ABSTRACT

SHORT TERM TEMPERATURE FORECASTING USING LSTMS, AND CNN

by
Darshan Shah

Weather forecasting is a vital application in present times. We can use the predictions to minimize the weather related loss. Use of machine learning and deep learning algorithms for forecasting, can eliminate or reduce the necessity of big data and high computation dependent process of parameterization. Long Short-Term Memory (LSTM) is a widely used deep learning architecture for time series forecasting. In this paper, we aim to predict one day ahead average temperature using a 2-layer neural network consisting of one layer of LSTM and one layer of 1D convolution. The input is pre-processed using a smoothing technique and output is raw (un-smooth) next day average temperature. The smoothing technique improves the performance of LSTM substantially and meanwhile 1D convolution helps unsmooth the output of LSTM to obtain the raw answers. All the models are for particular locations only. The study shows significant improvement in the forecasting with use of smoothing technique. Our method outperforms other model in terms of MSE and MAE.

SHORT TERM TEMPERATURE FORECASTING USING LSTMS, AND CNN

by
Darshan Shah

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Masters in Data Science

Department of Computer Science

May 2021

APPROVAL PAGE

SHORT TERM TEMPERATURE FORECASTING USING LSTMS, AND CNN

Darshan Shah

Dr. Chase Wu, Dissertation Advisor Date
Professor, Computer Science, NJIT

Dr. Jason T. L. Wang, Committee Member Date
Professor, Computer Science, NJIT

Dr. Senjuti Basu Roy, Committee Member Date
Assistant Professor, Computer Science, NJIT

BIOGRAPHICAL SKETCH

Author: Darshan Shah

Degree: Master of Science

Date: May 2021

Graduate and Undergraduate Education:

- Master of Science in Data Science,
New Jersey Institute of Technology, Newark, NJ, 2021

- Bachelor of Engineering in Information and Technology Engineering,
L.D. College of Engineering, Gujarat Technical University, India, 2018

Major: Data Science

DEDICATION

This paper is gracefully dedicated to me, who have worked honestly and faithfully, despite the unforgiving challenges presented. I dedicate this thesis to myself for finding the question and seeking the answers. Seeking an answer takes patience, persistence, hard work and many more challenges; it is a very difficult and tiresome part. Meanwhile, finding a question is impossible without inspiration of curiosity and eagerness for improvisation. Finally, I thank myself for maintaining the core quality of willingness to try, observe and try again, and again.

ACKNOWLEDGMENT

I, as a research student wish to express deepest and genuine gratitude towards my mentor and a well-known professor, Dr. Chase Wu for never undermining even the smallest ideas of mine, for always keeping the open end communication and helping without constraints of time. I am inspired by his simplicity about his achievements and sincerity towards mentoring throughout the research period. I would also like to thank to my committee members Dr. Senjuti Basu Roy and Dr. Jason Wang for giving me honest feedbacks.

I would also like to convey the feeling of thankfulness for my senior colleagues (especially Wuji), for advocating with technical issues I faced.

And finally, a huge thanks to my friends and family, for supporting me during difficult times.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

MSE	Mean Squared Error
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
NMSE	Normalized Mean Squared Error
LSTM	Long-Short Term Memory
FFBP	Feed-Forward Back Propagation
RBF	Radial Basis Function
GRNN	Generalized Regression Neural Network
RNN	Recurrent Neural Network
RCNN	Recurrent Convolutional Neural Network
JSPN	Jordan Pi-Sigma Neural Network
MLR	Multiple Linear Regression
ANN	Artificial Neural Network
MIMO	Multiple-Input Multiple-Output
MISO	Multiple-Input Single-Output
$\cong$	approximately equals to
Σ	summation
$\hat{y}$	predicted value

CHAPTER 1

INTRODUCTION

Climate change is a crucial challenge in the recent era. It affects many factors of the environmental ecosystems such as, soil erosion [20], bio-diversity, and changes in sea-water level [19]. Weather forecasting mitigates the economic crisis and promotes better public health [21] to maintain the quality of life. Safety and well-being of human is highly impactable by weather changes. It is also useful in the agricultural domain as it is an essential part of planning the farming operations. Farmers can make optimal decisions for crops using prediction of weather [22] whether to undertake or withhold the sowing operation. The consequences of unseasonal changes in weather and their potential negative effects on host plants and pests are very well known. Unseasonably high temperatures may lead to lower plant productivity and more pests on farm. Industries such as energy consumption and food security can also benefit from weather forecasting.

As the key problem of weather forecasting, air temperature prediction has manifold benefits for the environment, industry and agriculture. The impact of temperature on morbidity and mortality can be assessed at both the seasonal and daily level. Extreme temperature changes Due to harsh environment, arises the lack of access to safe water and food, it can also cause Heat-aggravated and respiratory illness. Prediction of the energy consumption, soil surface temperature and solar-radiation is related to ambient air temperature forecasting. Air temperature forecasting is useful in understanding the probability of storm, wildfires, drought and flood occurrence in an area.

