M. C. Thrun, J. Märte, P. Böhme and T. Gehlert

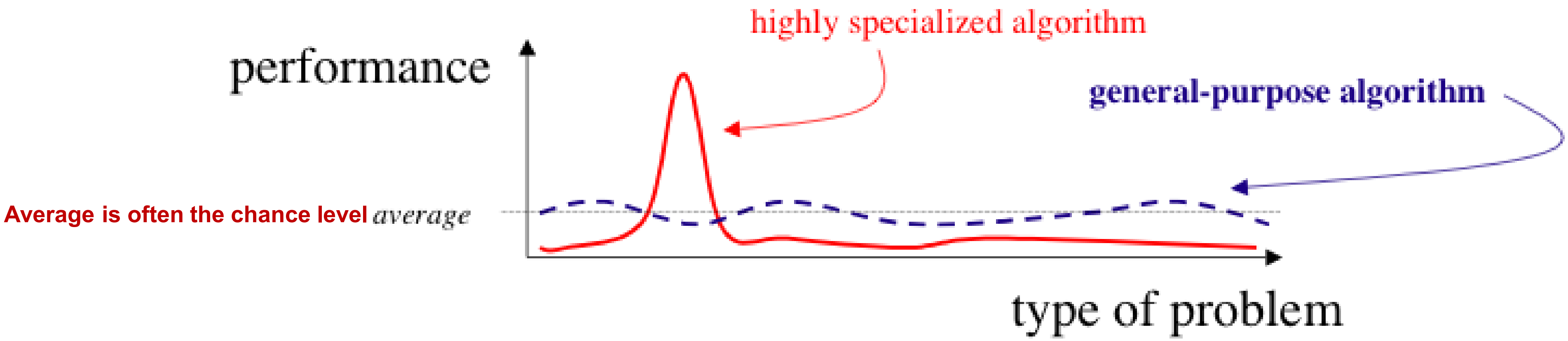
Applying Two Theorems of Machine Learning to the Forecasting of Biweekly Arrivals at a Call Center

Answering the Questions of How to Know the Appropriate Staffing Levels and Scheduling Plans Two Weeks Ahead

- Forecasting of arriving calls in call center plays a crucial role in determining appropriate staffing levels and scheduling plans [Ibrahim, 2016].
  - Installers of heating devices call the technical service for help
  - => Data is a collection of arriving calls of with a time-stamp and an information about the product type
  - Depending on the amount of incoming calls and the product type technical service employees answer the arriving calls (or is unable to answer)
- In Literature the number of calls per product is forecasted in a forecasting horizon
  - Best forecasting method is chosen by MAPE or MAE [Taylor 2008]
- This has the two following problems….

First Problem - No Free Lunch Theorem (NFL) - [Wolpert, 1996]

- 10 Timeseries of arriving calls are given, one per product type
- No solution given by an algorithm can be better than any other if the number of problems is high enough



⇒Choose the right Representation of Data meaning that you answer the question:
„What are my objects/smallest units in which I operate"?

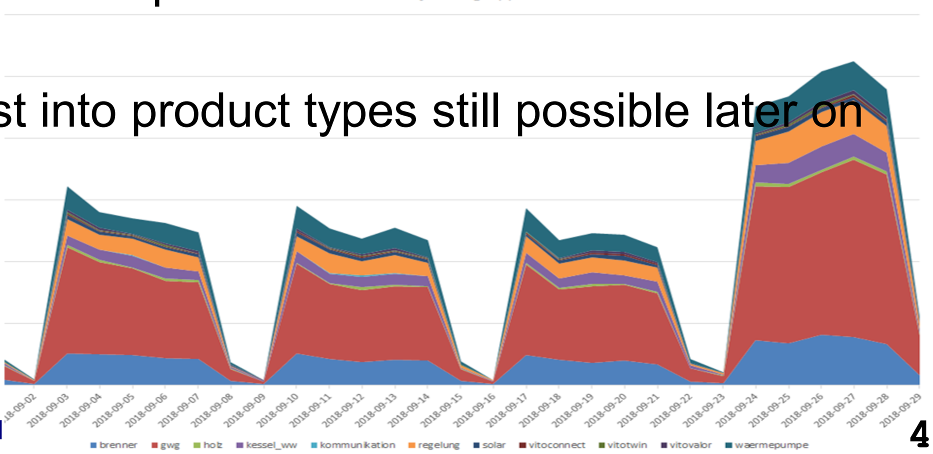
=> Select algorithm problem specific and think about data representation in order to change your problem

