Forecasting Long-Term Call Traffic Based on Seasonal Dependencies

Longchun Cao, Kui Ma, Bin Cao(✉), and Jing Fan

Zhejiang University of Technology, Hangzhou, China
bincao@zjut.edu.cn

Abstract. How to use future call traffic for scheduling different staffs to work in a month or a week is an important task for call center. In this problem setting, the call traffic should be predicted in a long-term way where the forecasting results for different periods are required. However, it is very challenging to solve this problem due to the randomness nature of the call traffic and the multiple forecasting in long term. Current methods cannot solve this problem since they either merely focus on short-term forecasting for the next hour or next day, or ignore call-holding time for call traffic prediction. In this paper, we propose an effective method for predicting long-term call traffic with multiple forecasting results for different future periods, e.g., every 15 min, and take both call arrival rate and call-holding time into consideration through the Erlang. In our method, the seasonal dependencies are summarized by performing data analysis, then different features based on these dependencies are extracted for training the prediction model.In order to forecast call traffic of multiple time buckets, we propose two strategies based on different features. The evaluation results show that the features, the prediction models and the strategies are feasible.

Keywords: Long-term · Multiple · Call traffic · Forecasting · Seasonal dependence

1 Introduction

Nowadays, more and more companies set up call centers to help them process customer requests through the telephone. Knowing future call traffic in different time buckets in advance can help call centers to further improve their service quality. However, it is far from trivial to perform effective long-term call traffic forecasting: First, the call arrival process is very complicated and it may be affected by different causes in different time buckets [1–3], e.g., many people may make calls to the taxi service center when they want to take rides to their offices in the morning. Second, some tasks in the call center usually require the awareness of call traffic in different time buckets within a future long term, e.g., the call center wants to schedule the staffs to different time buckets based on the corresponding call traffic for the next week, and as far as we know, the problem of forecasting call traffic of multiple time buckets haven't been well studied. Most existing solutions for call traffic forecasting merely focus on the prediction for the single time bucket of the next hour or day [3–5].

X. Wang et al. (Eds.): CollaborateCom 2019, LNICST 292, pp. 231–246, 2019.
https://doi.org/10.1007/978-3-030-30146-0_16

In this paper, based on the observed seasonal dependencies of historical call traffic data, we propose an effective method for predicting long-term call traffic with multiple forecasting results for different future time buckets. It is important to note that the call traffic here is calculated by Erlang formula [6] where both the call arrival rate and the average holding time are considered. In our method, we first extract the following three types of seasonal features: (1) Date time features, like the year, month, day, and the beginning of the time bucket, (2) Special days features, such as the day of week, whether it is the weekend, the beginning, middle or end of a month, (3) Intraday and interday features, which correspond to the call traffic of the same time buckets in the past few days and the call traffic of previous time buckets. Then, in order to forecast call traffic of multiple time buckets, we propose two strategies based on whether taking into account the third feature type, i.e., intraday and interday features. The first strategy that merely considers first two types of the feature is to directly use supervised classification method, i.e., train the model that connect features and corresponding call traffic, and then perform the prediction by inputting the corresponding features. The second strategy performs the prediction in an incremental way, i.e., we first forecast the next call traffic, and then use the predicted result as the intraday and interday feature values for the next time bucket.

The classification method used in our work is the Random Forests (RF) [7] due to its robustness in real applications. Moreover, to demonstrate the effectiveness of our method, we use the real-world dataset from a China Telecom branch throughout the paper. The experimental evaluation shows that considering different correlated dependencies play an important role in the call traffic forecasting. The contributions in this paper can be summarized as follows:

– We extract a variety of features based on different types of seasonal dependencies, including date time type features, special days features, and the intraday and interday features.
– We propose an incremental strategy for forecasting call traffic of multiple time buckets when intraday and interday features are considered.
– We perform extensive experiments and prove the effectiveness of the proposed method.

The rest of the paper is organized as follows. In Sect. 2, we briefly introduce the dataset. The feature construction based on the seasonal dependencies is detailed in Sect. 3. Then we present the method for forecasting call traffic in Sect. 4. Section 5 shows the experimental results. Related work is reviewed in Sect. 6. Finally, in Sect. 7, we conclude the paper and describe future work.

2 Dataset

The data we use is from a call center of China Telecom which has millions of customers, and its background database records all the information about the call center service. Note that the original data we get from the background database of the call center is from 1 January 2016 to 31 December 2018, and each time bucket that collects call traffic is 15-min intervals.

In order to settle down our problem with minimal cost, we only extract the following five fields from the original database:

- *callID*, the unique identification for the call services in each time bucket.
- *callDate*, the begin date of each time bucket.
- *callTime*, the begin time of each time bucket.
- *callArrivals*, the volume of the call arrivals which are collected in each time bucket.
- *callDuration*, the average call-holding time (the average time of a phone call) in each time bucket.

For the sake of describing the records of the call service more simply and effectively, we generate a new filed to identify the records. We name *callDate* plus *callTime* as *callDatetime* as the unique identification for the call services over 15-min intervals. Furthermore, we name the result of multiply *callArrivals* per second by *callDuration* as *callTraffic*. Finally, we get two fields that we need through data preprocessing: *callDatetime* and *callTraffic*, which can be computed by Eqs. 1 and 2, respectively.

$$callDatetime = callDate + callTime \tag{1}$$

$$callTraffic = \frac{callArrivals}{15 \times 60} \times callDuration \tag{2}$$

3 Feature Engineering

In this section, firstly, we analyze the data and introduce what features to be extracted. The rest of the section, we describe how to extract the features.

3.1 Data Analysis

The call traffic is different in different years, months, days and time buckets, and exhibit intraday, daily, weekly monthly and yearly seasonalities on the influence of people's normal routine and scheduling of the call center. We put the year, the month, the day and the time bucket into the features of the call traffic. Furthermore, we found that call traffic of some special days such as the day of the week, the weekend, the beginning, middle or end of a month, and holidays will suddenly increase or decrease. Therefore, we also take these factors as features. Moreover, under the influence of external factors like the weather or promotions, the impact will last for a period of time or days. As a result, the call traffic is related to the call traffic of previous time buckets, and the call traffic of the same time buckets in the past few days. Undoubtedly, we also consider these two factors as features. Taking these effective features which are related to the call traffic into consideration will make the forecasting more accurate.

Due to the call traffic influenced by date, time, external factors, and based on the analysis of the data mentioned above, we extract the following three types of features:

– Date time features, which have the date time dependencies, like the year, month, day, and total minutes of the begin time of the time bucket.
– Special days features, such as the day of the week, whether it is the weekend, the beginning, middle or end of a month, and whether it is a festival. Note that, the festival is the Spring Festival in this paper.
– Intraday and interday features, that is the call traffic of previous time buckets and the call traffic of the same time buckets in the past few days.

3.2 Date Time Features

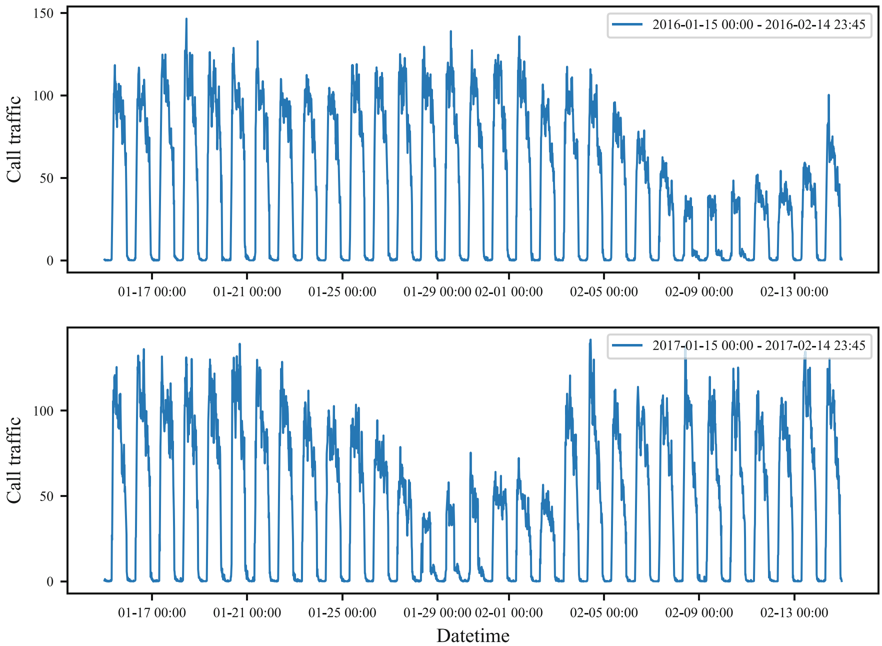

Fig. 1. 15-min call traffic from 15 January, 2016 to 14 February, 2016 and 15 January, 2017 to 14 February, 2017.

Figure 1 provides examples of the two time series. We plot 15-min call traffic from 15 January, 2016 to 14 February, 2016 and 15 January, 2017 to 14 February, 2017. Observe from different years, months, days and time buckets, we can easily find that the call traffic is completely different. So that we take the year, month, day and time bucket as the features. By calculating the timestamp, we can get the values of the date time features. For example, the timestamp is "2016-01-15

07:30", and the values of the year, month, day and total minutes of the time are "2016", "1", "15" and "450". Note that we named *totalMinuetes* as the total minutes of the begin time of the time bucket, which can be calculated by Eq. 3. To sum up, the date time features are the *year*, *month*, *day* and *totalMinuetes*.

$$totalMinuetes = hour \times 60 + minute \tag{3}$$

3.3 Special Days Features

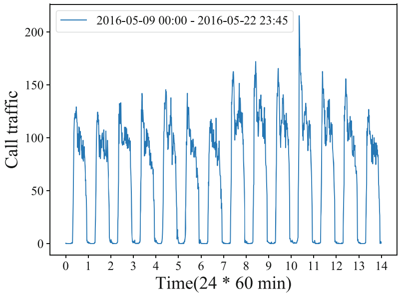

Fig. 2. 15-min call traffic over two consecutive weeks from 9 May 2016 to 22 May 2016.

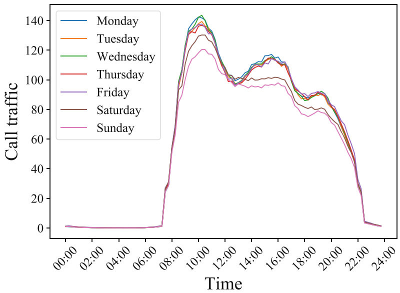

Fig. 3. Intraday profiles of call traffic by weekday and weekend from 1 January, 2016 to 31 December, 2017.

Weekly Features. From Fig. 2, we illustrate weekly seasonality by plotting daily call traffic from 9 May, 2016 to 22 May, 2016 including weekends. It is not hard to find that the call traffic is not equal every day of a week and less on weekends than other days, such as 14 May, 2016 and 15 May, 2016. In order to make this view more convincing, we analyze the data from 1 January, 2016 to 31 December, 2017 and plot 15-minutely average call traffic every day of the week, in Fig. 3. Although the call traffic of every day has a similar distribution, the call traffic at the same time bucket is still not equal and less on weekends than other days. We named *dayofweek* as the day of the week and *isweekend* as whether it is the weekend. The values are given in Table 1. To sum up, the weekly features are *dayofweek* and *isweekend*.

Monthly Features. In Fig. 4 we plot the call traffic per day arriving at the call center from 1 January, 2016 to 31 December, 2016 and from 1 January, 2017 to 31 December, 2017. It is very clear that the volumes of call traffic are particularly large at the beginning and end of every month and the call traffic at the middle of the month is also larger than the other days of the month. In Fig. 5, we plot the average daily call traffic of the same day of the different month from 1 January, 2016 to 31 December, 2017. Through the analysis of call traffic

Table 1. The values of the weekly features in the day of week.

	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.	Sun.
dayofweek	1	2	3	4	5	6	7
isweekend	0	0	0	0	0	1	1

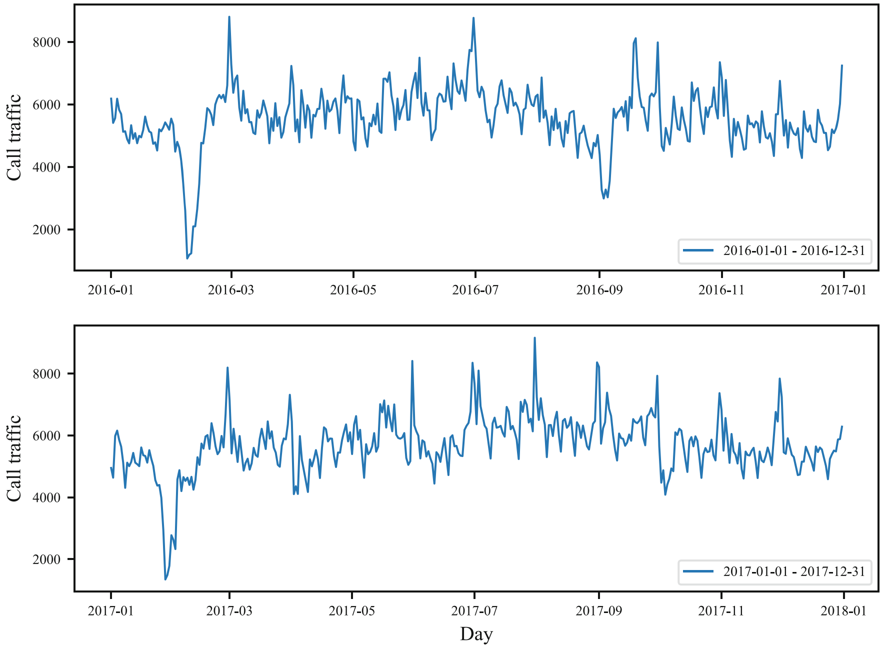

Fig. 4. Daily call traffic over successive months from 1 January, 2016 to 31 December, 2016 and 1 January, 2017 to 31 December, 2017.

distribution and the experience of staffs, The values of the beginning, middle, and end of the month are generated by Eq. 4, and the monthly feature here is named *sectionofmonth*.

$$sectionofmonth = \begin{cases} 0, & otherwise \\ 1, & day \in [1, 5] \\ 2, & day \in [16, 22] \\ 3, & day \in [days - 2, days] \end{cases} \tag{4}$$

where day is the day of a month, $days$ is the total days of a month.

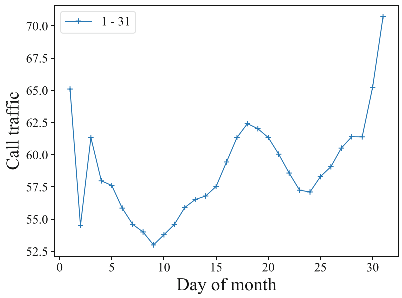

Fig. 5. Average daily call traffic of the same day of the different month from 1 January, 2016 to 31 December, 2017.

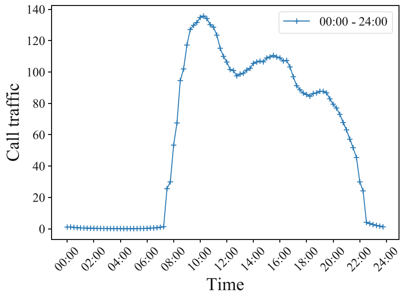

Fig. 6. Intraday profiles of average call traffic per day from 1 January, 2016 to 31 December, 2017.

Yearly Features. Observe from Figs. 1 and 4, the call traffic in the Spring Festival from 7 January, 2016 to 13 January, 2016 and from 27 January, 2017 to 2 February, 2017 is obviously less than the other days. The reason for the sharp decline in call traffic during the Spring Festival is that most people are on holiday and only a few staffs are on duty. Furthermore, the annual Spring Festival is not on fixed days. We refer to those days with unusual call traffic as special days of the year. In this paper, the special days we mainly consider is the Spring Festival. The yearly feature here is named *isfestival*, and the value can be calculated by Eq. 5.

$$isfestival = \begin{cases} 1, & is\ during\ the\ Spring\ Festival \\ 0, & otherwise \end{cases} \tag{5}$$

3.4 Intraday and Interday Features

Intraday Features. From the Figs. 1, 2 and 3, we know that intraday profiles of call traffic in every day is similar to each other. For a more microscopic view of intraday calll traffic, we plot intraday profiles of average call traffic per day from 1 January, 2016 to 31 December, 2017, in Fig. 6. We find that the trend of call traffic is almost the same for a consecutive period of time, e.g., from *07:00* to *11:00*, the call traffic increase at an almost same rate. In addition to the schedule of the call center, we divide the day into several sections with a similar trend. According to the trend of increase and decrease, here the sections are [22:30, 07:30), [07:30, 11:00), [11:00, 13:00), [13:00, 16:30), [16:30, 18:00), [18:00, 19:30) and [19:30, 22:30). The minimum duration is [16:30, 18:00) and [18:00, 19:30), and the count of the time buckets is *6*. Hence, the intraday features are the call traffic in the past up to *6* time buckets. For example, if we want to forecast the call traffic of [10:00, 10:15), the intraday features are the call traffic of [09:45, 10:00), [09:30, 09:45), [09:15, 09:30), [09:00, 09:15), [08:45, 09:00) and [08:30, 08:45). In Table 2, we illustrate the intraday correlation in consecutive 15-min intervals from *07:30* to *08:30*. The measure to be used to capture intraday

Table 2. Correlations between call traffic in consecutive 15-min intervals.

Time bucket	[07:30, 07:45)	[07:45, 08:00)	[08:00, 08:15)	[08:15, 08:30)	[08:30, 08:45)
[07:30, 07:45)	1	0.84	0.76	0.73	0.69
[07:45, 08:00)		1	0.81	0.78	0.76
[08:00, 08:15)			1	0.88	0.79
[08:15, 08:30)				1	0.79
[08:30, 08:45)					1

dependence in call traffic is Pearson's correlation coefficient. Table 2 illustrate two properties which are observed very commonly in reality:

– Correlations between the adjacent time buckets within a day are strong and positive.
– Intraday correlations are slightly smaller, with longer lags.

Interday Features. In Figs. 2 and 3, the call traffic exhibit daily and weekly seasonalities. The volumes and trend of call traffic are similar at the same time bucket on each day of the week. The interday features are the call traffic at the same time bucket in past up to 7 days. For instance, if we want to forecast the call traffic of [10:00, 10:15) on 8 January, the interday features are the call traffic of [10:00, 10:15) on 7 January, 6 January, 5 January, 4 January, 3 January, 2 January and 1 January. In Table 3, we illustrate the interday correlation in consecutive days from Monday to Sunday. The measure to be used to capture interday dependence in call traffic is Pearson's correlation coefficient. Table 3 illustrate two properties which are observed very commonly in reality:

– Correlations between successive days are strong and positive.
– Interday correlations are slightly smaller, with longer lags.

Table 3. Correlations between call traffic in consecutive days.

The day of the week	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.	Sun.
Mon.	1	0.89	0.80	0.74	0.72	0.64	0.60
Tues.		1	0.87	0.79	0.79	0.72	0.62
Wed.			1	0.85	0.81	0.75	0.63
Thurs.				1	0.87	0.78	0.65
Fri.					1	0.87	0.73
Sat.						1	0.83
Sun.							1

4 Forecasting Call Traffic

In this section, we will introduce the process of forecasting call traffic, including call traffic model training and call traffic forecasting. Firstly, we select the training data that we need from the existing call traffic data. Then we use the training data to train the model. Finally, input the features of the time buckets to be forecasted into the model to obtain the call traffic.

4.1 Call Traffic Model Training

Call traffic model training includes the training data preparation and model training. Firstly we introduce how to prepare the training data, then show the way of the model training.

Training Data Preparation. For the reason that the call traffic exhibit yearly and monthly dependencies, the training data we have prepared is the data for the first few months of the forecast month and the same months in the previous years. For example, in Fig. 7, if we want to forecast the call traffic of September 2018 and set the number of previous years to 2, the number of previous months to 3, the training data is from June 2018 to August 2018, from June 2017 to September 2017 and from June 2016 to September 2016.

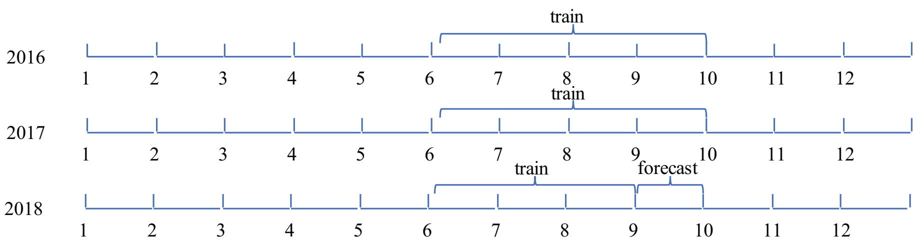

Fig. 7. Training data for September 2018.

Model Training. In this paper, we choose the RF to train model, which is a flexible and easy to use machine learning algorithm that always produces a good result even without hyper-parameter tuning. Since the RF is a supervised learning algorithm, we should know the features and target of each time bucket. Firstly, we can get the training data of the forecast month according to Sect. 4.1, then get the preprocessed data according to Sect. 2. Note that, if you select intraday and interday features to train the model, make sure the values are not null. Because the first few data has no intraday or interday features. Next, we can get the features of each training data according to Sect. 3, and the corresponding target is the call traffic of each time bucket. See Table 4, $tb_{d,i}$ is the ith time bucket on the dth day, $dt_{d,i}$ is the corresponding features of date time, $sd_{d,i}$

is the corresponding features of special days, and $t_{d,i}$ is the corresponding call traffic. Finally, we can select several of the features and target of each time bucket to train the model by using the RF. After training, we can get the model for forecasting call traffic.

Table 4. Features and target of each time bucket.

Time bucket	Features			Target
	Date time	Special days	Intraday and interday	
⋮	⋮	⋮	⋮	⋮
$tb_{d,i}$	$dt_{d,i}$	$sd_{d,i}$	$t_{d,i-1}, t_{d,i-2}, \cdots, t_{d-1,i}, t_{d-2,i}, \cdots$	$t_{d,i}$
$tb_{d,i+1}$	$dt_{d,i+1}$	$sd_{d,i+1}$	$t_{d,i}, t_{d,i-1}, \cdots, t_{d-1,i+1}, t_{d-2,i+1}, \cdots$	$t_{d,i+1}$
⋮	⋮	⋮	⋮	⋮
$tb_{d+1,i}$	$dt_{d+1,i}$	$sd_{d+1,i}$	$t_{d+1,i-1}, t_{d+1,i-2}, \cdots, t_{d,i}, t_{d-1,i}, \cdots$	$t_{d+1,i}$
$tb_{d+1,i+1}$	$dt_{d+1,i+1}$	$sd_{d+1,i+1}$	$t_{d+1,i}, t_{d+1,i-1}, \cdots, t_{d,i+1}, t_{d-1,i+1}, \cdots$	$t_{d+1,i+1}$
⋮	⋮	⋮	⋮	⋮

4.2 Call Traffic Forecasting

This section firstly introduces the incremental forecasting, then presents two strategies to forecast the call traffic. The first strategy uses the model trained without intraday and interday features to forecast, while the second strategy uses the model trained with intraday and interday features to forecast.

Incremental Forecasting. As we all know, the external events such as the weather or promotions will influence the call traffic, and will increase the call traffic for most of the time. Moreover, the impact will last a period of time or days. The time of the external events occur is not fixed, hence there are no date and time seasonalities. However, the call traffic has the intraday and interday dependencies, i.e., it related to the call traffic of previous time buckets and the call traffic of the same time buckets in the past few days. Therefore, we can forecast the call traffic according to the previous. For one month forecast, we can forecast the first call traffic based on the previous call traffic, and the remaining call traffic should be forecasted in an incremental way. Since when we forecast the first call traffic, the previous call traffic has been given, but the previous call traffic of the rest call traffic to be forecasted is unknown. Consequently, before forecasting the next call traffic, we must first forecast the previous. For example, the month to be forecasted is September, we first forecast the call traffic of [00:00, 00:15) on 1 September according to the previous call traffic that is given, then we forecast the call traffic of [00:15, 00:30) on 1 September, according to the call traffic of [00:00, 00:15) on 1 September that has been forecasted and the previous call traffic that is given. So that we can forecast the next call traffic in this incremental way. Finally, we get the call traffic of every time bucket in the month.

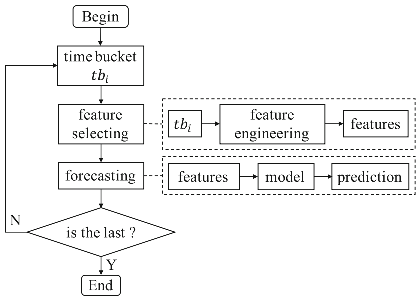

Fig. 8. The flowchart of forecasting without intraday and interday features.

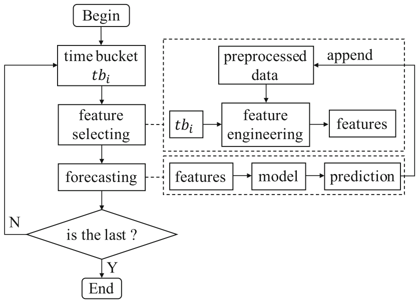

Fig. 9. The flowchart of forecasting with intraday and interday features.

Forecasting Without Intraday and Interday Features. In this method, the features of each time bucket we can get by calculating the beginning timestamp of the time bucket as we mentioned in Sect. 3. Then we can forecast the call traffic of each time bucket by the model trained without intraday and interday features in Sect. 4.1. The flowchart of Fig. 8 details the adopted forecast strategy.

Forecasting with Intraday and Interday Features. As we mentioned in Sect. 3.4, the associations between the call traffic of adjacent time bucket play a role in forecasting the next call traffic. The call traffic is related by the previous call traffic. Hence, the incremental forecasting method we mentioned above is a good choice. First of all, we could forecast the first time bucket of call traffic, because it's features could be extracted from the preprocessed data. Then, before we forecasting the second time bucket of call traffic, we should append the call traffic of the first time bucket to the preprocessed data so that when extracting the features for the second time bucket, we can get the intraday features. Hence, we can forecast the next in the same way. In the end, we could forecast all the call traffic using the incremental method. The flowchart of Fig. 9 details the adopted forecast strategy.

5 Experimental Evaluation

In this section, we conduct an empirical study using the data set described in Sect. 4 and quantify the accuracy of the forecasts generated by the candidate models. To analyze the impact of each group of features, we compare the models with different features based on their forecasting performance. We perform detailed experimental evaluations from the following models:

- *model_d*, trained by date time features.
- *model_ds*, trained by date time and special days features.
- *model_di*, trained by date time features, intraday and interday features.
- *model_dsi*, trained by date time features, specials days, intraday and interday features.

Learning and Forecasting Period. The period we want to forecast is from 1 September, 2018, to 31 December, 2018. That is, we make forecasts for 4 months, and each month include weekends. We generate $24 \times 4 = 96$ predicted values for each day, and accuracy from *07:30* to *22:30*, which is the normal operating time of the call center. For the learning period, in order to get more training data, we set the previous years to 2 and the previous months to 11, that is all the data before the forecast month. In the *model_di* and *model_dsi*, we set the previous time buckets of intraday features to 1 and the previous days of interday features to 1. We set the *n_estimators* to 100, which is the parameter of the random forest and represents the number of trees in the forest.

Performance Measures. We quantify the accuracy of a point prediction by computing the *mean squared error* (MSE) per 15-min intervals, defined by

$$MSE = \frac{1}{N}\sum_{d,i}(V_{d,i} - \hat{V}_{d,i})^2, \tag{6}$$

where N is the total number of predictions made, $V_{d,i}$ is the volume of call traffic in the ith 15-min intervals of a given day d and $\hat{V}_{d,i}$ is the predicted value of $V_{d,i}$. Consistent with standard practice, we also consider the square root of the MSE, the *root mean squared error* (RMSE), given by

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{d,i}(V_{d,i} - \hat{V}_{d,i})^2}. \tag{7}$$

In addition to the MSE and RMSE, we compute, for a relative measure of accuracy, the *mean absolute percentage error* (MAPE) defined by

$$MAPE = \frac{1}{N}\sum_{d,i}\frac{\left|V_{d,i} - \hat{V}_{d,i}\right|}{V_{d,i}}. \tag{8}$$

Forecasting Performance. With the same parameter and training data, Table 5 shows the performance of all the models in different forecast months. In the performance of these four forecast months, *model_ds* always generates the most accurate point forecasts among all models considered. It means that the *model_ds* trained by date time features and special days fit well with the call traffic data which exhibit different types of seasonal dependencies. Compare *model_d* with *model_ds*, although *model_d* has considered the year, month, day and period of the day, and the performance is also very good, it has not considered the specials days. And the performance proves that special days features could help improve the accuracy of the forecasting. Compare *model_d* with *model_di*, the performance in November, 2018 shows that intraday and interday features have a positive effect on improving accuracy, but has no significant effect in the other three months. Hence, we can know that the call traffic mainly depends on date and time and associations between the adjacent call traffic have a little effect from the performance of *model_d*, *model_ds* and *model_di*. The *model_dsi* has the lowest accuracy in three months, it can be inferred that the specials days and intraday, interday features have conflicting relationships, which lead to a decrease in accuracy.

Table 5. Performance of each model in different months

n_estimators	Month	Model	MSE	RMSE	MAPE
100	2018.09	*model_d*	159.7	12.6	13.4
		model_ds	**133.7**	**11.7**	**12.1**
		model_di	141.2	11.9	13.5
		model_dsi	178.9	13.4	15.2
	2018.10	*model_d*	239.5	15.5	20.0
		model_ds	**170.6**	**13.1**	**17.2**
		model_di	242.9	15.6	20.9
		model_dsi	276.2	16.6	22.2
	2018.11	*model_d*	237.6	15.4	15.6
		model_ds	164.6	12.8	**13.1**
		model_di	172.3	13.1	14.1
		model_dsi	**135.9**	**11.7**	13.6
	2018.12	*model_d*	213.3	14.6	18.6
		model_ds	**147.9**	**12.2**	**16.4**
		model_di	168.3	13.0	18.4
		model_dsi	203.4	14.3	19.8
	average	*model_d*	212.5	14.5	16.9
		model_ds	**154.2**	**12.5**	**14.7**
		model_di	181.2	13.4	16.7
		model_dsi	198.6	14.0	17.7

The results of this section show that the model trained by date time and special days features usually lead to the more accurate point than the other models.

6 Related Work

As far as we know, there is little work directly related to forecasting call traffic for a month. However, call arrivals forecasting has been studied in a variety of other research. The call arrivals process can be modeled as a Poisson arrival process, and has been shown to possess several features [1,2,8–10]. Moreover, call center arrivals typically exhibit a significant dispersion relative to the Poisson distribution. Thus, a doubly stochastic Poisson arrival process may be more appropriate [3,11–14]. The method based on Poisson is modeled in a day, and do not consider the seasonal dependencies. The call center arrivals exhibit different types of dependencies. A reasonable forecasting model needs to account appropriately for some or all of the types of dependencies that exist in real data [3]. In the case where the call arrival rate has intraday and interday dependencies, standard time series models may be applied for forecasting call arrivals [3], such as *autoregressive integrated moving average* (ARIMA) models and *exponential smoothing* [15]. It forecast the next value, but in our problem, we should forecast the values of each time bucket in a month. In addition, some studies have proposed fixed-effects models [13,16–18] and mixed-effects models [12,13] to account for the intraday dependence, interday dependence, and inter-type dependence of call arrivals. Dimension reduction [17,19,20] and Bayesian techniques [16,21,22] have also been adopted in the literature. Although it takes the intraday and interday features into consideration, the other seasonal dependencies, e.g., date time dependencies, are ignored. Many forecasting models assume that specials days are outliers, and remove such days [23], or describe the application of singular vector decomposition for outlier detection but provide no empirical evaluation [20]. However, many other studies avoid this problem by assuming the data pre-cleansed [11,18,24–26]. However, in our problem, the period we forecast is one month, including special days, such as weekends, holidays.

In summary, the problem we want to solve in this paper could be distinguished from previous research from four main aspects:

– As far as we know, there is little work had been done on the problem of forecasting long-term call traffic of multiple time buckets.
– Different types of seasonal dependencies are not all considered.
– We choose the RF to train model, which is a supervised learning algorithm.

7 Conclusions

In this paper, we proposed the problem of forecasting long-term call traffic of multiple time buckets, which has not been well addressed so far. In order to solve this problem, we first take the call arrival rate and the average holding time into consideration. Then we extracted three groups of features, named date time features, special days features, intraday and interday features. Next we trained the models based on the features by using the method of the RF. At last, we proposed two strategies to forecast the call traffic based on whether taking into account the intraday and interday features. According to the experiments, we obtained following conclusions:

– All features we extracted work well in our problem;
– The intraday and interday features have a little effect on the performance.
– The second model, trained by date time and special days features is the best choice for our problem.

For future work, we are going to further study the dependencies of the call traffic and explore more reliable features to improve the quality of forecasting.

Acknowledgment. This research was partially supported by the National Key Research and Development Program of China (2018YFB1402802).

References

1. Aksin, Z., Armony, M., Mehrotra, V.: The modern call center: a multi-disciplinary perspective on operations management research. Prod. Oper. Manage. **16**(6), 665–688 (2007)
2. Cezik, M.T., L'Ecuyer, P.: Staffing multiskill call centers via linear programming and simulation. Manage. Sci. **54**(2), 310–323 (2008)
3. Ibrahim, R., Ye, H., L'Ecuyer, P., et al.: Modeling and forecasting call center arrivals: a literature survey and a case study. Int. J. Forecast. **32**(3), 865–874 (2016)
4. Oreshkin, B.N., Réegnard, N., L'Ecuyer, P.: Rate-based daily arrival process models with application to call centers. Oper. Res. **64**(2), 510–527 (2016)
5. Barrow, D.K.: Forecasting intraday call arrivals using the seasonal moving average method. J. Bus. Res. **69**(12), 6088–6096 (2016)
6. Rowlett R.: How many?: A dictionary of units of measurement (1999)
7. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
8. Garnett, O., Mandelbaum, A., Reiman, M.: Designing a call center with impatient customers. Manuf. Serv. Oper. Manage. **4**(3), 208–227 (2002)
9. Gans, N., Koole, G., Mandelbaum, A.: Telephone call centers: tutorial, review, and research prospects. Manuf. Serv. Oper. Manage. **5**(2), 79–141 (2003)
10. Wallace, R.B., Whitt, W.: A staffing algorithm for call centers with skill-based routing. Manuf. Serv. Oper. Manage. **7**(4), 276–294 (2005)
11. Avramidis, A.N., Deslauriers, A., L'Ecuyer, P.: Modeling daily arrivals to a telephone call center. Manage. Sci. **50**(7), 896–908 (2004)
12. Aldor-Noiman, S., Feigin, P.D., Mandelbaum, A.: Workload forecasting for a call center: methodology and a case study. Ann. Appl. Stat. **3**(4), 1403–1447 (2009)
13. Ibrahim, R., L'Ecuyer, P.: Forecasting call center arrivals: fixed-effects, mixed-effects, and bivariate models. Manuf. Serv. Oper. Manage. **15**(1), 72–85 (2013)
14. Ding, S., Koole, G.: Optimal call center forecasting and staffing under arrival rate uncertainty. working paper (2014)
15. Hyndman, R., Koehler, A.B., Ord, J.K., et al.: Forecasting with Exponential Smoothing: The State Space Approach. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-71918-2
16. Weinberg, J., Brown, L.D., Stroud, J.R.: Bayesian forecasting of an inhomogeneous Poisson process with applications to call center data. J. Am. Stat. Assoc. **102**(480), 1185–1198 (2007)
17. Shen, H., Huang, J.Z.: Interday forecasting and intraday updating of call center arrivals. Manuf. Serv. Oper. Manage. **10**(3), 391–410 (2008)
18. Taylor, J.W.: A comparison of univariate time series methods for forecasting intraday arrivals at a call center. Manage. Sci. **54**(2), 253–265 (2008)
19. Shen, H., Huang, J.Z.: Forecasting time series of inhomogeneous Poisson processes with application to call center workforce management. Ann. Appl. Stat. **2**(2), 601–623 (2008)
20. Shen, H., Huang, J.Z.: Analysis of call centre arrival data using singular value decomposition. Appl. Stoch. Models Bus. Ind. **21**(3), 251–263 (2005)
21. Soyer, R., Tarimcilar, M.M.: Modeling and analysis of call center arrival data: a Bayesian approach. Manage. Sci. **54**(2), 266–278 (2008)

22. Aktekin, T., Soyer, R.: Call center arrival modeling: a Bayesian state-space approach. Naval Res. Logist. (NRL) **58**(1), 28–42 (2011)
23. Taylor, J.W., De Menezes, L.M., McSharry, P.E.: A comparison of univariate methods for forecasting electricity demand up to a day ahead. Int. J. Forecast. **22**(1), 1–16 (2006)
24. Pacheco, J., Millán-Ruiz, D., Vélez, J.L.: Neural networks for forecasting in a multi-skill call centre. In: Palmer-Brown, D., Draganova, C., Pimenidis, E., Mouratidis, H. (eds.) EANN 2009. CCIS, vol. 43, pp. 291–300. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-03969-0_27
25. Taylor, J.W.: Triple seasonal methods for short-term electricity demand forecasting. Eur. J. Oper. Res. **204**(1), 139–152 (2010)
26. Jongbloed, G., Koole, G.: Managing uncertainty in call centres using Poisson mixtures. Appl. Stoch. Models Bus. Ind. **17**(4), 307–318 (2001)