Transportation Research Part F 70 (2020) 37–57

Contents lists available at ScienceDirect

Transportation Research Part F

journal homepage: www.elsevier.com/locate/trf

Modeling pedestrian-cyclist interactions in shared space using inverse reinforcement learning

Rushdi Alsaleh *, Tarek Sayed

Department of Civil Engineering, University of British Columbia, 6250 Applied Science Lane, Vancouver, BC V6T 1Z4, Canada

ARTICLE INFO

Article history:
Received 20 September 2019
Received in revised form 8 February 2020
Accepted 10 February 2020
Available online 28 February 2020

Keywords:
Shared space modeling
Overtaking behavior
Following behavior
Simulation
Cyclist and pedestrian
Reward function

ABSTRACT

The objective of this study is to model the microscopic behaviour of mixed traffic (cyclist-pedestrian) interactions in non-motorized shared spaces. Video data were collected at two locations of Robson Square non-motorized shared space in downtown Vancouver, British Columbia. Trajectories of cyclists and pedestrians involved in interactions were extracted using computer vision algorithms. The extracted trajectories were used to obtain several variables that describe elements of road users' behaviour including longitudinal and lateral distances, speed and speed differences, interaction angle, and cyclist acceleration and yaw rate. The road users behaviour was modeled as utility-based intelligent rational agents using the finite-state Markov Decision Process (MDP) framework with unknown reward functions. The study implemented Inverse Reinforcement Learning (IRL) using two algorithms: the Maximum Entropy (ME) algorithm, and the Feature Matching (FM) algorithm to recover/estimate the reward function weights of cyclists in two types of interactions with pedestrians: following and overtaking interactions. Reward function weights infer cyclist preferences during their interactions with pedestrians in non-motorized shared spaces, and can form the key component in developing agent based microsimulation model for road users. Furthermore, the estimated reward functions were used to estimate cyclists' optimal policy for such interactions. A simulation platform was developed using the estimated reward functions and the cyclist optimal policies to simulate cyclist trajectories for the validation dataset. Results show that the Maximum Entropy (ME) IRL algorithm outperformed the Feature Matching (FM) IRL algorithm, and generally provided reasonable results for modeling such interactions in non-motorized shared spaces, considering the high degrees of freedom in movement and the more-complex road users interactions in such facilities. This research is considered an important step toward developing a full Agent-Based Model (ABM) for road users in shared space facilities to evaluate the safety and efficiency of such facilities.

1. Introduction

Many cities have been adopting policies that aim at promoting active modes of transportation such as walking and cycling. Encouraging active transportation supports the cities' sustainability goals, reduces traffic-induced air pollution, and helps road users to adopt a healthy lifestyle and increase their level of physical activity. Such policies may involve

* Corresponding author.
E-mail addresses: Rushdi.alsaleh@ubc.ca (R. Alsaleh), tsayed@civil.ubc.ca (T. Sayed).

https://doi.org/10.1016/j.trf.2020.02.007
1369-8478/

the redesign of selected streets or public places into non-motorized shared spaces for social and recreational activities. The shared space design paradigm has received considerable attention as an alternative approach of classic design of streets. Shared space programs have been implemented in many cities around the world including Vancouver, Calgary, Vienna, Auckland, and many other cities in Germany and the Netherlands. Furthermore, the "Open Streets" program represents another policy that encourages active road users of all ages and abilities to share the road. This program involves the temporarily closure of selected city streets for motorized traffic and creating non-motorized shared spaces for social and recreational activities. By 2016, around 122 cities in the United States have hosted "Open Street" programs, including New York and Los Angeles with more than 100 thousands participants per event (Hipp, Bird, van Bakergem, & Yarnall, 2017).

Shared spaces are areas with no clear segregation between the road users, meaning that the right of way is shared between them. Despite the emerging popularity and the wide implementation of shared spaces around the world, only few studies have analyzed and modeled the behaviour of the road users and the interactions among them in such spaces. Most shared spaces have already been implemented without prior evaluations of their efficiency or safety. This gives rise for the need to investigate optimal shared space designs prior to real-world implementations to evaluate their efficiency (e.g., average road user delay), capacity, and safety (e.g. near misses and collisions). The modeling and simulation of road users' behaviour and their interactions in shared spaces provide planners and engineering's with a powerful tool for evaluating their operation prior to real-world implementation. However, modeling road users' behaviour in shared spaces is challenging because of their complex interactions, which are difficult to be accurately described by the available software such as VISSIM (Gibb, 2015). Unlike conventional roads, the shared space concept provides the road users with the freedom to move in the whole area of the facility without being restricted to predefine paths (e.g. sidewalk or bike lanes). Thus, the road user behavior in shared spaces can differ significantly from their behaviour in conventional streets.

Previous works dealing with the modeling and simulation of shared spaces are mainly based on physical models, e.g., the social force model (SFM) (Helbing & Molnar, 1995), or the cellular automata (CA) model (Nagel & Schreckenberg, 1992), with extensions for modeling mixed traffic conditions. These models are limited as they do not consider the fact that road users can logically assess the surrounding environment and take rational decisions. For example, in the social force model, road users are modeled as particles, and their interactions are modeled using physical forces. In the cellular automata approach, road users move from one cell to another based on predefined rules that mainly depend on the probability of choosing the target cell. The heterogeneity of a road user system, e.g., pedestrians or cyclists, is hard to capture in the cellular automata models, as when such models are used, the movement of road users is limited to a fixed cell size in each time step.

This paper proposes a Markov Decision Process (MDP) to model cyclist-pedestrian interactions in non-motorized shared spaces. The MDP models the behaviour of the decision maker as a sequential decision process in which the decision and the consequent action depend on the current state of the decision maker and the optimal policy that aims at maximizing the utility or reward function of the decision maker. To solve a MDP and compute the optimal policy, e.g., predicted decision and action sequence, the reward function must first be specified. However, specifying the reward function is very challenging and requires more effort than learning the policy itself. One approach for solving this problem is using Inverse Reinforcement Learning (IRL) (Ng & Russell, 2000) to recover/estimate the reward function given expert demonstrations, e.g., real-world road user trajectories that describe road users' decision and behaviour. The contribution of this paper is the recovery and estimation of the reward (utility) function and optimal policy of road users' involved in cyclist-pedestrian interactions in non-motorized shared spaces. Two cyclist-pedestrian interactions are considered in the analysis; the following and overtaking interactions. Estimating the reward function is important for several applications such as the development of Agent-Based microsimulation models (ABM) of cyclists. Two IRL algorithms are used to recover/estimate the reward functions: (1) the Feature Matching (FM) algorithm (Abbeel & Ng, 2004) which assumes optimal road user behaviour or decision process, and (2) the Maximum Entropy (ME) algorithm (Ziebart, Maas, Bagnell, & Dey, 2008) which assumes sub-optimal behaviour. Moreover, this paper gives insights about the preferences and behaviours of road users during their interactions in non-motorized shared spaces.

2. Previous work

Shared space is an emerging urban design approach that reduces the segregation between road users and supports pedestrian and cyclist movements with slower vehicles. These schemes of street design encourage the integration of road users (pedestrian-friendly environment) by reducing the segregation between road users and decreasing the dominance of motorized vehicles (Kaparias, 2012). Several previous studies investigated the benefits of shared spaces and the behaviour of road users in shared spaces. Studies showed documented benefits of shared spaces, including the increase in road users safety and pedestrian activity levels (Swinburne, 2005). The conversion of a large five-way intersection in Oosterwolde, Netherlands to a paved shared space area for all users resulted in a reduction in traffic speed and severe collisions at the shared space area despite the increase in traffic volume (Hamilton-Baillie, 2008). The conversation of a complex roundabout in Austria into a shared space area have led to a narrower speed distribution for all road users, which has been explained by the smoother movements and less stop and go conditions in the shared space area (Schonauer, Stubenschrott, Schrom-Feiertag, & Mensik, 2012). Previous studies classified the interactions between cyclists and pedestrians in shared space based on conflicting angle between road users into three type of interactions same direction interaction, e.g., angle difference $0^\circ \pm 30^\circ$, opposite direction interaction, e.g., angle difference $180^\circ \pm 30^\circ$, and crossing interaction which include

the remaining cyclist-pedestrian interactions (Alsaleh, Hussein, & Sayed, 2020; Beitel, Stipancic, Manaugh, & Miranda-Moreno, 2018).

Limited studies have investigated the development of microsimulation models of road users' behavior and interactions at shared spaces. Most of the existing microsimulation models were developed to model a single mode of transportation, e.g., vehicular traffic (Gipps, 1981; Wiedemann, 1974), pedestrian flow (Burstedde, Klauck, Schadschneider, & Zittartz, 2001; Helbing & Molnar, 1995) or cyclist flow (Jiang, Jia, & Wu, 2004; Liang, Baohua, & Qi, 2012). Nevertheless, microsimulation models of pedestrians and cyclists are less developed compared to the vehicular traffic. Microsimulation models of pedestrians or cyclists are mainly based on cellular automata models or physical analog models such as social force model. Some studies extend to use these models to develop microsimulation models for mix traffic conditions. For example, Luo et al. (Luo, 2015) proposed a modified cellular automata model to simulate bicycles and cars in heterogeneous traffic urban road. The model discretized the environment into cells and used an occupancy rule that depends on cars speed to consider the variable lateral safety distance of mixed vehicular traffic.

Anvari, Bell, Sivakumar, and Ochieng (2015) proposed a modified social force model and rule-based constraint model to simulate heterogonous traffic of pedestrians and vehicles in shared spaces. The social force model was used to assign social/physical force for road users in order to reproduce their interaction and negotiation, e.g., moving toward a target, etc. Schönauer (2012) used the social force model to model the interactions between pedestrian and vehicles in shared spaces. The model used a discrete single-track approach to model vehicle dynamics, and game theoretic approach for resolving conflicts between pedestrians and vehicles. Huang (2016) used a fuzzy logic and a modified social force models to model cyclist interactions at unsignalized intersection with heterogeneous traffic. Dias, Iryo-Asano, Nishiuchi, and Todoroki (2018) used a social force model to model the interaction between segway and pedestrian mixed traffic at shared sidewalk assuming the segway is most similar to pedestrian

The Agent-Based Modeling (ABM) approach is an appealing and powerful approach for realistic modeling of road users' behaviour and their complex interactions in shared spaces. The ABM approach accounts for road users intelligent and their ability to take logical decisions based on their experience and surrounding environments. This approach requires modeling of agents' objectives or goals (Jennings, 2000). One approach of modeling agents' goals or strategies is a rule-based model (Hussein & Sayed, 2017), however, in in this approach the rules that govern road user interactions are mostly extracted heuristically or are based on ad-hoc rules (Papadimitriou, Yannis, & Golias, 2009). Other approaches include modeling intelligent agents that can learn from experience of interactions with other agents, e.g., experts' demonstration (Plekhanova, 2002).

Under the Markov Decision Process (MDP) modeling framework of road user interactions in shared spaces, recovering the reward function is challenging. Learning from the demonstration of the task is easier. One approach to solve this problem without the need to extract the reward function is the behavioural cloning as an approach of imitation learning, where the main aim is to mimic the action of the expert. However, the main shortcomings of this approach is that (1) if the demonstrated agent is not the same as the agent trying to perform the task, the goal now becomes to achieve the outcome the expert achieved instead of mimic the same actions; (2) behavioural cloning provides no reasoning about the outcomes (how the agent achieved the goal at the end of the task); (3) the expert may have different degrees of freedom on how to accomplish that task (Bratko, Urbancic, & Sammut, 1995). Inverse Reinforcement Learning (IRL) that has been developed by (Ng & Russell, 2000) provides a tool of reasoning what the expert (i.e. road user) is trying to achieve.

3. Methodology

The expert data used as a demonstration in the training and testing of the Inverse Reinforcement Learning (IRL) algorithms were cyclist and pedestrian trajectories. Video data were collected at two locations of shared spaces. These locations are on the busy non-motorized shared space of Robson Square in downtown Vancouver, British Columbia. Computer Vision (CV) algorithms were used to track road users, e.g., pedestrians and cyclists, in the shared space (Saunier & Sayed, 2006). The extracted road user trajectories were used to compute several variables that describe road user behaviour, e.g., speed and acceleration profiles, longitudinal and lateral distances, and yaw angles. The following sub-sections summarize the details of each of the following tasks: data collection, road user tracking, extraction of spatial, speed, acceleration and yaw rate profiles, and interaction modeling using Inverse Reinforcement Learning (IRL).

