Time series prediction using the adaptive resonance theory algorithm ART-2

A V Gavrilov, O K Alsova

Novosibirsk State Technical University, 20, Karl Marx ave., Novosibirsk, 630073, Russia

E-mail: andr_gavrilov@yahoo.com, alsowa@mail.ru

Abstract. The algorithm of the adaptive resonant theory ART-2 is based on the ideas of dynamic clustering and the unsupervised learning model. The classic application of the ART-2 algorithm is related to the solution of pattern recognition problems in the framework of the neural network approach. The article proposes a modification of the adaptive resonance theory ART-2 as applied to the solution of the time series (TS) prediction problem. A description of the TS forecasting algorithm based on ART-2, its properties and application features, as well as the results of a study of TS free electricity prices of the “day-ahead market” (DAM) in Russia is here. The obtained results allow us to conclude about the prospects of using ART-2 to study the structure and prediction of TS with a periodic (seasonal) component.

1. Introduction

The task of forecasting is a cornerstone in the field of Data Mining and in general in the field of AI, if we consider the problem of constructing AI as a problem of building bio-inspired intelligence. The latter is explained by the fact that the main function of the brain is the function of predicting the behavior of the external environment in order to obtain advantages for survival. In applied artificial intelligence systems, the problem of forecasting is usually considered as the problem of predicting the values of a time series describing the change in the price of shares or currency, meteorological indicators, water level in rivers, seismic activity, etc.

Currently, there are many different predictive models, such as the autoregressive integrated moving average (ARIMA) models, regression trees, recurrent neural networks and finally, an ensemble time series prediction approach, combining different models [1-4].

All these forecasting models somehow come down to using correlation between some set of values \( y_i, i = 1, ..., L \) so called window and the following value \( y_{L+1} \). In the case of using recurrent neural networks, it is better to talk about associative connection than about correlation, but the essence does not change.

Let us try to consider the problem of forecasting time series from the point of view of bionics, i.e. problem solving in a bio-similar way.

When solving problems of prediction by the biological brain, he does not have at his disposal the entire time series in which he can learn. The time series appears dynamically in the process of activity. Of course, the time series in this case is not a sequence of values of a single parameter, but a sequence of multidimensional images described by many different types of parameters that enter the system at arbitrary points in time. But to begin with, we simplify the problem, reducing it to the usual time series,
and we will further consider this simplification. It is logical to assume that the brain solves the problem of forecasting as a problem of recognizing patterns (situations) and for a similar image it uses a previously predicted value for a previously encountered one. Moreover, in the process of the system functioning, new, previously unseen images appear, which must be remembered and used later, i.e. there is ongoing (incremental) learning. Such process can be viewed as a dynamic clustering process implemented in an Adaptive Resonance Theory, proposed by Grossberg in 1976 [5] and developed by him together with Carpenter [6].

Until now, the applications of the ART-2 algorithm are known only for solving problems of pattern and situation recognition.

The article solves the problem of modification (adaptation) of the algorithm in order to apply it to study the structure and predict the values of the time series.

The study of the algorithm was carried out on the real TS of the “day-ahead market” (DAM) free electricity price. The data is provided by the Siberian Branch of PJSC RusHydro. At present, the task of short-term (hourly) forecasting of the DAM prices is becoming increasingly relevant due to the reform of the energy industry and the creation of the wholesale electricity and capacity market (WECM). The task of forecasting free hourly electricity prices is new for Russia and requires the development of efficient algorithms and forecasting techniques.

2. Description of the ART-2 algorithm as applied to TS prediction

The task of predicting next value of the time series can be reduced to the task of clustering. In this case, the cluster is recognized by the vector-window, and this cluster corresponds to the following (predicted) value (in the future will be called as value of cluster). Lower we give a mathematical description of the algorithm.

