seer: R package for featue-based forecast model selection

Thiyanga S Talagala
Rob J Hyndman
George Athanasopoulos

Monash University, Australia

UseR, 2018

Large collections of time series

- Forecasting demand for thousands of products across multiple warehouses.

Time series features

Objective

Develop a framework that automates the selection of the most appropriate forecasting method for a given time series by using an array of features computed from the time series.

Time series features

Objective

Develop a framework that automates the selection of the most appropriate forecasting method for a given time series by using an array of features computed from the time series.

- **Basic idea:**
  Transform a given time series $y = \{y_1, y_2, \cdots, y_n\}$ to a feature vector $F = (f_1(y), f_2(y), \cdots, f_p(y))'$.

Time series features

Objective

Develop a framework that automates the selection of the most appropriate forecasting method for a given time series by using an array of features computed from the time series.

- **Basic idea:**
  Transform a given time series $y = \{y_1, y_2, \cdots, y_n\}$ to a feature vector $F = (f_1(y), f_2(y), \cdots, f_p(y))'$.
- Examples for time series features

Time series features

Objective

Develop a framework that automates the selection of the most appropriate forecasting method for a given time series by using an array of features computed from the time series.

- **Basic idea:**
  Transform a given time series $y = \{y_1, y_2, \cdots, y_n\}$ to a feature vector $F = (f_1(y), f_2(y), \cdots, f_p(y))'$.
- Examples for time series features
  - strength of trend

Time series features

Objective

Develop a framework that automates the selection of the most appropriate forecasting method for a given time series by using an array of features computed from the time series.

- **Basic idea:**
  Transform a given time series $y = \{y_1, y_2, \cdots, y_n\}$ to a feature vector $F = (f_1(y), f_2(y), \cdots, f_p(y))'$.
- Examples for time series features
  - strength of trend
  - strength of seasonality

Time series features

Objective

Develop a framework that automates the selection of the most appropriate forecasting method for a given time series by using an array of features computed from the time series.

- **Basic idea:**
  Transform a given time series $y = \{y_1, y_2, \cdots, y_n\}$ to a feature vector $F = (f_1(y), f_2(y), \cdots, f_p(y))'$.
- Examples for time series features
  - strength of trend
  - strength of seasonality
  - lag-1 autocorrelation

Time series features

Objective

Develop a framework that automates the selection of the most appropriate forecasting method for a given time series by using an array of features computed from the time series.

- **Basic idea:**
  Transform a given time series $y = \{y_1, y_2, \cdots, y_n\}$ to a feature vector $F = (f_1(y), f_2(y), \cdots, f_p(y))'$.
- Examples for time series features
  - strength of trend
  - strength of seasonality
  - lag-1 autocorrelation
  - spectral entropy

Time series features

- length
- strength of seasonality
- strength of trend
- linearity
- curvature
- spikiness
- stability
- lumpiness
- first ACF value of remainder series
- parameter estimates of Holt's linear trend method
- spectral entropy
- Hurst exponent
- nonlinearity
- parameter estimates of Holt-Winters' additive method
- unit root test statistics
- first ACF value of residual series of linear trend model
- ACF and PACF based features - calculated on both the raw and differenced series

Methodology: FFORMS

FFORMS: Feature-based **FOR**ecast **M**odel **S**election

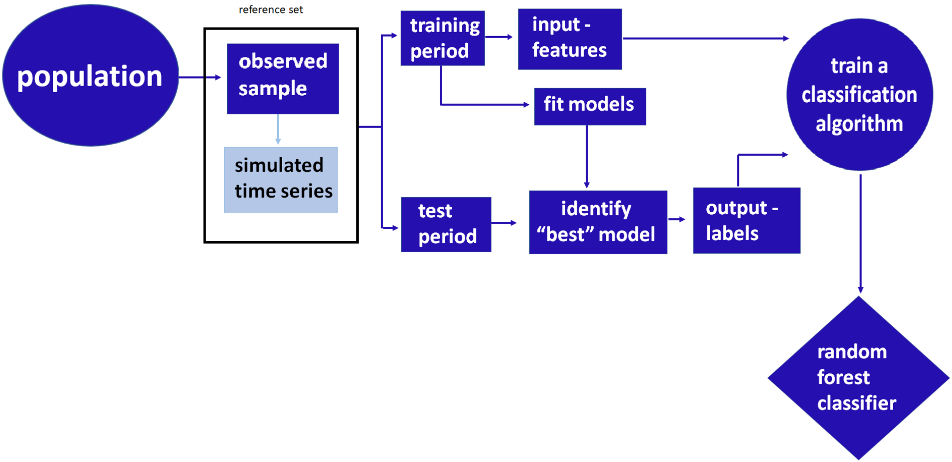
Offline

- A classification algorithm (the meta-learner) is trained.

