Bike Sharing Demand Forecasting Based on the Informer Model

Ruizhe Qu \textsuperscript{1,a}, Shouhuan Li\textsuperscript{2,b}

\textsuperscript{1} School of Economics and Management, Southwest Jiaotong University, Chengdu, China.
\textsuperscript{2} School of Economics, Southwestern University of Finance and Economics, Chengdu, China.
\textsuperscript{a} gdsjskq@gmail.com, \textsuperscript{b} 908362455@qq.com

Abstract. Advancements in machine learning and artificial intelligence have been instrumental in enhancing demand forecasting for shared transportation modes, significantly benefiting urban traffic management and promoting a balanced supply-demand dynamic. This study harnesses the deep learning Informer model to forecast bike-sharing demand, integrating multiple factors to provide an empirical foundation for transport scheduling and redeployment, thereby contributing to the advancement of the field. Our experimental analysis reveals that the Informer model outperforms LSTM in predictive accuracy, as evidenced by improved $R^2$ scores and MSE values, reflecting gains in training efficiency and predictive precision. Despite these advances, predictions for extreme demand values, particularly during peak usage periods, indicate potential areas for refinement. Future research directions are oriented towards optimizing model accuracy by leveraging multimodal data for a comprehensive analysis and enhancement of traffic management strategies. This approach aims not only to better predict demand and alleviate congestion but also to deliver more optimized solutions for businesses and convenience for users. Furthermore, the model's application extends beyond bike-sharing to other forms of shared mobility, such as electric scooters, potentially aiding in reducing carbon emissions and furthering the development of the sharing economy and environmental sustainability.

Keywords: Bike-sharing demand forecasting; Informer model; Time series analysis; Transportation management; Shared mobility systems.

1. Introduction

Bike sharing, an innovative transportation mode predicated on a sharing paradigm, has emerged as one of the predominant methods for short-distance travel. As of 2020, the market value of bike sharing stood at USD 3 billion, with projections suggesting an increase to USD 4 billion by 2026. This significant market valuation underscores the intrinsic value of bike sharing, which is directly reflected in the quality of services rendered to consumers.

Enhancing consumer experiences in bike-sharing fundamentally relies on the availability of adequate bicycles at necessary locations and times. Thus, accurately forecasting demand within dispatch operations is essential. From an organizational perspective, the essence of most bike-sharing operations lies in strategizing bicycle allocation to meet varying demands across different temporal and spatial dimensions. O'Mahony E et al. [1] have explored the deployment of intelligent bike-sharing systems in urban settings, where a pivotal aspect involves precise demand forecasting at diverse times and locales. In city centers, escalating traffic congestion and pedestrian volumes render bike sharing a feasible solution. However, the proliferation of bikes and suboptimal scheduling contribute increasingly to urban traffic chaos. Building on a bike-sharing flow model, Yang Z et al. [2] addressed the challenges of unequal bike distribution across stations. Insufficient bike availability or poor distribution during peak periods can lead to customer loss and economic detriment, highlighting critical issues in bike-sharing management. Consequently, accurately predicting bike demand in various regions becomes imperative. Within the intricate sharing economy, forecasting for bike sharing not only elevates service levels but also supports other shared mobility services. Bike sharing addresses the traditional mismatch between travel time and space, underscoring the importance of temporal and spatial scheduling. Demand forecasts can refine scheduling strategies,
enabling platforms to judiciously allocate resources, enhance service quality, and increase order volumes and economic returns.

In the realm of bike-sharing demand forecasting, analyses are predominantly conducted from a conventional machine learning perspective. Research by Jangwoo Park et al. [3] involves modeling through algorithms such as Support Vector Machines and Decision Trees. This study is dedicated to forecasting total demand and proposes a methodology based on data mining. It incorporates inputs such as weather condition data and previous demand metrics to predict future demand within a specified timeframe, thereby offering decision support to bike-sharing and other shared transportation entities. The model for this thesis is a variant of the Transformer model—the Informer model, introduced by Zhou H [4] and colleagues circa 2021. This model, specifically designed for long-time series forecasting, simplifies the attention mechanism of the standard Transformer model and enhances performance for extended time series predictions. The primary contributions of this paper are as follows:

1. Multi-feature variable prediction that leverages the interrelationships among various data fields to more accurately forecast ridership demand, thereby substantially augmenting predictive efficacy.
2. Application of the Informer model in conjunction with the Transformer, including refinements to the original attention computation mechanisms of the Informer to boost model effectiveness.
3. Execution of comprehensive data experiments and statistical validity tests to confirm the suitability of the Informer framework for addressing bike-sharing scheduling challenges, with demonstrated high precision in forecasts.

