View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by INRIA a CCSD electronic archive server

Probabilistic Decision-Making at Road Intersections: Formulation and Quantitative Evaluation

Mathieu Barbier, Christian Laugier, Olivier Simonin, Javier Ibañez-Guzmán

To cite this version:

Mathieu Barbier, Christian Laugier, Olivier Simonin, Javier Ibañez-Guzmán. Probabilistic Decision-Making at Road Intersections: Formulation and Quantitative Evaluation. ICARCV 2018 - 15 th International Conference on Control, Automation, Robotics and Vision, Nov 2018, Singapour, Singapore. pp.1-8. hal-01940392

HAL Id: hal-01940392

https://hal.inria.fr/hal-01940392

Submitted on 30 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Probabilistic Decision-Making at Road Intersections: Formulation and Quantitative Evaluation

Mathieu Barbier[1,2], Christian Laugier[2], Olivier Simonin[2,3] and Javier Ibañez-Guzmán[1]

***Abstract*— As drivers approach a road intersection, they must decide whether to cross it or to come to a stop. For this purpose, drivers make a situation assessment and adapt their behaviour accordingly. When this task is performed by a computer, the available information is partial and uncertain. Any decision requires the system to use this information as well as taking into account the behaviour of other drivers to avoid collisions. However, metrics such as collision rate can remain low in an interactive environment because of other driver's actions. Consequently, evaluation metrics must depend on other driving aspects. In this paper a decision-making mechanism and metrics to evaluate such a system at road intersection crossing are presented. For the former, a Partially Observable Markov Decision Process is used to model the system with respect to uncertainties in the behaviour of other drivers. For the latter, different key performance indicators are defined to evaluate the resulting behaviour of the system with different configurations and scenarios. The approach is demonstrated within an automotive grade simulator. It has showed at times, that whilst the vehicle can cross safely the intersection, it might not satisfy other key performance indicators related to highway code.**

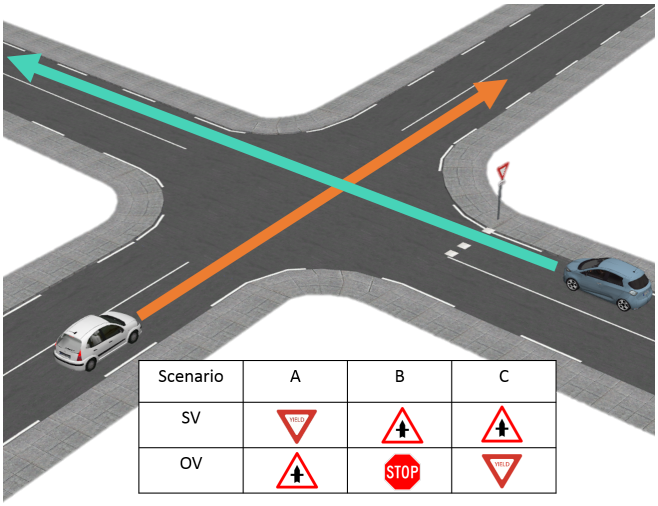


Fig. 1. Crossroad intersection representation including possible scenarios and the main actors: Subject Vehicle (SV) on the right side, Other Vehicle (OV) on the left side.

I. Introduction

Road intersections are likely the most complex segment in a road network. From an accidentology perspective, most major accidents occur at intersections, mainly due to human errors [1]. The failure to fully understand the situations encountered are the main reasons, particularly by young and elderly drivers [2].

Situation understanding is a major challenge, whether the vehicle is under manual or computer control. Thus, the understanding of situations and decisions taken are affected by the uncertainty associated to the digital representation of the environment. Further, interactions between drivers are done through a non-verbal exchange. For an autonomous vehicle this represents a major challenge. Given the uncertainties associated to the perception of the world by the vehicle on-board sensors and the unpredicted behaviour of other road network actors, approaches based on probabilistic methods are preferred to formulate this problem.

With non-deterministic methods, a particular attention should be paid to the evaluation process. It is not feasible to test every scenario on experimental grounds, hence simulators are used to test a larger variety of scenarios. In order to evaluate the system performance in a given scenario, a set of appropriate metrics has to be defined. Key performance indicators aim to evaluate performances and issues of decision-making for scenarios considering the highway code, navigation performances or passengers comfort. As the industry wants to deploy autonomous features in their vehicles, validation process is yet to be defined or updated in a rigorous manner.

This paper centres on the crossing of road intersections, these are considered complex but also where most accidents occur, in particular cross-cutting collisions. Fig.1 represents such scenarios, it includes the subject vehicle (SV) and the other vehicle (OV). Three types of scenarios are examined: Scenario A, where the SV is required to yield and Scenario B and C, where the SV has the right of way.

The purpose of the paper is twofold: first, a decision system is formulated that will control the subject vehicle acceleration as it approaches the intersection according to the situation and associated uncertainties. A partially observable Markov decision process and an online solver are used to find the action to be performed by the vehicle to react to the situation and future evolution. The state and observation spaces includes behavioural variables that represent three possible manoeuvres at an intersections: Pass, Stop and

[1]Renault S.A.S, 1 av. du Golf, 78288 Guyancourt, France. name.surmane@renault.com
[2]Inria Grenoble Rhône-Alpes , Chroma team, 655 Avenue de l'Europe, 38330 Montbonnot-Saint-Martin, France name.surmane@inria.fr
[3]INSA Lyon, CITI Lab., 6 avenue des Arts, 69680 Villeurbanne, France, France name.surmane@inria.fr

Yield. The reward function considers problems related to driver's behaviour, comfort and risk. Secondly, the formulation of key performance indicators (KPIs) is proposed to get a better understanding of the experiment results given the road intersection scenario.