Temperature prediction is an infamously sophisticated and resource consuming task. Temperature changes are caused by many factors. Parameterization [2] of those features is a difficult task to achieve due to their dynamic nature. Recent development in the field of artificial intelligence can help provide less computationally expensive solutions. We can approximate the forecasts using several black box methods without a need of extensive mathematical calculations by analyzing historical temperature data. Deep learning algorithms have been widely used for complicated data. Pattern analysis and recognition of temperature data can be simplified with use of deep learning algorithms. Air temperature data is classified as part of time series statistics up. Hence, use of Recurrent Neural Network (RNN) algorithms to estimate the future value of temperature seems a plausible solution.

CHAPTER 2

RELATED WORK

In our experiment, we came across various machine and deep learning methodologies. An early paper of this domain by B. Ustaoglu [9], tests three different kind of ANN based methods: (1) feed-forward back propagation (FFBP), (2) radial basis function (RBF) and, (3) generalized regression neural network (GRNN). It compares the answers with traditionally used multiple linear regression (MLR), they obtained notable improvements over MLR outputs. A paper written in 2015 by Z. Karevan [10] describes a black box idea: use of Machine learning methods such as k-NN and Elastic Net for the process of feature selection then trains model using Support Vector Machine Regression with Least Square loss function to predict minimum and maximum temperature. After 3 years, R. Isabelle used Recurrent Convolutional Neural Networks for weather forecasting and visualization [11] where they propose use of convolution filters + LSTMs. Their results were found substantially better in comparison with popular methods.

Another approach was used by P. Hewage [4], where they used multiple features like temperature, pressure, wind, humidity, precipitation and moisture to predict future value of the same feature. This was executed by implementing several machine learning and deep learning algorithms like TCN [13], LSTM with multi-input multi-output and multi-input single-output methods. S. Kendzierski used a novel approach by implementing Jordan Pi-Sigma Neural Network (JPSN) for time series data, introduced by N. Husaini [14]. In this paper they combined two methodology: Jordan Neural Network, Pi-Sigma Neural Network to predict the temperature. The MSE of the model is remarkably low, but model does not satisfy the criteria suggested by A. Kumar [3]—the performance of the model can be acceptable if: $NMSE \leq 0.5$.

Detail-metrics of work described in this section is presented in table 2.1. **Region of interest varies with different journals.

Table 2.1 Detail-Metrics Of Related Work

References	Input variables	Algorithm	Error Metrics
[9]	temperature	ANN	*RMSE = 1*
[10]	wind, snow, rain, fog, MIN, AVG and MAX temperature, wind speed, humidity, sea level pressure	k-NN + Elastic Net + LS-SVM	*MAE = 1.15*
[11]	temperature, pressure, wind directions (2D)	CNN+LSTM (RCNN)	*MAE = 0.88*
[4]	temperature, pressure, wind, humidity, precipitation and moisture	TCN, MIMO-LSTM, MISO-LSTM	*MSE = 3.4*
[15]	temperature	JPSN	*MSE = 0.006462, NMSE = 0.7710*

CHAPTER 3

PRELIMINIRIES

Accurate prediction of temperature requires knowledge of various parameters: longitude, latitude, sea level, pressure, wind, precipitation; and their internal correlations. Ecosystem of some parameters complete within small areas and for some-other, it occupies large areas. As consequences, it becomes difficult to define a region to accurately estimate those parameters. Due to this difficulty, parameterization becomes a vital part of the process. Parameterization is a complex and instantiate process. In the sense of computational liability, it is very resource consuming procedure. In different regions, temperature patterns may vary. However, it is generally repetitive with respect to time. Hence, with help of recent developments in the field of machine learning and deep learning, we can eliminate use of parameterization. In this thesis, we have used black box methods described in subsection 3.1 and 3.2 to forecast temperature using past temperature data points. Our aim is to create an accurate procedure which analyzes the patterns of past temperature data to predict future results.

3.1 Long Short-Term Memory (LSTM)

LSTM is a variant of Recurrent Neural Networks. RNNs are special types of Neural Networks with recurring properties. It takes current input example, and what they have perceived previously in time as their input as well.

LSTM adds three gate functions on the basis of the RNN: input, forgetting, and output gates which are used to control the input, memory, and output values, respectively.

$$z_t = \sigma(W_z.[h_{t-1}, x_t]) \tag{3.1}$$

$$r_t = \sigma(W_r.[h_{t-1}, x_t]) \tag{3.2}$$

$$\tilde{h}_t = tanh(W_r.[r_t * h_{t-1}, x_t]) \tag{3.3}$$

$$h_t = (1 - z_t) * (h_{t-1}) + z_t * \tilde{h}_t) \tag{3.4}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{3.5}$$

In the equations described above, z_t is the input; h_t is the output at the time t. W_z W_r and W are weights of input gate, forgetting gate and output gate respectively.

