mail: {b1931tn, hayashi-hisashi}@aiit.ac.jp

取る．

ワーカーは，配達者コミュニティ内での行動規範を順守することでサービスを利用できるが，遅配による低評価が続くとサービスを利用できなくなる可能性があるため，配達時間に関する厳しい要求による速度違反の要因にもなっている[2].

ワーカーの労働時間と行動選択には，配達の依頼を受けてから完了するまでの基本収入が大きな影響を与えている [3].

既存の多くのシェアリングプラットフォームでは，オンピーク/オフピーク時の割り増し料金や，格付けの高い人への追加インセンティブ等によって，より強力にワーカーをオンライン労働プロセスに引き入れ，ワーカーは目標以上のタスクを達成しようとする[4].

本論文では，速度超過に対し罰金によるペナルティを与える手法と，注文数の減少による機会損失のペナルティを与える手法を検討し，マルチエージェントシミュレーションによる評価を行う．

3. 提案手法と目的

本テーマでは，ワーカーの配達中の速度超過を抑制するためのタスク割当手法とインセンティブスキームを検討する．

また，エージェント・ベース・モデルで表現されたワーカーの一連の行動をシミュレーションすることにより，稼働時間あたりの収入と移動中の速度超過率を評価する．

シミュレーションにおける状況に応じたワーカーエージェントの合理的な行動選択は，強化学習で求める．

図 1 に，タスク割当とインセンティブ設計の概要を示す．速度超過などワーカーの行動履歴と個性に応じた配達タスク依頼を調整し①，これを受託したワーカーが配達時の行動方針（消極的：平均速度以下で移動 / 規制遵守:規制速度の範囲内で移動 / 規制無視：規制速度を超過して移動）に従って配達を行い②，速度超過があった場合は罰金を課す③.

このような手法により，速度超過を抑制することを検討する．

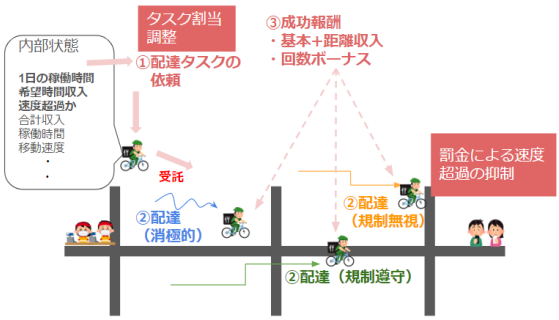

図 1：タスク割当とインセンティブ設計の概要

4. シミュレーションの概要

本稿では，プログラム言語 python によって，提案手法の検証ロジックを実装する．シミュレーション上でワーカーの個性と配達中の行動パターンを表現し，配達タスク中の行動選択を観察する．

4.1 配達エリア

配達エリアは 4×4 の座標上に位置するとし，隣接エリアとの距離が常に 1km とする．

配達の開始位置から配達先への移動について，配達収入と速度超過の有無を計算する．移動距離は開始座標から配達先座標までのユークリッド距離で計算し，移動経路探索は行わない．移動速度については 6.1 で述べる．

シミュレーションでの時間経過は 10 分を 1tick とし，ワーカーの行動による時間の経過と配達サービスが提供中かを計算する．

4.2 配達タスク

図 2 に，配達タスクについてのシナリオと状態遷移・行動について示す．1 日が始まるとワーカーは「依頼待ち（依頼なし）」の状態となり，待機を実行すると一定確率で「依頼待ち（依頼あり）」の状態となる．配達の依頼は必ず実行され，「配達」状態となる．配達を実行すると「依頼待ち（依頼なし）」状態に戻る．1 日のサービス提供時間が終了するか，ワーカーの稼働時間が終わると 1 日が終了する，

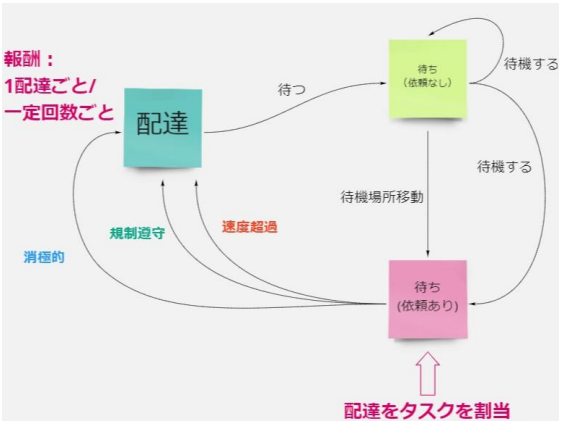

図 2：ワーカーの状態遷移

5. 強化学習による行動選択

ワーカーのそれぞれの状態での行動選択確率を求めるために，現実の行動データを用いることなく，シミュレーションの結果を学習データとして用い，より良い行動選択を学習させる．

