Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 155 (2019) 394–401

Procedia Computer Science

www.elsevier.com/locate/procedia

The 14th International Conference on Future Networks and Communications (FNC)
August 19-21, 2019, Halifax, Canada

Multi Features and Multi-time steps LSTM Based Methodology for Bike Sharing Availability Prediction

Xu Liu[a,*], Abdelouahed Gherbi[a], Wubin Li[b], Mohamed Cheriet[a]

[a] *Synchromedia Laboratory, École de Technologie Supérieure (ÉTS), University of Québec, Canada*
xu.liu.1@ens.etsmtl.ca, abdelouahed.gherbi@etsmtl.ca, Mohamed.Cheriet@etsmtl.ca
[b] *Ericsson Research, Ericsson, Montréal, Canada*
wubin.li@ericsson.com

Abstract

Most cities in the world promote bike-sharing services to encourage people to decrease carbon exhausting and to enhance their health. However, it is a big challenge for a bike-sharing service supplying corporation to re-balance bikes efficiently among different bike-sharing dockers without a forecasting ability. For solving this problem, we contribute two new approaches based on standard Long short-term memory (LSTM), which can not only take advantages of multi features inputs and multi-time steps outputs to improve the accuracy of predicting available bikes in one-time step, but also can forecast the number of bikes in multi-time steps. These approaches will help the bike-sharing agencies to make a better decision to distribute their bikes to each docker efficiently. The experimental results confirmed that our multi-feature and multi-time steps models outperform the standard LSTM model.

Keywords: Bike-sharing; LSTM; RNN; DNN; Neural Network; Prediction;

1. Introduction

Although there are many problems for a metropolis, traffic is always seated in the top five. As the cities developing, the population and personal cars increase much, the towns unavoidably become more and more crowded, and the traffic quickly turns into a jam. For easing this knotty problem, many governments encourage people to use public transportation, as this can dramatically decrease the usage of personal cars, which are the main reason causing a traffic jam. However, most public transportation stations are not near people's workplaces or home, and people still need to walk several hundred meters or even miles to their destinations, which hinders people from choosing public transit. It is also called the first/last mile transportation problem (Paul, 2009, [4]). Bicycle is cheap, small, convenient, flexible,

* Xu Liu. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000.
E-mail address: xu.liu.1@ens.etsmtl.ca

10.1016/j.procs.2019.08.055

easy to manipulate and environmentally friendly, which is a reasonable solution for the previous problem (Susan et al., 2010, [13]). However, before the Internet of Thing(IoT) emerging, the bike-sharing is not so popular, as it is hard to manage the bikes without the information from all kinds of sensors at that time. Now, as the IoT booming, it is easy to collect real-time data from bikes which can be used well for managing bikes (Razzaque et al., 2015, [12]).

Most of bike sharing systems contain a docker which is in charge of lending and retaking shared bikes. It is essential to keep every docker has a proper number of bicycles during a different period of the day. Because if it is full, people can not return their bikes, on the contrary, if it is empty, no bike can be borrowed. So almost every bike sharing company has its trucks to rebalance bikes among different dockers. However, it is often too late when the bike sharing agencies found some dockers are full or empty, as people have to wait or go to next docker to return or rent their bikes, which will limit the popular of bike-sharing. It would be beneficial if the agencies can forecast the number of bikes in each docker in some time earlier. However, it is not easy to build a model for forecasting by the traditional mathematical way as there are many factors such as weather, locations, and seasons, which will all have impacts on the prediction. Thanks to machine learning (ML)'s development, even we do not know much about the complex math relations of the different factors, we can build a forecasting model from the data (Bo Wang et al., 2018, [16]).

This paper proposed two solutions based on standard LSTM algorithm to build a model to predict bike availability of a bike sharing docker. Moreover, the rest of this article is organized as follows: Section 2 presents related work about bike sharing availability predicted by MLs. Section 3 elaborates the details of our methods of bike sharing availability prediction based on LSTM. Section 4 shows and discusses the results of the experiment of our methodologies. Section 5 gives a detail about the limitations of our methodologies. In Section 6, we conclude our work and briefly discuss our future research plan.

