Raport Badawczy
Research Report

RB/7/2016

Data-mining approach to priors in the analysis of short time series

K. Kaczmarek-Majer,
O. Hryniewicz

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 3810100

fax: (+48) (22) 3810105

Kierownik Zakładu zgłaszający pracę:
Prof. dr hab. inż. Olgierd Hryniewicz

Warszawa 2016

Abstract

Selection of an appropriate time series model and estimation of its parameters may become very challenging tasks for short series of observations. The well known information criteria often fail to adequately identify the predictive model for the small sample sizes, for which real-life applications are routinely made. Within this research, we focus on small sample sizes and first, analyze performance of selected information criteria. Secondly, we propose an automatic approach for the construction of the prior probability distributions. The proposed method incorporates selected data-mining techniques and similarity measures into the Bayesian averaging. The performance of the proposed method is illustrated with simulation study for stationary processes and experimental study for benchmark datasets.

1 Introduction

Although, there is a wealth of forecasting methods for time series and several well established information criteria, most of them require significant amount of historical data for the effective performance, and therefore, are inadequate for short time series. In [1], the authors review and investigate the performance of the common information criteria, namely Akaike information criteria (AIC, AICc) and Schwarz (Bayesian) Information criterion (BIC), concluding that the indiscriminate use of them in selecting the best predictive model may lead to inaccurate results, as their performance is model dependent. Especially, results for the small sample sizes, for which real-life applications are routinely made, are unsatisfactory, e.g., for the time series models ARMA(1, 1), ARMA(1, 2) or, ARMA(2,1) similar to the results obtained for the normal models, none of the considered information criteria selected the correct model (i.e., the model from which the data were generated), and in these cases, the true positives (TP) ratios are all below 45% [1]. Nonetheless, practitioners are often posed to the dilemma of model selection and forecasting despite the shortness of the available data.

At the same time, one of the main advantages of the Bayesian approach to time series analysis and the main challenge for practitioners, is the ability to describe the data imprecision in terms of prior probability distributions [2]. Many authors show the critical importance of the prior assumptions for the Bayesian inference. Hopefully, data mining techniques may support the selection of time series models. We use artificially generated template time series, and find these series, and in consequence these models, our data are similar to. Then, the degrees of similarity are used for the computation of prior model weights. The inspiration for the proposed approach comes from the idea of *imaginary training samples* and the *expected posterior priors* as introduced by [3].

The intelligent combination of data mining and Bayesian time series analysis and forecasting has been proven successful, see e.g., [4, 5]. or the recent application to process quality control [6].

Within this research, we focus on short time series of 50, 100 and 1000 observations, and first, analyze performance of selected well-known information criterion. Secondly, we propose an automatic approach for the construction of the prior probability distributions. The proposed method incorporates selected data-mining techniques and similarity measures for the training examples into the Bayesian averaging. The performance of the proposed method is illustrated with simulation study for benchmark datasets. The experimental results confirm that the proposed approach may outperform the well-known methods for short samples.

The structure of this report is as follows. Next Chapter explains the proposed approach. The numerical results of the similarity simulation analysis are gathered in Chapter 3. The experimental results for the benchmark data are presented in Chapter 4. This report concludes with general remarks and further research opportunities in Chapter 5.

2 Data-mining approach to prior weights

In this Section, we explain the proposed approach of calculating the similarity measures of training examples, and then the weights for model averaging. The approach assumes defining a set of considered predictive time series models $M = \{M_1, M_2, ..., M_J\}$). Next, for each of the J models (processes), its s realizations (training examples) are generated and considered for similarity calculations. For the clarity reasons, the length of generated series is the same as length of the considered time series.

According to [7] we define a set $M = \{M_1, M_2, ..., M_J\}$ of multiple competitive predictive models of a considered process.

