A LSTM Approach to Detection of Autonomous Vehicle Hijacking

Anonymous Author(s)
Affiliation
Address
email

Abstract

In the recent decades, automotive research has been focused on creating a driverless future. Autonomous vehicles are expected to take over tasks which are dull, dirty and dangerous for humans (3Ds of robotization). However, augmented autonomy increases reliance on the robustness of the system. Autonomous vehicle systems are heavily focused on data acquisition in order to perceive the driving environment accurately. In the future, a typical autonomous vehicle data ecosystem will include data from internal sensors, infrastructure, communication with nearby vehicles, and other sources. Physical faults/malicious attack or a misbehaving vehicle can result in the incorrect perception of the environment, which can in turn lead to task failure or accidents. Anomaly detection is hence expected to play a critical role in improving the security and efficiency of autonomous and connected vehicles. Anomaly detection can be simply defined as a way of identifying unusual or unexpected events and/or measurements. In this paper, we focus on the specific case of malicious attack/hijacking of the system which results in unpredictable evolution of the autonomous vehicle. We use a Long Short-Term Memory (LSTM) network for anomaly/fault detection. It is first trained on non-abnormal data to understand the system's baseline performance and behaviour, monitored through four vehicle control parameters namely velocity, acceleration, jerk and steering rotation. The model is next used to predict over a number of future time steps and an alarm is raised as soons as the observed behaviour of the autonomous car significantly deviates from the prediction. The relevance of this approach is supported by numerical experiments based on data produced by an autonomous car simulator, capable of generating attacks on the system.

1 Introduction

The past few decades have seen the automotive industry invest significant amount of resources in the development of autonomous driving and connected vehicles. It is expected that, with time, autonomous vehicles will find increasing use in real-world applications. With the advancement in sensor technology, information exchange networks, and ease of processing data, autonomous systems have become exceedingly capable and efficient at performing different driving tasks. As the whole autonomous environment is data driven, data acquisition and data reliability become an important aspect for smooth and efficient working of the system. In the future a typical autonomous vehicle data ecosystem will include data from internal sensors, infrastructure, communication with nearby vehicles, and other sources. A data based environment is a delicate structure and is vulnerable to error and hacking, which makes the autonomous and connected vehicles highly susceptible to malicious attacks and information tampering, along with system failures. Hence an anomaly detection scheme is essential, in particular to answer the question : Can the data received be trusted? In an autonomous dynamic driving environment where the vehicles do not receive all the information available to a driver, but instead rely on information gathered from sensors on the vehicle, it is impossible to foresee all the possible faults. Hence the system must be complemented by anomaly-detection systems, that can detect anomalies and trigger diagnosis or alert. Such a system has to be computationally light, and detect faults with high degree of both precision and recall. A too-high rate of false positives will lead operators to ignoring the system; a too-low rate makes it ineffective. In addition, the faults must be detected quickly after their occurrence, so that they can be dealt with before they become catastrophic. In this paper, we develop a LSTM approach to online hijacking detection for autonomous vehicles in two steps, based on the assumption that, in absence of attack on the system, the behavior of a self-driving car is smooth and highly predictible at a short term horizon. Precisely, the behavior of the self-driving vehicle is described by three parameters here: speed, acceleration, and rotation. The first step consists in training the LSTM network to understand the system's baseline performance and behaviour. The model is then used to predict these parameters over a number of future time steps and an alarm is raised as soon as the observed behaviour of the autonomous car significantly deviates from the prediction. The dataset used in this study arises from experiments performed on a treadmill based autonomous car simulator at University of Waterloo, Canada, see https://uwaterloo.ca/embedded-software-group/projects/adas-treadmill-demonstrator. The rest of this paper is organized as follows. Section 2 presents the LSTM approach promoted and related works. Section 3 describes the Treadmill Demonstrator we used to generate the dataset and the parameters of the LSTM model. In section 4, the performance of our approach is investigated and some concluding remarks are collected in section 5.

2 The LSTM Approach

This section presents the rationale behind our approach. We start by briefly describing LSTM network models. We next use LSTM to model the dynamic behaviour of the system (autonomous vehicle in our case) in order to gather knowledge about the baseline performance (model training stage). The model is then used to detect changes in the system as well as outliers using root mean square error metrics (prediction stage).

2.1 Long Short-Term Memory (LSTM) Networks

The persistence of information in our brain helps us in understanding any situation based on the memory of the past events. The human brain does not erase everything each time a new situation occurs and start from scratch. Recurrent neural networks use the same logic and in essence are neural networks with loops in them which allows information to persist. A loop allows information to be passed from one step of the network to the next.

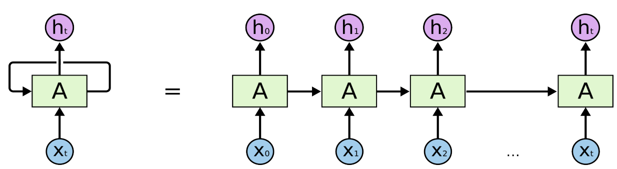

Figure 1: RNN.

