Calibration of Controlled Markov Chains for Predicting Pedestrian Crossing Behavior Using Multi-objective Genetic Algorithms

Jingyuan Wu[1], Johannes Ruenz[1], and Matthias Althoff[2]

***Abstract*— Pedestrian motion prediction is a core issue in assisted and automated driving and challenging to solve. In this work, controlled Markov chains are used for predicting pedestrian crossing behavior in urban environments with and without crosswalks. Intentions, such as crossing a road, are estimated by incorporating the probability of colliding with other traffic participants. On a public dataset, we calibrate the model parameters using genetic algorithms which we formulate as a multi-objective optimization problem. Rather than only minimizing the position deviation of the prediction, we also consider the classification performance for pedestrians' crossing intention. The conducted evaluation shows benefits of our approach: it achieves comparable intention recognition performance compared to a support vector machine, while additionally achieving accurate spatiotemporal predictions.**

I. Introduction

A. Motivation

Intelligent vehicles need to detect, assess, and react to dangerous situations [1]. Motion prediction of vulnerable road users, such as pedestrians in particular, is not only of utmost importance for safety in assisted and automated driving, but also a key for natural and smooth maneuvers of intelligent vehicles [2].

This work focuses on pedestrians intending to cross in front of an approaching vehicle, cf. Fig. 1. For the prediction we use the controlled Markov chains (MC) presented in [3]. Compared to pure machine learning approaches, our approach not only works in arbitrary situations, but also enables expert knowledge to be considered and the effort of system inspection to be reduced. However, calibrating MC is not easy, since many parameters cannot be observed directly. By using algorithms with an automatic fitting procedure for the calibration, we believe that the objective functions play an important role which will be investigated in this work.

B. Related Work

a) Intention recognition: Pedestrian intention recognition in urban environments can be formulated as a classification problem, such as whether to cross in front of a vehicle, and inferred from meaningful features [4]–[8].

[1]Jingyuan Wu and Johannes Ruenz are with Robert Bosch GmbH, D-74232 Abstatt, Germany, jingyuan.wu@de.bosch.com and johannes.ruenz@de.bosch.com

[2]Matthias Althoff is with the Department of Computer Science, Technical University of Munich, D-85748 Garching, Germany, althoff@in.tum.de

This work has received funding from EU H2020 interACT: Designing cooperative interaction of automated vehicles with other road users in mixed traffic environments under grant agreement No 723395.

Fig. 1. A typical scene where a pedestrian intends to cross in front of an approaching vehicle (the image was recorded using a camera mounted on a testing vehicle of the EU-funded project interACT).

b) Motion prediction: Predicting the path of pedestrians can be combined with their intention recognition, where the estimated intention serves as an input for the path prediction [9], [10]. A deep-learning-based system using visual features can be found in [11].

c) Social force model: Besides, pedestrians' motion can be described as if they would be subject to "social forces" [12]. A social force model outputs spatiotemporal results and hence implicitly integrates path prediction and intention recognition. With respect to interaction with others, the avoidance mechanisms of human beings as a core of social force models are usually modeled in repulsive potential forms [12]–[16]. Social force models can be integrated into other frameworks. For instance, a long short-term memory network in [17] incorporates collision risks based on repulsive potentials for mixed traffic trajectory prediction in a shared space. The work in [18] presents a planning-based approach that accounts for local social interactions.

d) Incorporating map information: There are different ways to interpret the influence of map information on the behavior of pedestrians. A set-based prediction for pedestrians is used in [19] which incorporates constraints based on traffic rules. Markov decision processes are used in [20], [21], where the local motion patterns are the so-called policy to a goal. Out of the above approaches, the ones relying on planning-based techniques, require a prelocation of goals. This problem of goal forecasting and motion planning is jointly solved in [22] via one single artificial neural network. In addition, [3] presents a heuristic method to infer potential goals to a pedestrian based on a semantic map automatically.

e) Clustering-based prediction: Clustering techniques can be applied to derive motion patterns. The work in [23] presents a probabilistic hierarchical trajectory matching approach to perform trajectory predictions. An augmented semi-nonnegative sparse coding algorithm is proposed in [24]

with an extension in [25] by incorporating semantic features, such as geometric information of an intersection, to learn the transition between motion patterns of pedestrian trajectories.

f) Calibration of prediction models: For calibrating social force models, the relative distance error is used as objective function in [26], while [16] considers both the relative distance- and angle error and performs a bi-objective optimization. Maximum likelihood estimation is combined with a regularization term in [7], [17]. For training a classifier, metrics regarding the receiver operating characteristic curve can be found in [4]–[6].

C. Contribution

The above literature review revealed that a large share of previous work either used pure machine learning approaches or model-based approaches. Pure machine learning approaches require large datasets and often only work well for scenarios similar to those used for training. Model-based approaches, however, often do not sufficiently calibrate their models with real-world data. This work aims to close this gap by calibrating model-based approaches using genetic algorithms (GAs).

In particular, we propose a novel approach to calibrate a model for predicting pedestrian crossing behavior by using two objective functions from different viewpoints—spatiotemporal accuracy and intention classification accuracy. The previous work described above considers only objectives from one of those two viewpoints, which has limitations as we will show. Instead, we perform a multi-objective optimization using GAs and select a trade-off solution from the Pareto frontier. The usefulness of considering the intention classification performance as the second objective function during optimization is confirmed by an evaluation using the Daimler dataset [27].

