IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 13 June 2023, accepted 14 July 2023, date of publication 20 July 2023, date of current version 26 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3297143

RESEARCH ARTICLE

Spatial-Temporal Correlation Neural Network for Long Short-Term Demand Forecasting During COVID-19

XIAOCHUAN GUO, WENBO XIE, AND XIN LI
Gaode Maps, Alibaba Group, Beijing 100012, China
Corresponding author: Wenbo Xie (daoqian.xwb@alibaba-inc.com)

ABSTRACT Demand forecasting is an important method for dealing with the supply-demand relationship in social resource management. The demands of daily life discussed in this study are mainly about hotels, restaurants, gas stations, drugstores, shopping malls, etc. Accurate demand forecasting can help enterprises meet people's needs properly, especially since COVID-19 has spread worldwide and significantly changed people's lives. Previous studies have proven the effectiveness of statistical and deep learning models in demand forecasting, including the Autoregressive Integrated Moving Average (ARIMA), Long Short-Term Memory (LSTM), and their variants. These methods focus only on one-step prediction of one demand and do not perform well in long-term or multi-step predictions. To address this issue, we propose a framework called the Deep Spatial-Temporal Neural Network (DSTNN) for multi-demand forecasting, which performs well in long short-term forecasting. The DSTNN adopts a transformer encoder-based structure and establishes a direct correlation between the demand distribution and spatiotemporal features. The transformer encoder is designed to learn high-order feature interactions. An auxiliary task of pandemic trend forecasting was used to help the main task learn about pandemic trends. As regional mobility affects the spread of COVID-19, a population mobility graph was designed to achieve a better spatial feature representation. The DSTNN eliminates the dependency on historical sequential data, which helps the model to perform well in long-term forecasting, and its performance does not obviously decline as the prediction period increases. The experimental results show that the DSTNN performs much better than conventional models, particularly in terms of long-term forecasting.

INDEX TERMS Demand forecasting, daily life, spatial-temporal, time-series, transformer encoder, deep learning.

I. INTRODUCTION

Since December 2019, COVID-19 has spread worldwide and has had a great impact on daily life [48], [50], [51]. To meet regional needs better, an accurate method of demand forecasting is needed to help enterprises make proper decisions in dealing with the supply-demand relationship. During the COVID-19 pandemic, personal demands were largely suppressed and more attention should be paid to regional group demands. Many existing studies have introduced methods to deal with demand forecasting, whereas existing methods are mainly time-series based models that depend heavily on historical sequential data and perform poorly in long-term and multi-step forecasting. In this study, we discuss a demand forecasting method that performs well in both short and long-term forecasting. We propose a method that can forecast multi-demand distributions in the long short-term by modeling the correlation between demands and spatiotemporal features directly while considering pandemic factors. To date, there are still some challenges in this task.

The associate editor coordinating the review of this manuscript and approving it for publication was Yiming Tang.

A. ACCURACY OF LONG-TERM DEMAND FORECASTING

Many previous studies have focused on short-term or one-step forecasting(with $x = x_0, x_1, x_2, \ldots, x_t$ as input and predicting x_{t+1}), whereas in our task, the ability to forecast long-term demand ($x_{t+30}, x_{t+31}, \ldots, x_{t+60}, \ldots, x_{t+n}$) is always needed. For example, we want to predict the demand distribution in December under different pandemic situations when we were still in September, and not only predict the demand distribution of x_{t+1}, but also x_{t+30}, x_{t+60} and even longer periods. Time-series models are good for short-term forecasting, but they do not perform well in long-term forecasting because of a lack of previous historical time-series data. Therefore, accurate long-term forecasting is a challenge.

B. COMPLEX SPATIOTEMPORAL CORRELATION

Many factors affect the demand distribution. Temporal factors include *Hour of Day, Day of Week, Month of Year, Holidays*. For spatial factors, each region is not independent, and they influence each other, although they are not adjacent. As previously mentioned, the proper modeling of spatiotemporal correlations is a significant factor in our task.

As shown in **Figure 1(a)**, a comparison of the demand distribution between *Sanya* and *Changchun* in January shows a significant difference. *Sanya* has a warm climate, whereas *Changchun* experiences cold weather in January. The cold weather affects the electric vehicles a lot which leads to the opposite demand distribution of charging stations and gas stations between these two cities. As one of the most popular tourist cities in January, the demand for tourist attractions and hotels in *Sanya* is expected to increase. **Figure 1(b)** shows that the demand for hospitals, kindergartens, and markets is stronger at approximately 7:00 a.m., and accommodations, massages, and barbeques are more attractive at night.

C. UNPREDICTABLE SOCIAL EVENTS (COVID-19 PANDEMIC)

In addition to spatiotemporal factors, social events are key factors. The mutation of COVID-19 is unpredictable and has a serious impact on daily life, which significantly changes demand distribution. COVID-19 significantly affects the economy, society and daily livelihoods [49], and presents significant challenges to the energy system [42], [43], [45], business [44], [46], [47], and online platforms. Many demands deviate from the normal situations. For example, *Shanghai* experienced a more serious COVID-19 pandemic in March 2022 than in January 2022. As shown in **Figure 1(c)**, the demand for entertainment activities will be restrained and livelihood demands will receive more attention during a serious pandemic. Therefore, modeling pandemic trends and encoding them using spatiotemporal features is another key factor.

As shown above, spatiotemporal and pandemic factors significantly affect the demand distribution, therefore an accurate forecast is meaningful. Many studies [1], [2], [3] have demonstrated the promising effect of using deep learning methods for demand forecasting and spatial-temporal forecasting.

Considering the previous limitations, we propose a method that directly models the demand distribution with spatiotemporal correlation and the pandemic situation. This method has the ability to forecast long short-term demand under different situations and helps enterprises make proper decisions. To demonstrate the simplicity and effectiveness of our framework, we conducted experiments using an industrial dataset, and our model outperformed conventional time-series models, particularly in terms of long-term forecasting. The main contributions of this study are as follows.

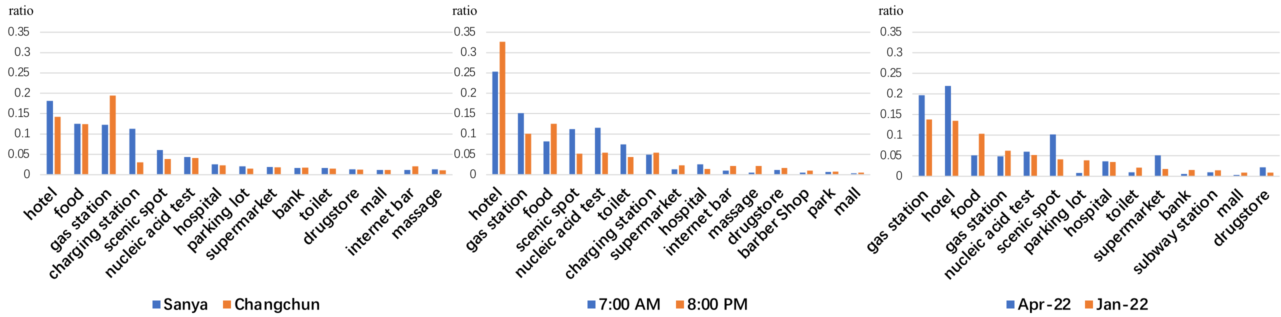
- We propose a novel framework for multi-demand distribution forecasting during the COVID-19 pandemic. This framework directly models the demand distribution with spatiotemporal and pandemic features, which eliminates the dependency on historical time-series data as input. This model has a strong ability for long-term prediction, and its performance does not decline obviously as the prediction period increases.
- We propose a new task for regional multi-demand distribution forecasting. This will help enterprises to make appropriate decisions in advance.
- We incorporate social events (the COVID-19 pandemic) into the model structure and build an auxiliary task to predict the trend of the pandemic, which helps the main task to have a better sense of pandemic influence.
- A population mobility graph network was built to learn regional representations, including those of non-adjacent regions.
- Modification of the standard KL divergence loss function enhances the prediction accuracy of the long-tail demands.

For this study, we collected demand distribution data from the Gaode Maps. This dataset consists of over 3 million samples with demand distribution labels. The experiments were designed and conducted carefully. The Deep Spatial-Temporal Neural Network(DSTNN) outperforms competitive models in demand distribution forecasting tasks, which demonstrates the effectiveness of our approach. We also apply this model to the Gaode Maps Recommendation system, which promotes Click-Through-Rate(CTR).

II. RELATED WORK

Demand forecasting is an important method for dealing with the supply-demand relationship. In this section, we discuss the methods, progress, and challenges of demand forecasting in different fields, including methods for modeling social events such as the COVID-19 pandemic.