3.1. Data collection

Video data were collected at two locations of a busy non-motorized shared space, located in Robson Square in downtown Vancouver, British Columbia (Fig. 1). The city of Vancouver considered the permanent closure of Robson Street, between Hornby Street in the west and Howe Street in the east, for motorized traffic in order to provide a comfortable and safe plaza for vulnerable road users in downtown Vancouver. The area is an active environment for walking and cycling and it is a commercial core and a place for many recreational facilities, including the Robson Square ice rink, Vancouver Art Gallery, and Vancouver Supreme Court. Cyclist-pedestrian interactions were frequently observed in the Robson shared space. Video data were obtained for the two locations from two cameras mounted on the edges of the Robson shared space area. The first camera was mounted at the shared space area near Howe Street and the video data were obtained for 21 h over five days in May

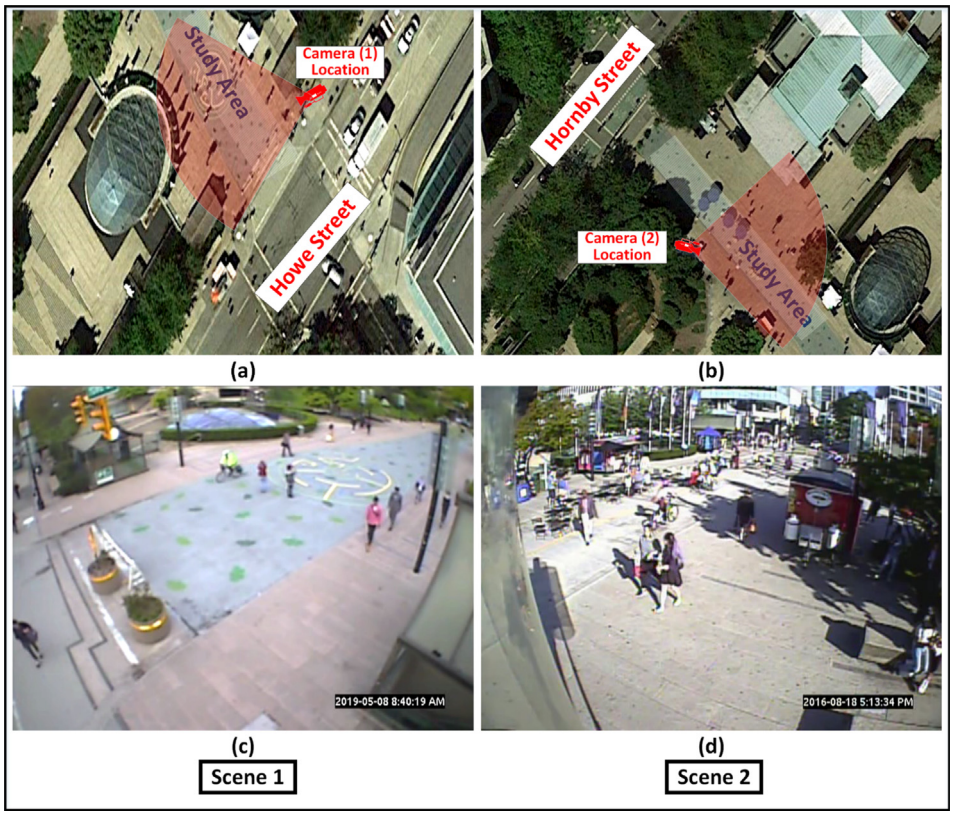


Fig. 1. Study locations (a) world image for the first study location (scene 1); (b) world image for the second study location (scene 2); (c) camera view for scene 1; (d) camera image for scene 2.

2019 (Fig. 1a and c). The second camera was mounted at the shared space area near Hornby Street and the video data were obtained for 18 h over nine days in August and September 2016 (Fig. 1b and d).

3.1.1. Data for training and testing

In this study, cyclist and pedestrian trajectories that are involved in the following and overtaking cyclist-pedestrian interactions were extracted during the 39 h of the analyzed video data of the two locations at the Robson shared space. A total number of 228 and 276 cyclist and pedestrian trajectories that are involved in following and overtaking cyclist-pedestrian interactions were extracted. Trajectories were extracted each time frame (1/30 s) and were associated with 18,376 and 28,068 data points for the following and overtaking interactions, respectively. The extracted data points of each interaction were divided into two sets; the training dataset which consists of around 80% of the data, while the remaining dataset, e.g., 20% of the data, was used as a testing dataset, similar to previous studies (Kuhn & Johnson, 2013; May, Maier, & Dandy, 2010). For the following interactions, the training dataset consists of 14,698 data points that are associated with 170 trajectories, while the testing dataset consists of 3678 data points that are associated with 58 trajectories. For the overtaking interactions, the training dataset consists of 22,636 data points that are associated with 222 trajectories, while the testing dataset consists of 5434 data points that are associated with 56 trajectories.

3.2. Road user tracking

The automated extraction of road user trajectories from the video footage were conducted using a video analysis system that has been developed at the University of British Columbia (Saunier & Sayed, 2006). As shown in Fig. 2, the procedure starts with the camera calibration process, in which a homography matrix is generated to create a mapping between the two-dimensional video image coordinates and the real-world three-dimensional coordinates, as described in details in Ismail, Sayed, and Saunier (2013). This enables transferring of the spatial and temporal information of the tracked trajectories to the actual coordinate system of the location being analyzed. In the next step, the feature tracking, computer-vision algorithms are used to detect road users in the traffic scenes. The algorithm detects distinct points (features) on moving objects in the video scene and differentiate between features that belong to road users (e.g. pedestrians, cyclists) and that are part of the environment. Features are identified and tracked using the implementation of the Kanade–Lucas–Tomasi

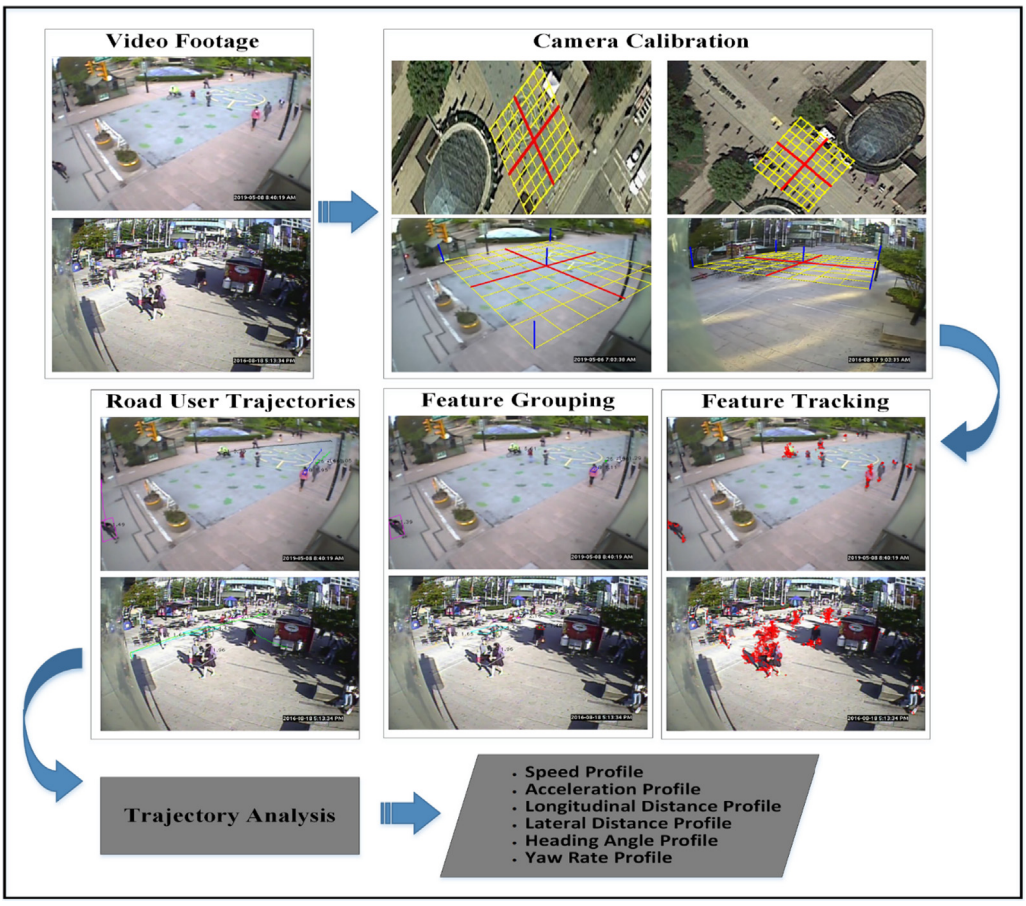


Fig. 2. Trajectory extraction process of road users and variable profiles extraction.

(KLT) feature-tracking algorithm (Lucas & Kanade, 1981; Tomasi & Kanade, 1991). Then, features are clustered in the feature grouping process to determine which group of features belongs to the same road user. The clustering algorithm uses different cues to cluster the tracked features, including their spatial proximity and movement pattern similarities. The grouped objects are then tracked on each video frame creating road users trajectories. The whole analysis procedures are illustrated in Fig. 2.

3.3. Variables extraction: Extraction of spatial, speed, acceleration and yaw rate profiles

The main set of variables that are used to describe the behaviour of cyclist and pedestrian interactions in shared space are based on previous cycling behaviour research (Gavriilidou, Daamen, Yuan, & Hoogendoorn, 2019; Ma & Luo, 2016), pedestrian behaviour research (Hussein & Sayed, 2017; Teknomo, 2006; Wang, Lo, Liu, & Kuang, 2014; Alsaleh, Sayed, & Zaki, 2018), cyclist-pedestrian interaction behaviour (Alsaleh, et al., 2019; Beitel et al., 2018) and mix traffic research in shared space (Anvari et al., 2015; Dias et al., 2018; Gorrini, Crociani, Vizzari, & Bandini, 2018; Luo, 2015). The previous studies used several variables to describe the behaviour of cyclist and pedestrian including longitudinal distance, lateral distance, road user speed, the speed difference between interacting road users, road user acceleration, interaction angle, and yaw rate or changes in steering angle.

Road user trajectories capture the movement of each pedestrian and cyclist in the form of a sequence of spatial coordinates and instantaneous speed at each video frame (1/30 s). A road user trajectory (T) is defined along the trajectory lifetime (n video frame) as a finite set of tuples, as shown in Eq. (1). The extracted road user trajectories and the derived variables were smoothed for noise using Savitzky-Golay filter (Savitzky & Golay, 1964).

$$T(t) = \{(X_1, Y_1, V_{X_1}, V_{Y_1}, \cdots, X_i, Y_i, V_{X_i}, V_{Y_i}, \cdots, X_n, Y_n, V_{X_n}, V_{Y_n})\} \quad (1)$$

where $i = \{1, \cdots, n\}$ is a discrete temporal index, $X_i$andY_i are the spatial coordinates of the road user at time frame (i), and V_{X_i}, V_{Y_i} are the corresponding velocities. A speed profile (S) for each road user is defined along the trajectory lifetime as$S(t) = norm(V_x, V_y)$, with V_x and V_y are the velocity vectors of length n, for the X and Y coordinates, respectively.

