Preprints, 10th IFAC International Symposium on
Advances in Automotive Control
Columbus, Ohio, USA, August 28-30, 2022

Cabin Load Prediction Using Time Series Forecasting in Long-haul Trucks for Optimal Energy Management

Satvik Khuntia*, Athar Hanif*, Qadeer Ahmed*
John Lahti, Maarten Meijer****

* *Center for Automotive Research, The Ohio State University, Columbus, OH, 43212 USA (e-mail: khuntia.2@osu.edu, hanif.6@osu.edu, ahmed.358@osu.edu).*
** *PACCAR Technical Center, 12479 Farm to Market Rd, Mount Vernon, WA 98273 USA (e-mail: John.Lahti@PACCAR.com, Maarten.Meijer@PACCAR.com)*

Abstract:
Predicting the electrical loads experienced by a battery pack during the 10 hour hotel period of a long haul class 8 mild hybrid truck with a sleeper cab, and using the information to achieve an optimal energy management strategy and controlling the State of Charge (SOC) of battery pack can help in improving it's freight efficiency. In this work, Machine Learning (ML) based algorithm has been proposed to predict the driver activity during the hotel period. Hence, the power load demanded from the auxiliaries can be predicted. A special kind of Recurrent Neural Network (RNN) called Long and Short Term Memory (LSTM) is used for the prediction task because of its ability to store recurrent information of a small and a large time horizon. To train the LSTM algorithm, the synthetic load profiles are synthesized using rules and observations derived from the existing baseline electrical power load profile of the hotel period. This paper entails the whole process of data synthesis to training the neural network on the synthesized data and the prediction and validation of the power load. The input to the network is a time series of 600 time steps. Dynamic Time Warping (DTW) is used to manipulate the time axis and point wise euclidean distance between the forecast and the test data is used to quantify the accuracy of the model. Then by performing hyper-parameter optimization we find the best combination for number of hidden units and the number of training days for the algorithm.

Keywords: Data synthesis, Deep Neural Network, Long and Short Term Memory, Dynamic Time Warping, Energy management.

1. INTRODUCTION

Long-haul (Class 8) trucks are meant for long distance (eg. coast to coast) freight transportation. These vehicles travel nearly 600 miles per day and are forced to be on the road for continuous 3 to 7 days [Smith et al. (2020)]. Federal Motor Carrier Safety Administration (FMCSA) has mandated heavy duty truck drivers to rest for an uninterrupted 10 hours for every 11 hours of driving [Administration (2011)]. Any extended periods of time that the drivers spend in their vehicle, often to take rest, is defined as "Hotelling". During this hotel period, drivers generally stay in the sleeper area provided with some electronically powered utilities and amenities by the truck manufacturers. In order to use these provisions, drivers tend to idle the vehicle's Internal Combustion Engine (ICE).

SuperTruck II is a program funded by the U.S. Department of Energy with a purpose of demonstrating cost-effective technologies that can increase freight efficiency by more than 100%. Idling the engine not only amounts to 0 freight efficiency, idling for long periods is also harmful for the engine, hence is very undesirable. Hybridization of the vehicle allows for the substitution of this idling (to power the auxiliaries) with any Energy Storage System (ESS) like battery packs. Moreover, it is important to ensure the battery pack has sufficient State-Of-Charge (SOC) at all times. This is not entirely possible because of the limitation in the battery sizing, hence there might be instances where some idling may be required to charge the battery back up. In this situation, it is helpful to know how much instantaneous power or the total electrical energy would be required in the hotel period so that the battery pack is only sufficiently charged and only charged at required times eliminating unnecessary idling.

For this purpose, we need to know the kinds of power loads that are expected in the hotel period. Figure 1 gives the detail of all the cabin electrical loads typically on a long-haul truck. These activities include the use of a lamp,

⋆ The authors thank the United States Department of Energy and PACCAR Inc. for the support for this project. This material is based upon work supported by the Department of Energy, under Grant Number DE-EE0008265.

