Long-Term Recurrent Predictive Model for Intent Prediction of Pedestrians via Inverse Reinforcement Learning

Khaled Saleh, Mohammed Hossny and Saeid Nahavandi
Institute for Intelligent Systems Research and Innovation (IISRI)
Deakin University, Australia
Email: {kaboufar, mhossny, saeid.nahavandi}@deakin.edu.au

***Abstract*—Recently, the problem of intent and trajectory prediction of pedestrians in urban traffic environments has got some attention from the intelligent transportation research community. One of the main challenges that make this problem even harder is the uncertainty exists in the actions of pedestrians in urban traffic environments, as well as the difficulty in inferring their end goals. In this work, we are proposing a data-driven framework based on Inverse Reinforcement Learning (IRL) and the bidirectional recurrent neural network architecture (B-LSTM) for long-term prediction of pedestrians' trajectories. We evaluated our framework on real-life datasets for agent behavior modeling in traffic environments and it has achieved an overall average displacement error of only 2.93 and 4.12 pixels over 2.0 secs and 3.0 secs ahead prediction horizons respectively. Additionally, we compared our framework against other baseline models based on sequence prediction models only. We have outperformed these models with the lowest margin of average displacement error of more than 5 pixels.**

I. Introduction

Recently, the development of autonomous vehicles (AVs) has reached major milestones and witnessed a number of success in highway traffic environments [1], [2]. That being said, they are still however faced with a number of challenges in urban traffic environments. More specifically when it comes to interacting with vulnerable road users (VRUs) such as pedestrians and cyclists [3]–[5]. Thus, the necessity for having predictive models within these vehicles that can infer and understand the VRUs intentions over longer time periods become inevitable. In the advanced driving assistance systems (ADAS) community, the intent prediction of pedestrians problem has been thoroughly investigated over the past few years [6]–[8]. Where the intent prediction problem is commonly accomplished based on forecasting the motion trajectories of pedestrians in traffic environments. Since the driver in ADAS context is still in command of the driving decisions most of the time. Therefore the predictive models for intent prediction proposed in the literature were only predicting shorter time horizons of pedestrians' intention. On the other hand, in the context of AVs, where the driver will be out of the decision-making loop, the need for long-term prediction is essential. As it was illustrated in [3], [9], longer time horizons of intent prediction of pedestrians were attributed to being one of the most strong cues for a trusted interaction between pedestrians and AVs.

Commonly, the approaches for the intent prediction introduced in ADAS field rely either on single linear dynamical models such as Kalman filter [6] or a switching multiple linear dynamical models [7], [8]. One constraint of linear dynamical models they need an explicit modeling of the agent (i.e.pedestrian) in the traffic scene. They are also challenged by the variable non-linear and uncertain dynamics of pedestrians over longer time periods. Another approach has been also investigated in the literature is the planning based models. Unlike linear dynamical models, planning based models do not need an explicit modeling of the pedestrians. Additionally, they have been proven to give a resilient prediction of pedestrians over a longer time horizon [10]–[12]. However, one challenge planning-based models are encountered with is their inherent reliance on a prior known end goal. Sequence prediction models [13], [14] are also another approach adopted for the intent prediction problem of pedestrians. Similar to planning-based models they do not need an explicit modeling of the pedestrians' motion dynamics. Unlike planning-based models, they do not need a prior end goal for the pedestrians to be known beforehand. Despite the promising results that data-driven approaches have shown for the intent prediction problem for pedestrians, they still need some improvements. For instance, so far the proposed sequence prediction models in the literature do not take into account the inherent uncertainty of the pedestrians' actions. Additionally, they are neglecting the effect of the physical environment on the pedestrians' actions.

Thus, in this paper, we are proposing a framework that combines the best of the two worlds of planning-based models and sequence prediction models. We first learn

978-1-5386-6602-9/18/$31.00 ©2018 IEEE

the reward function of the traffic environment by just observing a demonstrated trajectories of the pedestrians via inverse reinforcement learning (IRL). Then using the learned reward function alongside the motion trajectory of pedestrians we learn another RNN model. The RNN model in return infers a long-term trajectory without a prior information about the pedestrians' end goal. The rest of this paper is organized as follows. Section II presents a brief literature review of the problem. Section III describes the problem formulation and the proposed solution. Section IV describes the datasets used and the validation method. Finally, Section V concludes this paper.

