Predicting vehicle behaviour using LSTMs and a vector power representation for spatial positions

Florian Mirus, Peter Blouw, Terrence C. Stewart and Jörg Conradt
florian.mirus@bmwgroup.com, peter.blouw@appliedbrainresearch.com,
terry.stewart@appliedbrainresearch.com, jconradt@kth.se

Introduction

In this paper, we expand our previous work [1] on an automotive environment model based on Vector Symbolic Architectures (VSAs). Here, our main contribution is an encoding of spatial information for multiple objects in semantic scene vectors of fixed length. We hypothesize that this structured vector representation will be able to capture relations and mutual influence between traffic participants. For prediction the vehicle's future positions, we train a Long Short-Term Memory (LSTM) network using our vector-representation as well as other encoding schemes of the input data and compare their performance against each other as well as against a simple linear model based on a constant velocity assumption.

System architecture

We use a network-architecture consisting of one LSTM encoder and decoder cell with 150 hidden states each, for sequence to sequence prediction.

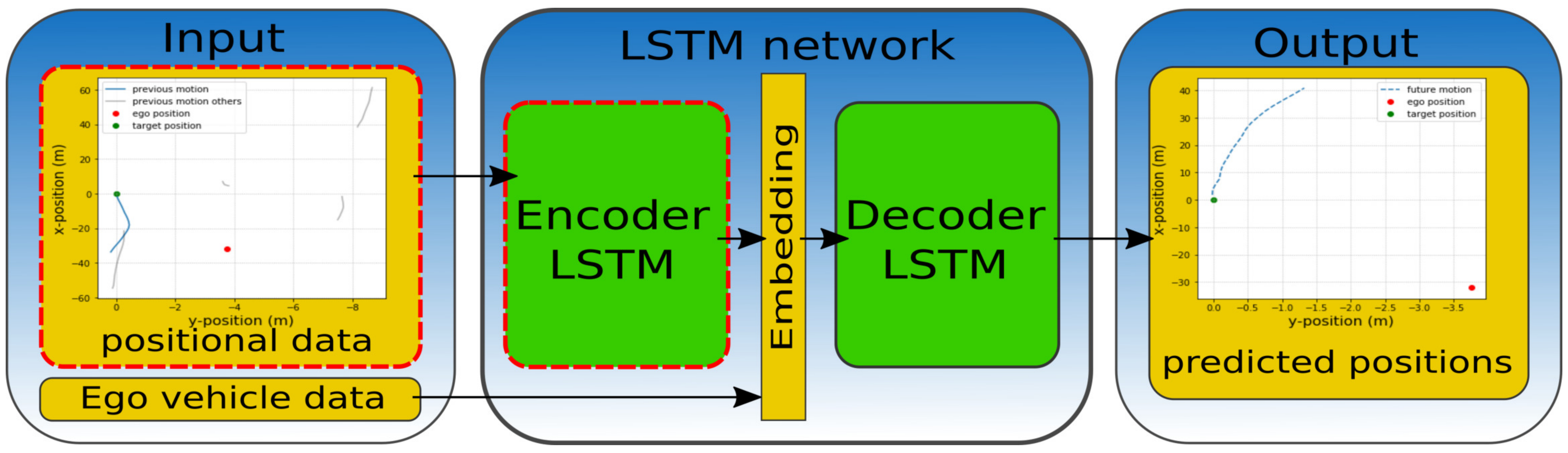

Figure 1: Visualization of our learning architecture. Modules that change with varying encoding of the input data are highlighted through dashed red borders.

