ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2. Available from http://www.i6doc.com/en/.

Detection of abnormal driving situations using distributed representations and unsupervised learning

Florian Mirus[1,2], Terrence C. Stewart [3] and Jörg Conradt[4]

1- BMW Group - Research, New Technologies, Innovations, Garching, Germany
2- Technical University of Munich - Department of Electrical and Computer Engineering, Munich, Germany
3- Applied Brain Research Inc., Waterloo, Ontario, Canada
4- KTH Royal Institute of Technology - Department of Computational Science and Technology, Stockholm, Sweden

Abstract. In this paper, we present an anomaly detection system employing an unsupervised learning model trained on the information encapsulated within distributed vector representations of automotive scenes. Our representations allows us to encode automotive scenes with a varying number of traffic participants in a vector of fixed length. We train a neural network autoencoder in unsupervised fashion to detect anomalies based on this representation. We demonstrate the usefulness of our approach through a quantitative analysis on two real-world data-sets.

1 Introduction

In this paper, we investigate the information encapsulated within distributed vector representations of automotive scenes and if they can be used to detect potentially dangerous driving situations from just the vector representation. Vector Symbolic Architectures (VSAs) have an intrinsic mechanism of comparing vectors through the dot product. However, it is not clear if simply comparing vectors in terms of similarity to, for instance, the mean pairwise-similarity of all known vectors or a subset of vectors considered "normal" will sufficiently differentiate outliers from the "normal data". A-priori, it might not even be clear what vectors belong to the baseline set of normal data or how to define vectors to be considered as inliers. One option could be to manually define metrics such as the number of vehicles in the scene or a threshold for the distance between the vehicles to detect crowded and potentially dangerous situations. However, such an approach suffers from the typical issues of manual engineering such as biases introduced by the human designer as well as poor scaling. Therefore, we employ an unsupervised learning approach based on fully-connected autoencoder neural networks similar to the one proposed by Chen et al. [1].

Related Work: Anomaly or outlier detection is an important research field that has been investigated across a variety of application domains using different approaches [2]. In automotive context, the main application domain beside production plant diagnosis [3] is the detection of abnormal driving situations. In contrast to our work, most of the current approaches perform anomaly detection based on camera images from the driver's perspective (instead of our abstract vectors) trying to detect unusual patterns [4, 5] or directly accidents [6]. Finally, Chandola et al. [2] present a general overview of different algorithmic approaches and other applications of anomaly detection.

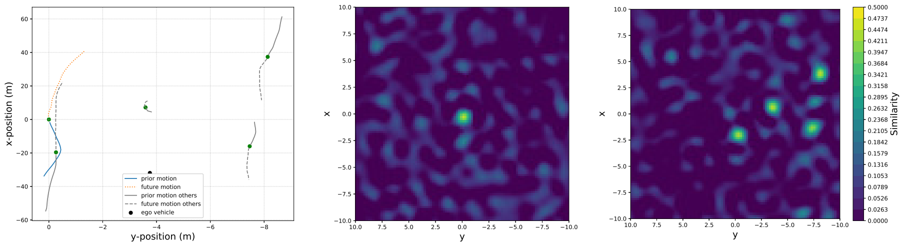

Fig. 1: Visualization of the convolutive power representation for 512-dimensional vectors. The left plot depicts a scene from the *On-board* data-set, while the middle and right plots visualize the similarity between the representation vector of that scene and auxiliary comparison vectors created from a sequence of discrete values as heat map for the target vehicle (middle) and other cars (right).

2 Methods

Data and Preprocessing: In this work, we use two different data sets for training and evaluation of our system, which we refer to as *On-board* or D_1, which is a proprietary data set containing real-world data gathered during test drives mainly on highways in southern Germany, and *Next Generation Simulation (NGSIM)* or D_2, which is a publicly available data set recorded using external cameras observing a segment of approximately 640 m length with 6 lanes on the US-101 freeway in Los Angeles, California. Both data sets contain object-lists with a variety of features such as position, velocity and acceleration but also object type probabilities and lane information. The *On-board* data set contains driving data from 3891 vehicles, whereas the *NGSIM* data set contains 5930 vehicles. For training and evaluating our model, we split both data sets into training $T_i \subset D_i$ and validation data $V_i \subset D_i$ containing 90 % and 10 % of the objects respectively to avoid testing our models on vehicles they have been trained with.