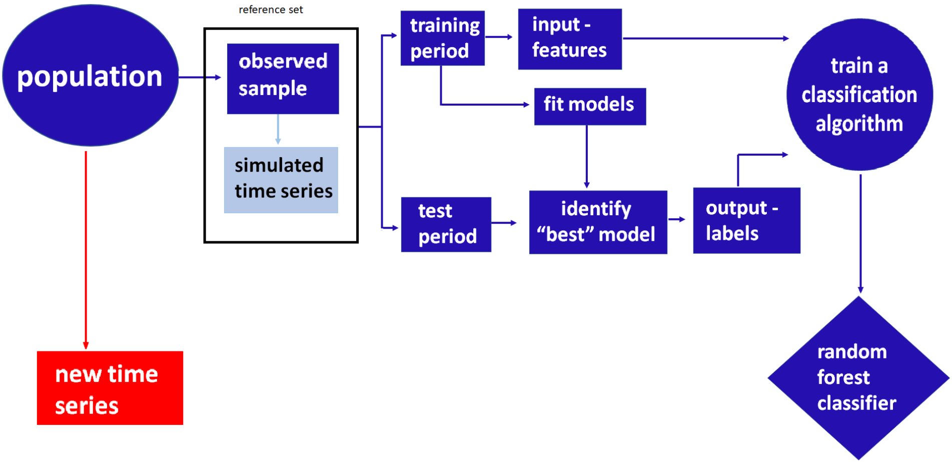
Online

- Calculate the features of a time series and use the pre-trained classifier to identify the best forecasting method.

