Ensemble Technique Based on Fuzzy Rules

Lenka Vavříčková, David Sikora, Martin Štěpnička

Institute for Research and Applications of Fuzzy Modeling
Centre of Excellence IT4Innovations
Division of the University of Ostrava

ISCAMI 2012, Malenovice

IRAFM

Outline

IRAFM

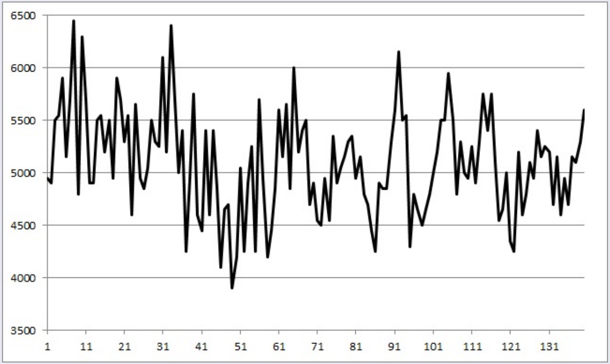
Outline

1. **Time series**
2. Ensemble techniques
3. Fuzzy Rule-Based Ensemble
4. Methodology

IRAFM

Time series

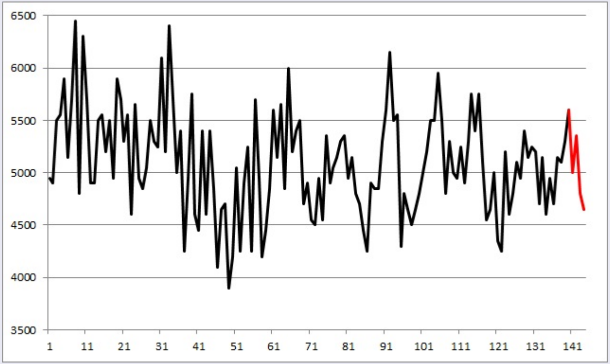
Time series: $y_1, y_2, \ldots, y_t$ - a finite sequence of real numbers



Time series

Task:

- to forecast the next values $y_{t+1}, y_{t+2}, \ldots, y_{t+h}$
- h - *forecasting horizon*



Forecasting methods

Standard approaches:

- Box-Jenkins methodology
- exponential smoothing methods
- decomposition
- etc.

Encounter difficulties

No single forecasting method that generally outperforms all other.

Consequence

- danger of choosing a method that is inappropriate for a given time series

Encounter difficulties

We stress that searching for methods that outperform any other for narrower specific subsets of time series have not been successful yet. . .

Collopy, Armstrong:

"Although forecasting expertise can be found in the literature, these sources often fail to adequately describe conditions under which a method is expected to be successful."

Outline

Ensemble techniques

Distinct ensemble techniques (“ensembles”) are proposed . . .

The main idea

- an appropriate combination of more forecasting methods
- to avoid the danger of choosing a single inappropriate one

Ensemble techniques

[Bates and Granger, 1969]

- significant gains in accuracy through combinations of more methods

[Newbold and Granger, 1974]

- combined various time series forecasts
- compared the combination against the performance of the individual methods
- showed that a linear combination of the forecasts could constitute a forecast with an error variance smaller than the individual forecasts

IRAFM

Ensemble techniques

...are constructed as a (linear) combination of the individual methods.

Let us assume

- a set of M individual methods
- a time series $y_1, y_2, \ldots, y_t$
- a forecasting horizon h
- the prediction provided by j-th individual method

$$\hat{y}_{t+1}^{(j)}, \hat{y}_{t+2}^{(j)}, \ldots, \hat{y}_{t+h}^{(j)}, \quad j = 1, \ldots, M$$

Ensemble techniques

Ensemble forecast

$$\hat{y}_{t+i} = \frac{1}{\sum_{j=1}^{M} w_j} \cdot \sum_{j=1}^{M} w_j \cdot \hat{y}_{t+i}^{(j)}, \quad i = 1, \ldots, h$$

$w_j \in \mathbb{R}$ - weight of the j-th individual method, $\sum_{j=1}^{M} w_j = 1$

How to determine weights

[Makridakis et al., 1982]

- taking a simple average outperforms taking a weighted average

Arithmetic mean ("Equal weights")

$$\hat{y}_{t+i} = \frac{1}{M} \cdot \sum_{j=1}^{M} \hat{y}_{t+i}^{(j)}, \quad i = 1, \ldots, h$$

- a hardly beatable benchmark

IRAFM

Outline

IRAFM

How to find non-equal weights

There are works that show the potential of more sophisticated approaches. . .

Motivation 1 [Lemke and Gabrys, 2010]

- use features of time series (length, skewness, kurtosis, . . .)
- time series are clustered on the basis of these features
- three best methods for each cluster are determined
- linear combination of these three methods is calculated

How to find non-equal weights

The approach of Lemke and Gabrys performed very well on sufficiently big set of time series.

Motivation 1

There exists a dependence between time series features and success of method.

