International Journal of Scientific & Engineering Research, Volume 10, Issue 9 Edition, September-2019
ISSN 2229-5518

Foresight – An intelligent forecasting engine

Avinash Karthik Sundarraj, Rishi Kumar

Abstract— Keeping pace with the advancement in technology that helps provide ground breaking solutions quickly, organizations force themselves to constantly search for smart and effective solutions to deliver without negatively impacting client satisfaction. This requires organizations to evolve and break loose from their primitive modus operandi. Organizations are now facing difficulty to service higher demand and with shorter response time which has led then to explore automation opportunities through artificial intelligence engines (AIEs) which also aids in decision making. The incorporations of these AIEs have revolutionized the industry to the extent that it has now become an integral part of any data driven decision making.

With cost savings at the helm of the organization's goals, predicting demand accurately is rapidly gaining importance. As is the case with call centers where-in the focus has shifted to identify the demand at a daily and at an intraday level. The Delphi method may have worked in the past, but with the increase in the complexity of data the identification of micro trend often becomes a challenge. The requirement of a more robust methodology which not only identifies the nature of the data but also reduces the human effort in the process is prevalent.

The developed algorithm employs Machine Learning techniques on R to deconstruct the past call-volume data thereby learning its seasonality and trend and then generates highly accurate forecast in a matter of minutes. This would help organizations to stay ahead of the curve and constantly evolve to the ever-changing market.

Index Terms— Automation, Neural network, Prophet, Machine learning, Manual effort, Forecasting, Data manipulation, R.

1 BUSINESS PROBLEM

THIS The need, in most industries, is to leverage the resourcefulness of technology to enhance user and/or stake holder experience. Not to mention the effort to reduce the cost incurred in service delivery. Both above asks can be achieved through meticulous research and efficient delivery. However, to do so, would incur a significant amount of time. To explain this in the client service industry, the call centers that are there to aid customers must operate at optimum efficiency i.e., they have to work on fulfilling customer requests efficiently and also not burn out in the process. This requires extensive planning in terms of the volume of requests (in our case: calls) coming in.

To obtain accurate call volume forecast, a forecaster should be privy to the nature of calls coming in, the day of the week they are hitting the queue and if there are any events and/or initiatives that might change the arrival pattern. To breakdown the historical data (time series data) and understand its seasonality and trend requires the employment of various statistical models. Once the nature of the time series data is decrypted and further analyzed, the forecaster is now able to predict its arrival pattern for the future by using methods as simple as weighted averages or a more complex multiple regression. The choice of the methodology is completely dependent of the nature of the data set. Once the choice is made and the forecasts are generated, it is tested with the actual call volume to evaluate the accuracy of the model. The model with the highest accuracy is then chosen and its output is sent out as the demand.

This exhaustive process is not only time consuming but also subjected to change as the data begins to evolve. Time is a commodity that organizations aren't likely to dispense. And as the data starts getting complex, it becomes strenuous for the forecaster to handle. In order to compete with fellow service providers, organizations are tightening service level agreements (SLA) to improve customer satisfaction. This narrows the tolerance interval which now would call for greater accuracy. To tackle these two pressing issues, an advanced automated forecasting system needs to be developed and deployed.

2 AUTOMATED FORECASTING

2.1 Data Collection and Processing

The information is fed in the form of a flat file which is populated from our local repository containing daily call volume data. This is then flagged for special days (day before holidays, day after holidays and the holiday). Weekends are added to the dataset to maintain weekly seasonality. Now the input dataset is ready to be read by the algorithm.

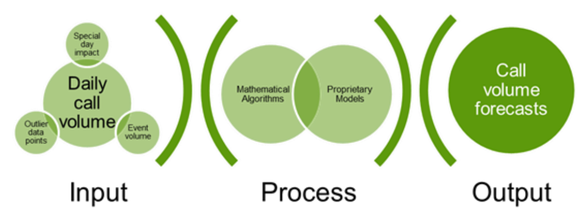

Fig 1: Workflow of automated forecasting

2.2 Forecast Generation

The data includes outliers, holiday impact and event volumes. This is then subjected to a series of algorithms and models to arrive at the final output. The mathematical models include namely three models: a. ARIMAX, b. Neural Network and c. Prophet. Along side these algorithms, we have included three proprietary models taking some of the best practices in

International Journal of Scientific & Engineering Research, Volume 10, Issue 9 Edition, September-2019
ISSN 2229-5518

forecasting that has been observed over the last few years.

The first algorithm adopted was ARIMAX as it has proven time and again to provide high accuracy forecast consistently. This model uses Alkaike Information Criterion (the model to have a unit root and for the time series to be stationary) and Maximum Likelihood Explanation to choose the best model. The Exogeneous variables, among others, is generated using a package called timetk which determines the place of the data point (day of the week or week of the month) from its time stamp. Like most ARIMAX models, this plots Autocorrelation function plots and Partial Autocorrelation function plots to determine the order of the model. After which it uses AIC to determine the best model then plots ACF to the residuals for the chosen model and after the residuals start behaving like white noise, the model predicts its forecast.