Vector representation

We assign a random ID-vector to each category of dynamic objects (e.g. car, motorcycle, truck) as well as random unitary vectors $\mathbf{X}$ and $\mathbf{Y}$ to encode spatial positions using the convolutive vector power [2] defined as

$$\mathbf{v}^p := \Re\left(IDFT\left(\left(DFT_j\left(\mathbf{v}\right)^p\right)_{j=0}^{D-1}\right)\right), \tag{1}$$

which results in the following vector representation of an automotive situation:

$$\mathbf{S}_t = \mathbf{THIS} \otimes \mathbf{TYPE}_{target} \otimes \mathbf{X}^{x_t} \otimes \mathbf{Y}^{y_t} \oplus \sum_{obj} \mathbf{TYPE}_{obj} \otimes \mathbf{X}^{x_{obj,t}} \otimes \mathbf{Y}^{y_{obj,t}}. \tag{2}$$

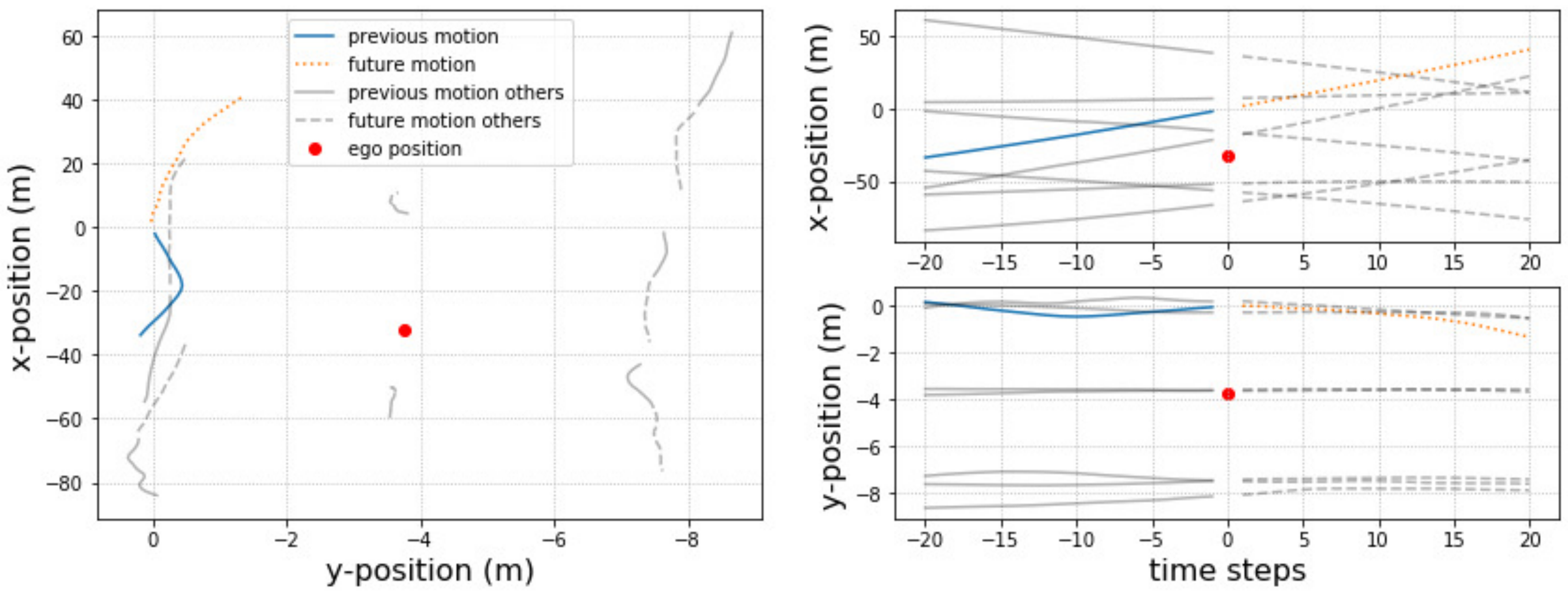

Figure 2: Data visualization of one driving situation example. The red dot indicates the position of the ego vehicle, blue and orange lines show past and future motion of the target vehicle whereas gray lines depict the other vehicles' motion.

