WILEY | Hindawi

Research Article

Travel Time Probability Prediction Based on Constrained LSTM Quantile Regression

Hao Li [ID],[1] Zijian Wang [ID],[2] Xiantong Li [ID],[3] Hua Wang [ID],[3] Yongxing Man,[3] and Jinyu Shi [ID][3]

[1]*Broadvision Engineering Consultants, Kunming 650041, China*
[2]*Shandong Key Laboratory of Smart Transportation (Preparation), Jinan 250014, Shandong, China*
[3]*School of Transportation Science and Engineering, Harbin Institute of Technology, No. 73 Huanghe Road, Nangang District, Harbin 150090, China*

Correspondence should be addressed to Hua Wang; wanghua@hit.edu.cn

Travel time reliability assessment has been widely used in recent years to evaluate the performance of transportation networks and measure the operation level of transportation systems. Weather, as one of the important factors influencing travel time reliability, affects the relationship between the supply and requirement of urban road networks. Considering the traffic characteristics under different traffic conditions, a study on the influence of weather on travel time reliability under different conditions is proposed to predict the probability of travelers completing their trips within the expected time under different weather conditions. Based on the urban road network data and cab trajectory data of Harbin city, this paper correlates the floating vehicle location with the road network information through a hidden Markov model to reduce the influence of vehicle trajectory errors on the calculation results of path travel time. To analyze the entire distribution of extreme travel time and its impact on the reliability of travel time under various traffic situations, it captures the tail features of the travel time distribution based on extreme value theory. Then, to increase the predictability of each quantile, it combines a deep-learning LSTM model and a quantile regression model to create a probabilistic travel time prediction model utilizing combined layers. The proposed model is compared with the linear quantile regression and neural network quantile regression models, and the model is evaluated in terms of point prediction results and probabilistic prediction results, respectively, to ensure the accuracy of predictions from the model. As a result, the prediction accuracy of the model in this paper is greatly improved, and the degree of violating quantile constraints is greatly reduced.

1. Introduction

Travel time reliability (TTR) is one of the most important factors to measure the stability of transportation systems. It also serves as a guide for travelers, advising them on when to begin their journey and how to get at their destination on time. According to a study [1], travelers are more concerned with the accuracy of journey time predictions than with the actual travel time. With the improvements of high quality and comprehensiveness of datasets by modern monitoring instruments and data collection mothers, it is now possible to anticipate travel time by using real-time traffic data [2].

Many factors influence travel time reliability, including the time of day, whether it is a weekday or not, traffic accidents, traffic capacity, traffic control measures, and weather. According to the supply-demand relationship of the transportation system, these factors can be divided into two categories: unrepetitive and repetitive influencing variables. For example, traffic accidents and weather circumstances fall under the first category, whereas the time of day, weekdays or not, falls under the second category [3]. It is critical to concentrate research efforts on these occurrences, which are important to travel time reliability. They are also fundamental requirements of advices about traffic policy or a traveler counseling [4].

Reliability of travel time prediction has always been a major issue in the intelligence transportation system (ITS). ASAKURA [5] proposes the concept of travel time reliability, which is thought to be one of the earliest studies. The following research focuses on ways to increase the accuracy of TTR, particularly TTR measuring indicators. Statistical approaches are based on statistics of travel time under different quantiles to verify TTR. Pu and Wenjing [6] define a coefficient of deviation as the measurement index, which is the ratio of travel time to the average travel time. Tu et al. [7] use the term travel time variability (TTV) to represent the range of the travel time, which is defined as the difference between the 90th and the 10th percentile travel time. Van Lint and van Zuylen [8] depict travel time reliability by the inclination and width of travel time distribution. The statistical characteristics of travel time variability, according to statistical methodologies, are more resilient than those based on the mean and variance. By computing the extra travel time, buffer time methods provide the uncertainty of travel time. The buffer time index and the travel time index are two of the buffer time methods that have been studied and applied extensively. Lomax et al. [9] investigate the associations between the 95th percentile and free flow travel time by using the buffer time index. Chen et al. [10] study the travel time index to investigate the various ratios of travel time between average and free popular. The buffer time index and the travel time index, on the other hand, accurately describe the reliability of travel time of the normal distribution, but they are insufficient to assess the nonnormal distribution. The pain index and the on-time arrival rate are two of the travel delay methods that are used to analyze the head and tail distributions in the entire span of travel time. The pain index is defined as the ratio between 20% of the higher travel rate and the average travel rate, and the on-time arrival rate is defined as the probability that the travel rate is less than 110% of the average travel rate.

Owing to the regular occurrence of adverse weather, academics have been paying increasing attention to the impact of weather on travel time in recent years. Many research studies have been conducted on the effect of weather on travel time reliability. Faouzi et al. [11] research the global framework of the expressway travel time prediction system, which confirms that meteorological conditions have an impact on travel time. Li et al. [12] propose a soft set theory-based function model to establish the relationship between weather and travel time. The model also considers how rainfall intensity and visibility affect expressway travel time. Andersen and Torp [13] collect 5-year high-speed road driving statistics in Denmark. According to this dataset's study, snow will increase travel time by 27%, while a strong headwind will increase travel time by 19%. Tsapakis et al. [14] observe that rain and snow have an effect on expressway travel time. The heavier the rain and snow are, the worse the travel time is influenced. Zou et al. [15] investigate the impact of weather on travel time on expressway and main roads, respectively, during rush hour and not during peak hour. The study discovered that bad weather has a greater impact during rush hours. Wang et al. [16] study the impact of snowfall on bus travel time and find that, for every 1 millimeter increase in snowfall accumulation, bus travel time increases by 0.483 minutes. Kamga and Yazıcı [17] divide weather types into sunny, light rain, heavy rain, light snow, and heavy snow and put forward different views: in terms of weather impact, bad weather may reduce the coefficient of variation (COV) and standard deviation in some time periods, improving the reliability of travel time. Shi et al. [18] discover that light rain has little influence on travel time variability; however, heavy rain increases unpredictability and so affects reliability.

The study of travel time reliability necessitates the use of travel time distribution (TTD). Scholars have been using many models to describe travel time distribution in recent years. Kim and Mahmassani [19] use a gamma distribution to simulate the model, which matches well. Furthermore, Castillo et al. [20] think beta distribution, Al Deek and Emam [21] think Weibull distribution, and Yang et al. [22] think bimodal distribution can better characterize travel time distribution. Ansari Esfeh et al. [23] propose a composite generalized extreme value distribution (CGEV) to study the impact of monthly and seasonal changes in extreme travel delay in the road network in their most recent work. The fitting of the travel time distribution is a variable in different research studies, which is due to the differences in traffic environments, traffic legislation, and driver psychology in various countries and locations. However, differing weather conditions will have an impact on the travel time distribution. For example, according to Shi et al. [18], in high-speed sections, the lognormal model fits best in good weather, light rain, and medium rainfall circumstances, and the Weibull model fits best in blizzard conditions.

The impact of weather on travel time reliability has been the subject of numerous studies. However, the majority studies focus on the travel time reliability of expressway, which is uncommon in urban roads [12]. Simultaneously, researchers place a greater emphasis on average travel time. Travelers in numerous cases are concerned about extreme travel time [24].

The paper aims to analyze the influence of different weather types on upper quantile travel time and predict travel time under different quantiles based on the quantile regression model and deep-learning model, combined with floating car data on urban roads in Harbin, in order to further explore the weather's influence on travel time reliability. The following are the paper's main contributions:

(1) The paper uses a hidden Markov model (HMM) to map match the track data of a floating automobile. The paper also develops a method for calculating journey time, as well as obtaining the travel duration of floating cars on Harbin's urban highways and the probability distribution of travel time.

(2) The paper divides the weather into categories and examines the impact of weather on the upper percentile travel time over time.

