Applying Two Theorems of Machine Learning to the Forecasting of Biweekly Arrivals at a Call Center

Michael Thrun (1), Julian Märte (2), Peter Böhme (1) and Tino Gehlert (1)
(1) Viessmann Werke GmbH & Co. KG, 35107 Allendorf
(2) Mathematics and Computer Science, University of Marburg, Germany

The forecasting of arriving calls in call centers plays a crucial role in determining appropriate staffing levels and scheduling plans[1]. Usually, the number of calls are forecasted in a period and the best forecasting method is chosen by MAPE/MAE[2] which has two problems. First, the incoming calls are dependent on the background of the call center and customer leading to the comparison of several types of problems which makes any specific choice of one forecasting algorithm impracticable[3]. Thus, we propose to change the data representation of the problem from arriving calls to issues because a low service level can result in multiple calls regarding the same topic. Second, evaluation of forecasting results is always biased if a quality measure (QM) is seen as a similarity measure between the forecast curve and the test set curve of data because any two different curves share the same number of properties(c.f.[4]). Thus, the QM should be chosen accordingly to the goal, namely, capacity planning with the key performance indicator defined as the service level[5]. A forecast smaller than the real value leads to undesired understaffing. Therefore, a forecast should be more similar, if it lies above the real value. Special events are usually known priorly by the call center manager. Hence less weight should be put on outliers.

Capacity planning of the call center using five years of daily historical data and weather data was performed by an ensemble of an additive decomposition model[6] combined with random forest regression 6]. For a forecast horizon of 14 days over a year of test data, the average forecasting quality is 91.3% outperforming[2,7]. For the decomposition model[6], all parameters were optimized w.r.t. MRE and bias[8]. This Pareto optimization problem was resolved using a radial basis function surrogate that is successively improving around areas of importance using a recursive inversion formula.

1. Ibrahim, R., et al., *Modeling and forecasting call center arrivals: A literature survey and a case study.* International Journal of Forecasting, 2016. **32**(3): p. 865-874.
2. Taylor, J.W., *A comparison of univariate time series methods for forecasting intraday arrivals at a call center.* Management Science, 2008. **54**(2): p. 253-265.
3. Wolpert, D.H., *The lack of a priori distinctions between learning algorithms.* Neural computation, 1996. **8**(7): p. 1341-1390.
4. Watanabe, S., *Knowing and Guessing: A Quantitative Study of Inference and Information*. 1969, New York, USA: John Wiley & Sons Inc.
5. Aksin, Z., M. Armony, and V. Mehrotra, *The modern call center: A multi-disciplinary perspective on operations management research.* Production and operations management, 2007. **16**(6): p. 665-688.
6. Geurts, P., D. Ernst, and L. Wehenkel, *Extremely randomized trees.* Machine learning, 2006. **63**(1): p. 3-42.
7. Ibrahim, R. and P. L'Ecuyer, *Forecasting call center arrivals: Fixed-effects, mixed-effects, and bivariate models.* Manufacturing & Service Operations Management, 2013. **15**(1): p. 72-85.
8. Kourentzes, N., J.R. Trapero, and I. Svetunkov, *Measuring the behaviour of experts on demand forecasting: a complex task*. 2014, Department of Management Science: Technical report published in the Lancaster University Management School. p. 1-23.