II. Related Work

The intent prediction problem of pedestrians has been commonly approached in the literature as a dynamical motion modeling problem solved using recursive Bayesian filters [6], [7], [15]. In [6], one of the early work on the intent prediction problem of pedestrians, they used an Extended Kalman filter (EKF)to model the linear dynamical motion model of pedestrians in four distinctive crossing scenarios. Keller et al. [7], relied on another dynamical motion model based on a Gaussian process to infer whether a pedestrian walking on a curb will cross or not. In their work, they developed two different motion models to identify the stopping and walking behavior of the pedestrians. They relied on optical flow fields so that they could predict the trajectory of the pedestrian.Another category of approaches utilized in the literature for the intent and trajectory prediction of pedestrians is the planning-based models. These models are inspired by the path planning approaches that are heavily used in the robotics field. Unlike the traditional path planning approaches where an ego-centric trajectory to be performed by a robot in an environment, they used it to plan a trajectory of other agents (i.e, pedestrian). In planning-based approaches, the inherent assumption is that the end goal the agent is trying to reach is known in advance which might not be foreseeable in case of pedestrians. In [16], a planning-based approach used for forecasting pedestrians' trajectories in traffic environments. Since the end goal of the pedestrians is not known beforehand, they firstly infer a set of possible goals using a combination of Gaussian Mixture Model (GMM) and Particle Filter (PF). Using these inferred end goals and an occupancy grid map of the environment, they can predict a probability distribution over the probable trajectories to these goals.

Yet another planning-based approach was introduced in [17] for on-road pedestrian avoidance system for an autonomous mobile robot. The pedestrian avoidance system took into account the pedestrians' intention and their associated uncertainty as part of the robot's motion planning framework. They formulated the problem as a Mixed Observable Markov Decision Process (MOMDP), where the motion model variables of the pedestrian are already given (fully observable) and the intention of the pedestrian is unknown (unobserved). They assume the pedestrian is directed towards his/her goal following a shortest path trajectory. They utilized a sampling based approximate algorithm called Successive Approximations of the Reachable Space under Optimal Policies (SARSOP). Using SARSOP, they solved the MOMDP model and inferred a probability distribution over the possible directions of the pedestrians.

Data-driven approaches specifically those ones based on recurrent neural network architectures such as LSTM were also investigated for the intent and trajectory prediction problem of pedestrians [13], [14]. In data-driven approaches and similar to planning-based approaches, no explicit modeling for the dynamics of the motion of pedestrians is needed to be performed firstly. However, unlike planning-based approaches, they do not require a prior information regarding the end goal of the pedestrians in the scene. In [13], recurrent neural network (RNN)-based approach was used for modeling human-to-human interactions in crowded environments from a surveillance camera's perspective. In [14], another RNN-based model for predicting trajectories of pedestrians in traffic environments was introduced. In their presented RNN model, they relied only on past positional information of pedestrians in order to predict their future motion trajectories. Similarly in [18], an RNN-based model was proposed for predicting whether pedestrians would cross or not in specific urban traffic scenarios. Based on two different input sequence features extracted from 3D LIDAR scans at an intersection, they tackled the problem as a binary classification problem. The two types of input features are namely, temporal (pedestrian's position over time) and geometrical (pedestrians' distance to curb).

III. Proposed Methodology

In this section, the proposed methodology for the intent prediction problem for pedestrians in urban traffic environment will be discussed. Firstly, we will begin with our formulation of the problem. Then, the building blocks of our proposed framework (shown in Figure 1) will be described.

A. Problem Formulation

In our formulation for the intent prediction problem of pedestrians in traffic environment, we cast the problem

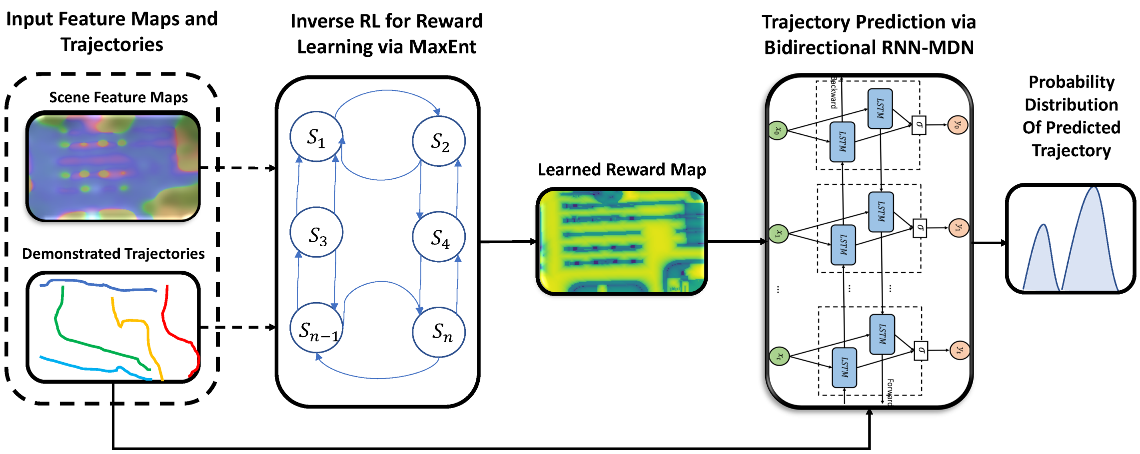

Figure 1. The proposed framework for long-term trajectory prediction of pedestrians in urban traffic environment. Firstly, demonstrated trajectories and contextual features maps are used for learning the reward map of the scene via IRL MaxEnt. The demonstrated trajectories along with the learned reward map of the scene are passed as the input sequences for training a probabilistic trajectory prediction B-LSTM-MDN model. The output of the B-LSTM-MDN model are probability density of future sequence trajectories of input pedestrians.

