BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES
DOI: 10.24425/bpasts.2024.152602

Human-like Decision Making for autonomous lane change driving: a hybrid inverse reinforcement learning with game-theoretical vehicle interaction model

Yalan Jiang[1], Xuncheng Wu[1]*, Weiwei Zhang[2], Wenfeng Guo[3], Wangpengfei Yu[2], Jun Li[3]

[1] School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai, 201620, China
[2] Shanghai Smart Vehicle Cooperating Innovation Center Co., Ltd., Shanghai, 201805, China
[3] School of Vehicle and Mobility, Tsinghua University, Beijing, 100084, China

Abstract. The development of automated driving vehicles aims to provide safer, comfortable, and more efficient mobility options. However, the decision-making control of autonomous vehicles still embraces limitations on human performance mimicry. These limitations become particularly evident in complex and unfamiliar driving scenarios, where weak decision-making abilities and poor adaptation of vehicle behaviour are prominent issues. This paper proposes a game-theoretic decision-making algorithm for human-like driving in the vehicle lane change scenario. Firstly, an inverse reinforcement learning (IRL) model is used to quantitatively analyse the lane change trajectories of the natural driving dataset, establishing the human-like human cost function. Subsequently, joint safety, comfort to build the comprehensive decision cost function. Use the combined decision cost function to conduct a non-cooperative game of vehicle lane changing decision to solve the optimal decision of host vehicle lane changing. The host vehicle lane-changing decision problem is formulated as a Stackelberg game optimization problem. To verify the feasibility and effectiveness of the algorithm proposed in this study, a lane change test scenario has been established. Firstly, we analyse the human-like decision-making model derived by the maximum entropy inverse reinforcement learning algorithm to verify the effectiveness and robustness of the IRL algorithm. Secondly, the human-like game decision-making algorithm in this paper is validated by conducting an interactive lane-changing experiment with obstacle vehicles of different driving styles. The experimental results prove that the human-like driving decision-making model proposed in this study can make lane-changing behaviours in line with human driving patterns in lane-changing scenarios of expressway.

Key words: expressway lane-changing scenarios; inverse reinforcement learning; Stackelberg game theory; human-like decision-making; interaction model.

1. INTRODUCTION

Autonomous driving decision-making methods have advanced significantly. In simple driving scenarios, vehicles can achieve safe passage. However, many challenges remain in achieving efficient and human-like driving decisions. vehicle decision planning in dynamic environments involves complex interactions among multiple traffic participants, especially in mixed-traffic situations where self-driving and human-driven vehicles coexist. Therefore, decision-planning algorithms must consider the human-likeness of autonomous driving vehicles' behaviour in addition to satisfying the basic requirements of safety and efficiency. This consideration is essential in dynamic environments with multiple traffic participants. As shown in Fig. 1, when faced with scenarios in which the surrounding vehicles (white vehicles) have uncertain motion states and random interaction behaviours, the host vehicle (purple vehicle) needs to have the ability to effectively deal with uncertainty in dynamic environments. The human-like decision-making algorithms for autonomous vehicles in this traffic state must incorporate the intentions of other drivers.

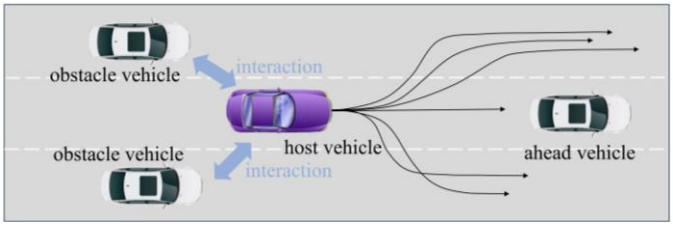


Fig.1. Example of high dynamic lane change interaction scenario

At present, interaction-based decision-making methods for uncertain scenarios are primarily classified into probabilistic reasoning-based methods, learning-based methods, game theory, and others. Firstly, probabilistic

*e-mail: jiangyl@sues.edu.cn

reasoning is a more traditional method, mainly outputting the behavioural probability of interacting vehicles. However, it struggles to handle highly dynamic, multi-participant traffic situations effectively. With the rapid development of machine learning technology, researchers have employed learning-based methods to model the interaction of obstacle vehicles. However, the learning-based method has the disadvantage of poor interpretation and cannot even deal with the algorithm failure in the scenario of high dynamic lane change. Consequently, researchers have turned to game theory approaches, which offer simple modelling and stable convergence, to address driving decision problems in interactive environments. Game theory-based methods are widely used to assess uncertainties in driving environments and can account for interactions among multiple participants.

1.1. Probability-based approaches

In previous studies, the behavioural uncertainty of other participants in highly dynamic interactive driving environments is usually expressed in terms of probabilities. This is the most traditional solution but cannot properly handle highly dynamic, multi-participant traffic situations. Probability-based decision-making methods include probabilistic graphical models, Bayesian networks, and Gaussian mixture models. Probabilistic graphical modelling approaches describe interaction characteristics by building functions that map to numbers with output probabilities of the behaviour of interacting vehicles[1]. Bayesian changepoint detection was used to estimate the possible strategies and behaviours performed by the surrounding vehicles [2], but the computational complexity of the algorithm skyrockets when confronted with a large number of random variables. To reduce the complexity of the model parameters and the dependence on prior knowledge, the researcher used Bayesian networks to model the uncertainty, such as potential strategy distribution of driving behaviour [3] and human psychology at multiple levels of abstraction [4]. Gaussian mixture models can also make inferences about the joint probability distribution for the future trajectory of vehicles based on the driver's intention [5]. During inference using Bayesian networks, uncertainty is continuously quantified during the prediction process to provide more reliable predictions.

1.2. Learning-based approaches

With the rapid development of machine learning in recent years, researchers have started to use learning-based approaches, such as deep learning and reinforcement learning, to solve vehicle-driving decision-making problems in highly dynamic environments. Long Short-Term Memory (LSTM) networks are effective in dealing with long-term dependencies, taking into account the behaviour and style of surrounding vehicles and the interrelationships between vehicles [6]. LSTM networks can also be used for pedestrian trajectory prediction in combination with attention mechanisms [7]. The PPO algorithm learns control strategies in a continuous motion planning space[8], simulates interactions with other vehicles, and solves the motion planning problem with multimodal driving intentions [9], and was used at unsignalised intersections in mixed traffic environments [10]. Raphael et al. proposed to learn pedestrian collision mitigation decision-making strategies for autonomous vehicles via deep reinforcement learning (DRL) to learn pedestrian collision mitigation decision-making strategies for autonomous vehicles [11]. However, deep learning has the drawbacks of inefficient sample usage and low robustness.

Reinforcement learning usually models the problem as a Markov Decision Process (MDP), which is used to solve sequential decision problems. The optimal policy for the host vehicle can be obtained by evaluating the behaviour of the other participants in the MDP framework [12].The high dynamics and uncertainty of the driving environment mainly stem from the uncertainty of human drivers' intentions and noise from sensors, so the driving task is usually described as a partially observable Markov Decision Process (POMDP). Uncertain driving intentions of surrounding vehicles are often used as a hidden variable of POMDP to address the impact of prediction uncertainty on the driving strategy of the auto-vehicle, which can be applied to intersection scenarios [13], expressway driving scenarios [14], and urban through-congestion scenarios [15]. POMDP has difficulties in solving Markov decision processes involving multiple spaces or multiple behaviours. Moreover, reinforcement learning is prone to overfitting or local optimal solutions due to the disadvantage of reward function setting.