Thus, RNNs use past information to understand the present situation. One major drawback of RNN's is how far in the past should we search. Sometimes, the recent past can provide enough information to execute the present task, but there are also times when we have to look further back in the memory to extract the required and relevant information. It's entirely possible that for certain applications or in certain scenarios this gap between the relevant information and the point where it is needed becomes very large. Performance of RNNs deteriorates as this gap grows. Long Short Term Memory networks are a special kind of RNN, capable of learning long-term dependencies. They were introduced by Hochreiter and Schmidhuber in Hochreiter and Schmidhuber [1997] and are explicitly designed to avoid the long-term dependency problem. LSTMs also have this chain like structure like RNN, but the repeating module has a different structure. Unlike RNN's that have a single neural network layer, LSTM comprises of four layers interacting in a special way. The LSTM has the ability to remove old information or add new information at any point, which is regulated by structures called gates. Gates are composed of a sigmoid neural net layer and a pointwise multiplication operation and are a way to exchange information. An LSTM has three of these gates, to protect and control the information.

- Forget Gate: to decide what information we're going to throw away from the block.
- Input Gate: to decide what new information we're going to update/store in the block "
- Output Gate: to decide what to output based on input and the memory of the block.

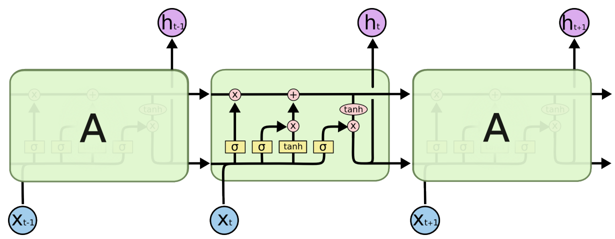

Figure 2: LSTM

2.2 Model Training and Testing

In Malhotra et al. [2015], LSTM is used to model time series data and proved to be efficient for detecting anomalies. In this paper we use a similar approach for on-line detection of malicious attacks on autonomous vehicles. We use a stacked LSTM architecture. In the training stage, a LSTM adapts its weights to mimic the training data. In our case we train the model using non abnormal data as we would like the model to learn and understand a normal driving behaviour. This model is next used for prediction: a significant deviation from the predicted behavior tends to indicate the occurence of an attack on the system. Root Mean Square Error between the prediction and observed values is used to set the threshold for hijacking detection. We use here a simple LSTM network architecture, since the goal pursued is not the accurate prediction of the driving behaviour but to investigate the use of LSTM model as a hijacking detection tool.

The parameters of the LSTM model are shown below

Table 1: LSTM Parameters

Layer (type)	**Output Shape**	**Param**
lstm 89 (LSTM)	(1, 3, 4)	96
lstm 90 (LSTM)	(1, 4)	144
dense 58 (Dense)	(1, 1)	5

3 Data Acquisition through the Treadmill Demonstrator

The Treadmill Demonstrator University was used to collect data for different driving scenarios. This demonstrator is a laboratory platform at the University of Waterloo, Canada and is used for research and validation of results on real-time safety-critical systems in the context of assisted and autonomous driving algorithms. The platform consists of treadmill which mimics the movement of a straight road. The position control places the vehicle on the treadmill without it drifting away. The car model is capable of emulating various driving scenarios like free fun, slalom, platooning and collision avoidance. The following data was collected for different driving scenarios with and without injection of attacks

- Position Data (Infrared Sensor)
- Vehicle Orientation (Infrared Sensor)
- Vehicle Commands (Steer/Throttle)
- Anomaly Information

The position data acquired from the different tests is used to calculate the

- Velocity: Rate of change of position

- Acceleration: Rate of change of Velocity
- Jerk : Rate of change of Accleration

The approach uses least-squares smoothing to locally fit a polynomial with a moving window and then evaluate the derivative of the polynomial. A Savitzky-Golay filter is used for this step. Savitzky-Golay filter is a digital filter used for smoothing of the data using a process known as convolution, i.e. fitting successive sub-sets of adjacent data points with a low-degree polynomial by the method of linear least squares. The algorithm calculates the velocity and acceleration of a given position signal based on two parameters:

1. the size of the smoothing window
2. the order of the local polynomial approximation

