An investigation of metrics to compare significantly different approaches to time series forecasting

1st Ulyana Osipova
Department of Appl. Math and Cybernetics
SibSUTIS
Novosibirsk, Russia
uljana.pavlova2012@yandex.ru

2nd Anton Agalakov
Department of Appl. Math and Cybernetics
SibSUTIS
Novosibirsk, Russia
anthony.agalakov@gmail.com

***Abstract*—This article briefly discusses two fundamentally different approaches to time series forecasting. The first method is based on universal coding. It has been studied in various real-life problems and shown to be effective. The second is based on a ten-headed finite state machine, which effectively predicts multilinear sequences. As a result of work the the considered methods of forecasting time series show the result that are incomparable with each other. This makes their direct comparison difficult.The authors propose to convert the results into one form and use widely used metrics for comparison, such as Mean Absolute Error (MAE), Mean Square Error (MSE) and others. This standardization will facilitate a comprehensive comparison of the two methods, revealing their relative strengths and weaknesses across diverse datasets. Also description of the application of metrics to forecasting methods is given. The result of the work is a conclusion on the possibility of using one or another metric.**

***Index Terms*—forecasting, R-measure, finite state machines, time series analysis, metrics**

I. Introduction

There is a lot of controversy about the choice of the most appropriate error measure to evaluate the performance of the forecasting method. Statisticians insiston the use of measures with excellent statistical properties, but practitioners prefer measures that are easy to communicate and understand. The task of time series forecasting is extremely important for scientific research, because it has not a theoretical definition only, but also an applied nature. Regardless of the purpose of a particular forecasting method, any developer or scientist should be able to evaluate the efficiency of the algorithm before applying it to real data. There are many ways to conduct such an assessment, but they are not universal.

This paper considers two significantly different approaches to forecasting time series: one is based on the R measure, and the other is based on a finite automaton. As a result of their work, the methods produce indicators that are incomparable to each other. Therefore, it is necessary to bring them to one form: so that the accuracy of the predictions can be unambiguously assessed.

II. Forecasting based on universal coding

The method discussed in this section was proposed in 1988 [1] and has been studied in many works where it has shown high efficiency. Its main disadvantage was the speed of forecast calculation. Improvements were proposed in [2] that reduced the computational complexity to linear. And in [3] its effective implementation was described, which allows you to calculate the predicted value online.

Let the following sequence $x_1, ..., x_t, x_i \in A$, where A is finite alphabet, is generated by static and ergodic source.

In [4], a predictor was proposed, on the basis of which a representation of the Krichevsky measure (1) was obtained for Markov sources with fixed memory $m \geq 0$.

$$K_m(x) = \begin{cases} \frac{1}{|A|^t}, & t \leq m, \\ \frac{1}{|A|^m} \prod\limits_{s \in A^m} \frac{\prod_{a \in A} \Gamma(v_x(sa)+1/2)/\Gamma(1/2)}{\Gamma(\bar{v}_x(s)+|A|/2)/\Gamma(|A|/2)}, & t > m, \end{cases} \tag{1}$$

where $v_x(sa)$ is number of sequence sa occurrences in the sequence x, $\bar{v}_x(s) = \sum_{a \in A} v_x(sa)$, $x = x_1, ..., x_t$, $\Gamma()$ is gamma-function.

Let's define the measure R:

$$R(x_1, ..., x_t) = \sum_{i=0}^{\infty} \omega_{i+1} K_i(x_1, ..., x_t). \tag{2}$$

This measure is universal for the set of all stationary and ergodic sources and is based on an asymptotically optimal universal code.

It uses a probability distribution $\{\omega_1, \omega_2, ...\}$ on integers $\{1, 2, ...\}$:

$$\omega_i = \frac{1}{\log_2(i+1)} - \frac{1}{\log_2(i+2)}. \tag{3}$$

For each $a \in A$ sequence $x_1, ..., x_t, a$ is constructed and the conditional probability is calculated using the measure R:

$$P(a|x_1, ..., x_t) = \frac{R(x_1, ..., x_t, a)}{R(x_1, ..., x_t)}. \tag{4}$$

Thus, as a result, the method produces a probability distribution.

III. Forecasting with automaton

One of the key concepts of the further description is the concept of a multilinear sequence. Any infinite multilinear sequence can be represented as:

$$q \prod_{n \geqslant 1} \prod_{i \geqslant 1}^{m} p_i s_i^n \tag{5}$$

for some $m \geq 1$ and strings q, p_i, s_i divided by blocks of type $p_i s_i$ and an initial block q.