Predictive model [2]
Predictive model A describes a vector of observable random variable y_t over a sequence of time $t = 1, 2, 3, \ldots$. Let Y_{t-1} denote a sequence $\{y_i\}_{i=t-1}^{i=1}$. A model A specifies a corresponding sequence of probability density functions $p(y_t|Y_{t-1}, \theta_A, A)$, where $\theta_A \in \Theta_A$ is a $k_A \times 1$ vector of unobservables.

Then, the posterior density of a vector of interest ω is defined as follows:

$$p(\omega|y, M) = \sum_{j=1}^{J} p(M_j|y, M)p(\omega|y, M_j) \tag{1}$$

where $p(M_j|y, M)$ are the prior model probability distributions constructed from weights.-

Within this research, various $AR(p)$ models are used as competitive predictive models.

$$y_i = \sum_{i=1}^{p} \phi_i y_{t-i} + a_t \tag{2}$$

where $a_t \sim N(0, \sigma^2)$ are normally distributed independent random variables with the expected value equal to zero, and the finite standard deviation $\sigma^2 \in (0, 1)$ and $\phi_i \in (-1, 1)$.
Thus, our assumed model describes a classical autoregressive stochastic process of the pth order $AR(p)$. For a comprehensive description of the $AR(p)$ process see e.g., the seminal book by Box and Jenkins [8] or a popular textbook by Brockwell and Davis [9].

Having defined the training models M, the weights are determined by the distances learned between the considered time series and the training examples where the training examples are generated from M. To sum up, the definition phase requires providing k - number of alternative models to be considered, p - max order of the AR process considered to training examples and α - min difference between autoregressive coefficients of AR models. The input for the approach is the considered (short) time series y and s - number of sample time series from each of the template AR processes. Its output is the value of prior model probability distributions (weights). The overview of the proposed procedure is presented in Fig. 1.

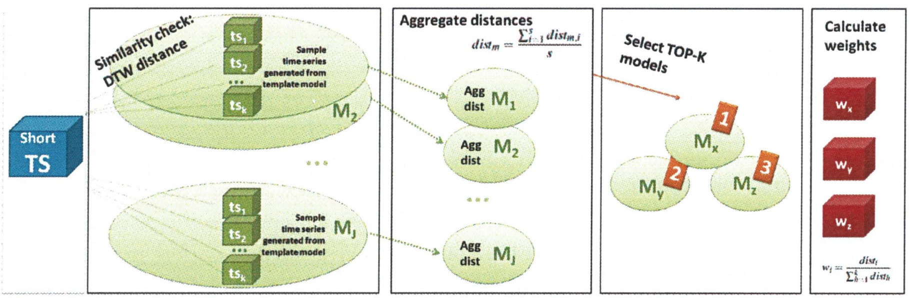


Figure 1: Overview of the proposed Bayesian forecasting with Data-Mining priors (B-DM) approach.

As presented, the data-mining methods are incorporated for the selection of alternative models in Bayesian averaging and their prior weights are calculated using DTW distance. DTW (*Dynamic Time Warping*) [10] is the elastic distance measure and enables to calculate the smallest distance between two series.

The approach may be divided into the following two steps:

1. checking the similarity between two time series;

2. aggregating similarities for training examples from the same model.

2.1 Similarity between time series

The similiarity of two time series is evaluated by calculating the distance between them. Within the proposed approach, the DTW (*Dynamic Time Warping*) measure as introduced by Berndt and Clifford [10] is adapted. DTW is the classical elastic measure and enables to calculate the smallest distance between two series of observations taking into account dilatation in time.

Let $X = \{x_1, x_2, ..., x_N\}$ and $Z = \{z_1, z_2, ..., z_M\}$ denote time series to be compared. The distance d between two points x_i and z_j, the so called local cost function, is defined as follows

$$d(i,j) = f(x_i, z_j) \geq 0 \tag{3}$$

The magnitude of the difference $d(i,j) = |x_i - z_j|$ (Manhattan) or square of the difference $d(i,j) = (x_i - z_j)^2$ (Euclidean) are some of the most common local cost functions considered in applications.