The subsequent sections of this document are structured as follows: Section II reviews existing research on demand forecasting in bike-sharing and broader transportation ridership contexts. Section III delineates the problem definition and explores potential solutions for this experimental endeavor. Section IV elaborates on the Informer model, its application to predicting bike-sharing ridership, and specifics regarding model optimization. Section V presents the experimental setup and findings. Finally, Section VI concludes the paper and discusses prospects for future research.

2. Background on relevant technological developments

Initially, forecasters predominantly employed linear predictive models, including the historical mean method, ARIMA (Autoregressive Integrated Moving Average), and Kalman filtering, to predict traffic flows. Kalman filtering, an efficient recursive filter, is utilized to estimate the system's current state. While these models are simple and effective, particularly with smaller datasets, their predictive accuracy deteriorates with increasing data volumes. This decline is attributed to the models' inability to capture the inherent nonlinearity and instability present within the data.

To address the constraints inherent in linear models, researchers have delved into nonlinear forecasting models encompassing fractal theory, wavelet analysis, and chaos theory. These models, designed to capture intricate data patterns, require extensive computational resources, are complex in their construction, and present challenges in application. With the advent and rapid advancement of machine learning and deep learning technologies, an increasing corpus of research has adopted these techniques for forecasting demand in ride-sharing services. The evolution from decision trees includes studies by Yexin Li et al. [5], who employed Gradient Boosting Regression Trees for stratified forecasting. Moreover, approaches such as random forests have been used by Young-Hyun Seo [6] and colleagues to forecast, particularly offering predictive capabilities for public broadcast systems. Extending to machine learning models ranging from support vector machines to artificial neural networks, Pei-Chann Chang [7] and others have explored combining AIS with neural networks to forecast demand, yielding substantial results. Advanced deep learning models, such as RNN, LSTM, and GRU, have demonstrated pronounced strengths in managing large datasets and deciphering complex nonlinear patterns. Notably, Long Short-Term Memory (LSTM) and GRU incorporate gating mechanisms that mitigate the long-standing issue of gradient vanishing in RNNs when processing extensive sequence data, thereby markedly improving the accuracy of forecasts.
Wang B et al. [8] have notably succeeded in employing these efficient models for bike-sharing demand forecasting, with LSTM particularly adept at managing time-related cyclical and seasonal variations and accommodating diverse factors like weather conditions and holidays.

To further refine predictive accuracy, scholars have extended their focus beyond the characteristics of time series data to include the influence of spatial features. For example, hybrid deep learning architectures that integrate Convolutional Neural Networks (CNN) with LSTM systems are adept at capturing both the local spatial relationships and temporal dynamics of traffic flows. Additionally, models such as Graph Convolutional Networks (GCN) exploit the graphical structure of urban road networks to advance the comprehension and forecasting abilities concerning intricate traffic networks. Notably, Xinwei Ma et al. [9] have developed a solution that synergizes LSTM and GCN. In this model, the GCN component extracts spatial data, while the LSTM component processes temporal data, collectively augmenting the model’s learning efficiency.

The Transformer model [10], developed by the Google team, represents a groundbreaking architecture that eschews both recurrence and convolution in favor of attention mechanisms, specifically designed for sequence transduction tasks such as machine translation. This architecture's formidable predictive capabilities are not confined to translation but are also efficacious in domains such as regression forecasting. The Transformer's capacity to model dependencies irrespective of their positional distance within input or output sequences furnishes a robust tool for addressing time series prediction tasks, potentially surmounting traditional challenges linked with long-term dependencies and the parallelization of computational processes. Moreover, the Transformer’s employment of self-attention mechanisms enhances the interpretability of the model's predictions, offering a significant advantage over more opaque models like deep RNNs. This attribute is particularly beneficial in fields such as demand forecasting, where discerning the factors that influence predictions is crucial for informed strategic planning and operational modifications.