Convolutive vector-power: The Semantic Pointer Architecture (SPA) [7] is one special case of Vector Symbolic Architectures [8], a family of modeling approaches based on high-dimensional vector representations. Here, atomic vectors are picked from the real-valued unit sphere, the dot product serves as a measure of similarity and the algebraic operations are component-wise vector addition $\oplus$ and circular convolution $\otimes$. In this work, we make use of the fact that for any two vectors $\mathbf{v}, \mathbf{w}$, we can write

$$\mathbf{v} \otimes \mathbf{w} = IDFT\left(DFT(\mathbf{v}) \odot DFT(\mathbf{w})\right), \tag{1}$$

where $\odot$ denotes element-wise multiplication, DFT and IDFT denote the Discrete Fourier Transform and Inverse Discrete Fourier Transform respectively.
Using Eq. (1), we define the *convolutive power* of a vector $\mathbf{v}$ by an exponent $p \in \mathbb{R}$ as

$$\mathbf{v}^p := \Re\left(IDFT\left(\left(DFT_j(\mathbf{v})^p\right)_{j=0}^{D-1}\right)\right), \tag{2}$$

where $\Re$ denotes the real part of a complex number. Furthermore, we call a vector $\mathbf{u}$ *unitary*, if $\|\mathbf{v}\| = \|\mathbf{v} \otimes \mathbf{u}\|$ for any other $\mathbf{v}$.

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2. Available from http://www.i6doc.com/en/.

Vector representation: In this paper, we adopt the vector representation for automotive scenes for trajectory prediction [9] with the goal to detect abnormal driving situations using an autoencoder neural network. 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. Let (x,y) denote the position of the target vehicle and (x_{obj}, y_{obj}) the positions of all other visible objects closer than 40 m to the target (to avoid accumulation of noise), we encapsulate this information in a scene vector

$$\mathbf{S} = \mathbf{TARGET} \otimes \mathbf{TYPE}_{target} \otimes \mathbf{X}^{x} \otimes \mathbf{Y}^{y} \oplus \sum_{obj} \mathbf{TYPE}_{obj} \otimes \mathbf{X}^{x_{obj}} \otimes \mathbf{Y}^{y_{obj}}, \qquad (3)$$

where **TARGET** denotes an additional ID-vector chosen at random to indicate the target object (relevant for trajectory prediction). Figure 1 visualizes one example scene from the *On-board* data-set and its representation vector queried for the target and other cars.

Network architecture and training: In this paper, we train a fully-connected autoencoder neural network [1] with 4 hidden layers consisting of 64, 32, 32 and 64 neurons in unsupervised fashion to generate replicates of the original scene vectors. Once the network is trained on a sufficiently large data set, we can calculate the element-wise error

$$\varepsilon_{\mathbf{v}} = \sqrt{\frac{1}{D}\sum_{i=0}^{D-1} (v_i - \tilde{v}_i)^2} \qquad (4)$$

between the original vector $\mathbf{v} = (v_0, \ldots, v_{D-1})$ and the replicate vector $\tilde{\mathbf{v}} = (\tilde{v}_0, \ldots, \tilde{v}_{D-1})$ generated by the neural network autoencoder. Vectors exceeding a certain threshold c for this reconstruction error, i.e., $\varepsilon_{\mathbf{v}} > c$ will be considered as outliers or anomalies. In this paper, we use 10 % for the amount of expected outliers within the data set to calculate c and train the network for 100 epochs on the vectors encoding the current scene as described in Eq. (3)

3 Experiments

Results: Since the data we are using to train the network is unlabeled, i.e., we do not have any information available which vectors belong to unusual or atypical situations, we have no way of comparing the results produced by the neural network with actual ground truth data. Hence, we analyze the anomalies detected by our neural network with respect to certain characteristic values describing the driving situation and compare them to the values of the complete data set. If this comparison uncovers significant differences between the detected anomalies and the entirety of all data samples, we can conclude that the anomalies are reasonably different and furthermore, that there is sufficient information encoded in our vector representation to unravel them. For this analysis, we use the same metrics already used in Mirus et al. [9] to characterize crowded and potentially dangerous driving situations, namely the distance between the target and the ego-vehicle, the distance between the target and the closest other vehicle and the number of other vehicles present in the scene. Figure 2 visualizes the distribution of these metrics on the set of anomalies produced by the neural network and on all data samples in the *On-board* data set.

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence
and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