The remainder of this paper is structured as follows. Sec. II introduces the framework of MC from our previous work [3] including extensions and refers to the model parameters to be calibrated. Sec. III handles the proposed multi-objective optimization procedure. The evaluation and discussion are found in Sec. IV. Finally, Sec. V concludes the paper.

II. Preliminaries

A. Controlled Markov chains

Controlled Markov chains are used in this work for motion prediction using the 2D position $x = (x_1, x_2)^T \in \mathbb{R}^2$ and the action $u = (\psi, v)^T$ consisting of the orientation $\psi \in \mathbb{R}_0^+$ and the speed $v \in \mathbb{R}_0^+$. The position cells are denoted by $\mathcal{X}_i \subset \mathbb{R}^2$ using Latin subscripts and the action cells by $\mathcal{U}^\alpha \subset \mathbb{R}^2$ using Greek superscripts, cf. Fig. 2. Let α_ψ and α_v be the separate indices for action intervals. Two separate indices determine a unique index of an action cell α and vice versa, e.g., $\alpha = n_\psi \cdot (\alpha_v - 1) + \alpha_\psi$ for n_ψ orientation intervals, cf. Fig. 2b.

The joint probabilities $P\left(x(t_k) \in \mathcal{X}_i, u(t_k) \in \mathcal{U}^\alpha\right)$ at points in time t_k are recursively estimated considering constraints. To this end, two steps are carried out in each time step: first, we compute the action transition probabilities

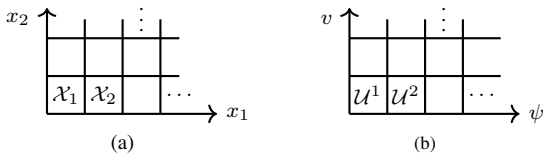

Fig. 2. (a) Position cells and (b) action cells [28].

$\Gamma_i^{\alpha\beta}(t_k) \coloneqq P\left(u(t_k) \in \mathcal{U}^\alpha \mid u(t_k) \in \mathcal{U}^\beta, x(t_k) \in \mathcal{X}_i\right)$, with which the action probability distribution is changed instantly at t_k; second, we propagate positions under the effect of actions according to the position transition probabilities $P\left(x(t_{k+1}) \in \mathcal{X}_i \mid x(t_k) \in \mathcal{X}_j, u(t_k) \in \mathcal{U}^\alpha\right)$, where $t_{k+1} - t_k = T \in \mathbb{R}^+$ is the time step size. While the position transition probabilities can be computed offline, the values $\Gamma_i^{\alpha\beta}(t_k)$ are computed during prediction as

$$\begin{aligned} \Gamma_i^{\alpha\beta}(t_k) &= \texttt{norm}\left(\hat{\Gamma}_i^{\alpha\beta}(t_k)\right) \coloneqq \frac{\hat{\Gamma}_i^{\alpha\beta}(t_k)}{\sum_\alpha \hat{\Gamma}_i^{\alpha\beta}(t_k)}, \\ \hat{\Gamma}_i^{\alpha\beta}(t_k) &= \lambda_{i,\mathrm{dyn}}^\alpha(t_k)\lambda_{i,\mathrm{stat}}^\alpha \Psi^{\alpha\beta}, \end{aligned} \tag{1}$$

with intrinsic action transition probabilities $\Psi^{\alpha\beta}$ as well as priority values $\lambda_{i,\mathrm{stat}}^\alpha$ and $\lambda_{i,\mathrm{dyn}}^\alpha(t_k)$ [3].

B. Intrinsic Action Transition

Similarly as in [3], the values $\Psi^{\alpha\beta}$ are computed with the parameters θ_1, θ_2, θ_3, and the index α_v^* representing an average walking speed as

$$\begin{aligned} \Psi^{\alpha\beta} \propto \exp\big(&- \theta_1 \cdot \mathrm{diff}(\alpha_\psi, \beta_\psi) - \theta_2 \cdot \mathrm{diff}(\alpha_v, \beta_v) \\ &- \theta_3 \cdot \mathrm{diff}(\alpha_v, \alpha_v^*)\big), \end{aligned} \tag{2}$$

where the operator $\mathrm{diff}(\cdot,\cdot)$ returns the absolute difference of interval centers either of speed or periodic orientation.

C. Constraints from Static Environments

For the considered scenario in this work, we assume the goal direction denoted by the index α_ψ^* is towards the road and compute the values $\lambda_{i,\mathrm{stat}}^\alpha$ with the parameter θ_4 as

$$\forall i : \lambda_{i,\mathrm{stat}}^\alpha \propto \exp\big(- \theta_4 \cdot \mathrm{diff}(\alpha_\psi, \alpha_\psi^*)\big). \tag{3}$$

D. Constraints from Dynamic Environments

By considering other traffic participants and following the collision checking process in [3], one can obtain the conditional collision probabilities $p_{i,\alpha}^C(t_{k+\kappa})$ for $\kappa = 1, 2, \ldots, K$ with $K \in \mathbb{N}^+$ collision checking steps for each joint event referring to a pedestrian $x^{\mathrm{ped}}(t_k) \in \mathcal{X}_i, u^{\mathrm{ped}}(t_k) \in \mathcal{U}^\alpha$. In this work, we propose a nonlinear relationship between dynamic priority values $\lambda_{i,\mathrm{dyn}}^\alpha(t_k)$ and conditional collision probabilities $p_{i,\alpha}^C(t_{k+\kappa})$ with K parameters γ_κ:

$$\lambda_{i,\mathrm{dyn}}^\alpha(t_k) = \min_{\kappa \in \{1,\ldots,K\}} \left(1 - p_{i,\alpha}^C(t_{k+\kappa})\right)^{\gamma_\kappa}. \tag{4}$$

