A Traffic Prediction Using Machine Learning: Literature Survey

Ji Yoon Kim
Johns Hopkins University – Carey Business School
Washington, DC
jkim624@jhu.edu

I. BACKGROUND

The Bid-Rent theory suggests that prospective land users, including residential, retail, or office users, compete to get the location most accessible to the Central Business District (CBD) where users can decrease their commute cost [1]. However, as recent technology development enables telecommuting, some studies suggested that CBD no longer exists since people can now work from home, which can reduce or eliminate the commute cost. At the same time, many urban economics scholars still support the Bid-Rent theory, finding that commuting cost determines the housing prices [2]-[3].

Also, while various studies have suggested that employers offer higher wages for employees who have a longer commute to attract high-quality labor, others suggested that low-skilled workers are not likely compensated enough for the higher commute cost as they do not have strong bargaining power in the job market [4]-[5], [6]. According to the U.S. Department of Transportation's research in 2003 [7], while middle-income workers use their car to commute to work and spend about 5% of their income on commuting, 66% of low-income workers commute by their car and spend more than 20% of their income on commuting. This implies that low-income workers have higher commute cost and lower income. In addition, studies have shown that longer commute times have a negative impact on people's physical health [8].

Many studies have already proven that machine learning can predict traffic and commute times. While different machine learning algorithms can be used, this study mainly considers using Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), which are based on the Recurrent Neural Networks (RNNs) architecture. Therefore, this paper will take a close look at case studies of two models, as well as case studies of urban economics theories.

II. APPROACH

A. Recurrent Neural Networks

1) Deep Learning with Keras: implement neural networks with Keras on Theano and TensorFlow. According to Gulli and Pal [9], RNNs are a powerful application for approximating and processing arbitrary non-linear functions, suitable for forecasting time series. However, RNNs have a major problem with gradients often vanishing and exploding. As the model backpropagates, gradients gradually increase and decrease, resulting in gradients exploding and vanishing [9]. Thereby, to improve the limitation of the RNNs, scholars developed two models (i.e., LSTM and GRU). The LSTM and GRU have similar characteristics as they can efficiently classify, process, and predict the time-series data. The main differences are that GRU does not have an output gate, which makes its structure simple. Also, GRU works efficiently with a relatively small dataset [9].

B. Long Short-Term Memory

1) Long Short-Term Memory. Hochreiter and Schmidhuber [10] introduced LSTM architecture in 1997, which is an advanced type of RNNs. The advantage of LSTM is that it can work with distributed representations, noise, and continuous values. Furthermore, LSTM requires no parameter tuning, as it automatically finds the optimal output [10]. According to Hochreiter and Schmidhuber, LSTM provides well-generalized output, as it has the following ability: "distinguish between two or more widely separated occurrences of a particular element in an input sequence, without depending on appropriate short time lag training exemplars" [10, pp. 1735-1780].

2) A Comparison of LSTM and GRU Networks for Learning Symbolic Sequences. Cahuantzi, Chen, and Güttel [11] conducted a study comparing the performances of LSTM and GRU on the time series data. The common approach for the time series was that a chunk of time series data is trained using global forecasting models (GFMs). However, Cahuantzi, Chen, and Güttel simplified the process. As a result, they found that GRU performs better on low complexity sequences, while LSTM outperforms GRU on high complexity sequences. They also found that the learning rate and the number of units in each layer are important parameters, but the larger depth of the layer does not have a significant impact on the model's prediction accuracy [11].

3) LSTM Network: A Deep Learning Approach for Short-term Traffic Forecast. Zhao *et al.* [12] found two dimensions (i.e., time and spatial domains) of the LSTM model with a temporal-spatial correlation as a way to predict short-term traffic. To improve the performance of the LSTM model, the origin destination correlation (ODC) matrix (i.e., determines the relationship between different road networks) is integrated into the model *via full connected layers and vector generators*. As the ODC matrix works as a parameter, the LSTM built in this study is differentiated from the conventional LSTM [12]. Zhao *et al.* described the advantage of using LSTM: "LSTM network can divide the long-term traffic forecast into a few short-term forecast processes and output multi-traffic flow forecast results in near future instead of a permanent forecast time" [12, pp. 68-75]. In this study, mean absolute error (MAE), mean square error (MSE), and mean relative error (MRE) are used as evaluation metrics to measure the LSTM model's performance. The RNNs, autoregressive integrated moving average (ARIMA) model, support vector machine (SVM), radial basis function (RBF) network, and stacked autoencoder (SAE) model are used to evaluate the LSTM performance. As a result, LSTM outperformed each of them except RNNs [12].