(3) To estimate trip times under different quantiles, a model incorporating quantile regression and LSTM is proposed in this study. Meanwhile, the research investigates and provides restriction constraints to avoid crossing between adjacent quantiles. The cumulative experience distribution of the travel rate

may be obtained based on the travel rate of each quantile, which can be used to record the variation of the travel rate at different times and provide guidance for passengers.

The rest of this paper is arranged as follows: Section 2 gives a description of the data and the travel time prediction model. The experimental description is given in Section 3. The conclusions of this effort are given in Section 4.

2. Methodology

The section introduces the data used and the model of travel time prediction.

2.1. Data Introduction. The investigation of the reliability of travel time is based on GNSS data. The information for this study came from Harbin Transportation Bureau, which contains the driving records of roughly 5000 taxis. Table 1 shows the comprehensive details for some of the data.

As given in Table 1, the data recorded the taxi's unique number (DEVID), LONGTITUDE, LATITUDE, SPEED, ORIENTATION, and UNIXTIME. The data for this study were collected from hundreds of millions of floating vehicles between June 1, 2016, and December 31, 2016. Previous studies have determined that time periods of 10 to 15 minutes are appropriate for analyzing traffic conditions [7, 25]. As a result, data in this paper are separated into 15-minute time slots, and the data for a day are divided into 96 time slots.

The data used in this paper meet the research requirements both in terms of time and space. In terms of time, floating car data include data from both working and nonworking days, as well as peak and flat peak times throughout the day. Because the travel time fluctuated with seasons, the collection of data was begun on June 1, 2016, and ended on December 31, 2016 when there was sufficient rain and snow. In terms of space, floating car data are fed into the road network and generally cover all Harbin's roadways.

There are inaccuracies in GPS equipment due to building occlusion and other factors, and some abnormal data are contained in the original data. Before using GNSS data, they must first be cleaned to remove any aberrant data interference of abnormal data. In this paper, anomalous data were filtered out based on their speed. Data with an instantaneous speed of 0 for more than 5 minutes, as well as data with an instantaneous speed substantially beyond the speed limit on the road, were regarded as abnormal and eliminated. It is important to match the trajectory points due to the presence of GPS track positioning mistake. This paper uses a hidden Markov model (HMM) to perform map matching. The effect diagram of the road network before and after matching is shown in Figure 1. The HMM algorithm can successfully match the track points to the road network based on the matching results. This study proposes using the travel rate (TR) instead of travel time for research focused on map matching. The travel rate refers to the time it takes the floating car to go one kilometer, and it can be used to reduce

errors caused by the length of a road section. This paper finds out all vehicles passing through the study path and acquires the starting and finishing points as well as the travel time stamps of vehicles based on the vehicle ID in order to determine the travel rate of various vehicles. The travel rate is the ratio of the total travel distance and travel time.

2.2. Influence of Rain and Snow on Travel Time Reliability. The variety of reliability measurement methodologies is due to the complexity of network features and impacting factors. Currently, researchers and scholars offer four types of reliability measurement methods: statistical methods, buffer time methods, travel delay methods, and probability measuring methods. The buffer time method offers a decent description for the normal distribution, but it is insufficient to analyze the "heavy tail" travel time distribution. According to the distribution of travel time in Harbin which is shown in Figure 2, the lognormal distribution best fits the path travel time of Harbin by distribution fitting and hypothesis testing. As a result, using the buffer time method to assess the reliability of trip times in Harbin is insufficient. Furthermore, the travel delay method and the probability measurement method are used to calculate the likelihood of trip delay and road congestion, and travelers are unable to quantify specific trip times using the above indicators. The statistical features of travel time changes are more robust than those based on the mean and variance in statistical methods [23]. As a result, travel time variability (TTV) is used to characterize the reliability of trip times at a given point of time.

In China, rainfall and snowfall are classified as follows: Rainfall less than 10 mm within 24 hours is classified as light rain, rainfall between 10 and 25 mm is classified as moderate rain, and rainfall larger than 25 mm is classified as heavy rain. A snowfall of 0.1–2.4 mm within 24 hours is considered light snow, a snowfall of 2.5–4.9 mm is considered moderate snow, and snowfall greater than 5 mm is considered heavy snow. The weather of Harbin city from June to December 2016 can be classified using this criterion, with clear, light rain, moderate rain, heavy rain, light snow, moderate snow, and heavy snow.

Since traffic patterns vary on different days of the week and at different times of the day, and the impact of weather on travel time reliability changes as traffic patterns change [24], it is important to evaluate traffic characteristics at different times of the week before analysis. The amount of time it takes to travel varies depending on the day of the week. The data from June to December 2016 are summarized in Figure 3 to create a thermal map of average travel time for different days of the week. The tendency is very similar from Monday to Friday, and tendencies are also very similar on Saturday and Sunday. Weekend travel times are often more consistent than weekday travel times, with no noticeable morning peak, which is in keeping with travelers' travel patterns. The impact of rain and snow weather on the reliability of travel time should be evaluated on weekdays and weekends, respectively, due to the clear cyclical tendency of travel time.

TABLE 1: Floating car data example.

DEVID	LONGTITUDE (E°)	LATITUDE (N°)	SPEED (100 m/h)	ORIENTATION (°)	UNIXTIME
100322318	126.61708	45.76433	18	7	1467327699
100307956	126.67012	45.72983	261	84	1467327842
100322118	126.61708	45.76433	18	7	1467328120
100322776	126.69008	45.79545	122	76	1467327639
. . .	. . .	. . .	. . .	. . .	. . .

Note. Orientation is a direction of a vehicle moving, with 0 degrees indicating due north, 90 degrees indicating due east, 180 degrees indicating due south, and 270 degrees indicating due west.
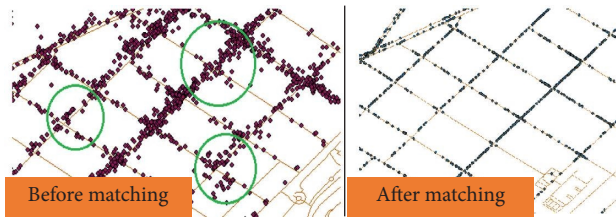

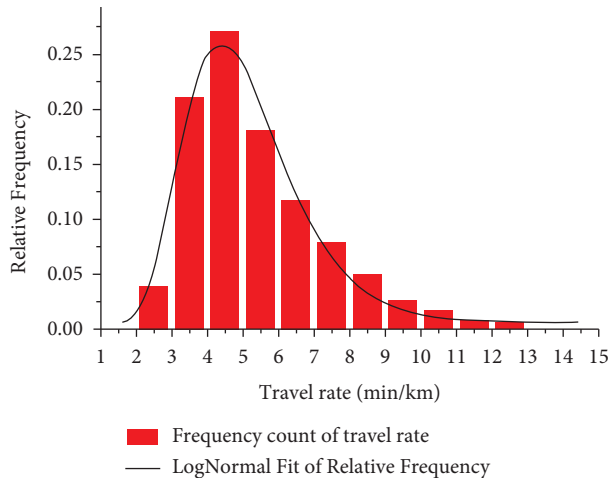FIGURE 1: Track diagram of the road network floating car.

FIGURE 2: Fitting diagram of travel time probability distribution.

FIGURE 3: Average travel time thermal map.

In this paper, the effect of rain and snow on travel time reliability is investigated by dividing datasets into distinct weather conditions. Figure 4 depicts how rain and snow affect TTV. It is deduced from the statistic graphic that, during morning rush hour, light rain days increase TTV compared to bright days, while other weather types decrease TTV. This implies that, in addition to light rain, additional weather conditions improve travel time reliability. The most significant increases were in moderate rain and moderate snowfall. Rain and snow have little impact on travel time reliability on weekends because there is no evident morning peak. Rain and snow, in general, reduce travel time reliability throughout the night periods (20:00–8:00) and also improve travel time reliability during the day. It is depicted in Figure 5.