2. Related work

Between 2012 and 2015, few related works use ML methodologies such as Junbin Xu et al. 's work [18], which only used statistical analysis to build their model to forecast the bicycle parking demand and Longbiao Chen et al. 's work [3], which proposed a dynamic cluster-based framework for over-demand prediction.

After 2015, most related works started to adopt MLs to predict the available bikes. In 2015, Divya Singhvi et al. used linear regression to predict the bike demands for New York City [14]. Moreover, in the same year, Oriol Cosp Arqué built a bike demand forecast model with random forest(RF) [1]. In 2016, Wen Wang forecasted bike rental demand based on RF too [17]. Similar works also have been done by Huthaifa I. Ashqar et al., 2017, [2], YouLi Feng et al., 2017, [5], Lei Lin et al., 2018, [7], Terry Liu et al., 2018, [8], etc. However, few works referred to Recurrent Neural Networks (RNNs) or LSTMs. We found one paper, Bo Wang et al., 2018, [16] has done some bike-sharing predicting work based on RNNs. However, their work can only predict one-time step, did not conduct multi-time steps prediction and did not consider other factors such as day of the week and time of the day.

Compared with previous related works, our main contributions are two, the first is to use multi-features and multi-time steps to train the model to improve the accuracy of one-time step prediction, the second is to take advantages of previous one-time step prediction result to generate next steps'input data to forecast the available bikes in multiple time steps later.

3. Methodology

During this section, we will quickly recap the theory of DNN, RNN, and LSTM, explain the reason why we selected LSTM as our primary solution and give out our detail solutions.

3.1. General background

3.1.1. DNN vs RNN

According to the general concept, DNN contains all kinds of deep neural networks, including convolutional neural network(CNN), RNN, and Reinforcement Learning (RL), and so on. However, within a specific definition, we only use DNN to stand for multiple layer perceptron (MLP) whose architecture is similar to Fig. 1(a), which only has static full connected layers, activation layers, and have no matter with time. They have simple structures and are easy to

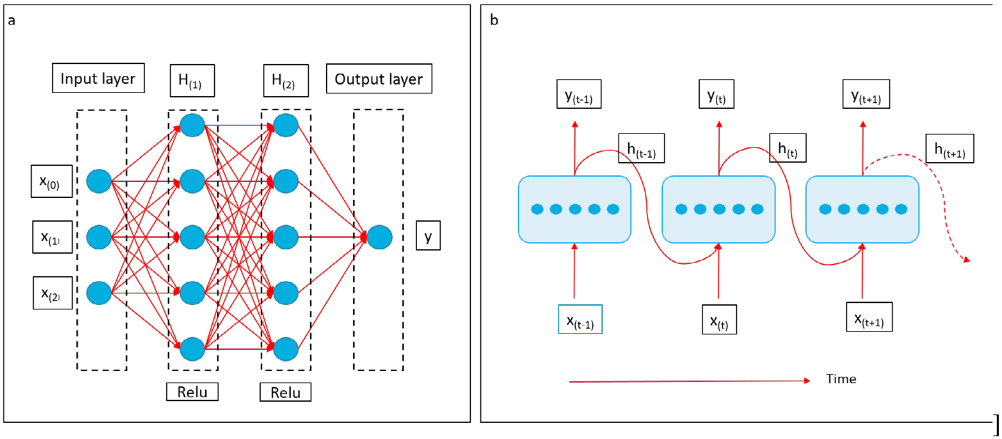

Fig. 1. DNN vs RNN.