The distance vector $\left(\vec{\boldsymbol{d}}\right)$ is defined as the distance between the cyclist and pedestrian and directing toward the pedestrian, as shown in Fig. 3. The angle θ is defined regarding the cyclist considering the pedestrian as a neighbour user as the angle between the velocity vector of the cyclist $\left(\vec{\boldsymbol{v}}\right)$ and distance vector $\left(\vec{\boldsymbol{d}}\right)$, and can be computed using Eq. (2). The longitudinal distance is the distance between the cyclist and pedestrian along the direction of the cyclist movement. Equation (3) shows the calculation of the longitudinal distance $(d_{Longitudinal})$ between the cyclist and pedestrian. Initially, the longitudinal distance is positive as the cyclist is behind the pedestrian, and negative values of the longitudinal distances indicate that the cyclist becomes a head of the pedestrian. The lateral distance is the absolute distance between the cyclist and pedestrian perpendicular to the direction of the cyclist movement. Eq. (4) shows the calculation of the lateral distance $(d_{Lateral})$ between the cyclist and pedestrian.

$$\theta = \cos^{-1}\left(\frac{\vec{\boldsymbol{d}} \cdot \vec{\boldsymbol{v}}}{\|\vec{\boldsymbol{d}}\| \cdot \|\vec{\boldsymbol{v}}\|}\right) \tag{2}$$

$$d_{Longitudinal} = \|\vec{\boldsymbol{d}}\| \cdot \cos\theta \tag{3}$$

$$d_{Lateral} = |\|\vec{\boldsymbol{d}}\| \cdot \sin\theta| \tag{4}$$

The acceleration profile of the road users are derived from the speed profile after smoothen it. Eq. (5) shows the calculation of the acceleration (a) as the change of the smoothed instantaneous velocity (S) of the cyclist as follows:

$$a(t) = \frac{d(S)}{dt} \tag{5}$$

where t is time.

The cyclists usually perform several swerving maneuvers while cycling in shared space to overtake slower road users, e.g. pedestrians, or to a void collision with other road users. The yaw motion around the yaw axis describes the rotation of the cyclist that changes direction to the left or right of its direction of motion. The yaw rate of the cyclist is the angular velocity of this rotation or the rate of change of the heading angle. The yaw rate signal profile is useful in quantifying the swerving maneuvers of the cyclists. Eq. (6) shows the calculation of the yaw rate (r) as the change of the heading angle Ψ of the cyclist, as shown in Fig. 3 (Ayres, Wilson, & LeBlanc, 2004):

$$r(t) = \frac{d\Psi}{dt} \tag{6}$$

where t is time.

3.4. Inverse Reinforcement Learning (IRL)

In this study, the road user decision is modeled as a finite-state Markov Decision Process (MDP). A Markov Decision Process (MDP) consists of a tuple $(S, A, P^{a}_{ss'}, \mathcal{R}, \gamma, D)$, where S is a finite set of states; $A = \{a_1, \cdots, a_k\}$is a set of k actions; $P^{a}_{ss'}$ is a set of the state transition probabilities; $\mathcal{R} : S \rightarrow A$ is the reward function; $\gamma \in [0, 1)$ is a discount factor, which describe how

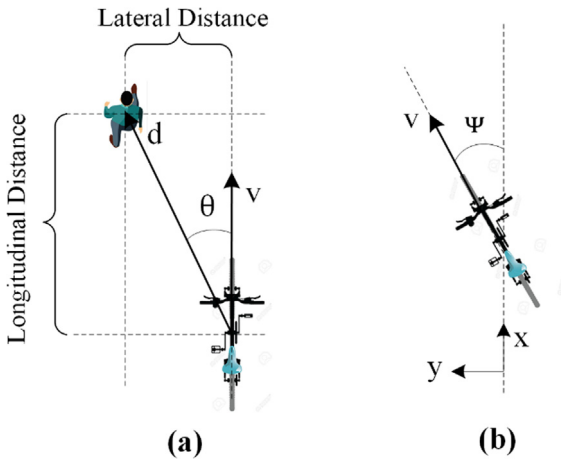


Fig. 3. Illustration of (a) longitudinal and lateral distance between cyclist and pedestrian; (b) Heading angle for a cyclist.

much a given reward is worth on step in the future compared with getting the same reward at current state; and D is the distribution of the initial state. In this study, a discount factor of $\gamma = 0.975$ is used assuming 10% effect of the reward at a state 3 sec later (90 time steps) from the current state. Let $MDP\{\mathcal{R}$ refers to a Markov Decision Process without a reward function.

In the forward Reinforcement Learning (RL), the reward function is known and is used to estimate the agent optimal policy, which maximizes the value of the agent action. However, the problem here is recovering the utility (reward) function that the agent was optimizing given some expert demonstrations (i.e. road user trajectories). Road users act to optimize their reward function and thus the problem is to find the reward weights θ that make their demonstrated behavior appear optimal or near optimal. The trajectories of the road users (expert demonstrations) ζ are assumed to represent the optimal or near-optimal behavior. A trajectory ζ of road user is a sequence of states and actions and is defined according to Eq. (7). Fig. 4 shows the structure of the IRL Problem applied in this paper.

$$\zeta = \{s_0, a_0, \cdots, s_i, a_i, \cdots, s_T\} \tag{7}$$

where s_i and a_i are the road user state and the observed action at a time step $i \in (0, T)$ of the trajectories lifetime (T frames).

It's assumed that there is a true reward function ($\mathcal{R}$) that linearly maps the features of each state f_s, to a state reward value which represent the utility of visiting that state. The reward function $\mathcal{R}_{\theta^*}(s) = \theta^T * f(s)$ is parameterized by the reward weights θ, which represent the weights on features over states $f(s)$. The reward is simply the sum of the state rewards (i.e. the reward weighs applied to the path features). A policy $\pi(\cdot; s)$ is a mapping from states to a probability distribution over the action space A. The value of the policy π (V^π) is the sum of the discounted reward and given by Eq. (8) (Abbeel & Ng, 2004).

$$E_{S_0\ D}[V^\pi(s_0)] = E\left[\sum_{t=0}^{\infty} \gamma^t \mathcal{R}(s_t)|\pi\right] \tag{8}$$

$$E_{S_0\ D}[V^\pi(s_0)] = \theta.E\left[\sum_{t=0}^{\infty} \gamma^t f(s_t)|\pi\right] \tag{9}$$

The expectation here is taken with a random state sequence draw by starting from state S_0 D, and picking and action according to the policy π. The feature expectations value vector $\mu(\pi)$, i.e., the expected discounted accumulated feature value vector, is defined by Eq. (10). Then the expectation of the value of the policy can be redefined as in Eq. (11) (Abbeel & Ng, 2004).

$$\mu(\pi) = \left[\sum_{t=0}^{T} \gamma^t f(s_t)|\pi\right] \tag{10}$$

$$E_{S_0\ D}[V^\pi(s_0)] = \theta.\mu(\pi) \tag{11}$$

The expert feature expectation $\mu_E = \mu(\pi_E)$ can be empirically estimated given a set of m trajectories $\{s_0^i, s_1^i, \cdots\}_{i=1}^m$ generated by the expert as presented in Equation (12) (Abbeel & Ng, 2004).

$$\widehat{\mu}_E = \frac{1}{m}\sum_{i=1}^{m}\sum_{t=0}^{T}\gamma^t f\left(s_t^{(i)}\right) \tag{12}$$

Two algorithms were used to recover the utility (reward) functions of cyclists during their interaction with pedestrians in shared spaces. The first algorithm is the Feature Matching (FM) IRL algorithm (Abbeel & Ng, 2004). This algorithm assumes a linear reward function that maps the feature in each state and optimal expert demonstrations. The algorithm estimates the reward function for which the feature expectation of the policy with respect to the reward function matches the feature counts of the expert trajectories. The second algorithm is the Maximum Entropy (ME) IRL algorithm (Ziebart et al., 2008). The algorithm assumes near-optimal behaviour of expert demonstrations to account for the inherent noise and imperfect trajectories. The algorithm uses a probabilistic approach based on the principle of maximum entropy to account for the noise and imperfect expert trajectories.

3.4.1. Feature Matching (FM) IRL algorithm

In this algorithm, the reward function is estimated for $MDP\{\mathcal{R}$ for which the feature expectation of the policy with respect to the linear reward function $\mathcal{R}_{\theta^*}(s) = \theta^T * f(s)$ matches the feature counts of the expert trajectories. The algorithm finds a policy whose performance is close to that of the expert's, on the unknown linear reward function $\mathcal{R}_{\theta^*}(s)$, i.e., $||\mu(\pi) - \mu_E||_2 < \varepsilon$. The steps of the algorithms to find the policy π is as follow (Abbeel & Ng, 2004):

1. Pick a random policy $\pi^{(0)}$, compute the feature expectation of the policy $\mu^{(0)} = \mu(\pi^{(0)})$, and set $i = 1$.
2. Compute $t^{(i)} = max_\theta min_{j\in\{0,\cdots,i-1\}}\theta^T(\mu_E - \mu^{(j)})$, then set $\theta^{(i)}$ the value of θ that attains this maximum.
3. Terminate if $t^{(i)} < \varepsilon$.
4. Compute the optimal policy $\mu^{(i)}$ for the MDP using rewards $\mathcal{R}_\theta(s) = \theta^T * f(s)$ through Reinforcement Learning (RL).
5. Compute$\mu^{(i)} = \mu(\pi^{(i)})$
6. Set $i = i + 1$, and go back to step 2.

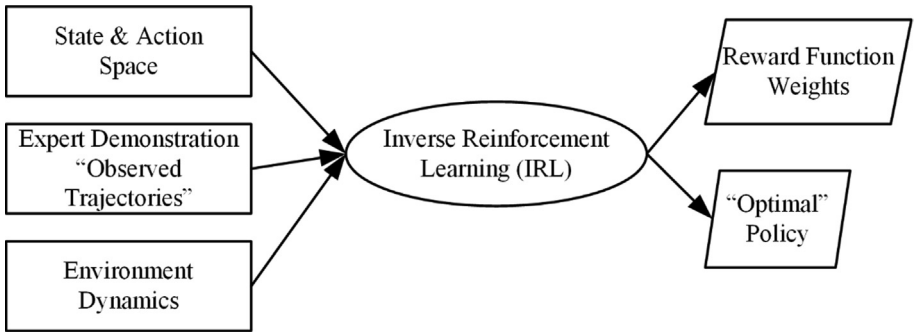


Fig. 4. Structure of the Inverse Reinforcement Learning (IRL).

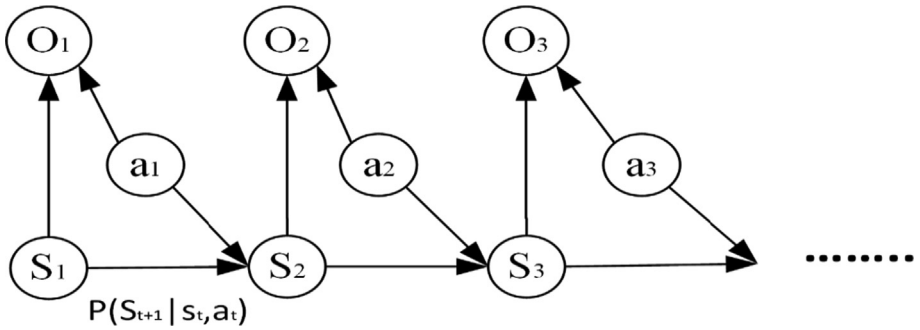


Fig. 5. Illustration of the optimality vector of the decision-making in MDP.

3.4.2. Maximum Entropy (ME) IRL algorithm

In the Maximum Entropy (ME) IRL approach, the problem of modeling road user behavior is formulated as a problem of recovering the reward function that makes the behavior induced by a "near-optimal" policy of road users closely mimic the demonstrated expert behavior. In the ME algorithm (Ziebart et al., 2008), the optimality of decisions is defined by a binary vector $O_{1:T}$ over a sequence of T decisions. Fig. 5 shows an illustration of the MDP process containing the optimality vector. The probability of optimality for each state and action is proportional to the reward associated with that state and action as given by Equation (13).

$$P(O_t|s_t,a_t) \propto e^{\mathcal{R}(s_t,a_t)} \tag{13}$$

Thus, the probability of a trajectory to be observed in the demonstration dataset given the optimality vector $P(\zeta|O_{1:T})$ is proportional to the probability of the trajectory to occurring times the exponential reward of that trajectory as presented in Equation (14). This means that trajectories of equal reward all have the same probability of being executed by the expert, and trajectories with lower reward are exponentially less likely, i.e., the expert can have some noise and not always outputting perfect optimal trajectories. Handling the uncertainty or noise in expert demonstrations in the ME algorithm can potentially lead to obtaining more robust and clean reward functions. Eq. (14) can be reformulated as given in Eq. (15).

$$P(\zeta|O_{1:T}) \propto P(\zeta)\prod_t e^{\mathcal{R}(s_t,a_t)} = P(\zeta)e^{\sum_t \mathcal{R}(s_t,a_t)} \tag{14}$$

$$P(\zeta|\theta,\tau) \approx \frac{e^{\theta^T * f(\zeta)}}{Z(\theta,\tau)} \prod_{s_{t+1},a_t,s_{t\in\zeta}} P_\tau(s_{t+1}|a_t,s_t) \tag{15}$$

where τ is the transition distribution, $Z(\theta,\tau)$ is the partition function and it is a normalization constant over all trajectories defined by Eq. (16).

$$Z(\theta,\tau) = \sum_{i=1}^{m} e^{\mathcal{R}(\zeta_i)} \tag{16}$$

where $i \in \{1,\cdots,m\}$ is a discrete index for the number of trajectories.