The DTW distance is based on the following recursive relation, which defines a cumulative distance $g(i,j)$ for $i \in \{1, ..., N\}$ and $j \in \{1, ..., M\}$

$$g(i,j) = d(i,j) + min[g(i-1,j), g(i-1,j-1), g(i,j-1)] \tag{4}$$

The cumulative distance is the sum of the distance between current elements and the minimum of the cumulative distances of the neighboring points. Two points (x_i, z_j) and (x_{i*}, z_{j*}) on the N-by-M grid are called neighboring if $(|i - i*| = 1 \; and \; |j - j*| = 0)$ or $(|i - i*| = 0 \; and \; |j - j*| = 1)$.

When two compared series are of the same length the value of $g(N, M)$ defines the distance between them. However, when $N \neq M$ the situation is more complicated, and the elements of both series have to be aligned in some way. The alignment of the elements from X and Z such that the distance between them is minimized is called a *warping path*. The DTW problem is defined as a minimization of cumulative distances over potential warping paths based on the cumulative distance for each path. This problem is solved using dynamic programming, and its complexity is $O(NM)$, and its solution is considered as the distance between two compared time series.

In Figure 2., the performance of the Euclidean and DTW distances is compared for exemplary series of observations from AR(-0,9), AR(-0,5) and AR(0,0) processes.

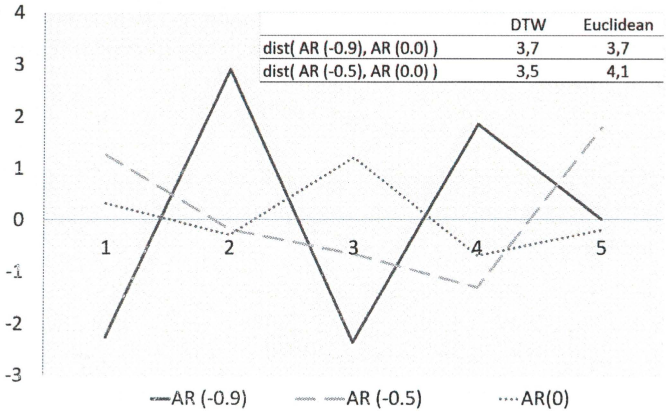


Figure 2: Euclidean and DTW distances for exemplary series of observations

As observed, the DTW distance between series from AR(0,0) and AR(-0,9) amounts to 3,7, whereas the distance between series from AR(0,0) and AR(-0,5) is smaller, and amounts to 3,5. On the other hand, the Euclidean distance between series from AR(0,0) and AR(-0,9) is also 3,7, but the distance between series from AR(0,0) and AR(-0,5) results 4,1, which is contradictory to intuition.

Our experiments confirm the good properties of the DTW measure, especially for time series with identified dilatation in time, because DTW seems to preserve trends. For further reading, we refer to, e.g., the recent survey and experimental comparison of representation methods and distance measures for time series data provided by Wang et al. [11]. Wang et al. conclude that especially on small data sets, elastic measures like DTW can be significantly more accurate than Euclidean distance and other lock-step measures.

Calculating distances between the monitored process y and the training time series from the template database using the DTW distance. For $m \in J$ and their realizations $i \in s$, the distance between the training time series and the considered monitored series of observations is calculated

$$dist_{m,i} = DTW(y_{m,i}, y) \tag{5}$$

2.2 Aggregation of the similarities for training examples from the same model

Aggregating similarities to establish weights corresponding to models $\{M_1, ..., M_k$ The mean aggregation operator is considered to construct weights for each model based on distances retrieved for each of the s sample time series. For model M_m where $m \in J$ having s realizations, the average distance between the training time series and the considered monitored series of observations is calculated as follows

$$dist_m = \frac{\sum_{i=1}^{s} dist_{m,i}}{s} \tag{6}$$

Having evaluated the average distance for each of the template models $\{M_1, ..., M$ the k models with smallest distance are selected. Then, the prior weights $\{w_1, ..., w_k\}$ are calculated

$$w_i = \frac{dist_i}{\sum_{h=1}^{k} dist_h} \tag{7}$$

One of the simplest scenarios assumes that prior model distributions are represented with a uniform distribution driven by w_i.

