Fuzzy Rule-Based Ensemble for Time Series Forecasting

Lenka Štěpničková[1], Martin Štěpnička[1], David Sikora[2]

[1]Centre of Excellence IT4Innovations - Division of the University of Ostrava
Institute for Research and Applications of Fuzzy Modeling
[2]Department of Informatics and Computers, University of Ostrava
Ostrava, Czech Republic

ISCAMI 2013, Malenovice

Outline

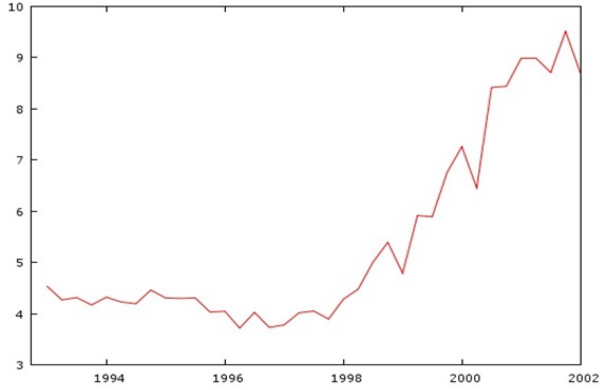
IRAFM

Outline

IRAFM

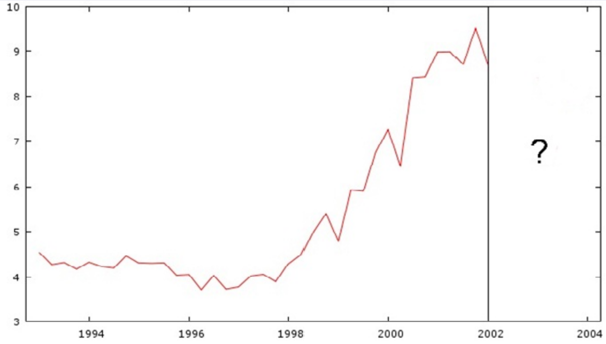
Time series

- Time series: $y_1, y_2, \ldots, y_T$ - a finite sequence of observations organized in time



Time series

- Task - to forecast h future values $y_{T+1}, y_{T+2}, \ldots, y_{T+h}$
- h – forecasting horizon



Forecasting methods

- Forecasting methods are used to forecast future values.
- Established many forecasting methods:
  - Exponential Smoothing methods,
  - Box-Jenkins Methodology,
  - Decomposition method,
  - etc.
- *"There are no algorithms that generally perform better or worse than random when looking at all possible data sets."*

IRAFM

Encounter difficulties

- Problem: Method $\times$ Time Series
  - Recommendations (to use Box-Jenkins methodology for long time series, do not use Single Exponential Smoothing for time series with distinct seasonality, etc.).
- However, these recommendations often fail.
- [Armstrong et al., 2001]

"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."

IRAFM

Outline

1. Time series
2. **Ensemble techniques**
3. Fuzzy Rule-Based Ensemble
4. Fuzzy GUHA
5. Implementation

Ensemble techniques

[Bates and Granger, 1969]

- significant gains in accuracy through combinations of more methods

"Ensembles"

- combine forecasts from M individual forecasting methods

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

$\hat{y}_{T+i}^{(j)}$ – forecast of y_{T+i} by the j-th individual method
$w_j \in \mathbb{R}$ – weight of the j-th individual method

Approaches for setting weights

Naïve yet powerful approach

- [Makridakis et al., 1982]

  "...taking a simple average outperforms taking a weighted average."

- Arithmetic mean ("equal weights") assigns equal weights to all individual forecasting methods ($w_j = 1/M$).
- It is demonstrated that arithmetic mean gives very good results and is hard to beat.

IRAFM

Approaches for setting weights

Motivating approaches

1. Meta-learning [Lemke and Gabrys, 2010]
   - uses features of time series (length, skewness, kurtosis, . . .)
   - time series are clustered using k-means algorithm
   - three best methods for each cluster are selected
   - for a given new time series, the closest cluster is determined and the given three best methods are combined

Approaches for setting weights

Motivating approaches

2. Rule-based forecasting [Collopy and Armstrong, 1992]
    - RBF is an expert system composed of 99 crisp rules that uses domain knowledge to combine forecasts from simple extrapolation methods.
    - RBF uses features of time series that are used in the conditional (if) part of the rules:

"IF there is a change in the basic trend THEN add 15% to the weight of random walk."

Motivation

We have followed the main ideas of approaches mentioned above:

1. To decrease the risk of choice of a wrong forecasting method, i.e. to achieve lower variance of forecast error.
2. To increase forecast accuracy on average.
3. To use only quantitative features of time series (no domain knowledge) which would allow to fully automatize the approach.

Outline

IRAFM

Fuzzy Rule-Based Ensemble (FRBE)

The nature of using fuzzy rules:

RBF: IF there is a change in the basic trend THEN add 15% to the weight of random walk.

FRBE: IF strength of trend is big THEN weight of ARIMA is big.

IRAFM

Fuzzy Rule-Based Ensemble (FRBE)

Important goal

We aim at transparent and interpretable model that uses fuzzy rules instead of crisp rules.

