RT&A, Special Issue № 3 (66)
Volume 17, January 2022

REDUCING RISKS THROUGH IMPROVEMENT OF PREDICTION MODELS

Zurab Gasitashvili[1], Merab Pkhovelishvili[2], Natela Archvadze[3]

•

[1]Georgian Technical University, Tbilisi, Georgia
zur_gas@gtu.ge
[2]Muskhelishvili Institute of Computational Mathematics, Georgian Technical University,
merab5@list.ru
[3]Ivane Javakhishvili Tbilisi State University, Tbilisi, Georgia
natela.archvadze@tsu.ge

Abstract

Management or avoidance of risks or mitigation of undesirable outcomes are linked to specific actions, as well as to prediction models. These prediction models should be improved to obtain "better" predictions and thus, manage risks, and take measures for their reduction. We consider such algorithm of event prediction, which, using parallel data, can obtain prediction with high reliability that, in its turn, helps to reduce risks or completely avoid them

Keywords: Risk management, Parallel data, Prediction models, model improvement algorithm.

I. Introduction

At this point, fundamental research studies on safety theory and risks are very important. They can be practically used for reducing risks in the fields of industry, energy, transport, construction, and agriculture [1].

There are many risks in our world. Some of them impact individuals, others pose danger for the entire society, and some specific risks impact only certain fields and activities. Because of damage caused by risks, protection from them and avoidance of negative outcomes are very important.

Risk management has great importance in the economy, as well as risk reduction. This mainly applies to those risks, which can be identified. These processes should be equipped with the appropriate models and procession of information needed for identification. As it is discussed in [2], systemic use of all existing information is one of the main parts of risk analysis, which allows to evaluate the risks of undesirable accidents and events.

For reduction of risks, we consider the algorithm for prediction improvement built by us, which helps to predict such events, as the risks associated with natural disasters. In particular, we single out prediction of earthquakes, landslides and mudflows.

II. Use of improved algorithm of earthquake prediction models for avoidance of undesirable risks

Let us review several models of earthquake prediction specifically for Georgia. The information was taken from the online map of earthquakes [3], where there are maps, lists, data and information on earthquakes, and a seismic map of the world. In Table 1, the list of earthquakes occurred on the territory of Georgia is given, which belong to the earthquakes with moderate strength (magnitude 4-5). We took earthquake magnitude, occurrence date, time and name of epicenter as characteristics of each earthquake. The table contains earthquakes occurred in 2020-2021. UTC means the Coordinated Universal Time.

Table 1: *Earthquakes occurred in Georgia in 2020-2021*

N	Magnitude	Date	Time	Epicenter
1	4.7	16.08.2021	00:49 (UTC)	Georgia, Kvemo Kartli region, Dmanisi municipality
2	4.1	15.08.2021	22:36 (UTC)	Georgia, region of Samtskhe-Javakheti, Ninotsminda municipality
3	4	14.07.2021	06:35 (UTC)	Georgia, region of Samtskhe-Javakheti, Ninotsminda municipality
4	4.1	17.04.2021	20:07 (UTC)	Georgia, Colchis National Park
5	4.3	13.03.2021	10:00 (UTC)	Georgia, region of Racha-Lechkhumi and Kvemo-Svaneti, Onsky municipality
6	4.3	21.04.2020	05:23 (UTC)	Georgia ('velo Sak art)

Designate the set of actually occurred earthquakes with p_{real}. Designate the earthquake prediction models with: $Mod_{1,}Mod_{2,}$ ... etc. which provide some predictions through their predecessors (for example, for earthquakes - when it would occur, at which location and with which magnitude). We must choose only those models from these models, which satisfy the necessary condition, i.e. Intersection of the set of model predictions with the set of actual events should result in the set of actual events. We call this condition a necessary condition for choosing a prediction model [4]. This condition in case of earthquake means the following: If during the T time there were occurred, for example, six earthquakes (as in our example), only those models should be considered that predicted all these six earthquakes. Assume that such are the following models: $Mod_{1,}Mod_{2,}$ $Mod_{3,}Mod_{4,}Mod_5$. In our case it is not essential, what specifically is each model and based on which predecessors of the earthquake it makes the prediction.