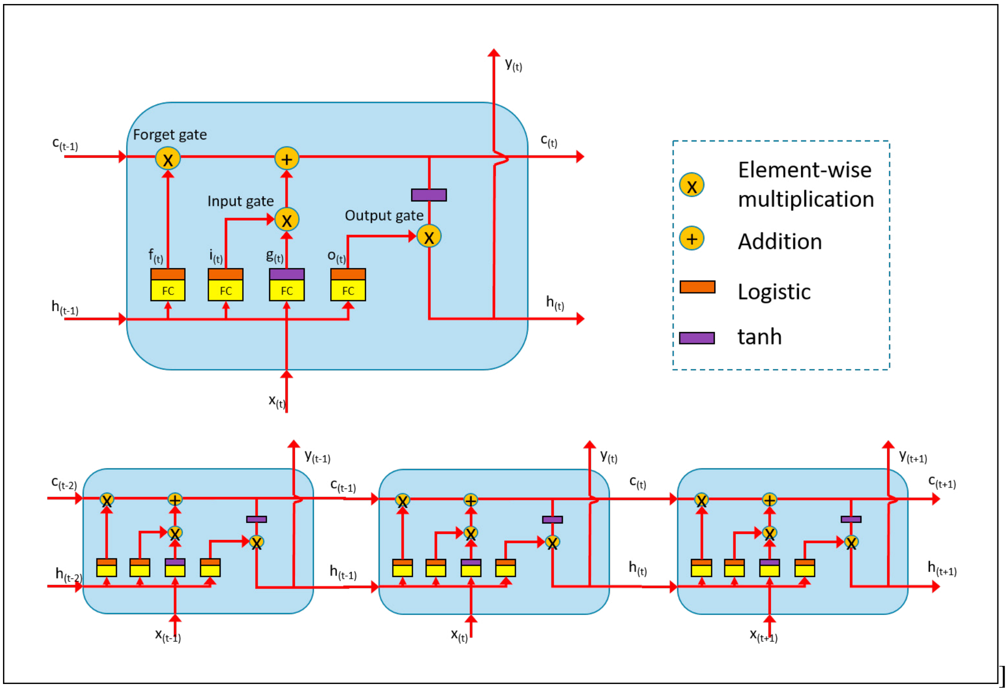

Fig. 2. The architecutre of LSTM.

train. However, their structures are fixed, and it is hard to learn features from previous and later time steps as it does not share the features learned across the different time steps. [11].

Developed from MLP, RNN not only has normal MLP's classical architecture but also can make a prediction based on the previous time step results. These advantage characteristics are shown in Fig. 1(b).

3.1.2. LSTM vs BiLSTM

Normal RNN has a shortcoming, it tends not to be very good at capturing long-range dependencies when the number of RNN's cell becomes more, and the gradient will vanish quickly. For overcoming this disadvantage, people developed the LSTM algorithm. A classical LSTM cell and an LSTM cells chain have been shown in Fig. 2.

From Fig. 2 we can find that each LSTM cell 's state can be split into two vectors, short-term state $h_{(t)}$ and long-term state $c_{(t)}$ and these states are controlled by three gates: forget gate, input gate, and output gate. The forget gate controls which part of last time step long-term state $c_{(t-1)}$ should be dropped in this time step, the input gate decides which part of $g_{(t)}$ should be added to the long-term state and the output gate selects which part of long term state should be output to $h_{(t)}$ and $y_{(t)}$. So at each time step, some memories are kept and some memories are added (Aurélien Géron(2017) [6]). The network learns by itself what to be dropped, what to be added, and what to be output. Based on these advantages, LSTM helps a lot in gradient vanishing problems and allow a neural network to learn even long-range dependencies.

The goal of our case is to predict the available bikes at different time steps, and it is a problem depending on the time series. Each of the previous time steps' information will help to predict the following time steps' results. According to our last analysis, LSTMs are the best choice. Standard LSTMs can solve the problem, but not the best. In the following sections, we will introduce our improved LSTMs.

3.2. Designs and Implementation

3.2.1. Data structure

The data is from a bike-sharing docker located in Suzhou, China. Every minute has a row record, and it has three columns: the number of available bikes per minute, the day of the week and the hour of the day, which looks like table 1. It has one month, 45949 rows (minutes) records totally (Bo Wang, 2017, [15]).

For training the RNNs' model, we should first convert the original data sets into the same length sequences. In our case, the sequence's length is 20. After converting, we have 45930 sequences. We divided these sequences into two parts: 43634 (95%) for training and 2296 (5%) for testing. These selections and configurations are the same as Bo Wang et al.'s work [16], which will be convenient to compare the results.