The distribution of the actions of each state over the paths provides a stochastic policy, where the probability of selecting an action is proportional to the sum of all probabilities of taking paths begin with that action as given by Eq. (17).

$$P(a|\theta,\tau) \propto \sum_{\zeta:a\in\zeta_t} P(\zeta|\theta,\tau) \tag{17}$$

In this algorithm, the objective of the expert instead of being maximizing the reward, is to maximize the difference between the expectation of the reward under the policy and the entropy of that policy. The entropy part in the objective function accounts for the uncertainty and noise in expert demonstrations. Estimating the reward function parameters θ is obtained by maximizing the likelihood of the expert demonstrations under the maximum entropy distribution as presented in Equations (18) and (19).

$$\theta^* = argmax_\theta(L(\theta)) = argmax_\theta \frac{1}{m}\sum_{i=1}^{m} logP(\zeta_i|\theta,\tau) \tag{18}$$

$$\theta^* = argmax_\theta \frac{1}{m}\sum_{i=1}^{m} \mathcal{R}(\zeta_i) - \log Z(\theta,\tau) \tag{19}$$

where $L(\cdot)$ is the likelihood function, and m is the number of observed trajectories.

In this study, the reward weights were estimated using a reward function in the linear form with intercept and taking the first level as a reference level of each feature as shown in Eq. (20).

$$\mathcal{R} = Intercept + \theta^T . f_{level} \tag{20}$$

where $\mathcal{R}$ is the reward function, *Intercept* is the intercept assuming the first level is the reference level for each feature, θ^T is the weight vector estimated for the features in relative for the reference level for each feature, and f_{level} is a dummy variable specifying the levels of each feature.

3.5. Evaluation metrics

The basis of the trajectory error computation is the distance between the predicted (simulated) trajectories and the true trajectories in the validation dataset. Two performance metrics are used to compare the simulated and the true trajectories as follow:

1. *Mean Absolute Error (MAE).* The MAE measures the average magnitude of the error in the simulated trajectory over the trajectories lifetime (n frames). Equation (23) shows the calculation of the MAE for a simulated trajectory.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |u_{sim,i} - u_{true,i}| \tag{23}$$

   where u_{sim} and u_{true} are the simulated and the true trajectories, respectively. The $i = \{1, \cdots, n\}$ is a discrete temporal index of the trajectory length.
2. *Hausdorff Distance (HausD).* The Hausdorff distance measures the degree of mismatch (deviation) between the simulated and the true trajectories. This measure computes the largest distance between the simulated and the true trajectories while ignoring the time step alignment. This measure relaxes the penalty for errors caused by the time step offset but emphasizes the overall parameter displacement between the simulated and the true trajectories. Eq. (24) shows the calculation of the Hausdorff distance between a finite point set of a simulated trajectory $A = \{a_1, \cdots, a_n\}$ and a true trajectory $B = \{b_1, \cdots, b_n\}$ (Huttenlocher, Klanderman, & Rucklidge, 1993; Rockafellar & Wets, 2009).

$$H(A,B) = \max\{h(A,B), h(B,A)\} \tag{24}$$

   where

$$h(A,B) = \ a \in Amax(\ b \in Bmin||a-b||) \tag{25}$$

$$h(B,A) = \ b \in Bmax(\ a \in Amin||b-a||) \tag{26}$$

   and $||\cdot||$ is the Euclidean norm (distance) between the points of A and B. The function $h(A,B)$ is called the directed Hausdorff distance from A to B, which identifies the largest distance of $a \in A$ from any point of $b \in B$, and measures the distance from $a \in A$ to its nearest neighbour in $b \in B$ using the Euclidean norm. Thus, $h(A,B)$ ranks each point of $a \in A$ based on its distance to the nearest point in B, and then use the largest ranked point distance, as it represents the most mismatched point of A. Similarly, $h(B,A)$ is the directed Hausdorff distance from B to A. The Hausdorff distance $H(A,B)$ is the maximum of $h(A,B)$ and $h(B,A)$.

4. IRL algorithm implementations: Analysis and results

In this study, the training dataset was used to estimate the reward function weights and to compute the cyclist optimal policy for each type of interactions. The estimated optimal policies were used to simulate road user trajectories, which were validated using the validation dataset.

4.1. Expert demonstrations and state and action discretization

Cyclist and pedestrian trajectories that are involved in following and overtaking interactions were analyzed separately and were used to estimate the variables (features) that describe the cyclist state and action at each time step. Most of the interactions analyzed in this work were between a single pedestrian/cyclist pair and took place in low shared space densities. Five features were used to describe the state of the cyclist at each time step which include: the longitudinal distance between the cyclist and the pedestrian, the lateral distance between the cyclist and the pedestrian, angle difference between the cyclist and the pedestrian, cyclist speed, and speed difference between the cyclist and the pedestrian. The action of the cyclist at each time step is defined by two variables; cyclist acceleration, and cyclist direction represented by the change in cyclist yaw rate, i.e., the change in cyclist steering angle between the next time step and the current time step. Analysis of these profiles showed multiple characteristics of cyclists and pedestrians in the following and overtaking interactions as illustrated in Fig. 6. Descriptive statistics for these characteristics are presented in Table 1. It can be seen that the average cyclist speed and the cyclist-pedestrian speed difference are higher for the overtaking interaction compared to the following interaction. The overtaking interaction is associated with a larger average lateral distance and a smaller average longitudinal distance comparing with the following interaction. Regarding the cyclist action, the average acceleration and change in yaw rate are higher for the overtaking interaction compared to the following interaction.

The space of the state was discretized for each interaction type by dividing each state feature into six levels based on equal frequency observation in each level. The space of the action was discretized for each interaction type by dividing the acceleration into five levels based on equal frequency observation in each level, and dividing the cyclist direction into five equal intervals length. The number of levels of the state features was selected based on a trade-off between the policy prediction accuracy and the computational cost, i.e., CPU computational time and required RAM. This discretization results in a number of states of $6^5 = 7776$ states, and a number of actions of $5^2 = 25$ actions. The discretization intervals of the state and action for the following and overtaking interactions are represented in Tables 2 and 3, respectively. Negative values for the angle difference (state feature) or change in cyclist direction (action) indicate a counter-clockwise angle. Negative speed difference values indicate higher pedestrian speed value. Negative longitudinal distance values indicate that the cyclist has overtaken the pedestrian.

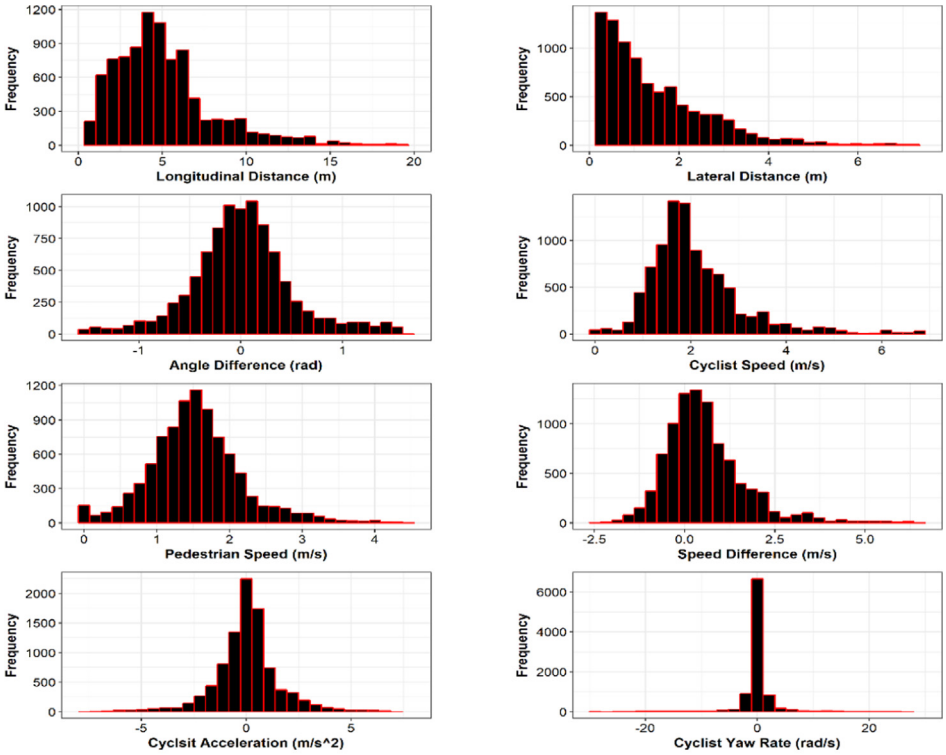
4.2. Reward function weights and optimal policy estimation

4.2.1. Cyclist-pedestrian following interaction

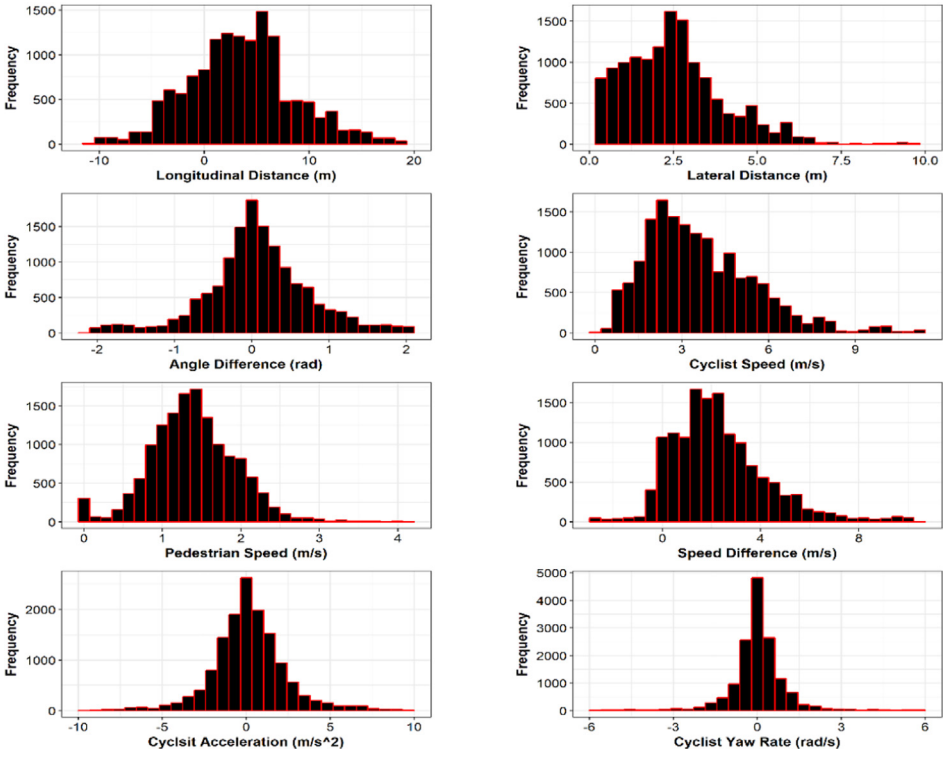
The estimated reward function weights for the following interaction using the Maximum Entropy (ME) and Feature Matching (FM) algorithms (Levine, Popovic, & Koltun, 2011) are presented in Fig. 7. The estimated reward function weights give insights about cyclists' preferences during their interaction with pedestrians in shared spaces. The estimated reward function weights for each state feature are estimated relative to the first level (level 1) of that feature. Reward function estimates suggest that states containing lateral distances in the largest-lateral distance level 6 have lower reward than the states containing lateral distances in the reference level 1, indicating a low value of being in a large-lateral distance level. Lateral distances in level 2 have the highest reward value, while reward weights for lateral distances levels 3 and 4 are slightly lower, suggesting a preferred lateral distance in level 2 [0.35, 0.657 m) according to the ME algorithm. However, the FM algorithm suggests that cyclists highly prefer intermediate lateral distances (level 4) which has a range of [1.06, 1.67 m). Both FM and ME algorithms predict similar behavior except for lateral distances levels 2 and 3.

