Long Short-Term Vehicle Dynamics

David Casterton  
Stanford University, CS230 Spring 2019  
david.casterton@stanford.edu / @gmail.com

Abstract

Vehicle dynamics simulation can achieve excellent results from manually crafted physical models, however manually crafting models can be time consuming, financially and computationally expensive, difficult to tune details, and difficult to generalize. This project focused on potential roles of Deep Neural Network’s (DNNs) to improve vehicle dynamics simulation by utilizing data to learn the model.

1 Introduction

A key challenge to bringing autonomous vehicles to market is proving they can safely operate within a very large state space of potential scenarios and associated sensor inputs. Simulation is one of the leading techniques being pursued to make this proof, due to the promise of scalable parallelization for rapid feedback and eventual deep exploration of the operational state space. However, before a simulator can meaningfully inform how a specific real-world vehicle will perform, the accuracy of the simulation must be well understood and refined to be within acceptable bounds.

This project evaluated if simulated vehicle movement can be improved by learning unmodeled attributes from vehicle specific data. Results show that training Long Short Term Memory (LSTM) Recurrent Neural Networks (RNN) for end-to-end learning of vehicle movement can output next movements with plausible results for some movement features, but that end-to-end learning results alone do not appear to outperform physics modeling. LSTM experiments were performed with the input of variable lengths of recent history from vehicle movement and control, and output as the next update for vehicle movement. Related work indicates that joint usage of physics modeling and novel LSTM architectures can achieve better results than physics modeling alone.

2 Related work

[4, 5, 8, 10] are examples from the mature field of vehicle dynamics modeling, which yields excellent results however is limited by the manual effort put in to the physical modeling. [1] explores end-to-end Deep Learning (DL) all the way from camera pixels to vehicle actuation output, this does not focus specifically on vehicle dynamics but presents a surprisingly successful extreme for the limit of end-to-end DL in vehicles. [11, 12] present techniques to unify physics models with DNN’s, with results outperforming physics models alone. [11] specifically presents a nominal dynamics model combined with a DNN to learn how to control multi-rotor drone movement and landing, which utilizes modeling for the well known aspects of the problem then extends modeling with a DNN resulted in better performance than a baseline control system.
3 Dataset and Features

Data for this project was acquired from [9], which contains 22 files of vehicle dynamics recorded from an expert driving a 1965 Ferrari in 2013 and 2014 at Monterey Motorsports Reunion and Targa Sixty-Six events. This data set includes the notable features of: physical inputs to the vehicle (steering, brake, throttle, clutch), GPS position (latitude, longitude, altitude), velocity and acceleration of the center of gravity (x, y, z), orientation angle and rate (roll, pitch, yaw), accuracy of position and orientation, vertical chassis accelerations, suspension deflections, and wheel accelerations.

![Graph 1](image1.png) ![Graph 2](image2.png)

Figure 1: Steering, brake, throttle from 2013. Figure 2: Steering, brake, throttle from 2014.

Figures 1 and 2 show zoomed-out examples of full files, which indicate idle time at the start and end of files, with repeated maneuvers in the middle from race laps.

![Graph 3](image3.png) ![Graph 4](image4.png)

Figure 3: Zoomed-in longitudinal signals. Figure 4: Zoomed-in lateral signals.

Figures 3 and 4 show zoomed-in example data from the data set with strong correlations. Figure 3 shows that throttle and brake have a strong correlation with longitudinal velocity (v_xCG). Figure 4 shows that hand wheel angle has a strong correlation with lateral velocity (v_yCG) and yaw angle, however these correlations have some unintuitive attributes: hand wheel angle has an inverted sign from lateral velocity (v_yCG), and yaw angle has a discontinuity at +/-180 where it inverts its sign.

This data set was recorded at 1000 Hz, however since some signals only update at 100Hz I chose to sample the data for training with a stride of 10. Some signals I considered critical (velocity, acceleration, orientation) appear to have originated from a GPS sensor that occasionally stopped updating, which resulted in those signals being temporarily incorrect then jumping with a discontinuity. To sanitize GPS discontinuities from this data set all rows were removed that contained null for GPS altitude.

4 Methods

4.1 Hand engineering features into data

The initial method pursued was to hand engineer features into the input data set that were considered to optimize for learning. The hypothesis was to train a neural network using all input columns available to output the change in position and orientation for the next time step. To accomplish this additional columns were appended to input data with the discrete derivatives of position and
orientation between the current time \( t \) and \( t+1 \). Then to avoid the idle times shown in Figures 1 and 2 disproportionately impacting the training, dev, or test set, all rows were shuffled to create an equal distribution across a file. Once a file had an equal distribution the rows were split 90% training, 5% dev, 5% test, then the columns were separated so original columns became input features and the new discrete derivatives became output labels.