Rain and snow have a greater impact on the reliability of travel time in the morning peak, as can be seen from the preceding statistics (7:00 am–9:00 am). Severe weather has also been shown to have a major influence on travel time only when traffic volume exceeds a particular threshold [26].
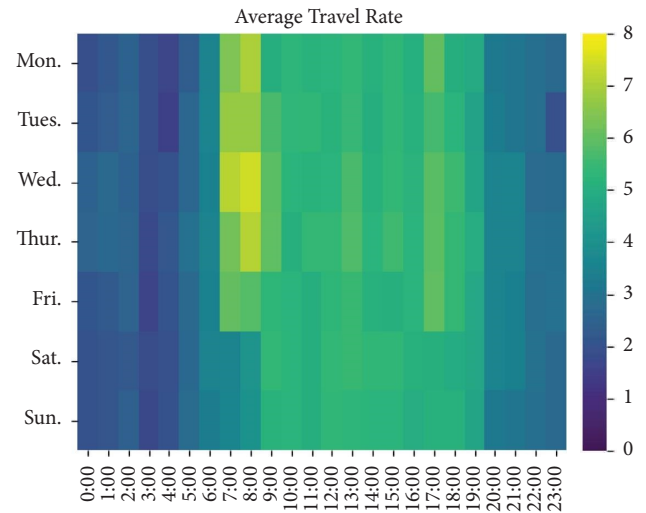
As a result, quantitative analysis is deployed on the morning peak time independently, taking the 90th percentile travel rate as an example to study the influence of rain and snow on the reliability of the morning peak travel time. Tables 2 and 3 demonstrate this. Except for light rain, all other weather conditions reduced the 90th percentile travel rate, with moderate rain and moderate snow having the greatest impact.

2.3. Probability Prediction Method Based on LSTM Quantile Regression. Rain and snow will have different impacts on the reliability of travel time at different times, as shown in the above analysis. It is critical for travelers to obtain the chance of travel arriving on time. This paper predicts the travel rate of different quantiles under different weather circumstances using a quantile regression model and an LSTM model and then constructs the empirical cumulative distribution curve of travel time. Quantile regression theory combines the memory characteristics of LSTM with the probability prediction function of quantile regression to directly explain the link between response variables (travel rate) and explanatory variables (weather and time) under different quantiles. The predicted values under different quantiles may cross due to the inherent properties of quantiles, affecting the accuracy of the model prediction results. The reliability of travel time can be determined by the quantile of travel time. For example, the 10% quantile and 90% quantile can be used to calculate TTV and thus
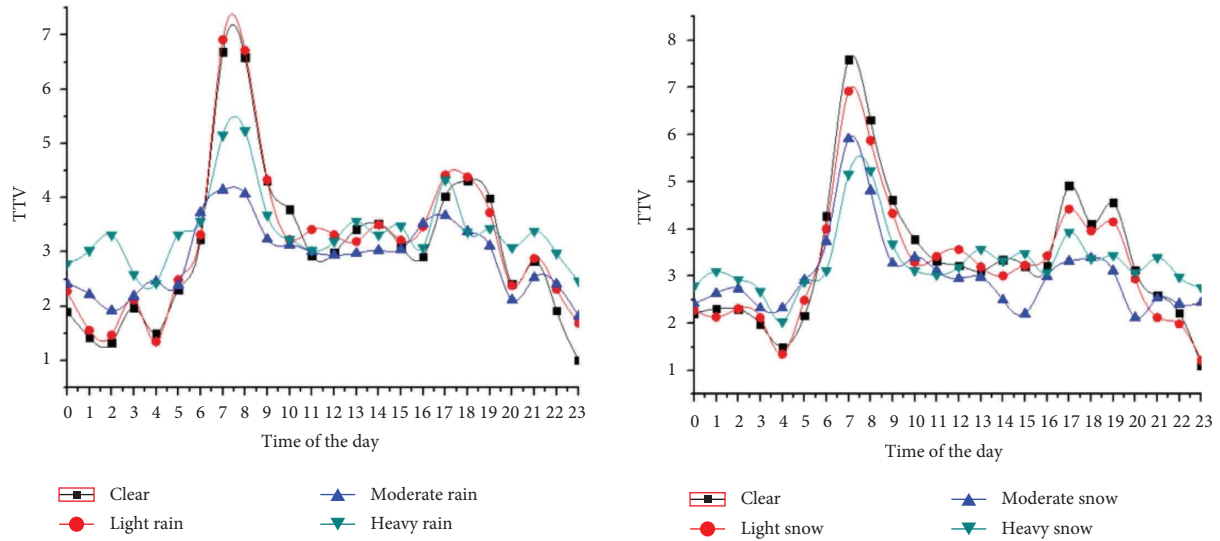
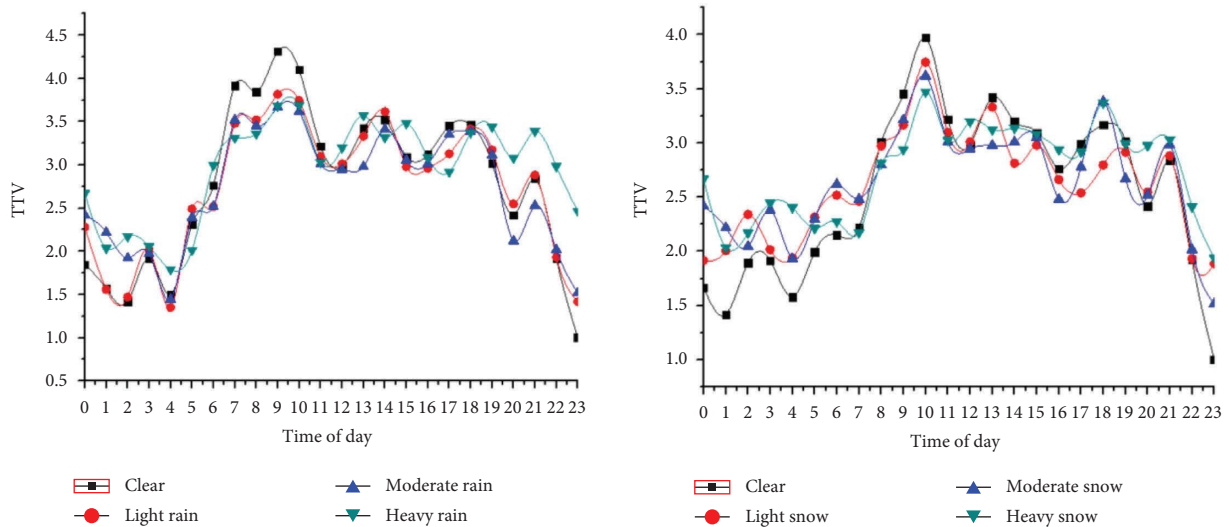FIGURE 4: Impact of rain and snow on TTV on weekdays.

FIGURE 5: Impact of rain and snow on TTV on weekends.

TABLE 2: Effects of rain during morning peak on the 90th percentile travel rate.

	Clear	Light rain	Light rain (%)	Moderate rain	Moderate rain (%)	Heavy rain	Heavy rain (%)
7:00–7:15	5.762	5.972	3.6	4.82	−19.5	5.382	−7.1
7:15–7:30	6.092	8.157	33.9	4.529	−34.5	5.641	−8.0
7:30–7:45	9.187	10.447	13.7	4.961	−85.2	8.775	−4.7
7:45–8:00	10.436	10.914	4.6	5.61	−86.0	8.169	−45.6
8:00–8:15	10.329	9.75	−5.9	5.233	−97.4	6.7	−54.2
8:15–8:30	8.957	9.698	8.3	4.98	−79.8	6.798	−31.8
8:30–8:45	8.869	9.067	2.2	4.966	−78.6	7.029	−26.2
8:45–9:00	7.164	7.686	0.9	4.572	−66.5	5.918	−28.7

determine the range of the distribution of travel time. As a result, this paper proposes a combination layer that violates the penalty term of the quantile prediction value constraint in order to minimize quantile crossing.