To compute conditional collision probabilities, we extend the dimensions of vehicles in their moving directions as in [3]. Two parameters θ_5 and θ_6 are used to compute the

weights $w_h(t_{k+\kappa})$ of the position cells $\mathcal{X}_h$ in the extended area of each vehicle:

$$w_h(t_{k+\kappa}) = \theta_5 \cdot \exp\left(-\theta_6 \frac{\mathrm{dist}\left(\mathcal{X}_h, x^{\mathrm{veh}}(t_{k+\kappa})\right)}{v^{\mathrm{veh}}(t_{k+\kappa})}\right), \quad (5)$$

where $x^{\mathrm{veh}}(t_{k+\kappa})$ and $v^{\mathrm{veh}}(t_{k+\kappa})$ denote the vehicle's position and speed, respectively; the operator $\mathrm{dist}(\cdot,\cdot)$ returns the distance between the center of a position cell and the vehicle's front along its driveway. For a more detailed explanation, we refer to [3].

E. Settings of Markov Chains

As a compromise between high resolution and low computation time, we have chosen the following values: time step size $T = 0.48\,\mathrm{s}$ (corresponding to 8 timestamps in the dataset [27]); *GridSize* $= 0.2\,\mathrm{m}$ for position cells; *SpeedInterval* $= 0.3\,\mathrm{m\,s^{-1}}$ and *OrientationInterval* $= \pi/8$ for action cells; collision checking steps $K = 5$. The average walking speed denoted by the index α_v^* is set to $1.5\,\mathrm{m\,s^{-1}}$.

III. Calibration

We calibrate model parameters using GAs for two major reasons. First, GAs search the solution space with a population of parameter sets; thus, the probability of getting stuck in a local optimum can be reduced [29]. Second, it is easy to handle multi-objective optimization problems, because the use of population of individual parameter sets also helps to find multiple non-dominated solutions [30].

A. Preprocessing Dataset

The Daimler dataset [27] is used for both calibration and evaluation. For each original trajectory, a labeled timestamp t_{event} denotes the time point when the pedestrian decides to stop at the curb or step onto the road soon. We shift the prediction to begin at $t_0 = t_{\mathrm{event}} - \nu T$ for $\nu \in \mathbb{N}^+$, so that several trajectories with different initial conditions are obtained. Moreover, each timestamp of the original trajectories annotates whether the pedestrian has seen the approaching vehicle. We treat this annotation as a cue for situation awareness and perform collision checking process (cf. Sec. II-D) after observing that the pedestrian has seen the vehicle. Therefore, we only consider those trajectories where the pedestrian either has already seen the vehicle until t_0 (denoted as *SV*–"seen vehicle") or has not seen it in the whole recording (*NSV*–"not seen vehicle"). Out of the 76 *SV*-trajectories the pedestrians cross in 41 cases and stop in the other 35, whereas all pedestrians cross in the 32 *NSV*-trajectories.

Due to the relative small size of this dataset, we repeat the experiment for 20 times for separate training and testing. For each experiment, we randomly chose 72 trajectories as the training set to calibrate model parameters; the remaining 36 trajectories comprise the testing set to evaluate the model performance.

B. Objective functions

Let us first introduce two objective functions from different viewpoints; later, we motivate the use of both.

1) Spatiotemporal Accuracy: Given the ground truth 2D positions $z^\eta(t_k) = (x_1^{\mathrm{meas}}, x_2^{\mathrm{meas}})^T$ of the η-th trajectory, the objective function f_1 is defined as the weighted mean absolute error [31]:

$$f_1 := \frac{1}{N} \sum_{\eta, t_k} f_1(\eta, t_k), \quad (6)$$

$$f_1(\eta, t_k) := \sum_{i=1}^{d} ||\mathrm{center}(\mathcal{X}_i) - z^\eta(t_k)||_2 \cdot P\left(x^\eta(t_k) \in \mathcal{X}_i\right), \quad (7)$$

where N is the total prediction steps over all trajectories in the training set; d is the number of position cells; the operator $\mathrm{center}(\cdot)$ returns the volumetric center of a set.

2) Intention Classification Accuracy: To quantify the predicted pedestrian crossing intention from the spatiotemporal outputs of MC, we utilize the predicted occupancy at a specific point in time $t_{\mathrm{eval}} := t_{\mathrm{event}} + \mathit{offset}$; as the ground truth position at t_{event} in most cases is located on the sidewalk, we set $\mathit{offset} = T$ to ensure as far as possible that the predicted occupancy at t_{eval} is on the road due to the underlying predicted crossing intention, or, is still on the sidewalk due to the underlying predicted non-crossing intention (cf. Fig. 5 for t_{event} and t_{eval}). Let the labeling of each *SV*-trajectory be $y^\eta \in \{0, 1\}$ for stopping and crossing, respectively. Then the cross-entropy as the second objective function f_2 is computed as [32]

$$f_2 := -\frac{1}{N_{SV}} \sum_{\eta=1}^{N_{SV}} y^\eta \cdot \ln(p) + (1 - y^\eta) \cdot \ln(1 - p), \quad (8)$$

$$p = \max\left(\min\left(p_{\mathrm{cross}}^\eta, 1 - \epsilon\right), \epsilon\right), \quad (9)$$

$$p_{\mathrm{cross}}^\eta = \sum_{i=1}^{d} P(x^\eta(t_{\mathrm{eval}}) \in \mathcal{X}_i) \cdot \mathbf{1}_{\mathcal{X}_i \subset \mathrm{road}}, \quad (10)$$

with N_{SV} as the number of *SV*-trajectories in the training set and the indicator function $\mathbf{1}_{(\cdot)}$ returning 1 if the condition is true, and 0 otherwise; to avoid the evaluation of $\ln(0)$ in (8), we ensure that the probability p_{cross}^η of being on the road at t_{eval} is at least $\epsilon = 0.01$ and not more than $1 - \epsilon = 0.99$ in (9).