In Table 2, the numbers of predictions for each of these models, the numbers of successful and failed predictions are given. Let's calculate the probability of success for each model.

It is obvious in this Table, that the sum of successful and failed predictions is equal to the total number of predictions. As for the probability of success [5, 6], it is calculated for each model and determines, how many times earthquake prediction was made and how many times an actual earthquake occurred. Assume that we consider the necessary predecessors and the models created for them: $A_1, A_2, ..., A_n$, where n is the number of considered predecessors. t is the time during which we make the analysis, and the number of actually occurred earthquakes is m. We calculated the number of earthquakes predicted by each predecessor: $p_1, p_2, ..., p_n$. For example, A_i the model, which was based on i predecessor, predicted earthquake occurrence p_i -times.

Table2: *The characteristics of "necessary models"*

Model	Number of predictions	Successful Number of predictions	Failed Number of predictions	Probability of success in %
Mod_1	100	6	94	6.00
Mod_2	95	6	89	6.32
Mod_3	99	6	93	6.06
Mod_4	98	6	92	6.12
Mod_5	99	6	93	6.06

For each p_i let's calculate quotients of the number of actually occurred earthquakes m, write it in % and designate with K_i:

$$K_i = \frac{m}{p_i}100\%$$

For example, if earthquake actually occurred 4 times, and we calculate the value $K_i = \frac{4}{20}100\% = 20\%$, then the probability of success for A_i will be 20%.

The probability of success also can be considered the probability of prediction correctness of specific model. Designate this last value with L_m and link the ratio $L_m = \frac{u}{v}$,*100% to its value, where u is a number of actually occurred events, and v is a predicted number of event occurrence obtained in the given p_m model.

Table 3: *The characteristics of the "necessary models" for the pairs*

Model	Number of predictions	Successful Number of predictions	Failed Number of predictions	Probability of success in %
$Mod_1 \cap Mod_2$	17	6	11	35.29
$Mod_1 \cap Mod_3$	20	6	14	30.00
$Mod_1 \cap Mod_4$	15	6	9	40.00
$Mod_1 \cap Mod_5$	13	6	7	46.15
$Mod_2 \cap Mod_3$	15	6	9	40.00
$Mod_2 \cap Mod_4$	10	6	4	60.00
$Mod_2 \cap Mod_5$	16	6	10	37.50
$Mod_3 \cap Mod_4$	17	6	11	35.29
$\boldsymbol{Mod_3 \cap Mod_5}$	**8**	**6**	**2**	**75.00**
$Mod_4 \cap Mod_5$	18	6	12	33.33

The theorem proved in [7]: From the given predictions, always can be chosen at least two such predictions, for which the probability of correctness of simultaneous occurrence is greater or equal than the probability of correctness of the best prediction model:

$\min(P_{ij}) \leq P_g$, where $i, j = 1 \dots n$.

In this theorem, P_{ij} designates the probability of correctness of simultaneous occurrence of two prediction models $P_{ij=}P_i \cap P_j$., and P_g designates the set containing the least number of predictions, which at the same time will be the best prediction model. $P_g \leq P_i$, where $i = 1 \cdots n$.

According to this theorem, we should consider pairs of models. Let's compose Table 3 with the values corresponding to Table 2, for each possible pair of all five models, considered in the example, whose total number will be 10.