TV, radio etc. Predicting these activities or/hence can give us an intuition on how much instantaneous power is expected in the next horizon (until the end of the hotel period). In a situation where the battery reaches the minimum allowable SOC, this prediction can be used to idle the vehicle to charge the battery just enough. In SuperTruck II, the HVAC is also powered by the battery pack hence also comes within the cabin loads. However, the HVAC power requirement estimation is not a part of the this predictive algorithm because of its dynamic nature. A physics based model is developed instead for the e-HVAC power load estimation modelled separately on MATLAB/Simulink [Khuntia et al. (2022)].

User activity prediction is a well studied focus area more centred towards predicting the user Activities of Daily Living (ADL). Sequence Prediction via Enhanced Episode Discovery (SPEED) had grown popular because of it's superior behaviour over other prediction algorithms like Active LeZi (ALZ) algorithm with temporal rule, Patterns of User Behavior System (PUBS). This method uses the temporal pattern of humans for the prediction [Aztiria et al. (2012)]. In [Marufuzzaman et al. (2015)], the authors introduce a modified-SPEED algorithm that achieved 96.8% accuracy, better than the above-mentioned algorithms for activity prediction. This prediction algorithm is also adopted for the prediction of activities in smart homes for gird power-cost manipulation. In [Goutham (2020)] the authors had explored an application of a multi layer Long and Short Term Memory (LSTM) algorithm to predict the time and duration of activities in a 24 hour period. The authors used two layer LSTM algorithm with 50 hidden units each can give good predictions with a learning on a moving 75 day window.

Time series prediction can be done using Recurrent Neural Networks (RNN) as the ability of having a "memory" which makes them good for long sequence prediction tasks. The RNN can remember contextual information through the hidden layer activations that are passed from one step in time to another. Different variants of RNN have been useful when temporal dependency of the data are important [Graves (2013)]. A popular variant of RNN is the Long and Short Term Memory (LSTM). LSTMs work in an iterative fashion like RNNs with the addition of a gating mechanism. It has three gates named as: (i) forget gate (ii) Update gate and (iii) Output gate which regulate the flow of information from input to activation, activation to activation and activation to output. This makes LSTMs robust against outliers in the data and learn long term dependencies and patterns in the time series. Deeper version of neural network, that is multiple (aka stacked) LSTM layers can be used to improve the performance [Goutham et al. (2021)].

Because of emerging opportunities of hybridization in freight vehicles, power load estimation can be seen as an emerging field of study in the near future. While electric load prediction in the form of activity prediction has seen myriad of applications in smart homes, there is still a void in research in the application to the electric/hybrid vehicles especially long haul trucks. Because of the same reason there are challenges in obtaining the data of user activity that would be essential for training the networks.

This paper presents a method to synthesise data from a single available data set generated from a survey conducted by PACCAR Inc. Using the synthesized data, Neural Networks toolbox in MATLAB is used to predict the future power load estimate. Leveraging the properties of Markov chain, an adaptation of a transition matrix is developed and along with temporal information of different activities, a prediction is made for the overall power demand in the next time step. This power estimate is then used in the estimation of the projected SOC for the duration of the hotel period, which in turn is used for the idle/ Engine On-Off control strategies.

2. PERFORMANCE QUANTIFICATION CRITERIA

In this section different metrics used for evaluating the predictions of a time series have been discussed. For the activity prediction as a classification problem, traditional classifier based methods can be used. In this work, the problem is formulated as a regression problem. Most popular approaches for evaluating such problems is a point-wise numeric distance between the prediction and the original values.

