Tax Service Volume Forecasting Based on Informer

Shengbo Wang[1], Chutong Deng[1], Jiale Chen[1], Ning Xiao[2], Shaobo Chen[2] and Yujuan Quan[2,3,*]
[1]International School (College of Information Science and Technology), Jinan University, Guangzhou 511400, China
[2]College of Information Science and Technology, Jinan University, Guangzhou 511400, China
[3]Guandong Provincial Key Laboratory of Traditional Chinese Medicine Informatization, Guangzhou 511400, China
[*]email: tquanyj@jnu.edu.cn

ABSTRACT

Intelligent resource planning is expected to be of significant help in promoting the efficiency in tax halls, while service volume forecasting is a basis for further research on intelligent resource allocation and optimization. This paper uses the historical data from a tax hall in one of the most developed regions in China as dataset and studies the tax service volume forecasting based on time series forecasting, including five different models based on traditional and deep learning methods: ARIMA, Prophet, DeepAR, Transformer and the newly-proposed Informer. Aiming at selecting the best model for forecasting, the performance of the models on different business scenarios is compared. The experiment results indicate that the Informer has shown good performance and required reduced complexity in univariate forecasting. The results are expected to provide reference and further promote the research on intelligent task scheduling and resource allocation to enhance the service quality in the tax hall.

Keywords: Time Series Forecasting, Deep Learning, Tax Service Volume

1. INTRODUCTION

In real-life practice, the tax halls are limited to the number of windows and resources. Therefore, how to improve the quality of services under such condition has become an essential problem [1].

Currently, intelligent resource planning with statistical and machine learning techniques are expected to play an important auxiliary role in improving business efficiency. The intelligent resource planning includes forecasting and allocation, while resource allocation is a pre-action that is only practical when the distribution of upcoming business volume is known in advance [2]. Thus, tax service volume forecasting is the primary task of research on intelligent planning.

Tax business volume forecasting is a typical time series forecasting problem. There is a strong dependence between time series data, and even the data in a long time before will have an impact on the future data [3]. Traditional time series forecasting models are based on statistical learning, including the Autoregressive Integrated Moving Average (ARIMA) model, exponential smoothing method. The Prophet proposed by Taylor et al. [4], as a generalized additive model, takes the forecasting of time series as a curve-fitting problem, which allows it to have better interpretability. Time series forecasting models based on traditional machine learning methods generally utilize multi-dimensional features in data to manually extract features related to the current time point for building regression models. However, deep neural networks pay attention to the periodicity of time series data and extract the features automatically [5]. The Deep State Space Model (DSSM) proposed by Rangapuram et al. [6]. uses Recurrent Neural Network (RNN) to generate the parameters of the Linear Gaussian State Space Model (LGSSM) at each time step. The Deep Autoregressive Model (DeepAR) proposed by Flunkert et al. [7] is an autoregressive RNN architecture designed for probabilistic prediction problems based on Long-Short-Term-Memory (LSTM). Unlike the RNN-based models, the Transformer proposed by Vaswani et al. [9] uses an Encoder-Decoder architecture based on the Attention Mechanism to capture the relationships between data in different parts of history. Moreover, the Informer proposed by Zhou et al. [10] made many changes based on the Transformer to increase performance and reduce complexity, and has made significant improvements in long-time series prediction.

The paper constructs and compares the accuracy of several tax service volume forecasting models based on traditional and deep learning time series forecasting models, including ARIMA, Prophet, DeepAR, Transformer and the newly-proposed

Informer. This is expected to provide the tax hall with a more accurate method to estimate the future business volume and raise the convenience of business management.

2. DATASET & MODELS

2.1. Dataset

The research of tax service volume forecasting is based on the historical data of a tax hall from January 1, 2019, to December 31, 2020 in one of the most developed regions in China. The dataset covers the information of every taxpayer, from the time when taking the ticket, calling the number to the completion of the transaction, with different types of tax services (a total of 22 kinds) on every working day. Therefore, the data in this paper is of great representativeness and influence for intelligent resource planning.

2.2. Models

Our forecasting model based on Informer is able to achieve significant improvement in long-sequence time series forecasting. To reduce the computational complexity, the information transmission process of the Attention is sparse. In the case of time t, not all time points before time t are correlated with t, and the complexity can be reduced if the calculations for these time points and t are ignored.