Section II presents related work on decision-making for road intersection crossing with POMDPs and metrics used in such a problem. A description of the proposed decision making model is presented in Section III. Section IV describes the different Key performance indicators , the testing system and an analysis of the obtained results.

II. Related Work

Probabilistic and Bayesian approaches include uncertainty into their representation. Partially observable Markov decision process (POMDP) is a model that relies on a Bayesian network to represent dynamics of the system (with a special action variable) and a reward model to keep track of action values. Planning with POMDP models is well known and has been experimented for various robotic applications [3]. The complexity of problems, like autonomous driving, with large number of possible states, observations and actions affect the possibility to find the optimal policy in a acceptable time. Efforts on methods to keep track of the belief [4] and to solve POMDPs online [5, 6, 7], have made their usage attractive and effective in the last couple of years. Applications with online solvers show that real life use is possible [8, 9].

The state space of a POMDP represents all the variable required to represent the situation. At road intersection crossing, the state space is composed of variables for vehicle pose, speed and behaviour. The behavioural variables can represent, for example, the motion intention of drivers separated between stopping, hesitating, normal and aggressive [9]. These were inferred from a previously learned context. The behaviour variable is modelled to either constant velocity or constant acceleration that are then used to update the pose of vehicles[10]. However, these behavioural variables are not used to calculate the reward. Drivers intention enforced by regulation (traffic signs) in place is not taken into account in referred works. A coarse discretization is often used to represent the pose and velocity of vehicles as it simplifies the problem without a performance loss. However, in [10, 11] continuous representation is used but required an exta-discretization step learned or computed online.

The observation model, that represents the uncertainty of system to observe the situation, mostly takes into account measurement noises. It can also include a more complex analysis with, for example, occluded obstacles [11]. Therefore, the situation understanding is done within the POMDP, but in a more complex system this function can be done by other systems (especially for behaviour inference) such as [12, 13].

The POMDP reward model promotes states that are desirable for the system. Thus, states after the intersection are highly rewarded and penalties are given to collision states [8, 11]. However, these long-term reward systems are not sufficient for decision making. It can be enhanced by penalizing changes of acceleration to guarantee the comfort of passengers or deviations from a reference speed [8, 9]. With recent improvement in deep reinforcement learning, such reward functions could be constructed from human demonstration or an annotated data-set [14].

In the approaches previously described, the behaviour model for the other vehicle assumed that the other driver is moving as expected from the situation. It is different from the aggressive behaviour defined for vehicle driving above a reference speed [9] but rather a driver that attempts a manoeuvre (for example to yield) when it did not concord with what is imposed by the situation. This idea to divide the behaviour between intention and expectation has been proposed in [15], where vehicle to vehicle communications are used by the SV to observe the scene and to estimate the level of risk based this difference. For this purpose, a Bayesian network is used to estimate probabilities over the possible states of the system but only for stop and go type of behaviour.

Most of referred works used simulation tools to test their approach and even before experimentation on close roads a simulator has been used to obtain preliminary results [9]. Simulation is convenient for that type of approach for two reasons: firstly, solving online a POMDP can be non-deterministic, thus the same scenario requires to be tested multiple times to ensure stability. Secondly, scenario variations are easier to test within a simulation and are also less costly to set up. To make sense of this large amount of results, some metrics are used in order to find successful tests. Selected metrics are often chosen arbitrarily and reflect the performances of the system and not its failures. Most of the times, authors looked at collision rate and travel time to gauge their approach. However, these do not take into account scenario specific metrics that could sometimes qualified the test a failure. With our proposed KPIs, we aim to have a more scenario driven approach and find what are the important performance indicators that could be used to judge a system.

III. POMDP Model Applied to Road Intersections

The underlying components of the model used in the decision making system are detailed in this section.

A. A Preliminary on POMDP

A Partially Observable Markov Decision Process models the decision process of an agent acting in an uncertain environment. Formally it is composed of $\{\mathcal{X}, \mathcal{A}, \mathcal{Z}, T, R, O\}$, with $\mathcal{X}$ the state space, $\mathcal{A}$ a set of actions that the agent can take, $\mathcal{Z}$ a set of observations that can be obtained by the agent. T is the transition function $T(x', x, a) = P(x'|x, a)$ describes how the system changes when the agent takes action a when in state x with x' the state at the next time step. The reward function $R(x, a)$ indicates the value obtained after performing an action in a given state. And $O(z, x, a) = P(z|x, a)$ represent the probability to obtain an

observation $z \in \mathcal{Z}$ given a state and an action. In an uncertain environment, the agent does not know the real state after an observation, it reasons with a belief $b \in \mathcal{B}$ with $b : \mathcal{B} \rightarrow \mathbb{R}_{\geq 0}$ and $\int_{x \in \mathcal{X}} b(x)dx = 1$. Thus, the goal of the agent is to maximize the value $V : \mathcal{B} \rightarrow \mathbb{R}$ for an initial belief. In a POMDP there exists an optimal policy $\pi^* : \mathcal{B} \rightarrow \mathcal{A}$ that maximize V^π. The value of a policy can be estimated as the expected future sum of rewards for an initial belief $b_o$is $V^\pi(b_0) = \sum_{t=0}^{\infty} \gamma^t R(b_t, a) = \sum_{t=0}^{\infty} \gamma^t E(R(x, a)|b_0, \pi)$. γ, the discount factor, penalized rewards more and more after each time step t. For the purpose of our approach the agent is our decision-making system that decides which acceleration is appropriate while approaching towards the intersection.

