Spectrum Usage Analysis And Prediction using Long Short-Term Memory Networks

Anneswa Ghosh
Microsoft
USA
anneswa.ghosh@gmail.com

Jacobus Van der Merwe
University of Utah
USA
kobus@cs.utah.edu

Sneha Kumar Kasera
University of Utah
USA
kasera@cs.utah.edu

ABSTRACT

The tremendous growth of wireless services has created an ever-increasing demand for the radio frequency spectrum. However, most of the spectrum, especially in the sub-6 GHz frequency ranges, have been allocated. Given the observation that a large part of the allocated spectrum remains unused in various locations and at different times, dynamic spectrum access technologies that allow for opportunistic use of the allocated bands when they are idle are being developed. In this paper, we study the spectrum usage in the frequency range of 700 MHz to 2.8 GHz at Salt Lake City, Utah. Our study indicates that several portions of these frequencies are under-utilized, with an average of only 19% usage. Furthermore, we observe that certain frequency bands demonstrate clear usage patterns, e.g., show higher utilization during the daytime compared to night-time, that can be exploited for opportunistic secondary usage of the spectrum. We propose a spectrum prediction system using Long Short-Term Memory (LSTM) neural networks to predict the occupancy of a channel in future time slots. We further introduce an LSTM based Window Selector to find the optimal window of future forecasts that increase the utilization of the network while minimizing the interference caused by the opportunistic user. Our experiments show that the Multivariate LSTM model can be reliably used to guide the choice of the channel for the opportunistic user. The multistep LSTM models can be used to forecast spectrum usage with approximately 96% accuracy on the frequency bands exhibiting discernible usage patterns.

CCS CONCEPTS

• **Networks** → **Network properties**; • **Computing methodologies** → **Machine learning**.

KEYWORDS

machine learning, wireless testbed, spectrum power, EWMA

ACM Reference Format:
Anneswa Ghosh, Jacobus Van der Merwe, and Sneha Kumar Kasera. 2023. Spectrum Usage Analysis And Prediction using Long Short-Term Memory Networks. In *24th International Conference on Distributed Computing and Networking (ICDCN 2023), January 4–7, 2023, Kharagpur, India.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3571306.3571412

ICDCN 2023, January 4–7, 2023, Kharagpur, India

ACM ISBN 978-1-4503-9796-4/23/01…$15.00
https://doi.org/10.1145/3571306.3571412

1 INTRODUCTION

The tremendous growth of wireless services has created an ever-increasing demand for the radio frequency spectrum. However, most of the spectrum, especially in the sub-6 GHz frequency ranges have been allocated. Given the observation that a large part of the allocated spectrum remains unused in various locations and at different times, dynamic spectrum access technologies that allow for opportunistic use of the allocated bands when they are idle and are being developed. In this paper, we study the spectrum usage in the frequency range of 700 MHz to 2.8 GHz at Salt Lake City, Utah. Our study indicates that several portions of these frequencies are under-utilized, with an average of only 19% usage. Furthermore, we observe that certain frequency bands demonstrate clear usage patterns, e.g., show higher utilization during the daytime compared to night-time, that can be exploited for opportunistic secondary usage of the spectrum. An opportunistic secondary usage of such frequencies involves frequently scanning the bands and determining their occupancy (spectrum sensing). An opportunistic user cannot transmit in a channel before sensing and determining its occupancy, as that may cause interference. This possesses a significant challenge, as these operations need to be performed in each time slot, causing substantial delays before the user gains access leading to reduced utilization. A system that can predict the state of the channel for future time slots can reduce the delay and the energy consumed in spectrum sensing and the decision making phase if the sensed channels are free and finalize the best channel for the opportunistic use (spectrum decision). We propose a spectrum prediction system using Long Short-Term Memory (LSTM) neural networks to predict the occupancy of a channel in future time slots. We further introduce an LSTM based Window Selector to find the optimal window of future forecasts that increase the utilization of the network while minimizing the interference caused by the opportunistic user. Our experiments show that the Multivariate LSTM model can be reliably used to guide the choice of the channel for the opportunistic user. The multistep LSTM models can be used to forecast spectrum usage with approximately 96% accuracy on the frequency bands exhibiting discernible usage patterns.

Our contributions in this paper are summarized as follows: (i) We build spectrum monitoring tools in the POWDER testbed and study the spectrum usage pattern for several frequency bands in Salt Lake City, Utah. (ii) We develop LSTM based models to forecast spectrum power values using real-world power spectral density (PSD) data that we collect. (iii) We also develop LSTM based architectures to forecast multiple timestep spectrum usages at once using real-world

PSD data that we collect. (iv) We use the data collected to evaluate the performance of various deep learning models (Stacked LSTM, Encoder-Decoder LSTM, Multivariate LSTM) and compare them with the performance of two baseline approaches (Exponential Weighted Moving Average (EWMA) and Zero Rule Algorithm). (v) We introduce a novel LSTM based Window Selector to select the best window to minimize interference and maximize throughput.

2 RELATED WORK

The US Government is taking initiatives for allowing sharing of the under-utilized frequency bands by supporting dynamic spectrum access technologies and facilitating further research on how to effectively access the spectrum holes [1]. However, without a proper understanding of the current and future spectrum usage, these initiatives would not be able to achieve their goals. Spectrum survey is an essential tool to determine the current spectrum usage and guide the policy makers to make the most informed decision. Spectrum surveys have been conducted in San Francisco [2], Denver [3], and San Diego [4]. Spectrum survey in Singapore [5] showed that except for the bands allocated for broadcasting (analog TV, digital TV, HDTV, and DAB) services and cellular networks, most are heavily underutilized, with only 4.54% average usage in the entire frequency bands ranging from 80 MHz to 5850 MHz. Spectrum surveys have indicated that the frequencies are being under-utilized because of static allocation schemes [6]. Cognitive Radio Network (CRN) has been introduced to enable unlicensed opportunistic users to communicate in idle time slots with no harmful interference to the licensed user [7]. To accomplish this, spectrum sensing must be performed to determine the current spectrum state, and to avoid any harmful interference. Spectrum prediction can alleviate the opportunistic user from conducting spectrum sensing in every timeslot by predicting the future spectrum states. This can save a lot of time and energy, thus improving the throughput of the network [8]. A survey [9] on spectrum prediction shows that most of the existing studies are based on classical statistical techniques or shallow architecture models. While deep learning has shown promising results in many applications of image recognition, machine translation, natural language processing, target detection, etc., its use in spectrum prediction is still in its budding state [10]. Hochreiter et al. [11] introduced the Long Short Term Memory (LSTM) network to learn long-term dependencies. Predicting the state of the channel is a time-series problem that can leverage long-term dependencies learning. The study in [12] applies deep learning to predict spectrum availability in cognitive aerospace communications; however, real-world data are converted into binary channel states like other prediction algorithms. Yu et al. [10] applies LSTM for spectrum prediction in the frequency hopping communication where frequency hopping sequence is also a binary time series artificially generated data. Spectrum prediction of one timestep on the power spectral density (PSD) values using LSTM is studied in [14], but one-time step prediction is not very useful for the opportunistic user.

