Development of an aggregator for choosing the best forecasting method for groups of time series

Vadim Moshkin, Irina Moshkina, Dmitry Yashin, Nadezhda Yarushkina
Information Systems department, Faculty of Information Systems and Technologies
Ulyanovsk State Technical University, UlSTU
Ulyanovsk, Russia
v.moshkin@ulstu.ru, i.timina@ulstu.ru, taurusrulez@yandex.ru, jng@ulstu.ru

***Abstract*—The paper presents a general algorithm for computing an aggregated time series forecast (TS), within which machine learning methods are used to adjust the parameters of a hybrid combined forecasting model. Also presented are the results of experiments on the application of the developed algorithm using the TS competition “Computational Intelligence in Forecasting” (CIF). The use of a neural network for choosing forecasting methods allowed, on average, for all experiments to reduce the error by 7.225%, as can be seen from the results of the experiments. The average error for the eight prediction methods chosen by the neural network turned out to be lower than the average error for all methods in 47 cases out of 50 (94%)**

Keywords*—Time series, forecasting, aggregated forecast, machine learning

I. Overview of forecast aggregation methods

There are many methods for predicting vehicles. Combined models are used to take advantage of several methods at once. According to [2], a combined forecasting model is a forecasting model consisting of several individual models, called a base set.

In [3], a number of factors are highlighted that emphasize the effectiveness of the combined model:

1. impossibility of choosing a single model based on experimental data, according to the theory of multiple models [4];
2. an attempt to choose the only best model leads to the need to choose from a group of models with similar statistical characteristics [5];
3. The choice of a forecasting model for a vehicle with a pronounced dynamics of level and properties leads to the choice of an averaged model [2]. It is impossible to quickly replace one forecasting model with another by analyzing its dynamics;
4. each forecasting model considers only one side of the dynamics of the analyzed process. The set of models allows a more detailed description of the dynamics. Any forecast rejected due to non-optimality contains information important for modeling [5].

According to [2], combined forecasting models are divided into selective and hybrid ones.

In the selective model, the current predicted value is calculated from the selected value according to the selective criterion of the model from the base set.

The selective criterion can be:

- the minimum of the absolute value of the forecast error of the current member of the series
- minimum of the absolute value of the error for the last K observations (K-test)
- the minimum of the exponentially smoothed squared deviation error (B-criterion).

Thus, when using a selective model, at each moment in time, the forecast is built according to a single method selected from the basic set, hybrid models, in turn, make it possible to build a forecast using several models at once, using the advantages of their joint application.

In a hybrid model, the predicted value is obtained by aggregating the predicted results from several models from the base set. As a rule, the final forecast is a weighted sum of individual forecasts.

In [6], for the first time, the idea of creating a hybrid model based on combining forecasts of several statistical models was substantiated; in [7], this idea was developed, and it was proposed to use the arithmetic mean of forecasting results of the models included in the base set [3] as the final forecast.

According to [5], the main problem of constructing hybrid forecasting models is to determine the optimal weights of individual forecasting models from the base set.

In [3], the following main directions of development of hybrid forecasting models are identified:

1. inclusion in the basic set of new (emerging) forecasting models
2. development of new methods for combining forecasts.

There are 7 main groups of methods for combining forecasts [8]:

1. Methods based on the arithmetic mean of particular predictions [6] [9] [10]. However, the presence of anomalous forecasts as part of a combined forecast

significantly reduces its accuracy [11]. It is proposed to exclude anomalous predictions by using a truncated arithmetic mean [12] [13].
2. Methods based on minimizing the final forecast error by the least squares method [14].
3. Methods based on minimizing the variance of the combined forecast error (works by Bates and Granger [6], Ershov [15], Baltrushevich [16]).
4. Methods based on retrospective forecasts. This group includes the AFTER method [17]. The weights of the private forecasts are calculated based on their own past values, conditional variance, and the past values of the private forecasts. The weights are updated after each new observation.
   The following disadvantages of the AFTER method were noted in [8]:
   - difficult applicability in practice;
   - strong dependence of the weights on the first set value.

