Noname manuscript No.
(will be inserted by the editor)

Analysis of Telecom Service Operation Behavior with Time Series

Civantos I. · García-Algarra J.

Received: date / Accepted: date

Abstract Operation of complex telecom services is a field that mixes technology, processes and teams. Despite the existence of detailed protocols and automation, the real behavior is hard to measure and predict. The human factor is a source of uncertainty, and this fact is of special relevance when facing stressful situations. Informal team working culture, time shifts or external stress are main sources of change. In this research we use time series analysis as a statistical proxy to detect this kind of drift in teams that solve network failures if three live services: IPTV, Cloud Infrastructure and IoT. This task known as incident management. This would provide not only a numerical evidence of the uncertainty in troubleshooting of digital services but also an assessment about the economic and operational impact of service releases. Changes in best fitting models may reflect different informal work cultures among the operation teams.

Keywords Time series · Telecom Services · Team behavior · Case management

1 Introduction

Service operation involves several human teams. Historical data show that measuring performance and conformance checking is not straightforward because of different operational tools and cultures. Incident management is a critical task in any telecom service, and protocols have to take into account technical and organizational details of each service.

Civantos I.
Telefónica I+D
E-mail: iciar.civantosgomez@telefonica.com

Garcia-Algarra J.
Universidad Pontificia Comillas (ICADE)
E-mail: fjgalgarra@comillas.edu

This research is focused on reaching a better understanding of operation performance differences among teams. The main objective is not predicting future incidents but finding remarkable changes in the models so we could identify them as significant events in human operational work teams. The starting point is that incident management data can be collected and treated as time series, applying statistical tools to characterize them. Time series are sets of samples carried out sequentially in time, in which the order of each observation is compulsory (Box, Jenkins, Reinsel, and Ljung, 2015). They are composed by values which are linked to instants of time, so that the analysis involves the joint management of two dimensions, the variable in study itself and time. Analysis of time series allow to extract regularities observed and compare operation of each service over the time but also to make comparisons among them.

Time series analysis is a pervasive technique in the telecom business, usually for forecasting purposes (Taylor, 2008a; Ye, 2010; Wang, Wang, Wang, and Wei, 2015). It is not common as a procedure to detect behavioral patterns, despite its application in other fields, such as forensics science (Pridemore and Chamlin, 2006), econometry (Frye and Gordon, 1980) or process control engineering (Alwan and Roberts, 1988).

We have studied the incident management of three live services deployed by Telefónica in Europe and Latin America: IPTV, Cloud Infrastructure and managed IoT connectivity.

The paper is organized as follows. Section 3 describes the dataset and the models we have use to fit the data series and assess external events impact. Section 4 shows the results of the application of different possible models to the three studied services. Finally, in section 4 we present our conclusions.

2 Materials and methods

The starting point to analyze the behavior of operation teams is the digital footprint they leave in the case management tool. This raw information is rich enough to build different types of time series, depending on the chosen indicator: response time, solution time, number of operators that deal with an individual ticket and so on. For our purposes, we have studied the number of individual open incidents per day and per operation team. In this section we describe the aggregation procedure and the data processing methodology.

2.1 Dataset

Data analytics is a powerful tool to discover operation patterns. In particular, the digital footprint that individual users leave in the internal management systems is the basis for powerful discovery techniques, such as Process Mining (Van der Aalst, 2011).

Raw data are extracted from ticketing system (called **UDo**, an in-house development). Each atomic operation (change of status, transfer, change of responsible...) is recorded with the operator, operation group identification and time stamp data. In the overall structure of operation there are two main roles, requester (the individual that issues a ticket) and operator (the set of individuals that solve it). Depending on the service and on the expertise, an individual may be part of different operation groups. Teams are sets of humans, but roles are encoded as operation groups. This information is enough to apply Process Mining techniques, as in fact we do on a daily basis (Sani, van der Aalst, Bolt, and García-Algarra, 2017), but long term evolution is difficult to detect.

Preprocessing produces a clean log of events of about 25.000 incidents per year. Final user claims are recorded in the front line systems of each country, and local operation teams only issue tickets to the global operation center when there is an evidence based belief of infrastructure malfunction. We have focused the study in the four services with more activity: TV over IP, internal Cloud Infrastructures and two Machine to Machine services.

2.2 Models

The goal of this research is to reach a better knowledge on how different team protocols or external events may affect operational performance in terms of time to fix the problem or in number of issued tickets. After preprocessing, the second step is to find out the model that best describes the time series properties. Our hypothesis is that sharp changes in the parameters of this model may be a symptom of drifts in operation.