Our work differs from the existing work in the following significant ways: First, we perform a spectrum usage study using the POWDER platform in Salt Lake City, Utah and analyze spectrum usage patterns. Second, we develop models for multiple timestep spectrum usage forecasts using real-world power spectral density (PSD) data that we collect. Third, we evaluate the performance of various deep learning models (Stacked, Encoder-Decoder, and Multivariate LSTM) and compare that with the performance of two baseline approaches (Exponential Weighted Moving Average and Zero Rule Algorithm). Last, we study the effects of the forecast window size and introduce a novel LSTM based Window Selector to select the best window to minimize interference and maximize throughput.

3 SPECTRUM USAGE ANALYSIS

We study the spectrum occupancy in Salt Lake City, Utah, using the the POWDER platform. POWDER [15] is a city-scale laboratory with radio equipment, fiber infrastructure, edge-compute, and datacenter/cloud resources for research on future wireless networks. We collect data using the Fixed Endpoint installation of the POWDER Platform located at the University of Utah's main campus. The Fixed Endpoint experiment setup consists of an ensemble of Software Defined Radio (SDR) equipment from National Instruments (NI) and a compute node. Figure 1 represents the high-level experiment setup. The Fixed-Endpoint equipment used for this study is NI USRP B210 SDR with its ports connected to dedicated Taoglas GSA.8841 wideband I-bar antenna. This antenna has a frequency range of 698-6000 MHz and has an approximately -2 dBi average gain across the range. The USRP is also connected to an Intel NUC, a small form factor PC, via USB 3.0.

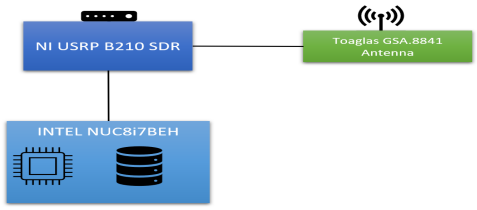


Figure 1: Experiment setup

3.1 Spectrum Power Measurement

The USRP SDR device is accessed using the Python APIs provided by the USRP Hardware Driver (UHD). These APIs are used to set receive gain and to acquire I/Q samples from a specific channel at the requested sample rate. The frequency range is divided into bands, each having a frequency width of 30MHz. Each 30MHz band is further divided into 200 points, i.e., the distance between two consecutive frequency points is 150KHz. These frequency points are represented by the center frequency of the 150KHz wide channel. The raw data collected for each frequency point is the signal power computed at the USRP. Figure 2 illustrates an example of how frequency division is performed for the spectrum measurement. The example frequency range 2300-2390 MHz is divided into three sections, each of which spans 30 MHz. This 30 MHz band is further subdivided into 200, 150 KHz channels, and the respective center frequencies are used to represent them. The upper boundary of each 150 KHz channel is shown in the top blocks of Figure 2.

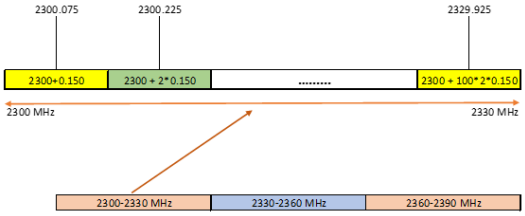


Figure 2: Illustration of frequency division in data collection

3.2 Energy Detector-Based Sensing

We use energy detector-based sensing for identifying the presence of signal transmission [16]. Energy detector-based approach does not need any prior knowledge of the signal. The signal is detected by comparing the output of the energy detector to a predefined threshold, T [19]. The decision metric of the energy detector can be written as follows:

$$S = \sum_{n=1}^{N} |y(n)|^2 \tag{1}$$

where N is the observation interval. The decision on the occupancy of the band is made as follows:

$$D = \begin{cases} 0 & S < T \\ 1 & S \geq T. \end{cases} \tag{2}$$

The threshold, T, is determined as a function of the Johnson-Nyquist thermal noise power (NP) [17, 18] and the noise figure (NF).

$$T = f(NP, NF) \tag{3}$$

where,

$$NP = 10log_{10}(k\tau\Delta f 1000).$$

k is the Boltzmann constant, τ is the temperature, and Δf is the noise bandwidth given in Hz. The Noise Figure (NF) of the system is defined as

$$NF = 10log_{10}(\frac{SNR_i}{SNR_o}). \tag{4}$$

SNR_i and SNR_o are the input and output signal-to-noise ratios (SNR), respectively. NF is obtained through calibrated measurements of the RF Hardware [13]. NP is the electronic noise generated by the thermal agitation of the electrons inside an electrical conductor at equilibrium. This noise is present in every electrical circuit. The noise figure represents the degradation in the signal to noise ratio as the signal passes through a device in the dB scale. Hence, these two quantities help determine the minimum power needed for a signal to be detected.