C. Gated Recurrent Unit

1) Deep Learning with Keras: implement neural networks with Keras on Theano and TensorFlow. In 2014, Cho *et al.* [13] developed the GRU, which is a modified version of the LSTM. The advantage of GRU is that although it is simple in structure, it is resistant to gradient vanishing and explosion, allowing GRU to be trained faster with small computational power. Since the model has a simple structure, GRU shows a well-generalized performance, especially with fewer data. However, if the dataset is large, then it may cause a computational burden, which reduces the performance of GRU. In this case, LSTM generally performs better [9].

2) Are GRU Cells More Specific and LSTM Cells More Sensitive in Motive Classification of Text?. Gruber and Jockisch [14] pointed out that the study of whether GRU outperforms LSTM or vice versa is still ongoing among many scholars. They conducted a study on the small text dataset to examine which model performs well in what kind of condition (e.g., memory cells). Gruber and Jockisch found the following: "GRU cells should show higher specificity as they do not have their own memory and therefore tend to learn more like an exclusion principle. LSTM cells, on the other hand, show higher sensitivity as they strongly adopt onto the data" [14, p. 5]. However, this experiment is only valid on the small datasets, and so further studies are required on the large datasets [14].

3) Using LSTM and GRU Neural Network Methods for Traffic Flow Prediction. Fu, Zhang, and Li [15] introduced GRU in real-time traffic prediction for the first time. The parameter model, such as the ARIMA model, has limitations (e.g., cannot perform on the stochastic and non-linear character of traffic flow) [15]. Fu, Zhang, and Li used GRU in this study, as it is fast in training and has a less complicated structure when compared to the LSTM. The GRU built in this study is based on the GRU that Cho *et al.* developed, and the MAE and MSE are used as evaluation metrics to compare the performance of GRU with LSTM and the ARIMA model. As a result, GRU is proved to be the best model in traffic prediction with slightly better performance than LSTM [15]. However, a dataset with different complexity might not yield the same result.

III. URBAN ECONOMICS

A. Relationship Between Prices and Distance to CBD

1) A Theory of the Urban Land Market. In 1960, Alonso [1] suggested the Bid-Rent theory that the land users (e.g., retail, office, or residential users) compete to get the most accessible location to CBD where users were willing to pay *bid rent* and decrease their commute cost. Another point made in the Bid-Rent theory is that the land that was most accessible to the CBD normally has the highest land price as the result of competition [1]. The Bid-Rent theory has some limitations, but it is one of the famous theories in urban economics and geoeconomics.

2) Quality of Urban Area, Distance from City Centre, and Housing Value. Case Study on Real Estate Values in Turin. In 2018, D'Acci [2] found what impacts housing prices in the city of Turin, Italy, by using a dataset collected in 2016. According to this study, the main factors that affect households' decisions in choosing dwellings are housing cost, transportation cost, and qualitative satisfaction with a location. This study implies that housing prices and distance from the city center have a negative relationship [2]. The relationship found in this study can be different in other cities.

3) The Center Restored: Chicago's Residential Price Gradient Remerges. McMillen [3] traced the housing prices and population change in Chicago's CBD (i.e., Loop). This study found that housing prices in this area declined until the 1980s, as urban employment declined. However, housing prices rapidly rose after the 1990s, as urban employment increased. McMillen found that the housing prices and distance from the city center have a negative relationship [3]. The relationship found in this study can be different in other cities.

B. Relationship Between Income and Commute Time

1) A Note on Commute Times and Average Income Levels. Johnston [4] found that employers were likely to compensate a higher commute cost by increasing the employees' wage (i.e., each additional minute of average commute time increases average household income by an additional $817.73). However, as the relationship found in this study may differ by demographic groups (e.g., age, gender, or education), their bargaining power may be different [6].