In electric power and energy areas, the demand system deviated from the normal system during the COVID-19 pandemic. Reference [31] designed a multiple linear regression model to eliminate the impact of COVID-19 and applied a Long Short-Term Memory(LSTM) multiseasonal net deseasonalized (LSTM-MSNet-DS) model to improve the prediction with the residuals from the first regression method. Reference [32] built a time-series forecasting model to predict the half-hourly electricity demand. This model is used to distinguish between lockdown periods and



(a) Effects of spatial factors
(b) Effects of temporal factors
(c) Effects of pandemic factors

FIGURE 1. Demand distribution under different situations. (a) The demands distribution of Sanya and Changchun in January shows the effect of spatial factors. (b) Demand distribution differs at different hours of day (c) Shanghai suffers from a serious pandemic in April 2022, and demand is much different with it in January.

non-restrictive periods, and aims to identify a variety of patterns that we show to be influential on electricity demand. A study [33] investigated the impact of lockdowns on the air quality index (AQI). A time-series regression model is used to obtain the residual series. LSTM is used to achieve a multi-step prediction of the AQI. Reference [34] developed a data-driven analysis to substantiate the impact of the COVID-19 pandemic from the perspectives of power system security, electric power generation, electric power demand and electricity prices and comprehensively assessed the impact of COVID-19 on existing U.S. bulk power systems and markets. The Prophet-LSTM [37] model was used to forecast hospital Intensive Care Unit(ICU) demand during the COVID-19 pandemic. This study loosely couples Prophet and LSTM, two trending machine learning approaches, to incorporate ex-post and ex-ante variables. Experiments show that Prophet-LSTM performs better than both the standalone models. Another study [36] investigated the impact of COVID-19 on the efficiency of water demand prediction models by considering lockdown measures and various exogenous features. This study provides a forecasting methodology for monthly water consumption based on ensemble and error prediction models. The Autoregressive Integrated Moving Average(ARIMA) model, with additional regressors [38] was used to investigate the influence of the COVID-19 vaccine on transportation in New York City. A time-series analysis was performed on driving, transit, and walking data from the Apple Mobility Trends Reports. In [41], an optimal bayesian regularization algorithm(BRA) method incorporating pearson correlation coefficient(PCC) was proposed for electrical load forecasting(LF) at different time horizons. According to its model structure, the calendar points, meteorological factors and sequential historical data are sent to the neural network to obtain a final prediction. Compared with our work, using historical demand data as input features makes it difficult to perform long-term forecasting, because the long-term forecasting has no recent historical demand data.

In the supply chain [21] area, existing methods [4] use LSTM, which uses the gates in each memory cell. Data can be deleted, filtered, or added to the next cell in the modeling time series prediction. Their results indicated that LSTM is a competitive method that effectively models the non-linearity of time series and outperforms the linear forecasting method.

For the recommendation area, some studies used an LSTM model combined with user historical behavior to predict the next point of interest(POI) [5], [6] which uses a modified LSTM cell unit to model user check-in behaviors for the next POI recommendation. Other studies have proven the effectiveness of using deep learning methods to predict user interests in recommendation systems.

In the field of transportation and travel, studies [7], [8] have used spatiotemporal deep learning for ride-hailing system supply-demand gap forecasting, which integrates a feature importance layer with a spatiotemporal deep learning architecture. A One-dimensional convolutional neural network (CNN) is used to detect spatial dependencies and a zone-distributed independent recurrent neural network (IndRNN) is used to detect temporal dependencies from time-series data, which also achieves better performance than the conventional time-series model. However, it still cannot eliminate time-series data with IndRNN. In addition, an Irregular convolutional Long-Short Term Memory model(IrConv+LSTM) [40] was designed to improve short-term bike sharing demand forecasting. Reference [35] discusses collecting and preparing data on air passenger traffic worldwide with the scope of analyzing the impact of travel bans on the aviation sector during the COVID-19 coronavirus outbreak. They hope that these preliminary results may be of help for informed policy making design of exit strategies from this global crisis. This study also presented the first assessment of recent changes in flight activity worldwide as a result of the COVID-19 pandemic.

Many spatiotemporal modeling methods have been designed for traffic flow forecasting. Spatio-Temporal Graph Convolutional Networks(STGCN) [27] have been proposed to tackle the time-series prediction problem in the traffic domain. Graph convolutional networks were used to extract spatial features and a gated CNN was used to extract temporal features from graph-structured time-series data. Attention based Spatial-temporal Graph Convolutional Networks(ASTGCN) [28] have developed a spatial-temporal attention mechanism to learn the dynamic spatial-temporal correlations of traffic data and use recent, daily periodic and weekly periodic dependencies of historical data for prediction. Spatial-temporal synchronous graph convolutional networks(STSGCN) [29] proposed a novel spatial-temporal graph convolutional module to synchronously capture localized spatiotemporal correlations directly, instead of using different types of neural network modules separately. The Spatio-Temporal Mobility Demand Forecaster(ST-MDF) [39] was designed to forecast mobility demand. It is based on LSTM layers, which are used to seize both the spatial and temporal aspects. In addition, the model incorporates two additional modules that extract features from weather data and temporal information.

A deep hybrid neural network [30] was used for COVID-19 hotspots prediction, which proposed a hybrid framework in which the CNN network extracts multiple time-scale features from the input data, and the multi time scale features extracted from two blocks are then concatenated and provided as input to the LSTM model. The LSTM model identifies short, medium and long-term dependencies by learning the representation of time-series data. Specifically, the prediction model considers the daily number of positive cases from the last ten days and predicts the number for the next day. In this study, the value of the predicting horizon was fixed at 10 days, and the data points were sampled for 10 days, and then the value was predicted on the 11th day. This model deals with long short-term dependencies, but predicts only the next day, which does not make long-term predictions, such as 30 days later or further.

Most studies have focused on one-step prediction, including the aforementioned application areas. In this study, a multi-step prediction is desired. An intuitive method to address this issue is to recursively reuse recent predictions as inputs in the subsequent prediction step. Reference [18] introduces an Ouroboros Training Scheme that optimizes the training process, and Double Spatio-Temporal Neural Network(D-STN) blocks enable accurate forecasting for extended periods of time (up to 10 hours). In study [20], a special transformer block was designed to capture long-range temporal dependencies using longer historical time-series data to achieve better performance in long-term forecasting. References [18] and [22] still use historical dependencies to recursively predict multi-step and do not solve the prediction error accumulation. Spatio-temporal graph neural networks [19] were used to forecast COVID-19, which build a spatial graph with region-level human mobility for each day and connections between graph slices to represent temporal relationships.

In conclusion, almost all existing methods treat demand forecasting as a time-series forecasting problem that uses a sequence model and takes recent historical sequential data as input features. Some traditional statistical models such as ARIMA [9], [10], and other deep learning models [11], such as LSTM, and its variants have always been used to model this task, including spatial-temporal graph neural networks [19]. These models focus on historical sequential data which always lack modeling with spatiotemporal correlation and social events directly, which may limit the capacity of long-term forecasting. As an emergency plan for public events, where input features are always uncertain and discrete, the model should have the ability of long-term forecasting under hypothetical conditions, not just a one-step prediction for the next hour or day. In addition, it is difficult to obtain highly sensitive historical demand distribution data. In the model training process, historical demand data can be collected from the industrial environment. However, after model deployment, these data are difficult to obtain. Furthermore, some regions have not experienced a serious COVID-19 pandemic, and do not have historical data to learn from. Accordingly, an accurate long short-term spatiotemporal multi-demand forecasting model that is not dependent on time-series data is necessary.

III. METHOD

In this section, we introduce our approach, including the data process and base model structure. Next, we present a comprehensive description of Deep Spatial-Temporal Neural Network(DSTNN).

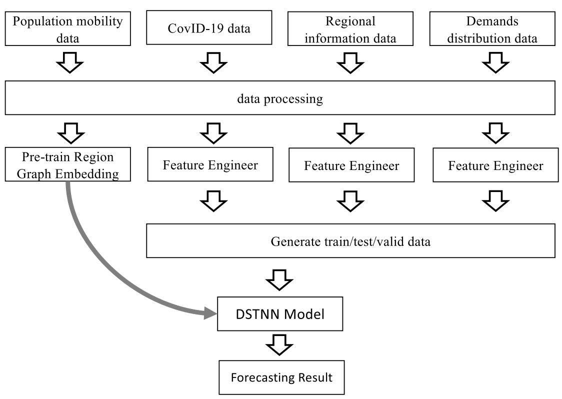


FIGURE 2. Pipeline of demands distribution forecasting.

As shown in Figure 2, the workflow of our task includes the following steps: data extraction, data preprocessing, feature engineering, regional embedding pre-training, and model training. After the entire pipeline was setup, adequate experiments were conducted to achieve the best performance. The experiments mainly consisted of several important parts, including finding the best feature set and proper model parameters, including the layer number and hidden size of the model layer, loss function, and hyper-parameters. A greedy strategy was used for these series of experiments, which means that only one variable parameter was conducted at a time and the other was fixed. After a series of experiments, the model structure with the best performance was found, and subsequent experiments were based on this model structure.