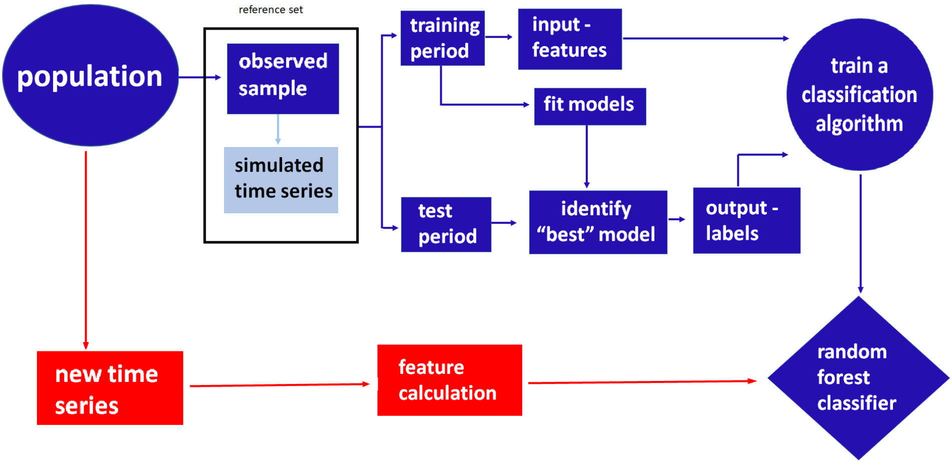
FFORMS: “offline” part of the algorithm



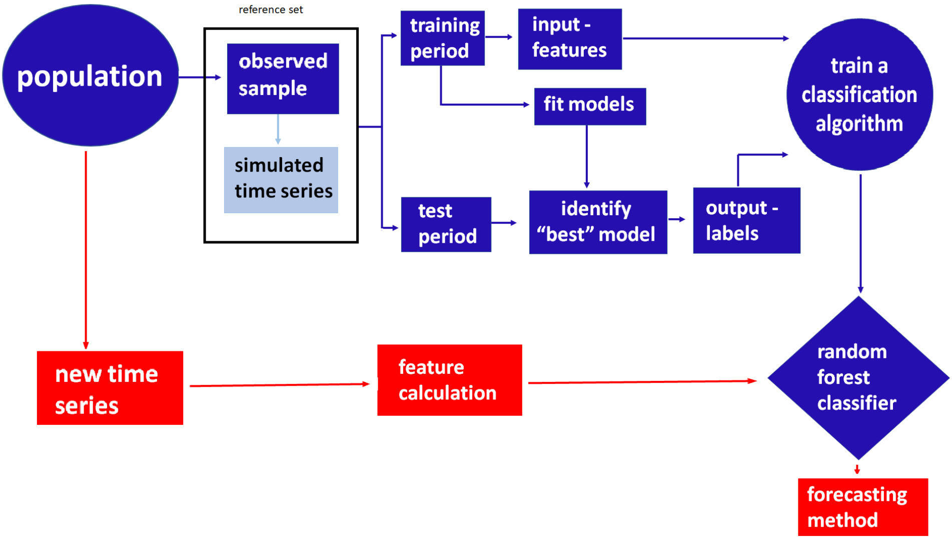
FFORMS: “online” part of the algorithm



FFORMS: "online" part of the algorithm



FFORMS: "online" part of the algorithm



R package: seer

Installation

```r
devtools::install_github("thiyangt/seer")
library(seer)
```

R package: seer

Installation

```
devtools::install_github("thiyangt/seer")
library(seer)
```

Example data set:

```
library(Mcomp)
yearly_m3 <- subset(M3, "yearly")
yearly_m3
```

```
M-Competition data: 645 YEARLY time series

        Type of data
Period   DEMOGRAPHIC FINANCE INDUSTRY MACRO MICRO OTHER
  YEARLY         245      58      102    83   146    11
```

Input: features

```
    entropy lumpiness stability     hurst     trend   spik
1 0.7729350         0         0 0.9710509 0.9950394    589
2 0.8374974         0         0 0.9473065 0.8687934 602056
3 0.8250352         0         0 0.9486339 0.8648297 668930
   curvature    e_acf1    y_acf1 diff1y_acf1  diff2y_acf1
1   531.9694 0.4124236 0.7623182   0.5974236 -0.004813322
2 -2823.8839 0.3240316 0.7507872   0.2399691 -0.398246929
3 -3078.0284 0.4571183 0.7687310   0.4461251 -0.211798893
  diff1y_pacf5 diff2y_pacf5 nonlinearity lmres_acf1      ur_
1    0.5483426    0.2301945     2.124405  0.4819001   1.329
2    0.1565805    0.3074159     1.998710  0.7227836  -3.735
3    0.3708305    0.1717048     1.449664  0.7645834  -3.978
     y_acf5 diff1y_acf5 diff2y_acf5     alpha      beta
1 1.0230152   0.4213774   0.3614128 0.9998869 0.9998869
2 0.9855601   0.1338517   0.5582498 0.9998999 0.2171013
3 1.0798980   0.3609988   0.7632291 0.9998999 0.5054122
```

Discussion and Conclusions

- FFORMS: framework for forecast model selection using meta-learning based on time series features.

Discussion and Conclusions

- FFORMS: framework for forecast model selection using meta-learning based on time series features.
- FFORMS algorithm uses the knowledge of the past performance of candidate forecast models on a collection of time series in order to identify the best forecasting method for a new series.

Discussion and Conclusions

- FFORMS: framework for forecast model selection using meta-learning based on time series features.
- FFORMS algorithm uses the knowledge of the past performance of candidate forecast models on a collection of time series in order to identify the best forecasting method for a new series.
- For real-time forecasting, our framework involves only the calculation of features, the selection of a forecast method based on the FFORMS random forest classifier, and the calculation of the forecasts from the chosen model.

Discussion and Conclusions

- FFORMS: framework for forecast model selection using meta-learning based on time series features.
- FFORMS algorithm uses the knowledge of the past performance of candidate forecast models on a collection of time series in order to identify the best forecasting method for a new series.
- For real-time forecasting, our framework involves only the calculation of features, the selection of a forecast method based on the FFORMS random forest classifier, and the calculation of the forecasts from the chosen model.
- We have also introduced a simple set of time series features that are useful in identifying the "best" forecast method for a given time series.

R package: seer

available at: https://github.com/thiyangt/seer

Installation

```r
devtools::install_github("thiyangt/seer")
library(seer)
```

R package: seer

seer

available at: https://github.com/thiyangt/seer

Installation

```r
devtools::install_github("thiyangt/seer")
library(seer)
```

paper: https://robjhyndman.com/publications/fforms/

Email: thiyanga.talagala@monash.edu