We see a limitation of using the objective function f_1 only, because computing the position error in (6) does not discriminate between the errors when the ground truth is close to the curb and those when the ground truth is away from the curb; yet, the former one can be more critical than the latter one from the viewpoint of an approaching vehicle. In contrast, the objective function f_2 emphasizes the predicted pedestrian's position error close to the curb from the perspective of the vehicle. Besides, it is difficult to distinguish a crossing trajectory from a stopping one by using f_1 in some cases where the recordings stop shortly after t_{event} (cf. Fig. 6). On the contrary, using f_2 enables conveniently evaluating such short recordings without extrapolating trajectories, since only the labeling of each trajectory y^η is required in (8). However, the spatiotemporal information is lost when using f_2 only. Therefore, we consider both objectives for the calibration.

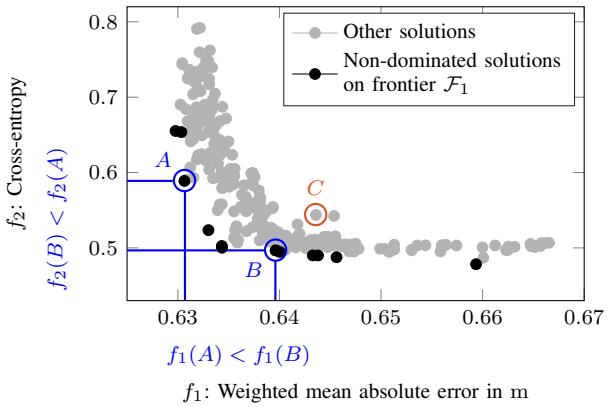

Fig. 3. Candidates of parameter sets in the objective function space in the 20th generation of the GA from an experiment. Smaller values of f_1 and f_2 are preferred.

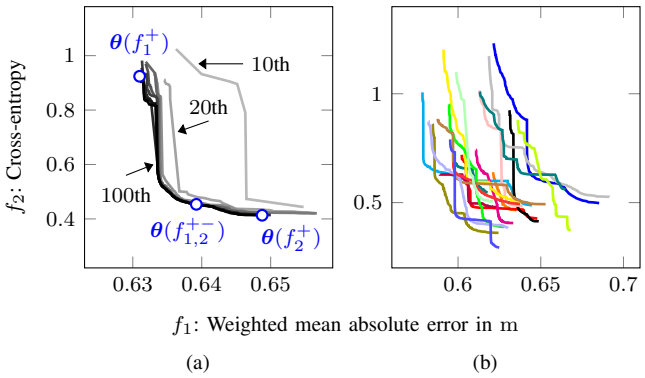

Fig. 4. (a) Evolution of frontiers $\mathcal{F}_1$ in the 10th, 20th, ..., 100th generation of the GA from an experiment in gray color with increasing darkness for later generations. The blue circles represent the positions of 3 selected parameter sets in the objective function space according to different criteria. (b) Frontiers $\mathcal{F}_1$ in their 100th generation of the GA from all 20 experiments.

C. Multi-objective Optimization

To consider both objective functions f_1 and f_2, we perform a multi-objective optimization, because it may be difficult to aggregate different objectives (due to e.g., their complex correlation) into a single synthetic objective a priori, that is, before alternatives are known [33]. We adopt GAs to handle the multi-objective optimization problem similarly as in [16]. The goal of multi-objective optimization is to find a set of non-dominated solutions, with convergence to the Pareto frontier while maintaining good diversity of solutions [34]. The non-dominated solutions are those which cannot be improved considering all objectives simultaneously. For instance, cf. Fig. 3, the candidate C is dominated by B, whereas A and B are not dominated by each other and hence they lie on the same frontier $\mathcal{F}_1$.

We use the MATLAB Global Optimization Toolbox[1] with the settings of the GA given in Table I (all other options are left at their default values) for calibrating the model parameters $\boldsymbol{\theta} := (\theta_1, \theta_2, \ldots, \theta_6, \gamma_1, \gamma_2, \ldots, \gamma_5)$.

TABLE I
THE CHOSEN PARAMETER VALUES FOR MULTI-OBJECTIVE GA.

Option	Value	Option	Value
PopulationSize	300	*MaxGenerations*	100
SelectionFcn	'selectiontournament'	*TournamentSize*	3
CrossoverFcn	'crossoverscattered'	*CrossoverFraction*	0.7
MutationFcn	'mutationadaptfeasible'	*ParetoFraction*	0.4

D. Optimization Results and Candidate Selection

Fig. 4a depicts the evolution of the frontiers $\mathcal{F}_1$ with the highest rank (the solutions on other frontiers are inferior to those on $\mathcal{F}_1$, cf. Fig. 3) in different generations of the GA from an experiment. The frontier $\mathcal{F}_1$ converges after about 30 generations. Fig. 4b shows the frontiers $\mathcal{F}_1$ in their 100th generation of the GA from all 20 experiments.

[1]MATLAB and Global Optimization Toolbox Release 2018b, The MathWorks, Inc., Natick, Massachusetts, United States.