Linguistic description

IF $Kurt$ is $\mathrm{ML\ Sm}$ AND CV is Sm THEN W_{GARCH} is $\mathrm{Ro\ Bi}$
IF $Kurt$ is $\mathrm{Ro\ Me}$ AND CV is Sm THEN W_{GARCH} is $\mathrm{ML\ Bi}$
....
IF $Kurt$ is Sm AND CV is $\mathrm{Ro\ Me}$ THEN W_{GARCH} is $\mathrm{Ro\ Bi}$

Time series features

Motivated by Lemke and Gabrys

- Length of the time series
- Strength of trend
- Strength of seasonality
- Skewness
- Kurtosis
- Coefficient of variation
- Stationarity
- Frequency

Note: For each method, different features play the significant role.

Fuzzy Rule-Based Ensemble

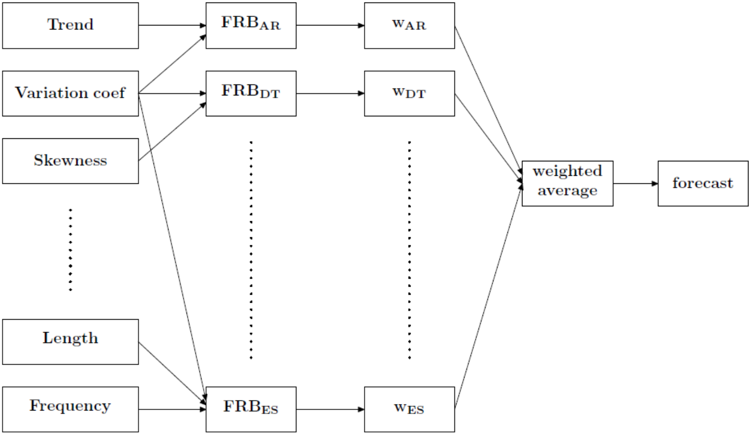
FRBE method

- It uses a single linguistic description (fuzzy rule base with evaluative linguistic expressions) for each forecasting method.
- Each of these linguistic descriptions determines the weights of a single individual method based on rules, such as:

  "**IF** *Strength of Seasonality is Small* **AND** *Coefficient of Variation is Roughly Small* **THEN** *Weight of the j-th method is Big*."

IRAFM

Structure of the FRBE method



IRAFM

Fuzzy Rule-Based Ensemble

FRBE method

- *Perception-based Logical Deduction* - to estimate (set up) a particular value of weight for each forecasting method
- *Weights*:

$$w_j = 1 - SMAPE_j$$

$SMAPE_j$ - normalized SMAPE forecasting error of the j-th method

IRAFM

Fuzzy Rule-Based Ensemble

Fuzzy rule base identification

- There are many distinct approaches.
- Because of the missing reliable expert knowledge, we focus on data-driven approaches.
- We employ the so called linguistic association mining firstly introduced as GUHA method.

Outline

IRAFM

(Fuzzy) GUHA method

It finds distinct statistically approved associations between attributes of given objects.

$$
\begin{array}{c|cccc}
 & X_1 & \ldots & X_m & Z \\
\hline
o_1 & a_{11} & \ldots & a_{1m} & z_1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
o_n & a_{n1} & \ldots & a_{nm} & z_n
\end{array}
$$

$o_1, \ldots, o_n$ – objects
$X_1, \ldots, X_m, Z$ – attributes
a_{ij}, z_i – values of j-th attribute measured on i-th object

GUHA method

Example of data set

	$\mathrm{BMI}_{<20}$	BMI_{20-25}	$\mathrm{BMI}_{>25}$	$\mathrm{Chol}_{5.2-6.2}$	$\mathrm{Chol}_{>6.2}$	$\mathrm{BP}_{>130/80}$
Smith	0	1	0	1	0	0
Clark	0	0	1	0	1	1
Bell	1	0	0	0	1	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Brown	0	1	0	1	0	0

Example of linguistic association

$$C(\mathrm{BMI}_{>25}, \mathrm{Chol}_{>6.2}) \simeq D(\mathrm{BP}_{>130/80})$$

IRAFM

Fuzzy GUHA method

Example of data set

	BMI_{low}	BMI_{middle}	BMI_{high}	Chol_{middle}	Chol_{high}	BP_{high}
Smith	0	0.9	0.1	0.7	0.3	0.1
Clark	0	0.2	0.8	0.1	0.9	0.9
Bell	0.6	0.4	0	0	1	0.6
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Brown	0	0.85	0.15	0.8	0.2	0.3

IRAFM

Application to time series

We search for implicative associations that may be directly interpreted as fuzzy rules.