3.2 1D Convolutions

Concept of convolution is adapted from digital signal processing [17]. Convolution is used for many purposes like smoothing, Image and pattern recognition. In deep learning, one dimensional convolution is helpful in time series analysis.

$$(f * g)(i) = \sum_{j=1}^{m} f(i) \cdot f(i - j + \frac{m}{2}) \tag{3.6}$$

In this thesis, we have used 1D convolution for smoothing the data as well as for adjusting back variations into the smoothed data.

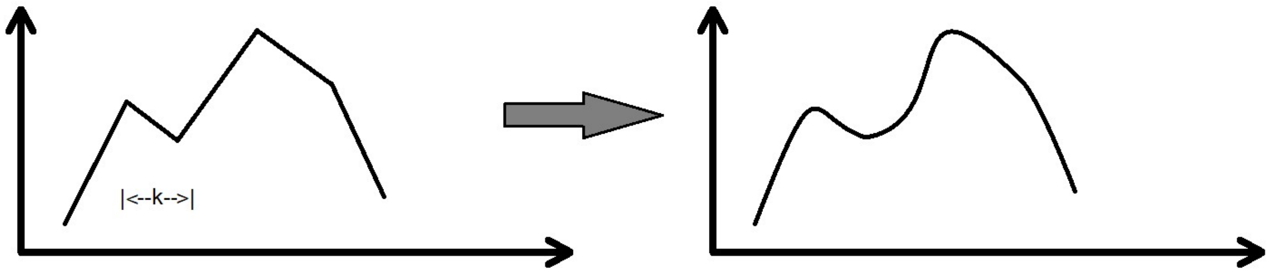

Figure 3.1 Shows how input gets smoothen by 1D convolution with *filter size = k*.

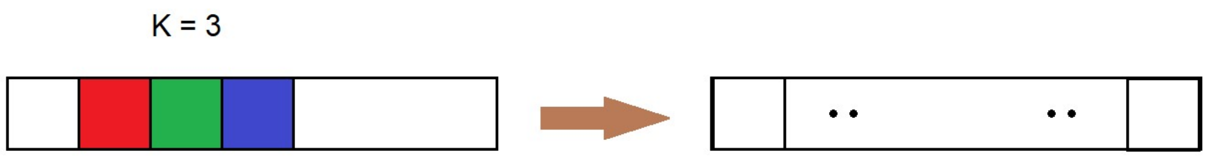

Figure 3.2 Shows 1D convolution across the signals having filter size *K = 3*.

CHAPTER 4

METHODOLOGY

4.1 Preprocessing: Smoothing

In this part, we used a smoothing technique to prepare data. Smoothing helps prevent over fitting. The local extreme differences reduce by use of this technique which consequently reduces variance in the data. It also handles outliers and generalizes them significantly .This causes increase in the performance of the training model.

In our experiment, we applied convolution smoother on the whole data set with *window size = 2*. Use of minimum necessary window size removes the harmful differences and cleans the data set for generalization. Smoothing effects on sample of *100* days temperature data is shown in the figure 4.1.

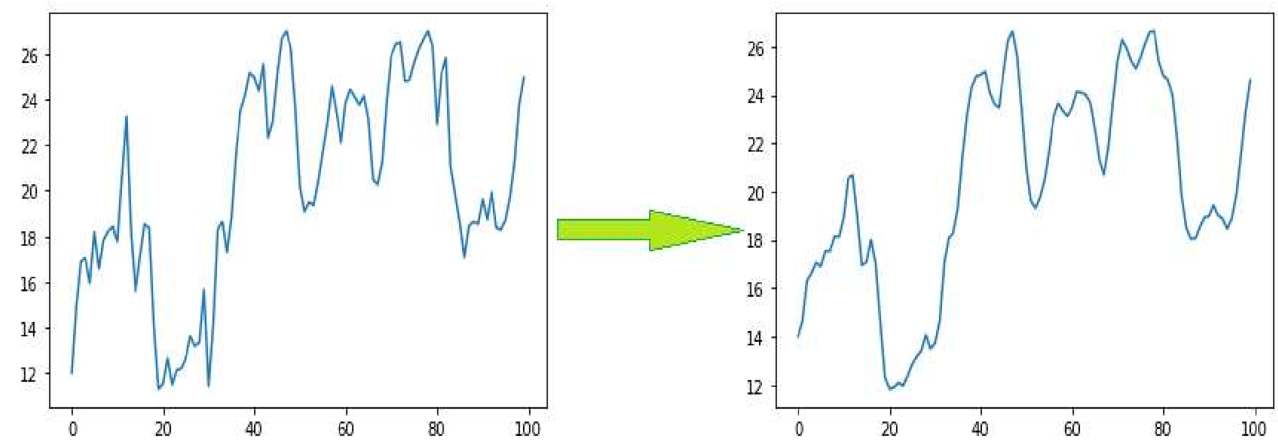

Figure 4.1 Smoothing the data of *100* days, using convolutional smoother with *window size = 2*.