In a word, this paper provides a constrained LSTM quantile regression model that properly accounts for the timing and nonlinearity of travel time prediction based on this. The following is an introduction to a specific model.

TABLE 3: Effects of snow during morning peak on the 90th percentile travel rate.

	Clear	Light snow	Light snow (%)	Moderate snow	Moderate snow (%)	Heavy snow	Heavy snow (%)
7:00–7:15	5.562	5.402	−3.0	4.768	−16.7	5.211	−6.7
7:15–7:30	7.677	7.594	−1.1	5.781	−32.8	6.412	−19.7
7:30–7:45	9.161	8.735	−4.9	8.69	−5.4	6.886	−33.0
7:45–8:00	10.302	10.015	−2.9	7.087	−45.4	5.952	−73.1
8:00–8:15	9.625	9.299	−3.5	5.001	−92.5	5.77	−66.8
8:15–8:30	7.684	7.632	−0.7	6.813	−12.8	5.499	−39.7
8:30–8:45	8.514	7.666	−11.1	6.059	−40.5	6.195	−37.4
8:45–9:00	7.493	7.482	−0.1	5.881	−27.4	7.049	−6.3

2.3.1. LSTM Neural Network. The LSTM model is a typical gated recurrent neural network (RNN). The memory unit of LSTM introduces a gating mechanism to preserve long and short-term memory, in contrast to the standard RNN loop structure. Figure 6 depicts its unit architecture, which consists primarily of an input gate, a forgetting gate, and an output gate. The input gate regulates the entry and exit of new data, the forgetting gate regulates the degree to which historical data are forgotten, and the output gate defines the final output data.

The computation method is as follows, given the current input x_t, the hidden layer state h_{t-1}, and the storage state C_{t-1} at the final moment. The detailed calculation process is shown as follows:

$$
\begin{aligned}
i_t &= \sigma\left(W_i[x_t, h_{t-1}]^T + b_i\right), \\
f_t &= \sigma\left(W_f[x_t, h_{t-1}]^T + b_f\right), \\
o_t &= \sigma\left(W_o[x_t, h_{t-1}]^T + b_o\right), \\
\widetilde{C}_t &= \tanh\left(W_c[x_t, h_{t-1}]^T + b_c\right), \\
C_t &= C_{t-1} \otimes f_t + i_t \otimes \widetilde{C}_t, \\
h_t &= o_t \otimes \tanh(C_t).
\end{aligned}
\tag{1}
$$

W stands for the associated weight matrix, and b stands for the corresponding bias vector. The hidden layer state h_t is used to determine the final output.

$$
\widehat{y}_t = W_s h_t + b_s,
\tag{2}
$$

where W_s and h_t are the hidden layer and output layer connection weight matrix and bias vector, respectively.

2.3.2. Quantile Regression. The quantile regression model for the response variable Y, which is affected by K factors X_1, $X_1 \ldots X_K$, is calculated as

$$
Q_Y(\tau \mid X) = X'\beta(\tau).
\tag{3}
$$

The estimation of the parameter vector is one of them that can be translated into solving the optimal problem:

$$
\begin{aligned}
\min_{\beta(\tau)} & \sum_{t \mid Y_t \geq X'\beta(\tau)} \tau \left|Y_t - X'\beta(\tau)\right| \\
& + \sum_{t \mid Y_t < X'\beta(\tau)} (1-\tau)\left|Y_t - X'\beta(\tau)\right|.
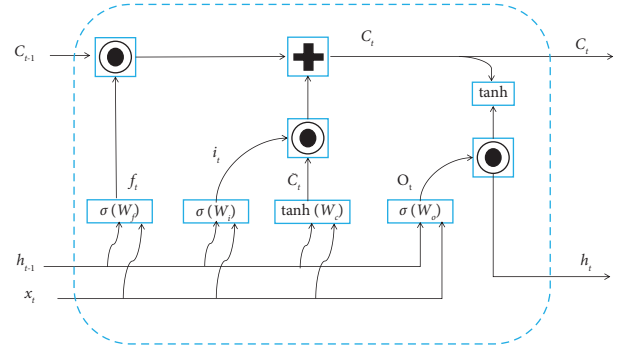\end{aligned}
\tag{4}
$$

FIGURE 6: Architecture of the LSTM unit.

2.3.3. Single-Point LSTM Quantile Regression Model. The value of the response variable Y at a quantile τ can be predicted using a single-point LSTM quantile regression model (Q-LSTM$_i$ model). Take any quantile τ_i. The loss function of the Q-LSTM$_i$ model under the quantile τ_i is as follows:

$$
\begin{aligned}
L(\tau_i) = & \min_{W(\tau_i), b(\tau_i)} \sum_{t=1}^{N} \rho_{\tau_i}(Y_t - f(X_t, W(\tau_i), b(\tau_i))) \\
& + \lambda_1 \sum W^2(\tau_i).
\end{aligned}
\tag{5}
$$

$W(\tau_i)$ and $b(\tau_i)$ are the weight parameter matrixes and the offset vectors under quantiles τ_i, respectively. λ_1 is the regular term penalty parameter, and $\rho_\tau(a)$ is the test function. Formula (6) is as follows:

$$
\rho_\tau(a) = \begin{cases} \tau a, & a \geq 0, \\ (\tau - 1)a, & a < 0. \end{cases}
\tag{6}
$$

The gradient descent method is used to update the parameters in order to obtain the optimal parameter $\{W(\tau_i), b(\tau_i)\}_{\text{opt}}$. The model's training process is shown in Figure 7.

2.3.4. Quantile Regression Model of LSTM considering Quantile Constraint. It is necessary to estimate the probability distribution of travel time through a succession of distinct quantiles $0 < \tau_1 < \tau_2 \ldots < \tau_i < \ldots < \tau_r < 1$ in order to gather more thorough probability distribution information. As a result, this paper established a constrained LSTM quantile regression model (CQ-LSTM model) using Q-LSTM$_i$ $(i = 1, 2 \ldots r)$ units. The CQ-LSTM model can
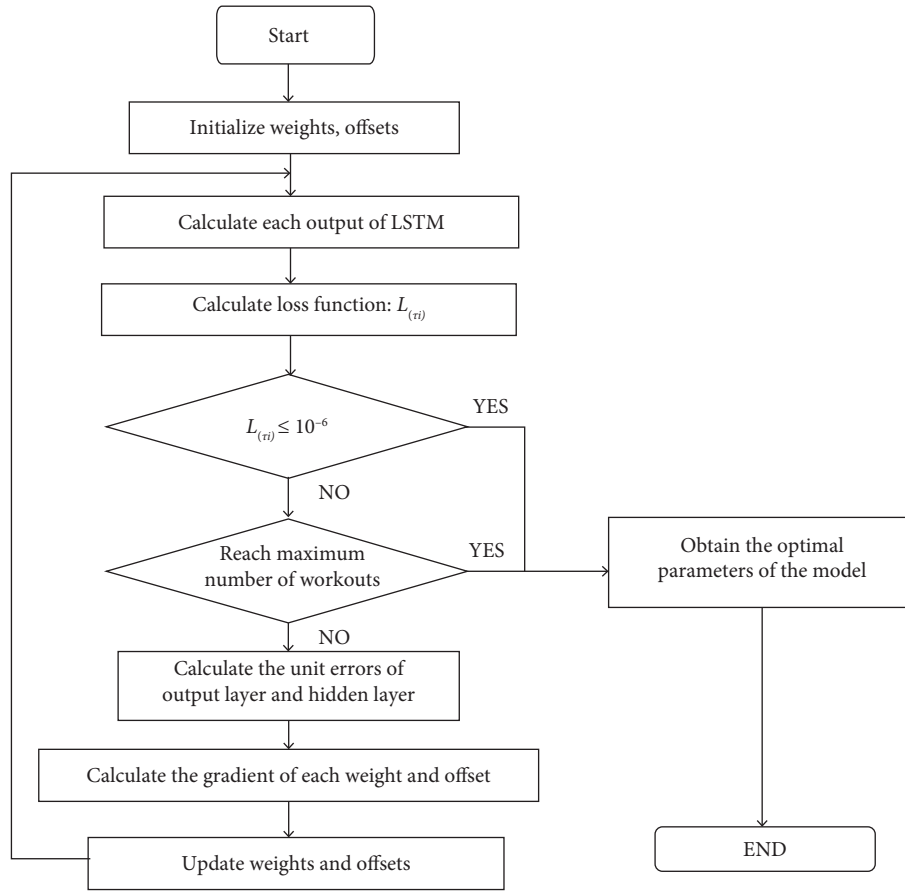
FIGURE 7: The process of Q-LSTM model training.

