ARIMA Feature-Based Approach to Time Series Classification*

Agnieszka Jastrzebska[1][0000−0001−5361−5787], Wladyslaw Homenda[1,2][0000−0001−7787−4927], and Witold Pedrycz[3,4],[0000−0002−9335−9930]

[1] the Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland {A.Jastrzebska,homenda}@mini.pw.edu.pl
[2] the University of Information Technology and Management in Rzeszow, Poland
[3] the University of Alberta, Edmonton, Canada wpedrycz@ualberta.ca
[4] the Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Abstract. Time series classification is a supervised learning problem that aims at labelling time series according to their class belongingness. Time series can be of variable length. Many algorithms have been proposed, among which feature-based approaches play a key role, but not all of them are able to deal with time series of unequal lengths. In this paper, a new feature-based approach to time series classification is proposed. It is based on ARIMA models constructed for each time series to be classified. In particular, it uses ARIMA coefficients to form a classification model together with sampled time series data points. The proposed method was tested on a suite of benchmark data sets and obtained results are compared with those provided by the state-of-the-art approaches.

Keywords: time series · ARIMA, classification · SVMs · Random Forest

1 Introduction

Time series classification emerged as a vital area of study in machine learning. A popular group of algorithms are feature-based methods that convert time series into a collection of attributes, which are then subjected to classification. Feature-based approaches reduce problem's dimensionality, as the number of features extracted for each time series is usually much smaller than its length, i.e., the number of data points of the time series. Feature-based methods typically do not need time series to be of equal length.

In this paper, a new approach to time series classification is proposed. The idea is to process time series models instead of the time series themselves. It has several advantages. First of all, time series models are defined as a set of parameters, which quantity is much smaller than the length of the original time series. Secondly, the approach is suitable to process sets of time series of differing

* The project was funded by POB Research Centre Cybersecurity and Data Science of Warsaw University of Technology within the Excellence Initiative Program - Research University (ID-UB).

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_26

lengths. Finally, operating on model parameters instead of temporal sequences allows employing general-purpose classifiers. The above properties make our approach a feature-based method, where model parameters can be seen as time series features.

The quality of the new method is evaluated in a series of experiments and compared with the quality of several state-of-the-art approaches. Results show that this technique provides satisfying results.

The particular novel contribution of this paper, not present in previous studies on feature-based approaches to time series classification, is the usage of ARIMA models describing time series and viewing their parameters as patterns to be classified. In this light, we may note that in this paper, we introduce a fusion of several approaches: baseline (plain), feature-based, and model-based.

2 Literature Review

The baseline approach to time series classification would treat each time series data point as a single attribute. Thus, we may apply any standard classifier, for example, random forest, to a data frame in which one row corresponds to one time series and one column corresponds to one moment in time. We would have to make sure that each time series is of the same length and starts at the same moment in time. Surveys show that this baseline approach achieves surprisingly satisfying results [2].

There are numerous algorithms dedicated specifically to the time series classification problem. They fall into two main categories: distance-based and feature-based. The distance-based approaches aim at computing distances between time series. The vast research volume in this area was devoted to studies on various distance measures [1]. In feature-based classification methods, time series are transformed into feature vectors and classified using a conventional classifier such as a neural network or a decision tree. Many approaches fall into this group. For instance, spectral methods, such as discrete Fourier transform [11] or discrete wavelet transform [9] are used to provide features of the frequency domain, which are the basis for classifier construction. We shall also distinguish feature-based methods that are called Bag of Patterns. These are dictionary approaches, in which one extracts specific attributes describing time series and uses those attributes to classify them. The literature offers a wide range of such approaches. For example, we can name methods that transform time series into strings and, at the same time, reduce the length of an input sequence. A particular example of such an approach is Symbolic Aggregate Approximation (SAX) [10]. This method discretizes time series that were previously normalized. Another example is the SAX and Vector Space Model (SAXVSM), which joins the idea of time series to string conversion via discretization, but it adds token (character) weighting [12]. SAX was also fused with DTW in the method named DTW Features. It uses distances computed with the DTW with SAX histograms [8]. It is worth mentioning a method called Bag-of-Features [3]. It produces random subsequences of a given time series from which features are extracted. Fulcher

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_26

and Jones presented a technique that is the most relevant in the context of the approach introduced in this paper [5]. It extracts time series summaries in terms of correlations, distributions, entropy, stationarity, and scaling. A classification model is constructed using such features.