A. DATASETS

The demands that we are talking about are always consumer demands in our daily lives. Online maps were used to search hotels, restaurants, gas stations, drugstores, shopping malls, vehicle maintenance, subways, supermarkets, massage, nucleic acid tests, and charging stations, etc. We call all of these actions our demands in this study.

A regional multi-demand distribution forecasting task was proposed in this study. We used an industrial dataset as our experimental datasets because no public datasets match our study. The industrial datasets are available from the corresponding authors. Our datasets consisted of four files: (a) COVID-19 information data, (b) Regional information data, (c)population mobility data for each region, and (d) demand distribution under specific spatial temporal conditions. The original dataset is presented in **APPENDIX A**.

COVID-19 information includes the pandemic details of different regions. Each line includes the *region ID, datetime, confirm num, death num, heal num, and confirm num*.

The regional information includes basic regional information. Each line includes the *region ID, population, number of hospitals at different levels, and adjacent region ID set*.

Population mobility describes the population mobility from Region A to Region B in the previous year. Each line includes the *ID of Region A, ID of Region B, and mobility number*.

Demand distribution data are related to the learning goal of our task, which was collected from the production environment. Each line includes the following: *Date, Year, month, day of the week, hour, region ID, holiday, holiday type, and demand distribution*. The demand distribution fields are as follows [*gasstation*:0.243,*hotel*:0.126,*subwaystation*:0.012, *hospital*:0.04,. . . . .,*drugstore*:0.39].

B. DATA PREPROCESSING

Original data have both categorical and numerical features which will be treated differently.

1) CATEGORICAL FEATURES

Some attributes includes *Hour of Day, Day of Week, Month of Year, Region* where values are always discrete integer values or string values. Because these categorical features cannot be directly used in most machine learning algorithms, they must be transformed into numerical features. We need to transform the weekdays into a numerical ID set. For example, we design a dict={*Mon*, *Tues*, *Wed*, *Thur*, *Fri*, *Sat*, *Sun*} to transform weekdays into a numerical id number, which is encoded by the embedding layer into a dense vector.

2) NUMERICAL FEATURES

These features such as regional population, number of hospitals, COVID-19 newly confirmed number are both numerical. Features highly varying in magnitudes, units and range. Most machine learning algorithms use the Euclidean distance in their computation. If left alone, features with high magnitudes will affect the distance computation more than those with a low magnitude. To overcome this problem, all the features must be brought to the same magnitude. This can be achieved by feature scaling [12]. In this study, we use feature scaling of standardization (also called z-score normalization), which transforms features such that the distribution has a mean of 0 and a standard deviation of 1, σ is the variance, and $\overline{x}$ is the mean.

$$x' = \frac{\overline{x}}{\sigma} \tag{1}$$

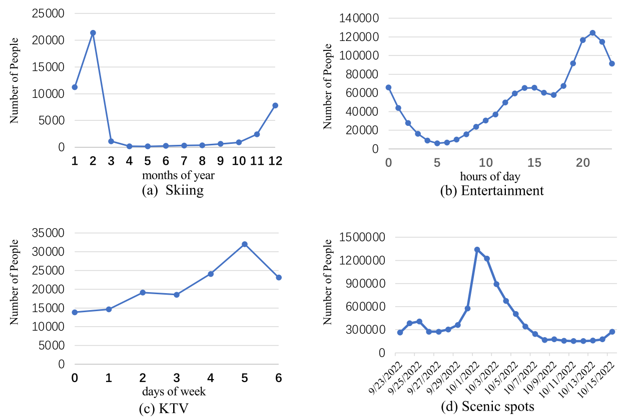


FIGURE 3. Change in demand under different temporal situations. (a) Skiing is a popular activity in winter (b) many people have entertainment after work time(c) KTV is much more like a weekend activity (d)October 1st is the National Day, many people travel around and visit scenic spots.

C. FEATURE ENGINEERING

The main goal of our study is multi-demand distribution forecasting under different spatiotemporal and pandemic situations. The original attributes of the dataset still have some limitations and can not be directly incorporated into our model. Excellent feature engineering is indispensable for improving the accuracy.

TABLE 1. Temporal features.

Feature Name	Feature Type	Description
Hour of Day	id feature	range [0,23]
Day of Week	id feature	range [1,7]
Month of Year	id feature	range [1,12]
Holiday/Workday	id feature	0/1
Holiday Type	id feature	-

Temporal features As shown in **Table 1**, including *Hour of Day, Day of Week, Month of Year*, as shown above demands

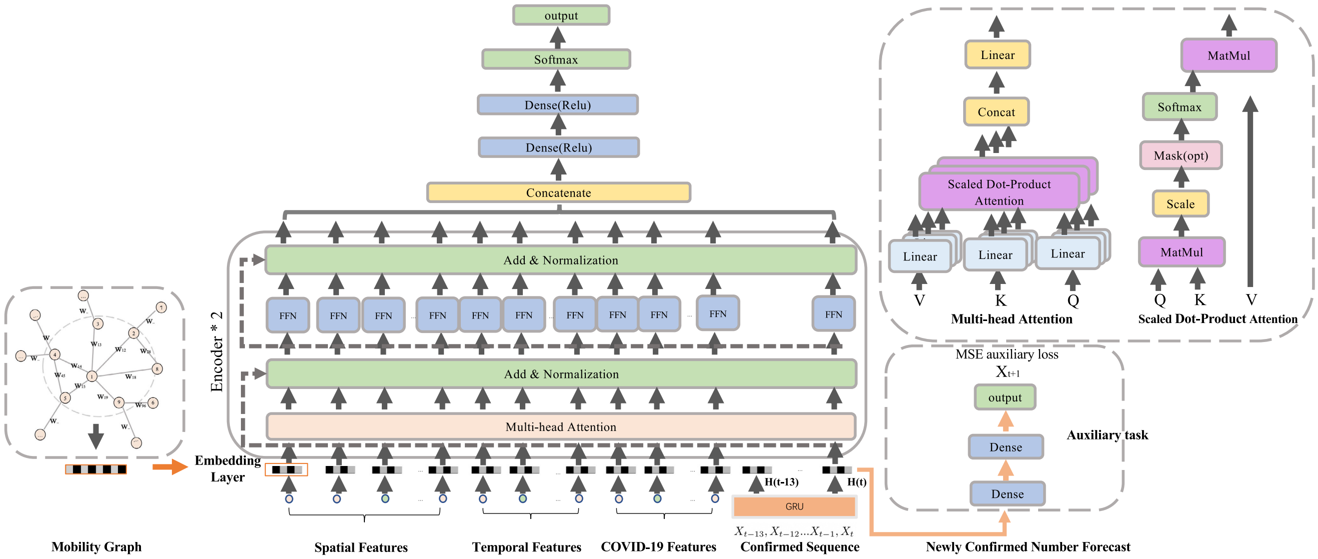


FIGURE 4. Detailed structure of DSTNN. The embedding layer transforms the spatial-temporal features and COVID-19 features into embedding vectors. A mobility graph is used to model spatial embedding, where each node represents a specific region and the weight of the edge represents the number of people's movements between two regions. An auxiliary task is built to forecast the pandemic trend, which ensures that the hidden state H(t) involves the trend of the COVID-19 pandemic. The backbone of the main task is a two-layer transformer encoder that can learn high-order feature interactions of the input features.

TABLE 2. Spatial features.

Feature Name	Feature Type	Description
region	id feature	-
population	numerical feature	-
nearby city	id feature	-
hospital number	numerical feature	-
hospitals of nearby city	numerical feature	-
population of nearby city	numerical feature	-

TABLE 3. Pandemic relative features.

Feature Name	Feature Type	Description
newly diagnosed	sequence feature	-
newly diagnosed diff	sequence feature	-
3 days avg diagnosed	numerical feature	-
7 days avg diagnosed	numerical feature	-
14 days avg diagnosed	numerical feature	-

vary according to temporal features. An obvious case indicates that people always engage in entertainment activities in the evening or weekends rather than weekday mornings, as shown in **Figure 3(c)**.

Spatial features As shown in **Table 2**, Region ID and some statistical features of this region also significantly impact demand. As shown in **Figure 1(a)**, the demand distribution in *Changchun* and *Sanya* were different. In addition to these original categories and statistical features, we also modeled the spatial correlation using the graph.

Social events (COVID-19) The features of social events are primarily related to the COVID-19 pandemic, which significantly affects people's lives. Therefore, it is necessary to design features that depict the severity of the COVID-19 pandemic. We used the number of confirmed cases in the current and past few days, as listed in **Table 3**.

