Using Different Types of Data Operations for Solving Complex Mathematical Tasks

Merab Phkhovelishvili

Niko Muskhelishvili Institute of Computational Mathematics of Georgian Technical University Tbilisi, Georgia,

e-mail: merab5@list.ru

Nato Jorjiashvili

Ilia State University Tbilisi, Georgia,

e-mail: nato_jorjiashvili@iliauni.edu.ge

Natela Archvadze

Department of Computer Sciences Ivane Javakhishvili Tbilisi State University Tbilisi, Georgia,

e-mail: natela.archvadze@tsu.ge

ABSTRACT

In the paper a new approach of solving complicated mathematical problems is presented, which is based on the notion of parallel data and means of definition of the new type relationship between data that is called parallelism between data. The parallel data are different data related to one event or existing different data influence (or are related to them) in specific time period (parallel by time) or event occurring in specific location (parallel by location) and/or parallel by additional feature. 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.

For storing and processing data, parallel data are elements of new type of matrix that is called a heterogenic matrix. In such a matrix columns can be numbers type, text type, images (geometric figures), sound records or video file types. In the work operations of heterogenic matrices are defined, the necessity of these operations is justified for forecasting problems. Operations are defined for the same type data of heterogenic matrix as well as for different type elements.

Keywords

Forecasting problem, algorithms, parallel data, heterogenic matrix.

1. INTRODUCTION

Many fields contain prediction type problems, such as: business, macroeconomics, weather, elections, results of sport competitions, finance, crimes, etc. Also there are many forecasting methods: expert judgment, analogy method, scenarios building, aim tree method, network planning and management, experts and logical analysis, direct interpolation and extrapolation, regression analysis, time series, etc. These methods can be attributed to the methods of heuristic, analytical and factographic groups. Apart from these methods, there are dozens of complex methods that are successfully used to solve larger problems.

The following basic principles are common for forecasting problems from different fields [1]:

• Sometimes difficult and statistic-based models do not seem to be more precise than the simplest models;

• Combination of models or computer predictions obtained by different models meanwhile increase the accuracy of forecasting;

• Accuracy of prediction will decrease according to the increase in the forecasting horizon.

We developed a new theory (algorithm) for solving forecasting problems, which implies to use new, parallel data strategies [2,3,4] instead of parallel computation. In this work, we will verify the parallel data definition, then we will talk about the construction of heterogeneous matrix from such data and explain the operations on such matrices.

Where do the signals come from?

Electromagnetic radiation, t radiation.

In this paper, a new algorithm is presented for solving prediction problems. According to this approach instead of common parallel computing strategy, a new parallel data strategy can be used. We will define parallel data, then introduce a dynamic prediction problem, then we will talk about definitions of parallel in time, different in space and the accuracy notation of prediction. The matrix, the elements of which are signals expressed by different type data.

2. PARALLEL DATA

Definition of parallel data: parallel data (set of data) are different type data, which give the same prediction of some events.

In case of earthquakes, parallel data (precursors) predict defined three parameters: time (T), location (D) and earthquake energy (E).

F(A) = G(B) = (T D E),

where F and G are the recognizable functions of occurrence of event by precursors, when A and B are precursor data. Often E can be different when time (T) and location (D) for both precursors are similar (at some time and space intervals). Matching time of events (T) can be expressed in different units. It can be minutes (for the very short time prediction), hours (medium short time prediction), days (medium time prediction), etc. Similarly, location matching can be expressed for some area in meters (exact prediction), kilometers (medium prediction) or hundreds of kilometers (for long time prediction).

If there is no complete match (matches by T or D or E), such data are called parallel in time (T), parallel in location (D) or parallel in energy (E). In case of earthquakes, matching by T and D are important, and E can be higher than some value (5 or higher by Richter scale).

In economics there is an approximate prediction in time or the amount of result. In this case prediction precursor data A and B give predictive quantity Pmt by m data in some given time with certain precision. Vm is the accuracy (in %) of the predictive quantity.

F(A)=Pat G(B)=Pbt
P = Pat * Va P= Pbt * Vb

P is the real value of the predictive quantity. Such A and B data are called parallel in time.