The analysis tests different models to find out which one provides the best fit. The idea of making comparisons of different forecasts implicitly assumes that there is not an uniform process and that behavior evolves in time. One model would not be good enough for all services and for all teams. Under this assumption, the best forecasting model is a distinctive mark. Firstly, several well-studied forecasting classical approaches were tested to find out drifts in operation teams. We started with ARMA, ARIMA, exponential smoothing ARIMA and seasonal ARIMA models (Valipour, 2015).

Moreover, we decided to analyze how the vector autoregression (VAR) performs when fitting univariate time series. The vector autoregression is a Bayesian model which has proven to be especially useful for describing the dynamic behavior of economic and financial time series (Zivot and Wang, 2006). It often provides superior forecasts to those from univariate time series models and elaborate theory-based simultaneous equations models. Forecasts from VAR models are quite flexible because they can be made conditional on the potential future paths of specified variables in the model. In case of univariate time series, forecasting from a VAR model is similar to forecasting from a univariate ARMA model.

The adopted procedure when implementing these models is the following. Time series are decomposed in three parts, two of them are referred to seasonality and trend patterns, and the last one is a purely random part, known as the residual component. Then, time series could be represented as a linear combination of past time components and residuals (Cleveland, Cleveland, and Terpenning, 1990).

Finally, additional modeling tests were implemented by means of the Prophet automatic forecast. Prophet is an automatic procedure designed by Facebook for forecasting time series data (Medina, Montaner, Tárraga, and Dopazo, 2006). It is based on an additive model where non-linear trends are fit with yearly and weekly seasonality, plus holidays. It works best with daily periodicity data with at least one year of historical data. Prophet is robust to missing data, shifts in the trend, and large outliers. Despite the possibilities this package offers to fit time series, it does not deploy such accurate inferences when applied to incidents data so this approach was decided not to be included in results section.

The accuracy of forecasts is evaluated using a variety of absolute and relative error indicators. It would help to identify the best fitting model for each studied service and measure the ability of each model to capture every work team behavior. To this aim we made use of several indicators, as follows: Mean Absolute Error (MAE), Residual Sum of Squares (RSS), Mean Square Error (MSE) and Root Mean Square Error (RMSE), maximum absolute error (MAXAE) and minimum absolute error (MINAE). Among these, the mean absolute error, the root mean square error, the residual sum of squares, and the mean square error are considered the most relevant and, therefore, presented in all the forecasting cases that follow. These four error indicators are used in the Appendix. As the volume of incidents differs among the four chosen services, absolute errors are only useful to make comparisons of model variation over time for each series. Relative magnitudes, on the other hand, allow to make the assessment among them.

We have used 18 months of data to have a a common prediction period for all the services. This period should be long enough to produce meaningful results but considering the objective of fitting the underlying time series within an acceptable error rate. As we stated in the introduction, operation conditions are not stationary but drift slowly. Long term predictions are not meaningful in this environment compared to other fields where time series are applied with high success (energy demand, commodities production, etc.) (Taylor, 2008b; Alvarez, Troncoso, Riquelme, and Ruiz, 2011; Marcellino, Stock, and Watson, 2006).

To this aim we choose incidents from January to July to fit the desired models. Then, based on these time series predictions from August to December were started up. Initially, as raw data were daily collected forecasts were carried out considering periods ranging from seven to thirty days. However, amounts of daily tickets were not significant in terms of building a suitable model. That is why available data were weekly grouped and forecasting periods were chosen to comprise from one to four weeks. According to this, models accuracy was measured taking into account the error indicators previously mentioned.

2.3 Causal Impact

It is very useful to measure the effect of the event, but quite complicated, because it is a counterfactual. A recent development offers this kind of tool to predict the evolution of the time series in the absence of the extraordinary event, using another uncorrelated process as the counterfactual basis. Causal Impact is a Google `R` package that, given a response and a set of control time series, constructs a Bayesian structural time-series model (Brodersen, Gallusser, Koehler, Remy, Scott et al, 2015). The model is then used to try and predict the counterfactual, i.e., how the response metric would have evolved after the intervention if the intervention had never occurred. In our research we measure the impact of exceptional events on one service taking as control series that of another one that we know was immune to that change.

3 Results

TV over IP is quite complex with central infrastructure in Spain and Argentina and local deployments over a dozen of countries. Operation teams are full time assigned to this service. The two Machine to Machine instances share a small percentage of operating engineers, so their behavior should have some degree of correlation, but they are fully independent of the video teams. Finally, the Cloud Infrastructure service has a strong centralized operation. It is a good control series because its work load is less dependent on fast commercial deployments.