accurately estimate the distribution of travel time over a given time interval. The CQ-LSTM model is shown in Figure 8.

The loss function of LSTM multiple quantiles is constructed using formula (5), and the quantile loss function under R quantiles is combined. The quantile, on the other hand, has its inherent characteristics. The conditional quantile of Y_t should satisfy

$$Q_{Y_t}(\tau_1 \mid X_t) \le Q_{Y_t}(\tau_2 \mid X_t) \le \ldots \ldots \le Q_{Y_t}(\tau_r \mid X_t). \quad (7)$$

The combination layer of penalty terms violating the quantile predicted value constraint relationship is added in this study to ensure that the model avoids intersecting neighboring quantiles as much as possible. Formula (8) is about the function of the CQ-LSTM model:

$$
\begin{aligned}
F_{\text{loss}} = \min_{W(\tau_i), b(\tau_i)} & \sum_{i=1}^{r} \sum_{t=1}^{N} \left\{ \rho_{\tau_i}\left(Y_t - Q_{Y_t}(\tau_i \mid X_t)\right) + \lambda_1 W^2(\tau_i) \right\} \\
& + \lambda_2 \sum_{i=1}^{r-1} \sum_{t=1}^{N} \zeta\left(Q_{Y_t}(\tau_i \mid X_t) - Q_{Y_t}(\tau_{i+1} \mid X_t)\right).
\end{aligned} \quad (8)
$$

The penalty parameter is λ_2 in this case. The rest of the model is identical to the single-point LSTM quantile regression model. The constrained penalty function may successfully avoid the cross between quantiles when using the CQ-LSTM
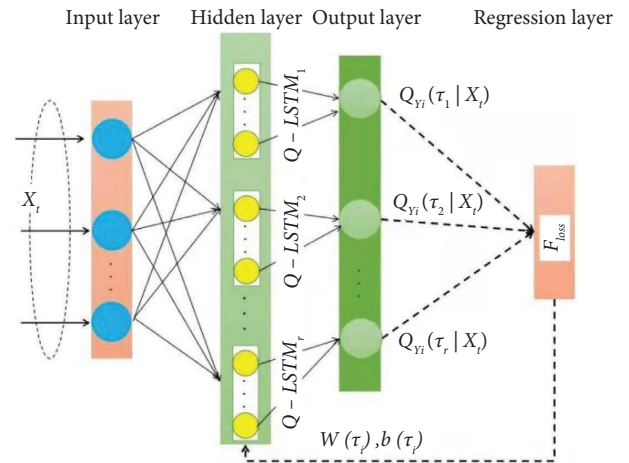

FIGURE 8: CQ-LSTM model.

model to determine the conditional quantile of trip times at several quantiles at the same time. Pretraining of the Q-LSTM model is also included before the overall training of the CQ-LSTM model. The initial parameter values of the CQ-LSTM model are $\{W(\tau_i), b(\tau_i)\}_{\text{opt}}$ obtained by training each Q-LSTM$_i$, which can improve the model training efficiency. The CQ-LSTM model's training and prediction procedure is shown in Figure 9. The following are the specific steps:
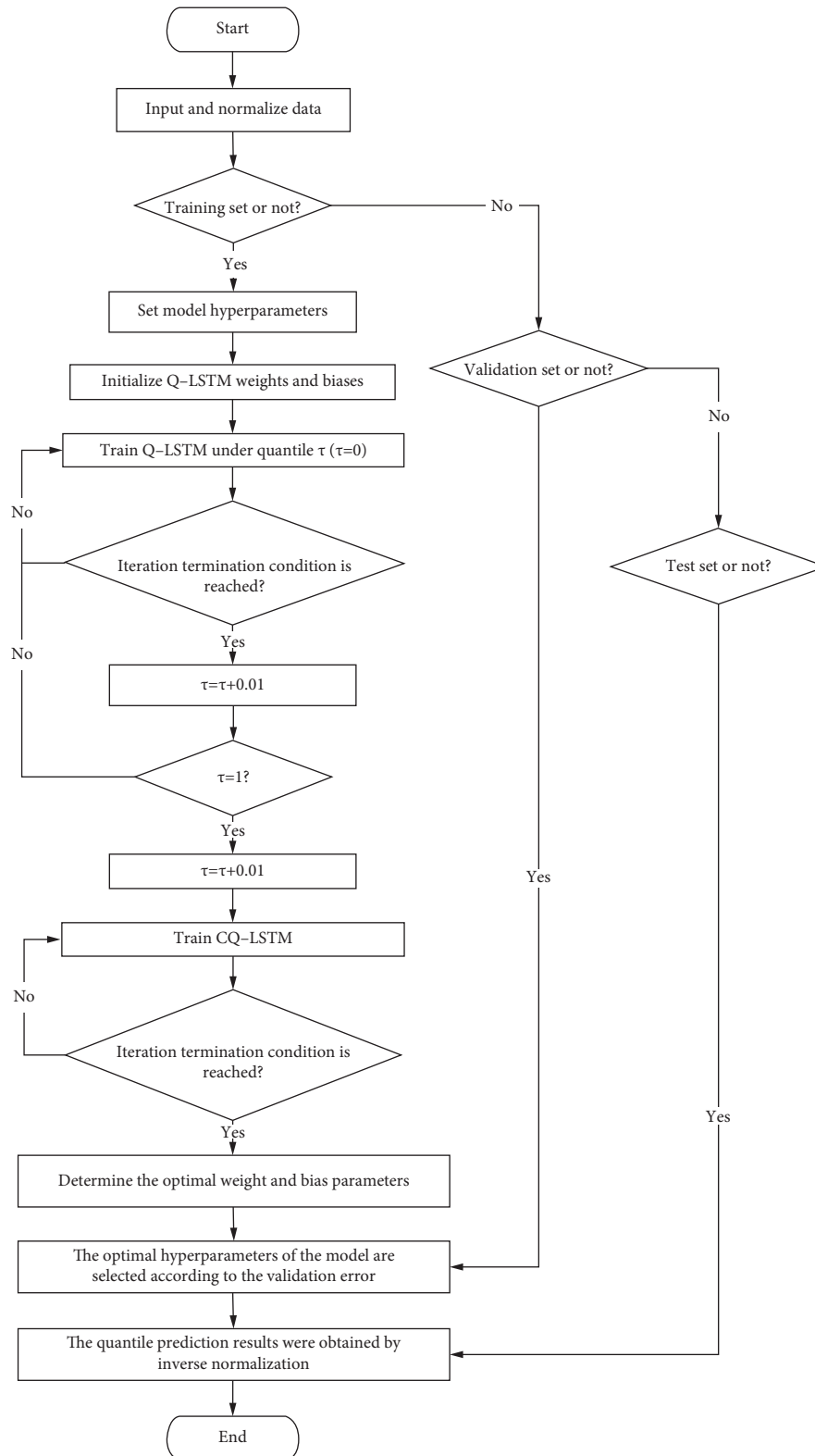
FIGURE 9: The process of CQ-LSTM model training.