The question of how to choose a final optimized solution among non-dominated solutions depends on the user preference. Therefore, for each experiment we select 3 parameter sets from the last generation according to the following criteria, cf. Fig. 4a:

- Let $\boldsymbol{\theta}(f_1^+)$ be the best parameter set on $\mathcal{F}_1$ w.r.t. objective function f_1.
- Let $\boldsymbol{\theta}(f_2^+)$ be the best parameter set on $\mathcal{F}_1$ w.r.t. objective function f_2.
- Let $\boldsymbol{\theta}(f_{1,2}^{+-})$ be the parameter set on $\mathcal{F}_1$ whose f_1 value is closest to $\frac{f_1(\boldsymbol{\theta}(f_1^+))+f_1(\boldsymbol{\theta}(f_2^+))}{2}$, i.e., the middle between the f_1 values of $\boldsymbol{\theta}(f_1^+)$ and $\boldsymbol{\theta}(f_2^+)$.

We denote the controlled Markov chains with the above parameter sets by MC-$\boldsymbol{\theta}(f_1^+)$, MC-$\boldsymbol{\theta}(f_2^+)$, and MC-$\boldsymbol{\theta}(f_{1,2}^{+-})$, respectively. Their performance will be evaluated in the next section.

IV. Evaluation

A. Spatiotemporal Accuracy

We first introduce a baseline model MC-UncAct using the same Markov chains from Sec. II but without applying action transition—the transition probabilities $\Gamma_i^{\alpha\beta}(t_k)$ in (1) are set to 1 if $\alpha = \beta$, and 0 otherwise, $\forall i, \alpha, \beta$; hence, this model suffers from the same discretization of MC for a fair comparison. Then we compare the position errors $f_1(t_k) := \frac{1}{N_{t_k}} \sum_\eta f_1(\eta, t_k)$ with N_{t_k} trajectories (*SV+NSV*) evaluated at different points in time t_k between MC-UncAct and our approach with different optimized parameter sets.

Their average performance (standard deviation in parentheses) in testing sets from 20 experiments is listed in Table II, where a smaller $f_1(t_k)$ is preferred. The spatiotemporal accuracy of MC with the parameter sets $\boldsymbol{\theta}(f_1^+)$ and $\boldsymbol{\theta}(f_{1,2}^{+-})$ is better than that of others.

B. Intention Classification Accuracy

We evaluate the performance of our approach regarding recognizing whether pedestrians will cross in front of vehicles. First, we convert the crossing probability p_{cross}^η from

TABLE II

AVERAGE PERFORMANCE IN TESTING SETS.

	MC-$\boldsymbol{\theta}(f_1^+)$	MC-$\boldsymbol{\theta}(f_{1,2}^{+-})$	MC-$\boldsymbol{\theta}(f_2^+)$	MC-UncAct
$f_1(t_1)$	0.292(0.02)	**0.290**(0.02)	0.290(0.02)	0.293(0.02)
$f_1(t_2)$	0.474(0.04)	**0.472**(0.04)	0.473(0.03)	0.500(0.03)
$f_1(t_3)$	0.631(0.04)	**0.623**(0.04)	0.623(0.04)	0.694(0.04)
$f_1(t_4)$	0.786(0.05)	**0.784**(0.04)	0.795(0.04)	0.891(0.05)
$f_1(t_5)$	**0.960**(0.07)	0.987(0.06)	1.013(0.06)	1.072(0.07)
$f_1(t_6)$	**1.207**(0.12)	1.254(0.09)	1.290(0.11)	1.294(0.12)

MC in (10) of each trajectory into the four categories of the confusion matrix [35] depending on the threshold $\rho = 0.5$:

$$\text{(true positives)} \quad TP = \sum_{\eta=1}^{N'_{SV}} \mathbf{1}_{p^{\eta}_{\text{cross}} \geq \rho} \cdot \mathbf{1}_{y^{\eta}=1},$$

$$\text{(false positives)} \quad FP = \sum_{\eta=1}^{N'_{SV}} \mathbf{1}_{p^{\eta}_{\text{cross}} \geq \rho} \cdot \mathbf{1}_{y^{\eta}=0},$$

$$\text{(false negatives)} \quad FN = \sum_{\eta=1}^{N'_{SV}} \mathbf{1}_{p^{\eta}_{\text{cross}} < \rho} \cdot \mathbf{1}_{y^{\eta}=1},$$

$$\text{(true negatives)} \quad TN = \sum_{\eta=1}^{N'_{SV}} \mathbf{1}_{p^{\eta}_{\text{cross}} < \rho} \cdot \mathbf{1}_{y^{\eta}=0},$$

where N'_{SV} is the number of *SV*-trajectories in the testing set. Based on that, one obtains the accuracy $ACC = \frac{TP+TN}{TP+FP+FN+TN}$, the false positive rate $FPR = \frac{FP}{FP+TN}$, and the true positive rate $TPR = \frac{TP}{TP+FN}$ [35].

For comparison[2] we implement a support vector machine with the radial basis function kernel (SVM-RBF) [36]. The used features $X(t_0)$ for SVM-RBF are similar to those[3] in [4], [5]:

$$X(t_0) = \left(v_1^{\text{ped}}, v_2^{\text{ped}}, v_1^{\text{veh}}, v_2^{\text{veh}}, d_{\text{curb}}^{\text{ped}}, d_{\text{veh}}^{\text{ped}}\right)^T,$$

where $v_1 = v\cos\psi$ and $v_2 = v\sin\psi$; $d_{\text{curb}}^{\text{ped}}$ represents the distance between the pedestrian and the curb; $d_{\text{veh}}^{\text{ped}}$ denotes the distance between the pedestrian and the vehicle's front, at the prediction beginning t_0. For both training and testing, those features are normalized with unified mean values and standard deviations. In the training set from each experiment, we use three-fold cross-validation and grid search [37] to obtain optimized parameters for SVM-RBF by maximizing the accuracy ACC.