For the first time, an automaton capable of fully recognizing various multilinear sequences was presented in the work [4] in the form of pseudocode, which left room for its various practical implementations. A detailed implemented algorithm for the operation of such an automaton was described in the work [5]. Therefore, we will only present the operation schemes of the two main methods of the automaton - Correction and Matching.

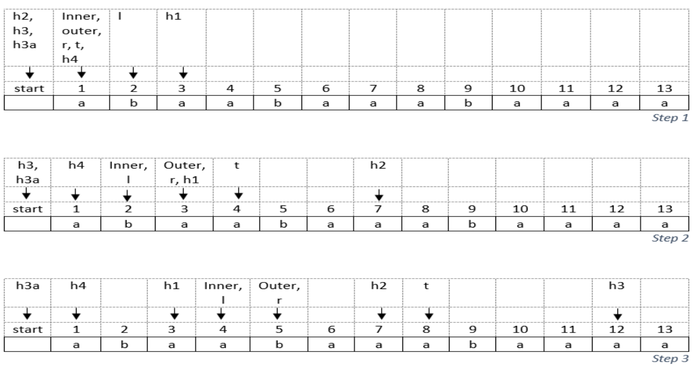

Fig. 1. The Correction's work scheme.

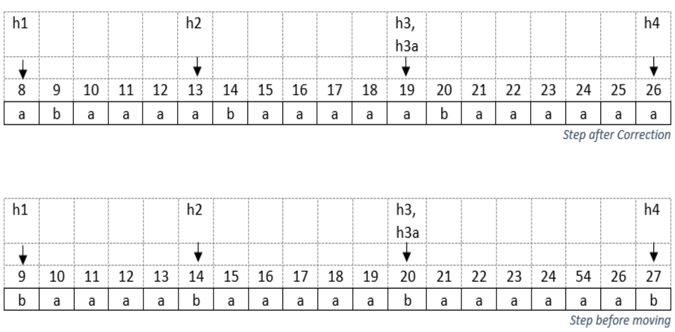

Fig. 2. The Matching's work scheme.

One of the main problems of such a finite automaton is its low efficiency on sequences that differ from the multilinear template described above. We can conclude that applying this algorithm in real task if impossible practically. Several modifications have been theoretically considered, described, and implemented as possible solutions to this problem. The most important of it is presented in [6].

The example below shows that as a result of its work, the automaton returns a predicted sequence of symbols from the input alphabet.

$$\prod_{n=1} \prod_{i=1}^{3} 0_i 1_i^n \longrightarrow 010110111 \text{ predicted sequence} \qquad (6)$$

Thanks to the algorithm improvements made by one of the authors, it is possible to see the number of correct and incorrect predictions in the resulting sequence. Obviously, such a result is incomparable with the result of the algorithm described in the previous section.

IV. Metrics

There are many opinions in modern scientific research about which metric is better to use. For example, the work [7] examines the impact of forecasting methods on forecast error for seven developed countries. The authors suggest that some of the differences in the resulting metrics can be explained by cultural differences, which suggest differences in the way forecasts are developed, the way forecasts are used, the extent to which models are used, and the subjective factors used.

In the work [8], the authors analyzed a large number of articles published between 2006 and 2017. The dataset under study was collected from forty articles selected according to the following criteria: the accuracy of the results presented in the selected articles is presented in terms of the normalized mean square root error (nRMSE) and the normalized mean bias error (nMBE). This allows comparison of samples of different sizes with data of different sizes, thereby providing a common basis for cross-comparative analysis. In fact, the authors proposed two new indicators to assess the quality of forecasts. In [9], the authors point out that forecasting methods are able to detect some anomalies in time series, but they conclude that existing quality assessment metrics overestimate the soft and underestimated cases. That is, a method cannot detect any anomalies if the forecast points to an incorrect range that is not associated with any anomalies (i.e., called a soft case). In addition, if the forecast indicates an insufficient range of anomalies(i.e., the so-called underestimated cases), then the anomaly cannot be detected. In addition, existing indicators do not take into account the duration of an incorrect forecast when evaluating, although long incorrect forecasts cause experts more inconvenience than short ones. The authors propose a new method to address these issues using two cross-referencing concepts and a weighting scheme. Based on these concepts, new metrics (i.e., eTaV and eTaff) are proposed.

V. Metrics for assessing forecast accuracy

The forecasting methods described above produce results in various forms. The first method is a probability distribution, the second is the directly predicted sequence. To obtain the predicted value in the first method, consider two approaches.

The first approach is to select the value with the highest probability.

$$x_{t+1} = argmax(P_1, ..., P_t), \qquad (7)$$

where $argmax()$ is the search function for a with the maximum value in the probability distribution P.