Imitation learning enables fast optimisation of strategies by imitating expert demonstrations and is often used to solve problems where the reward function is difficult to define. Zhu et al. modelled pedestrian interaction at intersections and proposed a multi-task imitation learning framework for safe and efficient crossing at intersections [16]. Huang et al. proposed a vehicle interaction model by inverse reinforcement learning (IRL) to achieve accurate prediction of surrounding vehicle trajectories [17]. Wen et al. used IRL to derive reward learning for following driving behaviour and behavioural strategies that consider driving style in following behaviour [18].

1.3. Game theory-based approaches

Game-theoretic approaches have been extensively studied in modelling vehicle interactions due to their stability and convergence. The decisions of host vehicle is influenced not only by the cost functions but also by the future strategies of interacting vehicles [19]. The host vehicle decision problem in lane changing scenarios usually uses game theory to analyse and model the interaction behaviour of vehicles[20-22]. Zhang et al. establishes a game model using fusion model predictive control for forced lane change scenarios and proposes a method to evaluate the aggressiveness of other drivers, thereby maintaining a good balance between driving safety and intelligent decision-making [23]. In [24], an adaptive robust control strategy using the level-k game framework was used to model uncertainty during vehicle interaction in order to increase the safety of lane-changing behaviours in hybrid driving scenarios. Li et al. established a hierarchical inference game theory formulation that can be extended to multiple vehicles to model interactions between drivers and other participants in various driving scenarios [25].

In the process of gaming, the vehicle decision cost function directly affects the decision tendency of the game method. The challenge in modelling human-like decision-making lies primarily in the complex interplay of uncertainties in human driving behaviour. In highly dynamic traffic scenarios, interactions among traffic participants increase the uncertainty in human driving behaviour.

1.4. Contributions.

The paper aims to design a human-like decision-making game theory method that incorporates human driving behaviour models for expressway lane change scenarios and enables vehicles to perform safe and efficient human-like decision-making behaviours. For the quantitative analysis of human-like behaviour, this paper combines a human-like driver model derived from maximum entropy inverse reinforcement learning with the Stackleberg game theory to establish a decision-making model that simulates more realistic traffic flow interaction behaviour. The main contributions of this paper are as follows:

- Modelling human driving behaviour using maximum entropy inverse reinforcement learning algorithm on next generation simulation (NGSIM).
- Establish the cost function of human-like lane change behaviour based on the derived real human driving behaviour model and construct the comprehensive decision-making cost function using the Stackleberg non-cooperative game.
- Implement human-like driving decision-making behaviour in expressway lane-changing scenarios using the Stackelberg game.

1.5. Paper Organization

The remainder of the paper is organized as follows: Section 2 introduces the general architecture of the human-like driving decision-making model. In Section 3, the human-like driving behaviour model based on maximum entropy inverse reinforcement learning is presented. Subsequently, Section 4 discusses the Stackelberg game theory interaction model. Section 5 covers the evaluation and analysis of the proposed model's performance. Finally, the summary of this research work is drawn in Section 6.

2. FRAMEWORK

Figure 2 describes the proposed human-like decision-making framework for autonomous vehicles in expressway lane change scenarios, considering vehicle interactions. First, key information such as ID, position, and state of the interacting vehicles is obtained from the natural driving dataset NGSIM. Then, the interaction scenario's key data is calibrated and preprocessed, and a probability-based inverse reinforcement learning method (maximum entropy inverse reinforcement learning) is used to model real lane change behaviours. This method learns human lane change behaviours under multi-participant interactions and derives the parameters of human feature vectors. Thirdly, a non-cooperative game optimisation problem is constructed based on the interaction process between the main vehicle and the obstacle vehicle during lane-changing. The design constructs a comprehensive decision cost function based on safety, comfort, and human-like characteristics for Nash equilibrium solving. This results in human-like lane change decision-making behaviour for autonomous vehicles in expressway scenarios.

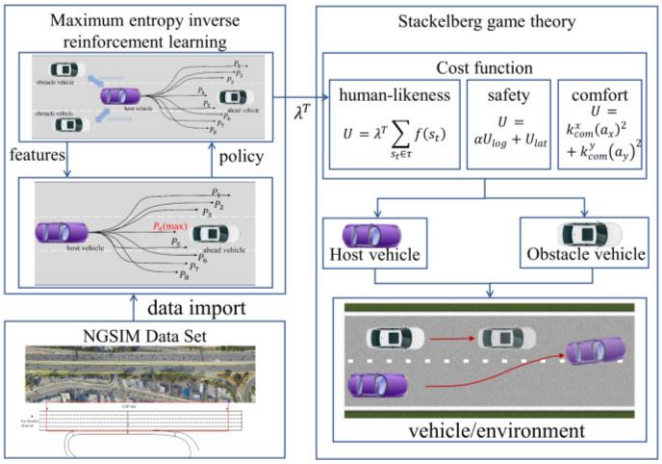


Fig.2. Algorithmic framework for human-like driving decisions

3. HUMAN-LIKE DRIVING MODEL

The inverse reinforcement learning method interprets the Markov decision-making process as the interaction between the agent and the environment, aiming to solve the mapping relationship between the driver's behavioural characteristics and the driving environment, which is essentially the agent's behavioural strategy. We assume that the reward function for the vehicle expressway lane change behavior is linear and is a weighted sum of the selected features. So that the reward function $r(s_t)$ for the s_t state can be set as:

$$r(s_t) = \boldsymbol{\lambda}^T \mathbf{f}(s_t), \tag{1}$$

where $\boldsymbol{\lambda}^T = [\lambda_1, \lambda_2, \lambda_3, \dots \lambda_K]$ is a k-dimensional weight vector, $\mathbf{f}(s_t) = [f_1(s_t), f_2(s_t), \dots, f_K(s_t)]^T$ are the feature vector extracted in the state s_t. The selection of feature vectors is based on the definition of the internal reward function for driving behavioura. Therefore, the reward function of the trajectory is:

$$R(\tau) = \sum_t r(s_t) = \boldsymbol{\lambda}^T \sum_{s_t \in \tau} \mathbf{f}(s_t), \tag{2}$$

where $\mathbf{f}_\tau$ denotes the accumulative characteristics of trajectory τ. The feature information is specified in Section 4. The natural dataset $\mathcal{D} = \{\tau_1, \tau_2, \dots \tau_N\}$ contains information about N trajectories. The problem of solving the maximum entropy model is equivalent to solving the optimization problems. The optimization problem is formulated as follows:

$$\max_p \sum_{\tau \in \mathcal{D}} -P(\tau|\boldsymbol{\lambda}) \log P(\tau|\boldsymbol{\lambda}),$$

$$\text{subject to:} \begin{cases} \sum_\tau P(\tau|\boldsymbol{\lambda}) \mathbf{f}_\tau = \tilde{\mathbf{f}} \\ \sum_\tau P(\tau|\boldsymbol{\lambda}) = 1 \end{cases}, \tag{3}$$

where $P(\tau|\boldsymbol{\lambda})$ represents the state $\tau = \{s_1, s_2, s_3, \cdots, s_T\}$ probability of a trajectory under the parameter $\boldsymbol{\lambda}$. $\mathbf{f}_\tau = \sum_\tau \mathbf{f}(s_t)$ represents the feature expectation of this trajectory, $\tilde{\mathbf{f}}$ represents the characteristic expectation of the expert

trajectory. Thus, the optimization problem of Eq. (3) can be transformed into the following expression:

$$\min_{p} \sum_{\tau \in \mathcal{D}} P(\tau|\boldsymbol{\lambda}) \log P(\tau|\boldsymbol{\lambda}). \quad (4)$$

