Usage on Different Types of Data to Solve Complex Mathematical Problems

ZURAB GASITASHVILI
Deputy Rector for Science
of Georgian Technical University, Professor
0175, Tbilisi, Kostava 77, GEORGIA,
zur_gas@gtu.ge http://gtu.ge/

MERAB PKHOVELISHVILI
Niko Muskhelishvili Computing Mathematic Institute
of Georgian Technical
0131, Tbilisi, Grigol Peradze str. #4, GEORGIA
merab5@list.ru http://micm.gtu.ge/

NATELA ARCHVADZE
Department of Computer Sciences
Ivane Javakhishvili Tbilisi State University
0179, Tbilisi, I.ChavChavsdze str.#3, GEORGIA
natela.archvadze@tsu.ge https://tsu.ge/

Abstract: - A new approach to solve some complicated problems using supercomputers is presented which is based on the notation of parallel data and means of definition of new relationship between them that is called parallelism between data. Parallel data are different data related to one event existing in different time (time parallel) and/or additional feature parallel. In practice parallel data can be used for the prediction of earthquakes or any other hazards, economics (business, macroeconomics), prediction of political events (elections, distribution of political forces), for effective solutions of some prediction problems in medicine or other fields.

Key-Words: supercomputers, prediction, parallel data, expandable matrices

1 Introduction

Creation of supercomputers, widespread adoption of cluster systems, connection of computers with each other by means of local and global networks, caused attraction of users to computation process and made possible to perform various tasks. This includes a task of weather forecast, simulation of nuclear test, etc. It can be said that supercomputers bring new possibilities for automation of prediction processes.

In its turn, working with supercomputers requires knowledge of characteristics of parallel computing, effective setting of tasks for multiprocessing systems, familiarization with various instrumental means, which facilitate this process. Although, it is also necessary to conduct advanced theoretical works, in relation to algorithms and software with parallel structure [1,2].

In the result of development of parallel computing, the paradigm of parallel programming was formed [3,4,5]. Today there is two approaches to the paralleling of computing process [6]. They are data parallelism and message passing. Both approaches are based on distribution of computing on processors of parallel computer, available to the users. At the same time, it is necessary to solve various problems. But no one of them is associated with the specifics of computing on supercomputers.

In this article we consider, at the one hand, the models of new type of predictable processes (called conditional, temporal, expandable matrix of vectors) and, on the other hand, an new fundamental style of programming (called fundamental style of parallel data) for processing of these models .

2 Forecasting Problems

Our main goal is to develop algorithms to solve some forecasting problems of predictable processes. We consider three type forecasting problems: earthquake prediction problem, prediction in economy and problems of diagnostics in medicine.

2.1 The Earthquake Prediction Task

Our goal is to develop algorithms, to solve the task of forecast of predictable processes, in particular, an earthquake. A relevance of this task is defined by the fact that earthquakes are natural disasters with most number of victims, losses and destructive results of impact on human environment.

Under the earthquake prediction we mean determination of the location, time and magnitude of earthquake. A prediction is divided into long-time (throughout decades), middle-term (years in advance), short-term (days, months in advance) and operative (minutes, hours in advance) earthquakes. This classification is quite conditional. Each stage of prognosis is based on a certain complex precursor – geophysical phenomenons (mainly), which occur prior to the earthquake.

At the same time, the geophysical precursors are divided into seismic, hydro, geodynamic, deformation, geochemical, thermal, gravitational, electromagnetic precursors and those obtained by means of remote monitoring with use of satellite technologies, developed in recent years [8].

Despite the fact that there are many precursors. Neither of them gives a correct prediction about the time, location and magnitude of the future earthquake. The possibility of successful prediction by means of each precursor does not exceed 0.5% [9]. One way to solve this problem is to use simultaneously several precursor for prediction, but there is not a sufficient information in this regard. The practice of recent years shows that their integrated use would improve the reliability and efficiency of prediction, at least in the middle-term prognosis [10].

There are up to five thousand indicators and precursors of forthcoming earthquake and neither of them guarantees a high level of correct prediction. We call geophysical phenomenon an unusual, special event, which in general is a precursor of earthquake.

2.2 Forecasting Models in Economy

Business-forecasting tasks may include: demand, intermittent demand, time and space hierarchies, shares, macroeconomic indicators, commodity groups, new products and more. Consider two tasks: predicting demand and interruption request. That's the solution of these two tasks will be discussed in the example when "parallel data".