3. Problem Definition and Description

3.1 Development of the Solution

Initially, forecasting future demand necessitates reliance on historical data, supplemented by a range of environmental factors, including temperature, humidity, and wind speed. Consequently, this task evolves into a long-term time series forecasting project, offering valuable insights and recommendations for the scheduling operations of bike-sharing and other shared transportation services. Several methods and models exist for time series forecasting, among which convolutional neural networks, recurrent neural networks, and LSTM have been extensively employed. While LSTM has been instrumental in addressing the challenge of maintaining distant memories in long time series, it exhibits certain limitations for more extended sequences. The introduction of the Informer model has substantially enhanced the resolution of long-term memory issues and introduced more effective methods. Hence, it is anticipated to exhibit robust performance in the field of bike-sharing demand forecasting.

3.2 Predictive Features

In the field of bike-sharing demand forecasting, it is imperative to comprehend and analyze a variety of potential predictors. Historically, research has primarily utilized historical demand data for forecasting purposes. However, it has been acknowledged that additional factors, such as weather conditions and holidays, exert a significant influence on demand patterns. Consequently, our project is dedicated to constructing a multivariate predictive model that incorporates the following determinants and their potential impacts on demand:

- Temporal Factors: The influence of specific dates and times on bike-sharing demand is considered, focusing on daily peak and off-peak intervals, as well as variances between weekdays and weekends.
Advances in Engineering Technology Research  
CVMARS 2024  
ISSN:2790-1688  
Volume-11-(2024)

- Meteorological Conditions: The impact of weather elements, including wind speed, humidity, and overall weather conditions, is analyzed. Elevated wind speeds may prompt a preference for ride-sharing services, whereas increased humidity and adverse weather may heighten vehicle demand.

- Urban Events: Large-scale events, holiday celebrations, and other municipal activities that temporarily augment demand are taken into account, potentially enhancing bike-sharing utilization in particular areas and times.

- Special Events: The effects of unique occurrences, such as urban construction or temporary road closures, are examined for their potential to alter conventional travel patterns and influence bike-sharing utilization.

- Seasonal Variations: The enduring impact of seasonal shifts on demand is assessed. For instance, demand for bike-sharing may significantly exceed other periods during the summer tourist peak.

4. Model Structure

The model employed in this study evolves from the Transformer architecture, consisting of an encoder and a decoder. The encoder converts the input sequence into a high-dimensional representation, utilizing a multi-head self-attention mechanism to address challenges associated with long sequences. Residual connections and normalization layers are utilized to facilitate the flow of gradients and enhance generalization capabilities. The decoder introduces a masking mechanism within the self-attention layer to ensure dependency solely on previously generated elements, and incorporates an encoder-decoder attention layer that captures the interdependencies between the input and output, culminating in the generation of the target sequence.

Despite the Transformer's considerable success in long-sequence prediction, it inevitably presents several structural redundancies and inherent mechanistic shortcomings:

- The weight calculations within the self-attention mechanism are characterized by a long-tail distribution, indicating inherent sparsity, as evidenced in the research conducted by Haoyi Zhou et al. This sparsity results in quadratic computational complexity, which hinders the efficient production of meaningful weights in the self-attention mechanism.

- The decoding process in the Transformer architecture is dynamic, wherein each subsequent output relies on the predictive results of the preceding time step. This dependency significantly slows down the inference process.

Consequently, this study utilizes the Informer model, an adaptation of the Transformer architecture with numerous effective modifications, to forecast the positioning and scheduling of bike-sharing systems.

The input for long-term time series forecasting extends beyond local timestamps, which include position embeddings as presented in the Transformer, to encompass global timestamps, such as information on year, month, and day. Predicting the locations of bike-sharing systems necessitates acknowledging long-term dependencies, such as intraday cyclical patterns and variations in ambient conditions. These requirements result in the traditional Transformer's attention computation mechanism being slow and, at times, prone to inaccuracies.