Sequential TensorFlow models composed of a variable number of dense layers were trained against the data strategy described above, with limited success. After spending more time with the data it became apparent that this strategy was not adequately accounting for the delay between human inputs (steering, brake, throttle) being given to the vehicle and the resulting actuation (motor, brake, steering response) - training against labels for the discrete derivative of the next time step only did not give adequate time for the input to result in a learnable actuation response. A brief attempt was made to hand engineer features with longer input/output relationships, however this was quickly abandoned in favor of pursuing a network architecture that could learn temporal sequences.

4.2 Long Short-Term Memory Recurrent Neural Networks

To handle the problem of unknown temporal delays between vehicle input and output actuations, a RNN was chosen due to its ability to make decisions from a temporal sequence [6], with the hope that would enable relationships to be learned with unknown temporal delays.

![LSTM network architecture](image1)

![LSTM cell from [7]](image2)

Figures 5 and 6 show an LSTM’s network architecture and the contents of a single cell within the network models. The LSTM cell is primarily composed of the following attributes: forget gate, update gate, and output gate. These gates are defined by the following equations from [7]:

4.2.1 Forget Gate

\[
\Gamma_f^{(t)} = \sigma(W_f[a^{(t-1)}, x^{(t)}] + b_f)
\]  

4.2.2 Update Gate

\[
\Gamma_u^{(t)} = \sigma(W_u[a^{(t-1)}, x^{(t)}] + b_u)
\]

\[
\tilde{c}^{(t)} = \tanh(W_c[a^{(t-1)}, x^{(t)}] + b_c)
\]

\[
c^{(t)} = \Gamma_f^{(t)} \ast c^{(t-1)} + \Gamma_u^{(t)} \ast \tilde{c}^{(t)}
\]

4.2.3 Output Gate

\[
\Gamma_o^{(t)} = \sigma(W_o[a^{(t-1)}, x^{(t)}] + b_o)
\]

\[
a^{(t)} = \Gamma_o^{(t)} \ast \tanh(c^{(t)})
\]
Data was then separated into 3 dimensions: # of input features (n_x), # of data samples (m), sequence length (T_x).

5 Experiments/Results/Discussion

Appendix A shows a table of attempted model parameters, hyperparameters, and input/output data sets (further defined in Appendix B), and the resulting loss and mean absolute error for each model. Source code for models is available [3]. The models that resulted in optimal results involved an input sequence length from 100-200 (1-2 seconds worth of history), and took about 5 hours to train on an NVIDIA GTX 1080 Ti.

For example, Model ID 80 from Appendix A achieved a mean absolute error of 0.95 when trained to data set group x3 (Appendix B.3) and output group y4 (Appendix B.8). Model 80 was composed of 3 layers:

1. LSTM layer 1
   - input sizes: n_x = 20, m = 132,645, T_x = 160
   - output sizes: n_y = 160, m = 132,645, T_x = 160

2. LSTM layer 2
   - input sizes: n_x = 160, m = 132,645, T_x = 160
   - output sizes: n_y = 160, m = 132,645

3. Dense layer
   - input sizes: n_x = 160, m = 132,645
   - output sizes: n_y = 10, m = 132,645

Figures 7, 8, 9 are zoomed-in plots of Model 80 predictions. These figures visualize that different vehicle dynamic features were learned with variable success:

- **longitudinal velocity**: (Figure 7) Mean Squared Error (MSE) of 1.88 with an input value range of 61.72, resulting in MSE of 3.04% of the input range. Longitudinal velocity appears to have been successfully learned as hoped due to strong correlation with the throttle and brake pedal input, after some time delay. However the model appears to accumulate most of its error when input values go beyond 50, future work would be needed to understand why the model appears to be failing to output values much beyond 50.

- **lateral velocity**: MSE of 0.27 on input range of 4.04, resulting in MSE of 6.91% of range. Lateral velocity was not as successful at being learned as longitudinal velocity, this is likely due to a weaker correlation between steering input and lateral velocity as well as a significantly smaller input range for this feature.

- **yaw angle**: (Figure 8) MSE of 2.96 on input range of 359.98, resulting in MSE of 0.82% of range. Yaw angle appears to have been quite successfully learned as hoped due to strong correlation with the steering wheel angle. It is surprising how well the model was able to handle yaw angle discontinuities when it exceeds +/-180, how this is being accomplished would benefit from further investigation.