3 The Method

Feature-based approaches rely on extracting attributes from a raw time series. In this paper, we propose to use the parameters of a time series theoretical model as attributes. We use the AutoRegressive Integrated Moving Average model (ARIMA). The ARIMA model consists of three elements: autoregressive model (AR), moving average model (MA), and model integration (I).

An autoregressive model of order p, denoted as AR(p), uses p previous values to describe (predict) the current value. Its general form is given as:

$$z_t^{'} = c + b_1 z_{t-1} + b_2 z_{t-2} + \ldots + b_p z_{t-p} \tag{1}$$

where $z_t^{'}$ is a description (prediction) of the current value z_t, c is an intercept, describing the drift. The autoregressive model is analogous to the multiple regression model, but with lagged values of time series data points, instead of standard attributes used as predictors. Parameters $b_1, b_2, \ldots, b_p$ describe how strong is a relationship between history and a current value. Autoregressive models are typically applied to stationary data. Thus, time series with a trend or seasonal regularities need to be preprocessed. The description of the current value is an approximation of the real value z_t with the error ε_t:

$$z_t = c + b_1 z_{t-1} + b_2 z_{t-2} + \ldots + b_p z_{t-p} + \varepsilon_t \tag{2}$$

A moving average model, another component of ARIMA, uses past forecast errors in a regression-like model:

$$a_{t-1}\varepsilon_{t-1} + a_{t-2}\varepsilon_{t-2} + \ldots + a_{t-q}\varepsilon_{t-q} \tag{3}$$

q denotes the order of the moving average model; we denote it as MA(q). $a_1, a_2, \ldots, a_q$ are discovered coefficients. While the autoregressive model uses past values, the moving average uses past distortions to model a time series.

The third component of the ARIMA model is integration (I). Integration, in this context, is the action opposite to differentiating. If we join the three components together, we obtain ARIMA(p, d, q) model, where d is the degree of first differentiating applied. The model can be written as:

$$\begin{aligned} \text{ARIMA}(p,d,q) = c + b_1 z'_{t-1} + b_2 z'_{t-2} + \ldots + b_p z'_{t-p} + \\ a_{t-1}\varepsilon_{t-1} + a_{t-2}\varepsilon_{t-2} + \ldots + a_{t-q}\varepsilon_{t-q} + \varepsilon_t \end{aligned} \tag{4}$$

To automatically detect the structure of the model, that is p, d and q, one may use the Hyndman-Khandakar algorithm [6]. It combines unit root tests, minimization of the Akaike information criterion, and Maximum Likelihood Estimation to obtain an ARIMA model.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_26

Parameters of the ARIMA model describe the properties of the time series. In particular, they describe the intensity of the influence of historical values and distortions of the time series on the current state of the process. We postulate to use ARIMA model parameters to distinguish between time series belonging to different classes. We fit an ARIMA model for each time series separately and use computed parameters as attributes: d (the differentiating degree), c (the intercept), $b_1, \ldots, b_p$, (coefficients associated with consecutive values in the autoregressive model), and $a_1, \ldots, a_q$ (coefficients from the moving average model). The following coefficients are generated for a particular training set:

- $1 + 1$ (for d and c)
- p_j (the highest discovered order of autoregressive model in time series in a given training set, j is the number of this time series $j = 1, \ldots, K$))
- q_r (the highest discovered order of moving average model in time series in a given training set, r is the number of this time series $r = 1, \ldots, K$)

If for a given time series, we have obtained an autoregressive model's order lower than p_j or a moving average model's order lower than q_r, then irrelevant coefficients are set to 0. ARIMA coefficients make the elementary data frame that can be subjected to classification using a standard classifier.