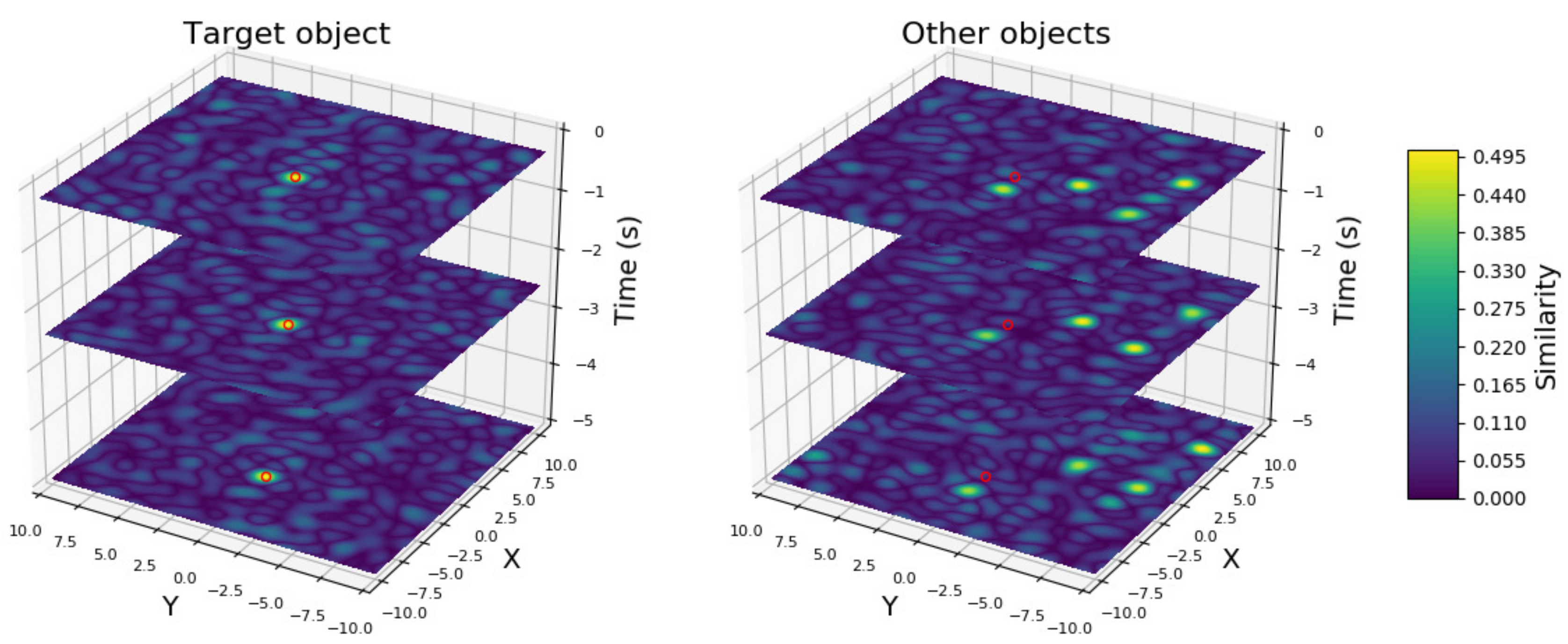

Figure 3: Visualization of the convolutive vector-power representation over time as a heat map. The red circles indicate the measured position of the target vehicle.