(i) Travel time data and influencing factors for N sample days (mostly rainfall, snowfall, holiday data, etc.) are entered. The data to get $\{X_t, Y_t\}_N$ are normalized.

(ii) The dataset is divided into three sections: training set, test set, and validation set, and model hyperparameters like the neuron number M and penalty parameter λ_1 are set.

(iii) The model parameters of Q-LSTM$_i$ ($i = 1, 2 ..., r$) are randomly initialized, and Q-LSTM$_i$ is trained. The initial parameters of CQ-LSTM model training were set to $\{W(\tau_i), b(\tau_i)\}_{\text{opt}}$ acquired through training.

(iv) The CQ-LSTM model is trained, $\{W(\tau_i), b(\tau_i)\}_r$ is fine-tuned, and the optimal weight and bias parameters are established.

(v) The validation set sample is input into the trained CQ-LSTM model, and the best hyperparameter of the model based on the validation error is selected.

(vi) The best hyperparameters were used to input test samples into the CQ-LSTM prediction model with t, and the output prediction results were reverse normalized to obtain r quantiles of travel time at each time on the anticipated day.

(vii) The empirical accumulative distribution function is calculated based on the aforementioned results, the travel time probability prediction result is compared to the actual travel time value, and the model is reviewed and analyzed.

2.3.5. Model Evaluation Index. The paper proposes an evaluation index that takes quantile constraints into account. In addition, one of the evaluation indices of the model in this study is the quantile score (QS), which is a typical index for evaluating probability prediction results:

(i) The quantity score (QS) is

$$P_{t,\tau_i}(\tau, \widehat{y}_t(\tau_i)) \&9 := \max((\tau_i - 1)(y_t - \widehat{y}_t(\tau_i)),$$
$$\&9; \quad \tau_i(y_t - \widehat{y}_t(\tau_i))), \tag{9}$$

$$X_{\text{QS}} = \frac{1}{r \times N} \sum_{i=1}^{r} \sum_{t=1}^{N} P_{t,\tau_i}(\tau, \widehat{y}_t(\tau_i)). \tag{10}$$

Here, P_{t,τ_i} is the pinball loss at the τ_i quantile, $\widehat{y}_t(\tau_i)$ is the predicted value at the time t and quantile τ_i, and N is the number of all test moments.

(ii) Constraint Score (CS)

This index can consider the constraint relationship between quantiles and is the normalization coefficient of the quantile constraint error squared. As can be seen from equation (7), quantiles have their inherent characteristics. Therefore, the paper needs to consider the relationship between these quantiles and propose the following indicators:

$$v_{t,i} = \begin{cases} 0, & \widehat{y}_t(\tau_i) \le \widehat{y}_t(\tau_{i+1}), \\ \widehat{y}_t(\tau_i) - \widehat{y}_t(\tau_{i+1}), & \widehat{y}_t(\tau_i) \ge \widehat{y}_t(\tau_{i+1}), \end{cases}$$
$$X_{\text{cs}} = \sqrt{\frac{2\theta}{N} \sum_{t=1}^{N} \sum_{i=1}^{r-1} v_{t,i}^2}. \tag{11}$$

$v_{t,i}$ is 0 when adjacent quantiles meet the relationship; otherwise, it is the different between adjacent quantiles, reflecting the degree of constraint violation. θ is the step size

between adjacent quantiles. When X_{QS} and X_{CS} are both low, it indicates that the predicted quantile is performing well. In this paper, the model evaluation indexes X_{QS} and X_{CS} are applied.

3. Experiments

The section discusses the test results of the proposed CQ-LSTM model. The comparisons between CQ-LSTM with other models are also performed.

To perform the probability prediction model research on trip times, the floating car dataset for Harbin city from June 1, 2016, to December 31, 2016, was used. The data were collected as training samples from June 1, 2016, to November 30, 2016, with each sample interval of 15 minutes and 192 samples per day. There were 35,136 samples in the training set. Floating automobile data from December 1 to 15 are included in the test set. There were 2880 samples in the test set. The training set covers summer time and winter time. Adding summer time to the training data is to increase the size of the training data. Deep neural networks need to use a large amount of data to complete training. In addition, the model constructed in this paper only predicts the next 15 minutes of travel time, and seasonal variation of travel time is not taken into consideration. This article iterates 100 epochs of LSTM under each quantile using the Keras deep-learning framework, and the LSTM structure is a 64-gate structure. Data should be standardized before training using equation (9), which can enhance training efficiency:

$$X_t = \frac{X_t - X_{\min}}{X_{\max} - X_{\min}}, \tag{12}$$

where X_t is the sample vector at time t and $X_{\max}$ and $X_{\min}$ denote the maximum and minimum values, respectively.

According to the provided information, the CQ-LSTM model was used to forecast 2880 intervals between December 1 and 15, 2016. With a 0.01 gap, the prediction varied from 0.01 to 0.99. Figure 10 depicts the model loss. The model has a decent effect, as shown in Figure 10. The loss of both the training and test sets converges to a low level after 100 epochs. The test set's loss converges to 0.0336. The expected values of distinct quantile travel rates may be acquired using the model's prediction findings, and empirical cumulative distribution curves of predicted points can be constructed using the model's prediction results. From the prediction points 0 to 2880, four intervals (248, 640, 1656, and 2656) are chosen at random. The cumulative distribution curves of the four intervals are shown in Figure 11. The blue curve represents the actual cumulative travel rate distribution curve for the specified time slot, while the red curve represents the CQ-LSTM model's anticipated curve. The model can completely predict the cumulative distribution of journey time under the 15-minute time slot, as shown by the anticipated cumulative distribution curve, and the predicted value is quite comparable to the true value. It clearly demonstrates that the CQ-LSTM model developed in this paper can accurately estimate trip times at different quantiles and forecast travel time fluctuations.