   This group includes the following methods:
   - ARM, developed by Yang [18];
   - the Bunn method [19], which assumes finding the distribution function for the weight coefficient through the beta distribution;
   - an adaptive method based on exponential smoothing [2], [20].
5. Methods based on factor analysis. These methods were proposed by Frenkel [21] and Gorelik and Frenkel [5]. The idea of using factor analysis is based on the fact that particular forecast results using a separate forecasting method are an external expression of some really existing but immeasurable forecast value, which is taken as a combined forecast [8].
6. The method of Gupta and Wilton, based on finding the optimal weights of the coefficients of particular predictions using a matrix of pairwise preferences, has been placed in a separate group [22].
7. Methods based on quadratic programming. The paper [23] describes a method for calculating the weights of particular predictions by minimizing the retrospective relative errors of particular predictions using quadratic programming methods.

The main advantage of the method is efficiency and ease of implementation. The main disadvantage is the obligatory preliminary selection of particular forecasting methods in order to comply with the requirement of error independence [8].

Most of the methods for combining forecasts are based on the assumptions about the independence of the absolute forecast errors and their distribution in accordance with the normal law with zero mathematical expectation and unknown variance. However, these assumptions are often not met [3], and therefore, methods based on fuzzy logic and stable statistical estimates are currently being actively developed, for example:
1. method of combining forecasts by Kovalev [24] based on a system of fuzzy rules;
2. the Davydov union method [25], based on the use of a robust M-estimate;
3. Methods for combining particular forecasts by Vasiliev [26] based on the robust Huber estimate of the truncated mean type and on the basis of the Hodges-Lehmann R-estimate.

Thus, despite a significant number of publications on the topics of forecasting methods for time series and methods for aggregating individual forecasts, the question of choosing the most appropriate aggregating method and its constituent forecasting models for the predicted time series remains.

II. Developed algorithm for calculating the aggregated forecast of time series

Figure 1 shows a schematic description of the developed algorithm for calculating the aggregated forecast of time series.

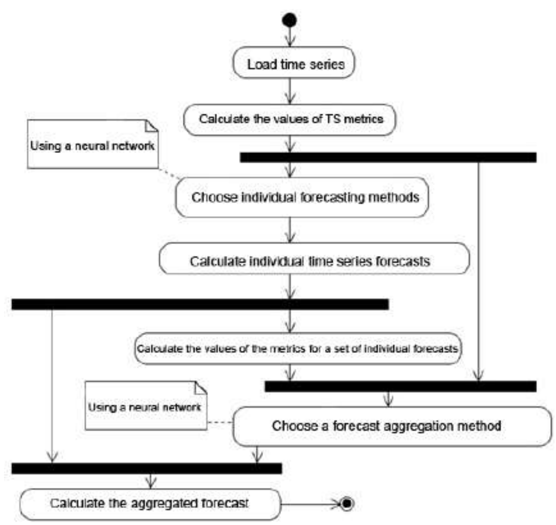

Figure 1. Algorithm for calculating the aggregated forecast of time series

In this paper, 2 methods of setting the forecast weights are considered:
- the first method is based on the values of the prediction error on the control part of the time series;
- the second method is based on the error values assumed by the neural network for choosing a prediction method.

The structure of the neural network for choosing the aggregating method is close to the structure of the neural network for choosing individual prediction methods, but it includes more input neurons corresponding to the metrics. Neurons corresponding to individual prediction methods have been replaced with neurons corresponding to aggregation methods. The structure of the neural network of the aggregating method is shown in Figure 2.