3 Numerical Results

In this Chapter, the simulation study checks the effectiveness of the well-known and commonly used information criterion - BIC for small samples from various autoregressive processes in line with [8]. The performance of the criterion is evaluated through Monte Carlo simulations for samples of size 100 and 1000. Table 1. and Table 2. summarize the obtained results for exemplary AR(1) and AR(2) processes of moderate and strong negative autocorrelation. For respective positive autocorrelation, the results are similar and lead to the same conclusions.

Table 1: Performance for the BIC for simulated autoregressive processes of order 1 and 2 with moderate negative autocorrelation (-0.4) for samples of size 100 and 1000.

Size	Act. par.	Stat.	AR(1) est. par	AR(1) BIC	AR(2) est. par.	AR(2) BIC	True Ord	AR(1) sel.
1000	[-0.4 0.9]	TRUE	[-0.99991534]	-	[-1.99999993 -0.99999993]	nan	2	TRUE
1000	[-0.4 0.8]	TRUE	[-0.99994173]	-	[-1.99999973 -0.99999973]	nan	2	TRUE
1000	[-0.4 0.7]	TRUE	[-1.]	-	-	-	2	TRUE
1000	[-0.4 0.6]	TRUE	[-0.99744928]	3 271	-	-	2	TRUE
1000	[-0.4 0.5]	TRUE	[-0.85011151]	3 104	[-0.45143124 0.46918402]	2 862	2	FALSE
1000	[-0.4 0.4]	TRUE	[-0.66324755]	3 017	[-0.4237208 0.36110914]	2 885	2	FALSE
1000	[-0.4 0.3]	TRUE	[-0.52760544]	2 906	[-0.37152362 0.29543919]	2 822	2	FALSE
1000	[-0.4 0.2]	TRUE	[-0.47480786]	2 919	[-0.37817226 0.20249097]	2 885	2	FALSE
1000	[-0.4 0.1]	TRUE	[-0.43145405]	2 840	[-0.40223421 0.06763289]	2 843	2	TRUE
1000	[-0.4]	TRUE	[-0.35939091]	2 888	[-0.36264945 -0.00906047]	2 894	1	TRUE
1000	[-0.4 -0.1]	TRUE	[-0.40491605]	2 898	[-0.45627706 -0.12656484]	2 889	2	FALSE
1000	[-0.4 -0.2]	TRUE	[-0.30765432]	2 925	[-0.37266501 -0.21062889]	2 887	2	FALSE
1000	[-0.4 -0.3]	TRUE	[-0.28993272]	2 941	[-0.37257875 -0.28450227]	2 864	2	FALSE