In order to further determine the accuracy of the CQ-LSTM regression prediction model, this paper selects the same dataset and adopts the linear quantile regression
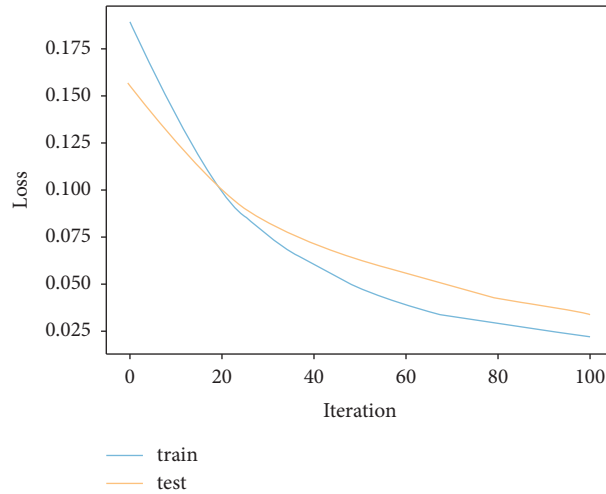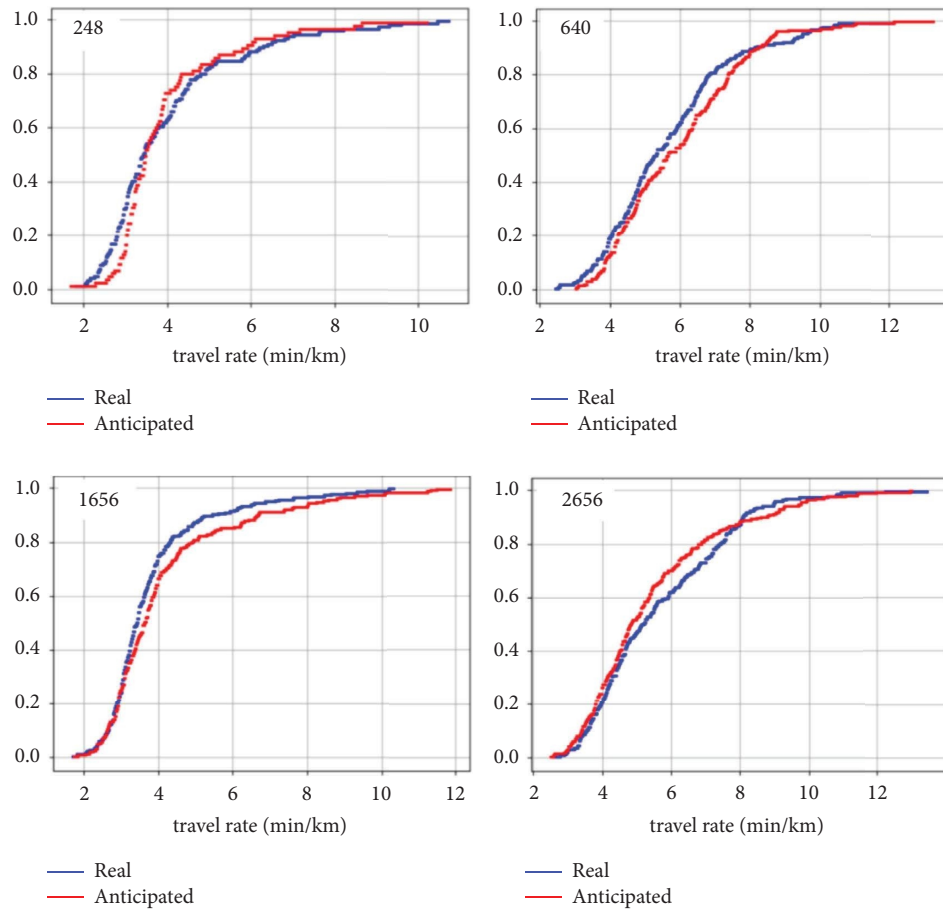
FIGURE 10: CQ-LSTM model loss diagram.

FIGURE 11: Cumulative distribution function of travel rate prediction experience at four time points.

model (L-QR) and the neural network quantile regression (QRNN) model for training prediction. The QRNN model selects the number of hidden layer nodes as 64 and the penalty parameter as 1. The L-QR model, QRNN model, Q-LSTM model, and CQ-LSTM model are compared. The statistical pairs of the predicted results of each model are shown in Table 4. X_{QS} and X_{CS} of all sample days in the test set are shown in Figures 12 and 13. Comparing the prediction results of different models, X_{QS} of the Q-LSTM model and the CQ-LSTM model is much lower than that of the QRNN model and the L-QR model, indicating that the quantile regression model based on LSTM has better prediction performance than the QRNN model. Although the X_{QS} index of CQ-LSTM is slightly higher than that of

TABLE 4: Comparison of model evaluation indexes.

Models	X_{QS} (min/km)	X_{CS} (min/km)
L-QR	4.731	3.67
QRNN	1.713	0.792
QR-LSTM	0.354	0.861
CQR-LSTM	0.479	0.135
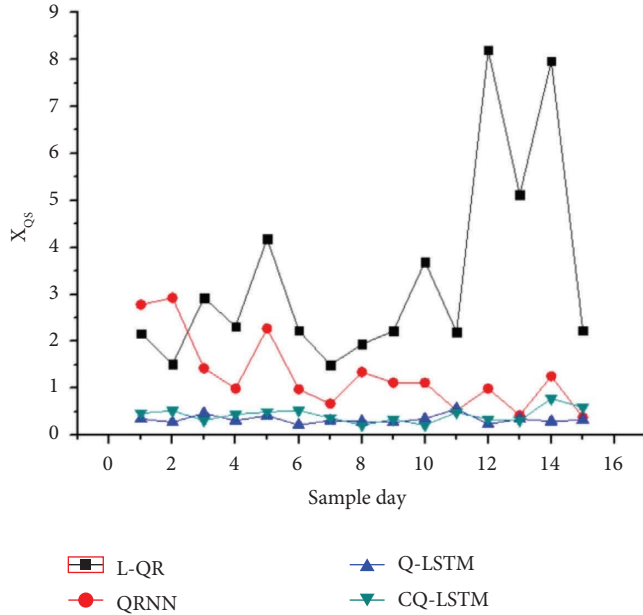
FIGURE 12: X_{QS} for each sample day of the test set.
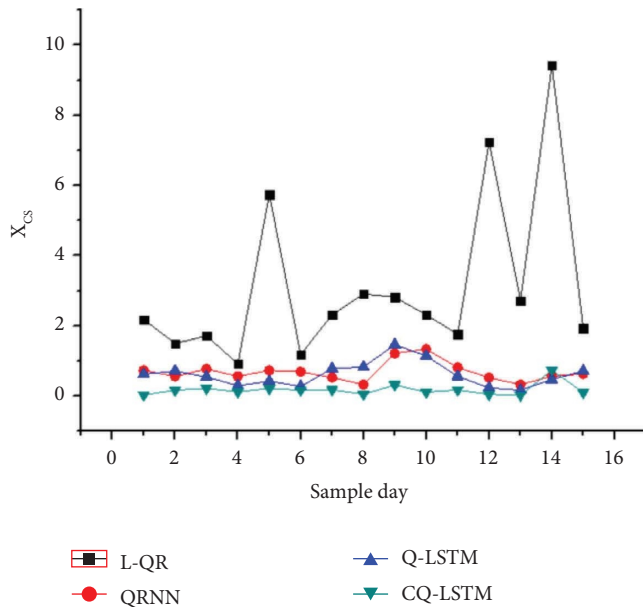
FIGURE 13: X_{CS} for each sample day of the test set.

Q-LSTM on some sample days, the X_{CS} index of CQ-LSTM is lower than that of Q-LSTM. The CQ-LSTM model can effectively avoid the crossover of quantiles and improve the rationality of prediction quantiles on the premise of basically not reducing prediction accuracy.

To demonstrate the application potential of the CQ-LSTM method proposed in this paper, the CQ-LSTM model is compared with the gated spatiotemporal attention model (GSTA) proposed by Khaled and Alfateh [27]. The GSTA method predicts based on the temporal correlation and spatial correlation of travel time and is able to predict the average travel time of the target road segment over a future period. CQ-GSTA can only predict the average travel time and cannot provide travel time reliability. CQ-LSTM does not consider the spatial correlation of travel time but can provide travel time reliability. The advantage of GSTA is that it provides travel time reliability for travelers and enables travelers to judge the probability of reaching their destination on time.

Studies [10] have shown that cabs follow the same traffic rules as private cars, and using cabs to represent private cars does not introduce additional errors. The driving pattern of buses is different from that of private cars, but buses account for a smaller share of the overall traffic volume and have a smaller impact on the prediction results.