C. Relationship Between Income and Commute Time on Disadvantaged Workers

1) Analysis of Commuting Distances of Low-Income Workers in Memphis Metropolitan Area, TN. Antipova [5] found the disparity in commuting patterns using a Tennessee Memphis Metropolitan dataset. One of the findings of the study was that employees who travel more than 50 miles to get to work are mostly disadvantaged workers [5]. This implies that disadvantaged workers with low income take more time commuting compared to higher-earning workers. The relationship found in this study can be different in other cities.

D. Relationship Between Long Commute Time and Physical Health

1) Relationship Between Commuting and Health Outcomes in a Cross-sectional Population Survey in Southern Sweden. In 2011, Hansson *et al.* [8] found a negative relationship between the commute time and the employees' health condition, regardless of the mode of transport using the Southern Sweden dataset. While this study suggested that longer commute time has a negative impact on people's health, reverse causality can exist (e.g., people who are unhealthy might need to commute long distances as their bargaining powers in the job market would not be strong as healthy people). In addition, exogenous variables might have an impact on both commute time and a health condition (e.g., a person's educational achievement might impact both commute time and health condition). The relationship found in this study can be different in other cities.

IV. CONCLUSION

Accurate calculation of the commute cost is crucial for the government to decide whether housing subsidy will be provided to disadvantaged workers, or to create a new method that can reduce the commute cost of the disadvantaged workers by offering mass transit (e.g., bus). As the traffic data is normally large and likely to have a pattern (e.g., data has seasonality as traffic on weekdays at rush hour would be highly congested), a model that can efficiently handle large amounts of time series data is required. As the result, this study will use LSTM and GRU to predict the traffic flow.

References

[1] W. Alonso, "A Theory of the urban land market," The Regional Science Association, vol. 6, pp. 149-157, 1960.
[2] L. D'Acci, "Quality of urban area, distance from city centre, and housing value. Case study on real estate values in Turin," Cities, vol. 91, pp. 71-92, 2019.
[3] D. McMillen, "The center restored: Chicago's residential price gradient remerges," Economic Perspectives, pp. 1-11, 2002.
[4] A. Johnston, "A note on commute times and average income levels," CrossMark, vol. 13, pp. 151-153, 2019.
[5] A. Antipova, "Analysis of commuting distances of low-income workers in Memphis metropolitan area, TN," Sustainability, vol. 12, no. 3, p. 1209, 2020.
[6] P. Ruppert, E. Stanecanelli, E. Wasmer " Commuting, wages and bargaining power, Annals of Economics and Statistics : Labor Market Outcomes: A transatlantic Perspective", GENES, vo. 93, pp. 201-220, 2009
[7] U.S. Department of Transportation, U.S. Department of Transportation "Commuting expenses: disparity for the working poor". 2003.
[8] E. Hansson, K. Mattisson, J. Björk, P. Östergren, K. Jakobsson, "Relationship between commuting and health outcomes in a cross-sectional population survey in southern Sweden," BMC Public Health, vol. 11, p. 834, 2011.
[9] A. Gulli, S. Pal, Deep Learning with Keras: implement neural networks with Keras on Theano and TensorFlow, Birmingham, UK: Packt, 2017.
[10] S. Hochreiter, J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.
[11] R. Cahuantzi, X. Chen, S. Güttel, A comparison of LSTM and GRU networks for learning symbolic sequences, pp. 1-12, 2021.
[12] Z. Zhao, W. Chen, X. Wu, P. Chen, J. Liu, "LSTM network: a deep learning approach for short-term traffic forecast," IET Intelligent Transport Systems, vol. 11, no. 2, pp. 68-75, 2017.

[13] K. Cho, B.Merrienboer. C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio , "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation", in EMNLP 2014, Doha, Qatar, 2014

[14] N. Gruber, A. Jockisch, "Are GRU cells more specific and LSTM cells more sensitive in motive classification of text?" Frontiers in Artificial Intelligence, vol. 3, p. 40, 2020.

[15] R. Fu, Z. Zhang, L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in YAC, Wuhan, China, 2016.