B. State Space

The state space contains all the information required to model the system. Our state vector is $x = [d_{sv}, d_{ov}, s_{sv}, s_{ov}, i_{ov}, e_{sv}, e_{ov}]$. With, the physical state of the subject vehicle is represented with its speed s_{sv} and distance towards the intersection d_{sv}. Similarly d_{ov}, s_{ov} represent the physical state the of the other vehicle. Since small variations on these variables are hard to perceive and do not add meaningful information, range is $d_{sv}, d_{ov}, s_{sv}, s_{ov} \in \mathbb{N}$. Distances are taken from 0 m that correspond to the entrance to the intersection to 50m that represents the point at which drivers start to be aware of the approaching intersection. The considered vehicle velocity ranges from 0 m/s to 14 m/s (50 km/h), the maximum legal velocity in urban areas.

For the behavioural variables, a difference is made between intention of the i_{ov} and expectation e_{ov} as in [15]. The longitudinal expectation e_{ov}, represents what a driver should do according to priority rules at the intersection and the situation. In a perfect world, drivers should always follow the expected behaviour to avoid risky situations. However, there could be differences caused by misinterpretation of the scene or if the SV or road users are occluded. As a consequence, another variable, the longitudinal intention i_{ov} represents what the driver is doing approaching the intersection. The SV, assumed driven by an autonomous system, will always comply with the current situation. Thus, the behaviour of the SV is only represented by its expectation e_{sv}.

The behavioural variables represent high level manoeuvres that drivers can do when approaching an intersection and are defined on the same ensemble $i_{ov}, e_{ov}, e_{sv} \in \{Stop, Yield, Pass\}$. $Stops$ means that a driver will brake to stop at the intersection. $Yield$ is a manoeuvre when drivers regulate their speed in order to let another vehicle to cross an intersection. And the $Pass$ manoeuvre corresponds to a behaviour when a driver intends to cross first the intersection.Our State space differs from [15] with the introduction of the $Yield$ manoeuvre, that will allow the selection of smaller deceleration while the SV approach the intersection, and to evaluate the reward on states with traffics laws (see section III-E).

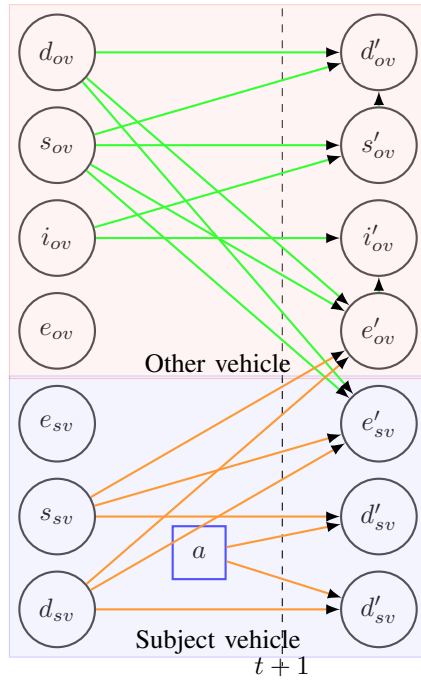


Fig. 2. The transition function represented as Bayesian network. The square node represent the action chosen by the framework

C. Action Space

The action space contains a range of accelerations that the SV can do to change its state. Approaching an intersection, a vehicle is most likely to decelerate. Slow deceleration helps the vehicle to widen the gap with another vehicle. High deceleration helps to avoid collision or to adapt the speed approaching towards the intersection entrance. Small accelerations help to reduce the travel time. From these observations, actions were chosen as follow: $a \in \{-2, -1.5, -1, -0.5, 0, 1\} ms^{-2}$. A larger range of accelerations would increase model complexity. The proposed action range is sufficient to stop the vehicle to avoid collision or to come to a stop. The step between each action is also small enough to avoid jerk.

D. Transition Model

This model represents the probability of going to a state x' given a previous state x and an action a. The structure of this model can be represented by the Bayesian network illustrated in Fig. 2. The action performed by the SV has the most impact on s_{sv} and d_{sv}. Kinematic equations are used to update the speed and the distance to the intersection. However, the vehicle might not be able to exactly perform the action. Thus, the transition probabilities are

$$\begin{aligned} P(s'_{sv}|a, s_{sv}) &= \mathcal{N}(\mu_s, \sigma_a) \\ With \quad \mu_s &= s_{sv} + a\Delta t \end{aligned} \tag{1}$$

$$\begin{aligned} P(d'_{sv}|a, d_{sv}, s_{sv}) = \mathcal{N}(\mu_d, \sigma_a) \\ With \quad \mu_d = d_{sv} - (s_{sv}\Delta t + a0.5\Delta t^2) \end{aligned} \tag{2}$$

With $\Delta t = 0.5s$ the time when acceleration is applied. The execution of the action could be delayed or altered due to the control systems, thus $\mathcal{N}(\mu_s, \sigma_a)$ represents this uncertainty with $\sigma_a = 1$. A gap acceptance model [16] is used to find the transition of expectation for the OV and SV. It uses the time gap between the two vehicles, to calculate the probability for the vehicle to come to a stop. It requires the distance and speed from each vehicle. Consequently expectation of both vehicle will change with the subject vehicle actions and the other vehicle states.