1000	[-0.4 -0.4]	TRUE	[-0.30391497]	3 027	[-0.42029523 -0.38294085]	2 875	2	FALSE
1000	[-0.4 -0.5]	TRUE	[-0.24862351]	3 199	[-0.37213851 -0.49696284]	2 922	2	FALSE
1000	[-0.4 -0.6]	TRUE	[-0.23204219]	3 303	[-0.37391891 -0.6183415]	2 830	2	FALSE
1000	[-0.4 -0.7]	TRUE	[-0.21919818]	3 495	[-0.36879389 -0.68095906]	2 878	2	FALSE
1000	[-0.4 -0.8]	TRUE	[-0.22079662]	3 774	[-0.39617456 -0.78151739]	2 850	2	FALSE
1000	[-0.4 -0.9]	TRUE	[-0.20823812]	4 780	[-0.401198 -0.91832112]	2 952	2	FALSE
100	[-0.4 0.9]	TRUE	[-0.99914832]	10 658	-	-	2	TRUE
100	[-0.4 0.8]	TRUE	[-0.99941467]	7 269	-	-	2	TRUE
100	[-0.4 0.7]	TRUE	[-0.99969738]	3 648	-	-	2	TRUE
100	[-0.4 0.6]	FALSE	[-0.99360436]	339	[-0.469274 0.52803043]	313	2	FALSE
100	[-0.4 0.5]	TRUE	[-0.77632671]	332	[-0.41117353 0.46866421]	312	2	FALSE
100	[-0.4 0.4]	TRUE	[-0.81702453]	299	[-0.45408944 0.4471683]	282	2	FALSE
100	[-0.4 0.3]	TRUE	[-0.64732309]	321	[-0.4614985 0.28315013]	318	2	FALSE
100	[-0.4 0.2]	TRUE	[-0.44157287]	294	[-0.42128782 0.04624576]	298	2	TRUE
100	[-0.4 0.1]	TRUE	[-0.5655]	287	[-0.60549867 -0.07029883]	291	2	TRUE
100	[-0.4]	TRUE	[-0.46947771]	266	[-0.48555754 -0.03821165]	270	1	TRUE
100	[-0.4 -0.1]	TRUE	[-0.45031099]	271	[-0.46648451 -0.03653231]	276	2	TRUE
100	[-0.4 -0.2]	TRUE	[-0.36502912]	294	[-0.4087339 -0.12137979]	297	2	TRUE
100	[-0.4 -0.3]	TRUE	[-0.31313784]	292	[-0.44678491 -0.4136535]	278	2	FALSE
100	[-0.4 -0.4]	TRUE	[-0.41397367]	290	[-0.58269877 -0.40021461]	276	2	FALSE
100	[-0.4 -0.5]	TRUE	[-0.24781814]	321	[-0.35923199 -0.44982645]	303	2	FALSE
100	[-0.4 -0.6]	TRUE	[-0.38289503]	320	[-0.56846743 -0.49257865]	298	2	FALSE
100	[-0.4 -0.7]	TRUE	[-0.13958258]	345	[-0.2435961 -0.68104896]	288	2	FALSE
100	[-0.4 -0.8]	TRUE	[-0.23766029]	421	[-0.44467018 -0.82658981]	315	2	FALSE
100	[-0.4 -0.9]	TRUE	[-0.22014092]	464	[-0.42187128 -0.9059012]	291	2	FALSE