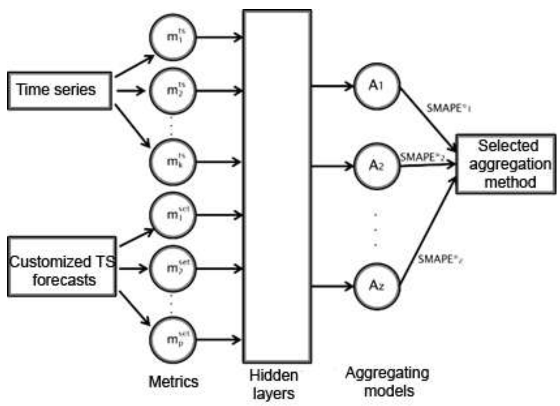

Figure 2. The structure of the neural network for choosing the aggregating method

Input neurons $m_1^{ts}, \ldots, m_k^{ts}$ correspond to time series metrics. Neurons $m_1^{set}, \ldots, m_p^{set}$ set correspond to the metrics of the aggregated set of individual predictions. The output values correspond to the estimated prediction error (SMAPE) values calculated by the neural network for each aggregator $(A_1, \ldots, A_z)$ from the base set. The aggregator from the base set of the combined model with the lowest estimated error value is used to obtain the final forecast.

The following main reasons for choosing just such a set of metrics for the input layer of the neural network can be distinguished:

- it is difficult to correctly train a neural network if the input neurons correspond to individual forecasting methods, since different individual methods will be selected for each time series. This means that the values of the signals arriving at the input of the neural network under consideration will be equal to zero for the unselected methods;
- the choice of the aggregator depends on the values of the time series metrics, but transitively;
- the direct dependence of the choice of an aggregator based on particular forecasting results is implemented through the metrics of a set of individual forecasts.

An error backpropagation algorithm with a logistic activation function is used to train the neural network. The training sample file contains the metric values and prediction errors (SMAPE) for each time series included in the set for each aggregation method.

This method of setting the weights includes dividing the time series into training and control parts, followed by forecasting by each individual method using the training part of the control values and calculating the prediction error. The weights of individual forecasts as part of the aggregator are set in inverse proportion to the magnitude of the error of each method.

III. Software system and experimental results

The developed program is designed to solve the problem of obtaining an aggregated forecast for the time series of the states of a technical system.

The software product is developed on the .NET Framework 4.6.1 platform in the C # language. The development environment was Microsoft Visual Studio 2015.

The "neuralnet" library [27] for the R language was used to work with neural networks. It made it possible to create neural networks with the structure of a multilayer preceptron, trained by the method of back propagation of the error (ordinary or elastic propagation). This library has a user-friendly interface and a high degree of configuration flexibility, allowing you to select the activation function and the error function.

The R library "ForecastComb" was used to compute the aggregated forecast. This library contains more than 12 aggregation methods (fig. 3).

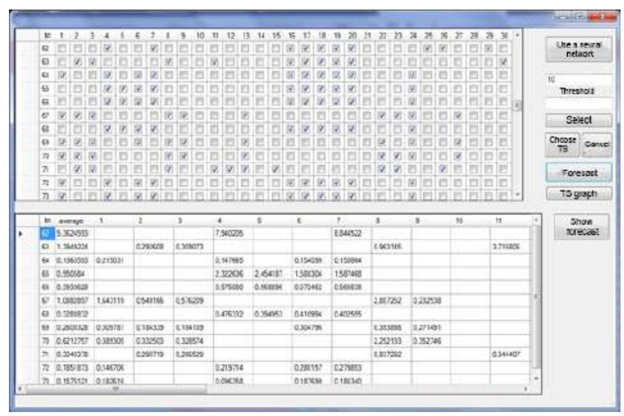

Figure 3. Forecast aggregation system form

Time series from the Computational Intelligence in Forecasting (CIF) competition [28] were selected to test the effectiveness of the developed solution.

- The first CIF benchmark contains 91 time series of different lengths (from 12 to 1089 observations) and different frequency of observations: day, month, quarter, year.
- The second CIF benchmark includes 72 time series with a frequency of a month and a length of 28 to 120 observations.

Five experiments with identical algorithm were carried out. The averaged final result was obtained for them.

1. A set of 152 time series of the benchmark was randomly divided into training (142 time series) and control parts (10 time series) during each experiment.
2. The time series of the control part were excluded from the general table of the training sample.