Some of these metrics are:

- Root mean squared error (RMSE): provides an average error throughout the predicted time series in real units.
$$RMSE = \sqrt{\frac{\sum(\hat{y}_i - y_i^*)^2}{T}} \tag{1}$$
- Normalized RMSE (NRMSE): a method that uses a normalized error over the total range of values of the test set. This is particularly useful in the case when the prediction of the time of the activity is not critical.
$$NRMSE = \frac{RMSE}{max(y_i^*) - min(y_i^*)} \tag{2}$$
- Mean Absolute Percentage Error (MAPE): another metric to evaluate the predictions where the error is normalized over the actual value. This error is a percentage of the true value.
$$MAPE = \frac{\sum(\frac{|\hat{y}_i - y_i^*|}{y_i^*}}{T} \tag{3}$$
- In [Minor et al. (2015)], the authors introduce a method to overcome the drawbacks of the "averaged error" methods listed above and also account for the outliers by introducing an error threshold, called Error Threshold Function (ETF). In this method, the error function, $I(\hat{y}_i, y_i^*) = 1$ if $|\hat{y}_i - y_i^*| < v$, v being a non negative threshold, i.e., error in a particular prediction contributes to the total error only if it is significant enough.
$$ETF = \frac{\sum I(\hat{y}_i, y_i^*)}{T} \tag{4}$$
- The predicted time series can also be seen as a signal and the test data as another signal. Using Dynamic Time Warping (DTW), we can determine the similarity between the two signals. DTW warps the x-axis (in this case, time axis) between the two signals to match the best y axis values irrespective of the lengths of the two signals [Müller (2007)]. For

two inputs $Y_1 \in \mathbb{R}^N$ and $Y_2 \in \mathbb{R}^N$ DTW computes a cost matrix $J \in \mathbb{R}^{(N+1)\times(M+1)}$ such that,

$$J_{i,j} = d(Y_{1,i}, Y_{2,j}) + min \begin{cases} J_{i-1,j-1} \\ J_{i-1,j} \\ J_{i,j-1} \end{cases} \quad (5)$$

Where, d is the distance between the Y_1 and Y_2 at time steps i and j respectively. It can be defined in any method like euclidean, absolute or squared. This cost matrix 'J' is then used to trace back from $J_{N,M}$ to $J_{0,0}$ which gives the best mapping of y values for the two time series.

For the way the current problem is formulated it is reasonable to choose DTW as the performance metric. We are interested in predicting a particular power load to be happening in a broad time horizon rather than focusing on the exact time in which it should be happening. To understand this better, we can think of it in this way, the driver might is likely to use the microwave to heat food after they wake up from their sleep. It is important to predict the load corresponding to the microwave anytime in the later half of the 10 hr hotel period. Using DTW we can warp the time axis such that the similarity between the power load predicted and the power load profile of any test day can be checked and they can be compared.

3. DATA GENERATION

The utmost requirement for the development and training of the prediction algorithm is the available data. This data can be collected using on board data loggers or OBD scanners and connecting it to the devices via Bluetooth which is not easy. In order to circumvent this limitation, the data was generated using some judgments and past literature. LSTM is a data hungry algorithm and hence there is a need to generate a lot of data sets in order to give enough input to the algorithm to learn from it. At the moment 1000 days data is synthesized from the base data.

A survey was conducted on the various drivers about their usual activities in a 10 hour hotel period and data is recorded. An auxiliary power usage is then generated. Figure 1 shows the activity profile for each device used in the cabin of long-haul truck.

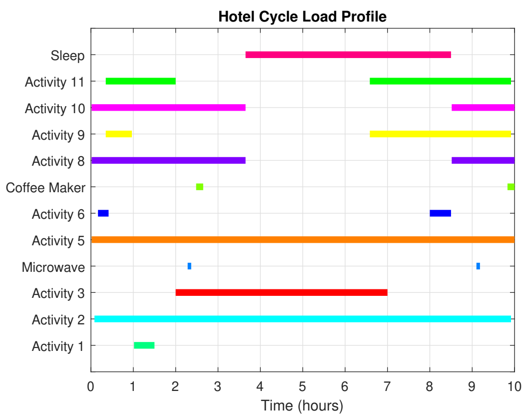

Fig. 1. Auxiliary load and activity Profile for a driver during hotel period

Some of these activities involve the usage of an electronic device (for example, eating food would involve the use of microwave to heat up the food, and drinking coffee would involve the use of coffee maker) which would have some power rating to it, which in this paper is not disclosed due to proprietary reasons. Also activities other than food (microwave) and coffee (coffee maker) are coded as Activity 4 and 7 respectively for the same reasons.