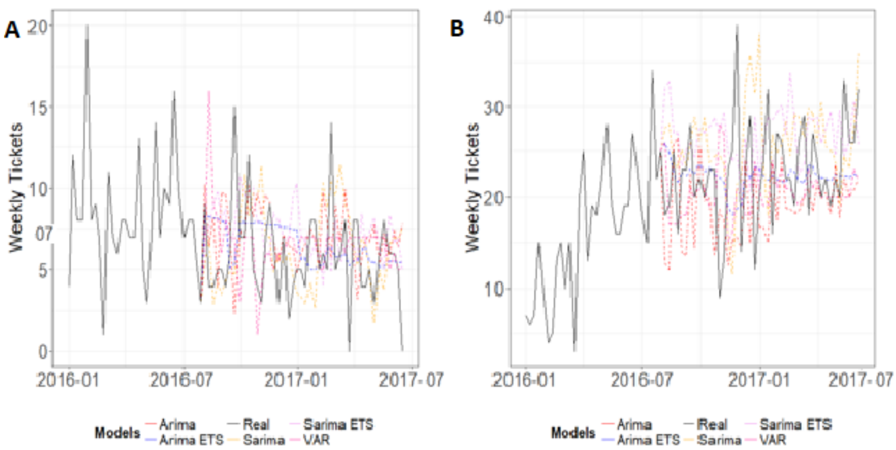

Fig. 1 A: Cloud Infrastructure tickets forecasts using August as the starting point. The solid line is the real number of incidents for the given service whereas forecasts are plotted as color dashed lines. B: TV over IP tickets forecasts using August as the starting point

For Cloud service, ETS ARIMA model with no autoregressive coefficients but third order moving average component is the top performer (Table 1). Since incidents in this service presented so regular behavior with an important seasonal component, selecting a moving average model would help to detect monthly raises.

On the contrary, finding a suitable model for the TV over IP service was not so straightforward due to the step shaped time series. The roll out in new countries during the first quarter of 2016 is the origin of this increase of the absolute number of managed claims. This change in operation conditions implied also a new structure of working teams.

We found out that a vector autoregression of time series is the best fit this model. Although forecasting from a VAR model is similar to forecasting from a univariate autoregressive model, this algorithm resulted to be more accurate than classical approaches when catching particular features of this service due to it generates not only a forecast but a complete, multivariate probability distribution for future outcomes that appears to be more realistic than those generated by other approaches. Unless the Cloud Infrastructure service, in which VAR model gave place to a fitted model that mainly consisted in the mean value of the original time series, in this case VAR forecasting model was able to accurately emulate service evolution.

Machine to Machine services, were independently analyzed. The first service deals with a lower number of incidents compared to the other three. The irregular shape, with monthly increases and decreases, imposed the necessity of using a seasonal ARIMA model. This model was composed of a first order autoregressive element as well as a differencing step for the integrated component. However, we also found out that an exponential smooth of time series was the best t this model. Then an exponential smoothing of the ARIMA (1,1,0) model was applied. Using this approach a rst order autoregressive integrated model would explain this first Machine to Machine service behavior.

On the other hand, the second service did not display such a predictable performance. Several releases deployed during last year as well as the incorporation of half million users from Germany to the service resulted in both abrupt and sudden changes. For this reason, a vector autoregressive model was not enough to fulfill this particular behavior. According to this we made us of the vector autoregression model to fit irregular time series.

Service	**Model**
Cloud Infrastructure	ARIMA ETS (0,0,3)
TV over IP	Vector Autoregression (VAR)
Machine to Machine 1	ARIMA ETS (1,1,0)
Machine to Machine 2	Vector Autoregression (VAR)

Table 1 Best fitting models.

Finally we have performed the Causal Impact study in order to assess how the Machine to Machine service number 2 would have evolved without the step

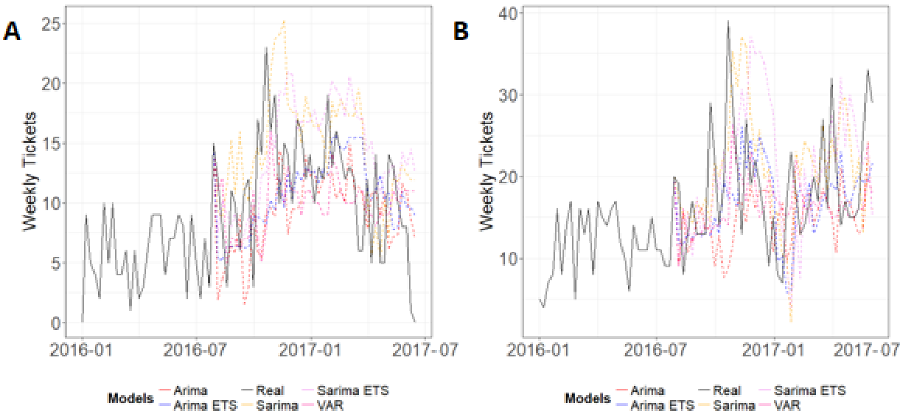

Fig. 2 A: Machine to Machine Service 1 tickets forecasts using August as the starting point. Black lines represent the real registered incidents for the given service whereas different forecasts are dashed. B: Machine to Machine Service 2 tickets forecasts using August as the starting point.