3.3 Analysis of the Collected Data

We collected spectrum data for 5 days from March 8th, 2020, 11:00 PM to March 13th, 2020, 11:00 PM over the frequency range 700-2800 MHz. Figures 3-5 show the spectrum usage patterns for three different frequency ranges. Each figures shows usage for 300 frequencies. The red dots indicate that the channel is occupied. We also show the US Department of Commerce spectrum allocation categories for each frequency range. From these figures, it becomes evident that a vast range of frequencies is either under-utilized or

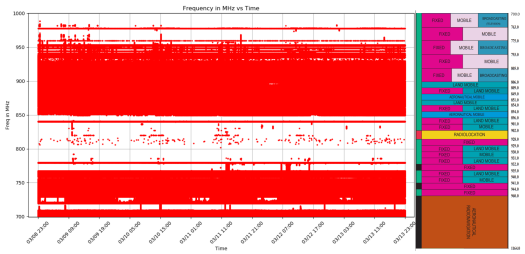


Figure 3: Spectrum allocation and usage for 700-1000 MHz

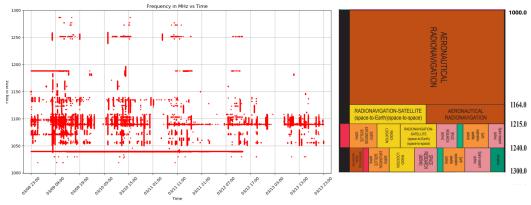


Figure 4: Spectrum allocation and usage for 1000-1300 MHz

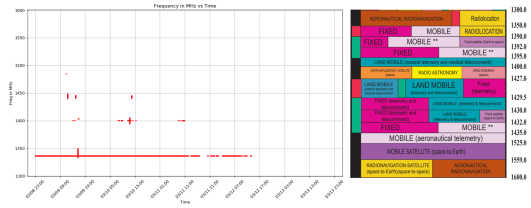


Figure 5: Spectrum allocation and usage for 1300-1600 MHz

not used at all. This behaviour is further illustrated in Figure 6, where we show the allocation categories, the frequency range, and their usage percentage. Additionally, Figure 7 shows that the usage is significantly low at night times in various frequencies. Specifically, from this figure we observe that there are 12 frequency ranges that show significant usage difference between day and night times. We consider 11:00 PM to 7:00 AM as our night hours. We make the following observations from our spectrum usage analyses. First, low or no occupancy is observed in bands allocated to radio navigation, aeronautical radio navigation, earth exploration, space research, amateur, and fixed satellite services. Second, the average usage of the whole spectrum [700-2800 MHz] is only 19.08%. Third, the majority (71.4%) of the bands have 0-20% usage, while only 5.7% of the bands have 80-100% usage, as shown in Figure 8. Fourth, relatively high occupancy is observed in bands allocated for broadcast services. Last, among the bands that have significantly lower night-time usage, fixed and mobile are the common services.

4 SPECTRUM PREDICTION

Spectrum prediction allows predicting the power of a channel and, in turn, the occupancy state of the channel for future time slots. It provides the following three key advantages:

- Conserves time and energy spent on spectrum sensing
- Gives an insight into the best frequency to be used

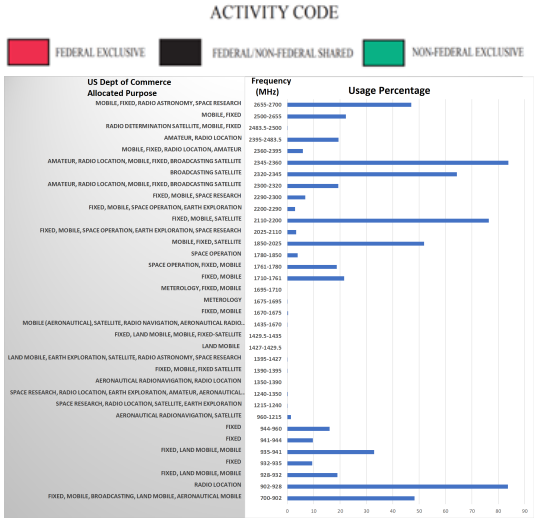


Figure 6: Spectrum allocation category and usage percentage

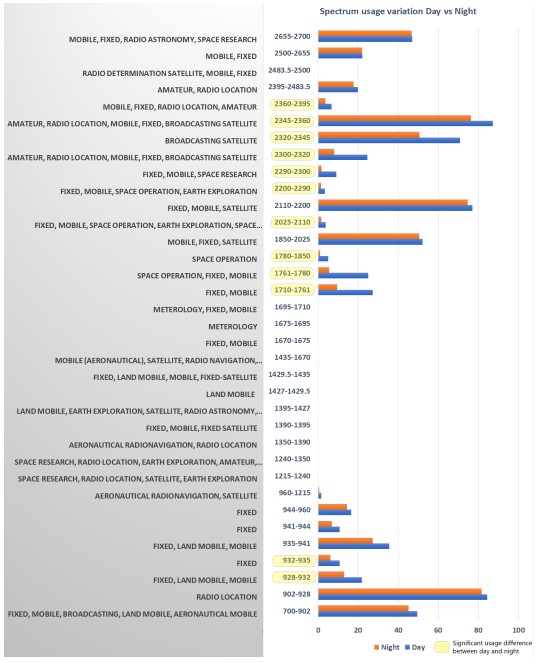


Figure 7: Spectrum usage difference between day and night

- Shows the state of the frequency bands for a large number of future slots

Without spectrum prediction, an opportunistic user (OU) must perform spectrum sensing for a large set of frequencies and determine the presence of a signal in each of them before actually using the frequency to transmit. Spectrum prediction also allows an OU to select the best channel that has the highest potential to be available when predicted by different mechanisms. We consider a wireless communication system where transmissions are performed in well-defined

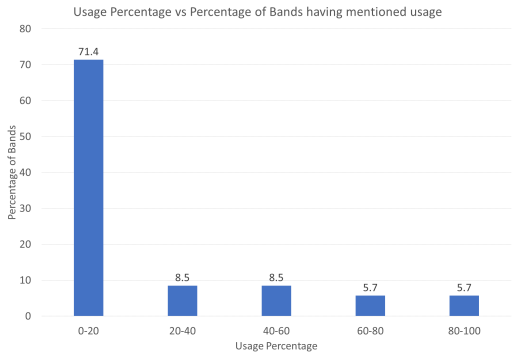


Figure 8: Usage percentage versus percentage of bands

time slots as in a time-division multiplexed system. The OU needs to perform spectrum sensing and spectrum decision in each time slot before the transmission. Our motivation for spectrum prediction is based on the presence of temporal variations in the spectrum usage, as shown in Section 3.3. We develop a spectrum prediction framework with different Long Short-Term Memory (LSTM) architectures in deep learning and present extensive experimental results with real-world spectrum data that we collect. Deep learning has proven to be a very valuable tool in various fields, including image processing and natural language processing. Applying deep learning methods for spectrum prediction remains an active research area with the potential for significant improvements [10]. Complex LSTM models in deep learning with multilayered networks have the ability to perform well on time-series prediction [21, 22]. LSTM is an artificial recurrent neural network architecture that can learn long-term dependencies. An LSTM unit is composed of a cell, an input gate, an output gate, and a forget gate. The cell remembers values over arbitrary time intervals, and the three gates regulate the flow of information into and out of the cell [11]. We explore the following LSTM architectures for spectrum time-series prediction.