as a probabilistic sequence prediction problem. Given a sequence of past trajectory observations x as well as a reward map r that represents the pedestrian's preference in an urban traffic environment. In return, we will anticipate the probability density $P(y|x,r)$ of the pedestrian's future trajectory y. In order to achieve a probabilistic sequence prediction model, we will utilize a bidirectional recurrent neural network model based on LSTM architecture [19] with a mixture density network on top of it [20]. For recovering a reward map that can accurately capture the pedestrians' preferences, we will rely on an inverse reinforcement learning (IRL) technique [21].

B. IRL and Markov Decision Process

Markov Decision Process (MDP), is one of the most widely used frameworks for modeling the dynamics of a decision making process [22]. MDP can be defined as $\mathcal{M} = \{\mathcal{S}, \mathcal{A}, \mathcal{T}, r\}$, where $\mathcal{S}$ is the state space of the system, $\mathcal{A}$ is the possible actions, $\mathcal{T}$ is the transition model that describes the system dynamics and r is the reward function. Typically acting in an MDP, results in a sequence of states and actions $\{s_0, a_0, s_1, a_1, s_2, \ldots\}$. A policy π, is the mapping sequences $(\mu_0, \mu_1, \mu_2, \ldots)$, where, at time t the mapping $\mu_t(\cdot)$ determines the action $a_t = \mu_t(s_t)$ to take when in state s_t. The ultimate goal in an MDP, is to find an optimal policy π^*, that maximizes the expected sum of rewards accumulated over time.

In IRL context, the specifications of MDP are available except the reward function r is unknown. Alternatively, a set of demonstrations $\mathcal{D} = \{\zeta_1, \zeta_2, \ldots, \zeta_N\}$ are provided by a demonstrator. Each sample trajectory ζ_i from the set of demonstration $\mathcal{D}$ is described by a pair of state-action according to $\zeta_i = \{(s_0, a_0), (s_1, a_1), \ldots, (s_T, a_T)\}$. Given, the demonstration $\mathcal{D}$, the goal of IRL is to recover the reward function r that can ultimately capture the preference of the agent. Since in real life applications, it would be difficult to observe a reward function for each action-state pair in the set of demonstration $\mathcal{D}$, especially if the state space is large. Thus, a common approach in IRL methods is collecting a feature values vector f that best characterize each possible action from the set of demonstrations $\mathcal{D}$.

C. Reward Learning for Pedestrian Intent and Trajectory Prediction

One of the most commonly used approaches for IRL is the maximum entropy IRL approach (MaxEnt) proposed in [21]. MaxEnt was successfully utilized in a number of applications such as learning driver behaviors [11], planners for social robotics [21], [23] and activity forecasting from surveillance data [10], [24]. In the formulation for the MaxEnt, the reward function can be calculated as a weighted linear combination of the feature values vector f according to Eq. 1.

$$r = \theta^T f, \tag{1}$$

where θ is a vector of unknown weights.

In this work, we will be focusing on the contextual physical information in urban traffic environment as our feature values vector for parameterizing the reward function. More specifically, we will utilize the vision-based contextual information extracted from the environment

Algorithm 1: Maximum Entropy IRL

input : $\mathcal{S}, \mathcal{A}, \mathcal{T}, f, \bar{f}$
output: Optimal set of weights $\hat{\theta}$
$\theta^1 = IRL\_initWeights()$
for n=*1:N* **do**
  $\pi^n = IRL\_valueIteration(\mathcal{S}, \mathcal{A}, \mathcal{T}, f, \theta^n)$
  $\hat{f}^n_\theta = IRL\_stateVisitFrequency(\mathcal{S}, \mathcal{A}, \mathcal{T}, f, \pi^n)$
  $\nabla\mathcal{L}^n_\theta = \bar{f} - \hat{f}^n_\theta$
  $\theta^{n+1} = IRL\_updateWeights(\theta^n, \nabla\mathcal{L}^n_\theta)$
end