Table 1. Part of bike-sharing data set.

Number of Available Bikes per Minute	Day of the week	Hour of the day
0	5	17
1	5	17
1	5	17

3.2.2. One time step prediction

Fig. 3 (a) is a benchmark design from Bo Wang et al.'s work [16]. It is a standard LSTM and has two LSTM layers, each layer has 19 time steps, and each LSTM cell contains 100 neutrons and uses the last time step value of the sequence as a training label and uses the previous time steps values as the training data. In each time step, it has only one input, the current step's number of available bikes in a docker and only one output, the 20_{th} step's number of available bikes.

For improving the forecasting accuracy, we supposed two new architectures which have been shown in Fig. 3 (b) and Fig. 3 (c). In Fig. 3 (b), we add two more input features, the day of the week and the hour of the day besides the current step's number of available bikes. Generally, more features will lead to higher accuracy. In Fig. 3 (c), we did not only use the last time (20_{th}) step value as the training label but also use the previous time (18) steps'results in the sequence as training labels which will extract more useful information from the data sequence.

3.2.3. Multiple time steps prediction

In most cases, it is beneficial to know what will be in multiple time steps later, which will help us to make a better decision. To the best of our knowledge, few people conducted this work before. In our paper, we proposed a method as Fig. 4 showing to do the multi-time steps forecast.

This work can be divided into two phases: data generation and prediction. During the data generation phase, we use the time steps (from 0_{th} to $(n-1)_{th}$) except the last one (n_{th}) in a sequence to forecast the end time step (n_{th}) value and then we drop the first step (0_{th}) and put forecast one (n_{th}) at the end of the new sequence to forecast the next time step value.

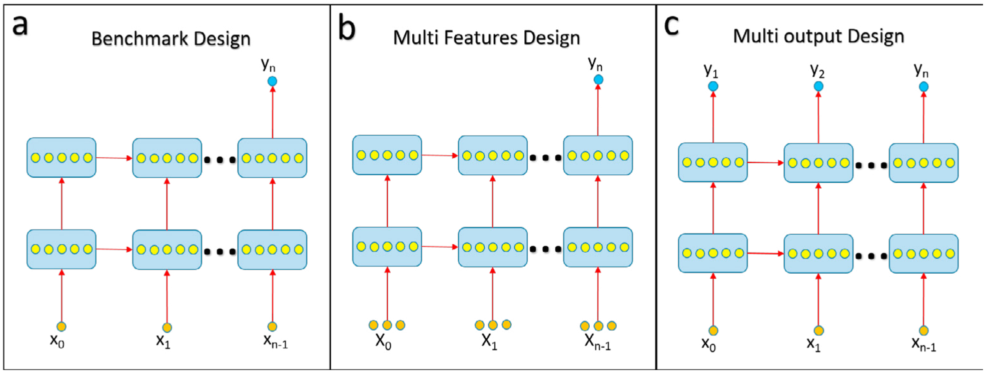

Fig. 3. (a) The design of standard LSTM benchmark; (b) The design of multi-features inputs LSTM; (c) The design of multi-output LSTM.