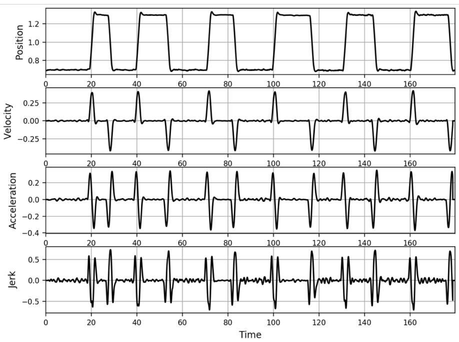

Figure 3: Non Anomalous Data

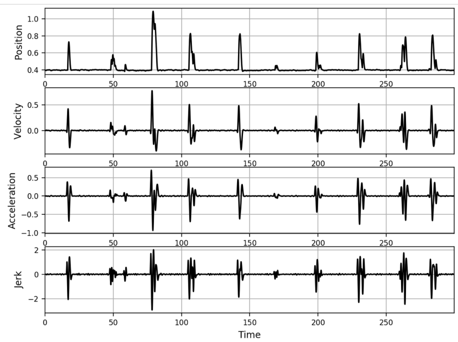

Figure 4: Anomalous Data

4 Results and Discussion

We present the results of LSTM on four vehicle control parameters which have different levels of difficulty as far as detecting anomalies in them is concerned. The non anomalous data was collected for free run and slalom driving scenarios whereas for anomalous dataset, anomalies were injected in the free run scenario. The injected anomaly is called compound injection and it simulates a scenario where a malicious attacker manages to gain access to the car's transmission control wirelessly, by causing the throttle value to be multiplied by the specified positive factor.

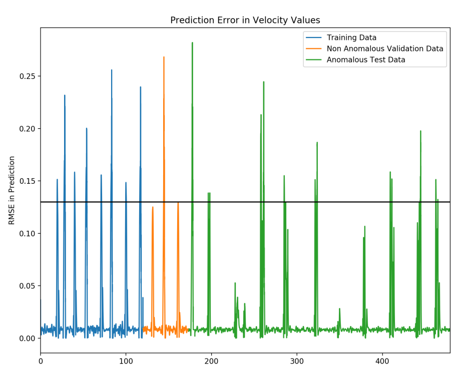

Figure 5: Prediction Error in Velocity Data

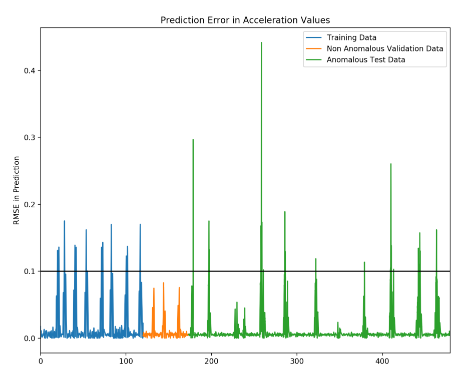

Figure 6: Prediction Error in Acceleration Data

The figures illustrate the prediction errors in the training, validation and testing stages for the four parameters under study. The errors tend to converge after 150 epochs. The spikes seen in the errors are areas where a driving manoeuvre was performed (increase in speed/acceleration/jerk or change of

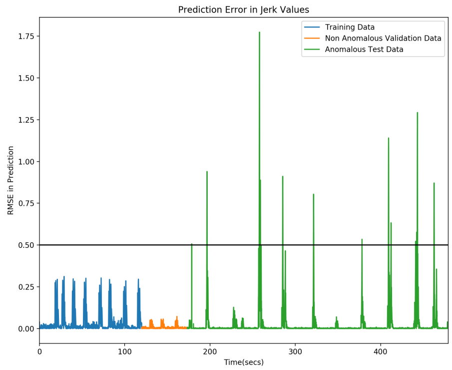

Figure 7: Prediction Error in Jerk Data

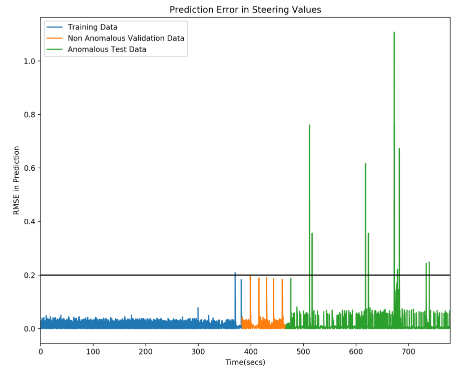

Figure 8: Prediction Error in Steering Data

vehicle direction). This manoeuvre can be a non anomalous driving behaviour (as in the training and validation set) or anomalous driving behaviour (as in the testing set).

By comparing the validation set and training set errors, we can clearly see that a certain threshold value can be used to detect attacks. The latter may vary with the efficiency of the prediction model of course. But the trend remains the same with prediction error being higher in the event of an anomalous driving manoeuvre.

5 Conclusion

In this paper, we have proposed a model for hijacking detection based on Long Short-Term Memory Recurrent Neural Network. We have provided empirical evidence that stacked LSTM networks are relevant to predict the normal behaviour of a self-driving vehicle at a short term horizon, and can be next used to detect possible attacks on the system. We showed that even a very basic LSTM Model approach yielded promising results on four different datasets. In future work, we will focus on:

1. LSTM model parameter tuning to improve the robustness
2. Improve the anomaly detection efficiency of the model
3. Model extension to be able to discriminate between different types of anomalies

References

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8): 1735–1780, 1997.

Pankaj Malhotra, Lovekesh Vig, Gautam Shroff, and Puneet Agarwal. Long short term memory networks for anomaly detection in time series. 2015.

Waterloo University. Treadmill Demonstrator. `https://uwaterloo.ca/embedded-software-group/projects/adas-treadmill-demonstrator`.