4. Conclusion

In order to study the influence of different weather on the reliability of urban road trip times at different times, this paper uses Harbin road network data and taxi data to calculate travel time and combines Harbin weather data to conduct an empirical study. In this paper, the log-normal distribution can better fit the trip rate, and travel time variability (TTV) is selected as the index to measure the reliability of travel time. The influence of the weather on the travel time on weekdays and nonweekdays is studied, respectively, and the quantitative analysis is made on the peak hours alone. In addition, this study also proposes a quantile regression model of LSTM considering quantile constraints. It can well predict the travel rate under different quantiles, and the model does better in prediction accuracy and reduces the quantile constraints compared with other models such as QRNN. Based on these studies, the empirical cumulative distribution of the travel rate can be predicted. The main findings of this paper can be summarized as follows:

(1) The travel rate of vehicles on urban roads in Harbin can be well fitted by the log-normal distribution. The travel rate distribution shows an obvious "heavy tail" trend.

(2) Different weather has different effects on the reliability of travel time in Harbin at different times. Overall, except for light rain, other weather has increased the reliability of travel time in early peak hours, among which moderate rain has the largest increase, up to 97.4%.

(3) The quantile regression model based on constrained LSTM well predicts the travel rate of each quantile at a specific time and then forms an empirical cumulative distribution curve. Compared with the QRNN model, the prediction accuracy is greatly improved and the degree of violating quantile constraints is greatly reduced.

For future work, based on the study of the impact of weather on travel time reliability, it is necessary to consider the impact of other nonrepetitive variables on travel time reliability, such as traffic accidents. In addition, the work of predicting the travel rate of different quantiles can be more studied in buses. This can provide travel basis for different types of travelers.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] M. A. Abdelaty, R. Kitamura, and P. P. Jovanis, "Investigating effect of travel time variability on route choice using repeated-measurement stated preference data," *Transportation Research Record Journal of the Transportation Research Board*, vol. 1493, no. 1493, pp. 39–45, 1995.

[2] S. Yang and S. Qian, "Understanding and predicting travel time with spatio-temporal features of network traffic flow, weather and incidents," *IEEE Intelligent Transportation Systems Magazine*, vol. 11, pp. 12–28, 2019.

[3] H. Caceres, H. Hwang, and Q. He, "Estimating freeway route travel time distributions with consideration to time-of-day, inclement weather, and traffic incidents," *Journal of Advanced Transportation*, vol. 50, no. 6, pp. 967–987, 2016.

[4] Z. Li, L. Elefteriadou, and A. Kondyli, "Quantifying weather impacts on traffic operations for implementation into a travel time reliability model," *Transportation Letters*, vol. 8, no. 1, pp. 47–59, 2016.

[5] Y. Asakura and M. Kashiwadani, "Road network reliability caused by daily fluctuation of traffic flow," *Presented at the PTRC Summer Annual Meeting*, University of Sussex, London, UK, 1991.

[6] W. Pu and Wenjing, "Analytic relationships between travel time reliability measures," *Transportation Research Record Journal of the Transportation Research Board*, vol. 2254, no. 1, pp. 122–130, 2011.

[7] H. Tu, H. W. van Lint, and H. J. van Zuylen, "Impact of Adverse Weather on Travel Time Variability of Freeway Corridors," in *Proceedings of the Transportation Research Board Meeting*, Washington, NJ, USA, January 2007.

[8] J. Van Lint and H. J. van Zuylen, "Monitoring and predicting freeway travel time reliability: using width and skew of day-to-day travel time distribution," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1917, no. 1, pp. 54–62, 2005.

[9] T. Lomax, D. Schrank, S. Turner, and R. J. Margiotta, *Report of Selecting Travel Reliability Measures*, Federal Highway Administration, Washington, DC, USA, 2003.

[10] P. Chen, R. Tong, G. Lu, and Y. Wang, "Exploring travel time distribution and variability patterns using probe vehicle data: case study in Beijing," *Journal of Advanced Transportation*, vol. 2018, Article ID 3747632, 13 pages, 2018.

[11] N. E. Faouzi, R. Billot, and S. Bouzebda, "Motorway travel time prediction based on toll data and weather effect integration," *IET Intelligent Transport Systems*, vol. 4, no. 4, pp. 338–345, 2010.

[12] H. Li, Q. Wang, and W. Xiong, "New model of travel-time prediction considering weather conditions: case study of urban expressway," *Journal of Transportation Engineering, Part A: Systems*, vol. 147, no. 3, Article ID 04020161, 2021.

[13] O. Andersen and K. Torp, "A data model for determining weather's impact on travel time," in *Proceedings of the Database and Expert Systems Applications: 27th International Conference*, pp. 437–444, Springer, Porto, Portugal, September, 2016.

[14] I. Tsapakis, T. Cheng, and A. Bolbol, "Impact of weather conditions on macroscopic urban travel times," *Journal of Transport Geography*, vol. 28, pp. 204–211, 2013.

[15] Y. Zou, T. Zhu, Y. Xie, L. Li, and Y. Chen, "Examining the impact of adverse weather on travel time reliability of urban corridors in Shanghai," *Journal of Advanced Transportation*, vol. 2020, Article ID 8860277, 11 pages, 2020.

[16] Y. Wang, Y. Bie, and Q. An, "Impacts of winter weather on bus travel time in cold regions: case study of Harbin, China," *Journal of Transportation Engineering, Part A: Systems*, vol. 144, no. 11, Article ID 05018001, 2018.

[17] C. Kamga and M. A. Yazıcı, "Temporal and weather related variation patterns of urban travel time: considerations and caveats for value of travel time, value of variability, and mode choice studies," *Transportation Research Part C: Emerging Technologies*, vol. 45, pp. 4–16, 2014.

[18] L. J. Shi, T. Jing, X. H. Chen, and D. X. Ou, "The effects of rainfalls on expressway travel time," *Applied Mechanics and Materials*, vol. 361, pp. 2255–2261, 2013.

[19] J. Kim and H. S. Mahmassani, "Compound Gamma representation for modeling travel time variability in a traffic network," *Transportation Research Part B: Methodological*, vol. 80, pp. 40–63, 2015.

[20] E. Castillo, M. Nogal, J. M. Menéndez, S. Sanchez-Cambronero, and P. Jimenez, "Stochastic demand dynamic traffic models using generalized beta-Gaussian Bayesian networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 565–581, 2012.

[21] H. Al-Deek and E. B. Emam, "New methodology for estimating reliability in transportation networks with degraded link capacities," *Journal of intelligent transportation systems*, vol. 10, no. 3, pp. 117–129, 2006.

[22] F. Yang, M. P. Yun, and X.-G. Yang, "Travel time distribution under interrupted flow and application to travel time reliability," *Transportation Research Record Journal of the Transportation Research Board*, vol. 2466, no. 1, pp. 114–124, 2014.

[23] M. Ansari Esfeh, L. Kattan, W. H. Lam, R. Ansari Esfe, and M. Salari, "Compound generalized extreme value distribution for modeling the effects of monthly and seasonal variation on the extreme travel delays for vulnerability analysis of road

network," *Transportation Research Part C: Emerging Technologies*, vol. 120, Article ID 102808, 2020.

[24] X. Zhang and M. Chen, "Quantifying the impact of weather events on travel time and reliability," *Journal of Advanced Transportation*, vol. 2019, Article ID 8203081, 9 pages, 2019.

[25] B. L. Smith and J. M. Ulmer, "Freeway traffic flow rate measurement: investigation into impact of measurement time interval," *Journal of Transportation Engineering*, vol. 129, no. 3, pp. 223–229, 2003.

[26] R. Zhang, Y. Shu, Z. Yang, P. Cheng, and J. Chen, "Hybrid traffic speed modeling and prediction using real-world data," in *Proceedings of the 2015 IEEE International Congress on Big Data*, pp. 230–237, IEEE, New York, NY, USA, August 2015.

[27] A. Khaled and M. Alfateh, "Tag Elsir, and Yanming Shen. "GSTA: gated spatial–temporal attention approach for travel time prediction," *Neural Computing and Applications*, vol. 34, pp. 1–16, 2022.