5.1 Q 学習を用いたワーカーの行動選択

ワーカーのすべての状態における最適な方策（Q 値）を知るため，状態を入力(Input: state)としてアクション(output: action)を出力するニューラルネットワーク（Q-Network）を作成して Q 値を近似し，学習を行う．

ニューラルネットワークは，入力層を全結合/128 ノード，中間層を全結合/128 ノードを 2 層，出力層を活性化関数に恒等関数を使用する全結合/1 ノードとする．

5.2 学習データの収集

1 日間のワーカーエージェントの行動を 1 回のシミュレーションとし，記録された状態/行動/報酬/次の状態を Experience Replay [5]によって集中学習させる．

シミュレーションは 5000 回実行され，シミュレーション 1 回毎に 1 回の学習を行う．学習に利用する行動記録のバッチサイズは 32 とする． ε -greedy アルゴリズムにより，epsilon を初期値 1 から 0.01 まで減衰率 0.9995 で漸減させながら，学習した方策を利用する[5].

行動選択のためには Q-Network を使って推論するが，学習時には Q-Network と同じネットワーク構造の Target Network[5]を用いて学習を安定化させている．

Q-Network の学習時に損失関数である TD 誤差を計算する際，現在の状態の Q 値を求めるために Q-Network を用いるが，次の状態の Q 値を求める際には Target Network を用いる．Target Network の重みは古い Q-Network の値を用い，定期的に（100 回学習するごとに）最新の Q-Network と一致させる．

6. タスク割当・インセンティブ設計

6.1 エージェント設計

シミュレーション内で，ワーカーエージェントはそれぞれの状態において，内部状態を考慮して行動を選択する．内部状態のうち， 1 日の稼働時間と配達時間あたりどれくらいの収入を得たいか（希望時間収入）は，ワーカーのタイプごとに固定されている．

表 1 にワーカーエージェントの内部状態，表 2 にワーカーエージェントのタイプを示す．

表 1 ワーカーの内部状態

	範囲	変化
稼働時間 (tick)	24～72	行動により増加
合計収入	0～	配達に応じて増加
移動速度 (km/h)	12，18，28	配達ごとに変化
速度超過	なし，あり	配達ごとに変化
1 日の稼働時間 (h)	4，8，12	なし
希望時間収入	1000, 1500, 2000	なし

表 2 ワーカーエージェントのタイプ

		1 日の稼働時間 (h)		
		4	8	12
希望時間収入	1000	タイプ 1	タイプ 2	タイプ 3
	1500	タイプ 4	タイプ 5	タイプ 6
	2000	タイプ 7	タイプ 8	タイプ 9

6.2 タスク割当調整

ワーカーエージェントが「依頼待ち（依頼なし）」の状態で待機を実行すると，確率 60%で配達タスクが割当てられる．配達タスクの割当ては，ワーカーエージェントのタイプ 1～9 の順番で 1 回ずつ行われる．

提案手法「タスク割当調整」では，ワーカーの配達中に速度超過が観察された場合，次回の「依頼待ち（依頼なし）」では配達タスクが割当てられず他の

タイプのワーカーへ配達タスクが割当てられる．

この結果 15 回配達毎の報酬 1 が得られ難くなるため，間接的にワーカーエージェントの速度超過が抑制される．

6.3 報酬とインセンティブ設計

ワーカーエージェントは配達依頼があると 3 つの移動行動方針（消極的・規制遵守・速度超過）に従って配達を行う．その結果に応じた報酬を表 3 に示す．

配達により基本収入 390 と 1km あたり 60 の収入を金銭収入として得る．配達のたびに移動時間と配達の金銭収入が計算され，時間あたりの収入がワーカーの希望額以上の場合は報酬 0.2 が与えられる．さらに 1 日に 15 回数以上の配達を行うと追加の報酬 1 が与えられる．

また，配達中に「速度超過」での移動を選択した場合確率 30%，それ以外の移動を選択した場合は確率 10%で体力切れによる負の報酬-0.2 が得られる．

提案手法である，速度超過が観察されると「罰金」を課す手法では，配達の金銭収入から罰金 125 または 390 が差引かれる．この結果 1 回の配達毎の報酬 0.2 が得られ難くなるため，間接的にワーカーエージェントの速度超過が抑制される．