How to find non-equal weights

Motivation 2: Rule-based forecasting [Collopy and Armstrong, 1992]

- a type of expert system composed of 99 crisp rules
- applies forecasting expertise and domain knowledge to combine forecasts from various extrapolation methods
- uses features of time series (e.g. length, trend, seasonality,...)
- features are identified in the conditional (if) part of the rules

IRAFM

How to find non-equal weights

- Only few rules in the RBF are directly used to determine weights
- Most of them rather set-up a specific method parameters
- *"IF there is an unstable recent trend, THEN add 30% to the weight on the random walk and subtract it from that on the other three methods."*

IRAFM

How to find non-equal weights

Our idea

- to follow the main ideas of rule-based forecasting by Collopy and Armstrong
- to follow the main ideas of using time series features by Lemke and Gabrys
- to obtain an interpretable and understandable model

Outline

IRAFM

Data sets and Evaluation

Time series data sets

- 198 time series (yearly, quarterly, monthly)
- 5 categories (micro, macro, industry, finance, demography)
- Training set (99 time series) - for an identification of our model
- Testing set (99 time series) - for a testing of the determined knowledge encoded in the fuzzy rules

Forecast error

$$SMAPE = \frac{1}{h} \sum_{T=t+1}^{t+h} \frac{|y_T - \hat{y}_T|}{(|y_T| + |\hat{y}_T|)/2} \times 100\%$$

Forecasting methods

Individual methods

1. Decomposition model (DT) - by NCSS®
2. Exponential smoothing (ES) - by ForecastPro®
3. ARIMA - by ForecastPro®
4. Moving average (MA) - by ForecastPro®
5. GARCH - by Gretl®
6. Random walk process (RW)
7. Random walk process with drift (RWd) - by Gretl®

Time series features

. . . motivated by Lemke and Gabrys

Features extracted from the given TS

- Trend
- Seasonality
- Length of the time series
- Skewness
- Kurtosis
- Coefficient of variation
- Stationarity
- Forecasting horizon
- Frequency

Regression analysis

Linear regression

- a technique for determining relationships among two or more variables
- represented by the system of linear equations

$$\mathbf{1}\text{-}\mathbf{s} = \mathbf{F} \cdot \beta + \varepsilon$$

s - a vector of the dependent variables (of normalized 99 SMAPE)
F - a matrix of the independent variables (of features of the 99 TS)
β - a vector of unknown regression parameters
ε - an error vector

...estimation of relationship between features of TS and SMAPE values for each individual forecasting method

Fuzzy rule base identification

1. The obtained models were sampled.
2. The nodes in the reduced features spaces with only significant features and SMAPE errors were obtained.
3. These nodes were used as learning data for the *Linguistic Learning Algorithm*.
4. Linguistic descriptions (LD) were generated.
5. LD with *perception-based logical deduction* derive conclusions based on imprecise observations.

Fuzzy rule base identification

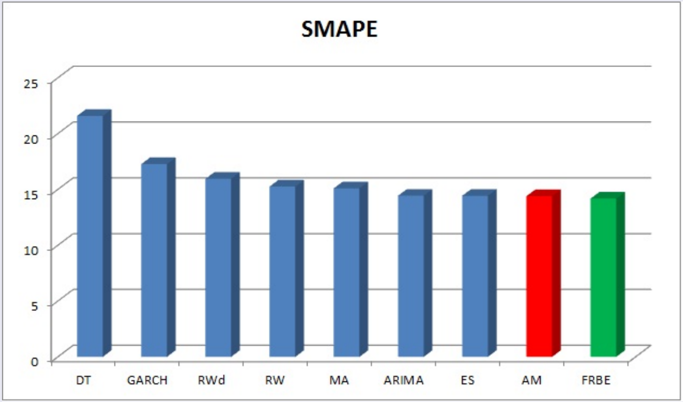
- Seven linguistic descriptions were obtained.
- Each of them determines weights of a single individual methods based on transparent and interpretable rules, such as

IF $Trend$ **is** Big **AND** *Variation* **is** $Small$ **THEN** w_{AR} **is** $ExBig$.

IRAFM

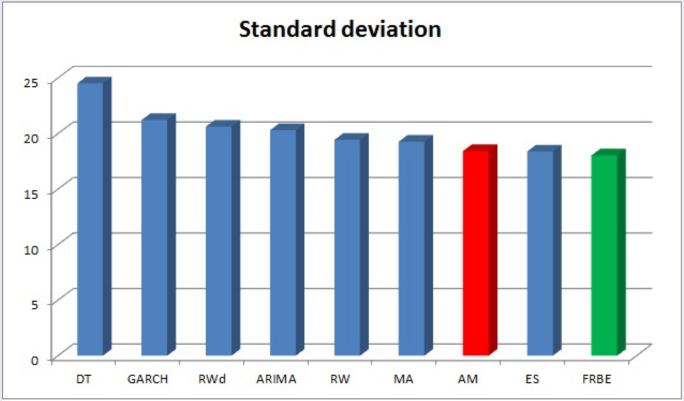
Results

Average of the SMAPE (in %)



Results

Standard deviation of the SMAPE (in %)



Results

Average of the SMAPE and standard deviation of the SMAPE

Methods	Average Error	Methods	Error Std. Deviation
DT	21.59	DT	24.52
GARCH	17.27	GARCH	21.22
RWd	15.95	RWd	20.62
MA	15.11	MA	19.27
ARIMA	14.44	ARIMA	20.31
ES	14.43	ES	18.39
AM	14.40	AM	18.42
FRBE	14.18	FRBE	18.03

Future work

Future work

- Using of other techniques for fuzzy rule base identification
  - fuzzy cluster analysis
  - linguistic associations mining
- Deep redundancy and/or consistency analysis of obtained fuzzy rule bases

IRAFM

Thank you for your attention.