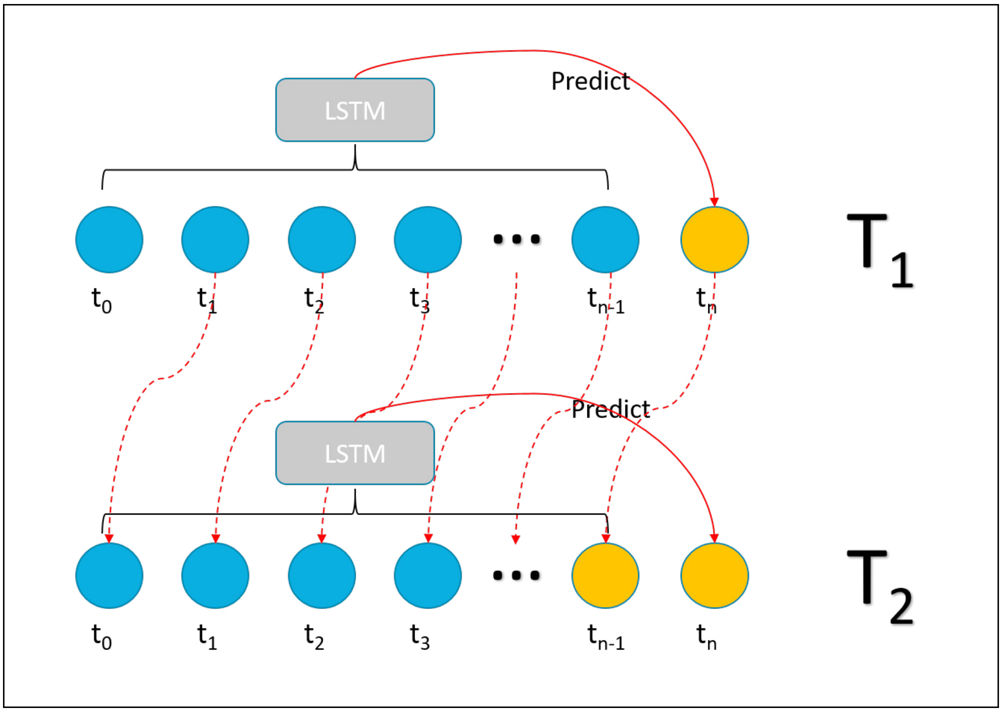

Fig. 4. The methodology of generating forecasting data.

4. Results and discussion

For verifying our approaches, we conducted five experiments and divided these experiments into two parts: one-time step prediction and multi-time steps forecast.

4.1. One time step prediction

During the one-time step prediction, we first implemented a standard LSTM architecture similar to Fig. 3 (a) as a benchmark. And then we create a multi-feature inputs LSTM as Fig. 3 (b) and a multi-time steps outputs LSTM as Fig. 3 (c). After we created all these three different kinds of neural networks, we feed them with the same data sequence which we have created in Section 3.2.1 and got results as Fig. 5 and Table 2 showing. For measuring the

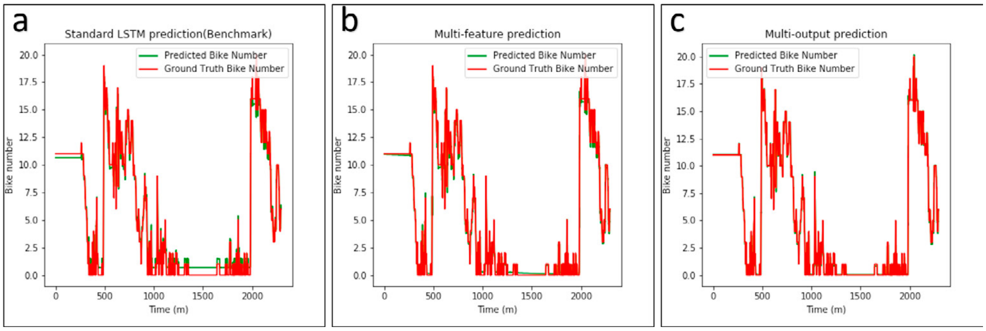

Fig. 5. Comparison of different one-time step prediction.

prediction accuracies of the three kinds of LSTMs, we chose three metrics, Mean Square Error (MSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE).

Table 2. Comparison of one-time step predictions

Metrics	Standard	Multi-feature	Multi-output
MSE	0.696499	0.398346	0.386967
MAE	0.649578	0.364747	0.296937
MAPE	394.8634	325.3479	290.8239

From Fig. 5 and Table 2, we can get a conclusion that the predictional accuracy of the multi-time steps outputs approach and the multi-feature approach are similar and have better prediction accuracy than the standard LSTM which has been shown by Bo Wang in 2017, [15].