Within the Informer model, the ProbSparse self-attention mechanism employs a probabilistic approach whereby each query vector selects a subset of keys correlated to it via a probability distribution. This distribution may be generated through several methods, most notably by scoring the keys using a softmax function followed by sampling.

The incorporation of this probabilistic selection process in the ProbSparse mechanism enables more efficient computation by selectively focusing attention on the most relevant keys, thus mitigating the computational load typically associated with traditional attention mechanisms. This approach not only streamlines the attention process but also enhances model performance in handling large datasets.
Consequently, each query interacts with only a select number of relevant keys to form attention weights, which significantly reduces computational demands. The ProbSparse self-attention mechanism substantially lowers computational complexity encountered in forecasting bike-sharing location scheduling, bolsters computational efficiency, and minimizes inaccuracies, enabling companies to perform real-time predictions of bike-sharing distribution.

Furthermore, the model employs a self-attention distillation strategy, utilizing a Conv1D layer on top of the output from each attention layer, followed by Maxpooling to halve the outputs, thus emphasizing predominant attentions and adeptly managing overly extended input sequences. Through these modifications to the self-attention mechanism and structure, the Informer model adeptly addresses the quadratic computational complexity and substantial memory consumption inherent in the Transformer model's self-attention mechanism as previously discussed.

![Figure 1 Architecture of a Multi-Head Attention Neural Network for Sequential Data Processing](image)

Furthermore, within the Informer model, batch generative forecasting is utilized, which outputs multi-step predictions directly. This is achieved by predicting the entire output sequence of a long series in a single forward pass, dynamically selecting segments of the sequence proximal to the prediction target to act as "start tokens." Consequently, this simultaneous computation of multiple outputs effectively addresses the slow inference process associated with the earlier mentioned Transformer model.

![Figure 2 Encoding and Decoding Sequential Model](image)

5. **Experimental Results and Model Performance Evaluation**

This chapter details the precise parameter settings of the experiments, the procedural execution, and the resultant findings. The experiments assess the Informer prediction model's accuracy in forecasting bike-sharing demand. In Section V-A, the dataset, model parameter configurations, and the rationale behind their adjustment are introduced. Section V-B elaborates on the training process
of the experiment and its convergence properties. Section V-C displays the predictive results and the fit on the test set, accompanied by an array of metrics, and offers a comparative analysis of the Informer's performance against alternative forecasting algorithms, including LSTM.

5.1 Dataset & Experimental Settings

In this experiment, an authentic bike-sharing operational dataset was sourced from Kaggle to facilitate the forecast of demand for rides. Revisiting the potential demand-influencing factors identified in Chapter Three, the dataset effectively encapsulates the subsequent features:

- dteday: date
- season: season (1:springer, 2:summer, 3:fall, 4:winter).
- yr: year (0: 2011, 1:2012)
- mnth: month (1 to 12)
- holiday: weather day is holiday or not
- weekday: day of the week
- workingday: if day is neither weekend nor holiday is 1, otherwise is 0
- weathersit: 1: Clear, Few clouds, Partly cloudy, Partly cloudy.2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist.3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds.4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog
- temp: normalized temperature in Celsius
- atemp: normalized feeling temperature in Celsius
- hum: normalized humidity
- windspeed: normalized wind speed. The values are divided to 67 (max)
- cnt: count of total rental bikes including both casual and registered

5.2 Training Parameters

Modifications to hyperparameters were conducted on the Informer model to optimize it for the specific conditions of the project experiment. The feature settings were adjusted with options M (multivariate predict multivariate), MS (multivariate predict univariate), and S (univariate predict univariate). Through empirical validation, option M was found to significantly excel over options S and MS. Parameters for sequence length were configured with the encoder's input sequence length at 192 and the decoder's prior sequence length at 96. The prediction length was determined to be 48 time points. The model underwent training for 6 epochs, implementing an early stopping criterion with a patience setting of 3 to circumvent overfitting. The dataset, which logs usage statistics hourly, employs timestamps to retain the temporal sequence data. For instance, the initial timestamp in the dataset is recorded as [2011, 01, 01, 00], denoting year, month, day, and hour sequentially.