It can be seen from Figure 1 that the Activity 5 and the Activity 2 are on throughout the 10 hours hotel period and the activity 12 (sleep) is also done for a fairly long time. The driver seems to be using microwave and coffee maker for very short intervals (before and once after Sleeping). It was seen that the load requirement for these two Activities are relatively higher than the rest.

Figure 2 gives the normalized electric power for the 10 hour hotel period. It can be seen that the load requirement increases to high values once in either halves of the hotel period. This was noticed to be mainly because of the Microwave and Coffee maker.

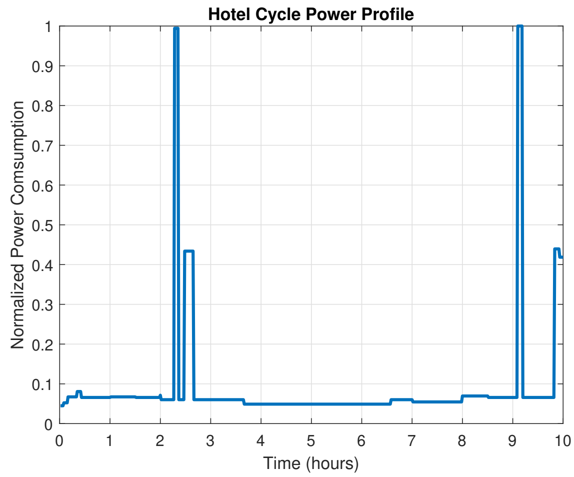

Fig. 2. Power load profile for a 10 hours hotel load period

Keeping in view the above observations, more data sets are generated. First, the activities are divided into four groups, that are, i) sleep, ii) Food, iii) Coffee and iv) miscellaneous. As mentioned before, the miscellaneous activities do not have very high load contribution as compared to Food and coffee, hence individually do not influence the total power load profile unlike the the usage of coffee maker and microwave which are rated relatively high. Hence the combined total load requirement from the miscellaneous activities is kept as same and is not varied throughout the 1000 different activity profile generated. Variability in the data is introduced using the food and coffee consumption and sleeping behaviour and is discussed in the paragraphs below.

For the sleep activity, Mitler, Merrill M., et al. in their paper, study the sleep patterns of 80 long-haul truck drivers with a total of 400 principal sleep periods [Mitler et al. (1997)]. The authors conduct a survey of how many hours do the drivers sleep during their long journey. They found that though the drivers desired an average ± SD sleep of 7.2 ± 1.2 hours, in reality they could average 5.34 hours.

