Self-learning Trajectory Prediction with Recurrent Neural Networks at Intelligent Intersections

Julian Bock[1], Till Beemelmanns[1], Markus Klösges[1] and Jens Kotte[2]
[1]*Institute for Automotive Engineering, RWTH Aachen University, Aachen, Germany*
[2]*Forschungsgesellschaft Kraftfahrwesen mbH, Aachen, Germany*

Keywords: Neural Networks, Machine Learning, Prediction, Intersection, V2x.

Abstract: We present the concept and first results of a self-learning system for road user trajectory prediction at intersections with connected sensors. Infrastructure installed connected sensors can assist automated vehicles in perceiving the environment in complex urban scenes such as intersections. An intelligent intersection with connected sensors can measure the trajectories of road users using multiple sensor types and store the trajectories. Our approach uses this information to collect a large dataset of pedestrian trajectories. This dataset is again used to train a pedestrian prediction model with Recurrent Neural Networks. This model learns intersection specific pedestrian movement patterns. Through a self-learning process enabled by the measurements of connected sensors, the system continuously improves the prediction during operation while keeping the dataset preferably small. In this paper, we focus on the prediction of pedestrian trajectories, but as the approach is data-driven, the system could also predict other road users such as vehicles or bicyclists if trained with the respective data.

1 INTRODUCTION

As automated driving and advanced driver assistance systems will play a more and more important role, anticipating the pedestrian's future movements is a valuable task for improving road safety and trajectory planning (Brouwer et al., 2016; Keller and Gavrila, 2014). Human movement patterns are often uncertain and depend on many individual influencing factors. The movement of pedestrians is highly dynamic and especially urban scenes require an accurate prediction (Schneider and Gavrila, 2013) Due to the different driving directions and diverse road participants, intersections are among the most complex scenarios for automated driving. Thus, it is a challenging task to design a model that is able to forecast future movements of pedestrian at intersections with long time horizons.

For short-time pedestrian predictions, the head orientation and arm movement are highly relevant characteristics, while long-time predictions are rather goal oriented (Rehder and Kloeden, 2015). Due to their high relative velocity, vehicles typically monitor a pedestrian only for a short time making it hard to interfere the goal of the pedestrian. At an intersection, connected stationary sensors allow long-time observations of the pedestrian movement. Those sensors not just allow a longer observation of a single pedestrian, but also can be used for the generation of datasets with large amounts of historical data. Using this pedestrian historical data, position-dependent movement patterns of this intersection can be learned with machine learning approaches to implicitly model pedestrian goals.

However, the position-dependent movement patterns might change over time or rare situations need to be considered, which are not be sufficiently covered in the learned model. As the infrastructure sensors are permanently monitoring the intersection, new measurement data is permanently created. Our proposed self-learning system is intended to permanently improve the prediction by re-training the models with these new measurements. In this re-training process, the prediction error is considered. The error of every prediction can easily be calculated by comparing the prediction with the measurement after the prediction time. With this continuous re-training process, rare situations can be incorporated and the prediction can adapt to changes at the intersection.

Bock, J., Beemelmanns, T., Klösges, M. and Kotte, J.
Self-learning Trajectory Prediction with Recurrent Neural Networks at Intelligent Intersections.
DOI: 10.5220/0006374003460351
In *Proceedings of the 3rd International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2017)*, pages 346-351
ISBN: 978-989-758-242-4

2 RELATED WORK

Our system builds upon a structure of locally connected sensors together with a computing unit. Thus, a short overview on the use of infrastructure sensors is given first. The data collected by the infrastructure sensors is processed with machine learning algorithms for temporal data. This leads to the current state of research on recurrent neural networks with focus on sequence prediction. Although the data-driven method makes it exchangeable to any other type of road users, the current state of pedestrian prediction models is provided. Finally, approaches on continuous learning are presented.

2.1 Infrastructure Sensors

Research projects such as Ko-FAS (Wertheimer and others, 2014), I2EASE (I2EASE Consortium, 2016) and the SADA (Consortium of Project SADA) examine the use of infrastructure installed sensors for e.g. cooperative perception. In I2EASE, the information from stationary infrastructure sensors as well as moving sensors carried by any road user are sent via X2X communication to a central intersection computer. The information from all sensors at the intersection are combined to a fused list of road user objects. Project SADA is working on a similar approach to fuse the data from any sensor.