3. The neural network for choosing prediction methods (with automatic selection of the optimal number of neurons) was trained using the remaining data.
4. The resulting neural network was used to select the 8 best forecasting methods from the base set for the time series of the control part of the general table [29].

Figure 4 is a diagram showing, for each of the five experiments, the average SMAPE error for best practices and SMAPE for all methods.

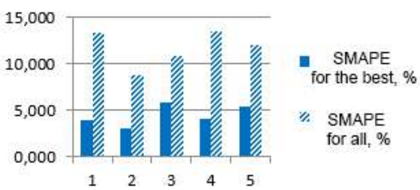

Figure 4. Results of the conducted experiments

IV. Conclusion

The use of a neural network for choosing forecasting methods allowed, on average, for all experiments to reduce the error by 7.225%, as can be seen from the results of the experiments. The average error for the eight prediction methods chosen by the neural network turned out to be lower than the average error for all methods in 47 cases out of 50 (94%).

Acknowledgment

This work was supported by the RFBR. Projects No. 18-47-730022, No. 19-07-00999 and 18-47-732007.

References

[1] Averkin, A.N. Soft computations and measurements // Intelligent systems. 1997.Vol. 2, no. 1-4. pp. 93-114.
[2] Lukashin, Yu.P. Adaptive methods of short-term forecasting of time series / Moscow, Finance and statistics, 2003. – 416 p.
[3] Vasiliev, A.A. Genesis of hybrid forecasting models based on the combination of forecasts /Bulletin of TVGU. Series "Economics and Management". - 2014. - No. 23. - pp. 316-331
[4] Ivakhnenko A.G. Inductive method of self-organization of models of complex systems: monograph / Kiev: Naukova Dumka, 1981 .—296 p.
[5] Gorelik N.A. Statistical problems of economic forecasting / N.A. Gorelik, A.A. Frenkel // Statistical methods of analysis of economic dynamics: textbook. app. on statistics, vol. 46. - M .: Nauka, 1983. - pp. 9-48.
[6] Bates J.M., Granger C.W.J. The Combination of Forecasts. – Oper. Reser. Quart., 1969, v.20, No. 4. P. 451-468
[7] Newbold P., Granger C.W.J. Experience with Forecasting Univariate Time Series and Combination of Forecasts. – J. of Royal Statistical Society. Ser. A. 1974, v. 137, No. 2. P. 131-164
[8] Frenkel, A.A. Methodological approaches to improving forecasting accuracy by combining forecasts / A.A. Frenkel, A.A. Surkov. - M: Questions of statistics. - 2015. - No. 8. - pp. 17-36
[9] Barnard G.A. New methods of quality control //J. R. Statist. Soc.. 1963. Vol. 126. P. 255-259
[10] Makridakis S. and Winkler R.L. Averages of forecasts: some empirical results //Management Science. 1983. Vol. 9. P. 987-996
[11] Winkler, R.L., &Clemen, R.T. Sensitivity of weights in combining forecasts // Operations Research. 1992. Vol. 40. P. 609-614
[12] Goodwin P. New evidence on the value of combining forecasts // FORESIGHT. 2009. Vol. 12. P. 33-35
[13] Jose V.R.. and Winkler R.L. Simple robust averages of forecasts: some empirical results //International Journal of Forecasting. 2008. Vol. 24. P. 163-169
[14] Granger C.W.J. and Ramanathan R. Improved methods of combining forecasts // Journal of Forecasting. 1984. Vol. 3. P. 197-204
[15] Ershov E.B. On one method of combining private forecasts // In the book: Statistical methods of analysis of economic dynamics. Uch. app. according to statistics / E.B. Ershov. Moscow: Nauka, 1973.T. XXII-XXIII. pp. 87-105
[16] Baltrushevich, T.G. Models and methods for assessing the effectiveness of flexible production systems // Avtoref. dis. ... Cand. econom. sciences. - M., 1991.pp. 17-20.
[17] Yang Y. Combining forecasting procedures: some theoretical results // Department of Statistics and Statistical Laboratory Iowa State University. 2001