Algorithm 2: IRL_valueIteration

$V(s) = -\infty$
for n=*N:1* **do**
  $V(s_{goal}) = 0$
  $Q^n(s,a) = \theta^T f_{s,a} + \mathbb{E}_{\mathcal{T}(s,a,s')}[V^n(s')]$
  $V^{n-1}(s) = \text{soft}\max_a Q^n(s,a)$
end
$\pi(a,s) = \exp^{Q(s,a)-V(s)}$

Algorithm 3: IRL_stateVisitFrequency

$\phi(s_{start}) = 1$
for n=*1:N* **do**
  $\phi_{s_{goal}} = 0$
  $\phi^{n+1}_s = \sum_{s',a} \mathcal{T}(s,a,s')\pi(a,s')\phi^n(s')$
end
$\phi_s = \sum_n \phi^n_s$
$\hat{f}_\theta = \sum_s \phi_s f_s$

by means of image semantic segmentation techniques. The common contextual information that could have a potential influence on the future actions of pedestrians are trees, buildings, sidewalks, and roads.

Using demonstrated trajectories of pedestrians in urban traffic environments along with contextual physical information, MaxEnt can be adopted for learning the reward function parameters. In MaxEnt, the probability distribution of a trajectory ζ_i is proportional to the exponentiated sum of rewards along the trajectory ζ_i, which can be formulated as in Eq 3 after the substitution in Eq 1 to produce

$$P(\zeta_i) \propto \exp \sum_{(s,a)\in\zeta_i} r_{s,a} \tag{2}$$

$$P(\zeta_i|\theta) = \frac{\exp \sum_{(s,a)\in\zeta_i} \theta^T f_{s,a}}{Z(\theta)} \tag{3}$$

where $Z(\theta)$, is the normalization function. By maximizing the entropy of Eq 3, learning from demonstration

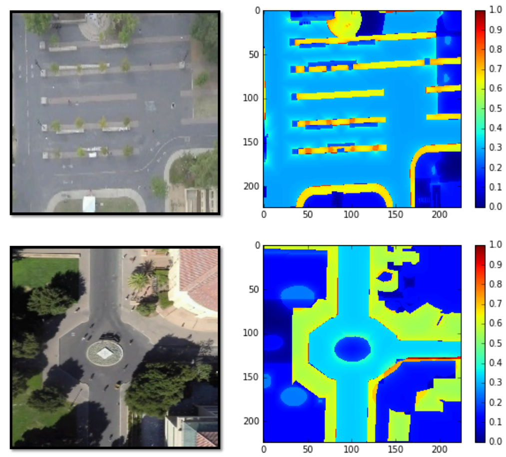

Figure 2. Sample of the learned reward maps from the Stanford Drone Dataset (SDD) using MaxEnt IRL. Highest reward value is 1.0 and the lowest is 0.0.

trajectories in MaxEnt can be accomplished. Additionally, the maximization of the entropy of Eq 3 can be interpreted as minimizing the log-likelihood of the same equation, which in returns can be calculated using learning algorithms such as conjugate or stochastic gradient descent. That been said, we will be using the similar forward-backward algorithm introduced and discussed in [10] for training the MaxEnt framework and obtaining the weights θ of the reward function. In the forward-backward algorithm (Algorithm 1, 2 and 3), the objective is to minimize the gradient of the log-likelihood of Eq 3 using a gradient descent algorithm. The gradient is calculated based on the difference between the empirical cumulative feature count $\bar{f}$ and the expected cumulative feature count $\hat{f}_\theta$ in Eq 4.

$$\nabla\mathcal{L}_\theta = \bar{f} - \hat{f}_\theta \tag{4}$$

The expected cumulative feature count is the average accumulated features according to the trajectories generated by the weights. The expected cumulative feature count can be expressed using the expected state visitation frequency property ϕ_s according to Eq 5. The sum of visits that have been to each state s are described over time by ϕ_s and can be calculated as shown in Algorithm 3. At the convergence, when $\bar{f}$ equals $\hat{f}_\theta$, an optimal set of weights $\hat{\theta}$ can be obtained as described in Algorithm 1. Which in return, can be used to have reward functions as the ones visualized in Figure 2 (right column).

$$\hat{f}_\theta = \sum_s \phi_s f_s \tag{5}$$

D. Probabilistic Trajectory Prediction via Bidirectional LSTM

Due to their capabilities in modeling complex temporal dependency of their input sequence information, Recurrent neural networks (RNN) have been achieving resilient results in sequence prediction tasks [13], [25], [26]. Thus, in our proposed framework for trajectory prediction of pedestrians in urban traffic environment, we will be capitalizing on their powerful sequence-to-sequence modeling capabilities. In specific, we will be utilizing one variant of RNN, the bidirectional long short-term memory (B-LSTMs) architecture [27]. In general, the operation of conventional LSTM architecture is governed by three main internal gates which dictate which information to be persisted over time and which to be forgotten. Therefore, LSTMs are considered one of the best RNN architectures for memorizing longer-term information. The aforementioned conventional LSTMs are usually referred to as unidirectional LSTM (U-LSTM), that because they process the information in only one direction which is the forward direction. On the other hand, B-LSTM can process the information in two directions, namely forward and backward which make them more capable of understanding a much higher level of abstraction of their input information [27].