Under the strategy of parallel data we consider construction of heterogenic matrices for given processes and their processing. The main idea is as follows:

We present all events related to predictive event as features, functions, the values of which can be sorted as individual vectors according to time. The function that describes the event, we call the event function. Thus, each event function (it can be software) defines data-vectors. We call them event vectors.

We build a matrix with event vectors (we consider such vectors as columns of matrix). It is previously unknow how many events have an influence on predictive event (i.e., how many columns are stored in the matrix). Every new event adds columns in the matrix. Thus, the matrix is dynamic. In some time period some columns will be deleted from the matrix, some columns will be added and some of them will be moved.

In fact, the solution to the prediction problem starts for the problem by obtaining of heterogenic matrices and working with them.

3. EQUATIONS

We call heterogeneous matrix (the same as non-homogeneous) (next Het-matrix) a dynamic expandable matrix, columns of which are event vectors and at the same time they are heterogeneous (its elements are of different type). However, in each column elements of the same type are stored. Let's specify each term.

We call the dynamic expandable matrix [2] a matrix with a variable dimension, the number of elements of which depends on time. So, if the X matrix has n rows and m columns for ti time period, then for the next ti+1 time period it already has n` and m` dimensions. Of course, these values can be increased or decreased or remain unchanged.

Generally, in the real problems, the new data are added in the matrix in the form of rows or columns. For this reason the term "expandable matrix" appeared. A row will be added when new values be added to one or several events. In the matrix a column will be added in case if a new event is added to the prediction process (function-software describing new event, which is needed for the prediction process). Here is should be mentioned that the matrix does not constantly expand: the matrix processing contains the process of removing some of the columns or rows of the matrix. The column is removed when no specific event is needed and the row is removed when the data is out of date and the processing is unnecessary.

As for the moment of time, it is determined according to the addition of new data in the matrix and according to the objectives it may be 1 minute, 1 hour or other.

3.1. Types of the Het-Matrix Columns

Columns of Het-matrix can be of different type. However, elements of each columns are of the same type as follows:

I. Simple type;

Vector type;

II. Het-matrix type that means that Het-matrix can be defined recursively.

Let’s review each of them.

Het-matrix columns belong to the simple type if they have the following types in their standard formats:

a) Numerical type;
b) Text type;
c) Image (geometrical figures);
d) Sound records;
e) Video file types.

Data type or simply type is a data classification that indicates the compiler or interpreter how the programmer is going to use them [5].

Text data is string type information (sequence of typing symbols). The open text in calculations - is a wide term for data (for example file contents), which represents only the signs of the material, but not their graphic representation and other objects (numbers with floating comma, images, etc.)[5].

A large number of graphics, audio and video formats are used and any of them are allowed in Het-Matrix. The description of these types is given in [6].

The Het-matrix columns are vector type, if its values are the same type or array, or list, or vector, or any other container. The selection of the container depends on the customer who chooses the container according to the purpose.

Also, the Het-matrix can be Het-matrix columns that indicate to the recursive nature of Het-matrix.

Let's agree that at the initial moment the matrix construction data are stored in the zero line. Data measured at the next moment of time will be stored in the first line, which will be located above the zero line, at the next moment - in the second line, etc. The matrix will be filled from above with new data. Let's agree that the first column of the matrix is a period of time.

Every new data group in the matrix is added from above. Consequently, the matrix lines are numbered from the bottom to the upward. If an information about the event history is added in the matrix (occurred before the observation) and/or archive data, then they will be stored on the rows with negative indices. So, in the Het-matrix rows are sorted by time.

As already mentioned, data of each column of the Het-matrix should be of the same type. This requirement is natural because as we have already noted, during the real prediction problems values are introduced into the event functions, of which vector columns are formed so the column elements have the given event function type. The second column elements - the type of the second event function, etc. To calculate the values of each event function and filling with these values of Het-matrix can be done for different time periods. Because of this, the matrix may be a sparse. Let's agree that the sparse element of the matrix will be determined by the value of the previous moment of time. This means that if there is no calculation of the event function at a certain moment, then we mean that the previous value has not changed.

Consider the simplest example when there are only three events of numerical type, string type and char type:

3	64	Asd	(5 7 9 1 12 ABA)
2	234	Rasd	(1 5 34 d DFG)
1	7	Gh	(A SD T 7 90)
0	15	Cba	(A X V Y M 890)

Fig.1. The simple example of the Het-matrix.