Equation 3 is used for the transition of the OV intention, if the previous intention matches the current expectation, we assumed that the driver will not change its intention (thus the probability of 0.9 in eq.3). Otherwise, the intention of the OV is unknown, thus uniformly distributed over all possible intentions.

$$P(i_{ov}^{'}|e_{ov}, i_{ov}) \begin{cases} P(i_{ov}^{'} = i_{ov}) = 0.9 & i_{ov} = e_{ov} \\ P(i_{ov}^{'}) = \mathcal{U}(\mathcal{I}) & otherwise \end{cases} \tag{3}$$

Then the speed of the OV is inferred from the previous speed and previous intention. Intentions to stop or to yield will decelerate induce a deceleration, whereas the pass manoeuvre correspond correponds to a null acceleration. The distance from intersection entrance depends on the speed and the previous speed as for the SV.

E. Reward Function

The reward function is a weighted sum of 5 components:

- Comfort reward (r_c) aims to reduce strong deceleration. Thus, $r_c(a = -2m/s^2) = -R_{max}$ for other actions no penalty is applied
- Risk reward (r_r in eq.4) penalizes small time gaps between two vehicles. As the distance has for origin an entrance of the intersection, this gap represents the remaining time before the two vehicles reach their respective entrance at the same time.

$$r_r(s_{sv}, d_{sv}, s_{ov}, d_{ov}) \begin{cases} R_{max} & |\frac{d_{sv}}{s_{sv}} - \frac{d_{ov}}{s_{ov}} - 5s| > 0 \\ |\frac{d_{sv}}{s_{sv}} - \frac{d_{ov}}{s_{ov}}| & otherwise \end{cases} \tag{4}$$

- Speed reward (r_s in eq.5) is calculated as the difference between the speed of the SV and a reference speed and if the action brings the speed toward the reference. This reference is obtained with a vehicle driving in the same intersection but without another vehicle.

$$r_s(s_{sv}, d_{sv}, a) \begin{cases} R_{max} & |V_{ref}(d_{sv}) - s_{sv}| < 2 \\ R_{max}/2 & V_{ref}(d_{sv}) - s_{sv} > 2 \wedge a > 0 \\ R_{max}/2 & V_{ref}(d_{sv}) - s_{sv} < -2 \wedge a < 0 \\ 0 & otherwise \end{cases} \tag{5}$$

- Intention reward (r_i) penalizes the situations when the intention of vehicles is in contradiction regarding the traffic laws. The relation between the intention of each vehicle is shown in table I.

TABLE I
REWARDS RELATED TO THE BEHAVIOUR OF DRIVERS WITH RESPECT TO TRAFFIC LAWS

e_{sv}/i_{ov}	Stop	Yield	Pass
Stop	0	0	R_{max}/2
Yield	R_{max}/2	0	R_{max}/2
Pass	R_{max}	R_{max}/2	—R_{max}

- Expectation reward (r_e in eq.6) penalizes states where the expectation of the OV is different from its intention (similar to the notion of risk introduced in [15]).

$$r_e(e_{ov}, i_{ov}) \begin{cases} R_{max}/2 & e_{ov} = i_{ov} \\ -R_{max} & otherwise \end{cases} \tag{6}$$

R_{max} is the maximum obtainable reward and is set to 10. Consequently, any component has a higher importance than another. The table I represents the function r_i that has been designed to reward states where the behaviour of OV and SV are compatible. For example, if one vehicle intends to stop/yield and the other to pass the situation is rewarded because collisions will be avoided and travel time optimized. Situations where both vehicles have the same intention are not rewarded since both vehicles is trying to let the other cross the intersection. If both of them intend to pass a collision could happen, thus these situations are penalized.

The last two components are important for the cooperation between two vehicles. If r_i is low, the SV will try actions that modify its expectation in order to cooperate. If r_e is low the SV can modify its speed to change the expectation of the OV.

The resulting reward function is:

$$\begin{aligned} R(x, a) = {} & w_c(d_{sv}) r_c(a) + w_r(d_{sv}) r_r(s_{sv}, d_{sv}, s_{ov}, d_{ov}) \\ & + w_e(d_{sv}) r_e(e_{ov}, i_{ov}) + w_s(d_{sv}) r_s(s_{sv}, d_{sv} a) \\ & + w_i(d_{sv}) r_i(e_{sv}, i_{ov}) \end{aligned} \tag{7}$$

Functions w are used to weight each reward component. They can be changed to adapt the behaviour of the system with a scenario or the driving style of the SV. These functions are dependent of the distance towards the intersection of the SV. For example, the risk of collision could be less important far away from the intersection and one would prefer to prioritize the comfort of the passengers. However, close to the intersection it is important to avoid collisions even if it requires strong braking. Linear functions, like $w(d_{sv}) = k_1 d_{sv} + k_2$, are used to weight each reward. Different values of k_1 and k_2 will be tested in order to find appropriate values that guarantee safety of vehicles and possible ranges for personification. Compared with a more conventional approach (assign a high reward to terminal states and penalties for failures), this approach can take into account more subtle aspects of driving such as the one involving the behaviour of vehicles.

F. Observation Space and Model

States of the subject vehicle are observed with proprioceptive sensors less subject to uncertainty than the measurement of the OV that comes from perception systems. Thus, the noise level is smaller for the SV than for the OV. This information is added to observation model to take into account measurement uncertainties as Gaussian noises on the states of each vehicle $\mathcal{N}(0, 0.5)$ for the SV and $\mathcal{N}(0, 1)$ for the OV). For the intention of the OV, in our previous work [13] we trained classifiers to recognize the intention of vehicle approaching an intersection. This framework is used to observe the intention of the other vehicle, whereas, [15] assumed that driver's intentions were hidden. The observation model for this variable is deduced from classification matrices of the classifier. The resulting observation vector is $z = [\hat{s}_{sv}, \hat{s}_{ov}, \hat{d}_{sv}, \hat{d}_{ov}, \hat{i}_{ov}]$ and variables are defined on the same ensemble as the state space.

