Overview of "Sufficient" Forecasting Models

Merab Pkhovelishvili
Faculty of Informatics and Control Systems
Georgian Technical University
Tbilisi, Georgia
e-mail: m.pkhovelishvili@gtu.ge

Natela Archvadze
Faculty of Exact and Natural Sciences, Ivane Javakhishvili Tbilisi State University,
Tbilisi, Georgia
e-mail: natela.archvadze@tsu.ge

Zaza Tsiramua
Faculty of Informatics and Control Systems
Georgian Technical University
Tbilisi, Georgia
e-mail: zaza@gtu.ge

Mariam Pkhovelishvili
GTS Group - Director
Tbilisi, Georgia
e-mail: pkhovelishvilimariami@gmail.com

https://doi.org/10.62343/csit.2024.14

***Abstract* — A new hybrid forecasting model based on the use of parallel data is considered, which allows you to select a pair of forecasting models (or a combination of three, four, etc. models), the "joint" forecast probability of which gives a much better result than each model separately. We eliminated models that were not "Necessary" from consideration and used a "sufficient" number of models together, thereby further increasing the accuracy of the forecast. "Necessary" predictive models are models whose set of predictions always includes events that occur, while "sufficient" predictive models are models whose predictions always come true.Keywords—Computer science, informatics, information technology.**

***Keywords* — Forecasting models, hybrid model, decision making**

I. INTRODUCTION

When modeling any processes, when it is necessary to test hypotheses about asynchronous processes, forecasting models are often used, which makes it possible to analyze their effectiveness. Based on existing models, it is possible to create a new hybrid model [1], the purpose of which is to increase the accuracy of the forecast. This increases the relevance of the topic under discussion.

One popular forecasting method is the Bayesian method, developed by many scientists (Student, Fisher, Neumann, and Pearson, as well as the conditional Bayesian method) and is mainly used to prove asynchronous hypotheses. Our goal is to develop a new method based on the use of parallel data, thereby reducing the number of models used in the Bayesian method and, consequently, the computation time.

II. CLASSIFICATION OF FORECASTING MODELS AND HYBRID MODEL

The hybrid forecasting model we created [1, 2] using a parallel data algorithm [3] is effective in various economic and financial problems, especially in those tasks for which there are several forecasting models. Each of these models has its own advantages, but none of these models provide the ability to determine predictive value with the smallest possible error.

Concurrent data (or data sets) are different types of data that influence (or predict) the same event. For example, parallel data during a particular disaster is a collection of the following data: geological, meteorological, and hydrological parameters, historical and current data, including interpreted information from satellite, radar, and aerial photographs, as well as data obtained from hydrometeorological/geological field surveys.

The likelihood of each pattern occurring may be small, but under certain conditions, when a coincidence occurs in time and place, this is already a sign of tragedy.

We consider the intersection of sets of forecasts given by two, three, and usually n forecasting models. The superiority of selecting pairs of models over any single best forecasting model is theoretically justified. The advantage of the model obtained from the ratio of predictions of the n number of models compared to the n-1 number model [4] is shown.

The best pair of models is a pair that has no or very little overlap in predictions.

The time for recalculation of pairs of forecasting models is set - after each past event, and also to locally take into account "approximate coincidences" of forecasts when building a new model.

When constructing a hybrid model, existing forecasting models were analyzed and two types of models were identified: "necessary" forecasting models and "sufficient" forecasting models. Their characteristics are given in the article [5, 6]. "Necessary" predictive models are those whose set of predictions always includes events that occur, while "sufficient" predictive models are those whose predictions always come true.

Suppose there are $A_1, A_2, \ldots, A_n$ models for predicting a certain event, where n is the number of models being considered. Let's assume that each model is "necessary", that is, it always predicts an event that will occur. These models do not take into account models that are "not needed" because they cannot predict the event.

In practice, there are very few "sufficient" models, and there is no single model that predicts all events, so the question is whether we can use these "sufficient" models in such a way that a combination of them can predict all events. that is, a set of "sufficient" models will simultaneously become "necessary".

III. "SUFFICIENT" PREDICTIVE MODELS

A model in which all the predictions are correct, as we have already said, is "sufficient", but there are events that were not predicted, although various combinations of these models lead to a model close to the required one. For example, consider a joint forecast of two models using a hybrid model (Fig. 1). Let the rectangle denote all events, and the circles denote 2 "sufficient" models that are built according to P_i and P_j on the antecedents and assume that they have a non-empty intersection P_{ij} (green in the figure), then combining them results in a set of event predictions. The joint forecast is outlined in the figure with a black line. In the figure, we see the regions that were not predicted by the joint prediction of these two "sufficient" models (purple rectangle in the figure).

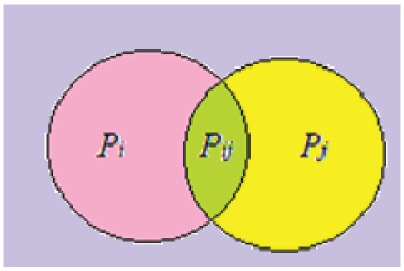

Figure 1. Joint prediction of two "sufficient" models.

Similarly, if we increase the number of models and consider three models (Fig. 2), we will see that the forecast when three models work together is larger, and closer to all combined forecasts (the purple rectangle is smaller than with two models).