For the longitudinal distance between cyclists and pedestrians, the estimated reward function weights from both the FM and ME algorithms suggest that cyclists have two preferences which are either to be in states with low-longitudinal distance (level 2) or intermediate/high-longitudinal distance (levels of 4 and 5 for ME and level 6 for FM). The first preference is associated with the cases of following at relatively higher shared space density condition where the average following distance was measured to be 2.36 m according to a shared space behavioral study (Alsaleh, et al., 2020). This case is suggested to be associated with moderate preferences of the cyclists to be in states with angle difference level 3 which has a range around the zero. However, the other preference is suggested to be associated with a low density shared space condition where cyclists prefer to keep moderate longitudinal distance with pedestrians and prefer to change the steering angle slightly to avoid the pedestrians, which is shown in cyclists preferences to have an angle difference with pedestrians deviated from the zero, i.e., angle difference level 5 (9–22 deg) for ME and level 1 (<−22 deg) for FM algorithms. Both FM and ME algorithms generally predict similar behavior except for levels 4 and 5 for longitudinal distances and levels 5 and 6 for angle difference.

For the cyclist speeds, both the FM and ME algorithms suggest that cyclists do not prefer to have very low speed, instead, they prefer to be in states with high-speed levels. The cyclists highly prefer to be in speed levels of 4 and 5 [1.87–2.77 m/s) according to ME algorithm, and speed level 6 (>2.77 m/s) according to FM algorithms. For the speed difference, both FM and Me algorithms generally suggest that cyclists do not prefer to be very slow compared with pedestrians. The ME algorithm suggest that cyclists prefer to be within the speed of the hindered pedestrians as the preference is to be in speed difference level 2 and 3 which have a speed difference range from −0.394 to 0.381 m/s. The FM algorithm is less consistent across levels as it suggests that cyclists prefer to have intermediate speed difference levels 3 and 4, however, their highest preference to be in high-speed difference level 6. The results of the ME algorithm are generally consistent with the observational behavioral study (Alsaleh, et al., 2020) as the cyclists in the following interactions try to maintain following distance with the hindered pedestrians with small fluctuation in their speeds. Overall, the ME reward function estimates are more stable



(a) Following Interaction



(b) Overtaking Interaction

Fig. 6. Characteristics of the cyclist and pedestrian behaviour in following and overtaking interactions.

Table 1
Data Descriptive Statistics (mean [standard deviation]).

Variable	Following Interaction	Overtaking Interaction
Cyclist Speed [m/s]	2.099 [0.993]	3.697 [1.852]
Pedestrian Speed [m/s]	1.541 [0.647]	1.392 [0.547]
Speed Difference [m/s]	0.558 [1.130]	2.305 [1.973]
Angle Difference [Rad]	0.013 [0.507]	0.0789 [0.688]
Lateral Distance [m]	1.452 [1.369]	2.587 [1.880]
Longitudinal Distance [m]	5.132 [3.219]	3.611 [5.223]
Cyclist Acceleration [m/s^2]	0.057 [1.533]	0.1772 [2.216]
Δ Cyclist Yaw Rate [rad/s]	−0.012 [3.167]	0.111 [3.438]

Table 2
States and actions discretization intervals for the following cyclist-pedestrian interaction.

State Level	Longitudinal Distance [m	Lateral Distance [m]	Angle Difference [Rad]	Cyclist Speed [m/s]	Speed Difference [m/s]
1	$(-\infty^*, 2.27)$	$(-\infty^*, 0.35)$	$(-\infty^*, -0.396)$	$(-\infty^*, 1.32)$	$(-\infty^*, -0.394)$
2	[2.27, 3.59)	[0.35, 0.657)	[−0.396, −0.164)	[1.32, 1.64)	[−0.394, 0.0399)
3	[3.59, 4.54)	[0.657, 1.06)	[−0.164, 0.00565)	[1.64, 1.87)	[0.0399, 0.381)
4	[4.54, 5.71)	[1.06, 1.67)	[0.00565, 0.169)	[1.87, 2.24)	[0.381, 0.767)
5	[5.71, 7.48)	[1.67, 2.58)	[0.169, 0.397)	[2.24, 2.77)	[0.767, 1.44)
6	$[7.48, \infty^*)$	$[2.58, \infty^*)$	$[0.397, \infty^*)$	$[2.77, \infty^*)$	$[1.44, \infty^*)$
Action Level	**Acceleration [m/s^2]**	**Δ Yaw Rate [rad/s]**			
1	$(-\infty^*, -0.876)$	$(-\infty^*, -23.6)$			
2	[−0.876, −0.153)	[−23.6, −10.3)			
3	[−0.153, 0.281)	[−10.3, 2.95)			
4	[0.281, 0.902)	[2.95, 16.2)			
5	$[0.902, \infty^*)$	$[16.2, \infty^*)$			

*$(-\infty^*)$ and (∞^*) represents the possible smallest and largest rational values.

Table 3
States discretization intervals for the overtaking cyclist-pedestrian interaction.

State Level	Longitudinal Distance [m]	Lateral Distance [m]	Angle Difference [Rad]	Cyclist Speed [m/s]	Speed Difference [m/s]
1	$(-\infty^*, -1.44)$	$(-\infty^*, 0.963)$	$(-\infty^*, -0.47)$	$(-\infty^*, 2)$	$(-\infty^*, 0.483)$
2	[−1.44, 1.36)	[0.963, 1.78)	[−0.47, −0.132)	[2, 2.66)	[0.483, 1.38)
3	[1.36, 3.47)	[1.78, 2.42)	[−0.132, 0.0546)	[2.66, 3.4)	[1.38, 2.02)
4	[3.47, 5.59)	[2.42, 2.9)	[0.0546, 0.287)	[3.4, 4.23)	[2.02, 2.78)
5	[5.59, 8.18)	[2.9, 3.91)	[0.287, 0.678)	[4.23, 5.47)	[2.78, 3.98)
6	$[8.18, \infty^*)$	$[3.91, \infty^*)$	$[0.678, \infty^*)$	$[5.47, \infty^*)$	$[3.98, \infty^*)$
Action Level	**Acceleration [m/s^2]**	**Δ Yaw Rate [rad/s]**			
1	$(-\infty^*, -1.29)$	$(-\infty^*, -20.5)$			
2	[−1.29, −0.276)	[−20.5, −6.71)			
3	[−0.276, 0.467)	[−6.71, 7.11)			
4	[0.467, 1.59)	[7.11, 20.9)			
5	$[1.59, \infty^*)$	$[20.9, \infty^*)$			

*$(-\infty^*)$ and (∞^*) represents the possible smallest and largest rational values.

compared with the FM estimates. The estimated reward function using the Feature Matching (FM) algorithm has higher intercept to parameter weight ratio compared to the estimated weights using the Maximum Entropy (ME) algorithm. This can indicate that the estimated reward function using FM algorithm did not adequately learn the following interaction behavior compared with the estimates of the ME algorithm, which can limit the transferability of the FM model.

A visualization of the reward function value estimates over states from applying both the ME and the FM algorithms are presented in Fig. 8. The figures show the differences in the reward value across the different states based on the features value of each state. The ME figure shows that having low lateral distance (level 2) combined with a low or intermediate longitudinal distance (level 2, levels 4 or 5) provides the highest reward for the ME algorithm when compared to other combination of lateral and longitudinal distances. However, The FM figure shows that having intermediate lateral distance (level 4) combined with a low or large longitudinal distance (level 1 or 6) provides the highest reward. Moreover, the figures show that having intermediate/high cyclist speed (levels 4 and 5), e.g, cyclist speed range from 1.87 to 2.77 m/s, combined with

Fig. 7. Estimated reward function weights for the following interaction.

intermediate speed difference (levels 2 and 3) provides the highest reward for the estimated of ME algorithm comparing with having very low or very high cyclist speed with low or high speed difference. However, having high cyclist speed (level 6) combined with high speed difference (level 6) provides the highest reward for the estimates of FM algorithm. The states associated with an angle difference around zero (levels 3) and angles deviated from zero (level 5) or (level 1) have higher rewards compared with the rewards associated with other states for the estimates of ME and FM, respectively.

4.2.2. Cyclist-pedestrian overtaking interaction

The estimated reward function weights for the overtaking interaction using the Maximum Entropy (ME) and Feature Matching (FM) algorithms are presented in Fig. 9. Similar to the previous analysis, the estimated reward function weights for each state feature are estimated relative to the first level (level 1) of that feature. The ME reward function estimates suggest that states containing intermediate lateral distances have higher reward than the states containing lateral distances in the reference level 1, indicating a low value of being in the lowest-lateral distance level. Lateral distances in the intermediate levels of 2–5 [0.963, 3.91 m) have higher reward values than the extreme largest lateral distance of level 6, with the highest reward value for lateral distance level 5, suggesting a preferred intermediate lateral distance especially in level 5 [2.9, 3.91 m). This result is expected as the overtaking maneuver involves swerving and increase of lateral distance to overtake slower pedestrians. However, the FM algorithm suggests that cyclists prefer lateral distances in the low lateral distance level 1 [0, 0.963 m) with a lower preference for lateral distance in the intermediate level 5. Both algorithms predict different lateral distance preferences.

For the cyclist speeds, the ME algorithm reward weights indicate that states that have speed in the intermediate and high speed levels have higher reward weight than the states containing the speed reference level 1. This suggests that cyclists preferring to be in states with intermediate to high speed levels of 4 to 6 which have speeds range of (>3.4 m/s). However, the FM estimates suggest that cyclist prefer to be in states with very high cycling speed (level 6) compared to the intermediate speed level 4, while their highest preference is to be in the low speed reference level 1. For the speed difference, both the FM and ME algorithms generally suggest that cyclists do not prefer to be much faster than pedestrians as they assign the lowest weight of being in the highest speed difference level 6. The ME and FM algorithms assign high weights of being in speed difference level 2 comparing with the lowest speed difference reference level 1, suggesting that cyclists prefer to be faster than pedestrians within the speed difference level 2 which has a speed difference range from 0.483 to 1.38 m/s. However, the FM algorithm is less consistent across states and assigns a relatively higher weight of being in speed difference level 5 [2.78, 3.98 m/s) compared to the speed difference level 2, suggesting a slightly higher preference in being in speed difference level 5 compared with level 2 according to FM algorithm.