The solution of the optimization problem is the solution of the maximum entropy inverse reinforcement learning model and also the solution of the reward function sought. We define a Lagrange function $L(P, \boldsymbol{\mu})$:

$$L(P, \boldsymbol{\mu}) = \sum_{\tau \in \mathcal{D}} P(\tau|\boldsymbol{\lambda}) \log P(\tau|\boldsymbol{\lambda}) + \mu_0 \left(1 - \sum_{\tau} P(\tau|\boldsymbol{\lambda})\right) + \sum_{i=1}^{n} \mu_i \left(\tilde{f}_i - \sum_{\tau} P(\tau|\boldsymbol{\lambda}) f_{\tau,i}\right), \quad (5)$$

where $\mu_0, \mu_1, \mu_2, \ldots, \mu_n$ is the Lagrange operator, and n denotes the number of features in the feature vector $f(\tau)$. The original optimization problem of Eq. (4) is expressed as:

$$\min_{p} \max_{\boldsymbol{\mu}} L(P, \boldsymbol{\mu}). \quad (6)$$

Its dyadic expression is:

$$\max_{\boldsymbol{\mu}} \min_{p} L(P, \boldsymbol{\mu}). \quad (7)$$

The Lagrange function $L(P, \boldsymbol{\mu})$is convex function which can be used to solve the dual problem.

$$\Psi(\boldsymbol{\mu}) = \min_{p} L(P, \boldsymbol{\mu}) = L(P_{\mu}, \boldsymbol{\mu}). \quad (8)$$

$$P_{\boldsymbol{\mu}} = \operatorname{argmin}_{p} L(P, \boldsymbol{\mu}) = P_{\mu}(\tau). \quad (9)$$

Combined with the formula $\sum_{\tau} P(\tau|\boldsymbol{\lambda}) = 1$, the partial derivative of the Lagrange function $L(P, \boldsymbol{\mu})$ with respect to $P(\tau|\boldsymbol{\lambda})$:

$$\frac{\partial L(P, \boldsymbol{\mu})}{\partial P(\tau|\boldsymbol{\lambda})} = \sum_{\tau \in \mathcal{D}} (\log P(\tau|\boldsymbol{\lambda}) + 1) - \sum_{\tau} \mu_0 - \sum_{i=1}^{N} \mu_i \sum_{\tau} f_{\tau,i}$$

$$= \sum_{\tau \in \mathcal{D}} \left(\log P(\tau|\boldsymbol{\lambda}) + 1 - \mu_0 - \mu_i \sum_{\tau} f_{\tau,i}\right). \quad (10)$$

$$P(\tau|\boldsymbol{\lambda}) = \exp(\mu_0 - 1) \cdot \exp\left(\sum_{\tau} \boldsymbol{\mu}_i f_{\tau,i}\right). \quad (11)$$

$$Z_{\boldsymbol{\mu}}(\tau) = \sum_{\tau} \exp\left(\sum_{i=1}^{n} \mu_i f_{\tau,i}\right) \approx \sum_{i=1}^{N} \exp\left(\boldsymbol{\mu}^T f_{\bar{\tau}^i}\right). \quad (12)$$

$$P(\tau|\boldsymbol{\lambda}) = \frac{1}{Z_{\boldsymbol{\mu}}(\tau)} \exp\left(\sum_{i=1}^{N} \boldsymbol{\mu}_i f_{\tau,i}\right). \quad (13)$$

$$P(\tau|\boldsymbol{\lambda}) = \frac{1}{\sum_{i=1}^{N} \exp\left(\boldsymbol{\lambda}^T f_{\bar{\tau}^i}\right)} \exp\left(\boldsymbol{\lambda}^T \mathrm{f}(s_t)\right). \quad (14)$$

Where $P(\tau|\boldsymbol{\lambda})$ denotes the probability of trajectory τ under parameter $\boldsymbol{\lambda}$, $Z_{\boldsymbol{\mu}}(\tau)$ denotes normalization factor, $f_{\bar{\lambda}^i}$ denotes the eigenvectors of the trajectory, N denotes the number of generated trajectories. Note that the partition function is easy to solve when the space in which it is located is low-dimensional and predictable. However, the state space in which it is located is high-dimensional and dynamic, making solving the partition function very complicated.

The solution problem of maximum entropy inverse reinforcement learning is equivalent to the constrained optimization problem. The maximum entropy model is subjected to great likelihood estimation:

$$\max_{\boldsymbol{\mu}} \Psi(\boldsymbol{\mu}) = \max_{\boldsymbol{\mu}} \sum_{\tau \in \mathcal{D}} \log P(\tau|\boldsymbol{\lambda}). \quad (15)$$

Combine Eq. (14) and Eq. (15) to obtain the objective function:

$$\mathcal{L}(\boldsymbol{\lambda}) = \sum_{\tau \in \mathcal{D}} \left[\boldsymbol{\lambda}^T f_{\tau} - \log \sum_{i=1}^{N} \exp\left((\boldsymbol{\lambda}^T f_{\bar{\tau}^i})\right)\right]. \quad (16)$$

We optimize Eq. (16) via the gradient approach.

$$\nabla_{\lambda} \mathcal{L}(\boldsymbol{\lambda}) = \sum_{\tau \in \mathcal{D}} \left[f_{\tau} - \sum_{i=1}^{M} \frac{\exp\left((\lambda^T f_{\bar{\tau}^i})\right)}{\sum_{i=1}^{M} \exp\left((\lambda^T f_{\bar{\tau}^i})\right)} f_{\bar{\tau}^i}\right]. \quad (17)$$

$$\nabla_{\lambda} \mathcal{L}(\boldsymbol{\lambda}) = \sum_{\tau \in \mathcal{D}} \left[f_{\tau} - \sum_{i=1}^{M} P\left(\bar{\tau}^i \middle| \boldsymbol{\lambda}\right) f_{\bar{\tau}^i}\right]. \quad (18)$$

Where f_{τ} is the trajectory feature vector of the human driver's driving, and $\bar{\tau}^i$ is the trajectory generated based on the initial conditions of the trajectory τ. In the gradient update, we add L2 regularization on the weights into the objective function. L1 regularization is suitable for sparse coding and feature selection. However, L2 regularization reduces the model parameter complexity and slows down overfitting. Then, the objective function and gradient are:

$$\mathcal{L}(\boldsymbol{\lambda}) = \sum_{\tau \in \mathcal{D}} \left[\boldsymbol{\lambda}^T f_{\tau} - \log \sum_{i=1}^{M} \exp\left((\boldsymbol{\lambda}^T f_{\bar{\tau}^i})\right)\right] - \gamma \boldsymbol{\lambda}^2, \quad (19)$$

$$\nabla_{\lambda} \mathcal{L}(\boldsymbol{\lambda}) = \sum_{\tau \in \mathcal{D}} \left[f_{\tau} - \sum_{i=1}^{M} P\left(\bar{\tau}^i \middle| \boldsymbol{\lambda}\right) f_{\bar{\tau}^i}\right] - 2\gamma \boldsymbol{\lambda}, \quad (20)$$

where γ is the regularization parameter and $\gamma > 0$. objective function gradient is expressed as the sum of the expected eigenvalue difference and the regularization gradient.