Consequence of NFL – Change the Problem by Preprocessing Data

- Incoming calls are dependent on the background of the call center and customer -> comparison of several types of problems -> any specific choice of one forecasting algorithm is impracticable [3]

⇒ Change the data representation of the problem from arriving calls to issues

- Not every call is relevant:
  - Low service level can result in multiple calls regarding the same topic
    - Installer calls several times a day if the service level is low
    - Service level: Proportion of arriving calls answered by the technical service during a day
- Data Aggregation of calls over all product types into one time series:
  - Several product type have high cross correlation (Lag=0)
  - 14 days forecast horizon requested but sporadic amount of calls in some product types
  - Proportional distribution of forecast into product types still possible later on (see on the left)



Selection of an adequate forecasting algorithm

Obviously:

- Depends on the available indicators, e.g. only historical data or also others
- Depends on the forecast horizon, e.g. how many days, weeks,.. in advance you require the information

But why does it

- Depend on the properties of the time series?
  - Depends on the data representation
- Depend on the choice of the quality measure?

-> Second problem

Second Problem – The Ugly Duckling Theorem (UDT) - [Watanabe, 1969]

- One of the key question in Data Science is similarity
- Relevant for every pattern recognition algorithm
- (Dis-)similarity is most often Euclidean distance, but thats most often incorrect

UDT states: classification is impossible without some sort of bias
- Depends on the features chosen
  - Example on the right

=> Similarity depends on the Representation of data

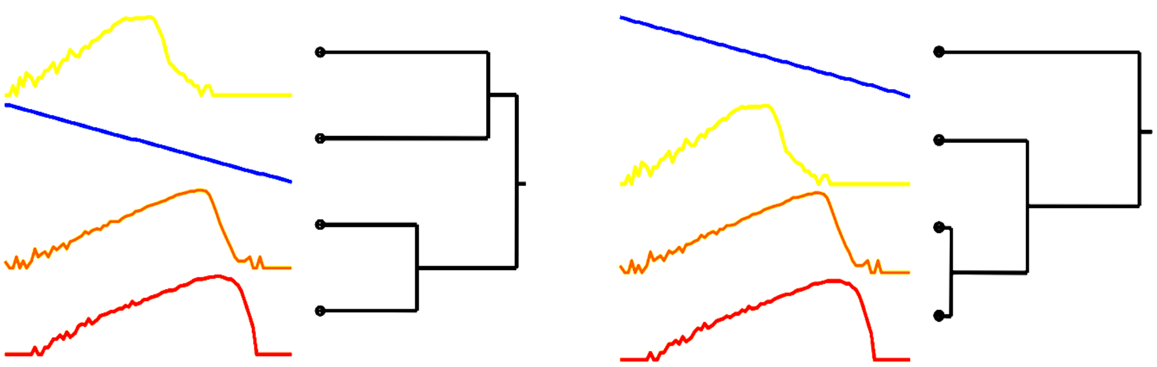
Are this two pictures is similar to each other?

Example: Evaluating the Quality of Forecasting

- Time series can be drawn like a curve
- Quality Measurement is a type of similarity measurement:
- Comparison of curve of forecasted values with curve of historical data

=> UDT: 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

=> UDT: any two different curves share the same number of properties (c.f. [4]).