For the longitudinal distance between cyclists and pedestrians, the estimated reward function weights from the ME algorithms show that states containing longitudinal distances in the large longitudinal distance levels have lower weights than longitudinal distances in the reference level 1 which have a range of (<-1.44 m), indicating a low reward value of being in large-longitudinal distance states. This suggests that cyclists prefer to overtake pedestrians with their highest preference to

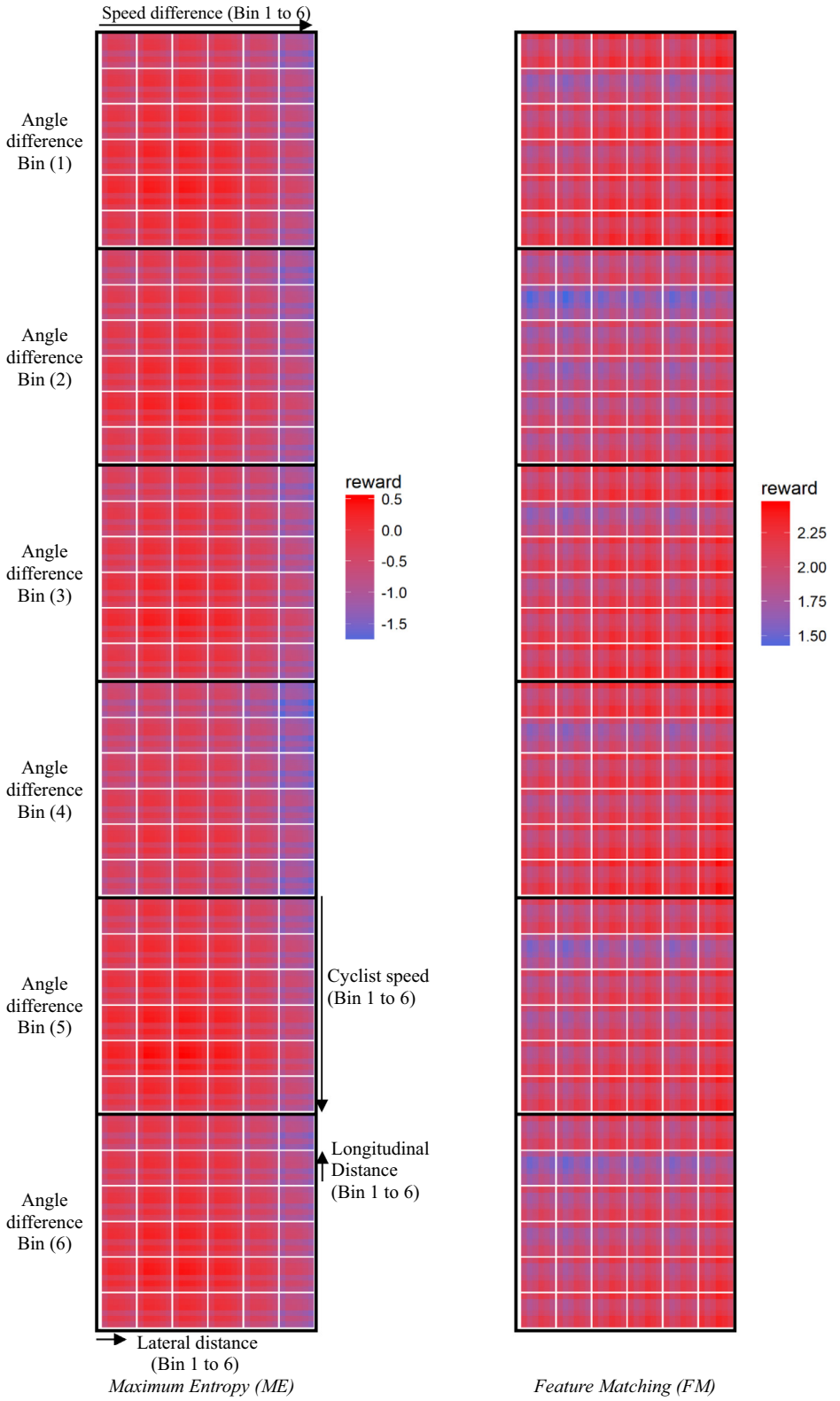


Fig. 8. Estimated reward function for the following interaction using ME and FM IRL.

Fig. 9. Estimated reward function weights for the overtaking interaction.

be in states with longitudinal distance level 1 (<−1.44 m), where the negative value of the longitudinal distance indicates that cyclists have overtaken pedestrians. Similarly, the FM algorithm assigns lower weights of being in high longitudinal distance levels 4, 5, and 6 compared to the low longitudinal distance level 1. However, the FM algorithm is inconsistent across levels and assigns higher weight of being in longitudinal distance level 3 [1.36, 3.47 m) compared to longitudinal reference level 1.

For the direction (angle) difference between the cyclists and pedestrians, the reward function weights of the ME algorithm assigns higher reward weights for states that contain angle difference levels of 2, 3, and 5 which have ranges of [−27, −7 deg), [−7, 3 deg), and [16, 38 deg), respectively. This suggests that cyclists prefer to be within these angle difference levels. The angle difference levels of 2 and 5 are associated with cyclists during the overtaking maneuvers where the overtaking takes place from the left (counter clockwise) or the right (clockwise), respectively. The angle difference of level 3 is suggested to be associated with the end of the overtaking maneuvers as the cyclist return to the original heading direction. Similarly, the reward functions estimated using the FM algorithm assign higher reward weights for states that containing angle difference levels of 3, and 5, suggesting that cyclists prefer to be within these angle difference levels. However, the FM algorithm agrees with ME preferences except for the angle difference level 2. Similar to the modeling of the following interaction, the ME algorithm provides a more stable estimates of the reward function weights, while the FM algorithm provides less consistent reward weights. Moreover, the estimated reward function using the Feature Matching (FM) algorithm has a higher intercept to parameter weight ratio compared with the estimated weights using the Maximum Entropy (ME) algorithm, which can limit the transferability of the FM model.

A visualization of the reward function value estimates over states from applying the ME and the FM algorithms are presented in Fig. 10. The ME figure shows that having intermediate to high lateral distances (level 4, 5, and 6) combined with a low longitudinal distance (level 1) provides the highest reward for the ME algorithms, when compared to states of having very low or very large lateral distances and large longitudinal distances. While the FM Figure shows that having very low lateral distances (level 1) with intermediate longitudinal distances (level 3) provides the highest reward. Moreover, the figures show that having high cyclist speed (levels 5 and 6 for ME) combined with speed difference (levels 2 for ME) provides the highest reward value compared to other combinations of speed and speed difference. The FM Figure shows that having low cyclist speed (level 1) combined with low or high speed difference (level 2, 3 or 5) provides the high reward value. The states associated with an angle difference in levels 2, 3 and 5 for ME and FM (except for level 2 for FM) have higher reward compared with the states associated with other levels.

4.3. Trajectory prediction

The cyclist optimal policy estimated from applying the ME and FM IRL algorithms is used to simulate the cyclist trajectories for each type of interaction. The simulation was run on an Intel® Core i7 with 16 GB RAM at a resolution of 30 HZ (1/30 sec). A simulation tool was developed using the R-software (R Core Team, 2018) to simulate the cyclist trajectories given the surrounding environment of the validation data set, which includes unconstrained pedestrians flow. Fig. 11 shows the workflow of the simulation tool/code. The simulation tool first initializes the simulation environment that includes the pedestrian

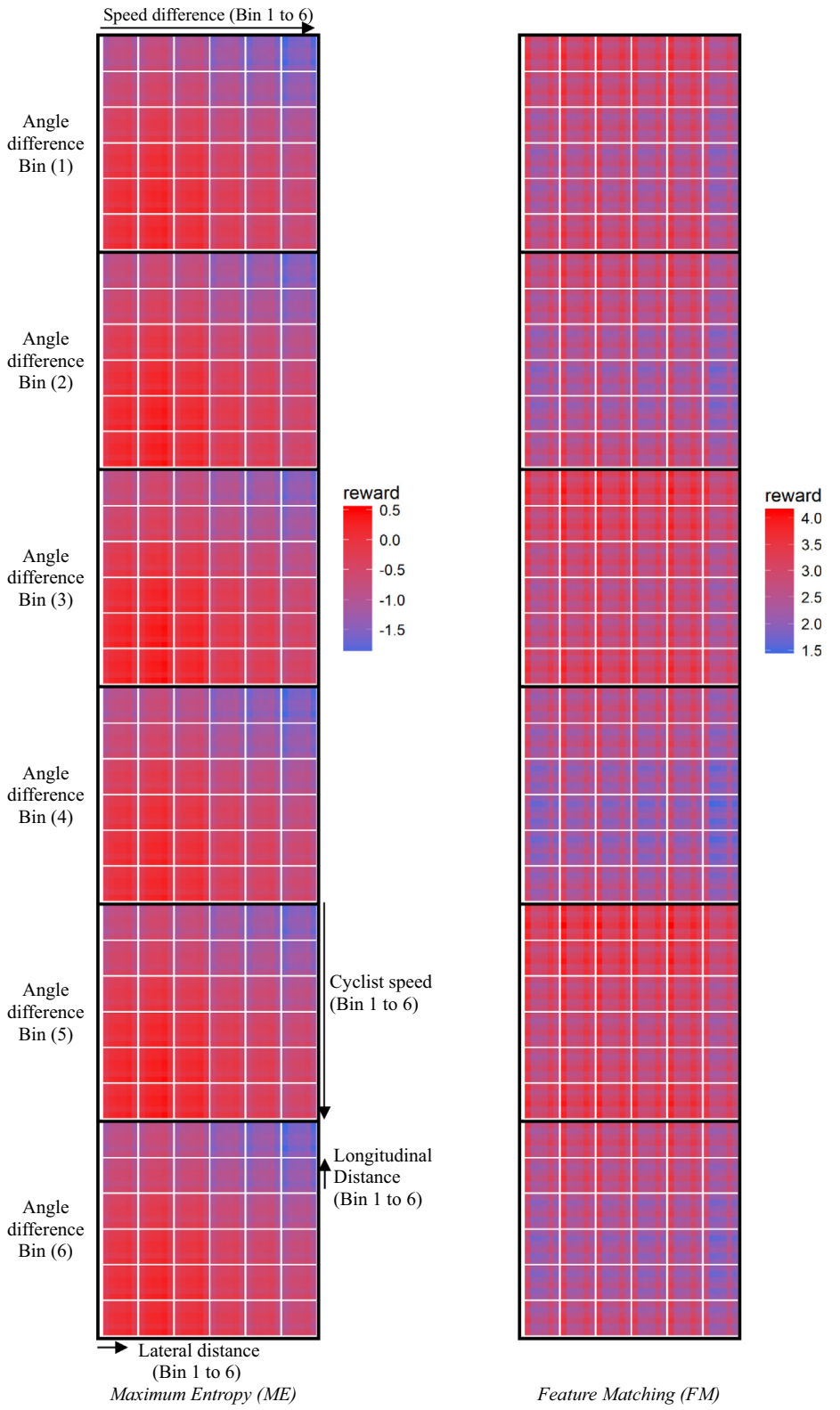


Fig. 10. Estimated reward function for the overtaking interaction using ME and FM IRL.