increase of users in Germany by July. The outcoming graphs from the package are composed by three panels (Fig. 3). The first panel shows the data and a counterfactual prediction for the post-treatment period. The second panel shows the difference between observed data and counterfactual predictions. This is the point wise causal effect, as estimated by the model. The third panel adds these contributions from the second panel, resulting in a plot of the cumulative effect of the intervention.

The evolution of the number of tickets, shows that the subjective perception of the operation team of an overload due to the event had a real basis. First one was not unexpected. The new operation structure has always a short learning period (around two months) were there is an increase in the number of issued tickets, but in this case figures nearly doubled by September. When the daily business was returning to normality, as predicted by the Causal Impact packet, two new episodes arised. Post mortem analysis found that the root cause of major problems during the first increase was due to a bottleneck. This bug only triggered by a software component under very special high load conditions. Corrective measures were taken to mitigate the effect and finally a new release solved the problem by late November.

4 Conclusions

Troubleshooting of complex digital services is a field with high uncertainty, due to the combination of many technologies and human factors. Studying four major services, we have found that the issued ticket series do not share a common underlying model (Table 1).

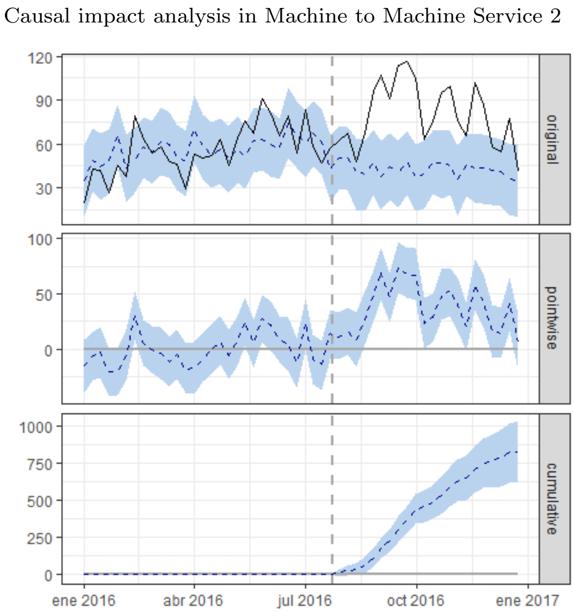

Fig. 3 Machine to Machine Service 2 tickets forecasts. In first panel, black lines represent the real registered incidents for the given service whereas blue lines represent the service evolution unless the incorporation of German users had happened.

Services less affected by explosive growth events, like Cloud and Machine to Machine 1 are quite predictable with simple ARIMA. On the contrary, TV over IP and Machine to Machine 2 are rather difficult to model and predict due to the fast changing environment conditions. Then, they required a vector autoregression modeling process to fit incidents data, which resulted into an advantage over standard univariate time-series methods. Operation teams had the perception of higher uncertainty when dealing with these services. Time series analysis provides a numerical evidence of this feeling and may be a useful tool in the allocation of human and technical resources as a function of changing conditions.

Causal impact analysis is a technique with a very high potential, because it helps to predict the counterfactual evolution of one service compared to other more stable. This feature will help to assess the economic and operational impact of unexpected and planned major service changes. This is important in a field that deals with uncertainty in a daily basis, and needs accurate measures of it to work properly.

Telecom services operation analysis has been proven to be a field that entails a range of facts to take into consideration. In order to achieve a deeper knowledge of these teams behavior it will be necessary both to take strong stands on what the boundaries are, and keep lined up with updated data. However, incidents data have not just originated by a unique point of failure but several sources of change such us time shifts or external events. Even uncertainty is regarded as a restriction, forecasting stress rates or productivity and performance indicators would become a reliable statistical proxy. As future work, outer facts predictions could be considered to find out significant events in human operational work teams.