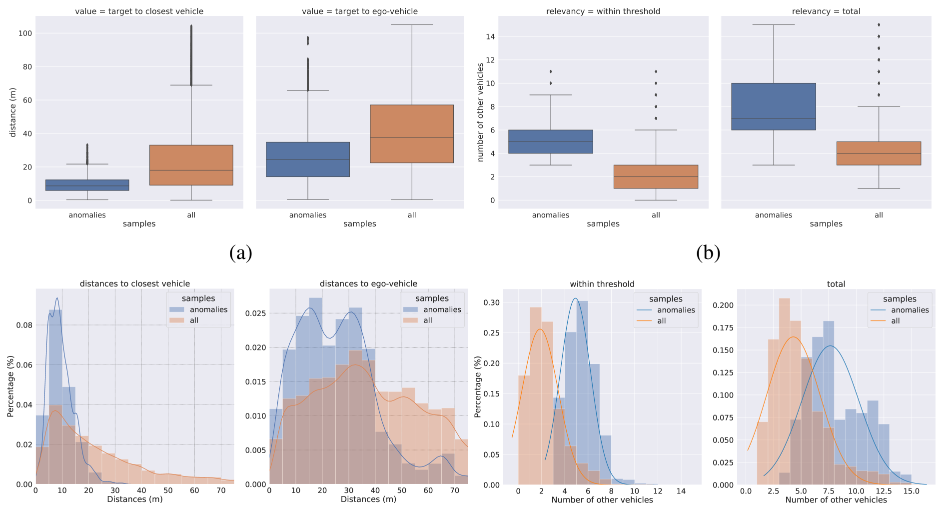

Fig. 2: Visualization of the results of the autoencoder neural network used for unsupervised anomaly detection on the *On-board* data set. The figures show the distribution of distances between the target and other vehicles (Fig. (a) and (c)) as well as the number of other vehicles (Fig. (b) and (d)) for situations classified as anomalies.

We observe a clear difference for both, the evaluation of the distances and the number of other vehicles, between the anomalies detected by our neural network and the complete set of data samples. Focusing on the distance information, the number of instances with smaller distances between the ego-vehicle or the closest other vehicle and the target is significantly higher for the anomalies than for the complete data set. While the mean distance between the target and closest other vehicle is slightly below 20 m for the complete data set, the mean distance for the anomalies is 10 m (cf. Fig. 2a) with clear concentration below 0 m to 15 m (cf. Fig. 2c). We observe a similar distribution for the distance between the target and the ego-vehicle, where the distances are more or less equally distributed around the mean of 40 m for the complete data set. For the anomalies, we observe a concentration of the distances between the target and the ego-vehicle below 40 m around the mean of 25 m. Regarding the number of other vehicles, the difference between the complete data set and the anomalies detected by our neural network is even clearer. For the complete data set, the mean number of other vehicles within 40 m to the target vehicle is 2, while the total mean number of other vehicles is around 4. Both numbers are significantly higher for the anomalies with a mean number of 5 other vehicles within 40 m and a mean of 7 other vehicles in total (cf. Fig. 2b). Looking at the distribution shown in the histograms in Fig. 2d, the picture becomes even clearer. There are no situations with less then 3 other vehicles within 40 m to the target vehicle in the set of anomalies, whereas in this same range fall the majority of samples of the complete data set. We observe a similar distribution for the total number of other vehicles in the scene with the anomaly samples having at least 3 and the majority of examples having at least 4 other vehicles present in the scene. In contrast, the great

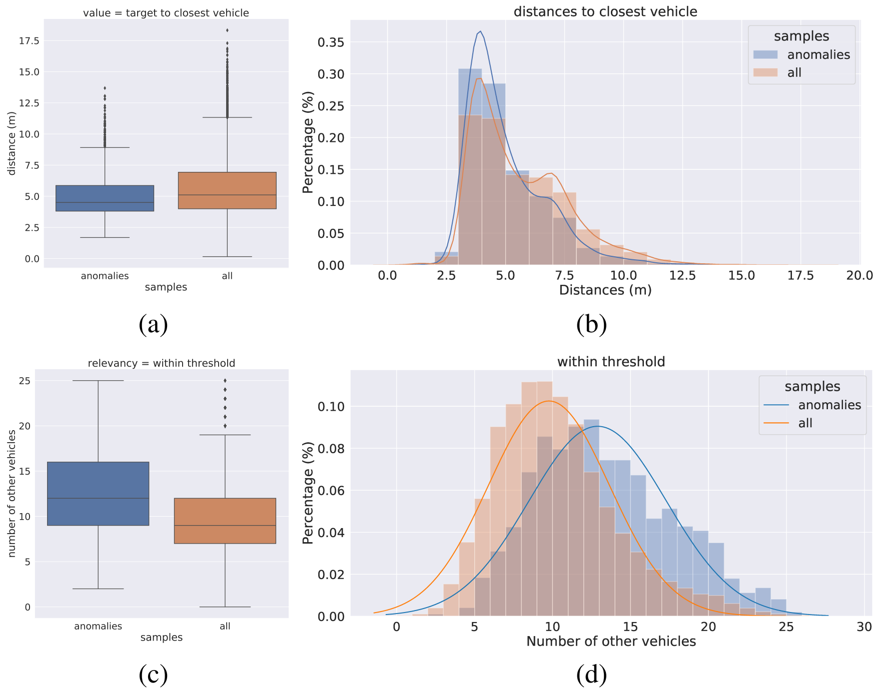

Fig. 3: Visualization of the results of the autoencoder neural network used for unsupervised anomaly detection on the *NGSIM* data set.