4. INTERACTIVE HUMAN-LIKE DECISION-MAKING

The human driver's lane changing behaviour is the drivers' decision-making behaviour after considering the host vehicle state (including speed, acceleration, heading angle, etc.), the state of the surrounding vehicles, the predicted trajectory information. This process is recognised as the interaction process between the host vehicle and the environment. The decision-making process made by the vehicles during the interaction is equivalent to the optimal solution process of the game model between the vehicles.

We choose the Stackelberg game method to model the dynamic game between the interacting parties in the non-cooperative game method. In this paper, according to the vehicle location information and driving environment, we designate the vehicle identity as either a leader or a follower in the game. The game process is dynamic, and the decision-making process of each vehicle adjusts according to the changes in revenue. The Nash equilibrium is eventually reached through the iterative process, forming the optimal strategy that minimizes the cost (or maximizes the benefit) for multiple participants.

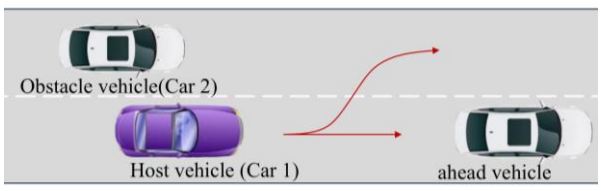


Fig.3. Lane Change Scenario

TABLE 1. Cost magnitude of different interaction behaviours

action cost		obstacle vehicle	
		accelerate	yield
host vehicle	change lane	5,1	3,2
	stay	6,2	4,3

Figure 3 shows a simple driving environment we set up for the lane-changing behaviour. A low-speed vehicle is assumed to exist in front of the host vehicle, which needs to decide whether to change lanes or not. Usually, we consider the host vehicle as car 1, and the obstacle vehicle as car 2. Table 1 lists the example gains for all scenarios during the game between the host vehicle and the obstacle vehicle, with specific number of costs. The host vehicle chooses the optimal strategy by predicting the action of the obstacle vehicle. The host vehicle is the leader in the Stackelberg game and can choose to make the left lane change or stay in lane. The obstacle car is the follower and responds to the leader. It is important that the driving strategies of both vehicles are rational and based on the principle of cost minimization. For example, when the host vehicle chooses to stay in lane, the obstacle vehicle chooses to accelerate in response. Similarly, when the host vehicle chooses to change lanes, the obstacle vehicle chooses to yield. In this example, the optimal strategy of the game is for the host vehicle to change lanes, and for the obstacle vehicle to yield, which is the result of the predictions made by the main vehicle about the behaviours of the obstacle vehicle.

4.1. Cost function design.

The feasibility evaluation of the host vehicle lane change behaviour is quantified through a cost function, typically encompassing driving factors such as safety, traffic efficiency, ride comfort, and spatial benefits of lane-changing. The dynamic decision-making process of lane-changing can be seen as a dynamic game among participants with varying driving strategies. This study considers a combined cost function incorporating driving safety, ride comfort, and human likeness in driving behaviour for Nash equilibrium determination.

The safety cost function of vehicle travel during lane-changing behaviour is manifested in both lateral and longitudinal directions. When the vehicle opts to keep the lane, the cost function primarily concerns the longitudinal cost. The formula for the safety cost function is:

$$U_{\text{saf}} = \alpha U_{\text{lat}} + U_{\text{log}}. \tag{21}$$

U_{log} and U_{lat} in the above equation denote the longitudinal and lateral safety cost functions, respectively. α denotes the driving decision-making behaviour of the host vehicle, $\alpha \in \{1,0\} := \{change\ lane, keep\ lane\}$.

The safety cost function in the transverse direction is defined utilizing the split-axis theorem. Vehicles are presumed to be rectangles of specific length and width, and lateral safety cost function between vehicles is evaluated based on the overlap between these two rectangles. The positions of the rectangles representing the two vehicles are illustrated in Fig.4. The probability of collision is computed as:

$$U_{\text{lat}} = \exp\left(-\sqrt{\frac{{D_{\text{col,v}}}^2 + {D_{\text{col,u}}}^2}{2}}\right). \tag{22}$$

$$D_{\text{col,v}} = \sqrt{{D_{\text{proj(v,1)}}}^2 + {D_{\text{proj(v,2)}}}^2}. \tag{23}$$

$$D_{\text{col,u}} = \sqrt{{D_{\text{proj(u,1)}}}^2 + {D_{\text{proj(u,2)}}}^2}. \tag{24}$$

$$D_{\text{proj(v,i)}} = \begin{cases} \min\left(\left|\frac{v_i \cdot v}{v_i}\right|\right) & if \forall \frac{v_i \cdot v}{|v_i|} < 0 \\ \min\left(\left|\frac{v_i \cdot v}{v_i}\right| - |v_i|\right) & if \forall \frac{v_i \cdot v}{|v_i|} > |v_i| \\ 0 & otherwise \end{cases}. \tag{25}$$

Where $D_{\text{proj(v,i)}}$ denotes the gaps along the separating axis, v_i represents a vector defining the rectangle and v represents the vector to the opposite corner. Similarly, $D_{\text{proj(u,i)}}$ is calculated in the same way. U_{lat} in the above equation represents the index value of the collision with $U_{\text{lat}} \in [0,1]$. A larger separation axle clearance of the vehicle corresponds to a U_{lat} value close to 0, indicating greater safety.

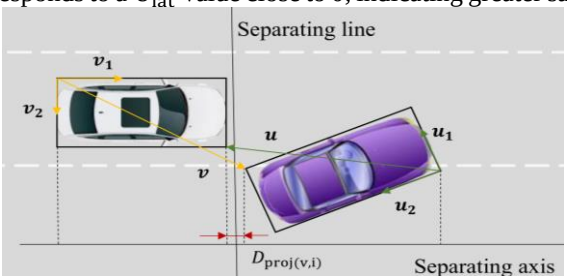


Fig.4. Vehicle position diagram

We define the longitudinal safety cost function based on information such as relative speed of interacting vehicles and longitudinal gap.

$$U_{\text{log}} = k_{\text{saf}}^{\text{v}} \lambda_v (\Delta v_{\text{x}})^2 + \frac{k_{\text{saf}}^{\text{x}}}{[(\Delta X_{\text{x}})^2 + \epsilon]}. \tag{26}$$

$$\Delta v_x = v_{\text{f,x}} - v_{\text{r,x}}. \tag{27}$$

$$\Delta X_x = X_{\text{f}} - X_{\text{r}} - L_{\text{f}}. \tag{28}$$

$$\lambda_v = \begin{cases} 0 & v_{\text{f,x}} \geq v_{\text{r,x}} \\ 1 & v_{\text{f,x}} < v_{\text{r,x}} \end{cases}. \tag{29}$$

where $v_{\text{f,x}}$ and $v_{\text{r,x}}$ denote the longitudinal velocities of the front and rear vehicles, X_{f} and X_{r} denote the longitudinal positions of the front and rear vehicles, respectively. $k_{\text{saf}}^{\text{v}}$ and $k_{\text{saf}}^{\text{x}}$ are the correlation weight coefficients of speed and distance. ϵ is a very small value to avoid the case where the denominator is zero; L_{f} is the length of the front vehicle. The longitudinal positional relationship of the vehicle is shown in Fig. 5.