- **wheel acceleration**: (Figure 9) MSE of 1.53 on input range of 83.88, resulting in MSE of 1.83% of range. Wheel acceleration data appears to be reporting with too high of a frequency for the model, which is resulting in the model appearing to low-pass filter this feature.

Additional models were trained to single outputs and were able to achieve the following results:

- Model 95 was trained to longitudinal velocity only and achieve MSE of 0.43.
- Model 90 was trained to lateral velocity only and achieved MSE of 0.09.
- Model 97 was trained to yaw angle only and achieved MSE of 1.89.
6 Conclusion/Future Work

From these results LSTM’s outperformed attempts to hand engineer temporal relationships into the data then train with simple multi-layer networks, and LSTM models appear generally capable of learning vehicle dynamics with temporal delays. However, LSTM results still had a significant MSE between predicted future results and actual future results in the test set, so the investigated network architecture does not appear to be a candidate to threaten replacing hand modeling vehicle dynamics for simulation.

For future work I would like to investigate:

1. **Integrating physical modeling with a RNN**: as proposed by [11, 12]. The approach proposed in both [11] and [12] is to utilize a physics model to implement the known features of movement then integrating with a learning model to fine-tune the unknown features or hardware unit specific features. Specifically, I would like to integrate a TensorFlow LSTM model with the PyBullet [2] physics simulator’s vehicle model, then apply common steering / brake / throttle input in simulation as from a real-world data capture and plot the resulting differences in position and orientation over time.
2. **Predict movement farther into the future**: this project continually predicted only the next step of movement based on recent history, it would be interesting to understand how much future movement could be accurately predicted based on the recent past.

References


7 Appendix

A Model Hyperparameters and Results

<table>
<thead>
<tr>
<th>ID</th>
<th>X</th>
<th>Y</th>
<th>dropout</th>
<th>rate</th>
<th>decay</th>
<th>time</th>
<th>norm</th>
<th>units</th>
<th>seq len</th>
<th>files</th>
<th>loss</th>
<th>mean error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>x1</td>
<td>y1</td>
<td>0.1</td>
<td>0.02</td>
<td>0.0</td>
<td>0:08:22</td>
<td>False</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>12.05</td>
<td>1.47</td>
</tr>
<tr>
<td>2</td>
<td>x1</td>
<td>y1</td>
<td>0.2</td>
<td>0.02</td>
<td>0.0</td>
<td>0:10:17</td>
<td>False</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>19.03</td>
<td>1.63</td>
</tr>
<tr>
<td>3</td>
<td>x1</td>
<td>y1</td>
<td>0.0</td>
<td>0.02</td>
<td>0.0</td>
<td>0:10:46</td>
<td>False</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>84.33</td>
<td>3.23</td>
</tr>
<tr>
<td>4</td>
<td>x1</td>
<td>y1</td>
<td>0.0</td>
<td>0.02</td>
<td>0.0</td>
<td>0:03:49</td>
<td>True</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>327.15</td>
<td>5.29</td>
</tr>
<tr>
<td>5</td>
<td>x1</td>
<td>y1</td>
<td>0.0</td>
<td>0.02</td>
<td>0.0</td>
<td>0:06:36</td>
<td>False</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>8.86</td>
<td>1.36</td>
</tr>
<tr>
<td>6</td>
<td>x1</td>
<td>y1</td>
<td>0.0</td>
<td>0.02</td>
<td>0.0</td>
<td>0:04:08</td>
<td>True</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>88.96</td>
<td>3.4</td>
</tr>
<tr>
<td>7</td>
<td>x1</td>
<td>y1</td>
<td>0.1</td>
<td>0.02</td>
<td>0.0</td>
<td>0:11:04</td>
<td>False</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>11.65</td>
<td>1.43</td>
</tr>
<tr>
<td>8</td>
<td>x1</td>
<td>y1</td>
<td>0.1</td>
<td>0.02</td>
<td>0.0</td>
<td>0:05:12</td>
<td>True</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>327.02</td>
<td>5.3</td>
</tr>
<tr>
<td>9</td>
<td>x1</td>
<td>y1</td>
<td>0.1</td>
<td>0.02</td>
<td>0.0</td>
<td>0:04:49</td>
<td>True</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>64.09</td>
<td>2.67</td>
</tr>
<tr>
<td>10</td>
<td>x1</td>
<td>y1</td>
<td>0.0</td>
<td>0.02</td>
<td>1e-06</td>
<td>0:04:49</td>
<td>False</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>21.29</td>
<td>1.96</td>
</tr>
<tr>
<td>11</td>
<td>x1</td>
<td>y1</td>
<td>0.0</td>
<td>0.02</td>
<td>1e-06</td>
<td>0:03:16</td>
<td>True</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>326.67</td>
<td>5.28</td>
</tr>
<tr>
<td>12</td>
<td>x1</td>
<td>y1</td>
<td>0.0</td>
<td>0.02</td>
<td>1e-06</td>
<td>0:07:00</td>
<td>False</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>11.55</td>
<td>1.5</td>
</tr>
<tr>
<td>13</td>
<td>x1</td>
<td>y1</td>
<td>0.0</td>
<td>0.02</td>
<td>1e-06</td>
<td>0:04:09</td>
<td>True</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>42.52</td>
<td>2.58</td>
</tr>
<tr>
<td>14</td>
<td>x1</td>
<td>y1</td>
<td>0.1</td>
<td>0.02</td>
<td>1e-06</td>
<td>0:10:32</td>
<td>False</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>12.07</td>
<td>1.5</td>
</tr>
<tr>
<td>15</td>
<td>x1</td>
<td>y1</td>
<td>0.1</td>
<td>0.02</td>
<td>1e-06</td>
<td>0:02:22</td>
<td>True</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>327.1</td>
<td>5.3</td>
</tr>
<tr>
<td>16</td>
<td>x1</td>
<td>y1</td>
<td>0.1</td>
<td>0.02</td>
<td>1e-06</td>
<td>0:07:21</td>
<td>False</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>12.4</td>
<td>1.5</td>
</tr>
<tr>
<td>17</td>
<td>x1</td>
<td>y1</td>
<td>0.1</td>
<td>0.02</td>
<td>1e-06</td>
<td>0:05:17</td>
<td>True</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>49.21</td>
<td>2.61</td>
</tr>
<tr>
<td>18</td>
<td>x1</td>
<td>y1</td>
<td>0.0</td>
<td>0.01</td>
<td>0.0</td>
<td>0:04:59</td>
<td>False</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>15.06</td>
<td>1.67</td>
</tr>
<tr>
<td>19</td>
<td>x1</td>
<td>y1</td>
<td>0.0</td>
<td>0.01</td>
<td>0.0</td>
<td>0:05:09</td>
<td>True</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>229.6</td>
<td>4.31</td>
</tr>
</tbody>
</table>
B Data Set Groups