Experiments

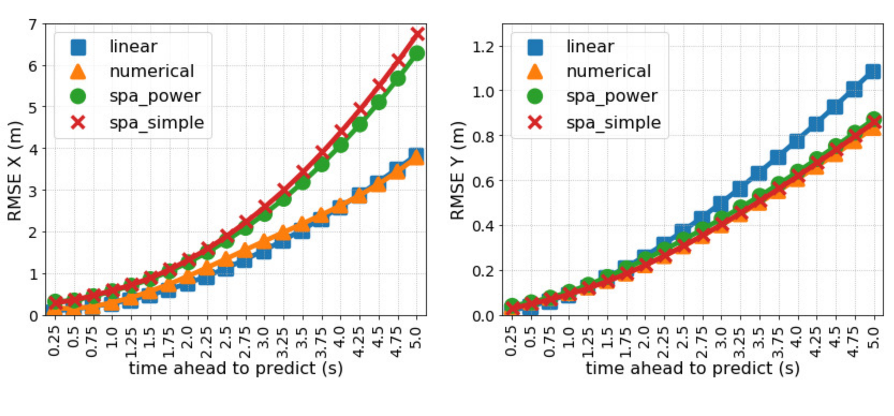

Figure 4: Validation-set V_1 complete

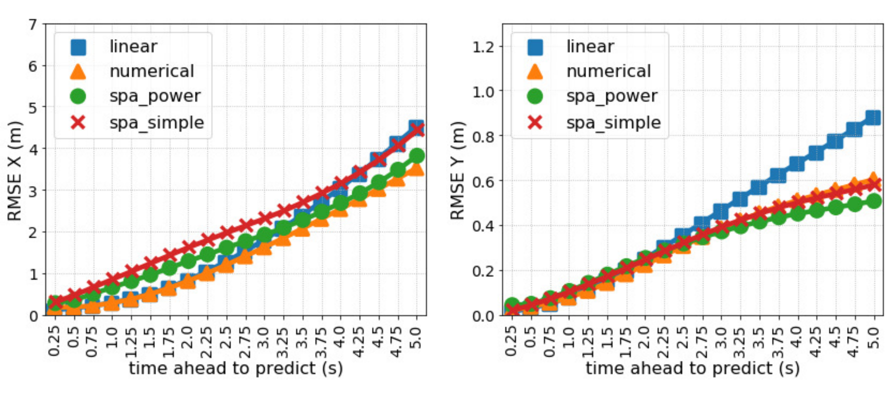

Figure 5: Validation-set V_1 subset

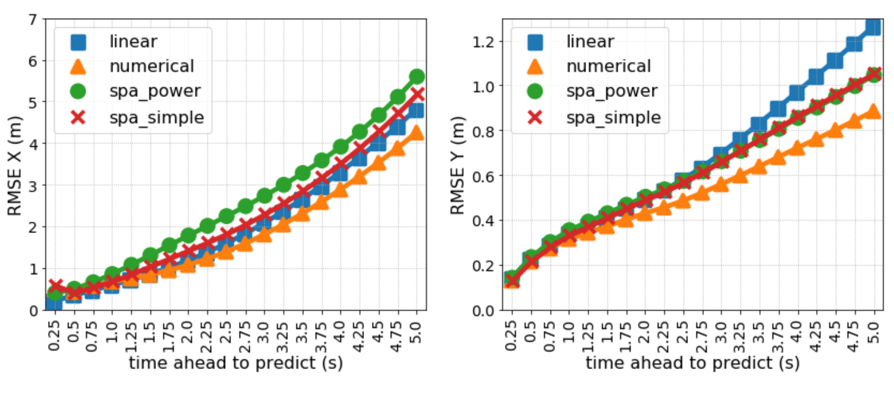

Figure 6: Validation-set V_2 complete

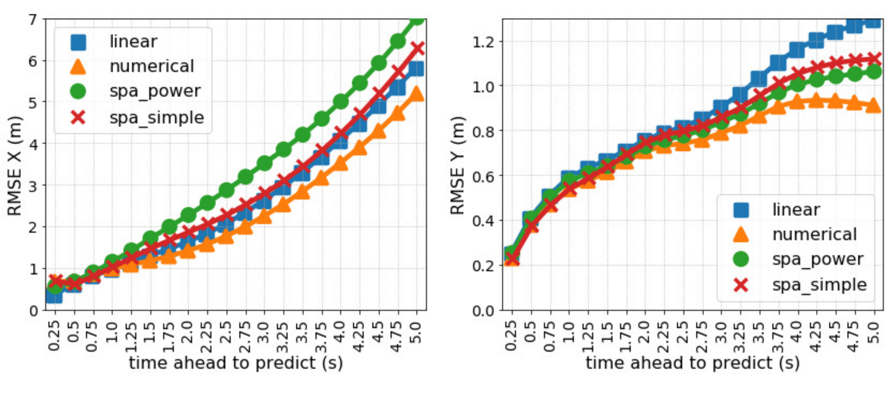

Figure 7: Validation-set V_2 subset

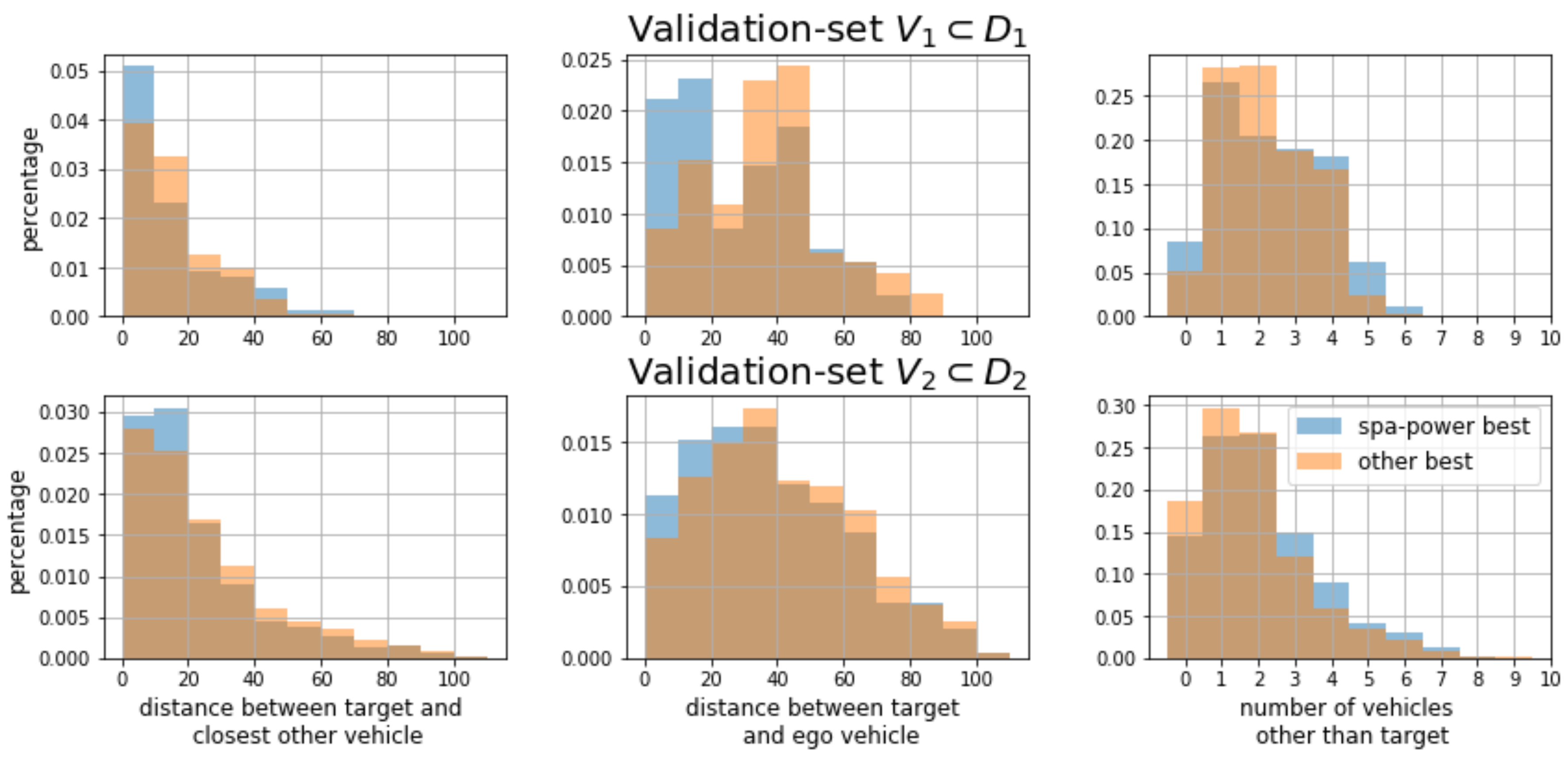

Figure 8: Metric evaluation based on the RMSE in y-direction regarding situations where the model using Semantic Pointer Architecture (SPA)-power representation outperforms all other approaches.

References

[1] F. Mirus, T. C. Stewart, and J. Conradt. "Towards cognitive automotive environment modelling: reasoning based on vector representations". In: *26th European Symposium on Artificial Neural Networks, ESANN 2018, Bruges, Belgium.* 2018-04-25, pp. 55–60.

[2] T. Plate. "Distributed Representations and Nested Compositional Structure". PhD thesis. University of Toronto, 1994.