Just recently (The City of San Diego, 09.03.2017) announced the installation of 3200 connected smart sensors in the urban area of San Diego. At least a part of those sensors will be installed at intersections. This shows that networked infrastructure sensors are a real future scenario.

2.2 Recurrent Neural Networks

Neural networks have shown to be capable of solving many tasks superior to previous methods. While feed forward neural networks are heavily applied on data without time dependency, recurrent neural networks are good for time dependent data through their possibility to save information from previous steps inside the network. In the early 90s, (Hochreiter, 01.01.1991) has shown, that there are issues on long-time dependency information with recurrent neural networks. A few years later, (Hochreiter and Schmidhuber, 1997) proposed a method called long short term memory (LSTM) extending RNNs for handling longer sequence information.

Learning tasks, which take a sequence of information over time as input with the goal to generate again a sequence of information over time, are called “Sequence to Sequence” learning. LSTMs have shown to successfully solve Sequence-to-Sequence problems for several problems such as language translation (Sutskever et al.) and handwriting speech recognition (Graves et al., 2013).

(Graves, 2014) used LSTMs for sequence generating tasks with complex long-ranged time dependencies. The LSTMs are trained with handwriting sequences based on tracked pen-tip trajectories. (Graves, 2014) showed, that the trained model can generate handwriting samples or computing the probability distribution of future pen tip locations. This approach was one of the first attempts of training X-Y positional data with a RNN-LSTM inspiring the application of this methodology to different problems and datasets.

2.3 Pedestrian Prediction

The prediction of pedestrian movement has been studied for quite some time. Already in the early 90s, first pedestrian models inspired by physical gas-kinetics were developed (Helbing, 1990). One of the first social-forces models was introduced by (Helbing and Molnar, 1995). This model describes the social-forces similar to energy potentials based on inter alia the distance to other pedestrians with respect to the sphere of privacy and an attraction effect. Helbing et al. demonstrated in computer simulations, that their model describes nonlinear interactions of pedestrians.

Furthermore, a recent publication (Brouwer et al., 2016) introduced four categories of models for pedestrian prediction, providing an overview of existing work:

• *Class I: Dynamics based pedestrian model*

These models are using dynamic information about the pedestrians such as position, velocity and moving direction.

• *Class II: Pedestrian model using physiological knowledge*

These models consider physiological constraints of the pedestrian such as the human’s capabilities to accelerate and change moving direction.

• *Class III: Pedestrian model using pedestrian head orientation information*

These models are using head orientation information of the pedestrian, which is being an important indicator, whether a pedestrian will cross the street.

• *Class IV: Pedestrian model using environment information*

These models rather focus on the environmental influence than just information about the pedestrian itself.

However, the pedestrian prediction problem cannot be considered as solved. Many of the pedestrian prediction models since 1990 are handcrafted models. In the last few years, more data-driven models were presented.

(Alahi et al.) proposed a data driven model for predicting human movement with a so called "Social-LSTM". In their approach, they considered that the movement of a person in a crowded scenario is usually influenced by its direct neighbours. In contrast to other social models they did not use handcrafted social forces functions, but designed a new end-to-end learning architecture that allows an interaction between spatially proximal sequences through a pooling layer. The pooling layer ensures that a LSTM cell has access to the hidden-states of all other LSTMs in a specific radius and this information is used for the prediction of the next time step. The model is evaluated on publicly available pedestrian tracking datasets showing that their model can anticipate future movements of individuals caused by social interactions among them.

2.4 Continuous Learning

Neural networks are commonly trained on fixed datasets. In the machine learning research community, static publicly available datasets are used to compare the performance of different net architectures and to develop new models (Pellegrini et al., 2009; Robicquet et al.). However, the human brain learns continuously something new, since we live in a permanently changing world (Käding et al.). There are some studies on neural nets that are trained with ongoing partially changing or growing datasets.