Common QMs

- Scale Dependent Errors
  - Mean Error (ME)
  - Mean average error MAE/
    - Mean Absolute Deviation(MAD)
  - Mean squared error (MSE) or
    - mean squared deviation (MSD)
  - Root-mean-square-error (RMSE) or
    - root-mean-square-deviation (RMSD)
- Scale Independent Errors
  - Mean percentage error (MPE)
  - Mean absolute percentage error (MAPE)
  - Symmetric Mean absolute percentage error (SMAPE)
  - Mean absolute scaled error (MASE)
  - Mean Directional Accuracy (MDA)
  - Sum of the root errors (SRE) with Mean of the root errors (MRE) and BIAS

Note: “Mean” can be defined as average, geometric mean, median…

Proposed Choice of Quality Measure (QM)

- QM should be chosen accordingly to capacity planning
  - Key performance indicator is the service level [Aksin, 2007].
  - Special events are usually known priorly by the call center manager
  - Forecast smaller than the real value leads to undesired understaffing.

  ⇒ Forecast should be more similar, if it lies above the real value with less weight put on outliers

  ⇒ MRE and bias are chosen for optimization

Proposed Choice of Forecasting Algorithms depends on Knowledge Discovery of the Data

- Strong Seasonality in data
  - Weekly – Less work on Friday and Saturday, only special problems on Sunday
  - Yearly: Heating season is between October and March leading to a higher amount of arriving calls
  - No trend in data: In Germany Sales are constant
- => In principal an explainable decomposition model is a good choice
  - https://CRAN.R-project.org/package=prophet as a shortcut

- Weather is a good predictor
  - High Cross Correlation 3 days ahead
  - Random Forests outperformed neural networks and generalized linear models in this case
    - Because relevance of weather features (various stations distributed in germany) is unknown
    - Still weaker than decomposition model
  - https://CRAN.R-project.org/package=ranger

Details of Parameter Optimization of Prophet

- Two Quality measures were relevant: MRE and bias [Kourentzes et al.,2014]
  - Pareto optimization problem
- Direct Parameter Optimization for decomposition model was unfeasible
  - Let Q be the quality function, i.e. it takes our parameter set as input and gives the quality of the corresponding forecast as output

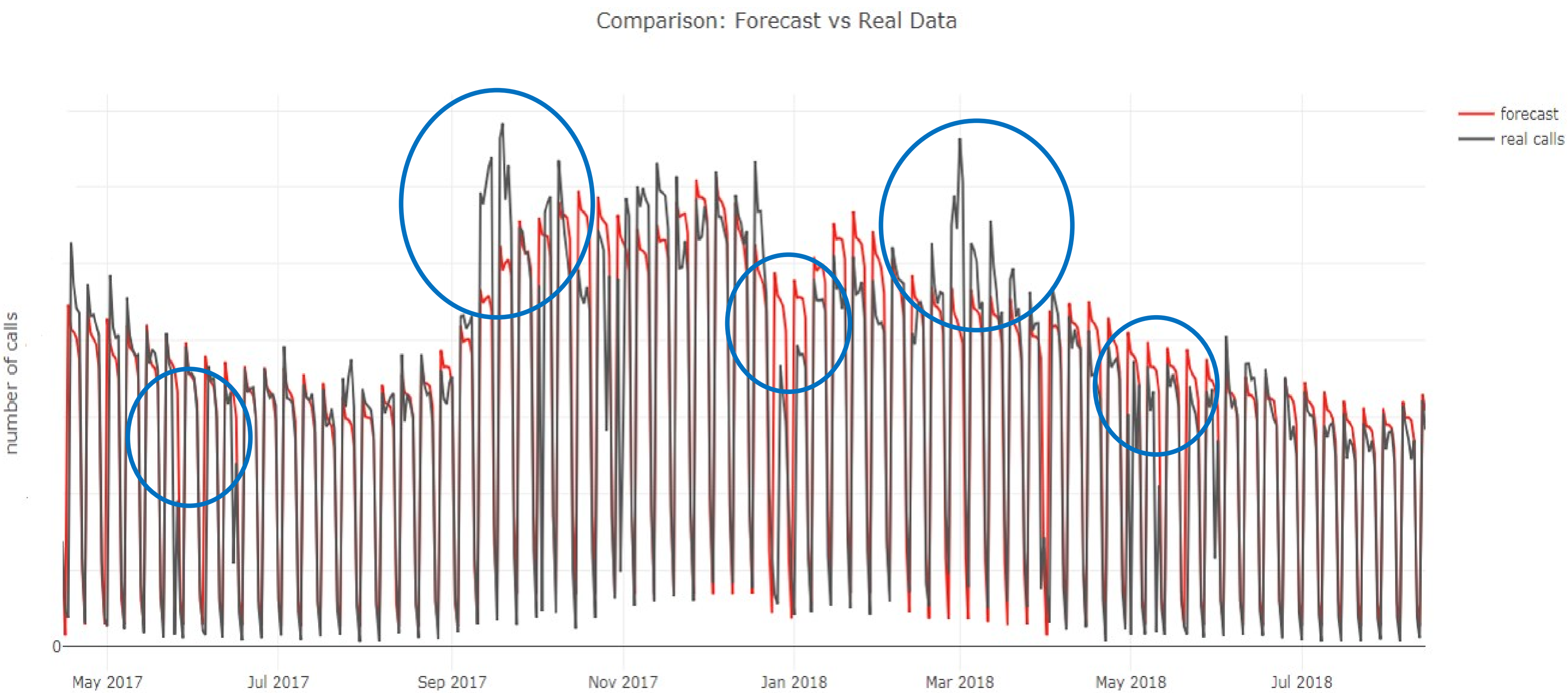
Optimization approach: Use radial basis function surrogate model:

1. Restrict paramaters to plausibe range
2. Approximate Q with a weighted sum of radial basis functions S
   - Chose an Initial set of base points around which Q will be approximated by S (Latin Hypercube)
   - Fit S to Q on the initial set by solving linear system of equations
3. Optimize S
4. Update the base point set and fit S
   - Use a recursive inversion formula instead of solving linear equations
   - Improves computation from O(n^3) to O(n^2)

-> Successively improves the parameters for the decomposition model

Decomposition Model using Prophet

- Three time periods of higher deviation between forecast and issues
  - Working Days before an after holidays, especially christmas
    - Non important because workforce manager can work around them using his experience
  - Beginning and Ending of heating season
    - Depending on weather and very important periods of time



Random Forest: Forecasting the Beginning and Ending of Heating Season using Weather

Data

- Trainings data 2013 - Okt 2017 for algorithm selection, feauture selection and parameter optimization using the two predefined QMs
    - Data preprocessed: TD issues, service level, temperatures
- Test data beginning in Okt. 2017 to Okt 2018 for Generalization

- Outdoor temperatures correlate with the volume of traffic => a forecast that takes temperatures into account should offer some benefits
    - Weather data of 15 locations in germany queried from dark sky API
        - Do not know the importance of location
        - Do not know the Temperature of a day (min, mean, max, …)
- Predictor: TD issues