After analysis of Table 3 we see that the best result is obtained from the combination of two models Mod_3 and Mod_5 (although the separate probabilities of success for them are not best, the combined probability of success is increased up to 75%, even though separately these models have significantly lower values of success: 6.06% and 6.06%. For the considered examples, it is possible that two pairs of the models show the same result. If this is the case, it should be decided by means of expert and material and technical resources needed for work of these models, which one should be used. The diagram corresponding to Table 3 see on Diagram 1:

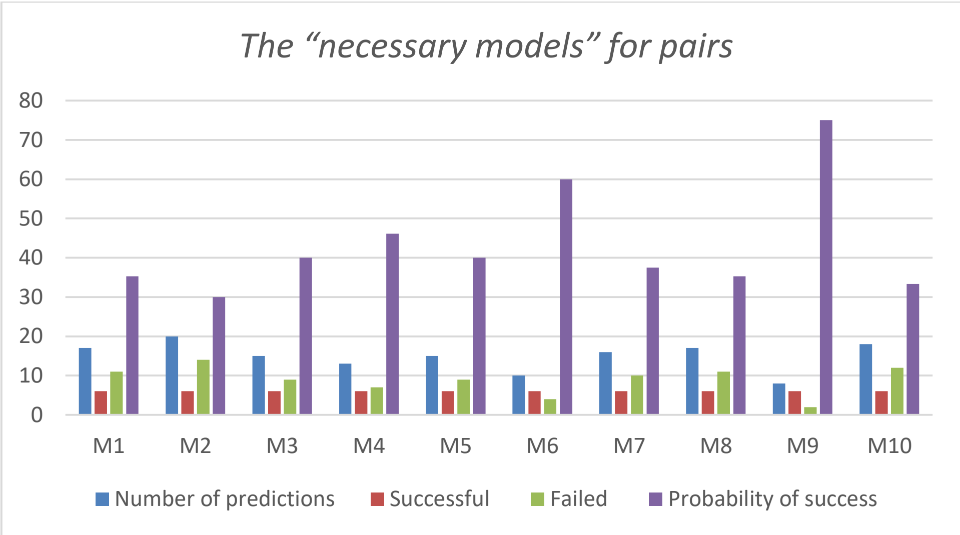

Diagram 1. *The characteristics of the "necessary models" for pairs*

Keys used in Diagram 1:
M1=$Mod_1 \cap Mod_2$; M2= $Mod_1 \cap Mod_3$;
M3= $Mod_1 \cap Mod_4$; M4=$Mod_1 \cap Mod_5$;
M5= $Mod_2 \cap Mod_3$; M6= $Mod_2 \cap Mod_4$;
M7= $Mod_2 \cap Mod_5$; M8= $Mod_3 \cap Mod_4$;
M9= $Mod_3 \cap Mod_5$; M10= $Mod_4 \cap Mod_5$.

The next stages of the prediction algorithm based on parallel data [papers] is consideration of model triples. See Table 4.

Table 4: *The characteristics of the "necessary models" for triples*

Model	Number of predictions	Successful Number of predictions	Failed Number of predictions	Probability of success in %
$Mod_1 \cap Mod_2 \cap Mod_3$	10	6	4	60.00
$Mod_1 \cap Mod_2 \cap Mod_4$	9	6	3	66.67
$Mod_1 \cap Mod_2 \cap Mod_5$	11	6	5	54.55
$\mathbf{Mod_1 \cap Mod_3 \cap Mod_4}$	**7**	**6**	**1**	**85.71**
$Mod_1 \cap Mod_3 \cap Mod_5$	9	6	3	66.67
$Mod_1 \cap Mod_4 \cap Mod_5$	8	6	2	75.00

After analysis of Table 4 we see that the best result is obtained from the combination of three models Mod_1, Mod_3, and Mod_5. The combined probability of success for them is increased up to 85.71%.