4.1 Multistep Univariate stacked LSTM (MSUL)

LSTM networks can learn to forecast long sequences in one shot. We leverage this to predict multistep spectrum power values of a selected frequency; i.e., given the historical power observations at the time, (t-1, t-2, ..., t-k), the model predicts power values for time (t, t+1, . . . , t+m) using only the power values as the input feature.

4.1.1 Supervised Data Creation. We collect data over consecutive slots and predict the spectrum occupancy of the future slots based on this data. Figure 9 shown an example where both input and forecast window size selected as 5. The values in the slots, (t1, t2, . . . , t5) of the first spectrum prediction point represent the power values in the shown timeslot. The corresponding label of the data point is represented by the values in the slots (t6, t7, . . . , t10). The power values are received over time, so the dataset is created by sliding a window with a fixed length, L. The power values are normalized to values between 0 and 1. This normalization makes the training and convergence faster and also helps in learning the problem effectively. The results we present in this document are

obtained using the input window size of 30 and the forecast window size set to 100.

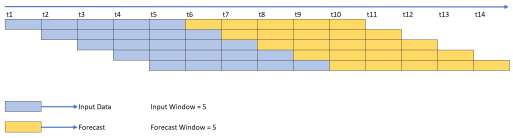


Figure 9: Dataset Collection and Prediction

4.1.2 Univariate Stacked LSTM Network. Our first model is based on a univariate unidirectional stacked LSTM architecture consisting of two hidden layers and one dense layer with 100 hidden states in every layer (see Figure 10). We use the learning rate of 0.001, RELU activation function, and mean squared error as the loss function. For our experiments, we create the dataset by moving the window one step at a time. In Figure 10, the input contains power value for a frequency for 9-timesteps (t0 to t8). As we have set the input step size or input window size as 3, the raw input data are reshaped into four samples. We use 50% of the collected data for training, 20% for validation, and the remaining 30% for testing. Each reshaped sample contains the input data for 3-time steps. In the training phase, we feed the next 3-time steps' data as the label.

4.2 Single-step Multivariate Stacked LSTM (SSML)

Neural Networks, like LSTM networks, can model problems with multiple input features. This is a great benefit of time-series forecasting over classical linear forecasting methods. We predict a single future power value for each of the input frequencies. This prediction can help the OU gain insight on which frequencies can be reliably predicted by the multistep forecast models.

4.2.1 Supervised Data Creation. The dataset is constructed using a sliding window, as shown in Figure 11. Suppose the sliding window

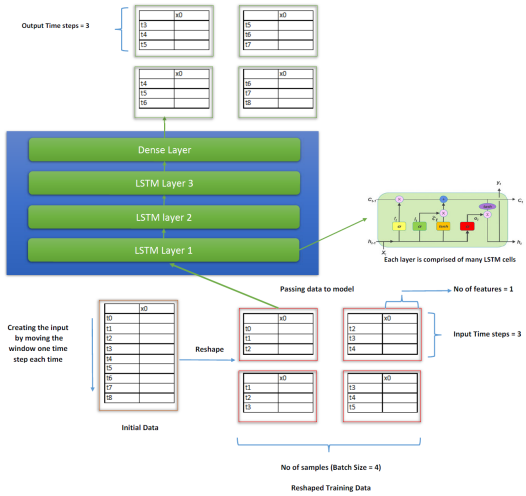


Figure 10: Univariate stacked LSTM architecture

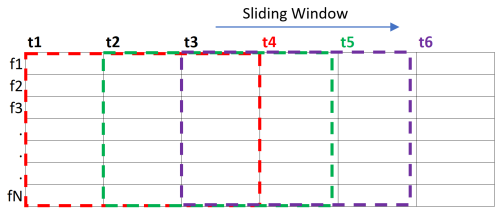


Figure 11: Dataset for single-step multivariate LSTM

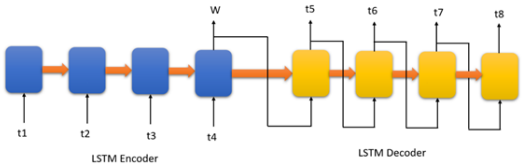


Figure 12: Encoder-decoder LSTM

length is 3; the label for the training is the power value of all the channels of time slot t4, and the model outputs the predictive power values of corresponding channels for the same time slot t4. The window is forwarded one timestep ahead to form the next datapoint.

4.2.2 Stacked LSTM Network. The structure of the LSTM network constructed in this case has five hidden layers, all of which are comprised of LSTM layers. The first layer has 300, the second layer 200 hidden, the third layer 100, the fourth layer 70, and the fifth layer has 50 hidden units. Each layer has RELU as the activation function, and the output of one layer is fed as the input to the next layer. The output of the last hidden layer goes into a Dense layer. We use Mean Square Error as the loss function, and the Adam optimizer is used for updating network weights. With input frequencies from $F_1, F_2, \ldots, F_N$, N input features are used, and the model predicts one power value corresponding to each of the N input frequencies.

4.3 Multistep Encoder-Decoder LSTM

The Encoder-Decoder LSTM is a special type of Recurrent Neural Network designed to solve sequence-to-sequence (seq2seq) problems. Given the multiple time steps as the input and multiple time steps as the output, this type of problem is referred to as many-to-many sequence prediction problem. The Encoder-Decoder LSTM is proven to be very effective in seq2seq prediction problems.

4.3.1 Supervised Data Creation. The dataset is constructed using a method similar to the one described in Section 4.1.1. Given the historical power observations at the time (t-1,t-2,...,t-k) for a particular frequency, the model predicts power values for time (t,t+1,...,t+m) using only the power values as the input feature. However, the label, in this case, has three dimensions - samples, timesteps, and features, unlike the previous case that had only two dimensions. The sliding window mechanism is used to create the entire dataset.