Evaluation of Forecasting Results

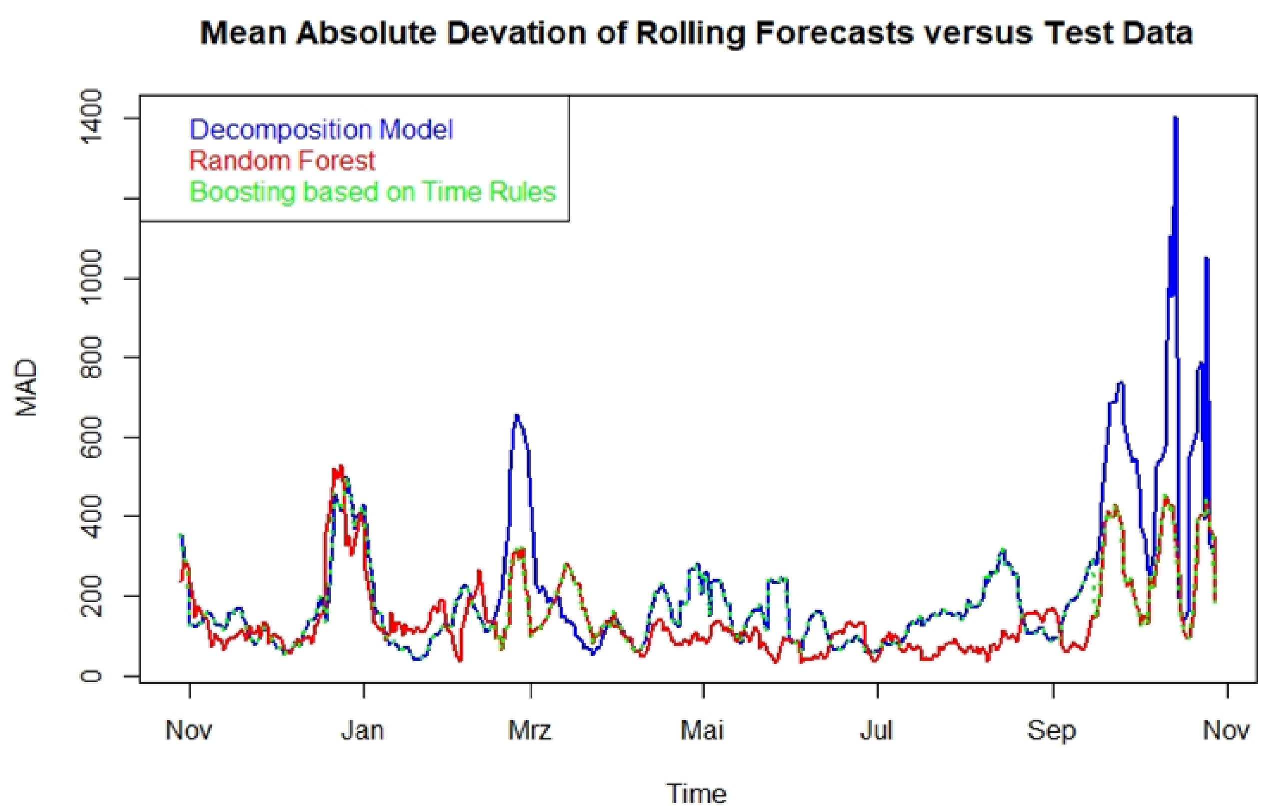
Checking the Generalization of the Ensemble Rolling forecasting

- For each method perform
- For each day use for the training of the data all data until the day before
- Forecast T days
  - For prophet compare to T days of test data with SRE and bias [Kourentzes et al., 2014] and optimize parameters
- Until 365 forecasts of ensemble are made

- T is 14 for prophet and 7 for random forest
  - weather forecast gets worse after 7 days

Ensemble with Boosting

- Random Forest forecast and prophet forecast combined to an ensemble based on time relevant rules
  - Transition of heating seasons are predicted with random forest (6 weeks in the beginning and ending of heating season)
    - Random forest adapts non-linear to temperatur changes
  - Other weeks of year decompositon model is applied