5.3 Figures and Tables

Subsequent to the optimization of hyperparameters, the model was executed, performing initial predictions and undergoing evaluation on the test set. The selected hyperparameters demonstrated superior performance, aligning well with the project's criteria for offering demand scheduling strategies to bike-sharing enterprises. The capacity to concurrently forecast 48 time intervals notably improved the foresight of predictions, ensuring accurate and immediate responses to unexpected demand peaks—a crucial aspect for optimizing the company's scheduling and enhancing customer satisfaction.

Figure 3 corroborates that the Informer model's predictions correspond substantially with the actual demand, validating the feasibility of its impressive predictive prowess within the bike-sharing demand forecasting realm. Moreover, the model's forecasts are anticipatory, allowing for prompt predictions in the face of demand surges and validating its practical utility in real-world bike-sharing
operations. Training for 6 epochs, depicted in Figures 4 and 5, reached convergence on the training set while successfully circumventing overfitting.

To substantiate the Informer's enhanced performance in forecasting bike-sharing demand time series, we juxtaposed it with the LSTM, a prevalent model in the domain of time series forecasting. The experiment was structured around two principal criteria: (1) Both models projected forecasts across an identical temporal series length, predicting prospective ride-sharing demand for a subsequent 48 time intervals, with a subsequent analytical comparison being conducted, and identical sequence lengths were employed as the input for the time series. The LSTM and Informer models were calibrated with closely matched hyperparameter configurations to negate any disparities in performance outcomes stemming from hyperparameter variations, exemplified by setting both models' hidden layer dimensions to 512 and fixing the epoch count at 6.
The R² score serves as a metric of the correlation level between a model's forecasts and actual outcomes, spanning a range from 0 to 1, with 1 denoting perfect predictability. Upon comparative analysis, the Informer model exhibits an R² score exceeding that of the LSTM, suggesting enhanced fitting efficacy and superior testing performance on the test set, closely mirroring the true values. Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) are indicative of the variance between the forecasted and actual data points. Notably, the Informer model demonstrates a reduced margin of discrepancy in contrast to the LSTM model, reflecting an improved forecasting function. The preeminence of the Informer is attributable to several factors:

- Proficient handling of long-term sequences: The Informer model's self-attention mechanism adeptly captures extensive temporal dependencies within time series data, in contrast to the gradient issues persisting in LSTM architectures.
- Parallel computation: The Informer model facilitates concurrent data processing through self-attention and achieves predictive outcomes in a singular step via its Generative Style Decoder, divergent from the LSTM's iterative, incremental forecasting approach.
- Enhanced interpretability: The Informer model, utilizing attention weights, not only refines its predictive accuracy but also provides both a visualization and an explanation of the predictive procedure, augmenting model transparency.

6. Conclusion And Future Work

The progression of information technology has considerably elevated the influence of machine learning and artificial intelligence in the domain of shared transportation demand forecasting, ameliorating the dynamics of urban mobility and the equilibrium of traffic supply and demand, thus safeguarding the stability of urban transit operations. Employing the deep learning Informer model, this investigation predicts bike-sharing demand by integrating diverse factor conditions, furnishing empirical support for the allocation and redistribution of transportation resources and fostering progress in this arena.

In comparative experiments, the Informer outperformed the LSTM model, exhibiting a 0.055 enhancement in the R² Score and advancements in MSE, reflecting improvements in both training velocity and accuracy. Nonetheless, the model's predictions for peak values, especially during surges in cycling demand, manifest deviations, signaling a need for further refinement.

Prospective studies should aim to refine model accuracy by harnessing multimodal data for a comprehensive analysis and optimization of traffic management. Such methodologies promise not just improved forecasting of demand and traffic congestion but also proffer more streamlined.
solutions for enterprises, thereby facilitating commuter transit. Additionally, the applicability of this model transcends bike-sharing, extending to other shared mobility services like electric scooters, and bears significance in curtailing carbon emissions and catalyzing growth within the sharing economy and environmental sectors. Collectively, this research contributes to the enhancement of urban transportation frameworks and provides a technological foundation within the expansive context of the sharing economy.

References

[1] O'Mahony E, Shmoys D. Data analysis and optimization for (citi) bike sharing[C]//Proceedings of the AAAI conference on artificial intelligence. 2015, 29(1).