In this article we consider, at the one hand, the models of new type of predictable processes (called conditional, temporal, expandable matrix of vectors) and, on the other hand, an new paradigm of programming (called parading of parallel data) for processing of these models.

For the computer presentation of mathematical models of prediction processes, we will use so-called conditional temporal expandable matrices of vectors (hereinafter – the expandable matrix). For the computer presentation of mathematical models of prediction processes, we will use so-called conditional temporal expandable matrices of vectors (hereinafter – the expandable matrix).

Each event impacting an event to be predicted, is represented in the form of separate vector of data. The word "conditional" in the name of matrix means that it is now known in advance, how many events impact an event to be predicted (i.e. how many vectors of data are contained in the matrix). The matrix is dynamical, therefore, some events (and respective vectors) are erased from it, some are added and some are moved into another place. The word "temporal" is used in the name of matrix because the number of row depends on the time. These vectors are shown on the figure by directions of arrows, which are directed from the bottom upwards. In each vector, the data is arranged according to the time.

2.3 Problems of Medical Diagnostics

The objects are patients. The features characterize the results of the examination, the symptoms of the disease and the methods of treatment used. Examples of binary features: gender, presence of headache, weakness. An ordinal symptom is the severity of the condition (satisfactory, moderate, severe, extremely severe). Quantitative features - age, pulse, blood pressure, hemoglobin in the blood, the dose of the drug. The symptom description of the patient is, in fact, a formalized history of the disease. Having accumulated a sufficient number of cases in electronic form, different problems can be solved:

- classify the type of disease (differential diagnosis);
- determine the most appropriate method of treatment;
- predict the duration and outcome of the disease;
- assess the risk of complications;

- find the syndromes - the most characteristic for the disease set of symptoms.

The value of such systems is that they are able to instantly analyze and summarize a huge number of precedents - an option that is not available to the specialist doctor.

3 Models of predictable processes

During solving a task of prediction, at first, it is necessary to create a mathematical model of presentation [11, 12]. A mathematical model is a mathematical recording of some regularity (Fig. 1):

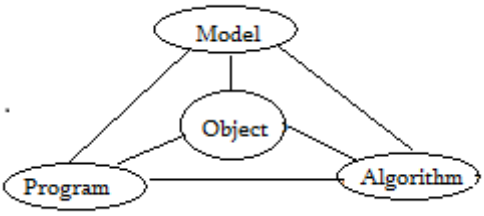

Fig 1. Computer simulation by Samarski

a) A hypothesis is expressed and it is verified an experiment.

b) Experiments are conducted, data is collected and then a model is formed. For computer modeling, more parameters are used.

3.1 Definition of conditional temporal expandable matrix of vectors

For the computer presentation of mathematical models of prediction processes, we will use so-called conditional temporal expandable matrices of vectors (hereinafter – the expandable matrix) (Fig. 2).

Each event impacting an event to be predicted, is represented in the form of separate vector of data. The word "conditional" in the name of matrix means that it is now known in advance, how many events impact an event to be predicted (i.e. how many vectors of data are contained in the matrix). The matrix is dynamical, therefore, some events (and respective vectors) are erased from it, some are added and some are moved into another place. The word "temporal" is used in the name of matrix because the number of row depends on the time. These vectors are shown on the figure by directions of arrows, which are directed from the bottom upwards. In each vector, the data is arranged according to the time. In the lower part of the model, there are values, corresponding to already occurred events. On each time section (depending on the task, it can be 1 minute, 1 hour or other unit), the new row is added.

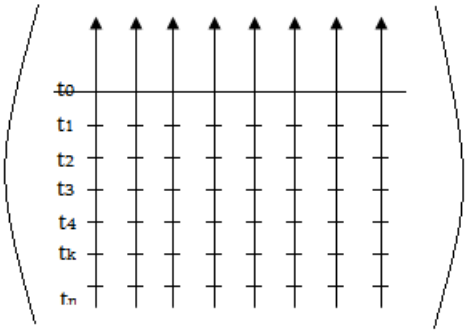

Fig 2. Expandable matrix

The matrix, corresponding to the model of predictable processes, as it was noted, is dynamically variable, which means a change of its sizes. Thus, the word "expandable" in its name. The matrix constantly expanses upwards, new data is added to it and the number of columns varies, the data is added or taken away according to the event function, which will be considered below. It is also possible that a matrix is expanded downwards – the data is added, corresponding to old, already occurred events or archive data.