Initial data:

Let \( Y = (y_1, y_2, \ldots, y_N) \) – time series, describing the change of a certain indicator at discrete points in time \( t=1, 2, \ldots, N \). Values of TS are normalized and brought to measurement range from 0 to 1.

Parameters of the algorithm:

- \( R \) – radius of cluster;
- \( L \) – length of vector-window.

Represent time series \( Y \) as sequence of fragments (vector-windows) length \( L \):

\[ Y_i = (y_{i1}, y_{i+1}, \ldots, y_{i+L-1}), 1 \leq i \leq N - L + 1. \]

In matrix form:

\[ Y = [Y_1; Y_{N-L+1}] = [y_{ij}]_{i,j=1}^{N-L+1,L} = \begin{pmatrix} y_1 & y_2 & \ldots & y_L \\ y_2 & y_3 & \ldots & y_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N-L+1} & y_{N-L+2} & \ldots & y_N \end{pmatrix}. \]

The first step of the algorithm:

The first fragment of TS is considered \( Y_1 \). Fragment \( Y_1 \) associated with the first cluster, the center of a cluster is declared a vector \( C_1 = Y_1 \), weight of the cluster: \( w_1 = 1 \), and the predicted value for first cluster is assigned value following the fragment \( Y_1: k_1 = y_{L+1} \).

\( m \)-th step of the algorithm:

\( m \)-th fragment of TS \( Y_m \) id considered. Calculates the Euclidean distance between the fragment \( Y_m \) and the centers of the clusters that are formed in the previous steps:

\[ d_{Y_m, C_k} = \sqrt{\sum_{j=1}^{L} (y_{mj} - c_{kj})^2}, 1 \leq k \leq K, \]

where \( K \) – number of clusters. Minimum distance is:

\[ d_{\text{min}} = \min\{d_{Y_m, C_k}\}, 1 \leq k \leq K. \]
The predicted value is the value of the cluster with the minimum distance of Euclidean $d_{\text{min}}$: $\hat{y}_{L+m} = k_{\text{min}}$.

If $d_{\text{min}} > R$, then a new cluster is formed corresponding to the $m$-th fragment of TS $Y_m$ and cluster value: $k_{K+1} = y_{L+m}$.

If $d_{\text{min}} \leq R$, then the center of the cluster and its weight are recalculated (adapted to the new recognized fragment) according to the formulas:

$$C_{\text{min}} = C_{\text{min(old)}} + \frac{y_m - C_{\text{min(old)}}}{w_{\text{min(old)}} + 1},$$ (4)

$$w_{\text{min}} = w_{\text{min(old)}} + 1,$$ (5)

where $w_{\text{min(old)}}$ – weight of cluster $C_{\text{min}}$ at previous step, $C_{\text{min(old)}}$ – center of cluster at previous step.

The cluster value $k_{\text{min}}$ corresponding with cluster $C_{\text{min}}$ is recalculated according to the formulas similar to (4) and (5).

Below is a description of the algorithm in the form of pseudo code:

**Initial parameters:**

- $\text{Inp}(I,k)$ – $i$-th value of vector-window $\text{Inp}(k)$ (normalized);
- $\text{Nex}$ – number of vector-windows;
- $\text{Out}$ – value, following after vector-window (normalized)
- $\text{Pred}(i)$ – predicted value for $i$-th window;
- $\text{NC}$ – number of clusters;
- $\text{Cc}(i,j)$ – $i$-th value of vector-center of $j$-th cluster $C(j)$;
- $\text{VC}(j)$ – predicted value, i.e. value of $j$-th cluster;
- $\text{NRC}(i)$ – number of recognized vectors for $i$-th cluster (weight of $i$-th cluster).

**Algorithm:**

**Function Prediction:**

\[
\text{NC} := 0;
\]

For $k := 1$ to $\text{Nex}$ do

If $\text{NC} = 0$ then

Create of cluster with center $C(1) := \text{Inp}(1)$.