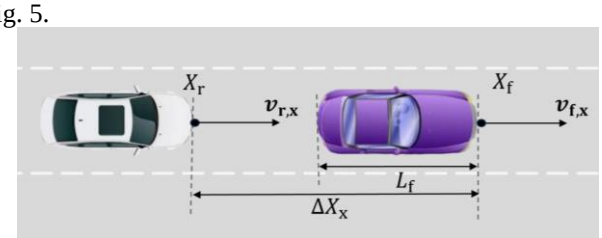


Fig.5. Vehicle longitudinal position relationship

The ride comfort is related to lateral and longitudinal acceleration, and the cost function expression for comfort is:

$$U_{\text{com}} = k_{\text{x,com}}(a_{\text{x}})^2 + k_{\text{y,com}}\left(a_{\text{y}}\right)^2, \quad (30)$$

where $k_{\text{x,com}}$ and $k_{\text{y,com}}$ are the weight coefficients of transverse and longitudinal acceleration, a_{x} and a_{y} are the magnitude of transverse and longitudinal acceleration.

Referring to the trajectory reward function in Eq. (2), we define the human-like cost function as:

$$U_{\text{hum}} = \boldsymbol{\lambda}^T \sum_{s_t \in \tau} \mathbf{f}(s_t). \quad (31)$$

$$\mathbf{f}(s_t) = \left[f_v, f_{ax}, f_{ay}, f_{jx}, f_{THWF}, f_{THWB}, f_{collision}\right]^T. \quad (32)$$

The selection of the trajectory features is based on the human driving state. The selection and definition of features is presented in section 5.1.

In summary, the cost function of the host vehicle is a linear combination of integrated driving safety and human-like properties, expressed as:

$$U_1 = w_{\text{saf}}U_{\text{saf}} + w_{\text{com}}U_{\text{com}} + w_{\text{eff}}U_{\text{eff}} + w_{\text{hum}}U_{hum}. \quad (33)$$

$$U_2 = w_{\text{saf}}U_{\text{saf}} + w_{\text{com}}U_{\text{com}} + w_{\text{eff}}U_{\text{eff}}. \quad (34)$$

where w_{saf}, w_{com}, w_{eff}, w_{hum} represent the weighting coefficients of the safety, comfort, effective and human-like cost functions, respectively. U_1 represents the integrated decision cost function containing the human-like cost function. And U_2 is the cost function considering without human-like cost function.

4.2. Non-cooperative decision making based on Stackelberg equilibrium.

In the Stackelberg game, vehicles adhere to the principle of cost minimization when making lane change decisions. The followers respond to the leader's actions based on their cost functions and the influence of the leader's behaviour. Therefore, the Stackelberg game problem is transformed into a two-layer optimization problem. The decision made by the leader affects the behaviour of the follower, but the leader cannot intervene in the follower's behavioural decision. Similarly, the follower's choice of strategy cannot alter the leader's decision.

A common lane-change scenario involves an interaction game between two participants, with the vehicle lane-change interaction game illustrated in Fig. 3. The optimization problem for the two-vehicle game is formulated as:

$$\gamma^{1^*} = (a_1{}^*, c_1{}^*) = \text{argmax} \left(\min_{(a_{x2}^{hv}, c_2) \in \gamma^2} U_1(a_1, c_l, a_2) \right). \quad (35)$$

$$\gamma^2(a_1, c_1) \triangleq \{\xi \in \Gamma^2: U_2(a_1, c_l, \xi) \geq U_2(a_1, c_l, a_2), \forall a_2, c_2) \in \Gamma^2\}. \quad (36)$$

subject to:

$$0 \leq V_{x,i} \leq V_{\text{x,max}}, i = 1,2. \quad (37)$$

$$a_{\text{min}} \leq a_{x,i} \leq a_{\text{max}}, i = 1,2. \quad (38)$$

Where a_1 denotes the possible acceleration, $a_1{}^*$ denotes the optimal acceleration of the host vehicle, c_l shows if car 1is changing lane, $c_l{}^*$ shows if changing lanes is beneficial for car 1. $\gamma^i, i = 1,2$ denotes the behavioural decision of the vehicle. $U_i, i = 1,2$ denotes the total cost function of the vehicle. $\gamma^2(a_1, c_l)$ denotes the optimal decision of car 2 under the influence of car 1. $\Gamma^i, i = 1,2$ is the set of possible actions of the vehicle.

5. TESTING RESULTS AND PERFORMANCE EVALUATION

This section focuses on verifying the feasibility and effectiveness of the human-like driving decision-making algorithm. We first perform data preprocessing. Next, we use the dataset to validate the effectiveness of the maximum entropy inverse reinforcement learning algorithm. Finally, we test and evaluate the human-like driving decision-making algorithm in a typical interaction scenario.

5.1. Data preparation and processing.

The NGSIM dataset is a publicly available dataset developed and released by the National Highway Traffic Safety Administration (NHTSA) of the U.S. Department of Transportation for the study of road traffic flow and driving behaviour. The dataset provides researchers with real road traffic information that can be used to simulate and analyse issues related to driving behaviour, traffic flow, and road safety, which is one of the indispensable fundamentals for the study of human-like driving decision-making algorithms.

In this paper, we use the NGSIM dataset to process and analyse the U.S. Highway 101. The structure and area recorded in the US-101 dataset are schematically shown in Fig.6. The highway section recorded by the dataset is about 2100 foot (about 640 m) and contains five lanes (lanes 1 to 5), with lane 6 connecting lanes 7 and 8 at both ends, which are the merge-in and merge-out lanes, respectively. The camera used for data acquisition records the traffic state of vehicles travelling in the form of snapshots according to a sampling period of 100ms (10 Hz). The dataset records information, including global and local position information, speed, and type of vehicles. This paper focuses on the human-like decision-making problem in the vehicle lane change scenario. Thus, the driving behaviour of vehicles on ramps is not considered. We randomly selected three hundred vehicle trajectories from the five main lanes as a secondary sampling dataset to serve as reward function learning samples for human-like driving.