Table 2: Performance for the BIC for simulated autoregressive processes of order 1 and 2 with strong negative autocorrelation (-0.9) for samples of size 100 and 1000.

Size	Act. par.	Stat.	AR(1) est. par	AR(1) BIC	AR(2) est. par.	AR(2) BIC	True Ord	AR(1) sel.
1000	[-0.9 0.9]	FALSE	-	-	-	-	2	FALSE
1000	[-0.9 0.8]	FALSE	-	-	-	-	2	FALSE
1000	[-0.9 0.7]	FALSE	-	-	-	-	2	FALSE
1000	[-0.9 0.6]	TRUE	-	-	-	-	2	FALSE
1000	[-0.9 0.5]	TRUE	[-0.99985545]	-	[-1.99999905 -0.99999905]	nan	2	TRUE
1000	[-0.9 0.4]	TRUE	[-1.]	-	-	-	2	TRUE
1000	[-0.9 0.3]	TRUE	[-0.99992074]	-	-	-	2	TRUE
1000	[-0.9 0.2]	TRUE	[-0.99995772]	-	[-1.99999994 -0.99999994]	nan	2	TRUE
1000	[-0.9 0.1]	TRUE	[-0.99882807]	2 890	[-0.87785919 0.12108994]	2 883	2	FALSE
1000	[-0.9]	TRUE	[-0.90467191]	2 967	[-0.92029911 -0.01725652]	2 974	1	TRUE
1000	[-0.9 -0.1]	TRUE	[-0.79518711]	2 809	[-0.89552481 -0.12628212]	2 800	2	FALSE
1000	[-0.9 -0.2]	TRUE	[-0.72293297]	2 884	[-0.86500172 -0.19637283]	2 852	2	FALSE
1000	[-0.9 -0.3]	TRUE	[-0.71352927]	2 931	[-0.93058083 -0.30453577]	2 841	2	FALSE
1000	[-0.9 -0.4]	TRUE	[-0.63585017]	3 070	[-0.90940107 -0.43097116]	2 871	2	FALSE
1000	[-0.9 -0.5]	TRUE	[-0.60884975]	3 151	[-0.90103684 -0.47976178]	2 897	2	FALSE
1000	[-0.9 -0.6]	TRUE	[-0.57421867]	3 410	[-0.93639121 -0.63227373]	2 911	2	FALSE
1000	[-0.9 -0.7]	TRUE	[-0.52757977]	3 507	[-0.89509946 -0.69765305]	2 849	2	FALSE
1000	[-0.9 -0.8]	TRUE	[-0.49835805]	3 884	[-0.89446149 -0.79650712]	2 886	2	FALSE
1000	[-0.9 -0.9]	TRUE	[-0.47955252]	4 487	[-0.91083249 -0.8997689]	2 842	2	FALSE
100	[-0.9 0.9]	TRUE	[-0.99749076]	27 726	[-1.99999998 -0.99999999]	nan	2	TRUE
100	[-0.9 0.8]	TRUE	[-0.99773483]	25 785	[-1.99999997 -0.99999998]	nan	2	TRUE
100	[-0.9 0.7]	TRUE	[-0.99799189]	22 902	-	-	2	TRUE
100	[-0.9 0.6]	TRUE	[-0.99826375]	20 209	-	-	2	TRUE
100	[-0.9 0.5]	TRUE	[-1.]	17 455	-	-	2	TRUE
100	[-0.9 0.4]	TRUE	[-0.99886386]	13 830	-	-	2	TRUE
100	[-0.9 0.3]	TRUE	[-0.99920196]	9 714	-	-	2	TRUE
100	[-0.9 0.2]	TRUE	[-0.99957583]	5 298	[-1.99999997 -0.99999998]	nan	2	TRUE
100	[-0.9 0.1]	TRUE	[-0.93616013]	300	[-0.86626565 0.07376972]	304	2	TRUE
100	[-0.9]	TRUE	[-0.95011361]	289	[-0.8464681 0.10994096]	293	1	TRUE
100	[-0.9 -0.1]	TRUE	[-0.76175558]	272	[-0.79385354 -0.04164955]	276	2	TRUE
100	[-0.9 -0.2]	TRUE	[-0.75246776]	292	[-0.83273737 -0.10592143]	296	2	TRUE
100	[-0.9 -0.3]	TRUE	[-0.64571773]	313	[-0.89541888 -0.39018598]	301	2	FALSE
100	[-0.9 -0.4]	TRUE	[-0.67723427]	348	[-1.00299091 -0.46411015]	330	2	FALSE
100	[-0.9 -0.5]	TRUE	[-0.61557138]	312	[-0.93634553 -0.51713239]	286	2	FALSE
100	[-0.9 -0.6]	TRUE	[-0.569693]	378	[-0.96289554 -0.69074195]	317	2	FALSE
100	[-0.9 -0.7]	TRUE	[-0.55650753]	387	[-0.99089992 -0.76781253]	301	2	FALSE
100	[-0.9 -0.8]	TRUE	[-0.48622115]	395	[-0.89560854 -0.82267709]	286	2	FALSE
100	[-0.9 -0.9]	TRUE	[-0.4724781]	375	[-0.82484756 -0.79240311]	288	2	FALSE

Finally, Table 3 shows the summary of the success rate for the BIC criterion when identifying of AR(1) and AR(2) stationary processes.

As demonstrated in Table 3, for samples of size 20 from the AR(2) processes, the BIC mostly identifies the AR(1) process, and the success rate amounts to only 45%. Also for sample sizes equal to 100, the proper discrimination between AR(1) and AR(2) success rate seems unsatisfactory and equals 38%.