$\text{NC} := 1$;

$\text{VC}(\text{NC}) := \text{Out}$;

$\text{NRC}(\text{NC}) := 1$;

Else

$J_{\text{min}} := \text{Min} (\text{Dist}(\text{Inp}(k), C(j)))$;

If $\text{Dist}(\text{Inp}(k), C(J_{\text{min}})) > R$ then

Create of cluster with center $C(J_{\text{NC}}) := \text{Inp}(k)$

$\text{VC}(\text{NC}) := \text{Out}(k)$;

Else

$\text{Pred}(k) := \text{VC}(J_{\text{min}})$;

For $i := 1$ to $\text{NC}$ do

$C(i,J_{\text{min}}) := C(i,J_{\text{min}}) + (\text{Inp}(i) - C(i,J_{\text{min}}))/(1 + \text{NRC}(\text{J_{\text{min}}}))$;

$\text{NRC}(\text{J_{\text{min}}}) := \text{NRC}(\text{J_{\text{min}}}) + 1$;

$\text{VC}(\text{J_{\text{min}}}) := \text{VC}(\text{J_{\text{min}}}) + (\text{Out}(k) - \text{VC}(\text{J_{\text{min}}})/(1 + \text{NRC}(\text{J_{\text{min}}}))$;

End Function.

To study the proposed prediction algorithm, a program was developed in the Delphi 7.

3. The study of electricity prices based on the application of the proposed algorithm

The developed algorithm was used to predict the TS of the free price for electricity of the “day-ahead market”.
For the study, a fragment of TS of hourly prices for January 2017 was selected, a total of 744 observations. The initial TS was divided into two parts: training (544 observations) and test (last 200 observations).

The TS of prices under investigation has a non-linear structure, namely, a trend of complex shape, periodic and random components. To use the algorithm, the initial TS was represented as an expansion:

\[ Y(t) = F(t) + E(t), \]  

(6)

where \( F(t) \) – trend function described by a fifth degree polynomial:

\[ F(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4 + a_5 t^5, \]  

(7)

where \( a_i, i = 0, 1, ..., 5 \) – parameters of trend, \( E(t) \) – function, including periodic and random components, for the model description of which the algorithm was used.

To estimate the parameters of the trend component of TS, the numerical iterative Gauss-Newton method was used. According to the results of identification, the coefficient of determination, explained by the trend function, was \(-0.21\). Figure 1 shows the TS of electricity prices with the imposition of the trend model.

![Figure 1. TS of electricity prices with the imposition of the trend model.](image)

Further, the model description of the component \( E(t) \) was performed using our proposed algorithm. To estimate the periodicity of TS \( E(t) \) and to determine of optimal parameters of the algorithms the periodogram was built. According to the results of the analysis, the peak values of the periodogram correspond to periods 24 and 12. In this way TS \( E(t) \) has a clearly defined period of oscillation, which describes the change in the price of DAM during the day. Therefore, a window length of 24 was set as an algorithm parameter.

The figure 2 shows the TS of electricity prices with the imposition of a TS model, built with a radius of cluster 0.4. Analysis of the graph allows us to conclude that there is a good visual correspondence between the studied TS and TS model, both on the training and on the test parts of TS.
The table 1 shows the accuracy of TS forecasting of DAM prices for different values of the cluster radius for the training and test parts of TS, namely, the mean absolute error, the standard deviation of the error, the coefficient of determination. Accuracy indicators are given for the initial TS prices DAM $Y(t)$, including trend component $F(t)$.

| Sample  | Radius $R$ | Number of created clusters | Mean absolute error | Standard deviation | Coefficient of determination |
|---------|------------|-----------------------------|---------------------|-------------------|-----------------------------|
| Training | 0.5        | 51                          | 23.42               | 24.11             | 0.87                        |
| Test    | 0.5        | -                           | 25.64               | 53.43             | 0.66                        |
| Training | 0.4        | 92                          | 18.77               | 29.32             | 0.90                        |
| Test    | 0.4        | -                           | 25.38               | 32.34             | 0.68                        |