G. Solver

The large number of possible states and observations made the proposed model hard to solve with an offline solver. Thus, an online solver has been chosen that provide an approximation of the optimal policy at anytime. The advantage of such a solution is that an evaluation of the value of an action is always available but might not be perfectly estimated. The POMCP [6] has been chosen for its simplicity and its scalability. This solver combines a search tree with a particle filter. A belief is represented as a density of particles. The filter uses the transition model, in section III-D, to simulate the transition of a single particle when an action is chosen. The tree is used to keep track of the history and to direct the exploration which is controlled by an Upper confidence bound (UCB1) algorithm. It finds a trade-off between exploration and exploitation (focusing the search on the best performing branch). The next action a to be explored is selected by $max_a V^{\oplus}(ha) = V(ha) + C\sqrt{\frac{log(N(h)}{N(ha)}}$with $N(.)$ being the number of times a node has been explored and h the history of observation and action that lead to the current node. Finally, C is used to control the exploration. With $C = 0$ the action that had the maximum value is always selected and with $C >> 0$ the action space is uniformly explored. With our implementation, γ=0.85 and a search time of 0.5s, an average of 1400 particles can be sampled from the current belief and simulated until the discounted reward becomes negligible. It is sufficient for studied scenarios, but to scale up the approach and include more vehicles or pedestrians, a better implementation using parallel processing or a learned estimation of $V(ha)$ is required.

IV. Quantitative Evaluation

The following section presents the key performances indicators alongside with the testing system. Then, an evaluation of two configurations of the proposed decision-making and a baseline algorithm shows the importance to use the proposed metrics to understand system performances.

A. Evaluation approach: Key Performance Indicators

The evaluation of the proposed decision-making system requires that vehicles within the test environment react to each other actions, consequently such vehicles are also able to avoid collisions with the SV. Thus, few collisions were observed even if some undesired behaviour were sometimes observed. It motivated us to propose some key performance indicators to evaluate a decision-making system in relation with the tested scenario. As explained in [17], it is a disputable designed choose to implement the metrics used to judge a system in its rewards function, as the system would only optimise the metrics. It is the reason why KPIs were not implemented in the reward function.

For an autonomous driving decision-making system, they can be divided in 4 categories: safety, navigation, trust and comfort. In this section, we propose some KPIs for each category with respect to the road intersection crossing context.

1) Safety: At cross road intersection crossing, it is unsafe and forbidden by the highway code to stop within the intersection. Thus, a first KPI can be formulated as the time the SV spent stopped within the intersection. Tests where this KPI is not null result in a failure.

2) Navigation: This category highlights the performances of the system to adapt its behaviours regarding the situation. First the travel time is looked at. A slow vehicle might be safe but will reduce the traffic flow. In the context of our experimentation, a vehicle driving at $8ms^{-1}$ covers the distance necessary to cross the intersection in $6.5s$. Taking into account that deceleration is required at road intersection, a travel time lower than 15s is considered a maximum for scenario B (SV has the priority). For Scenario A, a travel time lower than $20s$ is a success because the SV vehicle is required to decelerate more. In both scenario the SV is allowed to come to a stop before the intersection if the situation required it. The time stopped before entering the intersection is used as a second KPI for the navigation category. The best possible outcome for both scenarios is that the SV has adapted its speed to negotiate and did not come to a stop. However, if the SV stopped, it should not wait for a long time (in scenario A the waiting time should be enough to let the OV cross). Thresholds for this category could be learned from data obtained while driving.

3) Trust: Drivers tend to maintain a certain time gap between them and other obstacles. From [18], a gap of 4 seconds is a minimum not to frighten on-board passengers. With a smaller gap the trust of the driver in the system might be reduced and lead to dangerous system disengagements. At road intersections the gap between the two vehicles can be small during the approach, however when the SV enter the intersection this gap is required to be sufficient enough or non-existent (the OV has already crossed the intersection). Therefore, this KPI is formulated as the time gap when the SV enters the intersection. A successful test has the SV let the other vehicle pass first or the SV entered the intersection with a gap higher than 4 seconds.

TABLE II

CONDITION OF SUCCESS USING PROPOSED KPIS

	Scenario A					Scenario B and C				
KPI	Longitudinal jerk	Gap	Safe stop	Unsafe stop	Travel time	Longitudinal jerk	Gap	Safe stop	Unsafe stop	Travel time
Successful	Lower than $2ms^{-3}$	Over 4s	None	None	-	Lower than $2ms^{-3}$	Over 4s	None	None	-
Acceptable	-	-	Lower than $3s$	None	-	-	-	Lower than $5s$	None	-
Failed	Over than $2ms^{-3}$	Lower than 4s	Over 3s	Over 0s	Over 20s	Over than $2ms^{-3}$	Lower than 4s	Over 5s	Over 0s	Over 15s

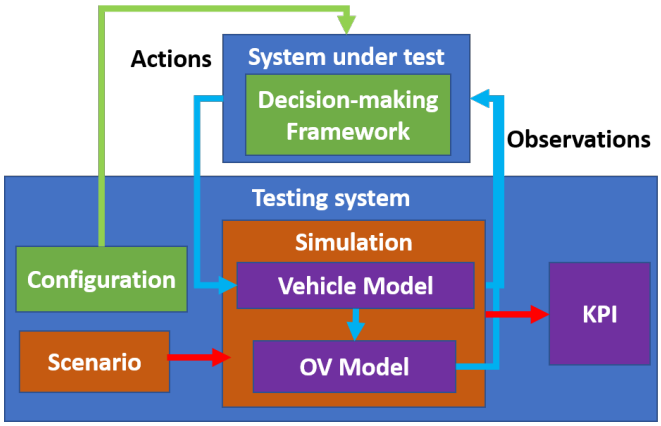


Fig. 3. Experimental setup, the bottom part represents elements of the testing system, and the top one the decision-making framework. The simulated environment is modified by selected actions and send back observations of the situation.