4.3.2 Encoder-Decoder LSTM Network. This architecture comprises of the encoder model and the decoder model. The encoder model

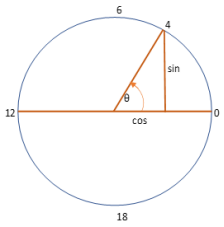


Figure 13: Time unit circle

is used for reading the input sequence and encode it into a fixed-length vector, and this fixed-length vector is then fed into the decoder model that outputs the predicted sequence. Let t1, t2, …, t4 be the input sequence that is fed into the encoder LSTMs and t5, t6, …, t7 be the predicted sequence from the decoder LSTM. The decoder model makes a one-step prediction for each element in the output sequence. An example encoder-decoder model is shown in Figure 4.4. This is the subtle difference from the previously seen stacked LSTM architecture in Section 4.1; as in practice, both approaches predict an output sequence of power values.

4.4 Multistep Multivariate Stacked LSTM (MSML)

So far, we have done multistep spectrum prediction with only a single input power variable or feature. Real-world spectrum prediction becomes more challenging when we need to include more than one feature and yet be able to predict power/occupancy across multiple time steps. This specific architecture of multistep multivariate stacked LSTM has the ability to handle multiple input variables and be able to predict power values for more than one timestep.

4.4.1 Supervised Data Creation. We use the power and time features to evaluate this architecture. Time, however, is cyclical in nature; that is, hour 23 and hour 0 are closer to each other than hour 0 and hour 3. We transform the time data such that this cyclical property is preserved. The time feature is transformed into two new features, x, and y where (x,y) represents a coordinate of a unit circle. We compute the x and y components using the sine and cosine trigonometric functions. Figure 13 shows an example unit circle for hours, where x is the cosine component, and y is the sine component. Midnight (0) is on the right, and the hour increases counterclockwise. Thus, hour 23 is very close to hour 0. The transformations (x,y) in our experiments are performed as follows.

$$x = cos\ (2\pi(^{\Delta s}/_{S})) \quad (5)$$

$$y = sin\ (2\pi(^{\Delta s}/_{S})) \quad (6)$$

where Δs represents the seconds since midnight, and S represents the total number of seconds in a day.

We follow the sliding window approach to create the entire dataset. Every data point has three features power, cosine transformed, and sine transformed components. Our sliding window approach is shown in Figure 14. In this figure, three timesteps are used to predict the next two time steps; t4 and t5 are predicted in the first data point represented by the red dotted box. Similarly, t5, t6, and t6, t7 are predicted in the second and third data points.

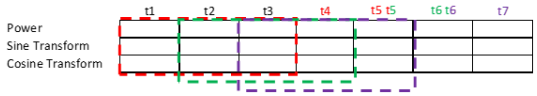


Figure 14: Data creation for multistep multivariate LSTM

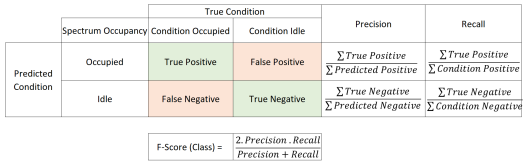


Figure 15: Confusion matrix elements

4.4.2 Multivariate Stacked LSTM Network. The structure of the LSTM network constructed in this case has three hidden layers, which are all comprised of LSTM layers. Each layer has 100 hidden units and RELU as the activation function, and the output of one layer is fed as the input to the next layer. The output of the last hidden layer goes into a Dense layer. We use Mean Absolute Error as the loss function, and the Adam optimizer is used for updating network weights. The model is trained for a particular frequency to predict power values of 100 timesteps with three input features.

5 EVALUATION

We evaluate the performance of all the models based on a walk-forward validation. The test dataset is provided to the model by progressing one timestep each time. Thus, the model is always fed with the latest k lag observations, where k is the input window size.

5.1 Evaluation Metrics

We use the Root Mean Square Error (RMSE) to evaluate the real-valued output of the models. We use the following additional metrics to evaluate the performance of the binary occupancy output: Accuracy, F-score, False Positives, False Negatives, True Positives, and True Negatives. The RMSE imposes a severe penalty on large errors in the prediction. In our evaluation, we calculate RMSE as follows for each forecast window:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (Predicted_i - Observed_i)^2}{n}}.$$

We also convert the predicted power values to binary occupancy states by applying an appropriate threshold. The binary values are then validated using precision, recall, and F-Score as described in Figure 15.

5.2 Baselines

We also establish the baseline performance of the spectrum prediction problem. This baseline performance provides a point of comparison when evaluating the more sophisticated LSTM models. We use two types of baseline algorithms, one type for the regression part of the problem and the other for the classification part. The two baseline algorithms are as follows:

(1) Exponential Weighted Moving Average for Regression
(2) Zero Rule Algorithm for Classification

5.2.1 Exponential Weighted Moving Average (EWMA). EWMA is used to predict the label of the spectrum power data corresponding to a frequency. EWMA for a spectrum power series, V_t is calculated recursively as follows:

$$V_0 = 0 \tag{7}$$

$$V_t = \beta V_{t-1} + (1-\beta)\theta_t. \tag{8}$$

Here, β represents the degree of the decrease in the weight of older datum, a constant smoothing factor between 0 and 1. θ_t is the current power value, and V_t is the value of the EWMA at the time t. Equation (8) shows that a single prediction can be made as the value of the prediction is dependent on the current true value. Also, the prediction is dependent on a single feature of power, so multiple variables cannot be incorporated to predict future power values. Therefore, to get a baseline for multistep models, we use the same prediction value for all the predicted time steps.

5.2.2 Zero Rule Algorithm. In the Zero Rule Algorithm, we predict the most common class label in the training set. This means that if the majority of the label in the training set is label "1," then this algorithm uses a single rule of predicting only label "1" for the testing set.

5.3 Evaluation of Frequency Bands using SSML

An opportunistic user needs to make the spectrum decision by selecting the best frequency channel for its application. The Single-Step Multivariate stacked LSTM (SSML) model can give insight about the best frequency band of width 150 KHz by associating an RMSE score with each frequency band. The OU can choose the best frequencies for its application by selecting frequency bands, which are lower than a required RMSE score. The selected frequency channel can be evaluated for multistep prediction by OU. Any choice of bandwidth can be used to evaluate the SSML models. We evaluate four different frequency bands with bandwidths of 5, 10, and 12 MHz. There are, for example, 67 150KHz frequency bins in a 10MHz band. Figure 16 shows the SSML model's performance for the four different bands. These plots show how the future multivariate LSTM models are likely to perform for various frequencies, and can guide OU to make the best choice in its spectrum decision phase. The red dot in the plots represents an example frequency, and it is the corresponding RMSE given by the model. We have highlighted one frequency from each of the range and its corresponding RMSE. We use these representative frequencies to further study other models.