4.2 Architecture

In this experiment, we used smoothed data as the dependent variable and the not smoothed data as target variable. There are two purposes of using unprocessed data as the output. (1) If unsmooth data is output of the model, then there is no need for having an extra process to get the desired answer. (2) Un-smoothing the variable would lead to irrelevant variance and increase the error rate.

After the smoothing step, we split both—smoothed and unprocessed dataset—into *m:n* ratio. We stored values of multiple past days of smoothed data as input variables and we kept unsmoothed data as output variable. Following that, we merged input and output variables for both train and test data. As the preprocessing completed, we trained data on the specific neural network architecture. In this experiment, neural network architecture was composed of LSTMs, 1D convolutions and a single perceptron.

We analyzed significance the error rate of predictions using the performance evaluation matrices, during testing. The whole process of the algorithm is shown in the figure4.2.

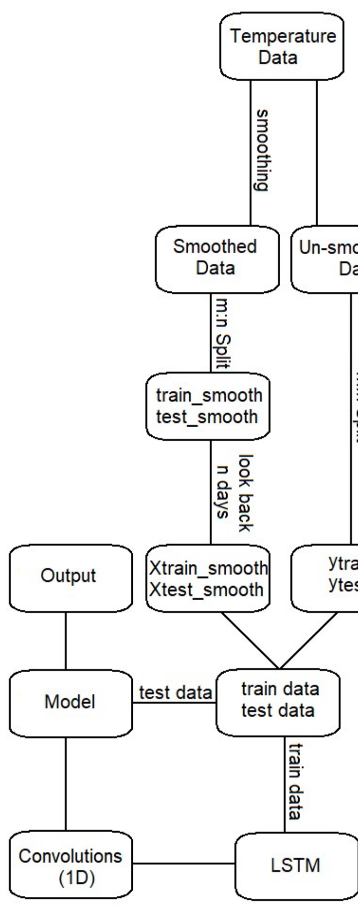

Figure 4.2 Architecture of our proposed method.

CHAPTER 5

EXPERIMENTAL RESULTS

We ran our experiment on Windows operating system having version 10. Processes were executed by Intel i5-9300H CPU, running at 2.4 GHz and 8 GB of RAM, and NVIDIA GeForce GTX 1660 Ti with Max-Q Design GPU, running at 1.34 GHz using 6 GB of RAM.

5.1 Performance Matrix

In this part of the study, we applied model evaluation based on the MSE, MAE error matrices. Since, the values of temperatures are in Celsius, the interpretation of MSE is the average squared difference between predictions and real values in terms of Celsius. MSE is defined by the following formula.

$$MSE = \frac{1}{n}\sum_{1}^{n}(\hat{y} - y)^2 \tag{5.1}$$

The interpretation of MAE is the average absolute difference between predictions and real values in terms of Celsius.

$$MAE = \frac{1}{n}\sum_{1}^{n}||\hat{y} - y|| \tag{5.2}$$

NMSE is used to avoid overestimate or underestimates values caused by bias in models. NMSE is defined by the following formula.

$$NMSE = \frac{1}{n}\sum_{1}^{n}\frac{(\hat{y} - y)^2}{y} \tag{5.3}$$

We also used, Pearson Correlation and R^2 score values to check the reliability of the model. Both matrices are described as below.

$$Correlation = \frac{\Sigma(\hat{y} - \bar{\hat{y}})(y - \bar{y})}{\sqrt{\Sigma(\hat{y} - \bar{\hat{y}})^2 + \Sigma(y - \bar{y})^2}} \tag{5.4}$$

$$R^2 = \frac{\sum(y - \hat{y})^2}{\sum(y - \bar{y})^2} \tag{5.5}$$

5.2 Data Preparation

In this thesis, we used Basel, Switzerland's average temperature of January 1989 to March 2012 as our dataset. Ratio was set *to traindata:testdata = 20:1*. The training data was used to build the model and the test set was used to evaluate the model.

The average temperature of the day is predicted using previous *n* days' average temperature. During our analysis, we found that decision process of the value of *n* is a greedy algorithm problem as there is a trade-off to be considered—increase in days, decreases the error and increases the resource engagement at same time.