Both U-LSTM and B-LSTM architectures output predictions of deterministic real target values, however in real-life there is usually an inherent uncertainty especially with respect to our pedestrian trajectory prediction problem. As a result, we will be augmenting B-LSTM architecture with an output layer of a mixture density network (MDN) [20]. The MDN layer will generate a weighted sum of numerous probability distributions that can account for the uncertainty of pedestrian trajectories in urban traffic environments.

In Figure 3, the proposed B-LSTM-MDN for pedestrians trajectory prediction is shown. It is comprised of two stacked LSTM layers (LSTM-1 & LSTM-2) each with 64 hidden nodes. At the output layer, it outputs two weighted MDNs. The information flow of the forward and backward iterations over time is denoted by the forward and the backward arrows. Given an input a sample sequence $X = \{x_0,..,x_T\}$ of length T to our B-LSTM-MDN model. Where X is comprised of two main information, the trajectory of the pedestrian in 2D dimension ($x_{0:T}$,$y_{0:T}$) and the k-neighbor reward features at each position of this trajectory ($r^k_{0:T}$). Then, the output of the model is the probability distribution over the future trajectory Y of the pedestrian. As we mentioned before, the output h_t of every LSTM memory cell is controlled by three internal gates at each time step t which in the case of our B-LSTM-MDN model will have two of them. The $\overrightarrow{h_t}$ for the forward layer and $\overleftarrow{h_t}$ for the backward layer. In return, the final output y_t from each LSTM cell is as follows:

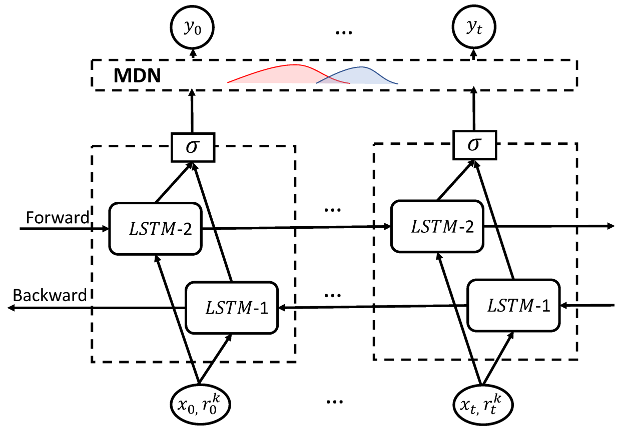

Figure 3. The probabilistic B-LSTM model for trajectory prediction of pedestrian in urban traffic environments.

$$y_t = \sigma(\overrightarrow{h_t}, \overleftarrow{h_t}), \tag{6}$$

where σ is a function to combine the outputs from the two inner LSTMs and it is in our model a concatenation function with the rectified linear unit (ReLU) as the activation layer.

For the MDN output layer, we chose the mixture of Gaussian as our probability density function (PDF), which is calculated as follows:

$$P(y_t|\mathcal{N}_t) = \sum_{m=1}^{M} \alpha_t^m \mathcal{N}(y_t|\mu_t^m, \sigma_t^m, \rho_t^m) \tag{7}$$

where y_t is the real target value, M the number of mixtures for the PDF of Gaussian which was two in our case, α_t^m is the weight for the m-th mixture and $\mathcal{N}$ is the normal Gaussian distribution.

Since the output values from our the B-LSTM model are real numbers, so transformations are needed before we use them as the parameters $\{\mu_t^m, \sigma_t^m, \rho_t^m\}$ for our normal distribution as follows:

$$\alpha_t^m = \frac{\exp(\tilde{\alpha}_t^m)}{\sum_{i=1}^{M}\exp(\tilde{\alpha}_i^m))}, \tag{8}$$

$$\sigma_t^m = \exp(\tilde{\sigma}_t^m) \tag{9}$$

$$\rho_t^m = \tanh(\tilde{\rho}_t^m) \tag{10}$$

where $\tilde{\alpha}_t^m$, $\tilde{\sigma}_t^m$ and $\tilde{\rho}_t^m$ are the PDF's weight, variance and the correlation values from the B-LSTM output layer of the m-th Gaussian mixture respectively.