Table III compares the average performance (standard deviation in parentheses) in testing sets from 20 experiments between SVM-RBF and MC with different parameter sets

[2]For a more generalized use case, we mixed the trajectories from all sub-scenarios in the Daimler dataset [27] for both training (cf. Sec. III-A) and testing, including the anomalous sub-scenario where pedestrians have seen the vehicles and cross in a critical situation. Since the proposed model in [27] performs trajectory predictions based on specific scenarios containing "normal" behaviors and fails to handle the above mentioned anomalous scenario, we chose not to compare it with MC.

[3]Some specific features w.r.t. zebra crossing are excluded, since no zebra crossing exists in the Daimler dataset [27].

TABLE III

AVERAGE PERFORMANCE IN TESTING SETS.

	MC-$\boldsymbol{\theta}(f_1^+)$	MC-$\boldsymbol{\theta}(f_{1,2}^{+-})$	MC-$\boldsymbol{\theta}(f_2^+)$	SVM-RBF
ACC	0.650(0.08)	0.720(0.07)	0.746(0.07)	**0.760**(0.07)
FPR	0.529(0.20)	0.365(0.15)	**0.263**(0.13)	0.290(0.15)
TPR	0.806(0.10)	**0.813**(0.12)	0.769(0.09)	0.806(0.12)

TABLE IV

AVERAGE PERFORMANCE IN TESTING SETS (SHORT-TERM PREDICTION).

	MC-$\boldsymbol{\theta}(f_1^+)$	MC-$\boldsymbol{\theta}(f_{1,2}^{+-})$	MC-$\boldsymbol{\theta}(f_2^+)$	SVM-RBF
ACC	0.704(0.12)	0.771(0.11)	**0.775**(0.15)	0.744(0.15)
FPR	0.485(0.28)	0.260(0.24)	**0.179**(0.21)	0.281(0.28)
TPR	**0.856**(0.13)	0.824(0.15)	0.779(0.17)	0.764(0.20)

(while ACC and TPR are to be maximized, a smaller FPR is preferred). While SVM-RBF achieves the highest ACC, MC-$\boldsymbol{\theta}(f_2^+)$ yields the lowest FPR and MC-$\boldsymbol{\theta}(f_{1,2}^{+-})$ the highest TPR. Moreover, cf. Table IV, if considering only those trajectories where pedestrians are not far away from the curb at the prediction beginning, i.e., $t_0 = t_{\text{event}} - T$, the performance of MC is further improved compared to their performance evaluated on all *SV*-trajectories (cf. Table III). Especially, MC-$\boldsymbol{\theta}(f_{1,2}^{+-})$ and MC-$\boldsymbol{\theta}(f_2^+)$ perform better than SVM-RBF regarding (all) ACC, FPR, and TPR.

C. Discussion

We demonstrate the calibration of MC using GAs and two objective functions. As shown in Table III, calibrating MC in the direction of minimizing the position error cannot ensure the best performance with respect to crossing intention recognition. In contrast, the spatiotemporal accuracy gets worse by treating MC as a classifier during optimization, cf. Table II. As a trade-off, we select the parameter set $\boldsymbol{\theta}(f_{1,2}^{+-})$ from the Pareto frontier according to the criterion in Sec. III-D, with which MC can achieve satisfactory results: while yielding better spatiotemporal accuracy than the baseline model MC-UncAct, it performs comparably to a SVM-RBF for pedestrian intention recognition. Moreover, as pedestrians get closer to the curb, the performance of MC is significantly improved regarding (all) ACC, FPR, and TPR, whereas SVM-RBF degenerates slightly (cf. Table IV). The reason for this can be in particular the influence of intrinsic action transition for short-term prediction.

Fig. 5 illustrates a dangerous situation (the vehicle would collide with the pedestrian if it does not react), where our approach predicts confidently ($98.9\,\%$) that the pedestrian will step onto the road in about $1.5\,\text{s}$. Another prediction example (using the trade-off solution $\boldsymbol{\theta}(f_{1,2}^{+-})$ as in Fig. 5) can be found in Fig. 6: although there is $12.3\,\%$ probability that the pedestrian will be on the road at t_5, this actually represents the intention that the pedestrian tries to cross behind the passing vehicle without stopping. This tendency is caused by the priority values $\lambda_{i,\text{stat}}^{\alpha}$.