D. DSTNN MODEL STRUCTURE

In contrast to previous related research, our task of regional demand distribution forecasting under a pandemic situation is not only used for one-step prediction, but also for a long period and unexperienced pandemic situations. We propose a model structure called the Deep Spatial-Temporal Demand Forecasting Network(DSTNN). The DSTNN was designed to forecast demand distribution by establishing a direct correlation between spatiotemporal features and demand distribution rather than using historical time-series data of the demand distribution. As shown in **Figure 4**, the DSTNN is composed of several core parts. The embedding layers project both categorical and numerical features into a dense vector. Population mobility graph using DeepWalk to achieve excellent regional representation. Multi-head attention was used to achieve high-order feature interactions. An auxiliary task was used to capture the pandemic trends. After the interaction of spatial, temporal and social event features with attention layers, high-order combinatorial features are concatenated and fed into a Multilayer Perceptron(MLP) for final prediction. The DSTNN has total 549,043 params. The core components of the DSTNN model are discussed in detail in this section.

1) EMBEDDING LAYER [13]

Because many features are categorical and id features that are very sparse and high-dimensional. High-dimensional sparse features will lead to an overfitting. To solve this problem, the embedding layer is a common method that enables the conversion of sparse category or ID features into a fixed-length dense vector. For the i-th feature group, we let

$$e^i = V_i x_i \tag{2}$$

V_i is embedding matrix of field i. We use lookup mechanism for embedding operation.

To achieve the interaction between categorical and numerical features, we should also convert numerical features to the same low-dimensional space, which is commonly used in Field Embedding, where different features share a uniform field embedding and then multiply this field embedding by their feature values:

$$e_m = v_m x_m \tag{3}$$

where v_m is an embedding vector for field m, and x_m is a scalar value.

2) MULTI-HEAD ATTENTION

The attention mechanism [14], [15] has been widely used in NLP areas which enables the dynamic highlighting of relevant features of the input data. In our study, scaled dot-product attention was used where the dot products are scaled down by $\sqrt{d_k}$, Formally we have a query Q, a key K and a value V and calculate the attention as

$$Attention(Q, K, V) = softmax(\frac{QK^t}{\sqrt{d_k}})V \tag{4}$$

Multi-head attention is a module for attention mechanisms that runs through an attention mechanism several times in parallel. The independent attention outputs are then concatenated and linearly transformed into the expected dimensions. Intuitively, multiple attention heads allow attending to different parts of the sequence (e.g., longer-term dependencies versus shorter-term dependencies).

$$MultiHead(Q, K, V) = [head_1, \ldots, head_h]W_0 \tag{5}$$

$$head_i = Attention(QW_i^q, KW_i^k, VW_i^v) \tag{6}$$

Above W are all trainable parameter matrices. In our model, the attention mechanism helps learn the weights of each feature field and achieve higher-order feature interactions.

In our task, attention mechanisms were used to learn high-order feature interactions of the input features. We define features i and feature j under a specific attention h as:

$$\alpha_{ij}^h = \frac{exp(Attention^h(e^i, e^j))}{\sum_{k=1}^m exp(Attention^h(e^i, e^k))} \tag{7}$$

$$Attention^h = \left\langle W_Q^h e^i, W_K^h e^j \right\rangle \tag{8}$$

$Attention^h$ is an attention definition, and we use the inner product for attention computation because of its simplicity and effectiveness.

$$\overline{e}_i^h = \sum_{k=1}^{m} \alpha_{ik}^h (W_V^h e_k) \tag{9}$$

$\overline{e}_i^h$ is a combination of the features of i and other relative features. $\overline{e}_i^h$ interact with each other to achieve a high-order combination. Different orders of feature combinations of input features can be modeled using different layers of multi-head self-attention neural networks.

3) MULTILAYER PERCEPTRON(MLP)

After the encoder blocks and concatenate all features, the concatenated vector is fed into the MLP for the final prediction.

4) SPATIAL CORRELATION MODELING

In this section, we describe how to encode different regions and model the relationship between the region and demand distribution using mobility graphs and other regional features. Each node in the graph represents a region and the edge represents the population mobility between the two regions. We chose the modified ***DeepWalk*** [22] method to learn the representations of each node, which is similar to the mobility of each region. The original DeepWalk algorithm comprises of two stages. The first step is a random walk generator that is used to uniformly sample a random vertex v_i as the root of the random walk sequence W_{v_i}. Then, samples are uniformly sampled from the neighbors of the previously sampled vertex until length is reached. In our task, we replaced the random walk with probability sampling, and the edge weight is used to sample, which perfectly simulated the flow of people. After sampling sufficient sequence W_{v_i} we use the ***skip-gram*** algorithm of word2vec [23] to obtain representations of the vertex. We believe that these node representations have spatial correlation and are not only correlated with spatial adjacency. We denote a random walk from vertex v_i and a stochastic walk step by weight w_i. Finally, the $v_1, v_2, \ldots, v_k$ sequences were constructed. We used skip-gram [24] to update these representations. More formally, given a sequence of nodes, the objective of the skip-gram model is to maximize the average log probability

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c<j<c}^{j\neq 0} log\ p(v_{t+j}|v_t) \tag{10}$$

where c is the window size of the training context. A larger window size resulted in higher accuracy.

5) TEMPORAL CORRELATION MODELING

Previous studies have always used historical observations to forecast the next time step with an RNN and its variants. To achieve long-term forecasting, we need to get ride of using historical sequential demand distribution data for demand forecasting and model temporal features with the demands distribution directly. As shown in **Figure 3**, demand distribution changes over time. Therefore, there must be a latent relationship between the temporal features and demand distribution. We designed temporal features such as *HourOfDay/Month/DayOfWeek* and interacted with spatial features and social event features. Finally, we establish a correlation for supervised learning.

6) MODELING WITH SOCIAL EVENTS(COVID-19)

a: AUXILIARY TASK

Previous studies on auxiliary tasks [16] and multi-task learning [25] have proved that in many application areas, designing a reasonable auxiliary task can help the main task achieve a better performance and generalization. In our study, COVID-19 is another important factor that affects demand distribution largely because the severity and trend of the pandemic will influence the scope of people's activities. Contingency plans should be implemented for every hypothetical pandemic. We used the average confirmed number of past few days (3,7,14) as the basic feature and used the sequential confirmed number and Gated Recurrent Unit(GRU) [26] to capture the pandemic trend. They work tremendously well on a large variety of problems, and are now widely used. The GRU formulations are as follows:

$$z_t = \sigma_g \left(W_z x_t + U_z h_{t-1} + b_z\right) \tag{11}$$

$$r_t = \sigma_g \left(W_r x_t + U_r h_{t-1} + b_r\right) \tag{12}$$

$$\widehat{h}_t = \phi_h \left(W_h x_t + U_h \left(r_r \odot h_{t-1}\right) + b_h\right) \tag{13}$$

$$h_t = z_t \odot \widehat{h}_t + (1 - z_t) \odot h_{t-1} \tag{14}$$

σ_g is the sigmoid function. ϕ_h is the hyperbolic tangent. x_t is the input of t step.However, the hidden state h_t only captures the dependency between the confirmed number of each day, which cannot represent the prediction of the next time-step. The ground truth only represents the demand distribution for final supervised learning. We use an auxiliary sequence modeling task to help model learn development trend of pandemic situation.Use newly confirmed number of $T + 1$ to supervise the learning of hidden state h_t. Compared with LSTM, GRU has fewer parameters, is computationally more efficient considering fewer parameters, and requires less data for generalization. Therefore, we used GRU to model the pandemic trend as an auxiliary task. With the help of an auxiliary task, the hidden stateh_t is sufficiently expressive to represent the pandemic state of $T + 1$.

b: MOBILITY GRAPH

In this graph, each node represents a specific region, and the weight of the edge represents the number of people moving between the two regions. Experiments show that the mobility of the two regions has a great impact on the spread of the pandemic, which will help the model learn about the impact of future pandemic.

c: MEDICAL RESOURCES

Regional medical resources, including the number of hospitals at different levels, will be obtained in spatial features, because excellent medical resources will help control the impact of the COVID-19 pandemic.