The usage of ARIMA parameters as indirect features in pattern mining in time series is already present in the literature. Kalpakis et al. [7] and Wang et al. [13] used it to cluster time series.

To reinforce the efficacy of classification, apart from considering ARIMA parameters, we may include randomly chosen samples from time series. We may append an arbitrary number of such attributes. The simplest way would be to select indexes of time series observations randomly to be appended. The randomization of indexes must be from the range $[1, M_f]$, where M_f is the length of the shortest time series in the training set.

Extracted attributes are subject to a standard classification, which may be preceded by removing correlated columns or thinning a training set.

4 Results

We have conducted experiments using a collection of publicly available time series from http://timeseriesclassification.com/ and compared our results with other methods. The data sets on the web page were already standardized and split into train and test sets. We implemented the method introduced in this paper in R language with the use of e1071, randomForest, and forecast packages.

First experiments were conducted for the scenario when only ARIMA parameters were used as attributes. Next, apart from ARIMA parameters, we added 1%, 5%, 10%, 20%, and 30% of randomly selected time series data points. We used two classifiers: Support Vector Machine (SVM) and random forest (RF). Those two classifiers are very popular. In all experiments, we removed columns with variance lower than 0.01. All SVMs were using Gaussian kernels. We individually tuned parameters with a simple search procedure. We used 500 trees

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_26

Table 1. Comparison of accuracy (in %) achieved with our method and two groups of other methods: plain and other feature-based approaches. We apply colors to improve visibility: the greener the color the better our model was in the comparison.

data set name	best accuracy feature	plain	ARIMA	differences: ARIMA − feature	− plain	ARIMA model
Beef	80.00%	93.33%	80.00%	0.00%	-13.33%	RF, 20% (94)
BeetleFly	90.00%	90.00%	90.00%	0.00%	0.00%	RF, 5% (26)
BirdChicken	100.00%	85.00%	85.00%	-15.00%	0.00%	RF, 30% (154)
ChlorineConcentration	71.98%	92.42%	78.05%	6.07%	-14.38%	SVM, 30% (50)
Coffee	100.00%	100.00%	100.00%	0.00%	0.00%	RF,SVM, 5% (14)
CricketY	75.38%	61.03%	61.03%	-14.36%	0.00%	RF, 20% (60)
DiatomSizeReduction	93.14%	96.41%	93.14%	0.00%	-3.27%	RF, 20% (69)
DistalPhalanxOutlnAgeG.	78.26%	82.25%	80.67%	2.41%	-1.58%	RF, 30% (24)
DistalPhalanxOutlineCor.	84.17%	80.58%	84.50%	0.33%	3.92%	RF, 10% (8)
DistalPhalanxTW	69.78%	70.50%	78.75%	8.97%	8.25%	RF, 30% (24)
Earthquakes	74.82%	74.82%	81.99%	7.17%	7.17%	SVM, 0
FordA	92.95%	84.47%	90.92%	-2.04%	6.45%	RF, 0
FordB	75.06%	77.16%	86.52%	11.46%	9.36%	SVM, 10% (50)
GunPoint	100.00%	94.00%	95.33%	-4.67%	1.33%	RF, 20% (30)
Herring	64.06%	65.63%	68.75%	4.69%	3.13%	RF, 1% (5)
InlineSkate	51.64%	37.09%	48.18%	-3.45%	11.09%	RF, 0
InsectWingbeatSound	63.28%	65.61%	64.85%	1.57%	-0.76%	RF, 30% (77)
ItalyPowerDemand	96.02%	97.28%	96.89%	0.87%	-0.39%	RF, 10% (2)
Lightning2	85.25%	75.41%	77.05%	-8.20%	1.64%	RF, 10% (64)
Lightning7	75.34%	72.60%	78.08%	2.74%	5.48%	RF, 20% (64)
MiddlePhalanxOutlnCor.	57.79%	61.69%	77.50%	19.71%	15.81%	SVM, 30% (24)
MiddlePhalanxTW	59.74%	62.99%	65.16%	5.42%	2.18%	RF, 20% (16)
MoteStrain	90.34%	88.90%	89.46%	-0.88%	0.56%	RF, 20% (17)
Plane	100.00%	99.05%	99.05%	-0.95%	0.00%	RF, 10% (14)
ProximalPhalanxOutlnCor.	84.88%	86.83%	86.34%	1.46%	-0.49%	RF, 30% (24)
ProximalPhalanxTW	81.46%	82.44%	80.50%	-0.96%	-1.94%	RF, 30% (24)
ScreenType	51.20%	44.80%	46.67%	-4.53%	1.87%	SVM, 0
Strawberry	97.57%	97.30%	94.45%	-3.11%	-2.84%	RF, 10% (24)
SwedishLeaf	92.16%	88.16%	87.04%	-5.12%	-1.12%	RF, 30% (38)
Trace	100.00%	93.00%	95.00%	-5.00%	2.00%	SVM, 10% (28)