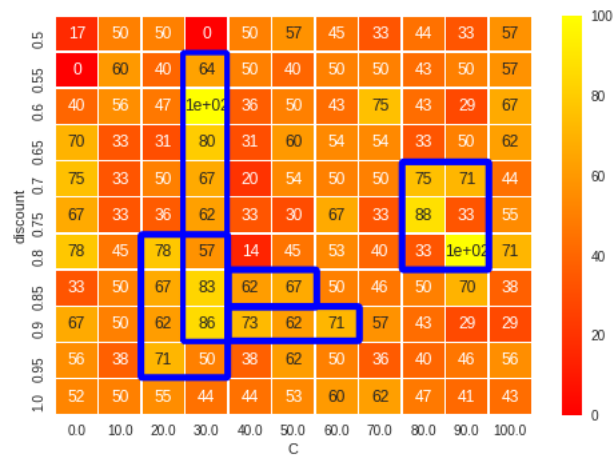


Fig. 4. Percentage of successful execution for different discount values and exploration coefficients. The area within the blue border represent area with high potential of success

4) Comfort: As the velocity of a vehicle is changed approaching towards the intersection, sudden changes of acceleration (jerk) are undesirable. From [19], passengers felt uncomfortable if the jerk is higher than $2ms^{-3}$. The last KPI is the mean jerk and should not be higher than $2ms^{-3}$.

Success conditions for each scenario and for each KPI are summarized in table II.

B. Testing System

The experimental setup, as shown in Fig. 3, works as follows: first a set of configurations and scenarios is created. A configuration is the set of k values used to compute weight functions. Next a scenario is loaded into a simulator and the decision-making system starts with a selected configuration. Each test last for 20 seconds, enough time for the SV to cross the intersection. KPIs defined in section IV-A are used at the end of the chain to find configurations that were most likely successful. To generate the environment of testing, we used an automotive grade simulator (Scaner) interfaced with a desktop computer that runs ROS and the decision-making system. Within the simulated environment, the observations of every vehicle could be precisely obtained. Thus, Gaussian noises were added on the speed and distance measurement to simulate the performances of a perception under real world conditions. For the SV a noise $\mathcal{N}(0, 0.5)$ is added for the measurement and a noise of $\mathcal{N}(0, 1)$ for the OV vehicle.

The layout in Fig. 1 is used for each test, it corresponds to an existing intersection at our test site. Regulation is changed regarding the requested scenario. The other two scenario changed the regulation and required the OV to come to a stop (Scenario B) or to yield (Scenario C). In order to change the encounter at the intersection, the initial speed for each vehicle is randomly assigned. Thus, the moment when the two vehicles reach the intersection is different.

Best configurations were found with an iterative process, with random configurations tested in different scenarios. Then, the search range was narrowed to best performing configurations and tested again with fewer configuration parameters free. As of now the process of selecting configuration range with successful outcomes is done manually, but in the future methods such as [20] or reinforcement learning could be used to improve this optimization.

For the POMCP solver, a discount of 0.85 (in our setup a horizon of 6 seconds) and C=30 gave the best results. In the Fig. 4, it can be observed that tests with these values are likely to be successful.

C. Results and Analysis

While optimising the success rate of the decision-making system, it appeared that some configurations were better used in a specific scenario. To demonstrate the usage of KPI to explain and compare the system performances , two configurations were chosen with different area of strength. The configuration 1 performed well across each scenario and

TABLE III

VALUES TO WEIGHT EACH COMPONENT OF THE REWARD FUNCTION

Configuration	1		2	
	k1	k2	k1	k2
w_c	0.5	0.5	0	0.6
w_r	0.1	0.3	0.7	0.9
w_i	-0.5	1	-0.5	1.4
w_e	0.5	0.7	0.5	1.5
w_s	-0.5	1.5	0	0.9

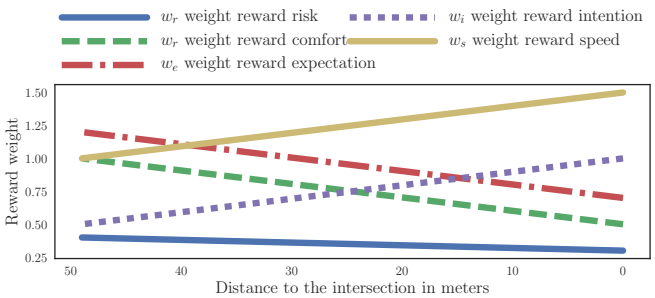


Fig. 5. Reward coefficient as function of the distance to the intersection for the configuration 1

the configuration 2 only on scenario A. The table III shows values selected for two configurations and the Fig. 5 the variation of these functions for the first configuration with respect to the distance to the intersection. For the baseline, we used the decision making within the simulator that is based upon used an Intelligent driver model to control the subject vehicle.