majority of samples from the complete data set has at most 7 other vehicles present and the overall distribution is somewhat shifted compared to that of the anomalies.
Figure 3 shows a similar analysis for the *NGSIM* data set with a few systematic differences. Since the *NGSIM* data set is recorded with external cameras observing highway traffic, we only analyze the distance between the target vehicle and the closest other vehicle (see Fig. 3a and 3b). Furthermore, we focus on vehicles within a distance of 40 m on lanes adjacent to the target vehicle's lane for the analysis of our anomaly detection network here as well (see Fig. 3c and 3d). While the differences between anomalies and the complete data set regarding the distance between the target and the closest other vehicle is not as significant in comparison to the *On-board* data set, we still observe a similar tendency for the anomalies to capture situations with smaller distances between the target and the closest other vehicle. For the number of other vehicles however, we also observe that the samples detected as anomalies by our autoencoder network tend to have more vehicles in the target vehicle's surroundings present than for all the samples within the *NGSIM* data set.

4 Discussion

Conclusion: In conclusion, our autoencoder neural network is able to detect a subset of anomalies consistently for both, the *On-board* and *NGSIM* data set, which show sufficiently significant differences to the complete data set regarding certain metrics.

The results indicate, that the anomalies detected by the network have a tendency towards crowded situations with rather small distances between the target vehicle and the other vehicles in its surroundings.

Future work: The results shown here offer interesting directions for future research. For instance, we could combine the anomaly detection network based on our vector representation presented in this paper with the behavior prediction networks proposed in [9] to decide whether the current driving situation is potentially hazardous and needs more accurate predictions than less crowded or dangerous situations. Our evaluation presented in [9] shows that Long Short-Term Memory (LSTM) models employing the convolutive power representation outperform the other models in such situations particularly in lateral direction. Furthermore, we could train trajectory prediction models particularly on lane changes, the outliers and a similar amount of "normal" samples to create a more balanced training data set. Finally, we could also investigate other anomaly detection algorithms in addition to the autoencoder models shown here to get a better understanding about what sort of data samples are actually outliers by evaluating how different models classify anomalies differently.

References

[1] J. Chen, S. Sathe, C. Aggarwal, and D. Turaga. "Outlier Detection with Autoencoder Ensembles". In: *Proceedings of the 2017 SIAM International Conference on Data Mining*. Society for Industrial and Applied Mathematics, 2017, pp. 90–98. DOI: 10.1137/1.9781611974973.11. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611974973.11.

[2] V. Chandola, A. Banerjee, and V. Kumar. "Anomaly Detection: A Survey". In: *ACM Computing Surveys* 41.3 (2009), pp. 1–58. ISSN: 0360-0300. DOI: 10.1145/1541880.1541882.

[3] C. Zhang, D. Song, Y. Chen, X. Feng, C. Lumezanu, W. Cheng, J. Ni, B. Zong, H. Chen, and N. V. Chawla. "A Deep Neural Network for Unsupervised Anomaly Detection and Diagnosis in Multivariate Time Series Data". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 33 (2019-07), pp. 1409–1416. DOI: 10.1609/aaai.v33i01.33011409.

[4] S. Bai, C. Han, and S. An. "Recognizing Anomalies in Urban Road Scenes Through Analysing Single Images Captured by Cameras on Vehicles". In: *Sensing and Imaging* 19.1 (2018-08), p. 34. ISSN: 1557-2072. DOI: 10.1007/s11220-018-0218-7. URL: https://doi.org/10.1007/s11220-018-0218-7.

[5] Y. Yuan, D. Wang, and Q. Wang. "Anomaly Detection in Traffic Scenes via Spatial-Aware Motion Reconstruction". In: *IEEE Transactions on Intelligent Transportation Systems* 18.5 (2017-05), pp. 1198–1209. DOI: 10.1109/TITS.2016.2601655.

[6] Y. Yao, M. Xu, Y. Wang, D. J. Crandall, and E. M. Atkins. "Unsupervised Traffic Accident Detection in First-Person Videos". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2019.

[7] C. Eliasmith. *How to build a brain: A neural architecture for biological cognition*. New York, NY: Oxford University Press, 2013.

[8] R. Gayler. "Vector Symbolic Architectures answer Jackendoff's challenges for cognitive neuroscience". In: *ICCS/ASCS International Conference on Cognitive Science*. Ed. by P. Slezak. University of New South Wales. CogPrints, 2003, pp. 133–138.

[9] F. Mirus, P. Blouw, T. C. Stewart, and J. Conradt. "An Investigation of Vehicle Behavior Prediction Using a Vector Power Representation to Encode Spatial Positions of Multiple Objects and Neural Networks". In: *Frontiers in Neurorobotics* 13 (2019-10), p. 84. ISSN: 1662-5218. DOI: 10.3389/fnbot.2019.00084. URL: https://www.frontiersin.org/article/10.3389/fnbot.2019.00084.