Underlining on the Fig. 2 designates the time t0, when the predictable process occurred (when the event occurred/will occur).

Each expandable matrix may describe the one territorial region, therefore, for the certain task of prediction, more than one expandable matrix can exist, which will be built for certain period of time.

3.2 Parallel data

We call the data of various types influencing the predicted event (data located on various vectors) the parallel data. If there are several pieces of such data and their aggregate forms the group, they are called a set (group) of parallel data.

Data in time is represented by each column of matrix, which characterizes the given forecast measurements. Data can be of various types. For example, if we consider a task of earthquake prediction, it is influenced by several important

characteristics of a seismic process, such as a seismic cycle, pre and post earthquake activity, qualitative value of attenuation of seismic events after the strong earthquakes, which is known as the Omori's law for aftershocks. At the same time, strong earthquakes are preceded with abnormal seismic activity or abnormal decrease of seismic activity, or its oscillation in various space-time zones. According to authors of [13], the main source of information about earthquakes in the near future can become an electromagnetic radiation of Earth VLF / LF. In particular, a monitoring of electromagnetic radiation of Earth in the period of earthquake preparation might prove to be very useful with the purpose of prediction of strong earthquakes (M5).

As it was noted, geophysicists name up to 5 000 signs of forthcoming earthquake, its precursors, but neither of them guarantees a high accuracy of prediction. Although, it is necessary tom compute possibilities of occurrence for each of them. The difficulty is also that an interconnection between these precursors should be determined and corresponding conditional possibilities computed. In addition to this, there are data, obtained by means of observation, from which it is difficult for specialists (geophysicists) to determine a regularity. Exactly in this we think it is necessary to use the methods of artificial intelligence, to formulate the hypothesis from the data, by means of "Mechanizing hypothesis formation" [14,15].

3.3 Event function

Location of columns in the matrix has its significance. That column is located to the left, data of which is more "important", i.e. the set of this data makes the process of prediction more reliable.

On the Fig. 1, each vector actually represents one of the function, on which the event depends. We call this an event function. Marked values are those data, which satisfy the event function, which can be a function-predicate (with true or false values), or unclear values. It is possible that the values of event function would be statistical data (Fig. 3):

Those data are marked on the figure, for which the event function takes the value, on which the occurrence of event is determined with the given probability.

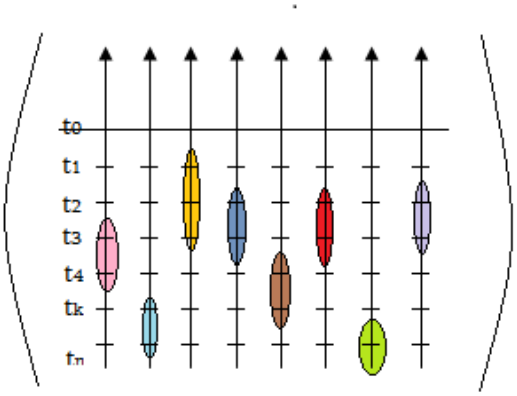

Fig. 3. The expandable matrix with event function

For example, during the prediction of earthquake, such regularity is fixed: F1(x1, x2,..., xk)=(A, B, C), where A designates "the location", B - "when" and C - magnitude.

Let's take that xk=3 * x1, and after some time, the data y is obtained, which returned to the value of xk, i.e. y=1/3 xk (Fig. 4):

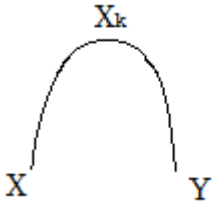

Fig. 4. A case when It is possible forecast (example).

This, the equation Fi(xi ,...,xj) = (A1, B1, C1) is obtained, which shows that after B1 days, at the location of A1, it is possible of occurence of the earthquake with magnitude C1. The column, corresponding to this function, should take the respective place in the expandable matrix. If the time of occurence of precursor events in the already existing matrices, including other matrices, is more early, than it is moved into the respective place. An arrangement of vectors in the matrix is done in such a way that those precursor vectors are on the first place, which give a prediction earlier, than others, then events, occured earlier than others, but later than first vectors, and so on.