In the two configurations, none of the weighting function was equal to zero. Thus, each aspect of the driving implemented in the reward function is useful to the system. For the first configuration, the low weight value for the risk related w_r reward can be explained by the effect of rewards related to the intention w_i and expectation w_e. If they prevented the vehicle to enter risky situations, the importance of the risk can be reduced. However, its importance is almost constant during the manoeuvre. The high importance for the speed reward w_s shows that a key to successful executions is to follow a reference speed. However, it is balanced by a reduction of the comfort as can be seen with a decreasing importance of the comfort.

Fig.6a, as the SV slows down after three seconds it gave enough time for the OV to cross without changing its speed. The more cautious behaviour of the SV in scenario B is due to uncertainties whereas, in scenario A, the OV drivers are much faster as the simulation behaviour model is different.

In Fig. 6 that shows the velocity of the two vehicles and acceleration actions selected by the system in two different scenario. It can be observed that high deceleration actions are selected between distances of 10 and 0 meters. During this test, in Fig.6b, it can be observed that from t=0s to t=4s the SV kept its speed because the OV is still in an acceleration phase. While being close to the intersection (<20m), the SV starts to adapt its speed and to be ready to come to a stop in case the gap is not sufficient if the OV does not come to a stop as it could be expected in the scenario B. In

TABLE IV

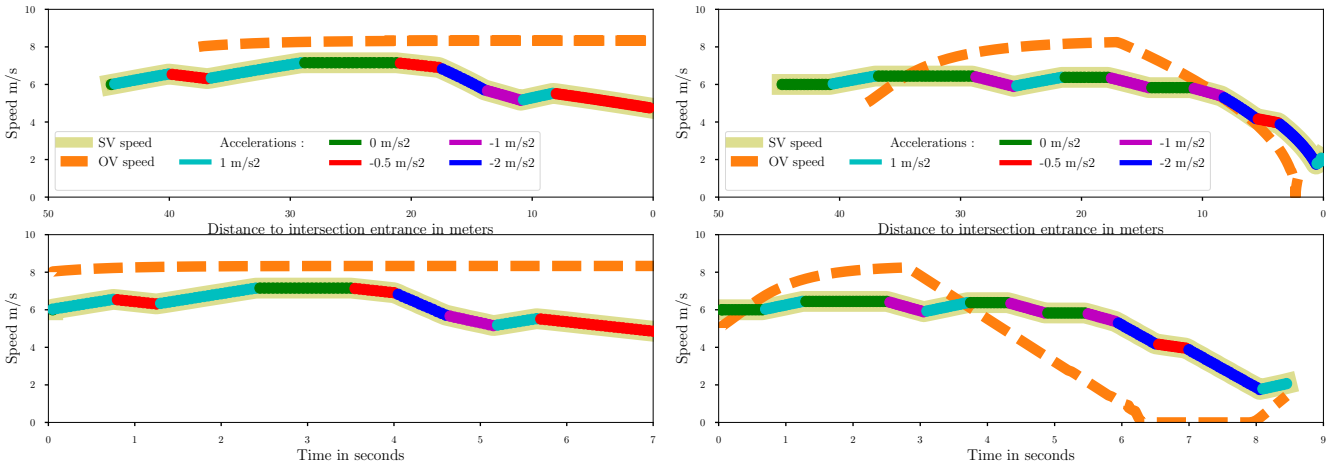
PERCENTAGE OF SUCCESSFUL TESTS AND REASON FOR FAILURE OUT 800 EXPERIMENTS FOR EACH CONFIGURATION

Metric	configuration	A	B	C
Success rate	1	75%	78%	86%
	2	82%	47%	56%
	baseline	2%	75	42%
Acceptable safe stop	1	0	4%	3%
	2	1%	8%	8%
	baseline	91%	0	6%
Failure Safe stop	1	0	0	0
	2	0	1%	1%
	baseline	0	0	0
Failure travel time	1	0	0	0
	2	0	25%	17%
	baseline	0	0	0
Failure jerk	1	3%	14%	8%
	2	4%	7%	8%
	baseline	0	0	0
Failure gap	1	19%	0	0
	2	11%	1%	0
	baseline	6%	25%	52%
Failure unsafe stop critic	1	3%	4%	3%
	2	2%	11%	10%
	baseline	1%	0	0

The table IV shows causes of failure found with the proposed KPIs. As for the baseline,the first configuration is that it is capable to cross the intersection at every test in every scenarios as for the baseline (no failure related to travel time). However, this is also the main causes of failure of the second configuration for scenario B and C. In the Scenario A, it appears that the baseline is most likely to stop at the giveaway intersection instead of slowing and pass as the proposed system is able to do. However, it is balanced with the gap related failure caused by our vehicle entering the intersection while the other vehicle is close. It can be observed that the failure related to the gap is less important for configuration 2 than configuration 1, but compare to the baseline in scenario B the POMDP is able to maintain a sufficient gap. In some of the test, with the POMDP the vehicle stopped within the intersection (Failure caused by unsafe stop critic), even if illegal, it might have helped the subject vehicle to react to the other vehicle unexpectedly crossing the intersection. From results obtained, it can be understood the area of improvement for each configuration are. The configuration 1 should reduce its tendency to stop within the intersection, whereas the second one should focus on reducing gap related failures in scenario A.

V. CONCLUSION AND PERSPECTIVE

In this paper, a POMDP based decision-making system has been proposed to solve road intersection problems taking into account the behaviour of another driver and associated uncertainties. Results of this approach have been evaluated with different key performances indicators that inform on system performances with respect to the road intersection crossing scenario. The evaluation demonstrated that the decision-making is able to cross the intersection, but the



(a) Scenario A with configuration 1

(b) Scenario B with configuration 1

Fig. 6. Evolution of the speed in function of distance (above) and time (below) Changes of acceleration are shown by the color of each segment.

analysis reveals that the behaviour does not conform with certain of the proposed criteria. Performances are better or at least equal to the baseline.