Table 3: Success rate of model order selection discriminating between AR(1) and AR(2).

True model	Criterion	N	*AR(1)*	*AR(2)*
AR(2)	**BIC**	**20**	0.55	0.45
AR(2)	**BIC**	**100**	0.38	0.63
AR(2)	**BIC**	**1000**	0.08	0.92

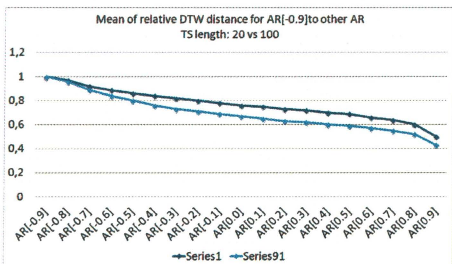
4 Experimental results for benchmark data

To show the performance of the proposed forecasting approach with data-mining priors, we run experiments on small samples from the popular benchmark repository. The experiments are performed for the subset of time series with 47 observations from the M3-Competition Repository by [12]. The 6-month-long forecast of B-DM are compared to best benchmark methods (ForecastPRO, Theta, Robust-Trend, ANN, ForecastX, Naive2) as referenced in [12].

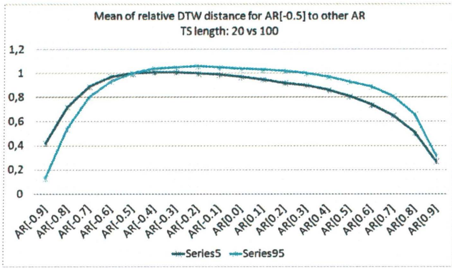
First, let us analyze the training examples. *Train dataset:* sample time series generated from template $AR(2)$ processes:

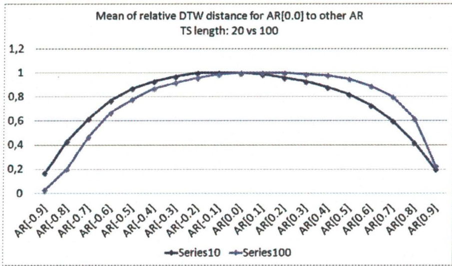
$$\tilde{y_t} = \sum_{i=1}^{p} \phi_i \tilde{y_{t-i}} + a_t \tag{8}$$

where $a_t \sim N(0, \sigma^2)$, $\tilde{y_t} = y_t - \mu$, $\phi_i \in (-1, 1)$, $\sigma^2 = 0.1$. We run the simulations on the database of the training examples to realize what is the impact of the sample size on the similarity measure performance. The comparative analysis for samples with 20 and 100 observations and the performance of the similarity measures for the training database is illustrated on the following charts.



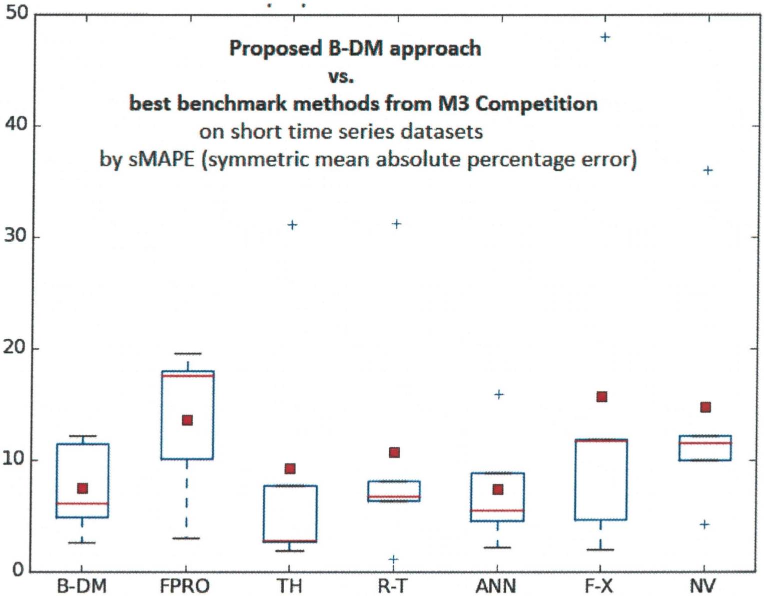
As illustrated, the DTW measure performs better for larger samples (n=100). However, in all considered examples for small samples (n=20),