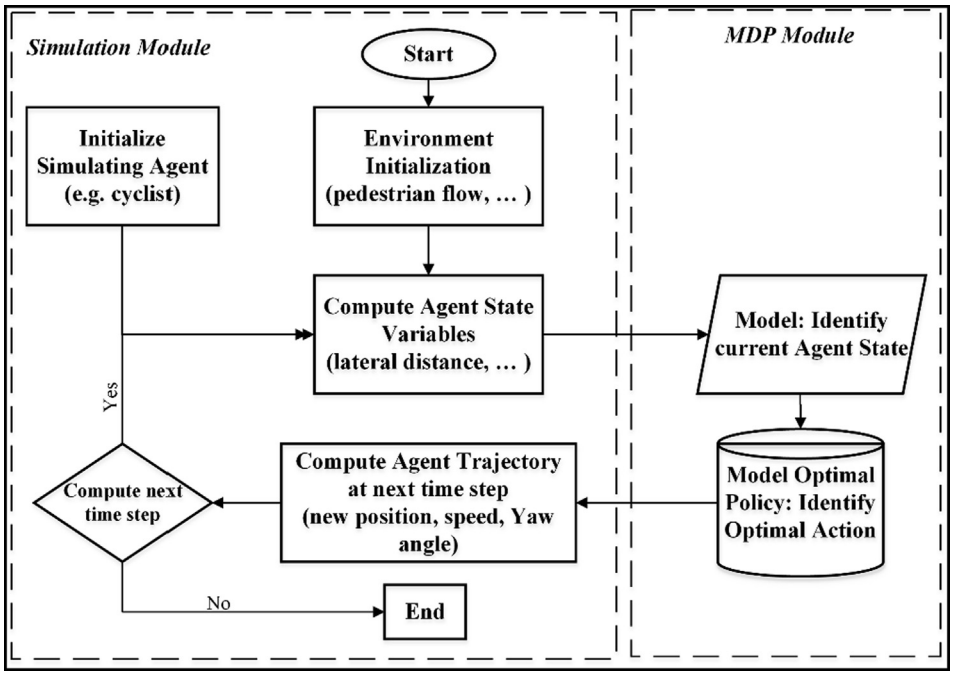


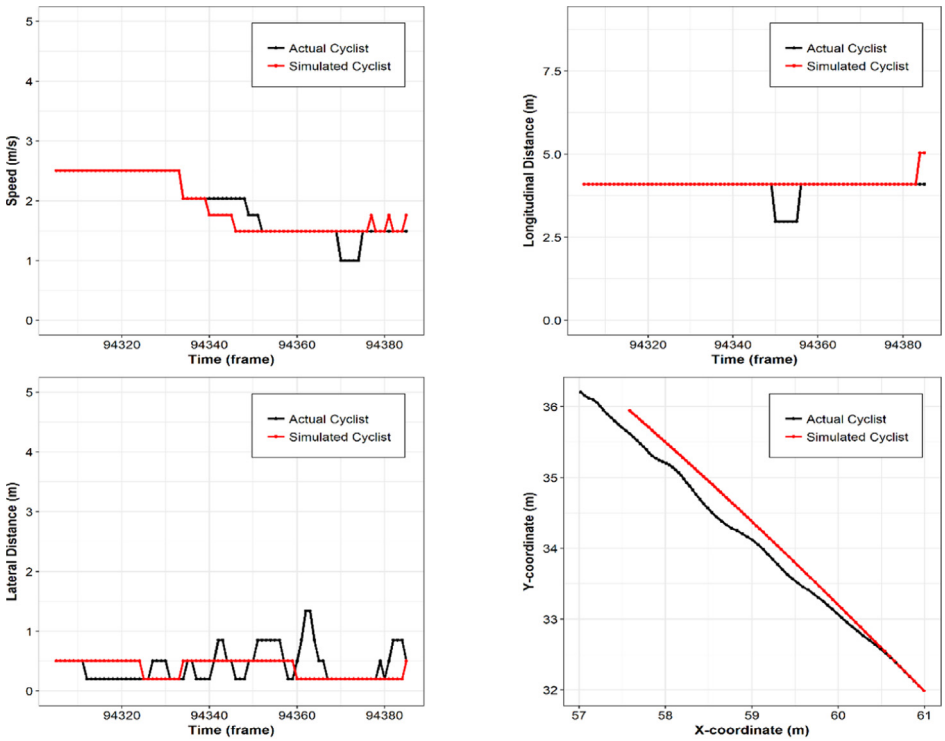
Fig. 11. Simulation tool workflow.

flow and their behaviour over time. The simulating agent, e.g., cyclist, is then initialized with information about its initial position, speed and yaw angle. Then, the simulation tool calculates the agent state variables based on the surrounding environment and assesses the state of the cyclist. The cyclist then takes an appropriate action based on the estimated optimal policy at each time step. The average value of the actions (acceleration and change in yaw rate) in each action interval (Tables 2 and 3) were used to represent the action for that interval for each type of interaction. The cyclist position, speed, and yaw angle is then updated at each time step based on the motion equations.

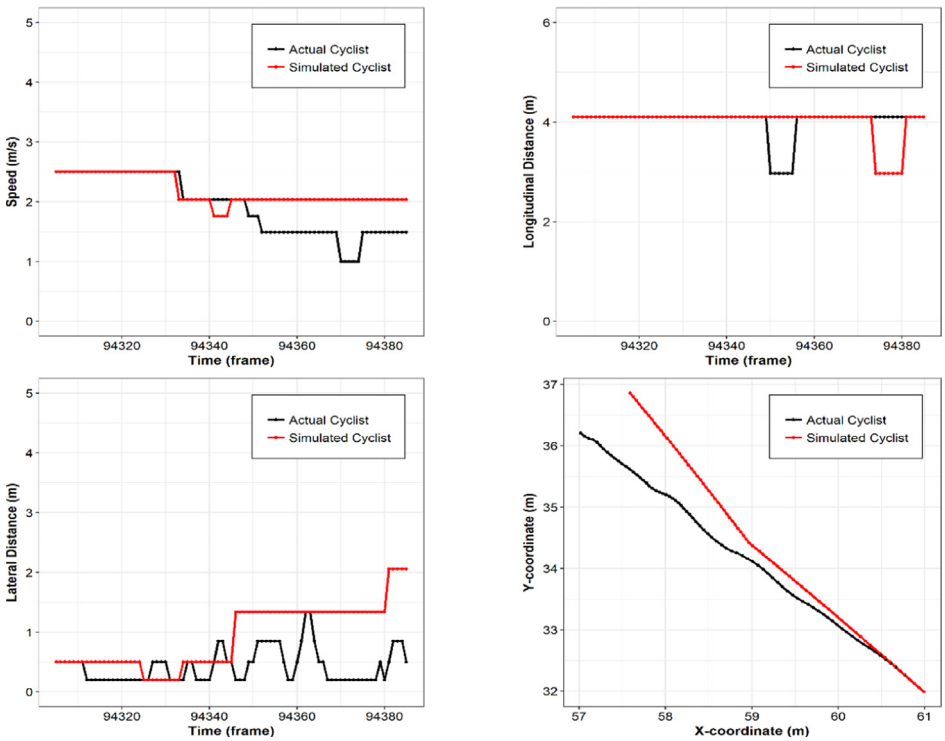
Trajectories in the validation datasets were simulated using the developed models from applying the two FM and ME IRL algorithms. The difference between the actual and simulated trajectories were evaluated using the absolute mean error and the Hausdorff distance (Table 4). Overall, the ME algorithm shows better accuracy and less dissimilarity between the actual and simulated trajectories for both the following and overtaking interactions. The ME algorithm also produces more accurate estimates of cyclist speed and longitudinal and lateral distances compared to with the FM algorithm. Both algorithms predict the cyclist speed more accurately than the cyclist position. For the following interaction, the ME algorithm achieved average improvement for the prediction of cyclist speed of about 12.3% (10.6% using Hausdorff distance) compared to the FM algorithm. While the average improvements in the estimation of longitudinal and lateral distances are more pronounced and is equal to 19.5% (28.7% using Hausdorff distance), and 20.7% (22.9%), respectively. For the overtaking interaction, the ME algorithm shows average improvements for the prediction of cyclist speed of about 20.5% (27.9% using Hausdorff distance) than the FM algorithm. While the average improvements in the estimation of longitudinal and lateral distances are 17.3% (21.6% using Hausdorff distance) and 20.6% (17.1%), respectively. Examples of actual and simulated trajectories and the corresponding speed profile, longitudinal and lateral distance profiles from applying both FM and ME IRL algorithms for the following and overtaking interactions are presented in Figs. 12 and 13 (as discrete intervals), respectively. As shown in the figures, the ME model is capable of reproducing more accurate following and overtaking behaviour compared to the FM models.

Table 4
Prediction errors for FM and ME IRL models for the follwoing interaction.

Variable		Maximum Entropy (ME)		Feature Matching (FM)	
		Avg. MAE	Avg. Hausdorff distance	Avg. MAE	Avg. Hausdorff distance
Following Interaction	Speed (m/s)	0.33	0.51	0.37	0.58
	Longitudinal distance (m)	0.64	1.10	0.80	1.54
	Lateral distance (m)	0.86	1.00	1.08	1.30
Overtaking Interaction	Speed (m/s)	0.66	1.11	0.83	1.54
	Longitudinal distance (m)	1.65	2.78	2.00	3.54
	Lateral distance (m)	1.47	1.74	1.86	2.10

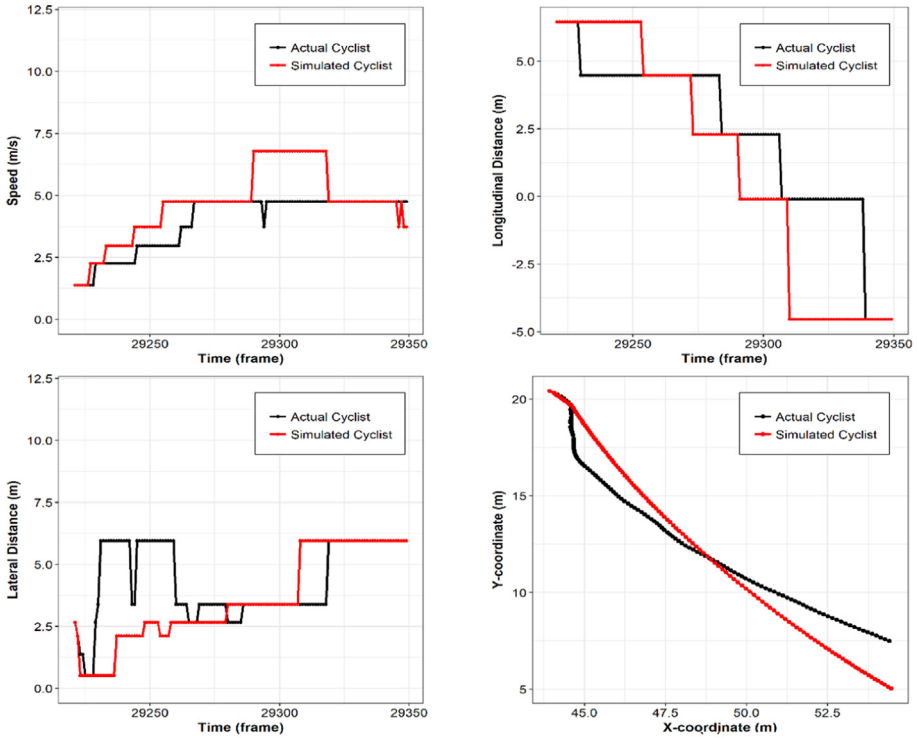


(a) Maximum Entropy (ME)

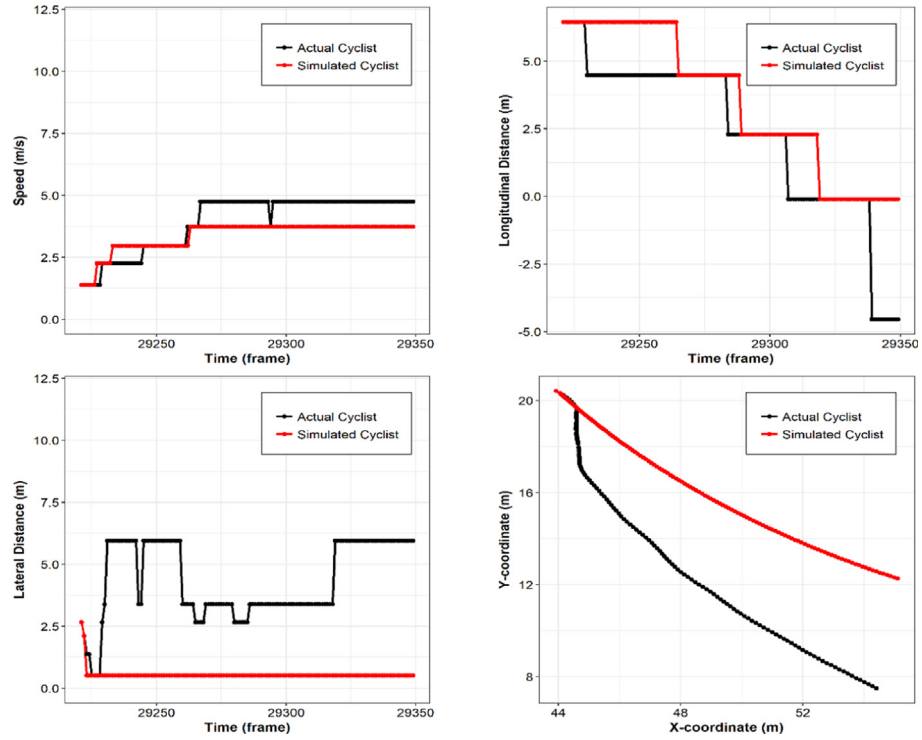


(b) Feature Matching (FM)

Fig. 12. Following interaction discrete simulation using ME and FM IRL algorithms.



(a) Maximum Entropy (ME)



(b) Feature Matching (FM)

Fig. 13. Overtaking interaction simulation using ME and FM IRL algorithms.