The Attention in Informer is defined as:

$$out_i = softmax(q_i K^T)V = \sum_{j=1}^{L_k} \frac{e^{q_i k_j^T/\sqrt{d}}}{\sum_{l=1}^{L_k} e^{q_i k_j^T/\sqrt{d}}} v_j = \sum_{j=1}^{L_k} p\,(k_j|q_i) v_j \tag{1}$$

When calculating the Attention, if the overall correlation between q_i and the key matrix is low, $p(k_j|q_i)$ degenerates into a uniform distribution, and the output of Attention degenerates into the average of the rows of the value matrix. Therefore, we can use the difference between $p(k_j|q_i)$ and the uniform distribution to measure the sparseness of query q_i. If KL divergence is low, the Attention can be replaced by averaging the rows with respect to V. Specifically, let q be a uniform distribution and let p be $p(k_j|q_i)$, the KL divergence is calculated as:

$$M(q_i, K) = ln \sum_{j=1}^{L_k} e^{\frac{q_i k_j^T}{\sqrt{d}}} - \frac{1}{L_k} \sum_{j=1}^{L_k} \frac{q_i k_j^T}{\sqrt{d}} \tag{2}$$

KL divergence can be calculated for each query, and the Top-k's queries are taken for the Attention. Since the complexity is still $O(L_k L_q)$ when sorting the Top-k query, the author proposes a method for approximate calculation for $M(q_i, K)$:

$$M(q_i, K) = ln \sum_{j=1}^{L_k} e^{\frac{q_i k_j^T}{\sqrt{d}}} - \frac{1}{L_k} \sum_{j=1}^{L_k} \frac{q_i k_j^T}{\sqrt{d}} \tag{3}$$

When calculating the upper bound, a part of k_j can be sampled randomly. Reducing k_j also reduces the complexity.

Due to the redundant information in Prob-Sparse self-attention, the “convolution + max pooling” method is adopted for the Encoder. Besides, the fully-connected network is replaced by Convolutional Neural Networks (CNNs), while the timestamp information is added as part of encoding, which allows Informer to be better at learning the periodicity of data. When decoding, the method of outputting multiple predictions at a time is adopted. The input to the Decoder is the sum of the intercepted part of the later part of input to the Encoder and the 0 matrices with the same shape as the predicted target.

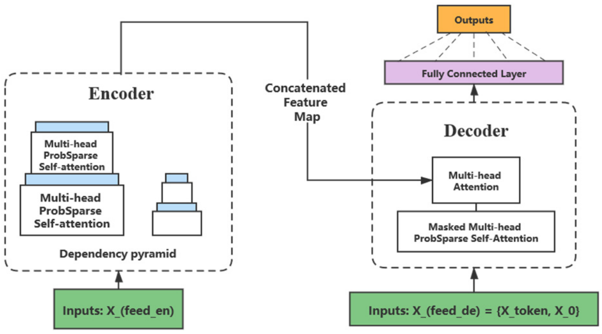

Figure 1. An overall graph of the Informer model

3. EXPERIMENTATION

3.1. Experiment Description

In this chapter, five time series forecasting models, based on ARIMA, Prophet, DeepAR, Transformer, and Informer, are constructed, and their performance is compared. The models based on ARIMA and Prophet adopted the traditional linear fitting model for training and prediction, while models based on DeepAR, Transformer, and Informer use a deep learning-based approach for training and forecasting.

To implement the models, the "statsmodels" library from Python is used for ARIMA, the "fbprophet" library provided by Facebook is used for Prophet, and the open-source library "GluonTS" provided by Amazon is used for DeepAR and Transformer. For Informer, the source code provided by authors is altered and applied.

Since the tax service volume is periodic, the data from 2019 are selected for training, and the data from 2020 are used for testing. Direct time series forecasting is used in ARIMA and Prophet, and when forecasting, the input is the timestamp of the predicted part, and the output is the value of the predicted part. Rolling time series forecasting is used in DeepAR, Transformer, and Informer. The length of the prediction window needs to be set in advance during training. The historical data of the previous period is input and the value of the predicted part is output.

Different from the traditional models, the accuracy of models based on deep learning is more likely to be affected by the length of the predicted part. Thus, performance of these models from two business scenarios, including daily forecast and monthly forecast, are compared. The length of prediction window on a daily-based forecast is 1. When forecasting by month, since the length of each month is variable, the latest historical data is used for training, and the length of the prediction window is the number of days in next month.

3.2. Data Preprocessing

Most time series forecasting models extract information based on dates. If the business volume data containing only working days are directly used, different models will adopt different ways to deal with the missing data on the holidays, which will be of high possibility to affect the learning of the overall trend of data. Therefore, the missing data on holidays is filled to ensure the continuity of data.