- MRE was used for evaluation
- Shown here is MAD for visualization purpose only

Discussion

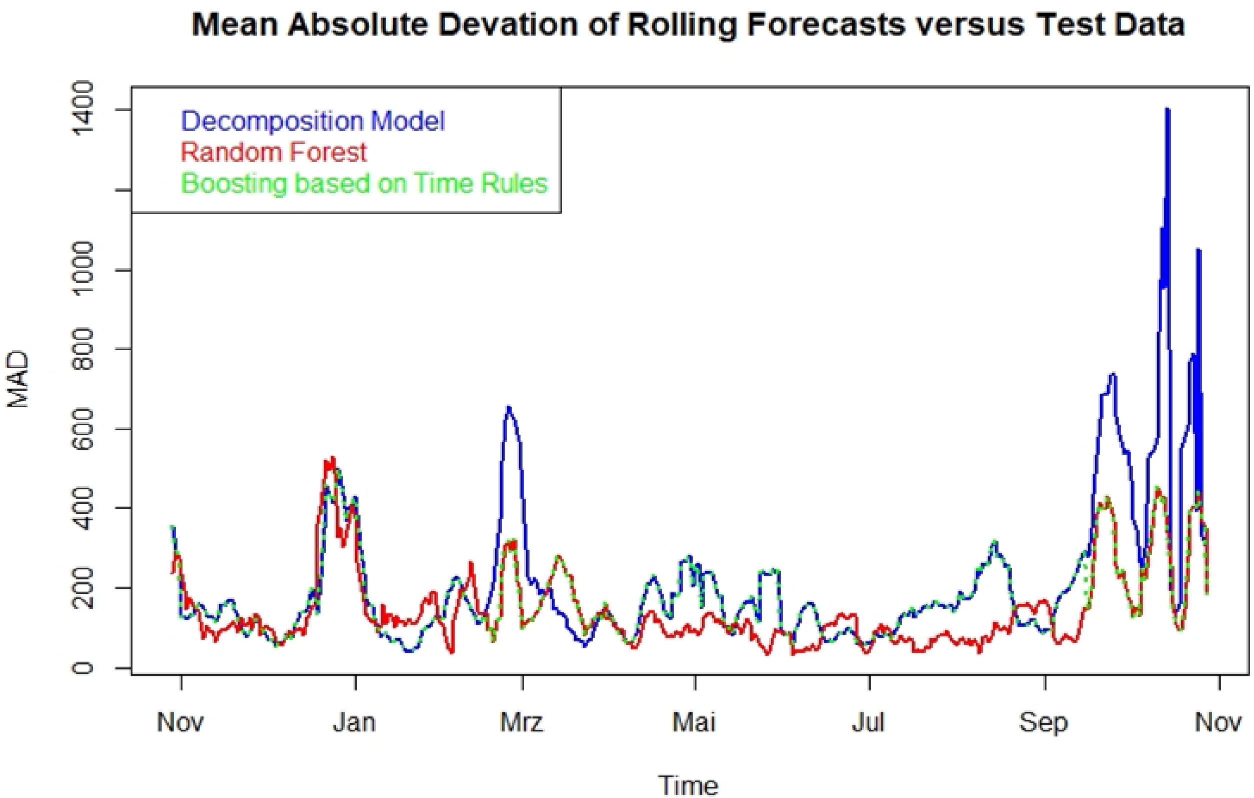
In an ensemble of decomposition model and random forest

- Bagging does not seem to provide good results
- Boosting does seem to provide good results

=> Ensemble performs better in september and march than prophet
=> Improves general quality from from 89.2% to 91.3%

Last two steps are post-processing

- Procentual distribution of issues to problem types is performed at the last step
- Capacity planning of the call center using our own modified Erlang C formel



Summary

⇒ For a forecast horizon of 14 days over a year of test data, the average forecasting quality is with 91% outperforming [Ibrahim/L'Ecuyer, 2007; Taylor, 2008; Ibrahim, 2016]

- There is no best-performing algorithm for a large variety of forecasting problems
  - Contrary to the claim of prophet [Taylor/Letham, 2017]
  - Sometimes it is better to aggregate multivariate time series, forecast this one predictor and thereafter distribute procentually
- Evaluating forecasting quality is the same as measuring the similarity between curves

  => It depends on the goal (c.f. clustering)

  => It has similar problems to clustering

Sources

General Sources:

Harvey, Andrew C., and Simon Peters. "Estimation procedures for structural time series models." *Journal of Forecasting* 9.2 (1990): 89-108

Breiman, Leo. "Random forests." *Machine learning* 45.1 (2001): 5-32.

Taylor, Sean J., and Benjamin Letham. "Forecasting at scale." *The American Statistician* 72.1 (2018): 37-45.