4.2. Multiple time steps prediction

During our multi-time steps prediction experiments, we first created two kinds of neural networks, one is with standard LSTM architecture which only outputs the last time step's value, the other is multi-time steps outputs LSTM which outputs all the sequence time steps' values, but we only use the last time step's output value as the next input. Then we feed the two neural networks with the same sequence which we generated according to the methodology described in Section 3.2.3, predicted 20 time steps, and got the results as Fig. 6 and Table. 3.

Table 3. Comparison of multi-time steps predictions

Metrics	Uni-time step output	Multi-time steps outputs
MAE	4.52	1.11
MSE	21.97	2.00

From Fig. 6 and Table. 3, we can get a conclusion that multi-time steps output LSTM is much better than the standard uni-time step output LSTM when making multi-time steps prediction.

5. Limitations

Even though the model of LSTM with three features has already been better than the model with one feature, however, there is still a lot of other factors which will have impacts on predicting the available bikes, such as weather, season, and holiday. If we can get more features, we have the confidence to get higher accuracy model. Moreover,

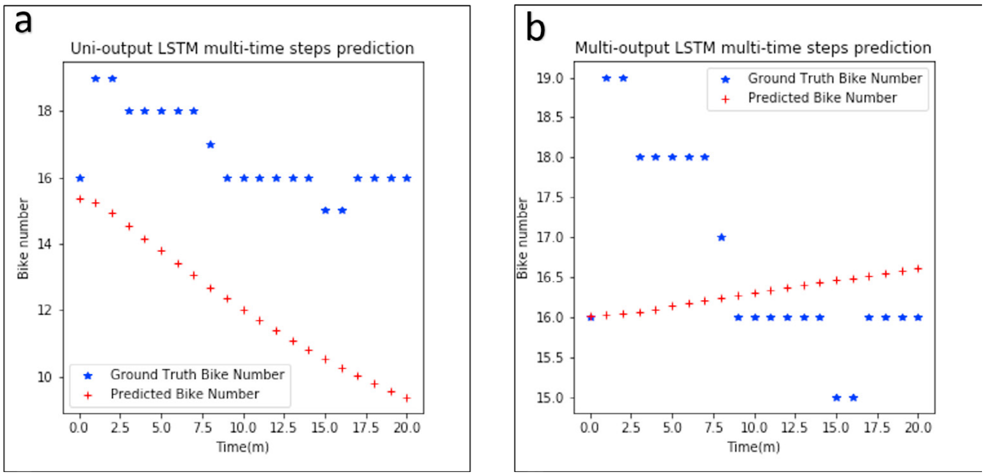

Fig. 6. Multi-time steps prediction.

the LSTM has many hyperparameters which can be tuned, limited by the time, we only select one configuration to implement, we are sure it is not the best one. In our future work, we will try to find the best setting.

6. Conclusion and future work

For improving the accuracy of prediction of bike-sharing availability in a docker, we analyzed all kinds of neural networks' characters and supposed two new solutions. One is adding more input features to the standard LSTM, and the other is using multiple time steps' values of sequence to train the model instead of only using the last time step's value of sequence to train. We also proposed an approach to do multi-time steps prediction. Our experiments verified all our proposed solutions outperformed the standard LSTM solution.

In the future, we will try to find a better data set which has more important features such as weather, season, and holiday. Moreover, we will extend our methodologies to other RNNs such as Gated Recurrent Unit (GRU) on the better data set to investigate how the prediction accuracy will be improved. Furthermore, when we get positive results, we also plan to transplant our verified solutions to the GPU and FPGA platforms to investigate the improvement of performance just as Xu Liu et al. (2018 and 2019) did in [9] and [10].

Acknowledgements

This work has been partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chair in Sustainable Smart Eco-Cloud.