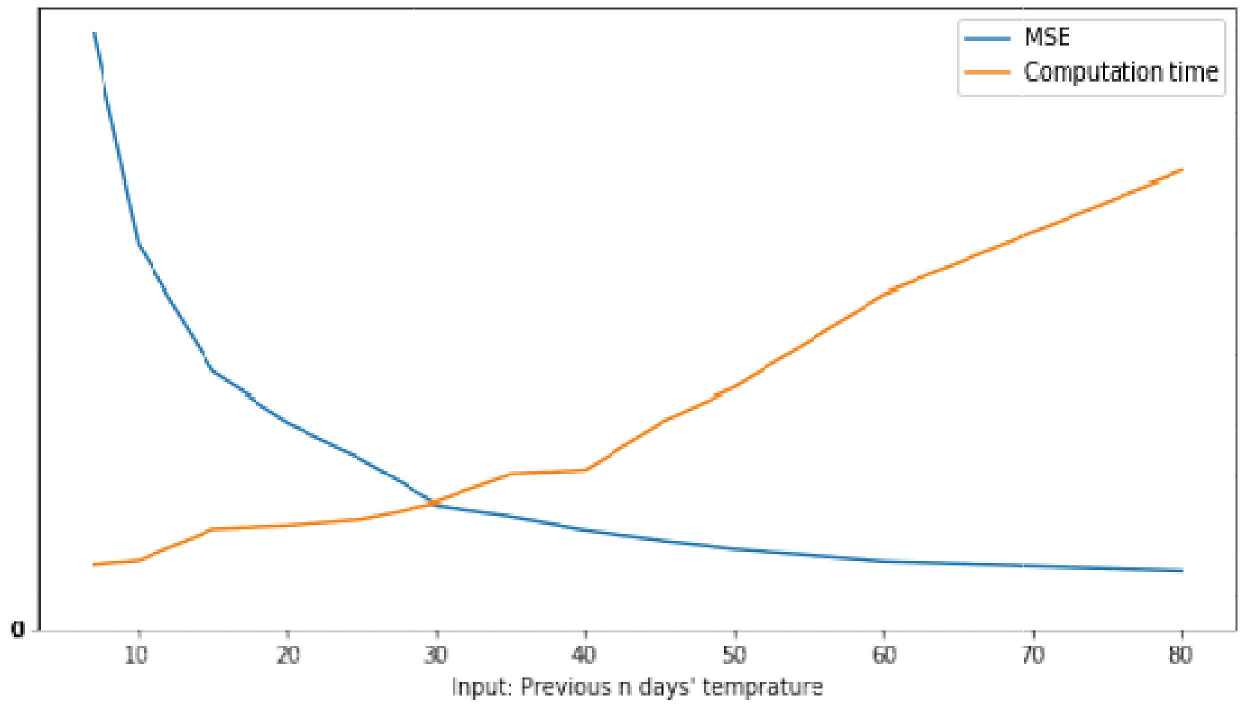

Figure 5.1 Visualization of how, different values of *n* affect error and training time.

Going by the steps of optimal trade-off resolution method, we found the minimum value of $n = 30$, to predict next day average temperature.

5.3 Training & Performance

The proposed model took temperature data of previous *30* days as input and passed through the architecture after hyper parameter tuning. In the model, *50* neurons of LSTM layer took input data and provided output for the 1D convolution layer. There were *50* kernels of 1D convolutions with *kernel size = 7* to train the data further. Output of this layer was then flattened. Finally, after flattening, it goes through single perceptron to predict the next day average temperature. Model's learning rate was set to *0.001*. Ridge regularizer with *importance = 0.01* was used to regularize LSTM layer. Furthermore, early stopping conditions were applied to avoid over fitting. This model was assigned *1000* epochs to run the training data.

For the comparison, we created two models: one model was trained on smoothed input and other one was trained on unsmooth input. Both models had similar training duration which was approximately *20* seconds and both models had $NMSE \cong 0$ which suggested the models were acceptable and significant. However, from the perspective of accuracy, the model created with unsmooth inputs, gave high error rates with $MSE = 3.4$ and $MAE = 1.44$. Meanwhile, performance of the proposed model which was train on smoothed inputs provided substantially low errors. The MSE of proposed model was about *0.064* and $MAE = 0.202$. Proposed method seems to give better performance of accuracy on test data, as shown in the figures. Our model outperformed other models except JPSN. JPSN had $MSE = 0.006462$. *NMSE* of JPSN model is *0.771*. Hence, it fails *NMSE* test which requires $NMSE < 0.5$ for model to be acceptable. Performance comparisons of different methods are shown in the figure 5.4, 5.5, and 5.6. The R^2 value of proposed model was found 0.9984 and $correlation = 0.9991$. Same experiment was done using ARIMA method where we found high error rates with $MSE = 3.52$, $MAE = 1.47$, $NMSE \cong 0$, $correlation = 0.9629$, and $R^2 = 0.9217$. Comparing performance of proposed model with ARIMA performance, our method outperforms ARIMA model.