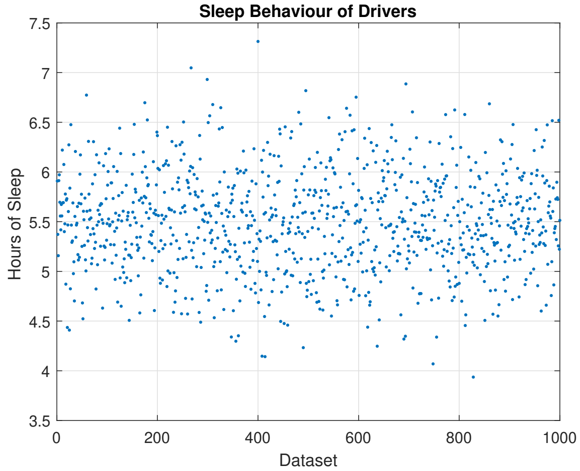

Fig. 3. Sleep duration distribution for the drivers on long haul journeys

In the above study the average time off duty was less than 8 hours (7.4 hours), however in our application, the number of OFF duty hours for the driver is 10 hours. In order to accommodate for the 2 hour increase, the sleep behavior of the driver is randomly generated using a Gaussian distribution $X \sim \mathcal{N}(\mu, \sigma) \sim \mathcal{N}(5.5, 0.5)$, that is, 5.5 hours of average sleep with a standard deviation of 0.5 hours. Figure 3 shows the random distribution of sleep hours for the 1000 days data set. It can be seen that the minimum can go up to 3.8 hours while the maximum can go up to 7.4 hours which also agrees to the study done in [Mitler et al. (1997)].

Usage of coffee maker varies a lot from person to person. This variability should be considered while synthesizing the activity profiles. In this activity, the driver is assumed to do either zero number of times to a maximum of one per hour of the awake time. The Activity is done for a fixed interval of 10 minutes each. It is assumed that the driver is very likely to do this activity in the last 10 minutes of the hotel period as well.

Microwave is assumed to be used twice during the hotel period. One out of which is between the start of hotel period and one is after waking up and before the end of the hotel period. The total time for this activity is done is for a fixed interval of five minutes. It is also assumed that the driver does not do this activity in the last 30 minutes of the hotel period just to account for the fact that they would be preparing for the journey and performing the last minute checks.

To avoid the redundancy in the synthesized data for training purpose, permutation and combination is used to calculate how many combinations of such activities are possible with the above discussed variability in the data. The total number of combination is equal to the number of possible unique data sets. For this calculation, the problem is conservatively simplified such that the sleeping activity can take 5 values between 4 to 8 hours. This introduces 5 different cases where different combinations of using microwave and coffee maker calculated. Adding up the number of combinations of these 5 cases, it is concluded that it is possible to generate $3.8e + 10$ unique data sets.

4. ALGORITHM SETUP

The LSTM algorithm is set up as a regression problem with the input as a sequence and the output as the next value in the series (many-to-one) in a predictor response format where the response for a time step is added to the predictor and the combined becomes the predictor for the response of the next time step. Activity prediction is a multi-variate problem, which can be converted to uni-variate by adding the load ratings of all the active devices at a particular time and predicting this total power load. This 1-D matrix would capture the information of multiple activities happening together and also would eliminate the need to create new categories in a classification problem. Time is discretized at 1min to produce a 1-by-600 matrix containing the total power load profile for a 10 hours of hotel period . The input to the algorithm for training is a 25-by-599 matrix representing one day. Out of the 25 rows, 12 rows store the Time Allocation Matrix [TAM] as a 12-by-599 matrix, 12 rows store the Transition Matrix [TM] in a 12-by-599 matrix and 1 row stores the power load profile, P is a 1-by-599 matrix, hence it is fed to the algorithm as $[\text{TAM TM P}]^{\text{T}}$. The response is the power load at the next time step. For the prediction, the activities at the first minute of hotel period of the test day along with the relational information is provided as a 25-by-1 matrix and the algorithm predicts the power load. This value is appended to the power load prediction series and fed to the algorithm again to predict the power load value next in line until the end of the 10 hours.

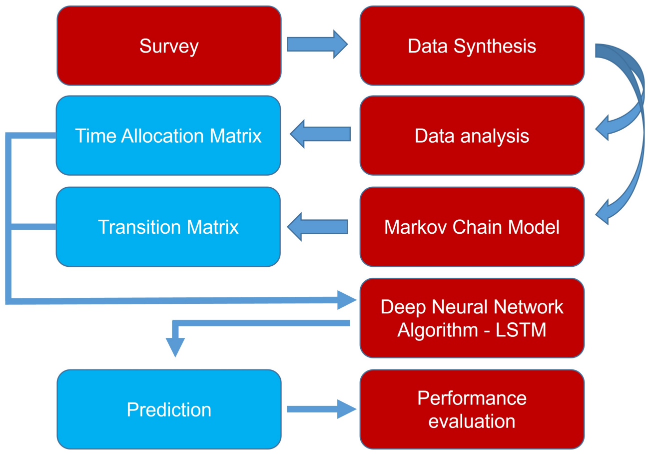

Fig. 4. Process description for algorithm setup

The temporal and relational information of the activities is very important for the algorithm to learn and are discussed in detail below.