7) OUTPUT LAYER

After feature interaction, high-order combinatorial features were obtained and fed into the MLP layer. After the last MLP layer, we obtained logits, such as $z = [z_1, z_2, \ldots, z_k]$ where k is the number of demand categories. Finally, we applied a softmax projection $\sigma(z)$ to achieve the final demand distribution. $\sigma(z)$ is defined as follows:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{k=1}^{m} e^{z_i}} \tag{15}$$

E. LOSS FUNCTION

1) MAIN TASK

For demand distribution forecasting, Kullback–Leibler(KL) divergence is widely used to measure the divergence between continuous probability distributions. Definition of KL divergence between distributions A and B, where A is the ground truth demand distribution and B is the forecasting result.

$$\begin{aligned} D_{KL}(A||B) &= \sum_{i=1}^{n} p_A(v_i) log\, p_A(v_i) \\ &\quad - p_A(v_i) log\, p_B(v_i) \\ &= \sum_{i=1}^{n} p_A(v_i) log\, \frac{p_A(v_i)}{p_b(v_i)} \end{aligned} \tag{16}$$

$p_A(v_i)$ is the ground-truth demand probability of demand v_i with probability distributions A. $p_B(v_i)$ is the demand prediction probability of demand v_i with probability distribution B. The right-hand side of the above equation is the cross entropy in terms of distributions A and B, defined as:

$$H(A, B) = -\sum_{i=1}^{n} p_A(v_i) log\, p_B(v_i) \tag{17}$$

$$H(A, B) = D_{KL}(A||B) + S_A \tag{18}$$

and entropy definition of S_A is:

$$S_A = -\sum_{i=1}^{n} p_A(v_i) log\, p_A(v_i) \tag{19}$$

The long-tail probabilities cause the loss to be too small, which may lead to inaccurate forecasting of long-tail demands. Therefore, we need to enhance the long tail probability effect by adding a relative value, which is defined as L_{abs} :

$$L_{abs} = -\sum_{i=1}^{n}(p_A(v_i)log((\frac{p_A(v_i)}{p_B(v_i)} - 1)^2 + 1) + p_A(v_i)log((\frac{p_B(v_i)}{p_A(v_i)} - 1)^2 + 1)) \tag{20}$$

$$L_{main} = L_{KL} + \alpha L_{abs} \tag{21}$$

In conclusion, the D_{KL} describes the difference in the distribution of A from B and from the perspective of A. However, a drawback of KL-Divergence is that it is not sensitive to small $P_A(v_i)$ and $P_B(v_i)$ which may cause the model to not predict a small probability exactly. To solve this problem, we propose a loss function L_{abs} that enhances the model sensitivity for the small probability part, and α is a hyper-parameter to balance the weight of the long-tail probability

2) AUXILIARY TASK

The number of newly confirmed cases was used to measure the severity of the pandemic. Prediction for the confirmed number of $t + 1$ is a regression problem and therefore we choose MSE as the loss function.

$$L_{aux} = MSE = \frac{1}{N}\sum_{i=1}^{n}(f(x_i) - y_i)^2 \tag{22}$$

3) TOTAL LOSS

The global loss of our model is L, and L_{main} means the loss of demand forecast, and L_{aux} is the loss of the auxiliary task for pandemic trend forecasting. λ is a hyper-parameter that balances interest representation and pandemic development trends.

$$L = L_{main} + \lambda L_{aux} \tag{23}$$

F. OUTLIER DETECTION

The DSTNN has high robustness because it does not use historical sequential data as the input. Some discrete outliers do not significantly affect on the model. However, we still used statistical methods and context data to detect outliers, such as the demands at the current time being much different from other days around this day or the corresponding period of the previous year. Because outliers were detected, we checked some of them to ensure that our data processing flow worked well. The same method was applied in the prediction phase, and outliers are reported for further analysis.

IV. EXPERIMENTS

In this section, we compare our model with a conventional time-series model using an industrial dataset. In addition, ablation experiments were designed to prove the effectiveness of the model structure, including an auxiliary task and graph embedding. Finally, we share the techniques for online serving.

Approximately 3 million data points are collected from industrial environments, ranging from 1/1/2021 to 7/31/2022. Each instance includes spatial-temporal and pandemic features, including the demand distribution as the ground truth.

Short-term data indicate that the forecasting period ranges from one day to 30 days. Long-term data means the forecasting period ranges from 30 to 60 days. Data (pandemic data only) indicate that all data are under the COVID-19 pandemic. Discrete data means that only half of the discrete training data were used in the training process, which proved that our model did not depend on sequential data and achieved stable performance with less training data.

To avoid data traversal, we split the dataset by time dimension, 1/1/2021 - 5/31/2022 as the training data and the other for testing and validation. We used Adam as the optimizer and searching for the best learning rate with 0.001 started and the decay rate was set to 0.95, and the batch size was set to be 32. Each experiment was performed five times and the average results were recorded.

A. NOTATION

Rolling type(Column 2 in Table 4).LSTM-based models make multi-step predictions by rolling prediction, which means that the model takes the previous prediction result as a new time-step appended to the new sequential input.

Daily rolling means that the sequential input is the same hour each day and predicts for the same time the next day, such as the input: $\{X_{day1\_12h}, X_{day2\_12h}, \ldots, X_{day14\_12h}\}$ predict for X_{day15_12h}.

Hourly rolling means that the sequential input is consecutive hours, such as $\{X_{day1\_0h}, X_{day1\_1h}, X_{day1\_2h}, \ldots, X_{day1\_22h}\}$ predict for X_{day1+1_23h} and then use the input: $\{X_{day1\_1h}, X_{day1\_2h}, X_{day1\_3h}, \ldots, X_{day1\_23h}\}$ predict for X_{day2_0h}. According to the data analysis shown in **Figure 5**, hourly rolling prediction is more difficult for the LSTM-based model because fluctuations in the data lead to a larger standard deviation.

B. EVALUATION METRICS

We choose three indicators as experimental evaluation metrics of multi-demand distribution task.

Mean Squared Error(MSE) defined as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{d}(p_i(v_k) - \widehat{p}_i(v_k))^2 \tag{24}$$

$\widehat{p}$ is forecasting result of demand-k and p is ground truth.

Normalized Discounted Cumulative Gain(NDCG@N) is a popular method for measuring the quality of a set of search result. **Cumulative Gain(CG)** is the sum of the graded relevance values of all the search results. The CG at a particular rank position p is defined as follows:

$$CG_p = \sum_{i=1}^{n} rel_i \tag{25}$$

In the standard definition of CG where rel_i is the graded relevance of the result at position i. In our task, rel_i is $P_A(v_i)$

TABLE 4. Comparison of different models on Industrial Dataset.

Model	rolling type	data	MSE	RMSE	NDCG@5	NDCG@10	Recall@5	Recall@10
LSTM	daily	next day	1.1557e-04	0.01532	0.9351	0.9373	0.7897	0.7999
LSTM	daily	30 days avg	1.67e-04	0.0129	0.8997	0.9039	0.7413	0.7495
LSTM	daily	60 days avg	2.35e-04	0.01532	0.8669	0.8744	0.7078	0.7178
LSTM	hourly	next hour	3.0514e-04	0.017468	0.8680	0.8681	0.7403	0.7483
LSTM	hourly	next day	3.1633e-04	0.017785	0.8600	0.8615	0.7298	0.7399
LSTM	hourly	next week	3.9862e-04	0.01996	0.8268	0.8343	0.6822	0.7083
CNN+LSTM	daily	next day	1.3743e-04	0.01172	0.9188	0.9214	0.7657	0.7745
CNN+LSTM	daily	30 days avg	1.57e-04	0.0125	0.9031	0.9087	0.7557	0.7622
CNN+LSTM	daily	60 days avg	2.23e-04	0.01493	0.8766	0.8890	0.7231	0.7190
CNN+LSTM	hourly	next hour	3.07e-04	0.01753	0.8647	0.8652	0.7358	0.7439
CNN+LSTM	hourly	next day	3.11e-04	0.017651	0.8615	0.8630	0.7309	0.7429
CNN+LSTM	hourly	next week	3.80e-04	0.01951	0.8405	0.8448	0.7008	0.7193
Spatiotemporal	-	30 days avg	1.44e-04	0.01199	0.917	0.92085	0.76755	0.7783
Spatiotemporal+pandemic	-	30 days avg	1.43e-04	0.01194	0.9175	0.9204	0.7684	0.7769
Spatiotemporal+pandemic + loss optimization	-	30 days avg	1.42e-04	0.0119	0.9172	0.9213	0.7663	0.778
Spatiotemporal+pandemic + graph embedding	-	30 days avg	1.39e-04	0.01177	0.9195	0.9232	0.7698	0.7796
DSTNN	-	next hour	1.33e-04	0.0115	0.9530	0.944	0.871	0.817
DSTNN	-	next day	1.29e-04	0.0113	0.9279	0.9298	0.7837	0.7845
DSTNN	-	next week	1.3726e-04	0.01171	0.9311	0.9324	0.7904	0.7920
DSTNN	-	30th day	1.3173e-04	0.01147	0.9214	0.9237	0.7658	0.7723
DSTNN	-	**30 days avg**	**1.34e-04**	**0.01157**	**0.9226**	**0.9261**	**0.7727**	**0.7822**
DSTNN	-	60 days avg	1.45e-04	0.01202	0.9175	0.9194	0.7618	0.7722
DSTNN	-	discrete data	1.39e-04	0.0118	0.918	0.9222	0.7664	0.7792
DSTNN	-	pandemic only	9.48e-05	0.00973	0.9357	0.9448	0.7977	0.8279

which means the probability of demand v_i under distribution A. Discounted Cumulative Gain **(DCG)** is defined as follows:

$$DCG_p = \sum_{i=1}^{n} \frac{rel_i}{log_2(i+1)} \tag{26}$$

Ideal DCG (IDCG) is defined as the DCG of the best results, therefore, NDCG is defined as

$$nDCG_p = \frac{DCG_p}{IDCG_p} \tag{27}$$

In this study, we draw on NDCG to evaluate our model quality, where rel_i is a single demand probability which indicates the relevance between the demand and spatiotemporal context.