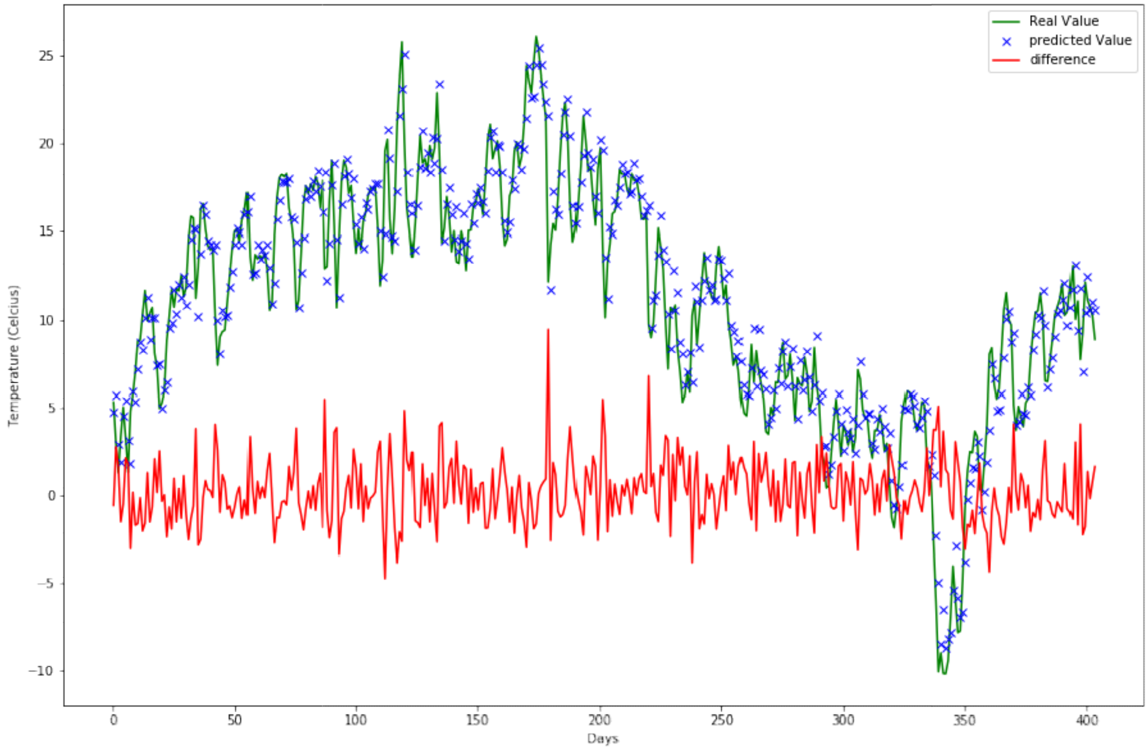

Figure 5.2 Comparison of actual values and predictions using unsmooth input.