The temporal information is given in the form of a time allocation matrix which stores in the information of the probability distribution of each of the activities. Figure 5 gives the probability distributions of three activities (sleep, Microwave, and Coffee maker). It can be seen that it is highly probable that the driver is sleeping at the middle of the hotel period. An intuition can be drawn from it. Since the mean of the sleep distribution is 5.5, hence, no matter when the driver starts to sleep, chances of them being asleep at the 5^{th} hour is very high. Also, it can be seen that the chances of microwave and coffee maker is high after sleeping. This behavior is also expected as the they used when the driver is not sleeping. Also, the probability of making coffee at the end of the hotel period is very high, this is because in the rule based data generation (discussed in section 3), very high weightage is given to the usage of coffee machine before the journey is resumed, i.e, end of the 10hr hotelling period.

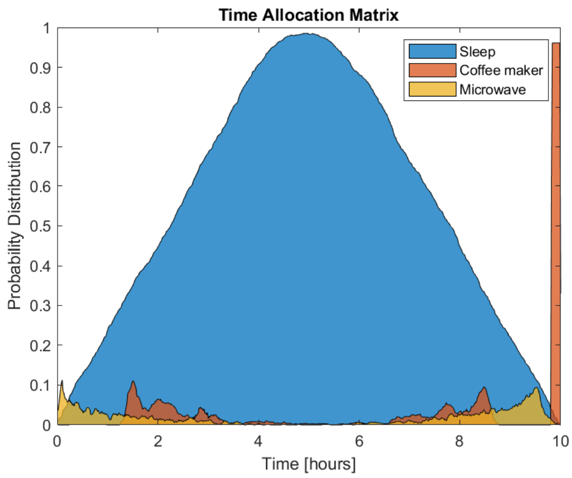

Fig. 5. Time Allocation Matrix for three main activities (Coffee Maker, Microwave and Sleep) during the hotel period

The transition matrix stores the relational information of the activities. It basically tells that the probability of the next activity in line [Gagniuc (2017)]. Using the Markov property , an n-by-n matrix for n-activities can be generated. In this n-by-n matrix, the element l-by-m $[l, m \in n]$ would store the probability of going from activity l to activity m at any instant. Hence, the rows of the transition matrix add up to 1.

The Markov property, however, is valid only when 1 activity is being performed at a given time instance. In the hotel period of long-haul truck, there can be multiple activities happening at the same time. This is handled by setting breakpoints on the time stamps and calculate the conditional probability of all the activity next in line. The fundamental difference in this method is that the sum of probabilities of all the activities following a particular activity will be more than 1. We can use normalization to force the sum of the rows to be equal to 1, but it was fond that doing so did not bring any improvement in the prediction, so was kept as is. In the presented work, there are 12 activities, hence a 12-by-12 transition matrix has been generated. As discussed above, in transition matrix, the rows do not add up to 1. Such as the device corresponding to Activity 2 is switched "ON" for the 10 hours hotel period, the probability it will be switched on after a given activity is always 1. Moreover, the second column of the transition matrix is always 1.

To feed this information to the learning algorithm, the activities that are currently being done are recognized, and the transition probabilities from the activities are added together. For instance, if Activity 5, Activity 10 and Activity 11 are being done at time 't', the rows of the transition matrix corresponding to these activities are added and this 12-by-1 matrix is transposed and appended to the input. Hence, for all 600 time steps, a 12-by-600 matrix represents the the probabilities of the activities using relational information.