5.4 Frequency Selection for Multistep models

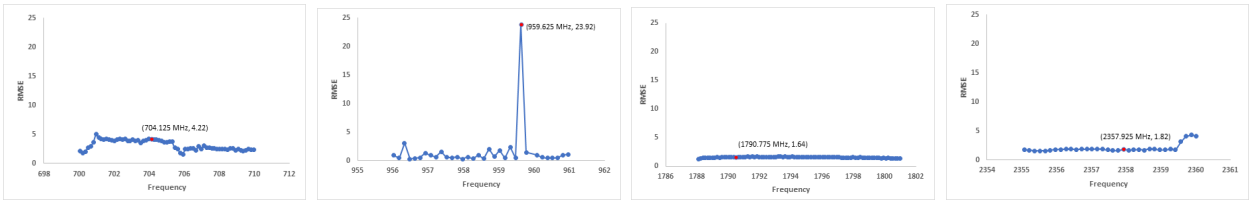
Our goal is to select frequencies that exhibit different usage patterns so that we can study the robustness of the multistep spectrum prediction models. The selection of the representative frequencies is guided by the visual inspection of the spectrum occupancy plots and the result of the SSML models. Our selected frequencies include 2357.925 MHz (very high usage in the day and significantly low usage at night), 1790.775 and 704.125 MHz (high usage in the day and relatively low usage at night) and 959.625 MHz (exhibits no clear usage patterns). The spectrum prediction is applicable and can be extended to any frequency. To be applied on a new frequency, the multistep models need to be trained on the desired frequency's usage data. The usage pattern of selected frequencies is illustrated in Figure 17. Each data point is collected every 6 seconds approximately. The red dot in these figures indicates the observed power value.

5.5 Evaluation of Multistep Models

In this section, we present our regression experiment results of the three multistep LSTM models and the EWMA base model. All the multistep models are based on 30 lag windows and a prediction length of 100 windows. Figure 18 illustrates the RMSE performance of 3 LSTM based models and the EWMA model across four representative frequencies. We compare the performance of all the LSTM models with that of EWMA in Figure 19. We see an improvement of almost 2% to 15%. From Figure 18, we observe that the 2357.925 MHz frequency band has the best forecast performance. This can be attributed to the fact that this band has regular high daytime usage and significantly low night-time usage, as shown in Figure **??**. However, given that frequency 959.625 MHz does not have any particular usage pattern, the LSTM models perform significantly poorly when compared to other frequencies. Interestingly the LSTM models perform much better than EWMA. We also see in Figure 20 that the difference between the MSML and MSUL models is minor, with only about 2.3%.

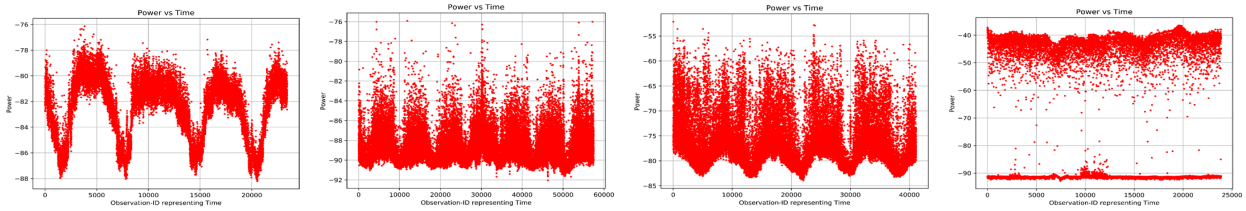
5.6 Adaptive Threshold Mechanism

To estimate the occupancy state of the spectrum, we convert the predicted power values to binary occupancy state using a power threshold. For our dataset, we observe that the LSTM's predicted power values are very smooth. The predictions do not scale to the highs and lows present in the data, but the cycles and seasonality present in the data are forecast correctly. The scaling issue poses a problem when we convert the power values to binary occupancy state using the original threshold. Figure 21 shows an example prediction of Frequency 704.125 MHz by stacked LSTM. We observe that with the correct threshold, the occupancy state can be determined effectively. We introduce an adaptive threshold mechanism that selects the best threshold by applying various thresholds in a predetermined range and selecting the one that provides the combined best F-Score and Accuracy across the forecast window. The adaptive threshold selection shown in Figure 22 explores the space ranging from -90 to -73 for selecting the appropriate threshold. It then identifies the threshold value that maximizes both accuracy and F-score across the entire forecast window. The original and selected adaptive thresholds are shown in Figure 23. Figure 23 provides a comprehensive plot of both training and test data. The blue marker indicates the true data, an initial 50% of the data is used for training, and the orange marker indicates the model's prediction on the training data. The next 20% of data is used for validation, the green marker indicates the LSTM's forecast on the validation data, and the red marker indicates the prediction on testing data. We study the impact of the adaptive threshold selection and determine that the adaptive threshold yields a better F-Score for both classes for different frequencies. Figure 24 gives an insight into the accuracy of the models after the adaptive selected threshold is applied to the LSTM models. It also compares the accuracy of the LSTM models with that of the baseline majority model. We find that the LSTM models perform significantly better than the



(a) 700-710 MHz (b) 956-961 MHz (c) 1788-1800 MHz (d) 2355-2360 MHz

Figure 16: RMSE for different frequency bands



(a) 2357.925 MHz over 4 days (b) 1790.775 MHz over 5 days (c) 704.125 MHz over 5 days (d) 959.625 MHz over 2 days

Figure 17: Usage patterns

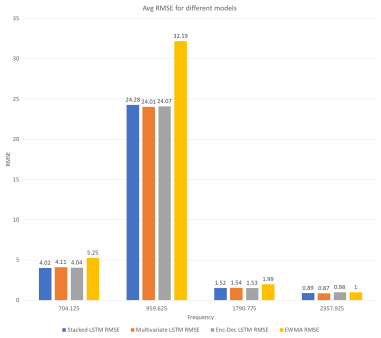


Figure 18: RMSE for 3 LSTM models and EWMA model

RMSE % Decrease in Comparison with EWMA			
Frequency	Stacked LSTM	Multivariate LSTM	Enc-Dec LSTM
704.125	11.94	9.49	11.39
959.625	9.02	10.25	9.97
1790.775	12.50	11.04	11.76
2357.925	12.36	14.94	2.04