Recall@N In this study, we define Recall@N as

$$recall@N = \frac{N_m}{N} \tag{28}$$

where N_m is the number of prediction results (TopN) occur in the ground-truth results (TopN).

C. COMPARED METHODS

We compare our model DSTNN with most popular time-series models:

•LSTM [17] uses the historical sequential demand distribution as a sequential input and predicts the next few time steps. In the model training phase, we used historical demand distribution data as inputs and predicted the next time step. In the prediction phase, we need to predict many observations forward, and the previously predicted result was treated as a new sequential input to the model so that we can predict N steps ahead.

•CNN+LSTM [30] used a CNN model to extract multi time-scale features and LSTM learned time series dependencies.

We also designed some ablation experiments to prove the effectiveness of our methods.

D. RESULTS AND ANALYSIS

We conducted more than 20 experiments on an industrial dataset to evaluate the performance of the baseline model and DSTNN. The ablation experiments showed that every enhanced part of the model worked as expected. The experimental results are presented in Table 4.

1) OVERVIEW OF THE EXPERIMENTAL RESULTS

As shown in Table 4, DSTNN outperformed in most of the comparison experiments except for the daily rolling prediction of the next day, because the time-series models have advantages in one-step prediction, especially when data change is relatively stable, such as the daily rolling data in **Figure 5(b)**.

For an LSTM-based model, hourly rolling type prediction performs worse than daily rolling. **Figure 5** shows that for hourly rolling time step prediction, the model has to learn both the 24 hours cycle and the weekly cycle. For daily rolling time step prediction, the model only needs to learn the day cycle, and the data are much smoother. Therefore, daily rolling forecasting performs better than hourly rolling for the LSTM-based models.

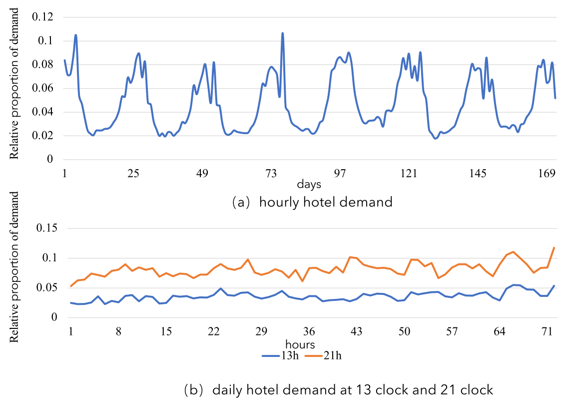


FIGURE 5. Figure.a shows the hotel demand for different hours, which changes over a 24-hour cycle. Figure.b shows the hotel demand during the same hour of each day.

Figure 6 shows the prediction performance of the different models over time. LSTM-based models perform well in short-term forecasting, but their performance declines much faster as the forecasting period increases. In comparison, the performance of the DSTNN does not decline much with the increase in forecast time, which proves that the DSTNN has the ability for long-term forecasting, and modeling demand distribution with spatial-temporal features directly is much more effective than taking historical sequential data as input for long-term forecasting.

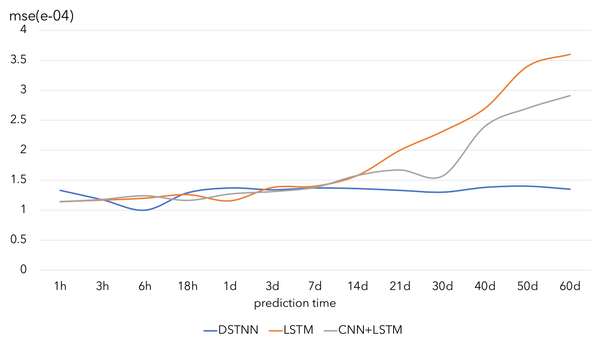


FIGURE 6. Performance comparison of different model with long short-term forecasting(h means hour and d means day). The performance of model DSTNN does not decline much as forecasting period grows. Performance of time series model LSTM and CNN+LSTM decline very fast after one week.

2) COMPARED WITH TIME SERIES MODELS

As expected, the conventional time-series model is suitable for short-term forecasting, especially when X_t does not differ significantly from $X_{t-13}, X_{t-12}, \ldots, X_t$. Our task is not just about short-term forecasting, but has no limitation on the forecast period, especially with unexpected social events and no historical data to reference. The conventional time-series model depends on the results $X_{t-13}, X_{t-12}, \ldots, X_t$ to predict X_{t+1}. In general, X_{t+1} is always very similar to X_t, therefore the time-series model performs well for short-term predictions. However, if we want to have a long-term forecast of X_{t+n}, conventional time series models have to predict $X_{t+1}, X_{t+2}, \ldots, X_{t+n-1}$ step by step, and errors will accumulate, which will lead to a large error of X_{t+n}.Thus, it does not perform well for long-term forecasting. As shown in Table 4, the LSTM for long-term forecasting is much worse than that for short-term forecasting. In comparison, the DSTNN models of demand distribution with spatial-temporal features directly perform much better for long-term forecasting, and there is no substantial decline in model performance. Meanwhile, experimentation with discrete datasets shows that using only half of the discrete training data, which is sampled randomly, the experimental performance does not substantially decrease and still performs well.

3) SPATIAL-TEMPORAL CORRELATION

The DSTNN forecasts demand distribution using spatial-temporal correlation and eliminates historical sequential data. The spatial-temporal model is more sensitive to spatial and temporal features in the future, and can easily adjust to spatial-temporal changes and achieve long-term forecasting. Spatial-temporal correlation is the key point that helps the model achieve stable performance as the forecasting period increases.

TABLE 5. Show cities similar to Beijing and Sanya using graph embedding.

City(Similar with Sanya)	Cosine	City(Similar with Beijing)	Cosine
Lingshui	0.169	Langfang	0.341
Baoting	0.216	Baoding	0.563
Ledong	0.235	Chengde	0.602
Wuzhishan	0.262	Zhangjiakou	0.618
Wanning	0.351	Tianjin	0.627
Haikou	0.369	Hengshui	0.717
Dongfang	0.435	Cangzhou	0.737
Qiongzhong	0.459	Wulanchabu	0.752

TABLE 6. Show cities similar to Beijing and Sanya using embedding learned from main task.

City(Similar with Sanya)	Cosine	City(Similar with Beijing)	Cosine
Shenzhen	0.027	Hangzhou	0.064
Fuzhou	0.111	Tianjin	0.107
Chongqing	0.137	Jinan	0.129
Wuhan	0.15	Shanghai	0.135
Changsha	0.161	Nanjing	0.143
Chengdu	0.187	Qingdao	0.154
Haikou	0.197	Chengdu	0.182
Zhuhai	0.218	Wuhan	0.206

4) REGIONAL EMBEDDINGS INSPECTION

With the training of the main task and pre-trained graph embedding, we achieved embedding in two different regions. The similarities of region embedding are presented in Tables 5 and 6. Region embeddings with graph neural networks showed that *Beijing* has a high correlation with *Langfang, Baoding, Chengde, and Zhangjiakou* which are around *Beijing* spatially, and *Sanya* has high correlation with *Haikou, Wanning, Lingshui, and Ledong* which are around *Sanya* spatially. Whereas region embedding trained by supervised learning is significantly different from graph learning, as shown in Table 6, *Sanya* is very similar to *Shenzhen, Fuzhou, Chongqing, Wuhan, and Haikou* which are almost southern areas with high temperatures in China. *Beijing* is similar to *Hangzhou, Tianjin, Jinan, Nanjing, Qingdao, and Chengdu* which are almost provincial capitals. These similar cities show that we have established a direct correlation between region and demand distribution.

5) INTRODUCTION TO SOCIAL EVENTS AND PANDEMIC TREND FORECASTING AS AUXILIARY TASKS

COVID-19 is one of the most important factors in regional demand forecasting, and it is unpredictable. Furthermore, many cities have never experienced COVID-19, let alone the historical sequence data. In our model, we build a pandemic correlation with social demands in case that we can mock the pandemic factor to prepare for contingency plans. The *Pandemic Only* dataset shows that our model also performs well with pandemic datasets.

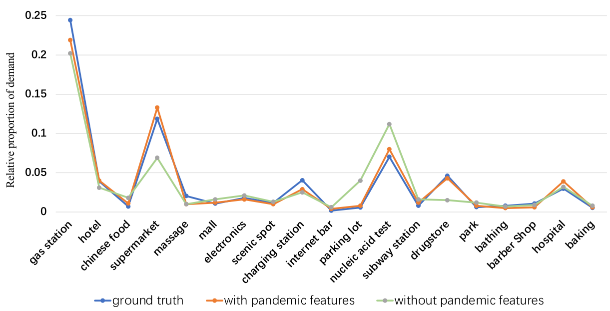


FIGURE 7. Demands distribution forecasting result shows that with pandemic features performs better. Supermarket, drugstore has stronger demand and finding parking lot has less demand (Shanghai in June is under serious pandemic).