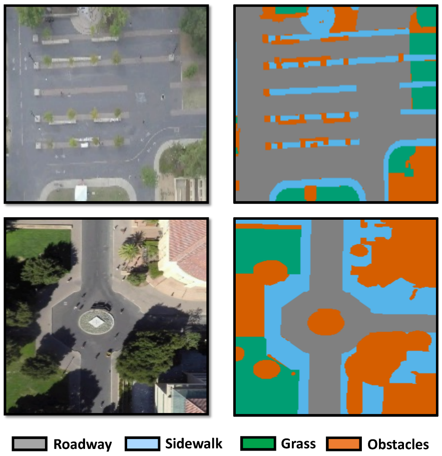

Figure 4. Semantic labels of two scenes from the SDD namely, bookstore ”top” and gates ”bottom”. The obstacles label is for any obstacle objects such as buildings or trees.

Eventually, the training of the B-LSTM-MDN model can be accomplished via minimizing the log likelihood of the normal Gaussian distribution against the input real-valued training data as follows:

$$L(X) = \sum_{t=1}^{T} -\log\left(\sum_{m=1}^{M} \alpha_t^m \mathcal{N}(y_t|\mu_t^m, \sigma_t^m, \rho_t^m)\right) \quad (11)$$

where T is the length of the input sequence, we empirically chose T to be of size 28 which corresponds to roughly 1 second of past trajectory of the VRUs. For optimizing the aforementioned loss function we used, the Adam optimizer with a learning rate of 0.005.

IV. Experiments

In this section, the data that has been utilized for training and testing our proposed framework for pedestrians’ trajectory prediction will be presented. Then, the performance of that framework will be evaluated against a number of evaluation metrics quantitatively. Additionally, it will be compared against a number of baseline models.

A. Dataset

Given the nature of the proposed framework which mainly relies on deep sequence prediction model (i.e, B-LSTM), the necessity for a relatively large amount of pedestrians’ trajectory in traffic environment is essential. Fortunately, recently the Stanford drone dataset (SDD), one of the largest datasets for agents’ behavior modeling has been made publicly available [28]. SDD was collected using a bird’s eye view camera mounted on a drone hovering over the vicinity of Stanford University campus. The dataset contains video images with frame by frame bounding-boxes annotations (at roughly frame rate of 28 FPS) for moving targets such as pedestrians, bikers, and cars. SDD was categorized into 8 scenes, each with a number of targets annotated videos. In our experiments, we focused on the scenes that had more number of pedestrians, which at the same time contain other static or dynamic objects similar to the ones found in urban traffic environments. These traffic objects are such as sidewalks, road/roundabouts, cars, grass, and buildings. Thus, we chose four scenes from the SDD for the training and testing of our framework, namely “bookstores, gates, deathCircle and little”. As a first preparation stage, for each pedestrian’s annotated bounding box coordinates over time in each scene, they were converted into a trajectory of (x, y) positions by calculating the bounding box’s center position.

B. Data Preparation for Training IRL MaxEnt

For the reward learning via IRL MaxEnt sub-system, we have further manually annotated the reference image for each scene from the four scenes with pixel-wise semantic labels. These semantic labels are the input features vector for the IRL MaxEnt as discussed in Section III-C. The number of pixel-wise semantic labels were scene specific but the common ones were: buildings, road, sidewalk and generic obstacles. Since the resolution of pixel-wise semantic label image for each scene is relatively large, so for tractable computation of the IRL MaxEnt algorithm, we resized all of the semantic label images of the four scenes into a size of (224×224). For training the IRL MaxEnt we used the entire pedestrian trajectories and semantic label images from each scene from the four scenes. In Figure 4, sample of the pixel-wise annotations is shown for two different scenes from the SDD.

C. Performance Evaluation and Discussion

For quantitatively evaluating the performance of our proposed framework for the pedestrians’ trajectory prediction problem, we adopted two different evaluation metrics. The first one is the average displacement error which was used in [13]. The average displacement error is essentially the averaged Euclidean distance between the future trajectory predicted and generated by our framework and the ground truth future trajectory over all the single steps of the pedestrians’ trajectories. The second metric is the Modified Hausdorff Distance (MHD) which was similarly adopted in [10]. MHD is used to evaluate the geometrical similarities between two non-linear sequences which in our case will be the

Table I

PERFORMANCE OF OUR PROPOSED FRAMEWORK (B-LSTM-MDN-REWARD) AGAINST A NUMBER OF BASELINE MODELS. OUR PROPOSED APPROACH WERE EVALUATED OVER TWO DIFFERENT PREDICTION HORIZONS (2 AND 3 SECS) OF THE PEDESTRIANS' TRAJECTORIES AND AGAINST TWO DIFFERENT EVALUATION METRICS. THE LOWER THE BETTER.