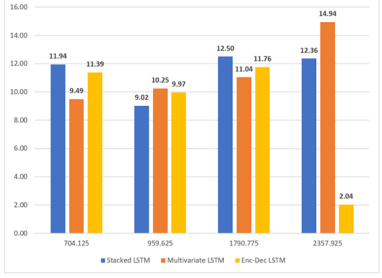


Figure 19: % increase in RMSE in EWMA compared to LSTM models

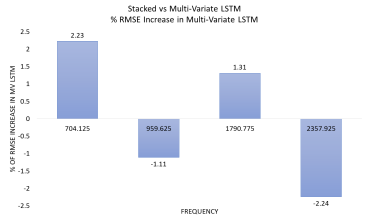


Figure 20: % increase in RMSE in MSML compared to MSUL

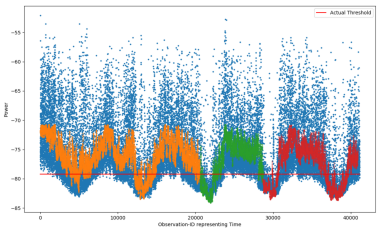


Figure 21: True vs predicted power (freq. 704.125 MHz)

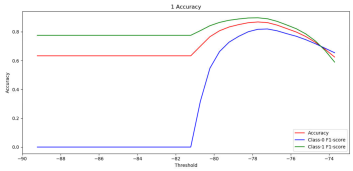


Figure 22: Adaptive threshold search

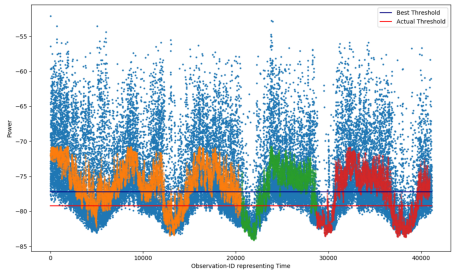


Figure 23: Adaptive threshold illustration

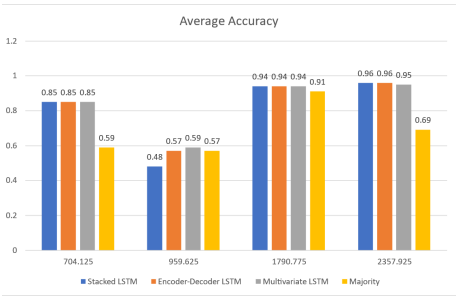


Figure 24: Accuracy of the models

baseline majority model. Approximately, 96% accuracy is observed in the frequencies of 2357.925 for all the LSTM models. All the LSTM models are almost identical in performance except at 959.625 MHz. As seen in the RMSE evaluation, prediction accuracy is also poor for 959.625 MHz, as no particular usage pattern is exhibited by this frequency. However, high prediction accuracy is seen in the bands that exhibit clear usage patterns.

6 FORECAST WINDOW LENGTH

While spectrum prediction has various advantages, the forecast window size remains a critical parameter for its applicability in real scenarios. A large forecast window increases the throughput by allowing an OU to access the channel continuously without having to pause transmission for determining the occupancy of the licensed user [20]. A large forecast window also reduces the overall energy spent by OU to sense and determine the occupancy state. However, when forecasting for larger windows using LSTM models, the error in the prediction increases with time. This phenomenon is displayed for Frequency 2357.925 MHz in Figure 25. This is also evident from the increasing rates of false-positives and false-negatives and a decreasing trend in accuracy and F-scores when considering binary spectrum occupancy states, as seen in Figures 26 and 27. Though the initial parts of the forecast window have the lowest RMSE, a small forecast window is not ideal as that involves sensing often. From our spectrum analysis, we observe that channels can have different usage patterns. Thus, each frequency may have a different forecast window that yields the best performance for that frequency.

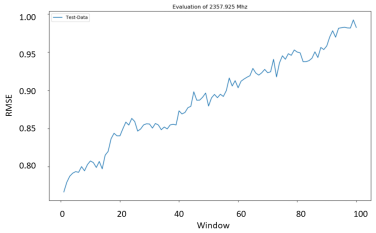


Figure 25: Increasing trend in RMSE

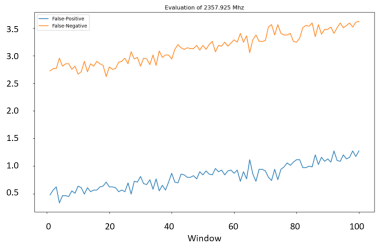


Figure 26: False positive and false negative over windows

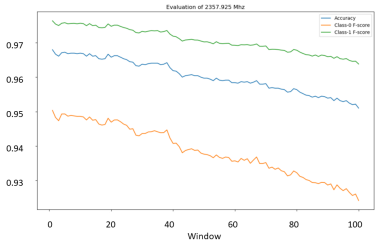


Figure 27: Accuracy and F-score over windows

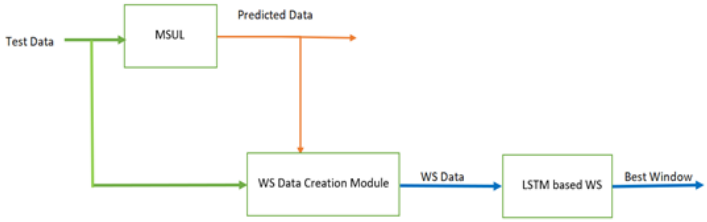


Figure 28: LSTM based window selector

6.1 LSTM based Window Selector

To find the best window length for the OU, we introduce a new LSTM based Window Selector that predicts the best window length given a sequence of past power values. This selector takes the real-world power value of a frequency and the predicted values from an already trained MSUL model and then outputs a window-length. The components of the selector are illustrated in Figure 28. 30% of the spectrum data, which is the test data of the MSUL model, is used in the training and evaluation of this window selector system.

Frequency	RMSE
704.125	6.8
959.625	1.97
1790.775	6.32
2357.925	4.55

Table 1: Window Selector Evaluation

The sliding window approach, described in Section 4.1.1, is used to create data points from the test data. The Window Selector (WS) Data Creation Module takes this created dataset as input, along with the predicted value from the MSUL model on the same dataset. It then labels each of the data points with a real number, which indicates the window length the MSUL model is able to predict correctly from the first prediction consecutively. This newly created dataset is then fed into the LSTM based WS model to train it.