in RF, which is the recommended default value in the used library. All models were trained using train sets. We only used test sets for quality evaluation.

In the designed experiment, we decided to compare the results achieved with our method with two kinds of algorithms: (i) plain classifiers, run directly on time series and (ii) other feature-based algorithms.

The first group of methods is assumed to provide a bottom-line efficiency in time series classification. We considered the following algorithms: Naive Bayes, C4.5 decision tree, SVM with linear (SVML) and quadratic kernel (SVMQ), Bayesian Network, RF with 500 trees, rotation forest with 50 trees, and multilayer perceptron.

The second group is made of feature-based algorithms. Those are the competitors belonging to the same group as the method addressed in the paper. These were: Bag of Patterns [10], Symbolic Aggregate Approximation – Vector Space Model (SAXVSM) [12], Bag of SFA Symbols (BOSS) [4], Time Series Forest (TSF) [11], Time Series Bag of Features (TSBF) [3].

Table 1 provides aggregated results. They concern test sets. We give the best accuracy achieved by a classifier from each group: plain, feature-based, and the proposed ARIMA feature-based. In the last column in Table 1, we outline which configuration produced our model: a percentage of time series data points added

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_26

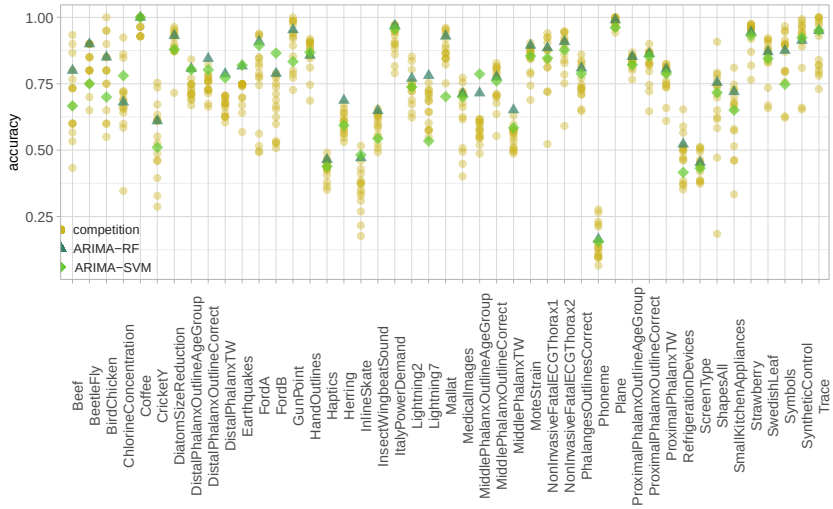

Fig. 1. The accuracy achieved using our method with SVM (light green diamonds) and RF (dark green triangles) as classifiers contrasted with particular results achieved by our 13 competitors (semi-transparent yellow circles). Results concern all 30 data sets.

as attributes, the name of a classifier that gave this particular result, and, in brackets, a specific number of data points that was added.