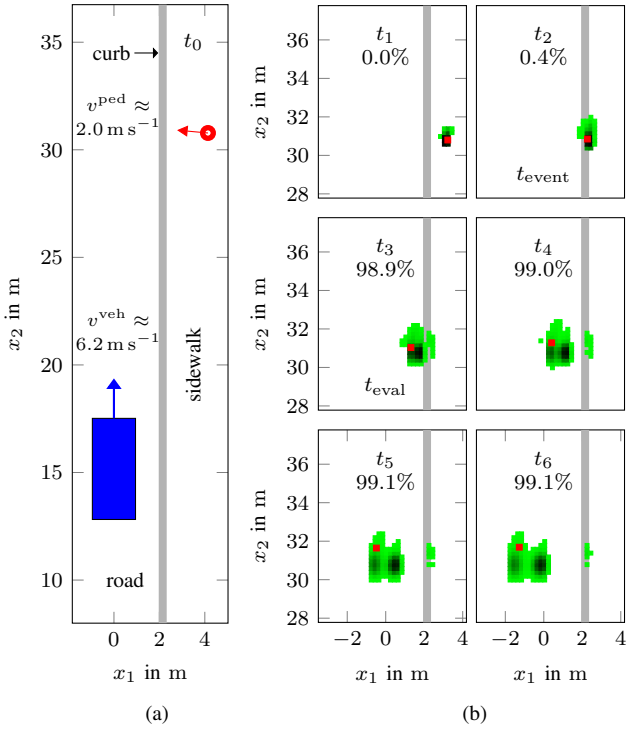

Fig. 5. A crossing trajectory in [27]. (a) The initial positions and velocities of the pedestrian (red circle) and the vehicle (blue rectangle) at t_0. (b) The green cells illustrate the predicted occupancies using MC-$\boldsymbol{\theta}(f^{+-}_{1,2})$ in different time steps (the darker the color, the higher the probability); the red squares represent the ground truth positions at different points in time; the numbers in each box are the summed probability of green cells on the road.

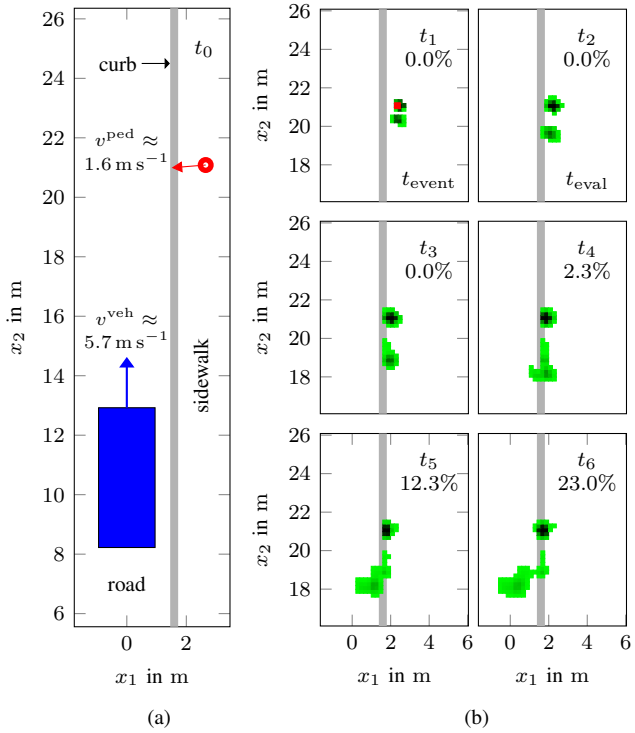

Fig. 6. A stopping trajectory in [27] with the same annotations as in Fig. 5. Although the recording stops after t_1 and the ground truth position therefore only exists at t_1 in part (b), we regard our prediction as a true negative w.r.t. crossing in front of the vehicle based on the labeling of this recording and the predicted occupancy on the road at t_2 (t_{eval}) of 0.0 % probability.

V. Conclusions

In the context of predicting pedestrian crossing behavior in urban environments, our approach achieves comparable performance compared to a support vector machine. But, rather than a binary result, our approach yields more detailed information about where the pedestrian will be and when, as well as the corresponding probabilities. As the pedestrian gets closer to the road, the performance of our approach is further improved. Thus, our approach is suitable for both long-term and short-term prediction. To sum up, it is beneficial to combine all aforementioned advantages in a single framework for a general use.

For calibrating such a model as ours which outputs probabilistic spatiotemporal results, we show the usefulness of considering two objective functions from different viewpoints—spatiotemporal accuracy and intention classification accuracy—by performing multi-objective optimization using genetic algorithms.

Acknowledgement

We gratefully thank Matthias Woehrle, Fanta Camara, Hendrik Berkemeyer, Timm F. Gloger, and Jihad Miramo for helpful conversations and suggestions.

References

[1] S. Lefèvre, D. Vasquez, and C. Laugier, "A survey on motion prediction and risk assessment for intelligent vehicles," *ROBOMECH Journal*, vol. 1, no. 1, pp. 1–14, 2014.

[2] B. Völz, H. Mielenz, I. Gilitschenski, R. Siegwart, and J. Nieto, "Inferring pedestrian motions at urban crosswalks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 2, pp. 544–555, 2019.

[3] J. Wu, J. Ruenz, and M. Althoff, "Probabilistic map-based pedestrian motion prediction taking traffic participants into consideration," in *Proc. of the IEEE Intelligent Vehicles Symposium*, 2018, pp. 1285–1292.

[4] S. Bonnin, T. H. Weisswange, F. Kummert, and J. Schmuedderich, "Pedestrian crossing prediction using multiple context-based models," in *Proc. of the IEEE Int. Conf. on Intelligent Transportation Systems*, 2014, pp. 378–385.

[5] B. Völz, H. Mielenz, G. Agamennoni, and R. Siegwart, "Feature relevance estimation for learning pedestrian behavior at crosswalks," in *Proc. of the IEEE Int. Conf. on Intelligent Transportation Systems*, 2015, pp. 854–860.

[6] Z. Fang and A. M. López, "Is the pedestrian going to cross? Answering by 2D pose estimation," in *Proc. of the IEEE Intelligent Vehicles Symposium*, 2018, pp. 1271–1276.

[7] A. T. Schulz and R. Stiefelhagen, "Pedestrian intention recognition using latent-dynamic conditional random fields," in *Proc. of the IEEE Intelligent Vehicles Symposium*, 2015, pp. 622–627.