5. Discussion and conclusion

The objective of this study was to model cyclist-pedestrian interactions in shared spaces. Two cyclist-pedestrian interactions were considered in the modeling, the following and overtaking interactions. The road users were modeled as utility-based intelligent rational agents. Such a modeling approach accounts for road users' intelligence and their ability to logically assess the surrounding environment and take optimal actions that maximize their utilities in order to achieve their goals. This is considered an important step in modeling road users' intelligence in microsimulation platforms, as most of the previous modeling frameworks ignored the intelligence of road users. Considering the intelligent of road users in microsimulation models is important especially in shared space modeling as they can have different degrees of freedom in accomplishing certain tasks, e.g., overtaking. As well, such models are transferable and capable of simulating agents with different characteristics than those used in models developments with reasonable accuracy. The utility (reward) function is the key component that represents how road users logically assess their surrounding environment.

The recovered reward functions using the IRL algorithms are important for estimating road users policy and developing agent-based microsimulation model for cyclists, i.e., utility-based intelligent rational agents. Such simulation tool can benefit urban designers and traffic engineering in visualizing the road users trajectories and evaluating the safety and efficiency of shared space facilities. In this study, two IRL algorithms, the Maximum Entropy (ME) and the Feature Matching (FM), were proposed to recover road users reward functions and estimate their optimal policies. Generally, the Maximum Entropy (ME) algorithm outperformed the Feature Matching (FM) algorithm in developing MDP for cyclist-pedestrian simulation in shared spaces. The ME reward function estimates were more consistent across levels and line with expectations than the FM estimates. The estimated ME reward function yielded to a higher prediction accuracy of road user trajectories than the FM algorithm. The ME IRL algorithm solves the ambiguity issue in reward function estimation, i.e., FM algorithm does not guarantee a unique solution, and accounts for imperfect (non-optimal) observed behaviour (Ziebart et al., 2008).

A platform was developed to simulate cyclist trajectories and the results were compared to actual data. The difference between the actual and simulated trajectories was evaluated using the absolute mean error and the Hausdorff distance. Generally, the ME algorithm outperformed the FM algorithm, and both algorithms predict the cyclist speed more accurately than the cyclist position. The validation approach used in this paper evaluates the accuracy of the cyclist-pedestrian interaction based on assessing the overall simulated trajectories accuracy compared to actual trajectories. Such a validation approach is suitable for simulation tools developed for the purposes of efficiency and operation evaluation of shared space facilities. However, for other applications such as traffic safety evaluation, accurate prediction of traffic safety indicators such as Time-To-Collision (TTC), Post-Encroachment Time (PET), and Evasive actions are of interest. The abrupt change in the cyclist behavior (e.g. speed, distance, and yaw rate) under the discrete MDP modeling framework shown in Figs. 12 and 13 can affect the safety assessment of these interactions as it may lead to more severe traffic conflicts.

Future research work can include the implementation of other IRL techniques that consider the expected nonlinearity in the data including the estimation of the reward function using deep neural networks. Other modeling MDP approaches, as the partially observed Markov Decision Process (POMDP), which accounts for unobserved effects on road users behaviour can also be considered in future research. Moreover, investigating the optimization of the cut-off value of the discretization, which can have a potential in increasing the accuracy of the developed MDP models is an important research area. As well, considering the continuous modeling approach can be useful for developing a simulation tool for traffic safety assessment as it can improve the trajectory prediction accuracy and avoid the abrupt changes in road users' behaviour. The MDP framework can also be extended to other types of interaction in shared spaces including the crossing and head-on interactions, which can be modeled using a multi-agent IRL framework. Most of the interactions modeled in this work were between a single pedestrian/cyclist pair and took place in low shared space densities. Cyclist behavior is defined based on factors related to relative position, speed, and yaw angle between the cyclist and the opponent pedestrian. However, other factors that can affect the decisions of road users in shared spaces as the neighbor condition (i.e. other pedestrians and cyclists) and shared space density can be explicitly considered in the model in future work.

CRediT authorship contribution statement

Rushdi Alsaleh: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing - original draft, Validation, Writing - review & editing. **Tarek Sayed:** Conceptualization, Methodology, Software, Formal analysis, Data curation, Validation, Supervision, Writing - review & editing, Funding acquisition.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.trf.2020.02.007.

References

Abbeel, P., & Ng, A. Y. (2004). Apprenticeship learning via inverse reinforcement learning. In *In the twenty-first international conference on machine learning, Banff, Alberta, Canada* (pp. 1–8).

Alsaleh, R., Hussein, M., & Sayed, T. (2020). Microscopic behavioural analysis of cyclist and pedestrian interactions in shared spaces. *Canadian Journal of Civil Engineering, 47*(1), 50–62.
Alsaleh, R., R., Sayed, T., T., & Zaki (2018). Assessing the Effect of Pedestrians' Use of Cell Phones on Their Walking Behavior: A Study Based on Automated Video Analysis. *Transportation research record, 2672*(35), 46–57. https://doi.org/10.1177/0361198118780708.
Anvari, B., Bell, M. G., Sivakumar, A., & Ochieng, W. Y. (2015). Modelling shared space users via rule-based social force model. *Transportation Research Part C: Emerging Technologies, 51*, 83–103.
Ayres, G., Wilson, B., & LeBlanc, J. (2004). Method for identifying vehicle movements for analysis of field operational test data. *Transportation Research Record, 1886*(1), 92–100.
Beitel, D., Stipancic, J., Manaugh, K., & Miranda-Moreno, L. (2018). Assessing safety of shared space using cyclist-pedestrian interactions and automated video conflict analysis. *Transportation Research Part D: Transport and Environment, 65*, 710–724.
Bratko, I., Urbancic, T., & Sammut, C. (1995). Behavioural cloning: Phenomena, results and problems. In *International Federation of Automatic Control IFAC, Berlin, Germany* (pp. 143–149).
Burstedde, C., Klauck, K., Schadschneider, A., & Zittartz, J. (2001). Simulation of pedestrian dynamics using a two-dimensional cellular automaton. *Physica A: Statistical Mechanics and its Applications, 295*(3–4), 507–525.
Dias, C., Iryo-Asano, M., Nishiuchi, H., & Todoroki, T. (2018). Calibrating a social force based model for simulating personal mobility vehicles and pedestrian mixed traffic. *Simulation Modelling Practice and Theory, 87*, 395–411.
Gavriilidou, A., Daamen, W., Yuan, Y., & Hoogendoorn, S. P. (2019). Modelling cyclist queue formation using a two-layer framework for operational cycling behaviour. *Transportation Research Part C: Emerging Technologies, 105*, 468–484.
Gibb, S. (2015). *Simulating the streets of tomorrow: An innovative approach to modeling shared space*. Bristol, UK: CH2M Hill Inc.
Gipps, P. G. (1981). A behavioural car-following model for computer simulation. *Transportation Research Part B: Methodological, 15*(2), 105–111.
Gorrini, A., Crociani, L., Vizzari, G., & Bandini, S. (2018). Observation results on pedestrian-vehicle interactions at non-signalized intersections towards simulation. *Transportation Research Part F: Traffic Psychology and Behaviour, 59*, 269–285.
Hamilton-Baillie, B. (2008). Shared space: Reconciling people, places and traffic. *Built Environment, 34*(2), 161–181.
Helbing, D., & Molnar, P. (1995). Social force model for pedestrian dynamics. *Physical Review E, 51*(5), 4282–4286.
Hipp, J. A., Bird, A., van Bakergem, M., & Yarnall, E. (2017). Moving targets: Promoting physical activity in public spaces via open streets in the US. *Preventive Medicine, 103*, S15–S20.
Huang, L. et al (2016). Cyclist social force model at unsignalized intersections with heterogeneous traffic. *IEEE Transactions on Industrial Informatics, 13*(2), 782–792.
Hussein, M., & Sayed, T. (2017). A bi-directional agent-based pedestrian microscopic model. *Transportmetrica A: Transport Science, 13*(4), 326–355.
Huttenlocher, D. P., Klanderman, G. A., & Rucklidge, W. A. (1993). Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 9*, 850–863.
Ismail, K., Sayed, T., & Saunier, N. (2013). A methodology for precise camera calibration for data collection applications in urban traffic scenes. *Canadian Journal of Civil Engineering, 40*(1), 57–67.
Jennings, N. R. (2000). On agent-based software engineering. *Artificial Intelligence, 117*(2), 277–296.
Jiang, R., Jia, B., & Wu, Q. S. (2004). Stochastic multi-value cellular automata models for bicycle flow. *Journal of Physics A: Mathematical and General, 37*(6), 2063–2072.
Kaparias, I. et al (2012). Analysing the perceptions of pedestrians and drivers to shared space. *Transportation Research Part F: Traffic Psychology and Behaviour, 15*(3), 297–310.
Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. New York: Springer.
Levine, S., Popovic, Z., & Koltun, V. (2011). Nonlinear inverse reinforcement learning with Gaussian processes. In *Advances in neural information processing systems 24 conference, Granada, Spain* (pp. 19–27).
Liang, X., Baohua, M. A. O., & Qi, X. U. (2012). Psychological-physical force model for bicycle dynamics. *Journal of Transportation Systems Engineering and Information Technology, 12*(2), 91–97.
Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *International joint conference on artificial intelligence, Vancouver, BC* (pp. 674–679).
Luo, Y. et al (2015). Modeling the interactions between car and bicycle in heterogeneous traffic. *Journal of Advanced Transportation, 49*(1), 29–47.
Ma, X., & Luo, D. (2016). Modeling cyclist acceleration process for bicycle traffic simulation using naturalistic data. *Transportation Research Part F: Traffic Psychology and Behaviour, 40*, 130–144.
May, R. J., Maier, H. R., & Dandy, G. C. (2010). Data splitting for artificial neural networks using SOM-based stratified sampling. *Neural Networks, 23*(2), 283–294.
Nagel, K., & Schreckenberg, M. (1992). A cellular automaton model for freeway traffic. *Journal de physique I, 2*(12), 2221–2229.
Ng, A. Y., & Russell, S. J. (2000). Algorithms for inverse reinforcement learning. In *International conference on machine learning, Stanford, CA, USA* (pp. 663–670).
Papadimitriou, E., Yannis, G., & Golias, J. (2009). A critical assessment of pedestrian behaviour models. *Transportation Research Part F: Traffic Psychology and Behaviour, 12*(3), 242–255.
Plekhanova, V. (2002). *Intelligent agent software engineering*. UK: IGI Global.
R Core Team (2018). *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
Rockafellar, & Wets (2009). Variational analysis (3rd ed.). Springer Science & Business Media.
Saunier, N., & Sayed, T. (2006). A feature-based tracking algorithm for vehicles in intersections. In *The 3rd IEEE Canadian conference on computer and robot vision, Quebec, Canada* (pp. 59).
Savitzky, A., & Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry, 36*(8), 1627–1639.
Schönauer, R. et al (2012). Modeling concepts for mixed traffic: Steps toward a microscopic simulation tool for shared space zones. *Transportation Research Record, 2316*(1), 114–121.
Schonauer, R., Stubenschrott, M., Schrom-Feiertag, H., & Mensik, K. (2012). Social and spatial behavior in shared spaces. In *17th International conference on urban planning and regional development in the information society, Schwechat, Austria* (pp. 759–767).
Swinburne, G. (2005). *Report on road safety in Kensington high street*. London: Royal Borough of Kensington and Chelsea.
Teknomo, K. (2006). Application of microscopic pedestrian simulation model. *Transportation Research Part F: Traffic Psychology and Behaviour, 9*(1), 15–27.
Tomasi, C., & Kanade, T. (1991). *Detection and tracking of point features*. Pennsylvania, USA: Carnegie Mellon University.
Wang, W. L., Lo, S. M., Liu, S. B., & Kuang, H. (2014). Microscopic modeling of pedestrian movement behavior: Interacting with visual attractors in the environment. *Transportation Research Part C: Emerging Technologies, 44*, 21–33.
Wiedemann, R. (1974). *Simulation des straßenverkehrsflusses, Schriftenreihe Heft 8*. Karlsruhe, Germany: Institute for Transportation Science.
Ziebart, B. D., Maas, A. L., Bagnell, J. A., & Dey, A. K. (2008). Maximum entropy inverse reinforcement learning. In *Twenty-third AAAI conference on artificial intelligence, Chicago, Illinois* (pp. 1433–1438).