Example of data set

	Tr_{ExSm}	Tr_{VeSm}	...	Seas_{ExSm}	...	$\mathrm{W}^{\mathrm{ARIMA}}_{ExSm}$	...	$\mathrm{W}^{\mathrm{ARIMA}}_{ExBi}$
TS_1	0.2	0.8	...	0.9	...	0.3	...	0
TS_2	0.1	0.6	...	0.5	...	0	...	0.8
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
TS_{99}	0.4	0.9	...	0.1	...	0.9	...	0

(Fuzzy) GUHA method

Goal

- to search for the linguistic associations of the form

$$C(F_1, \ldots, F_p) \simeq D(W)$$

C, D are (compound) evaluative predications,
$F_1, \ldots, F_p$ $(p \leq m)$ are all variables occurring in C

Four-fold table

	D	not D
C	a	b
not C	c	d

a, b, c, d - summations of membership degrees of data into fuzzy sets representing the antecedent C and consequent D or their complements

(Fuzzy) GUHA method

Quantifier $\simeq$

- a relationship between the antecedent and consequent
- a binary multitudinal quantifier $\simeq := \sqsubset_r^\gamma$ (implicational quantifier)

$$\frac{a}{a+b} > \gamma,$$
$$\frac{a}{m} > r$$

$\gamma \in [0,1]$ - a confidence degree
$r \in [0,1]$ - a support degree

(Fuzzy) GUHA method

Example

Let F_1 be skewness and F_2 be coefficient of variation.

Fuzzy GUHA method provided us with the following confirmed implicative hypothesis:

"*Skewness is roughly medium* AND *Coefficient of variation is extremely small* $\sqsubset^{0.7}_{0.04}$ *Weight of decomposition method is more or less big*"

Such an approved association may be viewed and thus, directly interpreted, as the following fuzzy rule:

"IF *Skewness is Roughly Medium* AND *Coefficient of Variation is Extremely Small* THEN *Weight of the decomposition method is More or less Big.*"

Outline

IRAFM

Data sets

	Training set			Testing set		
	Yearly	Quart.	Monthly	Yearly	Quart.	Monthly
Micro	5	7	7	5	7	7
Macro	7	8	8	7	8	8
Industry	7	7	6	7	7	6
Finance	7	7	5	7	7	5
Demo	7	4	7	7	4	7
Total	33	33	33	33	33	33

- Training set – for an identification of our model
- Testing set – for a testing of the determined knowledge encoded in the fuzzy rules

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®

Results

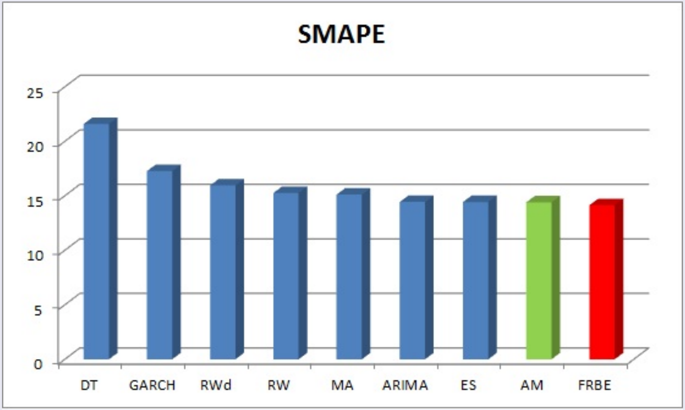
Big number of rules generated by fuzzy GUHA $\longrightarrow$ post-processing.

Process consist in search and deletion the redundant and duplicate rules.

Methods	No.of rules (GUHA)	Reduced no.of rules
ARIMA	7240	141
DT	9	3
ES	686	31
GARCH	17	7
MA	324	25
RW	234	23
RWd	152	20

Results

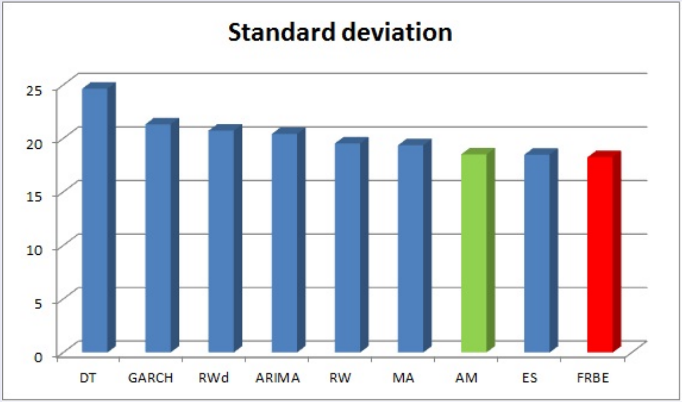
Average of the SMAPE (in %)



	DT	GARCH	RWd	RW	MA	ARIMA	ES	AM	FRBE
SMAPE	21.59	17.27	15.95	15.26	15.11	14.44	14.43	14.40	14.14

Results

Standard deviation of the SMAPE (in %)



	DT	GARCH	RWd	ARIMA	RW	MA	AM	ES	FRBE
SMAPE	24.64	21.32	20.73	20.41	19.53	19.36	18.51	18.48	18.24

Conclusions

Conclusions

- We have clearly stated the main motivations and ideas for fuzzy rule-based ensemble.
- We have demonstrated the promising potential of the fuzzy rule-based ensemble (the victory has been reached in the accuracy as well as in the robustness).
- Let us also recall the linguistic nature of the suggested approach.

Thank you for your attention.