Along with Het-matrix we will always consider the matrix (rows) with the dimension 1 x N (N) (by N numbers of Het-matrix columns are expressed), which is composed by the event functions, which define the Het-matrix:

(Fun1 Fun2 Fun3 - - - FunN)

Fig.2. Row of the event functions

At the first, identification of the event functions should be done for the prediction problem, and the next step is to fill the Het-matrix with their values for specific time periods.

As we have already mentioned, the data of one line of Het-matrix for the specific moment of time, which are actually the values of different events influence on one event, we call parallel data. Parallel data are data of different types, and the results of prediction are obtained by observing them.

4. OPERATIONS ON HETEROGENIC MATRICES

It should be noted that ordinary matrices and operations on them are defined in some of the widely distributed computer systems. For example, in Microsoft Excel numerical type matrices and the following operations are defined: multiplication of matrices and finding the inverse matrix.

Matlab system is focused on processing of variable arrays. Hence, the array (vector, matrix) is considered as the main object of the system. However, it should be noted that they only have a numerical type. The system Matlab performs operations in linear algebra and matrix theory using the library of numerous built-in functions. All simple operations are included here: addition, subtraction, multiplication of vectors or/and matrices, solving linear algebraic equation systems, calculation of Eigen vectors, etc.

Computer systems are permitted to build matrices only from numbers and are not allowed to build other type matrices such as sound or graphical type data. There are no operations defined for various types of data.

Since the Het-matrix contains different types of elements, let's define two types of operations: a) Operations on the same type elements (data) of the matrix; and b) operations on different types of matrix elements (data). Operations involve the following operations: addition, subtraction, multiplication, division, and comparison operations: equal, not equal, more than or equal, less, less or equal.

4.1. Operations on the Same Type Elements of the Het-Matrix

Suppose we have two Het-matrices, which have the same type of data. Operations will be determined in accordance with the types and, as far as possible, we will assert the need to determine this operation for the purpose of forecasting.

4.1.1. Addition Operation (+)

- Addition of numbers are determined according to the definitions in mathematics.
- Addition of texts is defined as the action of "union", the same "merge" or the concatenation. For example: "new" + "string" = "newstring".
- The necessary condition for addition of geometric figures is that these figures should be presented in their coordinate systems. Addition of two figures is defined as follows: both figures should be placed in the same coordinate system and then their union. These figures may have intersection, and in this case the result will be two independent figures (Fig.3):

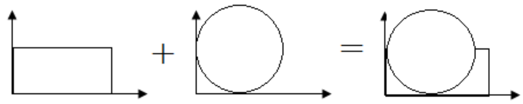

Fig.3. Addition of two figures.

The feasibility of the definition of this operation is derived from the earthquake prediction problem that also includes the location definition where an earthquake may occur. The other event function defines the other location, and for the united forecasting results it is necessary to add them.

- An addition of two sound recordings is defined as the sound recordings are placed on each other. For instance, if one record is only music, and the other recording-singer voice, the recording of two recordings is a song recording where the singer sings with music in the background.

While forecasting an earthquake, two sound recordings are done at some depth. Such sounds may not be detected by the human ear. When different voices are recorded, it is necessary to find out what they have in common (by the tone, wave). If geophysicists define a precursor that a certain level of noise leads to an earthquake, it may be possible to make a prediction by the sound level.

Operations of sound type can also be defined in another way: by addition of two sound elements an element can be obtained in which the second sound starts after the first ends; by subtraction – the second will be deleted from the first one, etc.

- Addition of two video files can be defined as video clips are placed on each other. Placement means to rewrite the second video clip on the first one. For example, in the movie, one actor plays the roles of two twin brothers. At first one brother will be filmed, then the other one, and by placement such movies to each other both brothers can be seen at the same time.

The need to work on the video information is well illustrated in the earthquake prediction problems. In this case filming takes place where there is a need to analyze the changes of images. For example, the length, width or depth of the rupture have changed. Another example is the recording of radiation in ionosphere. If the rising rate is observed, the expected earthquake area can be determined. At this time it is necessary to determine the movement rate and localization of this area.

- Addition of two vector type objects can be defined as a combination of both elements.