The basic objective of this research is to forecast the tax service volume on a daily basis. At the same time, based on the fact that the deep learning models support the use of multi-dimensional data for time series forecasting, daily tax service volume of different types of services (a total of 22 types) is extracted as the feature dimensions. And in order to facilitate the models to extract information, the missing data of public holidays is filled to ensure the continuity of the time series. The details of the preprocessed dataset are described as follow:

Table 1. Dataset Description

Duration	Number of Weekdays	Mean	Median
2019	253	887.3	935
2020	250	579.3	616.5
2019-2020	503	734.2	723

3.3. Evaluation Index

To evaluate the performance of different models, Rooted Mean Square Error (RMSE) is chosen as the evaluation index. The smaller RMSE is, the better the model works.

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \tag{4}$$

where m is the number of samples, y_i is the actual value and $\hat{y}_i$ is the predicted value. For evaluation, we will only evaluate the performance of the model on working days.

3.4. Experiment Results

Table 2 and Table 3 show the effects of each model on the dataset. Table 4 shows the model with the best performance (with RMSE score). Figure 2 and Figure 3 show the result of univariate forecasting result of Informer on a daily and monthly basis respectively.

Table 2. Experimental Results of Traditional Models (with RMSE score)

Model	**Univariate Forecasting**
ARIMA	263.59
Prophet	***182.08***

Table 3. Experimental Results of Deep Learning Models (with RMSE score)

Model	**Univariate forecasting by day**	**Univariate forecasting by month**	**Multivariate forecasting by day**	**Multivariate forecasting by month**
DeepAR	205.44	329.99	***155.01***	341.96
Transformer	225.01	322.86	198.43	***315.80***
Informer	***192.67***	***191.11***	407.25	323.34

Table 4. Best Model of Volume Forecasting (with RMSE score)

	Univariate forecasting by day	**Univariate forecasting by month**	**Multivariate forecasting by day**	**Multivariate forecasting by month**
best model (Deep Learning)	Informer 192.67	Informer 191.11	DeepAR 155.01	Transformer 315.80
best model	Prophet 182.08	Prophet 182.08	DeepAR 155.01	Transformer 315.80