Wolpert, D.H., *The lack of a priori distinctions between learning algorithms.* Neural computation, 1996. 8(7): p. 1341-1390.

Watanabe, S., *Knowing and Guessing: A Quantitative Study of Inference and Information*. 1969, New York, USA: John Wiley & Sons Inc.

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.

Call Center Forecasting:

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.

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.

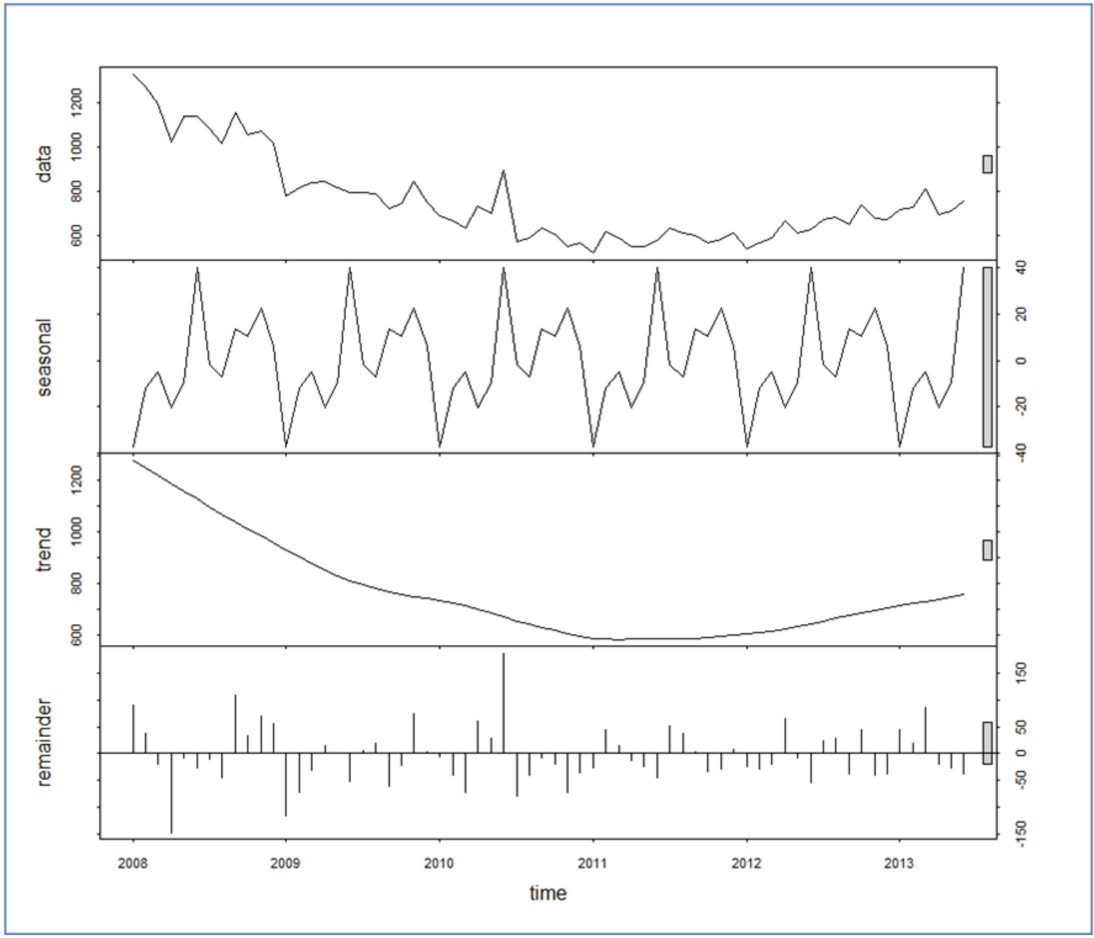
Ibrahim, R., et al., *Modeling and forecasting call center arrivals: A literature survey and a case study.* International Journal of Forecasting, 2016. 32(3): p. 865-874.

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.