References

[1] ARQU, O.C., 2015. DEMAND FORECAST MODEL FOR A BICYCLE SHARING SERVICE. URL: https://upcommons.upc.edu/bitstream/handle/2117/78121/Tesina.pdf?sequence=1&isAllowed=y.
[2] Ashqar, H., Elhenawy, M., Almannaa, M., Ghanem, A., Rakha, H., House, L., 2017. Modeling bike availability in a bike-sharing system using machine learning, in: MTITS, p. 1. doi:10.1109/MTITS.2017.8005700.
[3] Chen, L., Zhang, D., Wang, L., Yang, D., Ma, X., Li, S., Wu, Z., Pan, G., Nguyen, T.M.T., Jakubowicz, J., 2016. Dynamic cluster-based over-demand prediction in bike sharing systems, in: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM, New York, NY, USA. pp. 841–852. URL: http://doi.acm.org/10.1145/2971648.2971652, doi:10.1145/2971648.2971652.
[4] DeMaio, P., 2009. Bike-sharing: History, impacts, models of provision, and future. Journal of public transportation 12, 3.

[5] Feng, Y., Wang, S., 2017. A forecast for bicycle rental demand based on random forests and multiple linear regression, in: 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), IEEE. pp. 101–105.

[6] Gron, A., 2017. Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. 1st ed., O'Reilly Media, Inc.

[7] Lin, L., He, Z., Peeta, S., 2018. Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach. Transportation Research Part C: Emerging Technologies 97, 258–276.

[8] Liu, T., Fricke, L.M., 2018. Demand forecasting of the bike-sharing service in beijing. MIT Library .

[9] Liu, X., Ounifi, H.A., Gherbi, A., Lemieux, Y., Li, W., 2018. A hybrid gpu-fpga-based computing platform for machine learning. Procedia Computer Science 141, 104 – 111. URL: http://www.sciencedirect.com/science/article/pii/S1877050918318052, doi:https://doi.org/10.1016/j.procs.2018.10.155. the 9th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN-2018) / The 8th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH-2018) / Affiliated Workshops.

[10] Liu, X., Ounifi, H.A., Gherbi, A., Li, W., Cheriet, M., 2019. A hybrid gpu-fpga based design methodology for enhancing machine learning applications performance. Journal of Ambient Intelligence and Humanized Computing URL: https://doi.org/10.1007/s12652-019-01357-4, doi:10.1007/s12652-019-01357-4.

[11] NG, A., 2017. Lecture 8—Neural Networks: Representation. COURSERA: Machine Learning.

[12] Razzaque, M.A., Clarke, S., 2015. Smart management of next generation bike sharing systems using internet of things, in: 2015 IEEE First International Smart Cities Conference (ISC2), IEEE. pp. 1–8.

[13] Shaheen, S., Guzman, S., Zhang, H., 2010. Bikesharing in europe, the americas, and asia: Past, present, and future. Institute of Transportation Studies, UC Davis, Institute of Transportation Studies, Working Paper Series 2143. doi:10.3141/2143-20.

[14] Singhvi, D., Singhvi, S., Frazier, P.I., Henderson, S.G., O'Mahony, E., Shmoys, D.B., Woodard, D.B., 2015. Predicting bike usage for new york citys bike sharing system, in: Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence, p. 1.

[15] Wang, B., 2017. Forecasting with rnn model of keras. URL: http://resuly.me/2017/08/16/keras-rnn-tutorial/.

[16] Wang, B., Kim, I., 2018. Short-term prediction for bike-sharing service using machine learning. Transportation Research Procedia 34, 171 – 178. URL: http://www.sciencedirect.com/science/article/pii/S2352146518303181, doi:https://doi.org/10.1016/j.trpro.2018.11.029. international Symposium of Transport Simulation (ISTS18) and the International Workshop on Traffic Data Collection and its Standardization (IWTDCS18)Emerging Transport Technologies for Next Generation Mobility.

[17] Wang, W., 2016. forecasting bike rental demand using new york citi bike data. Dublin Institute of Technology .

[18] Xu, J., Zhang, Z., Rong, J., 2012. The forecasting model of bicycle parking demand on campus teaching and office district. Procedia - Social and Behavioral Sciences 43, 550 – 557. URL: http://www.sciencedirect.com/science/article/pii/S1877042812010099, doi:https://doi.org/10.1016/j.sbspro.2012.04.128. 8th International Conference on Traffic and Transportation Studies (ICTTS 2012).