B.1 x1

axCG, ayCG, azCG, brake, chassisAccelFL, chassisAccelFR, chassisAccelRL, chassisAccelRR, clutch, handwheelAngle, deflectionFL, deflectionFR, horizontalSpeed, pitchAngle, pitchRate, rollAngle, rollRate, throttle, vxCG, vyCG, vzCG, wheelAccelFL, wheelAccelFR, wheelAccelRL, wheelAccelRR, yawAngle, yawRate

B.2 x2

brake, clutch, handwheelAngle, throttle, horizontalSpeed, pitchAngle, pitchRate, rollAngle, rollRate, vxCG, vyCG, vzCG, yawAngle, yawRate

B.3 x3

brake, clutch, handwheelAngle, throttle, pitchAngle, pitchRate, rollAngle, rollRate, yawAngle, yawRate, vxCG, vyCG, vzCG, axCG, ayCG, azCG, wheelAccelFL, wheelAccelFR, wheelAccelRL, wheelAccelRR

B.4 x4

brake, clutch, handwheelAngle, throttle, pitchAngle, pitchRate, rollAngle, rollRate, yawAngle, yawRate, vxCG, vyCG, vzCG, axCG, ayCG, azCG
B.5  y1
axCG, ayCG, azCG, chassisAccelFL, chassisAccelFR, chassisAccelRL, chassisAccelRR, horizontal-Speed, pitchAngle, pitchRate, rollAngle, rollRate, vxCG, vyCG, vzCG, wheelAccelFL, wheelAccelFR, wheelAccelRL, wheelAccelRR, yawAngle, yawRate

B.6  y2
pitchAngle, rollAngle, vxCG, vyCG, vzCG, wheelAccelFL, wheelAccelFR, wheelAccelRL, wheelAccelRR, yawAngle

B.7  y3
pitchAngle, pitchRate, rollAngle, rollRate, yawAngle, yawRate, vxCG, vyCG, vzCG, axCG, ayCG, azCG, wheelAccelFL, wheelAccelFR, wheelAccelRL, wheelAccelRR

B.8  y4
pitchAngle, rollAngle, yawAngle, vxCG, vyCG, vzCG, wheelAccelFL, wheelAccelFR, wheelAccelRL, wheelAccelRR

B.9  y5
rollAngle, yawAngle, vxCG, vyCG, vzCG

B.10  y6
yawAngle, vxCG