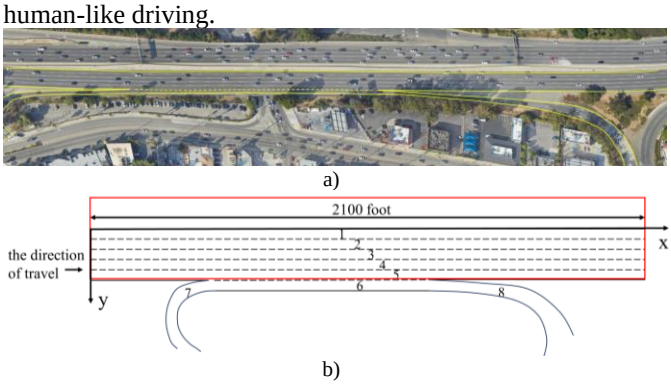


Fig.6. Dataset Preparation: a) US-101 dataset; b) Road structure schematic

Human driving trajectories are influenced by the driver's conscious control and the surrounding environment. For example, during driving, the driver's intent to accelerate and the level of safety threat from surrounding vehicles affect the

magnitude of vehicle acceleration. Therefore, the selection of trajectory characteristic variables will be considered in terms of efficiency, safety and comfort of the vehicle. Specifically, we chose speed v to represent vehicle travelling efficiency, lateral acceleration a_x , longitudinal acceleration a_y and longitudinal jerk $j_x$to represent comfort; and safety indicator was represented by the headways (*THWB* and *THWF*) and collision.

$$f_v(s_t) = v(t) \tag{39}$$

$$f_{ax}(s_t) = |a_x(t)| \tag{40}$$

$$f_{ay}(s_t) = |a_y(t)| \tag{41}$$

$$f_{jx}(s_t) = |jerk_x(t)| = |\dot{a}_x(t)| \tag{42}$$

$$f_{THWF}(s_t) = \frac{X_f(t) - X_{host}(t)}{v_{host}(t)} \tag{43}$$

$$f_{THWB}(s_t) = \frac{X_{host}(t) - X_r(t)}{v_r(t)} \tag{44}$$

$$f_{collision}(s_t) = \begin{cases} 1 & if\ collision \\ 0 & otherwise \end{cases} \tag{45}$$

Where $x_f(t)$ is the longitudinal position of the nearest front vehicle, $X_{host}(t)$ and $v_{host}(t)$ are the position and speed of the host vehicle, $X_r(t)$ and $v_r(t)$ are the longitudinal position and speed of the nearest rear vehicle, respectively. f_{THWF} and f_{THWB} represent the time headway between the host vehicle and the front and rear vehicles, respectively.

Note that the transverse and longitudinal driving models of surrounding vehicles use MOBIL (minimize overall braking induced by lane change) and IDM (intelligent driving model) to predict their future behaviours.

5.2. IRL model analysis.

In the process of vehicle lane-changing, we focus on lateral position change and longitudinal speed change to simplify the algorithm. Therefore, the vehicle decision sampling space is denoted as $\Phi = \{ v_{xo}, y_o \}$.And the transverse information collection only records its lane change information. The sample of the lateral lane change is $\{ y_L, y, y_R \}$, where y represents the initial lateral position, y_L and y_R are the position of the left lane and right lane, respectively. The simulated trajectories of the vehicle are generated using polynomial curves, where the trajectory horizons are 5 s. The driving model IDM parameters of the surrounding vehicles are set as follows: desired velocity $v = v_{\text{current}}$ m/s, maximum acceleration $a_{\max} = 5$ m/s^2, and comfortable acceleration $a_{\text{com}} = 3$m/s^2, minimum desired spacing $s_0 = 1$m.

The initial values of the parameter vectors of the IRL algorithm are randomly sampled using a normal distribution with a mean of 0 and a standard deviation of 0.5. The performance of traditional deep learning optimization algorithms, such as gradient descent and stochastic gradient descent, is often limited by fixed learning rates and parameter update strategies. The parameter optimization in this study uses an adaptive learning rate optimization algorithm, in which the optimization algorithm parameters are: the regularization parameter $\gamma = 0.01$, the learning rate α=0.05, the exponential decay rate of the first and second order β_1=0.9, $\beta_2 = 0.99$.

The raw data of a vehicle's driving trajectory is divided into multiple short trajectories with horizons of 5 seconds for learning. Thirty-five trajectory data will be randomly selected from the short trajectories to train the parameters of the reward function. In contrast, the remaining trajectory data will be used as the test data. During the training process, the difference between the feature expectations of the human driving trajectory and the simulated trajectory is used to establish the gradient for the iterative updating of the cost function parameters. The final displacement error (FDE) and the average displacement error (ADE) between the human driving trajectory and the simulated trajectories are used as the human-likeness evaluation indexes. The training process is illustrated in Figure 7.

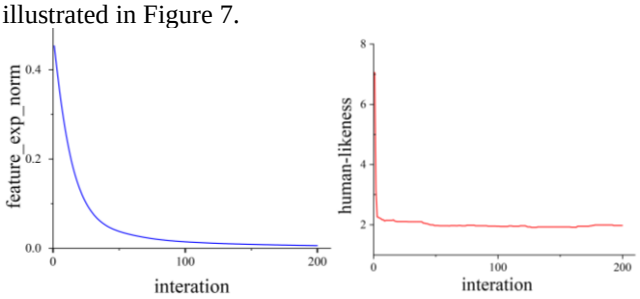
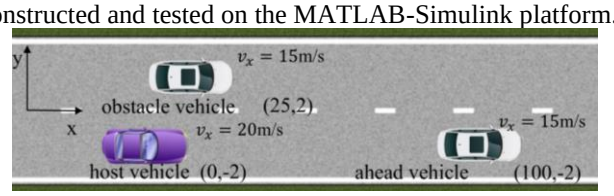


Fig.7. Training process: a) plot of trajectory feature expectation difference; b) plot of human-likeness.

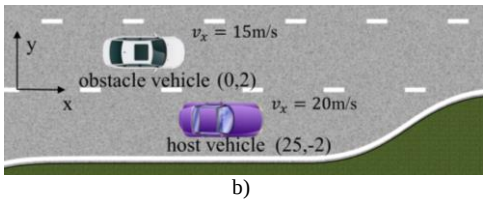
Fig. 7(a) plots the trajectory feature expectation difference curves between the IRL model and the real trajectory during the iteration process. The human-like driving strategy obtained by IRL model could be close to the real trajectory data and the differences between them converge to 0. Fig. 7(b) plots the human-likeness of trajectories curve, where the parameters of the learned reward function are gradually adjusted as the training period increases, allowing the IRL model to better fit human driver behaviour. Fig.7 indicate that the IRL algorithm is learning human-like strategies that increasingly resemble human-driving strategies with more iterations. The experimental results demonstrate the feasibility of the IRL algorithm in enabling the learning of human-like decisions.

5.3. Human-like driving decision-making algorithms

In this paper, we designed two simple expressway lane-change scenarios to analyse and evaluate the effectiveness and human-like nature of the lane-change behaviour of this algorithm. The scenario 1 is a two-lane highway with a lane width of 4 meters. The scenario 2 is a merger scenario. The initial positions and initial speeds of each vehicle are shown in Fig.8. As shown in Fig. 8(a), the driver of the host vehicle intends to change to the adjacent lane and interact with the obstacle vehicle, when the ahead vehicle moves slowly. Similarly, scenario 2 is the ramp import where the host vehicle interacts with the obstacle vehicle on the main road when it cuts into the adjacent lane. All driving scenarios were constructed and tested on the MATLAB-Simulink platform.



a)



b)

Fig.8. Expressway lane change scenario traffic information map: a) scenario 1; b) scenario 2.