表 3 ワーカーへの報酬設計

	報酬	行動への影響
1 回の配達毎	0.2（希望時間収入を考慮）	直接的
15 回配達毎	1	直接的
速度超過あり（罰金 125/390）	0	間接的（時間収入が減少）
速度超過あり（依頼減少）	0	間接的（配達回数減少による機会損失）
体力切れ	-0.2	直接的

7. シミュレーションの実行結果

図 3 に，シミュレーションによる提案手法の効果を示す．4 本のグラフのうち，一番左（1:改善策なし）が改善策の適用前，中央の 2 つ（2:罰金（小），3:罰金（大））が配達中の速度超過に対し罰金を課した結果を，右端（3:タスク割当調整）がワーカーの働き方のタイプによって依頼数を変化させた結果を示している．

配達中の速度超過に対し少額（125）の罰金を課した場合では速度超過が 5.0%改善されているが，高額（390）の罰金を課すと速度超過が 15.8%改善される一方で収入も大きく減っている．

ワーカーの働き方のタイプによって依頼数を変化させたグラフでは，速度超過が 5.5%改善するという結果になった．

タスク割当調整を行った場合よりも，少額の罰金を課した場合の方が平均配達収入の減少が低く速度超過率も同等に抑えられており，バランス良く改善されている．

図 3：シミュレーション結果

8. おわりに

本稿では，オンラインフードデリバリーでの交通課題に対して，ワーカーの行動を外部からの要因によって変化させることで，事故を回避する方法を提示するものである．

配達中の速度超過に対し「罰金」「タスク割当調整」の提案手法を適用した結果，速度超過が 5.0%～15.8%減少した．少額の罰金を課した場合，速度超過が抑えられ平均配達収入減も少なく済む結果となった．

以上の結果から，強化学習によりワーカーエージェントの合理的な行動選択を推論し，配達収入・移動中の速度超過率を観察可能なシミュレーションを構築することで，「罰金」「タスク割当調整」など速度超過を抑制する手法を評価することが可能になった．今後さらに，タスク割当・インセンティブ設計を改良していく．

現在の課題として，ワーカーが 1 日の稼働時間と配達時間あたりどれくらいの収入を得たいか（希望時間収入）が事前に分かっているとしている点がある．ワーカーの行動を観察し，ワーカーのタイプを推測する方法を検討する必要がある．

また，現状のワーカーへの報酬設計では，速度超過に対して直接的に負の報酬は与えず，金銭収入の

減少や配達タスクの減少によって報酬を得難くする，間接的な報酬の与え方になっている．

直接的に負の報酬を与えると，ワーカーが常に消極的な移動行動をとってしまうためそのような報酬設計となっているが，今後は直接的な報酬を与えてもワーカーがより多くの報酬を獲得しつつ，負の報酬を避けるような行動選択を模擬できる，エージェントモデルと強化学習のモデルを検討する．

参考文献

[1] Haruhiko Inada, Jun Tomio, Shinji Nakahara and Masao Ichikawa: “Area-Wide Traffic-Calming Zone 30 Policy of Japan and Incidence of Road Traffic Injuries Among Cyclists and Pedestrians,” American Journal of Public Health, Vol. 110, No. 2, pp. 237-243, (2020)

[2] Mayila Maimaiti, Xueyin Zhao, Menghan Jia, Yuan Ru and Shankuan Zhu: “How We Eat Determines What We Become: Opportunities and Challenges Brought by Food Delivery Industry in a Changing World in China,” European Journal of Clinical Nutrition, Vol. 72, pp. 237-243, (2020)

[3] Yuqian Xu, Baile Lu, Anindya Ghose, Hongyan Dai and Weihua Zhou: “How Do Ratings and Penalties Moderate Earnings on Crowdsourced Delivery Platforms?,” NYU Stern School of Business, (2020)
https://ssrn.com/abstract=3609132 or
http://dx.doi.org/10.2139/ssrn.3609132（visited in 2020）

[4] Qingjun Wu and Zhen Li, “Labor Control and Task Autonomy under the Sharing Economy: a Mixed-Method Study of Drivers’ Work,” The Journal of Chinese Sociology, Vol. 6, Article Number 14, (2019)
https://doi.org/10.1186/s40711-019-0098-9 （ visited in 2020）

[5] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg & Demis Hassabis: “Human-level control through deep reinforcement learning” Nature, Vol. 518, pp. 529–533, (2015)
http://dx.doi.org/10.1038/nature14236（visited in 2021）