In 15 out of 30 data sets, the proposed method outperformed plain classifiers. In the further 5 cases, the best plain classifier produced the same accuracy as our method. In 14 out of 30 data sets, the new method outperformed other feature-based approaches to time series classification. In many cases, the advantage of the proposed method is very high. In 4 cases, the best feature-based competitor gave the same accuracy as the method introduced in this paper. When we compare the ARIMA feature-based method with plain classifiers, in two cases, our technique achieved accuracy at least 10% greater than this group of approaches. If we compare our method with other feature-based approaches, the results are slightly worse. Still, in six cases, the ARIMA feature-based method outperformed its best-performing competitor by at least 5%. RF turned out to be frequently outperforming SVM-based models.

In Figure 1, we illustrate accuracy achieved using our method with SVM and RF as classifiers contrasted with particular results achieved by our 13 competitors. The plot shows that in several cases, there are noticeable differences between accuracy obtained using SVM and RF. We also see that the quality of classification highly depends on a data set. There are cases, such as Coffee, HandOutlines, Haptics, and others, where all algorithms reached a very similar accuracy. In contrast, for sets such as Beef, BirdChicken, and ChlorineConcentration the differences were substantial. For example, in the Beef data set, SAXVSM achieved accuracy equal to 0.43 while SVMQ achieved accuracy equal to 0.93. The figure demonstrates that the proposed method performs well.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_26

5 Conclusion

In the paper, we have studied a new method for time series classification. It combines three distinct methodologies:

- A plain approach working on raw time series data. We randomly pick a portion of data points from the time series.
- A feature-based approach, in which the order of elements does not matter and a standard classifier performs the final classification task. Two general-purpose classifiers (RF and SVM) were utilized in the study.
- Time series models providing their parameters as features used in the classification process. The ARIMA model was employed for this purpose.

The empirical analysis performed on a wide range of benchmark time series demonstrates a satisfying quality of the presented technique comparing to two standard approaches: plain and feature-based.

References

1. Abanda, A., Mori, U., Lozano, J.A.: A review on distance based time series classification. Data Mining and Knowledge Discovery **33**(2), 378–412 (2019)
2. Bagnall, A., Lines, J., Bostrom, A., Large, J., Keogh, E.: The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. Data Mining and Knowledge Discovery **31**(3), 606–660 (2017)
3. Baydogan, M.G., Runger, G., Tuv, E.: A bag-of-features framework to classify time series. IEEE Trans. on Pattern Analysis and Machine Intel. **35**(11), 2796–2802 (2013)
4. Deng, H., Runger, G., Tuv, E., Vladimir, M.: A time series forest for classification and feature extraction. Information Sciences **239**, 142–153 (2013)
5. Fulcher, B.D., Jones, N.S.: Highly comparative feature-based time-series classification. IEEE Trans. on Knowledge and Data Engin. **26**(12), 3026–3037 (2014)
6. Hyndman, R., Khandakar, Y.: Automatic time series forecasting: The forecast package for r. Journal of Statistical Software, Articles **27**(3), 1–22 (2008)
7. Kalpakis, K., Gada, D., Puttagunta, V.: Distance measures for effective clustering of arima time-series. In: Proceedings 2001 IEEE International Conference on Data Mining. pp. 273–280 (2001)
8. Kate, R.J.: Using dynamic time warping distances as features for improved time series classification. Data Mining and Knowledge Discovery **30**(2), 283–312 (2016)
9. Li, D., Bissyande, T.F., Klein, J., Traon, Y.L.: Time series classification with discrete wavelet transformed data. International Journal of Software Engineering and Knowledge Engineering **26**(09n10), 1361–1377 (2016)
10. Lin, J., Khade, R., Li, Y.: Rotation-invariant similarity in time series using bag-of-patterns representation. Journal of Intelligent Information Systems **39**(2), 287–315 (2012)
11. Schäfer, P.: The BOSS is concerned with time series classification in the presence of noise. Data Mining and Knowledge Discovery **29**(6), 1505–1530 (2015)
12. Senin, P., Malinchik, S.: Sax-vsm: Interpretable time series classification using sax and vector space model. In: 2013 IEEE 13th International Conference on Data Mining. pp. 1175–1180 (2013)
13. Wang, X., Smith, K., Hyndman, R.: Characteristic-based clustering for time series data. Data Mining and Knowledge Discovery **13**(3), 335–364 (2006)

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_26