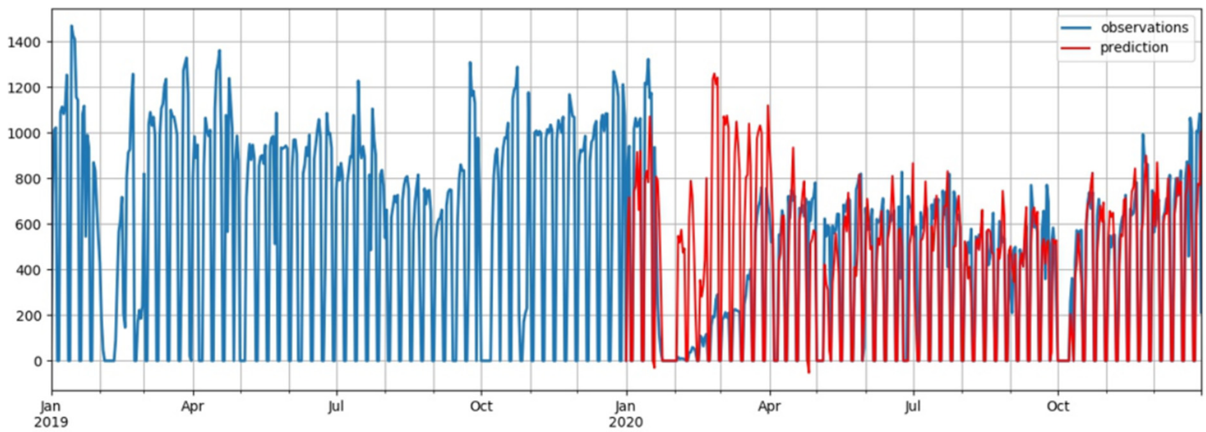

Figure 2. Informer Univariate Forecasting per-Day

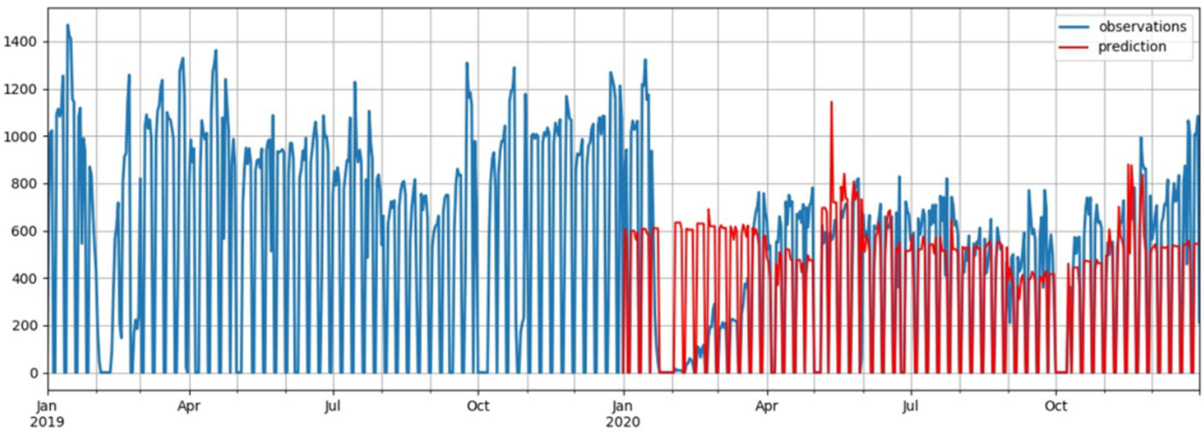

Figure 3. Informer Univariate Forecasting per-Month

Apart from the months where COVID-19 had unpredictable influence on tax service volume, the model based on Informer has shown advanced performance in forecasting for both daily forecast and monthly forecast. Besides, in practice, it has also shown a reduced complexity compared with Transformer, which makes it a model with higher efficiency.

4. CONCLUSION

It can be concluded from the experiments that the model based on Informer has achieved good performance. However, it does not show an overwhelming advantage, which is largely due to the lack of training data, making it difficult for the models to predict the unpredictable impact of COVID-19. Especially after adding another dimension, the performance of Informer on this dataset starts to deteriorate. The experimental phenomena possibly are a consequence of the lack of data in the dataset, since Informer may magnify the errors of the dataset after distilling the Attention.

The model based on Prophet shows satisfying performance, and this is because the maximum capacity of 2020 is predefined to be 80% of the maximum capacity of 2019, which is often set based on the understanding of experienced data analysts. Thus, the impact of COVID-19 had been anticipated.

For deep learning models, the model based on Informer shows a lowest RMSE score comparing with DeepAR and Transformer in univariate prediction in both daily forecasting and monthly forecasting. Moreover, the attention mechanism of the model based on Informer shows a reduced complexity in compared with Transformer, which makes Informer an efficient model for tax service volume forecasting.

In the future, in order to further enhance the quality of tax service, the task scheduling and allocation can be promoted based on the previous tax service volume forecasting. Starting from predicting and reducing the waiting time for taxpayers, machine learning methods can be applied on both tasks. Furthermore, with these models, an online-offline intelligent tax hall for tax service can be established.

ACKNOWLEDGEMENT

This paper is a substantive result of the National Key R&D Program of China under Grant 2018YFC2002500, Guangdong Basic and Applied Basic Research Foundation under Grant 2021A1515011999, Guangdong Key Laboratory of Traditional Chinese Medicine Informatization, Guangdong Province Big Data Innovation Engineering Technology Research Center, "Outstanding Future" Data Scientist Incubation Project of Jinan University, and the Fundamental Research Funds for the Central Universities, Jinan University under Grant 21619412.

REFERENCES

[1] Zhang M. 2020 Practical Thinking on the New Tax Service in the Era of Artificial Intelligence *International Conference on E-Commerce and Internet Technology: International Conference on E-Commerce and Internet*

Technology (ECIT) pp 201-203

[2] Wacker, J.G. and Lummus, R.R. (2002), "Sales forecasting for strategic resource planning", *International Journal of Operations & Production Management*, Vol. 22 No. 9, pp. 1014-1031. https://doi.org/10.1108/01443570210440519

[3] Faloutsos C, Flunkert V, Gasthaus J, et al. 2019 Forecasting Big Time Series: Theory and Practice *The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* New York United States Association for Computing Machinery pp 3209 **-**3210

[4] Taylor S, Letham B 2017 Forecasting at scale *PeerJ Preprints* 5:e3190v2 https://doi.org/10.7287/peerj.preprints.3190v2

[5] Längkvist M, Karlsson L, Loutfi A 2014 A review of unsupervised feature learning and deep learning for time-series modelling *Pattern Recognition Letters* **42** pp 11-24

[6] Rangapuram S, et al. 2018 Deep state space models for time series forecasting *The 32nd International Conference on Neural Information Processing Systems* Curran Associates Inc. United States pp 7796–7805

[7] Flunkert V, Salinas D, Gasthaus J 2017 DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks*. International Journal of Forecasting* **36** pp 1181-1191

[8] Alana de Santana Correia, et al. 2021 Attention, Please! A Survey of Neural Attention Models in Deep Learning *ArXiv*. 2103.16775v1

[9] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, Kaiser L, Plosukhin I 2017 *Attention Is All You Need ArXiv.* **abs/1706.03762**

[10] Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, Zhang W 2021 Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting *35th AAAI Conference on Artificial Intelligence* pp 11106**-**11115