A 3-layer stacked LSTM architecture is used for this model with RELU as the activation function of each layer. The Adam optimizer for updating network weights and mean square error as the loss functions are used for the training of the model. Early stopper is used on the validation set to train the model for the optimal number of epochs. The training data consists of 50% of the total data, the next 20% data is kept for the validation set and the last 30% for testing the model. A new model needs to be trained and created for each individual frequency. We evaluate this model by choosing the window length of the past values to be 20. We choose the prediction window length to be 20 as well. This model can be extended to other past and prediction window lengths. The model's RMSE performance for each frequency is shown in Table 1.

Discussion

SSML can predict the spectrum power value of multiple frequencies in a single go, whereas EWMA based model can be used only for a single frequency at a time. E.g., if one SSML model can predict future values for 100 frequencies at a time, 100 different EWMA models are needed to achieve the same purpose. Also, EWMA can predict only one step in the future, unlike robust multistep LSTM models, that can predict multiple steps in a one-shot. Our work gives an insight into how an opportunistic user can use the SSML model to evaluate various bands to narrow down the choices of the frequencies to be used for its transmission and then use multistep LSTM models to forecast the availability of the channels corresponding to the selected frequency over multiple timeslots. LSTM based Window Selector model can then guide the user to use the best length of the forecast window. All these models can be trained and used for any choice of bands and bandwidth. Multistep LSTM models can be trained and used to predict for any length of the forecast window. The hyperparameters of the models are optimized for the frequencies evaluated in this work. Multistep models can work better on a different set of hyperparameters instead of the ones used in this work when applied to a different frequency band.

7 CONCLUSION

We collected and analyzed spectrum usage data on the POWDER platform in Salt Lake City, Utah. We found that various segments of the spectrum remain underutilized. We also highlighted the segments of the spectrum where the usage pattern differs significantly between day and night. Motivated by our spectrum usage findings, we explored LSTM models to perform spectrum prediction on four bands, each exhibiting a different usage pattern over time. We introduced a new adaptive threshold parameter that significantly boosts the occupancy prediction performance. To minimize the chances of interference and maximize the throughput, we introduced a novel LSTM based Window Selector system. This system automatically outputs the best window length with the given power input data.

ACKNOWLEDGMENTS

This material is based upon work supported by the US National Science Foundation under Grant Nos. 1564287 and 1827940.

REFERENCES

[1] M. A. McHenry et al., "Chicago spectrum occupancy measurements & analysis and a long-term studies proposal," in Proc. of the First Int. Workshop on Technol. and Policy for Accessing Spectr., 2006.
[2] F. H. Sanders et al., "Broadband Spectrum Survey at San Francisco, California, May-June," US Dept. of Commerce, Nat. Telecommun. and Inf., 1995.
[3] F. H. Sanders and V. S. Lawrence, "Broadband spectrum survey at Denver, Colorado," US Dept. of Commerce, Nat. Telecommun. and Inf., 1995.
[4] F. H. Sanders, "Broadband spectrum surveys in Denver, CO, San Diego, CA, and Los Angeles, CA: methodology, analysis, and comparative results," IEEE Int. Symp. on Electromagn. Compat. Symp. Rec. (Cat. No. 98CH36253), 1998.
[5] M. H. Islam et al., "Spectrum Survey in Singapore: Occupancy Measurements and Analyses," in IEEE CrownCom, 2008.
[6] Vienna, VA, USA, Tech. Rep. 20100323, "Shared spectrum company general survey of radio frequency bands (30 Hz to 3 GHz)," 2009.
[7] J. Mitola, "Cognitive radio for flexible mobile multimedia communications," Mobile Networks and Appl., vol. 6, no. 5, pp. 435-441, 2001.
[8] X. Xing, T. Jing, Y. Huo, H. Li, and X. Cheng, "Channel quality prediction based on Bayesian inference in cognitive radio networks," in IEEE INFOCOM, 2013.
[9] G. Ding, Y. Jiao, J. Wang, Y. Zou, Q. Wu, Y.-D. Yao, and L. Hanzo, "Spectrum Inference in Cognitive Radio Networks: Algorithms and Applications," IEEE Communications Surveys and Tut., vol. 20, no. 1, pp. 150-182, 2018.
[10] L. Yu, J. Chen, and G. Ding, "Spectrum prediction via long short term memory," in IEEE ICCC, 2017.
[11] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.
[12] L. Yu, Q. Wang, Y. Guo, and P. Li, "Spectrum availability prediction in cognitive aerospace communications: A deep learning perspective," in CCAA, 2017.
[13] A. Orange, "Calibration-data-binder." https://gitlab.flux.utah.edu/alex_orange/.
[14] L. Yu et al., "Spectrum Prediction Based on Taguchi Method in Deep Learning With Long Short-Term Memory," IEEE Access, vol. 6, pp. 45923-45933, 2018.
[15] "Powder wireless research," University of Utah, https://powderwireless.net/.
[16] I. Akyildiz, et al., "A survey on spectrum management in cognitive radio networks," IEEE Commun. Mag., vol. 46, no. 4, pp. 40-48, 2008.
[17] H. Nyquist, "Thermal Agitation of Electric Charge in Conductors," Physical Rev. D, vol. 32, no. 1, pp. 110-113, 1928.
[18] J. B. Johnson, "Thermal Agitation of Electricity in Conductors," Physical Rev., vol. 32, no. 1, pp. 97-109, 1928.
[19] H. Urkowitz, "Energy detection of unknown deterministic signals," Proc. of the IEEE, vol. 55, no. 4, pp. 523-531, 1967.
[20] X. Xing et al., "Optimal Spectrum Sensing Interval in Cognitive Radio Networks," IEEE TPDS, vol. 25, no. 9, pp. 2408-2417, 2014.
[21] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in IEEE ICASSP, 2013.
[22] M. Hermans and B. Schrauwen, "Training and Analysing Deep Recurrent Neural Networks," in Advances in Neural Inf. Process. Syst. 26, 2013.
[23] I. Sutskever, O. Vinyals and Q. V. Le, "Sequence to Sequence Learning with Neural Networks," in Advances in Neural Inf. Process. Syst. 27, 2014.