Decomposition model (Komponentenmodel, [Harvey/Peters,1990])

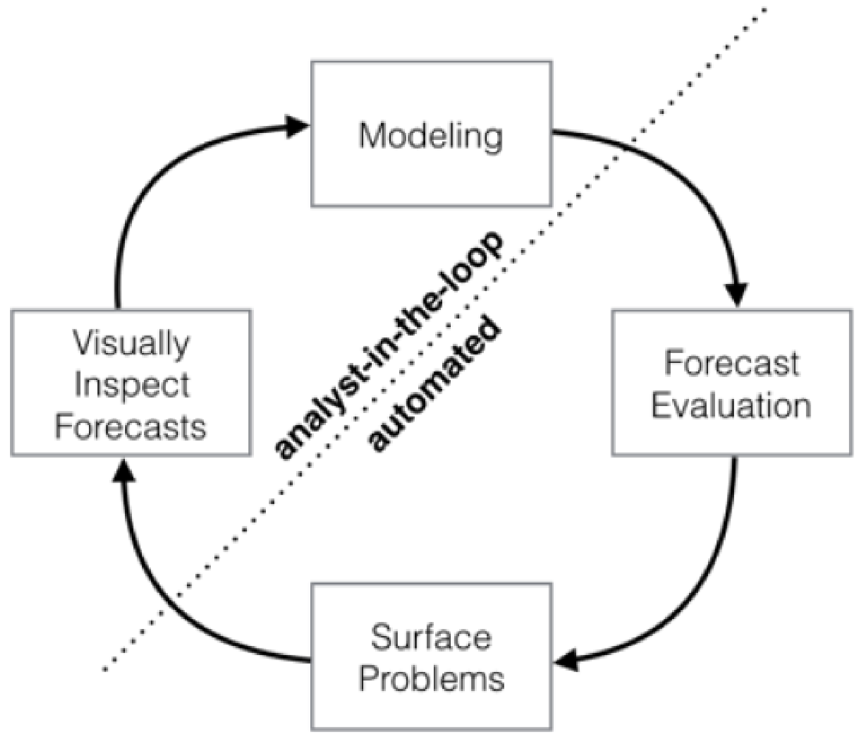
y(t)=trend(t)+seasonality(t)+holidays(t) +error(t)

- The part error(t) cannot be modeled by decomposition
  - should be normally distributed indicating randomness
- Holidays(t) represent repeating irregularities
- Seasonality(t) represents periodic values of time series y(t)
- Trend: monotone continuous line of ascending or descending values



Decomposition model of Prophet - [Taylor/Letham, 2017]

- Prophet is a decomposition model using modular regression with interpretable parameters
  - Forecasting "at scale" that combines configu
    performance analysis [Taylor/Letham, 2017]
    - Large number of non-experts doing for
    - Large variety of problems
    - Large number of automated forecasts
      poorly for human intervention
  - Regression: Transferring forecasting to curv
    - Prophet parameters are scalar factors



Taken from [Taylor/Letham, 2017]

Random Forest Method - [Breiman,2001]

- Usually random forest is a type of subsymbolic classifiers
- In forecasting random forests are used as a non-linear regression approach
- Random forests are an approach of bagging of an ensemble of simple decision trees predictors
  - Random selection of input features and than applying decision tree (i.e. CART with Gini)
  - Bagging in regression: Averaging over Tree results with the goal to reduce the variance of predicted values
    - [Breiman,2001]: “Due to strong law of large numbers, random forests always converge”
      - => “Overfitting is not a problem”
- Advantages
  - Good for datasets with many features containing only a small amount of information [Breiman, 2001]
  - Relative small training set and a large test set of data can be used [Breiman, 2001]
  - Overfitting is difficult problem in forecasting which can be disregarded here
- Disadvantages
  - Not understandable (subsymbolic)
    - Knowledge about the relevant features difficult to extract
    - Results cannot be explained in a way a decomposition model can
  - Stochastic Forecast (Slightly depends on the trail)