Approach	**2.0 (sec) Ahead**		**3.0 (sec) Ahead**	
	Avg. Disp. Error (pixels)	MHD (pixels)	Avg. Disp. Error (pixels)	MHD (pixels)
U-LSTM	12.12	10.96	15.16	13.48
U-LSTM-MDN	9.16	7.48	13.49	11.12
B-LSTM-MDN	8.13	6.48	11.29	8.94
U-LSTM-Reward	11.49	10.29	15.04	13.29
U-LSTM-MDN-Reward	3.22	1.93	4.35	**2.72**
B-LSTM-MDN-Reward (proposed)	**2.93**	**1.95**	**4.12**	2.90

predicted future trajectory from our framework and the future ground truth trajectory. It is worth noting, that as our framework predicts a probability distribution over each point of the future trajectory. Thus, we will use a random sampling technique to get the real numbers of the future predicted trajectories to evaluate it against the future ground truth trajectory. In order to further evaluate the performance of our framework, more specifically whether the learned reward map features had made an actual difference in the predicted trajectories from our B-LSTM-MDN model. In Table I, we compare against a number of variants of data-driven baseline models based on LSTM network over two different long-term future prediction horizons (2 seconds and 3 seconds ahead). The baseline models are:

- U-LSTM: traditional unidirectional stacked LSTM model similar to the one in [14], that relies only on the past trajectories of pedestrians in order to directly infer real-valued future trajectory.
- U/B-LSTM-MDN: unidirectional or bidirectional stacked LSTM network with MDN at the output layer (with the same layers as in Section III-D), that relies only on past trajectory positions to infer probability distributions over the future trajectory.
- U-LSTM-Reward: a traditional unidirectional stacked LSTM model but augmented by reward function along with the past future trajectories.
- U/B-LSTM-MDN-Reward: is the proposed probabilistic trajectory model described in Section III-D. In the case of U-LSTM-MDN-Reward, the LSTM layers are unidirectional instead of the bidirectional ones.

As it can be noticed from Table I, the proposed framework has outperformed all the other LSTM-based baseline models in terms of lowest average displacement errors and MHD. More specifically, the additional learned reward features were also proven to improve the performance of all the LSTM-based models that did not include it, namely (U-LSTM, U-LSTM-MDN and B-LSTM-MDN). Another observation is that the LSTM-based models with MDN output layer tend to be giving more accurate predictions in comparison to the LSTM model that was without it (i.e. U-LSTM). Moreover, the main proposed framework (B-LSTM-MDN-Reward), was also proved to be providing resilient results over two long-term prediction horizons (namely 2 and 3 seconds ahead).

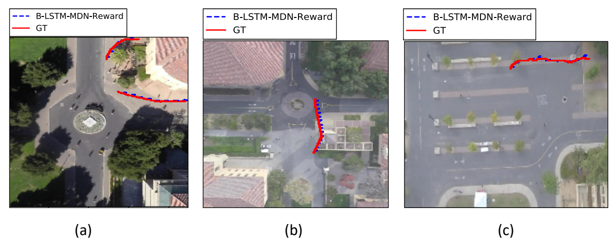

Figure 5. Qualitative sample predictions of our B-LSTM-MDN-Reward framework (dashed blue) against the ground truth trajectory (solid red) over three scenes of SDD, (a) gates, (b) "deathCircle", (c) "bookstore" .

For an additional qualitative evaluation of the predictions of our proposed framework. In Figure 5, some predicted trajectories of our proposed framework are plotted against the ground truth trajectories. As it can be shown, our framework can generate trajectories that are close enough to the ground truth trajectories. Moreover, the model can capture the non-linear motion pattern of the pedestrians in traffic environments while generating collision-free trajectories.

V. CONCLUSION

In this work, a framework for long-term prediction of pedestrians trajectories in urban traffic environment was proposed. Our proposed framework is based on a combination of planning-based models and sequence prediction models via inverse reinforcement learning (IRL) and deep recurrent neural networks. With the help of IRL, a reward function of the physical environment can be learned that perfectly capture the pedestrians preference in traffic environments. Then using the learned reward function alongside the motion trajectory of pedestrians in the environment we learn another RNN model that infer a long-term trajectory without a prior information about the end goal at inference time. We evaluated the proposed framework against two different evaluation metrics and in comparison to other baseline models. Our framework has shown a significant improvement over the baseline models in terms of lower average displacement errors and modified Hausdorff distance. Future directions would be investigating further factors or features to include within our reward function that could influence pedestrians behaviors in traffic environments. One example could be the interactions among a group of pedestrians such as family members pedestrians.

REFERENCES

[1] J. Ziegler, P. Bender, M. Schreiber, H. Lategahn, T. Strauss, C. Stiller, T. Dang, U. Franke, N. Appenrodt, C. G. Keller *et al.*, "Making bertha drivean autonomous journey on a historic route," *IEEE Intelligent Transportation Systems Magazine*, vol. 6, no. 2, pp. 8–20, 2014.