The longest common subsequence (LCSS) judging metric is introduced to quantify the similarity of lane change trajectories. Define two trajectories $\tau_1 = \{p_1, p_2, \ldots, p_n\}$ and $\tau_2 = \{p_1', p_2', \ldots, p_m'\}$ between the similarity function SF as well as for:

$$SF(\varepsilon, \tau_1, \tau_2) = \frac{LCSS_\varepsilon(\tau_1, \tau_2)}{\min(n, m)}, \tag{46}$$

where

$$LCSS_\varepsilon(\tau_1, \tau_2) = \begin{cases} 0, \ if\ n = 0\ or\ m = 0 \\ 1 + LCSS\big(Rest(\tau_1), Rest(\tau_2)\big), if\big(Head(\tau_1), Head(\tau_2)\big) \le \varepsilon. \\ \max\big(LCSS_\varepsilon\big(\tau_1, Head(\tau_2)\big), LCSS_\varepsilon\big(Head(\tau_1), \tau_2\big)\big), otherwise \end{cases} \tag{47}$$

where $\big(Head(\tau_1), Head(\tau_2)\big) = \sqrt{(x_n - x_m)^2 + (y_n - y_m)^2}$ represents the distance between two points $p_n = (x_n, y_n)$ and $p_m' = (x_m, y_m)$ on the trajectory. ε represents the matching similarity threshold. The output result interval of the similarity function SF is [0,1], which represents the trajectories have no similarity at all and are identical, respectively.

In this study, we select 10 volunteers with driver's licenses and driving experience to conduct lane change driving experiments in scenario 1 and scenario 2. Each volunteer performs 10 repetitions of lane changing behavior, and formed a comprehensive lane changing trajectory curve used to reflect the lane changing driving behavior of each volunteer, as shown in Fig. 9.

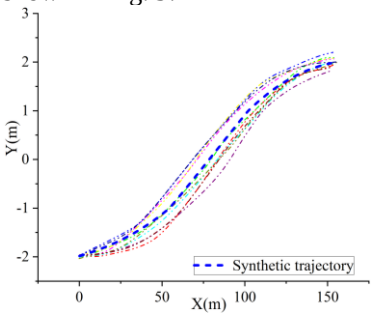
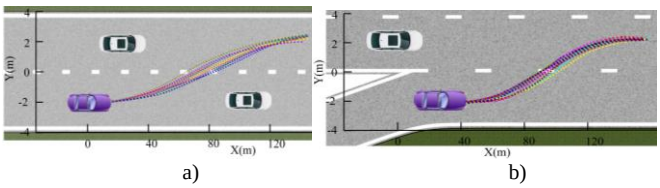


Fig.9. synthetic trajectory



a) b)

Fig.10. Track diagram of lane change driving experiment: a) scenario 1; b) scenario 2.

The results of the lane-changing experiments in the two scenarios are shown in Figure 10, where the synthetic trajectories of 10 volunteers are represented by dashed lines of 10 different colors.as shown in Fig. 10. We record each trajectory location information and calculate similar function separately.

In real traffic driving environments, the social behaviour of the obstacle vehicle affects the decision-making and planning of the host vehicle. Therefore, this study simulates the obstacle vehicle interaction in a real traffic environment by defining different driving styles for the obstacle vehicle. In obstacle vehicle modelling, different driving styles are expressed by setting different parameter weights $w_{saf}, w_{eff}, w_{com}$. The details of the three driving styles are shown in Table 2.

TABLE 2. Weighting coefficients of the cost function for different driving styles

Weighting coefficients	Driving style		
	aggressive	normal	cautions
w_{saf}	20%	50%	75%
w_{eff}	70%	25%	10%
w_{com}	10%	25%	15%

To verify the driving ability of the human-like driving decision-making algorithm in expressway lane change scenarios under the influence of social vehicles with different driving styles, we use U_1 and U_2 from Eq. (33) and Eq. (34), respectively, for the lane-changing decision-making in expressway scenarios. Table 3 show the human-like experimental results.

TABLE 3. Results of simulation

Driving style	Cost function	SF (scenario 1)		SF (scenario 2)	
		range	average	range	average
Aggressive	U_1	0.65~0.70	**0.68**	0.62~0.70	**0.65**
	U_2	0.62~0.68	0.65	0.61~0.66	0.63
Normal	U_1	0.69~0.78	**0.73**	0.65~0.74	**0.69**
	U_2	0.65~0.71	0.69	0.62~0.68	0.64
Conservative	U_1	0.65~0.72	**0.71**	0.63~0.73	**0.68**
	U_2	0.61~0.69	0.65	0.60~0.68	0.62

Faced with three driving styles in lane-changing experiments, 10 volunteers performed lane-changing driving experiments respectively. The experimental results show that different driving styles of the obstacle vehicle led to different interaction outcomes for the main vehicle, resulting in various lane-changing behaviours. According to the lane-changing trajectories of the main vehicle under the influence of the three driving styles, it can be seen that the lane-changing trajectories account for risk aversion. The aggressive driving style prioritizes efficiency and reduces the comfort and safety weighting in the cost function. In contrast, obstacle vehicles with conservative driving styles focus more on safety and driving comfort, leading to interaction lane-change results where the primary vehicle has faster lane-change acceleration. The interaction results of the obstacle vehicle with a normal driving style fall between the aggressive and conservative types.

As shown in the experimental results, the host vehicle can achieve make safe and efficient lane-changing decisions under the influence of the different social behaviours of surrounding vehicles. Moreover, the game algorithm, which

includes the human-like driving decision cost function, can realize human-like trajectories in lane-changing decisions.

6. CONCLUSIONS

In this paper, we set out to find a driving decision algorithm that is more human-like and applicable to highly dynamic, multi-participant interaction highway lane-changing scenarios. The method is validated and analysed for effectiveness through simulation. We use the inverse reinforcement learning algorithm to construct the cost function for human-like decision-making and combine the safety, comfort, and other basic cost functions to form a comprehensive decision in a lane change decision game. The Stackelberg game approach is employed to address the non-cooperative decision-making challenges posed by varying driving styles in social vehicle interactions. The algorithm demonstrates strong effectiveness and robustness in handling different interaction vehicle styles.

REFERENCES

[1] W. Wang, L. Wang, C. Zhang, C. Liu, and L. Sun, 'Social Interactions for Autonomous Driving: A Review and Perspectives', *FNT in Robotics*, vol. 10, no. 3–4, pp. 198–376, 2022, doi: 10.1561/2300000078.

[2] C. Dong, J. M. Dolan, and B. Litkouhi, 'Intention estimation for ramp merging control in autonomous driving', in *2017 IEEE Intelligent Vehicles Symposium (IV)*, Los Angeles, CA, USA: IEEE, Jun. 2017, pp. 1584–1589. doi: 10.1109/IVS.2017.7995935.

[3] E. Galceran, A. G. Cunningham, R. M. Eustice, and E. Olson, 'Multipolicy decision-making for autonomous driving via changepoint-based behavior prediction: Theory and experiment', *Auton Robot*, vol. 41, no. 6, pp. 1367–1382, Aug. 2017, doi: 10.1007/s10514-017-9619-z.