[8] F. Camara, O. Giles, R. Madigan, M. Rothmüller, P. H. Rasmussen, S. A. Vendelbo-Larsen, G. Markkula, Y. M. Lee, L. Garach, N. Merat, and C. W. Fox, "Predicting pedestrian road-crossing assertiveness for autonomous vehicle control," in *Proc. of the IEEE Int. Conf. on Intelligent Transportation Systems*, 2018, pp. 2098–2103.

[9] A. T. Schulz and R. Stiefelhagen, "A controlled interactive multiple model filter for combined pedestrian intention recognition and path

prediction," in *Proc. of the IEEE Int. Conf. on Intelligent Transportation Systems*, 2015, pp. 173–178.

[10] J. F. Kooij, F. Flohr, E. A. Pool, and D. M. Gavrila, "Context-based path prediction for targets with switching dynamics," *International Journal of Computer Vision*, pp. 1–24, 2018.

[11] M. Hoy, Z. Tu, K. Dang, and J. Dauwels, "Learning to predict pedestrian intention via variational tracking networks," in *Proc. of the IEEE Int. Conf. on Intelligent Transportation Systems*, 2018, pp. 3132–3137.

[12] D. Helbing and P. Molnar, "Social force model for pedestrian dynamics," *Physical review E*, vol. 51, no. 5, p. 4282, 1995.

[13] F. Zanlungo, T. Ikeda, and T. Kanda, "Social force model with explicit collision prediction," *EPL (Europhysics Letters)*, vol. 93, no. 6, p. 68005, 2011.

[14] P. Scovanner and M. F. Tappen, "Learning pedestrian dynamics from the real world," in *Proc. of the IEEE Int. Conf. on Computer Vision*, 2009, pp. 381–388.

[15] D. Yang, Ü. Özgüner, and K. Redmill, "Social force based microscopic modeling of vehicle-crowd interaction," in *Proc. of the IEEE Intelligent Vehicles Symposium*, 2018, pp. 1537–1542.

[16] W. Zeng, P. Chen, G. Yu, and Y. Wang, "Specification and calibration of a microscopic model for pedestrian dynamic simulation at signalized intersections: A hybrid approach," *Transportation Research Part C: Emerging Technologies*, vol. 80, pp. 37–70, 2017.

[17] H. Cheng and M. Sester, "Modeling mixed traffic in shared space using LSTM with probability density mapping," in *Proc. of the IEEE Int. Conf. on Intelligent Transportation Systems*, 2018, pp. 3898–3904.

[18] A. Rudenko, L. Palmieri, and K. O. Arras, "Joint long-term prediction of human motion using a planning-based social force approach," in *Proc. of the IEEE Int. Conf. on Robotics and Automation*, 2018.

[19] M. Koschi, C. Pek, M. Beikirch, and M. Althoff, "Set-based prediction of pedestrians in urban environments considering formalized traffic rules," in *Proc. of the IEEE Int. Conf. on Intelligent Transportation Systems*, 2018, pp. 2704–2711.

[20] K. M. Kitani, B. D. Ziebart, J. A. Bagnell, and M. Hebert, "Activity forecasting," in *European Conference on Computer Vision*. Springer, 2012, pp. 201–214.

[21] V. Karasev, A. Ayvaci, B. Heisele, and S. Soatto, "Intent-aware long-term prediction of pedestrian motion," in *Proc. of the IEEE Int. Conf. on Robotics and Automation*, 2016, pp. 2543–2549.

[22] E. Rehder, F. Wirth, M. Lauer, and C. Stiller, "Pedestrian prediction by planning using deep neural networks," in *Proc. of the IEEE Int. Conf. on Robotics and Automation*, 2018.

[23] C. G. Keller and D. M. Gavrila, "Will the pedestrian cross? A study on pedestrian path prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 2, pp. 494–506, 2014.

[24] Y. F. Chen, M. Liu, and J. P. How, "Augmented dictionary learning for motion prediction," in *Proc. of the IEEE Int. Conf. on Robotics and Automation*, 2016, pp. 2527–2534.

[25] G. Habibi, N. Jaipuria, and J. P. How, "Context-aware pedestrian motion prediction in urban intersections," *arXiv preprint arXiv:1806.09453*, 2018.

[26] A. Johansson, D. Helbing, and P. K. Shukla, "Specification of the social force pedestrian model by evolutionary adjustment to video tracking data," *Advances in complex systems*, vol. 10, no. supp02, pp. 271–288, 2007.

[27] J. F. P. Kooij, N. Schneider, F. Flohr, and D. M. Gavrila, "Context-based pedestrian path prediction," in *European Conference on Computer Vision*. Springer, 2014, pp. 618–633.

[28] M. Althoff, "Reachability analysis and its application to the safety assessment of autonomous cars," Ph.D. dissertation, Technische Universität München, 2010.

[29] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, 1989.

[30] L. Wang, A. H. Ng, and K. Deb, *Multi-objective evolutionary optimisation for product design and manufacturing*. Springer, 2011.

[31] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate research*, vol. 30, no. 1, pp. 79–82, 2005.

[32] K. P. Murphy, *Machine learning: A probabilistic perspective*. MIT press, 2012.

[33] J. Branke, J. Branke, K. Deb, K. Miettinen, and R. Slowiński, *Multiobjective optimization: Interactive and evolutionary approaches*. Springer Science & Business Media, 2008, vol. 5252.

[34] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[35] T. Fawcett, "An introduction to ROC analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[36] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[37] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," 2003. [Online]. Available: https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf