International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Analyzing Long Call Timer using ARIMA Model

Abhishek Raj Permani

IIT(ISM) Dhanbad,Jharkhand(826004),India
7-B-36 Mahaveer Nagar-III,Kota,Rajasthan(324005),India

A B S T R A C T

In this paper, I have analyzed long call timer using AutoRegressive Integrated Moving Average model to get a threshold value to end long calls. Firstly, I generated a sample data for 30 users for 25 days that contains average call duration of each user on each single day. Secondly, I visualized the data generated to find the best model for analyzing this. Finally, used ARIMA (AutoRegressive Integrated Moving Average model) for predicting the threshold value of long calls.

Keywords: Long call timer, ARIMA, predicting threshold

1. Introduction

Most of service providers have a threshold value of each call, once the call duration exceeds this threshold value, the call automatically gets disconnected, and this value cannot be very small (because this will end call very frequently, which will leave users unsatisfied) also cannot be very large (because this will increase the load on network lines), it should be optimal.

We have to find that value of time after which the user has to redial again in order to continue. Also, a fix percentage of users must be satisfied with that time value.

Satisfied Percentage of users- is the percentage of total users, who ends the call before that end call time value, which we have to calculate.

We have assumed that many users can simultaneously make phone calls over the network.

1.1. Introduction to ARIMA Model

An ARIMA model is a class of measurable models for investigating and gauging time series information.

It expressly obliges a set-up of standard constructions in time series information, and as such gives a straightforward yet incredible strategy for making talented time series figures.

ARIMA is an abbreviation that represents AutoRegressive Integrated Moving Average. It is a speculation of the easier AutoRegressive Moving Average and adds the thought of coordination.

This abbreviation is graphic, catching the critical parts of the actual model. Momentarily, they are:

- AR: Autoregression. A model that utilizes the reliant connection between a perception and some number of slacked perceptions.
- I: Integrated. The utilization of differencing of crude perceptions (for example taking away a perception from a perception at the past time step) to make the time series fixed.
- MA: Moving Average. A model that utilizes the reliance between a perception and a leftover mistake from a moving normal model applied to slacked perceptions.

Every one of these parts are expressly indicated in the model as a boundary. A standard documentation is utilized of ARIMA(p,d,q) where the boundaries are subbed with whole number qualities to rapidly demonstrate the particular ARIMA model being utilized.

* *Corresponding author.* Tel.: +0-000-000-0000 ; fax: +0-000-000-0000.
E-mail address: author@institute.xxx

The boundaries of the ARIMA model are characterized as follows

- **p**: The number of lag observations included in the model, also called the lag order.
- **d**: The number of times that the raw observations are differenced, also called the degree of differencing.
- **q**: The size of the moving average window, also called the order of moving average.

A direct relapse model is built including the predefined number and kind of terms, and the information is ready by a level of differencing to make it fixed, for example to eliminate pattern and occasional designs that contrarily influence the relapse model.
A worth of 0 can be utilized for a boundary, which shows to not utilize that component of the model. Thusly, the ARIMA model can be designed to play out the capacity of an ARMA model, and surprisingly a basic AR, I, or MA model.
Taking on an ARIMA model for a period series accepts that the basic interaction that produced the perceptions is an ARIMA cycle. This might appear glaringly evident, yet assists with inspiring the need to affirm the presumptions of the model in the crude perceptions and in the remaining mistakes of figures from the model.

Nomenclature

ARIMA AutoRegressive Integrated Moving Average
MAE Mean absolute errorMethodology
RMSE Root mean squared error

1.2. Preparing Dataset-

- Dataset of 30 users for 25 Days has been taken in consideration
- Average call duration of each user on each single day has been taken into account.
- Dataset is generated artificially using Python.

User ID	Date	Time	Duration
And an entry	01-05-2008	02:50	32
And another entry	02-05-2008	08:45	23
And another entry	03-05-2008	10:30	17

Table 1- Sample Dataset for user with User ID-1

1.3. Working on prepared Dataset-

Fig. 1 is the plot of a particular user's call duration versus date of that call. Fig. 2 is an autocorrelation plot of the time series. Also, when fitting the model, a lot of debug information is provided about the fit of the linear regression model. We can turn this off by setting the *disp* argument to 0.
Section 1.4, summarizes the coefficient values used as well as the skill of the fit on the on the in-sample observations. Fig. 4- Density plot of the residual error values, suggesting the errors are Gaussian, but may not be centred on zero. The distribution of the residual errors is displayed. The results show that indeed there is a bias in the prediction (a non-zero mean in the residuals).

1.4. ARIMA Model Results

Dep. Variable:	D.duration
No. Observations:	27
Model:	ARIMA(3, 1, 0)
Log Likelihood	-115.055
Method:	css-mle
S.D. of innovations	16.880
Date:	Thu, 23 Sep 2021
AIC	240.111
Time:	22:50:41
BIC	246.590
Sample:	05-02-2008
HQIC	242.037

```
==============================================================================
                     coef    std err        z      P>|z|       [0.025      0.975]
------------------------------------------------------------------------------
const               0.4179      1.287     0.325     0.745      -2.104       2.939
ar.L1.D.duration   -0.7309      0.184    -3.979     0.000      -1.091      -0.371
ar.L2.D.duration   -0.5816      0.203    -2.872     0.004      -0.979      -0.185
ar.L3.D.duration   -0.3311      0.193    -1.713     0.087      -0.710       0.048
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1           -0.0916           -1.3824j            1.3854           -0.2605
AR.2           -0.0916           +1.3824j            1.3854            0.2605
AR.3           -1.5737           -0.0000j            1.5737           -0.5000
```

```
              0
count  27.000000
mean   -0.441215
std    17.266405
min   -34.393738
25%    -7.747248
50%    -1.620101
75%     6.858619
max    47.286828
predicted=43.096186, expected=31.500000
predicted=37.267813, expected=29.000000
```

1.5. Mathematics of ARIMA Model

A non-seasonal ARIMA model is classified as an "ARIMA(p,d,q)" model, where:

- p is the number of autoregressive terms,
- d is the number of nonseasonal differences needed for stationarity, and
- q is the number of lagged forecast errors in the prediction equation.

The forecasting equation is constructed as follows. First, let y denote the d^{th} difference of Y, which means:

If d=0: $y_t = Y_t$

If d=1: $y_t = Y_t - Y_{t-1}$

If d=2: $y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$

Note that the second difference of Y (the d=2 case) is not the difference from 2 periods ago. Rather, it is the *first-difference-of-the-first difference*, which is the discrete analog of a second derivative, i.e., the local acceleration of the series rather than its local trend.

In terms of y, the general forecasting equation is equation (1).

After performing above calculations, and observing the above dataset, we will now forecast the values of call duration of a particular user shown in Fig. 5. A rolling forecast is required given the dependence on observations in prior time steps for differencing and the AR model. A crude way to perform this rolling forecast is to re-create the ARIMA model after each new observation is received. In Fig. 5, a line plot is created showing the expected values (blue) compared to the rolling forecast predictions (red). We can see the values show some trend and are in the correct scale.

Manually keep track of all observations in a list called history that is seeded with the training data and to which new observations are appended each iteration.

Fig. 3 is the list of predicted and expected values of call duration average of all 30 users taken into account, these values will be used for calculating - Root mean squared error (RMSE) value for dataset.

1.6. Errors

The forecast errors are on the same scale as the data. Accuracy measures that are based only on e_t are therefore scale-dependent and cannot be used to make comparisons between series that involve different units.

The two most commonly used scale-dependent measures are based on the absolute errors or squared errors, are shown in equation (2) and (3).

(where e_t is the error term)

When comparing forecast methods applied to a single time series, or to several time series with the same units, the MAE is popular as it is easy to both understand and compute. A forecast method that minimizes the MAE will lead to forecasts of the median, while minimizing the RMSE will lead to forecasts of the mean. Consequently, the RMSE is also widely used, despite being more difficult to interpret.

2. Illustrations

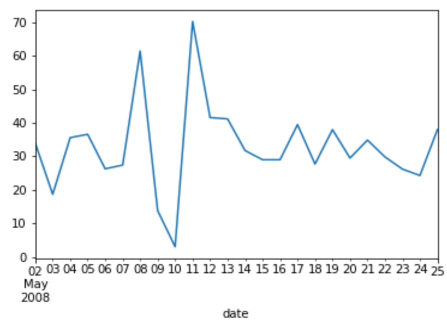

Fig. 1

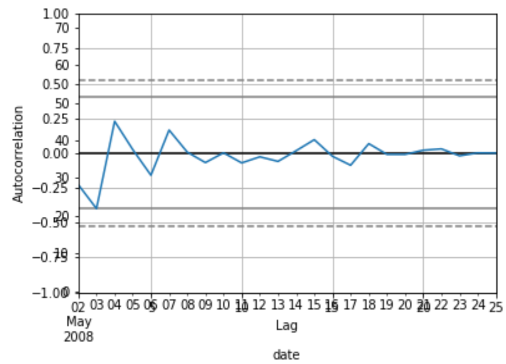

Fig. 2

	A	B
1	predicted	expected
2	32.704698	26.2
3	31.880349	24.285714
4	29.85682	38
5	28.876742	30
6	22.406223	15
7	20.310771	19.2
8	15.916917	25.75
9	26.10319	27
10	21.870625	16.5

Fig. 3

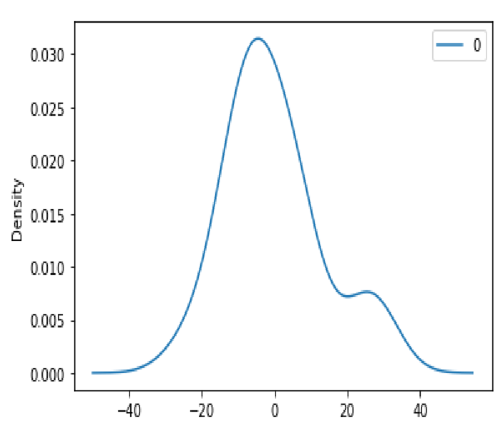

Fig. 4

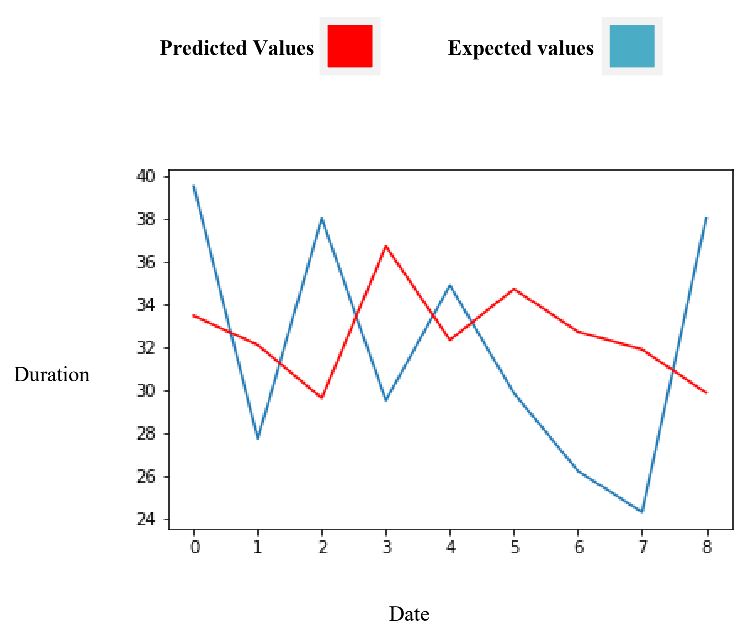

Fig. 5

3. Equations

$\hat{y}_t = \mu + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q}$ (1)

$\mathrm{MAE} = \mathrm{mean}(|e_t|)$ (2)

$\mathrm{RMSE} = \sqrt{\mathrm{mean}(e_t^2)}$ (3)

4. Conclusion

The RMSE value for the above Dataset comes out to be= 15.879966352

Once we have obtained the forecasted data using above ARIMA Model, we can calculate the threshold value. For the above dataset taken into consideration, the threshold time value to satisfy 95% of the users comes out to be = 55.09 minutes.

References

S. C. Hillmer & G. C. Tiao (2012). An ARIMA-Model-Based Approach to Seasonal Adjustment
Adebiyi A. Ariyo; Adewumi O. Adewumi; Charles K. Ayo…(2014),Stock Price Prediction Using the ARIMA Model, 14934124