[4] C. L. Baker and J. B. Tenenbaum, 'Modeling Human Plan Recognition Using Bayesian Theory of Mind', in *Plan, Activity, and Intent Recognition*, Elsevier, 2014, pp. 177–204. doi: 10.1016/B978-0-12-398532-3.00007-5.

[5] J. Wiest, M. Hoffken, U. Kresel, and K. Dietmayer, 'Probabilistic trajectory prediction with Gaussian mixture models', in *2012 IEEE Intelligent Vehicles Symposium*, Alcal de Henares , Madrid, Spain: IEEE, Jun. 2012, pp. 141–146. doi: 10.1109/IVS.2012.6232277.

[6] L. Hou, L. Xin, S. E. Li, B. Cheng, and W. Wang, 'Interactive Trajectory Prediction of Surrounding Road Users for Autonomous Driving Using Structural-LSTM Network', *IEEE Trans. Intell. Transport. Syst.*, vol. 21, no. 11, pp. 4615–4625, Nov. 2020, doi: 10.1109/TITS.2019.2942089.

[7] Z. Huang, J. Wang, L. Pi, X. Song, and L. Yang, 'LSTM based trajectory prediction model for cyclist utilizing multiple interactions with environment', *Pattern Recognition*, vol. 112, p. 107800, Apr. 2021, doi: 10.1016/j.patcog.2020.107800.

[8] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, 'Proximal Policy Optimization Algorithms', Aug. 28, 2017, *arXiv*: arXiv:1707.06347. Accessed: Jun. 13, 2024. [Online]. Available: http://arxiv.org/abs/1707.06347

[9] F. Ye, X. Cheng, P. Wang, C.-Y. Chan, and J. Zhang, 'Automated Lane Change Strategy using Proximal Policy Optimization-based Deep Reinforcement Learning', May 20, 2020, *arXiv*: arXiv:2002.02667. Accessed: Jun. 13, 2024. [Online]. Available: http://arxiv.org/abs/2002.02667

[10] Y. Shi, Y. Liu, Y. Qi, and Q. Han, 'A Control Method with Reinforcement Learning for Urban Un-Signalized Intersection in Hybrid Traffic Environment', *Sensors*, vol. 22, no. 3, p. 779, Jan. 2022, doi: 10.3390/s22030779.

[11] R. Trumpp, H. Bayerlein, and D. Gesbert, 'Modeling Interactions of Autonomous Vehicles and Pedestrians with Deep Multi-Agent Reinforcement Learning for Collision Avoidance', in *2022 IEEE Intelligent Vehicles Symposium (IV)*, Aachen, Germany: IEEE, Jun. 2022, pp. 331–336. doi: 10.1109/IV51971.2022.9827451.

[12] S. Brechtel, T. Gindele, and R. Dillmann, 'Probabilistic MDP-behavior planning for cars', in *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, Washington, DC, USA: IEEE, Oct. 2011, pp. 1537–1542. doi: 10.1109/ITSC.2011.6082928.

[13] C. Hubmann, M. Becker, D. Althoff, D. Lenz, and C. Stiller, 'Decision making for autonomous driving considering interaction and uncertain prediction of surrounding vehicles', in *2017 IEEE Intelligent Vehicles Symposium (IV)*, Los Angeles, CA, USA: IEEE, Jun. 2017, pp. 1671–1678. doi: 10.1109/IVS.2017.7995949.

[14] H. Yu, H. E. Tseng, and R. Langari, 'A human-like game theory-based controller for automatic lane changing', *Transportation Research Part C: Emerging Technologies*, vol. 88, pp. 140–158, Mar. 2018, doi: 10.1016/j.trc.2018.01.016.

[15] L. Li, W. Zhao, and C. Wang, 'POMDP Motion Planning Algorithm Based on Multi-Modal Driving Intention', *IEEE Trans. Intell. Veh.*, vol. 8, no. 2, pp. 1777–1786, Feb. 2023, doi: 10.1109/TIV.2022.3209926.

[16] Z. Zhu and H. Zhao, 'Learning Autonomous Control Policy for Intersection Navigation With Pedestrian Interaction', *IEEE Trans. Intell. Veh.*, vol. 8, no. 5, pp. 3270–3284, May 2023, doi: 10.1109/TIV.2023.3256972.

[17] Z. Huang, J. Wu, and C. Lv, 'Driving Behavior Modeling using Naturalistic Human Driving Data with Inverse Reinforcement Learning', Jul. 19, 2021, *arXiv*: arXiv:2010.03118. Accessed: Jun. 13, 2024. [Online]. Available: http://arxiv.org/abs/2010.03118

[18] X. Wen, S. Jian, and D. He, 'Modeling the Effects of Autonomous Vehicles on Human Driver Car-Following Behaviors Using Inverse Reinforcement Learning', *IEEE Trans. Intell. Transport. Syst.*, vol. 24, no. 12, pp. 13903–13915, Dec. 2023, doi: 10.1109/TITS.2023.3298150.

[19] X. Di and R. Shi, 'A survey on autonomous vehicle control in the era of mixed-autonomy: From physics-based to AI-guided driving policy learning', *Transportation Research Part C: Emerging Technologies*, vol. 125, p. 103008, Apr. 2021, doi: 10.1016/j.trc.2021.103008.

[20] J. Yoo and R. Langari, 'A Game-Theoretic Model of Human Driving and Application to Discretionary Lane-Changes'.

[21] H. Yu, H. E. Tseng, and R. Langari, 'A human-like game theory-based controller for automatic lane changing', *Transportation Research Part C: Emerging Technologies*, vol. 88, pp. 140–158, Mar. 2018, doi: 10.1016/j.trc.2018.01.016.

[22] A. Talebpour, H. S. Mahmassani, and S. H. Hamdar, 'Modeling Lane-Changing Behavior in a Connected Environment: A Game Theory Approach', *Transportation Research Procedia*, vol. 7, pp. 420–440, 2015, doi: 10.1016/j.trpro.2015.06.022.

[23] Q. Zhang, R. Langari, H. E. Tseng, D. Filev, S. Szwabowski, and S. Coskun, 'A Game Theoretic Model Predictive Controller With Aggressiveness Estimation for Mandatory Lane Change', *IEEE Trans. Intell. Veh.*, vol. 5, no. 1, pp. 75–89, Mar. 2020, doi: 10.1109/TIV.2019.2955367.

[24] G. S. Sankar and K. Han, 'Adaptive Robust Game-Theoretic Decision Making Strategy for Autonomous Vehicles in Highway', *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 14484–14493, Dec. 2020, doi: 10.1109/TVT.2020.3041152.

[25] N. Li, D. W. Oyler, M. Zhang, Y. Yildiz, I. Kolmanovsky, and A. R. Girard, 'Game Theoretic Modeling of Driver and Vehicle Interactions for Verification and Validation of Autonomous Vehicle Control Systems', *IEEE Trans. Contr. Syst. Technol.*, vol. 26, no. 5, pp. 1782–1797, Sep. 2018, doi: 10.1109/TCST.2017.2723574.