Following are the steps to summarized the developed algorithm:

(1) Split the data into training and test sets in a 9:1 ratio
(2) Create predictor $x^{<1:T_{x-1}>}$ and response $x^{<2:T_x>}$ for the training sequences where T_x is the length of the series.
(3) Row wise append the predictor with TAM and TM, i.e., create the input matrix. and train the network.
(4) Use the trained network to predict the test day. An initial guess of $x^{<1>}$ is 0 and a prediction $y^{<1>}$ is made. This prediction is column wise appended to $x^{<1>}$ as $x^{<2>}$ and a prediction for 3rd time step, i.e., $y^{<2>}$ is made until $y^{<T_y>}$

The power load is normalised before being fed as input to the algorithm. This was observed to have improved the performance. MATLAB neural network toolbox is used. The architecture of the LSTM is:

(1) Input Layer: With 25 features
(2) LSTM Layer: With 50 hidden units
(3) LSTM Layer: With 50 hidden units
(4) Fully Connected Layer: Multiplies the input weight matrix and adds the bias vector
(5) Regression Layer: Compares the mean squared error

It was seen that having a multi-layer LSTM worked better in capturing the minor trends in the data however having lot of LSTM layers increased the computation time exponentially. Two LSTM layers seemed to work perfectly well without compromise on the computation time. For training the algorithm, we use adam solver with initial learning rate of 0.005 and gradient threshold of 1 for 250 epochs.

5. VALIDATION AND SIMULATION RESULTS

As discussed in section 2 DTW is used to characterise the error. The number of hidden units and the period of training are chosen as the training parameters. The number of hidden units chosen are [20 50 100 150] while the training period is [10 20 40 70 100 140 200]. With large number of hidden units the model is able to learn more relations between the events of the time series. While this is a desirable feature, it compromises on the computation speed and the risks over fitting. Similar is the case with the length of training set. Figure 6 shows a surface plot of the error matrix, that is the euclidean distance measured point wise on the warped time axis, created using dynamic time warping of the test day power load profile and the predicted power load profile.

It was seen that increasing the number of hidden units or number of training days alone did not improve the accuracy. Having small number of hidden units and training days did just as good of a job in the predicting the series as

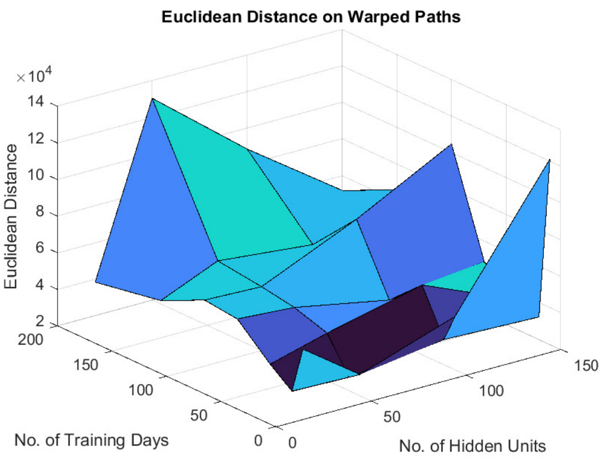

Fig. 6. Euclidean distance comparison for algorithm trained over test periods and hidden units.

very high number of hidden units and training days. The global minimum error was found to be with 50 hidden units and a training period of 20 days. Figure 7 (top) shows the forecast and the prediction as is while Figure 7 bottom shows the warped version. The warped version is a manipulation of the time axis such that DTW algorithm finds the minimum euclidean distance between the two signals. Though in some cases warping the time axis can be undesirable to check the accuracy of a prediction, however in this case however it is an acceptable metric. This is because it is more important for the algorithm to predict an event (in this case the total power load value) and in more general sense the pattern, rather than the exact time it is supposed to happen. In other words, it is important for the supervisory control on the truck's ECU to know if there is going to be a surge in the power demand and hence prepare the battery for it.