The introduction of pandemic information and the auxiliary task of sequence modeling for the prediction of future pandemic development trends can help the model achieve better forecasting results. The pandemic trend is always changing, and modeling this trend will significantly affect the precision of demand forecasting. We build an auxiliary task to predict the trend of the pandemic, which ensures that the hidden state $H(t)$ is a representation of the pandemic trend and not just a correlation between sequence input. As shown in Figure 7, the model with pandemic information is useful for forecasting the demand for *supermarkets, drugstores, and nucleic tests*.

6) LOSS FUNCTION

To enhance the long-tail demand forecast accuracy, we use L_{abs} as an optimization of KL which focuses on the optimization of the relative error. The hyper-parameter λ is used to balance the effect of the auxiliary task. In Table 4, we find that auxiliary loss can greatly improve demand distribution forecasting, which shows the importance of supervision information for the learning of sequential confirmed numbers. This helps the $H(t)$ to learn the representation of pandemic trends. Our goal is to forecast demand and the pandemic trend will help the main task gain more information for the future.

For the demand distribution forecasting goal, the loss function is defined as $L = L_{main} + \lambda L_{aux}$, L_{aux} is the loss of an auxiliary task. Balancing the loss weight of the auxiliary task is meaningful. An excessively large λ will affect the main task optimization, and after a series of experiments for hyper-parameter λ, we choose $\lambda = 0.5$ that achieves the best performance.

To conclude, Table 4 and Figure 6 show that DSTNN outperforms in most of the comparison experiments, and DSTNN's performance is much more stable as the period increases, especially when there is no obvious decline after a week. High performance long-term forecasting is useful in production environments. The LSTM-based time-series model performs well only within a very short period because the demand distribution does not change significantly, whereas the prediction performance declines rapidly as the prediction period increases. In this study, we discuss a model that can achieve excellent performance in both long and short-term forecasting. The DSTNN achieved this goal based on the experimental results.

V. CONCLUSION

In this study, we propose a task for forecasting the regional demand distribution during the COVID-19 pandemic. We designed a new deep neural network structure called the deep spatial-temporal neural network (DSTNN) for demand distribution forecasting. The experimental results show that our method outperforms the conventional time-series model, particularly for long-term forecasting. Some novel techniques have been proposed to enhance performance. We modeled this task to establish a direct correlation between the demand distribution and spatial-temporal features, instead of using historical demand distribution data for time-series forecasting. Social events were considered, such as the pandemic situation, and an auxiliary task was designed to forecast pandemic trends, which will help the model perceive the impact of the pandemic more accurately. The optimization of the loss function was designed to achieve a better performance with long tail demand forecasting. Furthermore, we designed a population mobility graph network to achieve better region representation instead of a region adjacent graph. Above all, our model does not depend on sequential historical demand distribution data, and it performs well in terms of long-term forecasting, which leads to a better application scope. In the future, we intend to design a more comprehensive regional graph network which nodes include more regional information and use graph convolutional network to achieve better spatial feature representation, which will enhance the performance of the demand forecasting model.

APPENDIX A
DATASETS EXAMPLE

This section provides some original data as a supplement to Section III-A. **Table 7** presents population mobility sample data which is used to build the mobility graph. **Table 8** presents COVID-19 relative data which are used to modeling with social events of COVID-19. **Table 9** presents some regional information data including population, hospital number and adjacent region which are used for spatial correlation modeling. **Table 10** and **11** presents the demand distribution data under specific spatial-temporal conditions. The demand distribution in the last column of **Table 10** is shown in detail in **Table 11** which are the learning objectives of our model. All datasets were processed by feature engineering to be fed into the demand forecasting model.

TABLE 7. Population mobility data sample.

region A	region B	mobility probability
3	217	0.00063
3	196	0.00139
3	186	0.00001
3	161	0.00009
3	155	0.00014
3	154	0.00044
3	112	0.00005
3	91	0.00021
3	71	0.00038
3	68	0.00008

TABLE 8. COVID-19 information data.

region	dt	confirm num	death num	heal num	confirm num
51	2022-05-07	3	0	55	630
51	2022-05-06	3	0	67	682
51	2022-05-05	2	0	40	746
51	2022-05-04	3	0	37	784
51	2022-05-03	3	0	56	818
51	2022-05-01	30	0	201	995
51	2022-04-30	16	0	103	1166
51	2022-04-29	19	0	126	1253
51	2022-04-28	34	0	176	1360
51	2022-04-27	52	0	252	1502
51	2022-04-26	46	0	905	1702

TABLE 9. Regional information data.

region	population	hospitals	adjacent region
271	small	few	257|262
7	middle	many	1|3|11|12|13|146
116	middle	middle	114|117|118
218	small	few	219|228|229
43	Large	many	39|44|45|47|49|50

TABLE 10. Demands distribution data.

date	hour	region	is holiday	holiday type	demands
2022-06-03	10	271	3	1	-
2022-06-26	7	7	2	0	-
2022-06-20	19	116	0	0	-
2022-06-01	21	218	0	0	-
2022-06-20	19	43	0	0	-

TABLE 11. Demand distribution sample.

demand ID	ratio	demand ID	ratio
1	0.071258	6	0.026128
2	0.067695	7	0.026128
3	0.041567	8	0.007125
4	0.034441	9	0.002375
5	0.030878	10	0.042755
...	...	...	...
25	0	50	0.008313

REFERENCES

[1] S. Gong, X. Mo, R. Cao, Y. Liu, W. Tu, and R. Bai, "Spatio-temporal parking behaviour forecasting and analysis before and during COVID-19," 2021, *arXiv:2108.07731*.

[2] J. Ke, H. Zheng, H. Yang, and X. Chen, "Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach," *Transp. Res. C, Emerg. Technol.*, vol. 85, pp. 591–608, Dec. 2017.

[3] W. Jiang and L. Zhang, "Geospatial data to images: A deep-learning framework for traffic forecasting," *Tsinghua Sci. Technol.*, vol. 24, no. 1, pp. 52–64, Feb. 2019.

[4] M. Pacella and G. Papadia, "Evaluation of deep learning with long short-term memory networks for time series forecasting in supply chain management," *Proc. CIRP*, vol. 99, pp. 604–609, Jan. 2021.

[5] P. Zhao, A. Luo, Y. Liu, J. Xu, Z. Li, F. Zhuang, V. S. Sheng, and X. Zhou, "Where to go next: A spatio-temporal gated network for next POI recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 5, pp. 2512–2524, May 2022.

[6] Y. Luo, Q. Liu, and Z. Liu, "STAN: Spatio-temporal attention network for next location recommendation," in *Proc. Web Conf.*, Apr. 2021, pp. 2177–2185.

[7] M. H. Rahman and S. M. Rifaat, "Using spatio-temporal deep learning for forecasting demand and supply-demand gap in ride-hailing system with anonymised spatial adjacency information," *IET Intell. Transp. Syst.*, vol. 15, no. 7, pp. 941–957, Jul. 2021.

[8] C. Wang, T. Hou, and M. Barth, "Data-driven multi-step demand prediction for ride-hailing services using convolutional neural network," in *Proc. Sci. Inf. Conf.* Cham, Switzerland: Springer, 2019, pp. 11–22.

[9] J. Fattah, L. Ezzine, Z. Aman, H. E. Moussami, and A. Lachhab, "Forecasting of demand using ARIMA model," in *Proc. Int. J. Eng. Bus. Manag.*, 2018, Art. no. 1847979018808673.

[10] P. Chujai, N. Kerdprasop, and K. Kerdprasop, "Time series analysis of household electric consumption with ARIMA and ARMA models," in *Proc. Int. MultiConf. Eng. Comput. Scientists*, vol. 1, 2013, pp. 295–300.

[11] H. Abbasimehr, M. Shabani, and M. Yousefi, "An optimized model using LSTM network for demand forecasting," *Comput. Ind. Eng.*, vol. 143, May 2020, Art. no. 106435.

[12] P. Juszczak, D. Tax, and R. P. W. Duin, "Feature scaling in support vector data description," in *Proc. ASCI*, 2002, pp. 95–102.

[13] Y. Qu, H. Cai, K. Ren, W. Zhang, Y. Yu, Y. Wen, and J. Wang, "Product-based neural networks for user response prediction," in *Proc. 16th Int. Conf. Data Mining*, 2016, pp. 1–12.

[14] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 5998–6008.

[15] W. Song, C. Shi, Z. Xiao, Z. Duan, Y. Xu, M. Zhang, and J. Tang, "AutoInt: Automatic feature interaction learning via self-attentive neural networks," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, Nov. 2019, pp. 1161–1170.