At the same time, the system cheks the data of other vectors and seeks, whether this is the situation, when the other event function takes the values, necessary for prediction. If so, then the probability of simultaneous occurence of these two events included here.

Therefore, each event function has its characteristic, which defines, what is its share in the

whole process of prediction (is given in percentage or probability). I.e. if the set can be find in the data of one column, which satisfies the event function, then the event will occur with the given reliability.

4 Fundamental Style of Parallel Data

We process the expandable matrix by using a new fundamental style, so called style of parallel data and this is important, because it is possible to warn users of various levels. The first level consists of limited number of senior managers. On the second level, there economic employees, who can plan an evacuation of the population and other activities, to avoid the loss of human life. The third level is formed by the mobilization of respective rescue workers, and the next level – warning of population by means of mass communication and their evacuation to the safe places.

We should especially highlight those matrices, which are designated for prediction of earthquake on those territorial zones, where there are the objects dangerous for the population: nuclear power plants, large scale hydro power plants and reservoirs, chemical plants, etc. Of course, an evacuation of population from this places should begin earlier.

4.1 Regularity function

There is the task, to process the event function, in which we mean, what vector corresponding to the event function should be moved to the left, or to the right and which one should be erased. At the same time, the time and other characteristics of prediction should be determined (for example, the magnitude).

It it necessary to check the work of the event function and, therefore, the regularity function to the already occurred events. Although, it is necessary to consider the reliability of recording these data (whether there were errors in recording).

As it was noted, an arrangement of vectors in the matrix has its significance, so that to have the time to conducting measures, warn the population, evacuate them and so on. Most left in the matrix should be earliest in terms of time.

Rearrangement of matrix:

a) According to time (preparation of population and rescue workers)

b) According to probabilities

c) According to impacting each other.

It also should be considered that the data have an impact on each other. Their parallelism means that one of them impacts other, that other impacts yet another and so on. For example, if the warming of sea gives the picture that magnetic radiation is increased, then in the specific place and in the specific time, the evaporation can be increased, and the level of water in the well can decrease. The decrease of water level can cause some process, which increase the density of underground waters and, proceeding from this, the displacement of plates with collision. Therefore, it is necessary to detect interactions in the existing earthquakes, compute possibilities and create the respective functions.

Most interesting thing is that this process occurs in parallel with each other, in terms of time. For example, an ionization can cause the decrease of levels, electromagnetic radiation, strong wind and other events, and it is necessary to detect these connections, in what time, what caused and represent them graphically.

d) Detect more regularities

It is often known that the specific event impacts some other event (we mean an earthquake), but how it occurs, is not known. Therefore, it is necessary to separate a tendency of changes from existing data for the given region, by means of which the respective function should be built. There is m such data: electromagnetic radiation, ionization, etc. The data should be measured and regularities detected.

4.2 Peculiarities of business forecasting and medicine prediction

Both processes have common characteristics that the results of the prediction are obtained not as exact values, but as probabilities. To be more precise:

A new approach for business forecasting is discussed in the article. This means usage of parallel data paradigm of programming. Parallel data are different kind of former data, which give us chance to predict an event in dynamic mode. Also, functioning of forecasting process online is being discussed. This method helps us to use super computers not only for original purpose - calculation big amounts, but for processing parallel data online.

Supercomputers are computers with a high level of performance, which are used to work with those applications, which require more intense computations. Creation of supercomputers, widespread adoption of cluster systems, connection of computers with each other by means of local and

global networks, caused attraction of users to computation process and made possible to perform various tasks. This includes a tasks of economic forecast, simulation of nuclear test, etc. It can be said that supercomputers bring new possibilities for automation of prediction process.

The matrix, corresponding to the model of predictable processes, as it was noted, is dynamically variable, which means a change of its sizes. Thus, the word “expandable” in its name. The matrix constantly expanses upwards, new data is added to it and the number of columns varies, the data is added or taken away according to the event function, which will be considered below. It is also possible that a matrix is expanded downwards – the data is added, corresponding to old, already occurred events or archive data.

Each expandable matrix for each business forecasting task may describe the one territorial region, therefore, for the certain task of prediction, more than one expandable matrix can exist, which will be built for certain period of time.