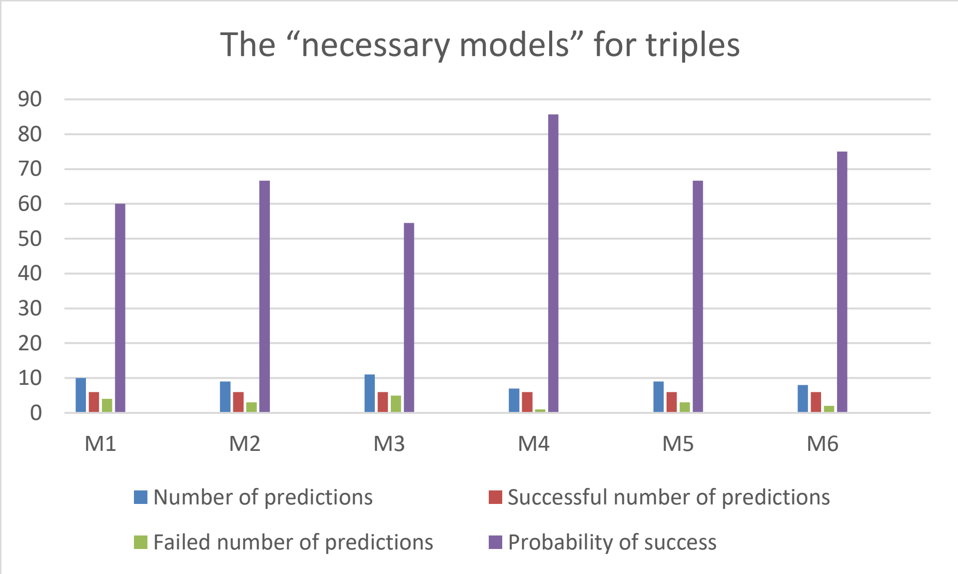

Diagram 2. *The characteristics of the “necessary models” for model triples*

We introduced the following keys for Diagram 2:

M1=$Mod_1 \cap Mod_2 \cap Mod_3$; M2=$Mod_1 \cap Mod_2 \cap Mod_4$;
M3=$Mod_1 \cap Mod_2 \cap Mod_5$; M4=$Mod_1 \cap Mod_3 \cap Mod_4$;
M5=$Mod_1 \cap Mod_3 \cap Mod_5$; M6= $Mod_1 \cap Mod_4 \cap Mod_5$;

III. Conclusion

It is obvious that the more is the number of intersections of prediction models, from which we choose the best, the better would be the result, compared to the case of less number of intersections. But we should take into account that greater number of models need greater number of data (predecessors), which can be obtained by spending considerable amount of material resources. Collection and analysis of large amount of data is an unresolved issue for small, low income states. Exactly for these cases it is important to theoretically choose two or three models of prediction, for which intersection of predictions would give best results. While collection of information, in this case, would be needed only for these chosen models, thus sharply reducing the costs of information procession.

References

[1] Gasitashvili, Z., Phkhovelishvili, M., Archvadze, N., Jorjiashvili N. An Algorithm of Improved Prediction from Existing Risk Predictions. / *Published by AIJR Publisher in “Abstracts of The Second Eurasian RISK-2020 Conference and Symposium” April 12- 19, 2020,* Tbilisi, Georgia. DOI: 10.21467/abstracts.93 pp. 31. 2020.

[2] Aliyev, V., Magerramova, S., Balaeva, A., Azeryar, L. Risk Assessment and Analysis of Accidents of Water Facilities, AIJR Abstracts, pp. 9-10, 2020.

[3] Latest Earthquakes. https://earthquaketrack.ru

[4] Gasitashvili, Z., Phkhovelishvili, M., Archvadze, N. New Algorithms for Improvement of Prediction Models Using Data Parallelism. / 13th International Conference on Computer Science and Information Technologies CSIT 2021. Proceedings. Armenia, Yerevan, September 27 - October 1, 2021. Pp.17-20.

[5] Gasitashvili, Z., Phkhovelishvili, M., Archvadze, N. New algorithm for building effective model from prediction models using parallel data. / Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk: UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8. Pp. 25-28.

[6] Chogovadze, G., Surguladze, G., Topuria, N., Archvadze, N. Implementation of a prediction model with cloud services. // *Bulletin of the Georgian National Academy of Sciences, 2020, 14(3),* pp. 29–35.

[7] Gasitashvili, Z., Phkhovelishvili, M., Archvadze, N. Prediction of events means of data parallelism. Proceedings - Mathematics and Computers in Science and Engineering, MACISE 2019this link is disabled, 2019, pp. 32–35, 8944725.