For example, (“asd“,”kjh”,”aas”) + (“sd”,”jh”,”KK”) = (“asd“,”kjh”,”aas”,“sd”,”jh”,”KK”).

- Operations on Het-matrix type values are defined as gathering all relevant elements. At this time, if the elements are not of the same type, then the result of the addition is the zero element.

4.1.2. Subtraction Operation (-)

- Subtraction of numbers is determined according to the definitions in mathematics.
- Subtraction of texts is defined as the removal of the second from the first srting. For example, "asdfghjkasd" - "fg" = "asdhjkasd". The subtrahend can be found in minuend and was deleted from minuend. "Asdfghjkasd" - "mno" = "asdfghjkasd. In this case, the subtraction cannot be performed because the symbols of subtrahend cannot be found. Subtraction from the right side and from the left side can be defined. For example: "asdfghjkasd" - "asd" = "fghjkasd" (subtraction from the

left side) and "asdfghjkasd" - "asd" = "asdfghjk" (subtraction from the right side).

Subtraction of geometric figures can be done in the same way as the addition, but instead of the figures union, intersection is taken. In other words, for the subtraction it is necessary to place geometric figures in their coordinate systems. Subtraction of two figures is defined as a placement of both figures in one coordinate system, and then the removal of their common part. The remaining figure will be the difference of two figures. These figures may do not have intersection and then the resulting figure is the minuend figure:

- Subtraction of two sound recordings is defined as a removal of the second sound record from the first one. For example, after a noise removing from the trail, the clean sound will remain.
- Subtraction of two video files as a removing of the second video from the first video image. For example, if the second video contains errors, then the first video will be cleaned from it.
- Subtraction of the vector type data is defined as a removal of the second vector elements from the first vector.

For example, (“asd“,”kjh”,”aas”)-(“asd”,”aas”)=(“kjh”).

- Operations on Het-matrix type values are defined as gathering all relevant elements. At this time, if the elements are not of the same type, then the result of the subtrcation is the zero element.

4.1.3. Multiplication Operation (*)

- Multiplication of numbers is determined according to the definitions in mathematics.
- Multiplication of the text to text is defined as a "union", but first the division to symbols of the multiplier should be done and then their union to the multiplicand and then union again should be performed. For example: "new" * "ab" = "newa" + "newb" = "newanewb".
- Multiplication of geometric figures means drawing one figure inside another figure. If they do not have a common part, it can be filled with other figures.

For graphical images, these four operations can also be defined in another way, such as operations on the pixels of the figures.

4.2. Logical Operations

We can also define logical operations. If the elements of Het-matrix are of the same type, then the usual rules will remain, and if the elements of the matrix are of different types, then they will be converted to the same type and then compared. For example, if we compare the number 7 and the string "7" we get a true value, and the number 7 and the string "abc" will get False.

5. CONCLUSION

In the paper, algorithms for buildup of a heterogeneous matrix, the operations on these matrices are defined. We use Het-matrices to solve prediction.

REFERENCES

[1] Makridakis, Hibon, 2000. Makridakis, S., Hibon, M., 2000. The M3-Competition: results, conclusions and implications. *International Journal of Forecasting* 16, 451-476.

[2] N. Archvadze, M. Pkhovelishvili. Prediction of Events by Means of Data Parallelism. *Proceedings of International Conference on Matematics, Informatics and Informtional Technologies (MITI2018).* 2018. pp.120-121. www.miti2018.usarb.md

[3] Natela Archvadze, Merab Pkhovelishvili, Lia Shetsiruli. A New approach to Constructing Parallel Algorithms. *Computer Sciences and Telecommunications.* 2018. No.1 (53). pp.30-34. http://gesj.internet-academy.org.ge/en/list_artic_en.php?b_sec=comp

[4] N. Archvadze, M. Pkhovelishvili. The use of parallel data for prediction complex processes. *Proceedings of the System Analysis and Information Technologies 20-th International Conference SAIT 2018.* pp. 210. http://sait.kpi.ua/books/

[5] https://en.wikipedia.org/wiki/Data_type

[6] Tipy i vidy graficheskikh, tekstovykh, audio i video formatov. http://obuchaemsja.com/informacionnaya-baza/tipy-i-vidy-graficheskix-tekstovyx-audio-i-video-formatov/