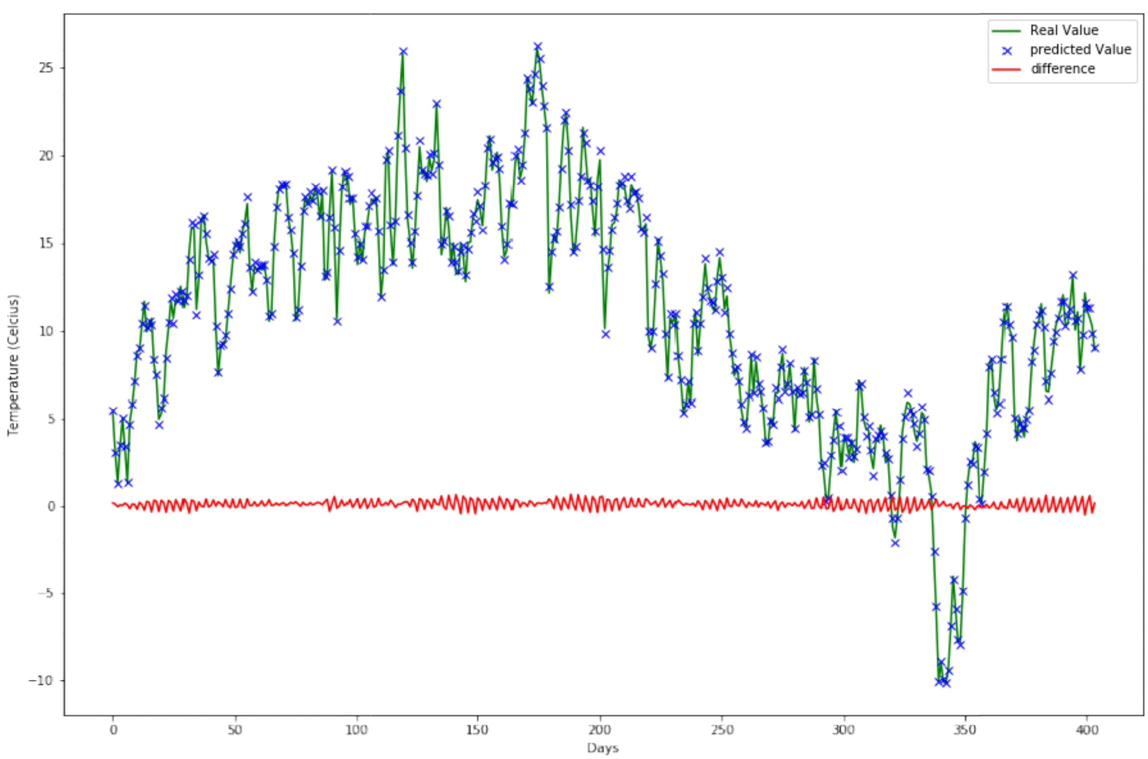

Figure 5.3 Comparison of actual values and predictions using smooth input.

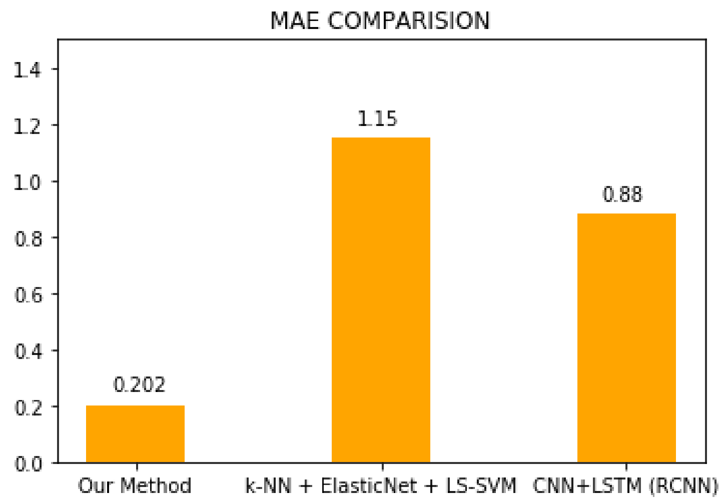

Figure 5.4 Comparison of actual values mean absolute error.

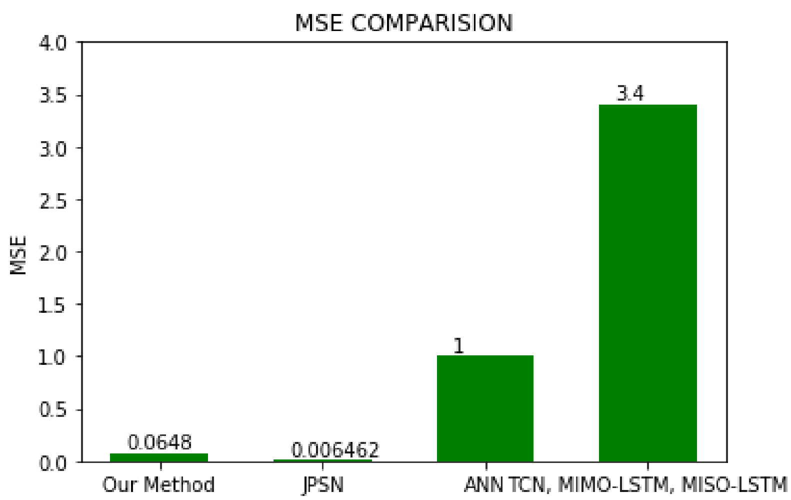

Figure 5.5 Comparison of actual values mean squared error.

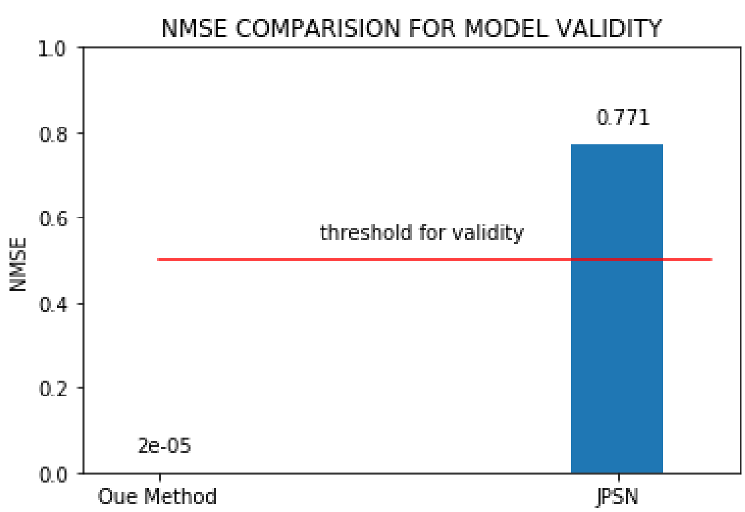

Figure 5.6 Comparison of actual values normalized mean squared error.