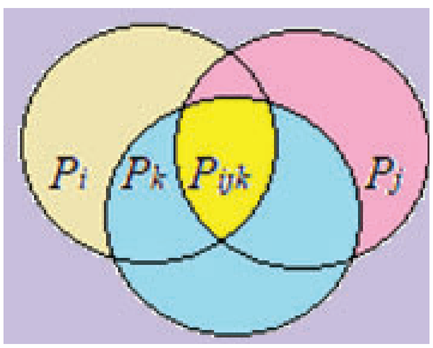

Figure 2. Joint prediction of three "sufficient" models

When using multiple "sufficient" models together, it is necessary to study which predictions were repeated in the intersection of their predictions (what happened in their intersection was repeated). This means that there are predictions that some or most models think will happen. Of the common parts, one part remains, and the other is not taken into account. Particular attention should be paid to the fact that if many models have the same predictions, then a given event will occur, but the probability of this prediction will need to be determined.

It is necessary to study the predecessor who repeats all the predictions. For example, in the case of an earthquake, several methods rely on the methods of light from a celestial body. For one model this was a prerequisite, for the second, third, fourth... This means that the brightness of the celestial body is one of the main prerequisites.

It's also important to consider how often these models make predictions. The model M_i predicted 10 times, Mj model did this 20 times, and they did everything correctly because there are "sufficient" models. But M_k model could only make a prediction once, and it was correct. We need to calculate the frequency of predictions of these models. Such a "sufficient" model, which produced only one prediction, does not need to be included in the discussion. It's good when the model made a forecast many times and it came true just as many times. Therefore, for "sufficient" models, the frequency S must be defined in % as the ratio of the number of predictions issued (n) to the number of predictions issued (m). The formula will look like:

$$S = \frac{n}{m} 100\%$$

For example, if M_1 model predicted 20 times, M_2 model has 10 times and a total of 40 earthquakes, then the frequencies of these models will be:

$$S_1 = \frac{20}{40} 100\% = 50\%\,,\ S_2 = \frac{10}{40} 100\% = 25\%$$

Let us consider the Bayesian method [7] for calculating a single forecast of "sufficient" models. Let's carry out calculations similar to those used in the Bayesian method - apply the same formula to the models and calculate the frequency of combining these two frequencies S_1=50% and S_2=25% of the frequency (Fig. 3):

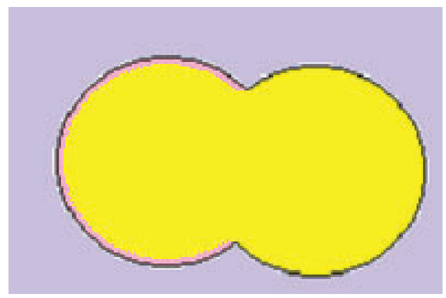

Figure 3. Calculation of the frequency of joint prediction of two "sufficient" models

where M_1 and M_2. A combination of models is mentioned:

$$P(M_i \cup M_j) = \frac{k}{m} 100\ \%\,,$$

where $k = n_i + n_J - n_p$, where n_i - this is a number of model predictions M_i, n_J —this is number of model predictions M_j, and - n_p is the number of intercept terms $M_i \cap M_j$ (we assume is 5). Let's calculate the joint prediction frequency for our example:

$$S_k = \frac{20+10-5}{40} 100\% = \frac{25}{40} 100\% = 62.5\%$$

We found that the prediction rates of individual models (50% and 25%) increased, which was expected since we used a hybrid forecasting model.

CONCLUSION

Based on the analysis of existing forecasting models, we divided the models into "necessary" and "sufficient", built a hybrid forecasting model with parallel data from pairs and triplets of models, excluded "unnecessary" models from consideration, and used "sufficient" models together, which further increases the accuracy of forecasting.

REFERENCES

1. Aliyev, V., Pkhovelishvili, M., Archvadze, N., Gasitashvili, Z., Tsiramua, Z. "Hybrid Prediction Method Using Experts and Models". 2023 5th International Conference on Problems of Cybernetics and Informatics, PCI 2023, 2023.
2. Gasitashvili, Z., Pkhovelishvili, M., Archvadze, N. Prediction of events means of data parallelism. Proceedings - Mathematics and Computers in Science and Engineering, MACISE 2019, 2019, pp. 32–35, 8944725. https://ieeexplore.ieee.org/abstract/document/8944725
3. Pkhovelishvili, M.G., Archvadze, N.N., Nikoleishvili, M.M. "Analysis of earthquake prediction models to obtain the best model". Geologiya i Geofizika Yuga Rossii, 2023, 13(1), pp. 162-172. DOI: 10.46698/VNC.2023.61.18.012
4. Prangishvili, A., Gasitashvili, Z., Pkhovelishvili, M., Archvadze, N. "Predicting Events by Analyzing the Results of the Work of Predictive Models". Communications in Computer and Information Science, 2022, 1562 CCIS, pp. 64–78.
5. Prangishvili, A., Gasitashvili, Z., Pkhovelishvili, M., Archvadze, N. "Theory of universal approach to improve predictive models using parallel data and application examples." IP Conference, 2023, 2757, 020004.
6. Aliyev, V., Gasitashvili, Z., Pkhovelishvili, M., Archvadze, N. Algorithm for Building a Hybrid Model of the Existing Risk model. Reliability: Theory and Applications, 2022, 17, pp. 106–110.
7. Stuart, A.; Ord, K. (1994), Kendall's Advanced Theory of Statistics: Volume I – Distribution Theory, Edward Arnold, §8.7