For diagnostic problems in medicine parallel data can be used, in this case observations on patient (patients' groups) can be done in time (minute, hourly, diary). The object of observation can be the vital data of the human: pressure, wrist, temperature, etc. In this case the data can be considered as predecessors which give predictions about human health state at a certain time point. As more predecessors are used to predict some people, it is more likely to take the treatment to avoid the forecasting data.

As in economics, we have a dynamic prediction for the forecasting problems in medicine. Here we have no exact, but the percentage compliance of the disease. For example, if the patient has a pressure as 160-90, when the normal one is 130-70, then the probability of stroke is defined by 5, for example, 15%.

5 Conclusion

The method proposed by us can be used not only for prediction of earthquakes, but also for other events, which are hard to predict and which use the set of precursors of various types (often with low probabilities). Using a parallel data in these methods allows us not only to perform prediction with more probability,but also divide it into the sequental stages, which makes it controlable to monitor the expected event and take appropriate measures. Also, by using the method of parallel data, it is possible to detect new regularities, which are not yet detected, because of limitation of sequential algorithms.

References:

[1] V.V.Veoyevodin, Vl.V.Voyevodin "Parallel'nyye vychisleniya" - SPb. : BKHV-Peterburg, 2002. -608 s. ISBN 5-94157-160-7. (in Russian)

[2] Voyevodin Vl. V.Chislennyye metody, parallel'nyye vychisleniya i informatsionnyye tekhnologii. Izdatel'stvo Moskovskogo Universiteta. 2008. (in Russian)

[3] Paradigmy parallel'nogo programmirovaniya - Blog programmista. https: //pro-prof.com/forums/topic/parallel-programming-paradigms. (in Russian)

[4] Obshchiye paradigmy parallel'nogo programmirovaniya - OpenNET. https://www.opennet.ru/docs/RUS/linux_parallel/node221.html. (in Russian)

[5] N.V. Shilov, L.V. Gorodskaya, Ye.V.Bodin. Paradigma parallel'nogo programmirovaniya: uchit' ili ne uchit'. http://agora.guru.ru/abrau2011/pdf/193.pdf (in Russian)

[6] Dve paradigmy parallel'nogo programmirovaniya. http://umk.portal.kemsu.ru/mps/html/node4_25.htm . (in Russian)

[7] Gul'yel'mi A. V. Forshoki i aftershoki sil'nykh zemletryaseniy v svete teorii katastrof. Uspekhi fiz. nauk. 2015. T. 185. № 4. S. 415-429. (in Russian)

[8] Zav'yalov Aleksey Dmitriyevich. Prognoz zemletryaseniy: chto novogo? https://www.nkj.ru/interview/11124/ (in Russian)

[9] Shuman V.N. Seysmicheskiy protsess i sovremennyye monitoringovyye sistemy. Geofiz. zhurn. 2014.T. 36. № 4. S. 50-64. (in Russian)

[10] V. N. Shuman.Seysmoelektromagnetizm i prostranstvenno-vremennyye struktury. Geofizicheskiy zhurnal № 6, T. 37, 2015 g. c. 24-41. (in Russian)

[11] *Samarsky*, A.A., Nikolaev, E.S.: Methods of Solving Grid Equations, p. 588. Science, Moscow (1978). (in Russian)

[12] A.A., Mikhaylov A.P. Matematicheskoye modelirovaniye: Idei. Metody. Primery. - M .: Fizmatlit, 2001. - 320 s. (in Russian)

[13] Kachakhidze M.K., Kachakhidze N.K. Prognoziruyemyye vozmozhnosti vydeleniya VLF / LF v kachestve osnovnogo predshestvennika zemletryaseniya. GESJ: fizika 2015 | № 1 (13) ISSN 1512-1461. pp.80-89.

[14] Mayyer R.V. Komp'yuternoye modelirovaniye: modelirovaniye kak metod nauchnogo poznaniya. Komp'yuternyye modeli i ikh vidy. http://econf.rae.ru/article/6722. (in Russian)

[15] Il'in A.V., Il'in V.D. Formalizatsiya znaniy o zadachakh: zadachnyye konstruktivnyye ob"yekty. Informatsionnyye tekhnologii v upravlenii. http://ubs.mtas.ru/bitrix/components/bitrix/forum.interface/show_file.php?fid=4933