[16] G. Zhou, N. Mou, Y. Fan, Q. Pi, W. Bian, and C. Zhou, "Deep interest evolution network for click-through rate prediction," in *Proc. AAAI Conf. Artif. Intell.*, Jul. 2019, vol. 33, no. 1, pp. 5941–5948.

[17] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, Oct. 2000.

[18] C. Zhang and P. Patras, "Long-term mobile traffic forecasting using deep spatio-temporal neural networks," in *Proc. 18th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Jun. 2018, pp. 231–240.

[19] A. Kapoor, X. Ben, L. Liu, B. Perozzi, M. Barnes, M. Blais, and S. O'Banion, "Examining COVID-19 forecasting using spatio-temporal graph neural networks," 2020, *arXiv:2007.03113*.

[20] M. Xu, W. Dai, C. Liu, X. Gao, W. Lin, G.-J. Qi, and H. Xiong, "Spatial–temporal transformer networks for traffic flow forecasting," 2020, *arXiv:2001.02908*.

[21] R. Carbonneau, K. Laframboise, and R. Vahidov, "Application of machine learning techniques for supply chain demand forecasting," *Eur. J. Oper. Res.*, vol. 184, no. 3, pp. 1140–1154, Feb. 2008.

[22] B. Perozzi, R. Al-Rfou, and S. Skiena, "DeepWalk: Online learning of social representations," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data mining*, Aug. 2014, pp. 701–710.

[23] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, *arXiv:1301.3781*.

[24] T. Mikolov et al., "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 26, 2013.

[25] S. Ruder, "An overview of multi-task learning in deep neural networks," 2017, *arXiv:1706.05098*.

[26] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*.

[27] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," 2017, *arXiv:1709.04875*.

[28] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial–temporal graph convolutional networks for traffic flow forecasting," in *Proc. 33rd AAAI Conf. Artif. Intell.*, vol. 33, Jan. 2019, pp. 922–929.

[29] C. Song, Y. Lin, S. Guo, and H. Wan, "Spatial–temporal synchronous graph convolutional networks: A new framework for spatial–temporal network data forecasting," in *Proc. AAAI*, 2020, vol. 34, no. 1, pp. 914–921.

[30] S. D. Khan, L. Alarabi, and S. Basalamah, "Toward smart lockdown: A novel approach for COVID-19 hotspots prediction using a deep hybrid neural network," *Computers*, vol. 9, no. 4, p. 99, Dec. 2020.

[31] Z. Cui, J. Wu, W. Lian, and Y.-G. Wang, "A novel deep learning framework with a COVID-19 adjustment for electricity demand forecasting," *Energy Rep.*, vol. 9, pp. 1887–1895, Dec. 2023.

[32] J. Wu, N. Levi, R. Araujo, and Y.-G. Wang, "An evaluation of the impact of COVID-19 lockdowns on electricity demand," *Electr. Power Syst. Res.*, vol. 216, Mar. 2023, Art. no. 109015.

[33] Z. Zhao, J. Wu, F. Cai, S. Zhang, and Y.-G. Wang, "A hybrid deep learning framework for air quality prediction with spatial autocorrelation during the COVID-19 pandemic," *Sci. Rep.*, vol. 13, no. 1, p. 1015, Jan. 2023.

[34] G. Ruan, J. Wu, H. Zhong, Q. Xia, and L. Xie, "Quantitative assessment of U.S. Bulk power systems and market operations during the COVID-19 pandemic," *Appl. Energy*, vol. 286, Mar. 2021, Art. no. 116354.

[35] S. M. Iacus, F. Natale, C. Santamaria, S. Spyratos, and M. Vespe, "Estimating and projecting air passenger traffic during the COVID-19 coronavirus outbreak and its socio-economic impact," *Saf. Sci.*, vol. 129, Sep. 2020, Art. no. 104791.

[36] M. A. Talib, M. Abdallah, A. Abdeljaber, and O. A. Waraga, "Influence of exogenous factors on water demand forecasting models during the COVID-19 period," *Eng. Appl. Artif. Intell.*, vol. 117, Jan. 2023, Art. no. 105617.

[37] D. Borges and M. C. V. Nascimento, "COVID-19 ICU demand forecasting: A two-stage prophet-LSTM approach," *Appl. Soft Comput.*, vol. 125, Aug. 2022, Art. no. 109181.

[38] J. Drummond and M. S. Hasnine, "Did the COVID-19 vaccine rollout impact transportation demand? A case study in New York city," *J. Transp. Health*, vol. 28, Jan. 2023, Art. no. 101539.

[39] I.-I. Prado-Rujas, E. Serrano, A. García-Dopico, M. L. Córdoba, and M. S. Pérez, "Combining heterogeneous data sources for spatio-temporal mobility demand forecasting," *Inf. Fusion*, vol. 91, pp. 1–12, Mar. 2023.

[40] X. Li, Y. Xu, X. Zhang, W. Shi, Y. Yue, and Q. Li, "Improving short-term bike sharing demand forecast through an irregular convolutional neural network," *Transp. Res. C, Emerg. Technol.*, vol. 147, Feb. 2023, Art. no. 103984.

[41] M. Waseem, Z. Lin, S. Liu, Z. Jinai, M. Rizwan, and I. A. Sajjad, "Optimal BRA based electric demand prediction strategy considering instance-based learning of the forecast factors," *Int. Trans. Electr. Energy Syst.*, vol. 31, no. 9, Sep. 2021, Art. no. e12967.

[42] A. Haxhimusa and M. Liebensteiner, "Effects of electricity demand reductions under a carbon pricing regime on emissions: Lessons from COVID-19," *Energy Policy*, vol. 156, Sep. 2021, Art. no. 112392.

[43] U. Şahin, S. Ballı, and Y. Chen, "Forecasting seasonal electricity generation in European countries under Covid-19-induced lockdown using fractional grey prediction models and machine learning methods," *Appl. Energy*, vol. 302, Nov. 2021, Art. no. 117540.

[44] D. Guan, D. Wang, S. Hallegatte, S. J. Davis, J. Huo, S. Li, Y. Bai, T. Lei, Q. Xue, D. Coffman, D. Cheng, P. Chen, X. Liang, B. Xu, X. Lu, S. Wang, K. Hubacek, and P. Gong, "Global supply-chain effects of COVID-19 control measures," *Nature Hum. Behav.*, vol. 4, no. 6, pp. 577–587, Jun. 2020.

[45] H. Zhong, Z. Tan, Y. He, L. Xie, and C. Kang, "Implications of COVID-19 for the electricity industry: A comprehensive review," *CSEE J. Power Energy Syst.*, vol. 6, no. 3, pp. 489–495, Sep. 2020.

[46] F. Pinzaru, A. Zbuchea, and L. Anghel, "The impact of the COVID-19 pandemic on business. A preliminary overview," in *Proc. STRATEGICA*, 2020, pp. 721–730.

[47] B. H. Meyer, B. Prescott, and X. S. Sheng, "The impact of the COVID-19 pandemic on business expectations," *Int. J. Forecasting*, vol. 38, no. 2, pp. 529–544, Apr. 2022.

[48] L. A. Panzone, S. Larcom, and P.-W. She, "Estimating the impact of the first COVID-19 lockdown on U.K. Food retailers and the restaurant sector," *Global Food Secur.*, vol. 28, Mar. 2021, Art. no. 100495.

[49] L. Szász, C. Bálint, O. Csíki, B. Z. Nagy, B.-G. Rácz, D. Csala, and L. C. Harris, "The impact of COVID-19 on the evolution of online retail: The pandemic as a window of opportunity," *J. Retailing Consum. Services*, vol. 69, Nov. 2022, Art. no. 103089.

[50] L.-Y. He, H. Li, J.-W. Bi, J.-J. Yang, and Q. Zhou, "The impact of public health emergencies on hotel demand—Estimation from a new foresight perspective on the COVID-19," *Ann. Tourism Res.*, vol. 94, May 2022, Art. no. 103402.

[51] H. Zhang and J. Lu, "Forecasting hotel room demand amid COVID-19," *Tourism Econ.*, vol. 28, no. 1, pp. 200–221, Feb. 2022.

XIAOCHUAN GUO is currently pursuing the degree in software engineering. He is also interning with Gaode Maps (Alibaba Group Amap Ltd.) as an Algorithm Intern. The main task is about demand forecasting. His research interests include time series analysis, machine learning, and data mining.

WENBO XIE received the M.S. degree in computer science and technology from Xidian University, Xi'an, China, in 2013. He is currently with Gaode Maps (Alibaba Group Amap Ltd.) and primarily engaged in information retrieval and recommendation systems. His research interests include applied deep learning and information retrieval.

XIN LI received the Ph.D. degree from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 2010. He is currently with Gaode Maps (Alibaba Group Amap Ltd.) as the Algorithm Director and mainly engaged in information retrieval, recommendation systems, online advertising, and large scale machine learning systems.