the actual model was identified properly according to the adapted similarity measure.

Finally, the forecasting accuracy of the proposed Bayesian Forecasting with Data-Mining (B-DM) approach is analyzed. To evaluate the performance of the proposed method, the symmetric mean absolute percentage error (sMAPE) forecasting accuracy measure is used. The following boxplot shows results scored by the proposed Bayesian with data-mining prioris (B-DM) approach and by some of the leading benchmark methods [12].



As observed, the forecasting performance of the proposed B-DM approach is highly competitive with the best state-of-the-art forecasting methods.

5 Conclusion

As repeated in the literature and shown in this report with simulations, the state-of-the-art information criteria may inadequately identify the actual probabilistic model when only small sample of a time series is available. Hopefully, the intelligent combination of data-mining methods and the Bayesian time series analysis seems very promising for the forecasting of small samples. In this report, we have proposed the B-DM method that incorporates selected data-mining techniques and similarity measures into the Bayesian averaging.

The experimental results show that the proposed approach delivers very accurate forecasts, especially for the time series with small number of observations (n=20, 50, 100). Furthermore, it is observed that the Dynamic Time Warping measure helps to identify predictive models better than the Euclidean distance. Future research assumes next experimental evaluations, the inclusion of other similarity measures to optimize their performance and extending the approach for the other classes of predictive models.

References

[1] P. C. Emiliano, M. J. Vivanco, and F. S. de Menezes, "Information criteria: How do they behave in different models," *Computational Statistics and Data Analysis*, vol. 69, pp. 141–153, 2014.

[2] J. Geweke and C. Whiteman, *Handbook of Economic Forecasting*, vol. 1. Elsevier, 2006.

[3] J. Perez and J. Berger, "Expected-posterior prior distributions for model selection," *Biometrica*, vol. 89, pp. 491–511, 2002.

[4] O. Hryniewicz and K. Kaczmarek, "Bayesian analysis of time series using granular computing approach," *Applied Soft Computing*, vol. 47, p. 644–652, 2016.

[5] K. Kaczmarek and O. Hryniewicz, "Forecasting of short time series with intelligent computing," *Challenging Problems and Solutions in Intelligent Systems, Studies in Computational Intelligence*, vol. 59-74, pp. 59–74, 2016.

[6] O. Hryniewicz and K. Kaczmarek, "Monitoring of short series of dependent observations using a control chart approach and data mining techniques," *(in review) Proc. of Intelligent Statistical Quality Control, Computational Statistics*, 2016.

[7] J. Geweke, "Contemporary bayesian econometrics and statistics," *Wiley series in probability and statistics*, 2005.

[8] G. Box, G. Jenkins, and G. Reinsel, *Time Series Analysis, 4th Edition*. Wiley, 2008.

[9] P. Brockwell and R. Davis, *Introduction to Time Series and Forecasting*. Springer, 2002.

[10] D. Berndt and J. Clifford, "Aaai-94 workshop on knowledge discovery in databases," *Using dynamic time warping to find patterns in time series*, pp. 359–370, 1994.

[11] X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann, and E. Keogh, "Experimental comparison of representation methods and distance measures for time series data," *Data Mining Knowledge Discovery*, vol. 26(2), pp. 275–309, 2013.

[12] S. Makridakis and M. Hibon, "The m3-competition: results, conclusions and implications," *International Journal of Forecasting*, pp. 451–476, 2000.