The second is based on calculating the mathematical expectation.

$$x_{t+1} = \sum_{i=1}^{t} a_i * P_i \qquad (8)$$

After this, a comparative analysis of the methods can be carried out for accuracy.

To assess the accuracy of forecasts, the MAE (Mean Absolute Error) metric is most often used in many works.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \bar{y_i}|, \tag{9}$$

where n is the number of observations, y_i is the actual values, and $\bar{y_i}$ is the predicted values. The advantages of this metric include ease of calculation and less sensitivity to outliers compared to others.

MSE (Mean Squared Error) is also used, which allows to show a large error (due to the quadratic dependence) and is also easy to calculate. Its disadvantages include sensitivity to emissions.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y_i})^2. \tag{10}$$

MAPE (Mean Absolute Percentage Error) is a dimensionless metric and is easy to interpret. Its weakness is the impossibility of application in cases where the values of the resulting variable are equal to zero.

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{|y_i - \bar{y_i}|}{|y_i|}. \tag{11}$$

SMAPE (Symmetric Mean Absolute Percentage Error) is a measure of accuracy based on percentage errors and is defined as follows:

$$SMAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{|y_i - \bar{y_i}|}{|y_i| + |\bar{y_i}|}. \tag{12}$$

Its advantages include correct processing of predicted values, regardless of whether they lag behind or exceed actual values. But the error increases sharply as the actual or predicted value approaches zero, since they are present in the denominator.

EXPERIMENT

The dataset from M3-Competition was chosen as a testing set. The set contains a little more than 3000 real numbers.For the first testing, the authors decided to include in the algorithms the calculation of SMAPE only.

The operation of the both methods involves a training stage at first and then a forecasting stage. The initial part of the sequence was used as a training sample. The testing was carried out using a training sample of various lengths. The table below shows the best results presented when predicting 20 values.

TABLE I
TEST RESULTS

Metric\method	Automaton	Univ.coding
sMape	7.3	6.8

It is worth noting once again that the table shows the best result among all those obtained for the selected data set. In general, the results are quite comparable with each other.

CONCLUSION

The article discusses several of the most popular metrics for assessing the accuracy of time series forecasting methods. Prediction based on the R measure gives a probability distribution, i.e. it shows the probability with which a certain element of the alphabet will appear next, while prediction based on state machines gives the predicted sequence directly. Based on the above, we can conclude that none of the presented assessment methods is suitable for both methods developed by the authors simultaneously. However, by additional modifications of the methods, the ability to calculate one metric was added to them. The results obtained are encouraging, but are not sufficient for specific conclusions. The authors propose to post-process the results of the first method so that other existing metrics can be used to assess the accuracy of the forecast.

REFERENCES

[1] B. Y. Ryabko, Prediction of random sequences and universal coding, Problems of information transmission, vol. 24, pp. 87–96, 1988.
[2] A. Rakitskiy, Efficient Algorithms for Time Series Prediction Method, 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), Russian Federation, pp.120–123, November 2022.
[3] A. Agalakov, ”Osobennosti practicheskoy realizatsii effectivnogo metoda prognozirovaniya na osnove universalnogo kodirovaniya [Features of practical implementation of an effective forecasting method based on universal coding]”. Rossijskaja nauchno-tehnicheskaja konferencija, Novosibirsk, April, 25–26, 2023, pp. 106–111.
[4] T. Smith, Prediction of infinite words with automata, Theory of Computing Systems, 2018. V. 62, p. 653–681.
[5] Pavlova U., Rakitsky A. “Issledovanie vozmozhnosti primenenija avtomatov dlja prognozirovanija vremennyh rjadov [Investigation of the possibility of using automata to predict time series]”. Rossijskaja nauchno-tehnicheskaja konferencija, Novosibirsk, April, 25–26, 2019, pp. 168–172.
[6] U. V. Pavlova, Rakitskiy A.A., ”Primenenie spetsializirovannih deteremenirovannix konechnix avtomatov dlya prognozirovaniya vremennix ryadov [Application of specialized deterministic finite automata for time series forecasting]”, Vectnik SibGUTI, vol. 2, pp. 12–22, 2022.
[7] J. G. Wacker, L. G. Sprague, Forecasting accuracy: comparing the relative effectiveness of practices between seven developed countries, Journal of operations management, vol. 16, pp. 271 - 290, 1998.
[8] R. Blaga, et al. A current perspective on the accuracy of incoming solar energy forecasting, Progress in energy and combustion science, vol. 70, pp. 119 - 144, 2019.
[9] W. S. Hwang, et al. Do you know existing accuracy metrics overrate time-series anomaly detections?, Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing, pp. 403 - 412, 2022.