On-going works involve using the Statistical Model Checking [20] to automatically explore the configuration and scenario space. As such, KPIs will be formulated as Bounded Linear Temporal Logic (BLTL). From that, the bounded probability of occurrence of a KPI value could be found and helps with the validation of decision-making system.

For the proposed POMDP model, the scenario complexity needs to be increase to include pedestrians and more vehicles. Thus, the implementation of the online solver needs to be improve and use parallel computing and learning to be better estimate the value function. The model used for the dynamics of the vehicle and the observation of the physical could state can be improve to be closer to systems used in a vehicle.

The domain of validation of AI-based systems is an emerging topics that requires more attention [21] and the definition of key performance indicators is important to understand the behaviour of such system [17].

References

[1] "Bilan de l'accidentalite de l'annee 2014," tech. rep., Observatoire national interministeriel de la securite routiere, 2015.

[2] J. Ibañez-Guzmán, S. Lefevre, A. Mokkadem, and S. Rodhaim, "Vehicle to vehicle communications applied to road intersection safety, field results," in *2010 13thInternational IEEE Conference on Intelligent Transportation Systems (ITSC),*, pp. 192–197, Sept 2010.

[3] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial intelligence*, vol. 101, no. 1, pp. 99–134, 1998.

[4] S. Thrun, "Monte carlo POMDPs," in *Advances in Neural Information Processing Systems 12* (S. Solla, T. Leen, and K.-R. Müller, eds.), pp. 1064–1070, MIT Press, 2000.

[5] S. Ross, J. Pineau, S. Paquet, and B. Chaib-draa, "Online planning algorithms for pomdps," *J. Artif. Int. Res.*, vol. 32, pp. 663–704, July 2008.

[6] D. Silver and J. Veness, "Monte-carlo planning in large pomdps," in *Advances in Neural Information Processing Systems 23* (J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, eds.), pp. 2164–2172, Curran Associates, Inc., 2010.

[7] A. Somani, N. Ye, D. Hsu, and W. S. Lee, "Despot: Online pomdp planning with regularization," in *Advances in Neural Information Processing Systems 26* (C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, eds.), pp. 1772–1780, Curran Associates, Inc., 2013.

[8] C. Hubmann, J. Schulz, M. Becker, D. Althoff, and C. Stiller, "Automated driving in uncertain environments: Planning with interaction and uncertain maneuver prediction," *IEEE Transactions on Intelligent Vehicles*, vol. 3, pp. 5–17, March 2018.

[9] W. Liu, S. W. Kim, S. Pendleton, and M. H. Ang, "Situation-aware decision making for autonomous driving on urban road using online pomdp," in *2015 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1126–1133, June 2015.

[10] M. Bouton, A. Cosgun, and M. J. Kochenderfer, "Belief state planning for autonomously navigating urban intersections," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, pp. 825–830, June 2017.

[11] S. Brechtel, T. Gindele, and R. Dillmann, "Probabilistic decision-making under uncertainty for autonomous driving using continuous pomdps," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 392–399, Oct 2014.

[12] G. Aoude, V. Desaraju, L. Stephens, and J. How, "Driver behavior classification at intersections and validation on large naturalistic data set," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, pp. 724–736, June 2012.

[13] M. Barbier, C. Laugier, O. Simonin, and J. Ibañez-Guzmán, "Classification of drivers manoeuvre for road intersection crossing with synthethic and real data," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, pp. 224–230, June 2017.

[14] D. Isele, R. Rahimi, A. Cosgun, K. Subramanian, and K. Fujimura, "Navigating Occluded Intersections with Autonomous Vehicles using Deep Reinforcement Learning," *ArXiv e-prints*, May 2017.

[15] S. Lefevre, C. Laugier, and J. Ibañez-Guzmán, "Risk assessment at road intersections: Comparing intention and expectation," in *Intelligent Vehicles Symposium (IV), 2012 IEEE*, pp. 165–171, June 2012.

[16] M. A. Pollatschek, A. Polus, and M. Livneh, "A decision model for gap acceptance and capacity at intersections," *Transportation Research Part B: Methodological*, vol. 36, no. 7, pp. 649 – 663, 2002.

[17] D. Amodei, C. Olah, J. Steinhardt, P. F. Christiano, J. Schulman, and D. Mané, "Concrete problems in AI safety," *ArXiv e-prints*, 2016.

[18] K. Fitzpatrick, "Gaps accepted at stop-controlled intersections," *Transportation Research Record*, vol. 1303, no. 11, pp. 103–112, 1991.

[19] K. Yi and J. Chung, "Nonlinear brake control for vehicle cw/ca systems," *IEEE/ASME Transactions on Mechatronics*, vol. 6, pp. 17–25, Mar 2001.

[20] A. Legay, B. Delahaye, and S. Bensalem, "Statistical model checking: An overview," in *Runtime Verification*, (Berlin, Heidelberg), pp. 122–135, Springer Berlin Heidelberg, 2010.

[21] L. Li, Y.-L. Lin, N.-N. Zheng, F.-Y. Wang, Y. Liu, D. Cao, K. Wang, and W.-L. Huang, "Artificial intelligence test: a case study of intelligent vehicles," *Artificial Intelligence Review*, Apr 2018.