The second algorithm is that of Neural Network using a package called nnetar from the forecast package offered which is a feed forward neural network with a single hidden layer and lagged inputs for forecasting univariate time series. Its pressure point is that it deciphers special weeks (holidays) and weekly criticality with greater accuracy. This model was chosen after experimenting with LSTM which provided lower accuracy. The model trains several networks with random initial weights to inputs, which are basically lagged values of the series itself. The it applies Box-Cox transformation to the series if explicitly mentioned for minimizing outlier effects. Lastly, it creates an ensemble of the outputs of all these networks to arrive at a comprehensive output.

The third algorithm is Prophet. This is a package incepted and developed by Facebook and includes many different forecasting techniques (ARIMA, exponential smoothing, etc.), each with their own strengths, weaknesses, and tuning parameters. There are smoothing parameters for seasonality that allow you to adjust how closely to fit historical cycles, as well as smoothing parameters for trends that allow you to adjust how aggressively to follow historical trend changes. This is deployed by a package called prophet which adopts a logic like arithmetic regression model. First, it identifies logistic growth curve by selecting change points in the data. Then the yearly trend is traced using Fourier series and weekly by dummy variables to obtain the final forecast.

The three proprietary models are those developed in house taking some of the best practices in forecasting. These models have various assumptions in place to arrive at a method to forecast. The proprietary models are those which have unique logic which a forecaster takes into consideration. This includes: outlier treatment, seasonality, trend, first week of the month, weekly seasonality, week over week change and day over day change. By including these various attributes, the generated forecast indeed, a very high accuracy with consistency.

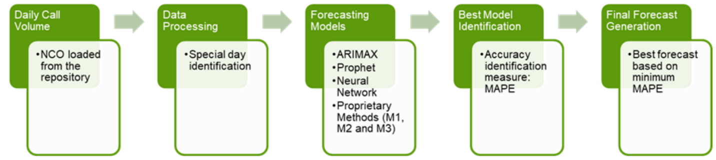

Fig 2: Automated forecasting process

The input data is divided into train and test data sets. The train usually consists of the 80% if the historical time series data. The train data set is passed through all the six above mentioned algorithms and models to obtain six sets of forecasts namely ARIMAX, NN (Neural Network), Prophet, M1, M2 and M3 (proprietary models). An ensemble (E1) forecast is calculated as well. Thus, a set of seven forecasts are produced in a single run

2.3 Model Selection

To test the performance of these seven models, each of their forecasts is compared to the test data set. MAPE (mean average percentage error) is used to measure accuracy. The final stage of forecast generation is to obtain a second ensemble: E2 using an in-house logic. E2 is then compared with the test data to obtain its MAPE. Now all the eight forecasts are stacked in ascending order of MAPE. The algorithm (Foresight) then generates the forecast from the best model it selects.

2.4 Results

With the model tending to learn after each iteration, we have been observing improved forecast accuracy for highly volatile time series as well. The model also has an acute understanding of micro trends which is often missed out in manual forecasting methods. With the introduction of exogeneous variables, the model can identify weekly, monthly and yearly seasonal patterns and replicate them while forecasting. Using the ensemble, we can leverage the short coming of some models over others to provide forecasts with the least error at different time periods of the year.

4 CONCLUSION

Using the advanced forecasting algorithm, we are now able to predict demand more accurately with only a fraction of effort. The engineering of this algorithm is designed in such a way that it adapts its model selection as and when the da-ta set develops a new trend. This shortens its learning with every iteration and enables it to identify error quicker and eliminate it soon after. Thus, it not only reduces manual ef-fort but also improves forecast accuracy which leads to in-formed staff planning and improved client satisfaction.

International Journal of Scientific & Engineering Research, Volume 10, Issue 9 Edition, September-2019
ISSN 2229-5518

REFERENCES

[1] https://research.fb.com/blog/2017/02/prophet-forecasting-at-scale/

[2] https://otexts.com/fpp2/nnetar.html

[3] https://beckmw.wordpress.com/2013/08/12/variable-importance-in-neural-networks/

- *Avinash Karthik Sundarraj, India, PH-+91 900 308 0232. E-mail: avi.karthik@gmail.com*
- *Rishi Kumar, India, PH-+91 855 314 1058. E-mail: kmr.rishi@yahoo.in*

(*This information is optional; change it according to your need.)*

Disclaimer: The views, opinions, findings or recommendations expressed in this paper are strictly those of the author. They do not necessarily reflect the views of any specific industry.