Analysis of the results indicates a high accuracy of the identification of TS $Y(t)$: the determination coefficient on the learning part of TS with a cluster radius of 0.5 is 0.87, on the test portion - 0.66. Excluding the percentage of variation attributable to the trend component of TS (21%), the contribution of the model description of the component $E(t)$ using the algorithm: 66% (training portion) and 45% (test portion). Accuracy of identification $E(t)$ increases with decreasing cluster radius and is up: 69% (training portion) and 47% (test portion). The number of created clusters when specifying a cluster radius of 0.5 and 0.4 is respectively 51 and 92. A relatively large number of clusters can be explained by the structure of the component, in which, apart from clearly defined oscillations with periods of 24 (day) and 12 (half a day), there are also periodicities of other orders, although their contribution to the change is much smaller.

4. Results and Discussion
The proposed algorithm was also tested on artificial TS of a periodic structure. According to the results of a comprehensive study of the algorithm, we can draw the following conclusions.

1. The accuracy of prediction of TS depends on the specified parameters of the algorithm (window length, cluster radius). As the cluster radius $R$ decreases, the number of clusters $K$ increases and the accuracy of the algorithm improves.
2. As the window length $L$ increases with the same cluster radius $R$, the prediction accuracy increases. The optimal window length for a TS with a clear periodic structure is equal to or a multiple of the TS period.

3. The proposed algorithm can be used to predict: seasonal TS with a clearly defined periodic structure, with a given oscillation period; TSs are characterized by repetition (within the statistical error) of certain sequences of values, and in the general case, the sequences can be repeated after an interval of time unequal to a constant.

The possibility of using an algorithm for predicting TS with non-equidistant moments of measuring TS values is most important, since virtually the entire classical mathematical apparatus of TS prediction is designed for discrete TS with equal intervals of measurement of the series.

The disadvantages of the algorithm include the absence of a collapsed parametric analytical description of the model, which complicates its interpretation. In addition, a sufficiently long observation period of the investigated TS is required for learning the algorithm, the accuracy of identifying the initial values of TS is low due to the small number of clusters.

5. Conclusion
The article provides a mathematical description and the results of a study of the time series prediction algorithm, based on the ideas of adaptive resonance theory.

The efficiency of the algorithm is illustrated by the example of analyzing the TS of free electricity prices. As a result of the identification, a model of the investigated TS was built, which has a high degree of adequacy to the initial data. The obtained results allow us to conclude about the prospects of using the algorithm for short-term forecasting of electricity prices for future periods.

The algorithm presented in the article can also be applied in any subject areas for the description and analysis of discrete time series with a periodic structure, and the moments of measurement of values of TS may not be equidistant. The absence of a tight binding of an algorithm to time is its most important property and significantly expands the scope of use of the algorithm.

References
[1] Newbold P and Granger C W J 1974 Experience with forecasting univariate time series and the combination of forecasts J. R. Statist. Soc. 137 131-164
[2] Granger C W J and Ramanathan R 1984 Improved methods of combining forecasts Journal of Forecasting 3 197-204
[3] Alsova O K, Gubarev V V and Abalov N A 2013 Toolkit for variative forecasting of time series Proc. the 8 international forum on strategic technologies – IFOST 2013 (Mongolia, Ulaanbaatar) vol 2 pp 280-282
[4] Batygin R I and Alsova O K 2016 Software system for different types of data classification based on the ensemble algorithms Proc. Int. Conf. Actual problems of electronic instrument engineering – APEIE–2016 (Novosibirsk: NSTU) vol 1 part 2 pp 506-509
[5] Grossberg S 1976 Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors Biological Cybernetics 23 121–134
[6] Carpenter G and Grossberg S 1991 Pattern Recognition by Self-Organizing Neural Networks (Cambridge, MA, MIT Press) p 712.