(Xiao et al., 2014) considered a convolutional neural network for image classification with an incrementally increasing dataset. In their approach, an algorithm hierarchically expands the convolutional network leading to bigger net capacities.

A study on continuously learning neural nets with incoming data streams was performed by (Käding et al.). The researchers investigated the impact of training parameters for newly added image data and their corresponding labels. They found out, that the effort of retraining a neural net with new data can be decreased by reducing the number of weight update iterations. Furthermore, (Käding et al.) state that neglecting already known data during retraining leads to overfitting of the new added data. Thus, robust retraining in a continuous fashion should be performed with a fraction of new and old data.

3 RESEARCH APPROACH AND METHODOLOGY

The analysis of related work shows a large potential in the prediction of road users using recurrent neural networks. Furthermore, first planned installations show that an intersection with several connected sensors measuring the positions of road users is a possible future scenario. We assume, that the combination of recent machine learning approaches such as LSTMs together with the on-going measurement of connected stationary sensors, can lead to highly accurate predictions in a local area

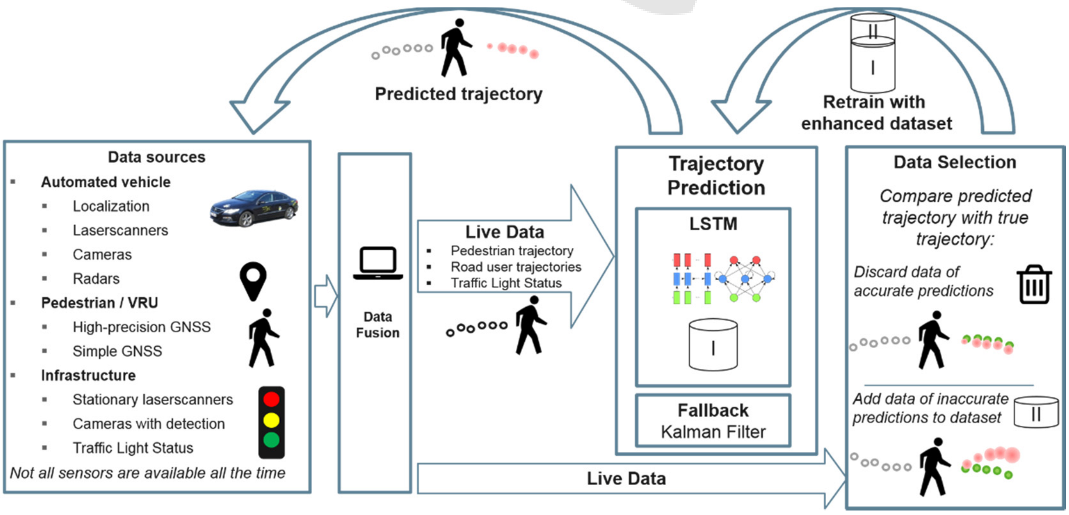

Figure 1: Continuously learning trajectory prediction system.

covered by the sensors. The accuracy can continuously improve through a self-learning process. Thus, we present here our concept for a self-learning trajectory prediction using LSTMs.

Our concept is illustrated in Figure 1. The proposed system can consider any type of sensor, which is measuring the positions of pedestrians. An automated vehicle could provide object lists from typical sensors such as cameras, radars and laserscanners. Pedestrians carrying devices with localization capabilities could provide their position. Infrastructure sensors such as laserscanners and cameras could provide object lists. The information is then sent via cellular X2X or 802.11p X2X to a local intersection computer, which is collecting all sensor data. This data is then fused to one description of the intersection in form of an object list.

For an initial phase, the object lists are collected at the local intersection computer to create a first dataset. This dataset does not need to be based on a consistent sensor setup if an object list is provided with accuracies. This first dataset A is then used to train an initial LSTM model LSTM.A. The model learns to generate a sequence Y with fixed length using the input sequence X with fixed length. As soon as the training is finished, LSTM.A can be applied to predict the future movement from a measured sequence.