[18] Yang Y. Adaptive regression by mixing //Journal of American Statistical Association. 2001b. Vol. 96. P. 574-588
[19] Bunn D.W. A Bayesian approach to the linear combination of forecasts // Operational Research Quarterly. 1975. Vol. 26. P. 325-329
[20] Dubrova T.A. Statistical analysis and forecasting of economic dynamics: problems and approaches // In the book: Methodology of statistical research of socio-economic processes. - M .: Unity, 2012.pp. 129-138.
[21] Frenkel A.A. Forecasting labor productivity: methods and models. - M .: Economics, 1989.pp. 142-154.
[22] Gupta S. and Wilton P. C. Combination of forecasts: an extension // Management Science. 1987. Vol. 3. pp. 356-371
[23] Beilinson, Ya.E. Combined forecast models / Ya.E. Beilinson, M.A. Motova // Expert information. Series: Modeling of socio-economic processes. M., 1990. Issue. 2.pp.110-121.
[24] N. G. Yarushkina ; A. A. Filippov ; A. A. Romanov ; V. S. Moshkin ; E. N. Egov Developing a system for time series data mining on the basis of F-transform and the domain-specific ontology // Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSA-SCIS), 2017 Joint 17th World Congress of International//IEEE, pp.1-6.
[25] Davydov, A.A. Forecasting social phenomena using "neural" networks / Sociological methods in modern sociological practice: collection of articles. materials Vseross. scientific. conf. in memory of A.O. Kryshtanovsky. - M .: GU-HSE, 2008 .– pp. 41-49.
[26] Vasiliev, A.A. Forecasting in logistics based on stable hybrid models / A.A. Vasiliev, V.M. Kurganov // Logistics: modern development trends: materials of the X international. scientific-practical conf. April 14, 15, 2011; ed. count .: V.S. Lukinsky (editor-in-chief) [and others]. SPb .: SPbGIEU, 2011. - pp. 52-55
[27] Yashin D., Moshkina I., Moshkin V. (2020) An Approach to the Selection of a Time Series Analysis Method Using a Neural Network. In: Gervasi O. et al. (eds) Computational Science and Its Applications – ICCSA 2020. ICCSA 2020. Lecture Notes in Computer Science, vol 12249. Springer, Cham. pp 692-703, https://doi.org/10.1007/978-3-030-58799-4_50
[28] Ayvazyan, S.A. Applied statistics and foundations of econometrics / S.A. Ayvazyan, V.S. Mkhitaryan. - M.: UNITI, 1998 .– 1024 p.
[29] Yarushkina N., Andreev I., Moshkin V., Moshkina I. (2019) Integration of Fuzzy OWL Ontologies and Fuzzy Time Series in the Determination of Faulty Technical Units. In: Misra S. et al. (eds) Computational Science and Its Applications – ICCSA 2019. Lecture Notes in Computer Science, vol 11619. Springer, Cham, pp 545-555, https://doi.org/10.1007/978-3-030-24289-3_40

Разработка агрегатора для выбора наилучшего метода прогнозирования групп временных рядов

Мошкин В.С., Мошкина И.А.,
Яшин Д.С., Ярушкина Н.Г.

В статье представлен общий алгоритм вычисления агрегированного прогноза временных рядов, в рамках которого используются методы машинного обучения для настройки параметров гибридной комбинированной модели прогнозирования. Также представлены результаты экспериментов по применению разработанного алгоритма с использованием временных рядов конкурса «Вычислительный интеллект в прогнозировании» (CIF). Использование нейронной сети для выбора методов прогнозирования позволило в среднем по всем экспериментам снизить ошибку на 7,225%. Средняя ошибка для восьми методов прогноза, выбранных нейронной сетью, оказалась ниже средней ошибки для всех методов в 47 случаях из 50 (94%).

Received 06.05.2021