CHAPTER 6

CONCLUSIONS

In this thesis, we proposed a model for temperature forecasting. The objective of this study was to forecast the daily mean temperature using three LSTM, 1D convolution and smoothing technique. Model showed acceptable and significantly high performance in terms of normalized mean squared error, mean squared error and mean absolute error on predicting average next day temperature of Basil region. Model used the previous n day's temperature where the optimal trade-off resolution method selected the value of n. Smoothing dataset played an important role in performance of the model.

REFERENCES

[1] S. HOCHREITER, AND J. SCHMIDHUBER, Long short-term memory. Neural Computation 9,8 (1997), 1735–1780.

[2] M. Norman. (2011). Parameterizations: Representing key processes in climate models without resolving them. Wiley Interdisciplinary Reviews: Climate Change. 2. 10.1002/wcc.122.

[3] A. Kumar, J. Luo and G. Bennett, "Statistical Evaluation of Lower Flammability Distance (LFD) using Four Hazardous Release Models", Process Safety Progress, 12(1), pp. 1-11, 1993.

[4] P. Hewage, M. Trovati, E. Pereira, et al. Deep learning-based effective fine-grained weather forecasting model. Pattern Anal Applic 24, 343–366 (2021).

[5] B. Mohammadi, S. Mehdizadeh, F. Ahmadi, et al. Developing hybrid time series and artificial intelligence models for estimating air temperatures. Stoch Environ Res Risk Assess (2020).

[6] J Cifuentes, G Marulanda, A Bello, J Reneses. Air Temperature Forecasting Using Machine Learning Techniques: A Review. Energies. 2020; 13(16):4215.

[7] K.H., Shin, et al. "Dynamical Prediction of Two Meteorological Factors Using the Deep Neural Network and the Long Short Term Memory \$(1)\$."

[8] P. Hewage, A. Behera, M. Trovati, et al. Temporal convolutional neural (TCN) network for effective weather forecasting using time-series data from the local weather station. Soft Comput 24, 16453–16482 (2020).

[9] B. Ustaoglu,H.K. Cigizoglu, and M. Karaca, (2008), Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods. Met. Apps, 15: 431-445.

[10] Z. Karevan, S. Mehrkanoon and J. A. K. Suykens, "Black-box modeling for temperature prediction in weather forecasting," 2015 International Joint Conference on Neural Networks (IJCNN), Killarney, Ireland, 2015, pp. 1-8, doi: 10.1109/IJCNN.2015.7280671.

[11] R. Isabelle and T. Günther. 2019. "Visualization of Neural Network Predictions for Weather Forecasting." Computer Graphics Forum 38 (1): 209–20. doi:10.1111/cgf.13453.
[12] Keras Metrics—Keras documentation [WWW Document]. [Cross Reference]

[13] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager. Temporal convolutional networks for action segmentation and detection. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2017.

[14] N. A. Husaini, R. Ghazali, N. M. Nawi and L. H. Ismail, "Jordan pi-sigma neural network for temperature prediction" in Ubiquitous Computing and Multimedia Applications, Springer Berlin Heidelberg, pp. 547-558, 2011.

[15] Understanding LSTM Networks. [Cross Reference]

[16] S. Stephen W (1997). "13.Convolution". The Scientist and Engineer's Guide to Digital Signal Processing (1 ed.). California Technical Publishing. ISBN 0-9660176-3-3.

[17] machine learning - Intuitive understanding of 1D, 2D, and 3D convolutions in convolutional neural networks - Stack Overflow. [Cross Reference]

[18] D. Marutho, S. Hendra Handaka, E. Wijaya and Muljono, "The Determination of Cluster Number at k-Mean Using Elbow Method and Purity Evaluation on Headline News," 2018 International Seminar on Application for Technology of Information and Communication, 2018, pp. 533-538, doi: 10.1109/ISEMANTIC.2018.8549751.

[19] Grimm, N., Groffman, P., Staudinger, M., and Tallis, H. (2015). Climate change impacts on ecosystems and ecosystem services in the United States: process and prospects for sustained assessment. Climatic Change, 135(1), 97-109.

[20] Y. Wu, W. Ouyang, Z. Hao, C. Lin, H. Liu, Y. Wang, Assessment of soil erosion characteristics in response to temperature and precipitation in a freeze-thaw watershed, Geoderma, Volume 328, 2018, Pages 56-65, ISSN 0016-7061.

[21] Lazo, J. K., Hosterman, H. R., Sprague-Hilderbrand, J. M., and Adkins, J. E. (2020). Impact-Based Decision Support Services and the Socioeconomic Impacts of Winter Storms, Bulletin of the American Meteorological Society, 101(5), E626-E639.

[22] Hammer G.L., J.W. Hansen, J.G. Phillips, J.W. Mjelde, H. Hill, H.A. Love, and A. Potgieter. 2001. Advances in Application of Climate Prediction in Agriculture. Agricultural Systems. 70(2-3)(November):515-553.