As the measurements at the intersection are ongoing, the true trajectory of the pedestrian can be recorded. This data can be used to enlarge the dataset for the training of the LSTM model. However, due to the continuous measurements, the dataset would become exceedingly large if storing all positions permanently. Thus, as soon as the prediction time has passed, the predicted trajectory is compared with the true trajectory giving an error measure for the prediction. If the prediction error exceeds a certain threshold, the trajectory is stored in a second dataset B. After a certain time or size of dataset B, this dataset B is then used for continued training based on LSTM.A creating a LSTM.B model. From this moment on, LSTM.B is applied in the regular prediction of trajectories and the process is repeated as just described.

Finally, our prediction concept considers a fallback on a Kalman Filter. Models learned with machine learning methods can provide very bad results if a new situation is not sufficiently covered in the training dataset and the model does not generalize well on the data. Our proposed system shall detect, whether the new input data is sufficiently covered by the trained model and if not, the prediction of the Kalman Filter is output by the system.

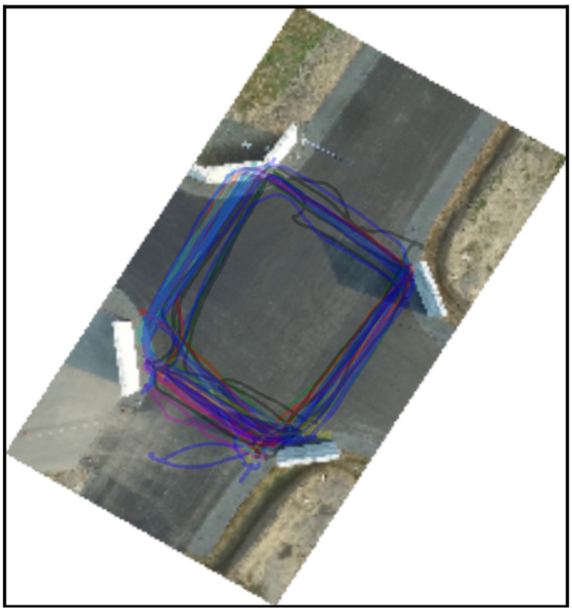

Figure 2: GNSS dataset on research intersection.

4 RESULTS

We implemented a LSTM learning system using the framework Keras (Chollet, 2015) with Theano (Al-Rfou et al., 2016) backend. For LSTM training, the data is normed to a zero mean and unit variance. As the settings in (Alahi et al.), we use a trajectory observation time of 3.2 secs and prediction time of 4.8 secs at a sampling rate of 2.5 Hz. At the selected sampling rate of 2.5 Hz, this results in 8 observation points and 12 prediction points. We use the mean displacement error (1) and final displacement error (2) as criteria for the evaluation of the prediction. The mean displacement error calculates the average Euclidean distance between the predicted and measured position over all time steps. The final displacement error calculates the Euclidean distance between the final predicted and measured position.

$$L_{MD}(\hat{Y},Y) = \frac{1}{T}\sum_{i=1}^{T}\sqrt{(x_i^N - \hat{x}_i^N)^2 + (y_i^N - \hat{y}_i^N)^2} \quad (1)$$

$$L_{FD}(\hat{Y},Y) = \sqrt{(x_T^N - \hat{x}_T^N)^2 + (y_T^N - \hat{y}_T^N)^2} \quad (2)$$

For first experiments, we created an own dataset with a high-precision pedestrian GNSS positioning device on a research intersection. This dataset consists of 9980 positions with 2.5 Hz measurement rate. The dataset contains the movement of a single pedestrian walking a distance of about 3.4 km. The trajectories are visualized in Figure 2.

The results are compared to a Kalman Filter with constant velocity model. Table 1 shows the results achieved on the GNSS pedestrian dataset. On both measurements, mean displacement error and final displacement error, the LSTM model achieves error rates similar to those in (Alahi et al.) and surpasses the Kalman Filter.

Table 1: Results on GNSS pedestrian dataset.

Model	L_{MD}	L_{FD}
Kalman Filter	0.674	1.257
LSTM model	**0.520**	**1.032**

However, for the use of machine learning methods, this dataset is quite small and by that not appropriate for investigations on continuous learning. Thus, we created two datasets with simulated pedestrian movements using the traffic simulation software PTV VISSIM. Both datasets are created using the same intersection model with the only difference, that in the second dataset one part of the pedestrian walkway path is changed to a small curve. This shall depict the change of the pedestrian path due to an obstacle.