Appendix: Numerical results

Table 2 Tickets forecast RMSE for model and service (period two weeks)

	ARIMA	ARIMA ETS	SARIMA	SARIMA ETS	VAR
Cloud Infrastructure	3,18%	2,8%	3,24%	3,5%	2,9%
TV over IP	6,9 %	6,38%	7,27%	7,17 %	5,19%
Machine to Machine 1	5,36%	4,58%	5,5%	5,51%	4,63%
Machine to Machine 2	7,4 %	6,8 %	7,71%	9,01 %	5,9%

Table 3 RMSE error for every service.

Forecast period	Cloud ARIMA ETS (0,0,3)	TV over IP VAR	M2M 2 VAR	M2M 1 ARIMA ETS (1,1,0)
One week	3,06%	4,99 %	5,73%	4,01%
Two weeks	3,08%	5,19 %	5,9%	4,58%
Three weeks	3,825%	5,95 %	6,42%	5,684%
Four weeks	4,79 %	6,78 %	12,09%	7,14%

Table 4 RSS error for every service.

Forecast period	Cloud ARIMA ETS (0,0,3)	TV over IP VAR	M2M 2 VAR	M2M 1 ARIMA ETS (1,1,0)
One week	24,03	68,75	117,2	42,04
Two weeks	25,25	76,63	123,52	56,21
Three weeks	39,375	92	367,14	77,6
Four weeks	55,72	142,73	400,91	135,08

Table 5 MSE error for every service.

Forecast period	Cloud ARIMA ETS (0,0,3)	TV over IP VAR	M2M 2 VAR	M2M 1 ARIMA ETS (1,1,0)
One week	12,01	34,37	60,38	21,02
Two weeks	12,625	38,31	61,76	28,1
Three weeks	19,69	46	105,4	38,8
Four weeks	27,86	71,36	200,45	67,54

Forecast period	Cloud ARIMA ETS (0,0,3)	TV over IP VAR	M2M 2 VAR	M2M 1 ARIMA ETS (1,1,0)
One week	2,703	4,44	5,31	3,73
Two weeks	2,75	4,68	5,67	4,28
Three weeks	4,93	7,27	7,42	9,06
Four weeks	5,13	7.43	15,375	9,21

Table 6 MAE error for every service.

References

Van der Aalst WM (2011) Process discovery: An introduction. In: Process Mining, Springer, pp 125–156

Alvarez FM, Troncoso A, Riquelme JC, Ruiz JSA (2011) Energy time series forecasting based on pattern sequence similarity. IEEE Transactions on Knowledge and Data Engineering 23(8):1230–1243

Alwan LC, Roberts HV (1988) Time-series modeling for statistical process control. Journal of Business & Economic Statistics 6(1):87–95

Box GE, Jenkins GM, Reinsel GC, Ljung GM (2015) Time series analysis: forecasting and control. John Wiley & Sons

Brodersen KH, Gallusser F, Koehler J, Remy N, Scott SL, et al (2015) Inferring causal impact using bayesian structural time-series models. The Annals of Applied Statistics 9(1):247–274

Cleveland RB, Cleveland WS, Terpenning I (1990) Stl: A seasonal-trend decomposition procedure based on loess. Journal of Official Statistics 6(1):3

Frye J, Gordon RJ (1980) Government intervention in the inflation process: The econometrics of" self-inflicted wounds"

Marcellino M, Stock JH, Watson MW (2006) A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series. Journal of econometrics 135(1):499–526

Medina I, Montaner D, Tárraga J, Dopazo J (2006) Prophet, a web-based tool for class prediction using microarray data. Bioinformatics 23(3):390–391

Pridemore WA, Chamlin MB (2006) A time-series analysis of the impact of heavy drinking on homicide and suicide mortality in russia, 1956–2002. Addiction 101(12):1719–1729

Sani MF, van der Aalst W, Bolt A, García-Algarra J (2017) Subgroup discovery in process mining. In: International Conference on Business Information Systems, Springer, pp 237–252

Taylor JW (2008a) A comparison of univariate time series methods for forecasting intraday arrivals at a call center. Management Science 54(2):253–265

Taylor SJ (2008b) Modelling financial time series. world scientific

Valipour M (2015) Long-term runoff study using sarima and arima models in the united states. Meteorological Applications 22(3):592–598

Wang M, Wang Y, Wang X, Wei Z (2015) Forecast and analyze the telecom income based on arima model. The Open Cybernetics & Systemics Journal 9(1)

Ye X (2010) The application of arima model in chinese mobile user prediction. In: Granular Computing (GrC), 2010 IEEE International Conference on, IEEE, pp 586–591

Zivot E, Wang J (2006) Vector autoregressive models for multivariate time series. Modeling Financial Time Series with S-PLUS® pp 385–429