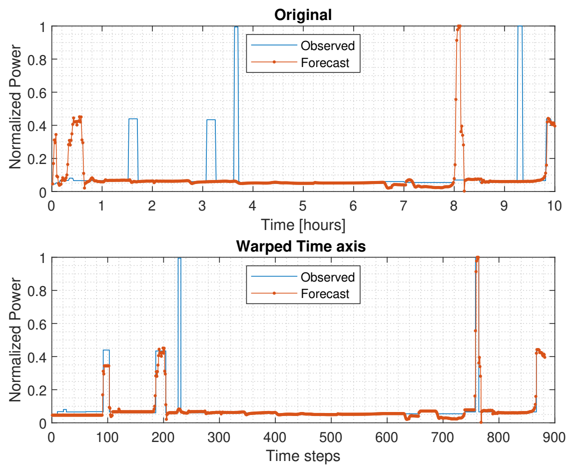

Fig. 7. Predicted power load profile vs test day load profile

6. CONCLUSION

In this paper a multivariate time series prediction problem is discussed and reduced to a uni-variate prediction problem. A combined effect of all activities as the total power load is predicted at a given time instead if predicting the individual power load ratings of the various devices inside the cabin of a long-haul truck. This algorithm is trained on synthetic data generated using observations and judgements from a baseline profile created from survey data. A multi-layer LSTM with each layer with 50 hidden units is trained on total of 20 days. The total time required for this training is about 4 minutes on a 32GB RAM, 2.2Ghz clock rate and 64 bit processor. Preference is given to predicting an event (an event being a particular load value) over the exact time of the event happening, hence a DTW is used as a performance qualification criteria.

ACKNOWLEDGEMENTS

The authors thank the United States Department of Energy and PACCAR Inc. for the support for this project. This material is based upon work supported by the Department of Energy, under Grant Number DE-EE0008265.

REFERENCES

Administration, F.M.C.S. (2011). Interstate truck driver's guide to hours of service.

Aztiria, A., Augusto, J.C., Basagoiti, R., Izaguirre, A., and Cook, D.J. (2012). Discovering frequent user–environment interactions in intelligent environments. *Personal and Ubiquitous Computing*, 16(1), 91–103.

Gagniuc, P.A. (2017). *Markov chains: from theory to implementation and experimentation.* John Wiley & Sons.

Goutham, M., Stockar, S., Blaser, R., and Hanumalagutti, P. (2021). User activity sequence prediction in smart homes using multi-layer long short-term memory networks. *IFAC-PapersOnLine*, 54(20), 901–906.

Goutham, M. (2020). *Machine learning based user activity prediction for smart homes.* Master's thesis, The Ohio State University.

Graves, A. (2013). Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*.

Khuntia, S., Hanif, A., Singh, S.P., and Ahmed, Q. (2022). Control oriented model of cabin-hvac system in a long-haul class 8 trucks for energy management applications. Technical report, SAE Technical Paper.

Marufuzzaman, M., Reaz, M., Ali, M.A.M., and Rahman, L. (2015). A time series based sequence prediction algorithm to detect activities of daily living in smart home. *Methods of information in medicine*, 54(03), 262–270.

Minor, B., Doppa, J.R., and Cook, D.J. (2015). Data-driven activity prediction: Algorithms, evaluation methodology, and applications. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 805–814.

Mitler, M.M., Miller, J.C., Lipsitz, J.J., Walsh, J.K., and Wylie, C.D. (1997). The sleep of long-haul truck drivers. *New England Journal of Medicine*, 337(11), 755–762.

Müller, M. (2007). Dynamic time warping. *Information retrieval for music and motion*, 69–84.

Smith, D., Ozpineci, B., Graves, R.L., Jones, P., Lustbader, J., Kelly, K., Walkowicz, K., Birky, A., Payne, G., Sigler, C., et al. (2020). Medium-and heavy-duty vehicle electrification: An assessment of technology and knowledge gaps. Technical report, Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States).