[2] B. Huval, T. Wang, S. Tandon, J. Kiske, W. Song, J. Pazhayampallil, M. Andriluka, P. Rajpurkar, T. Migimatsu, R. Cheng-Yue *et al.*, "An empirical evaluation of deep learning on highway driving," *arXiv preprint arXiv:1504.01716*, 2015.

[3] M. Mara and C. Lindinger, "Talking to the robocar - new research approaches to the interaction between human beings and mobility machines in the city of the future," pp. 86–91, 2015.

[4] M. Wagner and P. Koopman, "A philosophy for developing trust in self-driving cars," in *Road Vehicle Automation 2*. Springer, 2015, pp. 163–171.

[5] D. Rothenbücher, J. Li, D. Sirkin, B. Mok, and W. Ju, "Ghost driver: A field study investigating the interaction between pedestrians and driverless vehicles," in *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*. IEEE, 2016, pp. 795–802.

[6] N. Schneider and D. M. Gavrila, "Pedestrian path prediction with recursive Bayesian filters: A comparative study," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8142 LNCS, pp. 174–183, 2013.

[7] 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.

[8] 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.

[9] K. Saleh, M. Hossny, and S. Nahavandi, "Towards trusted autonomous vehicles from vulnerable road users perspective," in *Systems Conference (SysCon), 2017 Annual IEEE International*. IEEE, 2017, pp. 1–7.

[10] K. M. Kitani, B. D. Ziebart, J. A. Bagnell, and M. Hebert, "Activity forecasting," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7575 LNCS, no. PART 4, pp. 201–214, 2012.

[11] B. D. Ziebart, N. Ratliff, G. Gallagher, C. Mertz, K. Peterson, J. A. Bagnell, M. Hebert, A. K. Dey, and S. Srinivasa, "Planning-based prediction for pedestrians," *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2009*, pp. 3931–3936, 2009.

[12] V. Karasev, A. Ayvaci, B. Heisele, and S. Soatto, "Intent-Aware Long-Term Prediction of Pedestrian Motion," *International Conference on Robotics and Automation (ICRA)*, 2016.

[13] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social lstm: Human trajectory prediction in crowded spaces," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 961–971.

[14] K. Saleh, M. Hossny, and S. Nahavandi, "Intent prediction of vulnerable road users from motion trajectories using stacked lstm network," in *Intelligent Transportation Systems Conference (ITSC), 2017 IEEE International Conference on*. IEEE, 2017, pp. 327–332.

[15] E. A. Pool, J. F. Kooij, and D. M. Gavrila, "Using road topology to improve cyclist path prediction," in *Intelligent Vehicles Symposium (IV), 2017 IEEE*. IEEE, 2017, pp. 289–296.

[16] E. Rehder and H. Kloeden, "Goal-Directed Pedestrian Prediction," *IEEE International Conference on Computer Vision Workshops*, 2015.

[17] T. Bandyopadhyay, C. Z. Jie, D. Hsu, H. Marcelo, A. Jr, D. Rus, and E. Frazzoli, "Intention-Aware Pedestrian Avoidance," *The 13th International Symposium on Experimental Robotics*, pp. 963–977, 2013.

[18] B. Völz, K. Behrendt, H. Mielenz, I. Gilitschenski, R. Siegwart, and J. Nieto, "A data-driven approach for pedestrian intention estimation," in *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*. IEEE, 2016, pp. 2607–2612.

[19] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[20] C. M. Bishop, "Mixture density networks," Citeseer, Tech. Rep., 1994.

[21] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey, "Maximum entropy inverse reinforcement learning." in *AAAI*, 2008, pp. 1433–1438.

[22] R. Bellman, "A markovian decision process," DTIC Document, Tech. Rep., 1957.

[23] M. Kuderer, H. Kretzschmar, C. Sprunk, and W. Burgard, "Feature-Based Prediction of Trajectories for Socially Compliant Navigation," *Proceedings of Robotics: Science and Systems*, 2012.

[24] W.-C. Ma, D.-A. Huang, N. Lee, and K. M. Kitani, "A Game-Theoretic Approach to Multi-Pedestrian Activity Forecasting," vol. 1, no. c, 2016. [Online]. Available: http://arxiv.org/abs/1604.01431

[25] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in neural information processing systems*, 2014, pp. 3104–3112.

[26] Y. Zhao, R. Yang, G. Chevalier, R. C. Shah, and R. Romijnders, "Applying deep bidirectional lstm and mixture density network for basketball trajectory prediction," *Optik-International Journal for Light and Electron Optics*, vol. 158, pp. 266–272, 2018.

[27] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

[28] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, "Learning social etiquette: Human trajectory understanding in crowded scenes," in *European conference on computer vision*. Springer, 2016, pp. 549–565.