We evaluated based on mean displacement error for the predicted trajectories depending on the percentage of additional training data from the second dataset. Our results show that with just the initial dataset A, predicted trajectories on a split of dataset B have high errors of more than 1.1 m. Adding five percent of a disjoint split of dataset B already reduces the error significantly to less than 0.1m. Further adding data is still reducing the error but reaches saturation at about 40% of the dataset. This shows, that our method can adapt to changes in the intersection's structure and only a part of the new data is needed for the major error reduction.

5 CONCLUSIONS

A system for self-learning pedestrian trajectory prediction using LSTMs is introduced and developed. The system relies on continuous measurements of pedestrian's positions at an intersection using connected sensors. The system can learn local pedestrian movement patterns at the intersection. The mean prediction error is continuously reduced by training the LSTMs again with additional data from new measurements. The approach is independent of the used sensor and can be applied to any road user, for which the sensors deliver position measurements over time.

Our results show, that the LSTM prediction model is superior to a constant velocity Kalman Filter for pedestrian prediction even on small datasets. We showed that the prediction model can adapt to changes in the pedestrian walking path using only a small part of the new data. By that, the size of the dataset can be kept rather small although depicting the pedestrian's movement patterns.

However, as we are currently missing a large real dataset, our results still mainly rely on simulated data. Thus, an important focus for future work lies in the creation of real datasets. There are several requirements on the datasets. First, the datasets need to be larger with a duration of several hours. Second, the datasets need to contain information about the pedestrian's dynamic environment and the traffic light status. Third, the dataset must be a measurement from an intersection in real traffic.

For the core prediction model, it is planned to consider the pedestrians' dynamic environment as well as traffic light information as input for the prediction. The self-learning process also needs to be further enhanced with the goal to build up a good compromise between adaptability and stability. Finally, the prediction shall provide an area with presence probability distribution depending on the certainty of the measured trajectory.

ACKNOWLEDGEMENTS

This paper is part of the work in the project I2EASE funded by the German Federal Ministry of Education and Research based on a decision of the Deutsche Bundestag.

REFERENCES

Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L. and Savarese, S. 'Social LSTM: Human Trajectory Prediction in Crowded Spaces' [online] http://vision.stanford.edu/pdf/CVPR16_N_LSTM.pdf.

Al-Rfou, R., Alain, G., Almahairi, A., Angermueller, C., Bahdanau, D., Ballas, N., Bastien, F., Bayer, J., Belikov, A., Belopolsky, A., Bengio, Y., Bergeron, A., Bergstra, J., Bisson, V., Snyder, J.B., Bouchard, N., Boulanger-Lewandowski, N., Bouthillier, X., Brébisson, A.d., Breuleux, O., Carrier, P.-L., Cho, K., Chorowski, J., Christiano, P., Cooijmans, T., Côté, M.-A., Côté, M., Courville, A., Dauphin, Y.N., Delalleau, O., Demouth, J., Desjardins, G., Dieleman, S., Dinh, L., Ducoffe, M., Dumoulin, V., Kahou, S.E., Erhan, D., Fan, Z., Firat, O., Germain, M., Glorot, X., Goodfellow, I., Graham, M., Gulcehre, C., Hamel, P., Harlouchet, I., Heng, J.-P., Hidasi, B., Honari, S., Jain, A., Jean, S., Jia, K., Korobov, M., Kulkarni, V., Lamb, A., Lamblin, P., Larsen, E., Laurent, C., Lee, S., Lefrancois, S., Lemieux, S., Léonard, N., Lin, Z., Livezey, J.A., Lorenz, C., Lowin, J., Ma, Q., Manzagol, P.-A., Mastropietro, O., McGibbon, R.T., Memisevic, R., van Merriënboer, B., Michalski, V., Mirza, M., Orlandi, A., Pal, C., Pascanu, R., Pezeshki, M., Raffel, C., Renshaw, D., Rocklin, M., Romero, A., Roth, M., Sadowski, P., Salvatier, J., Savard, F., Schlüter, J., Schulman, J., Schwartz, G., Serban, I.V., Serdyuk, D., Shabanian, S., Simon, É., Spieckermann, S., Subramanyam, S.R., Sygnowski, J., Tanguay, J., van Tulder, G., Turian, J., Urban, S., Vincent, P., Visin, F., Vries, H.d., Warde-Farley, D., Webb, D.J., Willson, M., Xu, K., Xue, L., Yao, L., Zhang, S. and Zhang, Y. (2016) 'Theano: A Python framework for fast computation of mathematical expressions', *arXiv e-prints*, abs/1605.02688.

Brouwer, N., Kloeden, H. and Stiller, C. (2016) 'Comparison and evaluation of pedestrian motion models for vehicle safety systems', pp.2207–2212.

Chollet, F. (2015) 'keras', *GitHub repository*.

Consortium of Project SADA 'Project SADA'. http://www.projekt-sada.de/ (Accessed 26 January 2016).

Graves, A. (2014) *Generating Sequences With Recurrent Neural Networks*. http://arxiv.org/pdf/1308.0850v5.

Graves, A., Mohamed, A.-r. and Hinton, G. (2013) 'Speech recognition with deep recurrent neural networks', IEEE, pp.6645–6649.

Helbing, D. (1990) 'Physikalische Modellierung des dynamischen Verhaltens von Fußgängern (Physical Modeling of the Dynamic Behavior of Pedestrians)'.

Helbing, D. and Molnar, P. (1995) 'Social force model for pedestrian dynamics', *Physical review E*, Vol. 51, No. 5, p.4282.

Hochreiter, S. (01.01.1991) *Untersuchungen zu dynamischen neuronalen Netzen,* diploma thesis, institut für informatik, lehrstuhl prof. brauer, technische universität münchen.

Hochreiter, S. and Schmidhuber, J. (1997) 'Long short-term memory', *Neural computation*, Vol. 9, No. 8, pp.1735–1780.

I2EASE Consortium (2016) 'I2EASE Project'. http://www.cerm.rwth-aachen.de/i2ease (Accessed 26 January 2017).

Käding, C., Rodner, E., Freytag, A. and Denzler, J. *Active and Continuous Exploration with Deep Neural Networks and Expected Model Output Changes*. http://arxiv.org/pdf/1612.06129v1.

Keller, C.G. and Gavrila, D.M. (2014) 'Will the Pedestrian Cross? A Study on Pedestrian Path Prediction', *IEEE Transactions on Intelligent Transportation Systems*, Vol. 15, No. 2 [online] http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6632960.

Pellegrini, S., Ess, A., Schindler, K. and van Gool, L. (2009) 'You'll never walk alone: Modeling social behavior for multi-target tracking', pp.261–268.

Rehder, E. and Kloeden, H. (2015) 'Goal-directed pedestrian prediction', pp.50–58.

Robicquet, A., Sadeghian, A., Alahi, A. and Savarese, S. 'Learning Social Etiquette: Human Trajectory Understanding In Crowded Scenes', in , *Computer Vision – ECCV 2016*.

Schneider, N. and Gavrila, D.M. (2013) 'Pedestrian path prediction with recursive Bayesian filters: A comparative study', Springer, pp.174–183.

Sutskever, I., Vinyals, O. and Le V, Q. *Sequence to Sequence Learning with Neural Networks*. http://arxiv.org/pdf/1409.3215v3.

The City of San Diego (09.03.2017) 'Smart City San Diego'. https://www.sandiego.gov/sustainability/smart-city.

Wertheimer, R. and others (2014) 'Ko-PER Fahrerassistenz und präventive Sicherheit mittels kooperativer Perzeption: Partnerübergreifender Schlussbericht', *schlussbericht, Bundesministerium für Wirtschaft und Technologie (BMWi)*.

Xiao, T., Zhang, J., Yang, K., Peng, Y. and Zhang, Z. (2014) 'Error-driven incremental learning in deep convolutional neural network for large-scale image classification', ACM, pp.177–186.