ON THE METHOD OF DETERMINING LEARNING DESCRIPTIONS TO FORECAST NATURAL DISASTERS WITH THE PATTERN RECOGNITION SYSTEM

Nelly Tkemaladze,

PhD, Senior Scientific Worker, Department of Mathematical Cybernetics Georgia, Tbilisi, V. Chavchanidze Institute of Cybernetics

Violeta Jikhvashvili,

MA, Department of Mathematical Cybernetics Georgia, Tbilisi, V. Chavchanidze Institute of Cybernetics

Giorgi Mamulashvili,

MA, Department of Mathematical Cybernetics Georgia, Tbilisi, V. Chavchanidze Institute of Cybernetics

DOI: https://doi.org/10.31435/rsglobal_ws/31052020/7072

ARTICLE INFO

ABSTRACT

Received: 28 March 2020 Accepted: 04 May 2020 Published: 31 May 2020

KEYWORDS

System, Pattern recognition, Learning description, model, Natural disaster, Forecasting, Matrix. period T0 with SPRL – the System of Pattern Recognition with Learning (elaborated by us) it is necessary to have the data of the previous 12 months of period T0 and learning descriptions (LDs). To identify this latter, the fact of occurrence or non-occurrence of disasters in the same region and the period T0 should be known in other years and also, the above mentioned 12month date for each year. Determining LDs based on them is the aim of the article. For this purpose, the method which will be included in the first model of the SPRL is elaborated. The SPRL comprises: 1) preliminary elaboration of the initial information, 2) learning and 3) recognition models. This system is implemented on a PC. It is verified on the basis of the real data to recognize objects of different classis. Primary, additional and formal additional parameters are determined in the method given in the article. On the basis of their values in correlation with the aforementioned 12 months two matrices are determined. The first of them corresponds to the fact of occurrence of disasters and the second one - of non-occurrence. By using these parameter values given in these matrices LDs will be determined. The best LDs will be given to the learning model of the SPRL for transformation and increasing of informativity. Based on the LDs obtained after the transformation, the learning model will make knowledge and data bases.

To forecast natural disasters (floods, mud-slides) in the fixed region and in

Citation: Nelly Tkemaladze, Violeta Jikhvashvili, Giorgi Mamulashvili. (2020) On the Method of Determining Learning Descriptions to Forecast Natural Disasters with the Pattern Recognition System. *World Science*. 5(57), Vol.1. doi: 10.31435/rsglobal_ws/31052020/7072

Copyright: © 2020 **Nelly Tkemaladze, Violeta Jikhvashvili, Giorgi Mamulashvili.** This is an openaccess article distributed under the terms of the **Creative Commons Attribution License (CC BY)**. The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Introduction. One of the important problems is forecasting the natural disasters (floods, mudslides) [1,2]. To forecast natural disasters with the System of Pattern Recognition with Learning (SPRL) [3,4] in the fixed region in the given period T_0 , it is necessary to have initial learning descriptions. In case of objects, learning description [5,6] is the sequence of parameter values (characteristics) of an object for which the class, the object presented by this sequence belongs to, is known beforehand. The sequence of parameter values is also called realization [3, 4, 7], vector of *m* component (*m* – the number of parameters) [8]. In case of a natural event, occurrence or non-occurrence of such a natural event should be known in the fixed region and the given period T_0 . Let us call this period learning zero block and the sequence of parameter values characterizing the natural event in the first and second case separately in the previous 12 months of period T_0 – learning description.

We have elaborated the System of Pattern Recognition with Learning – SPRL [3, 4]. It includes: 1) model of preliminary elaboration of the initial information, also 2) learning and 3) recognition models. These models include the methods and algorithms of solving 21 main objectives. The SPRL was implemented on a PC. The system is verified on the basis of the real data to recognize objects of different classes. This system can recognize new objects from the list of the given classes even in case when descriptions corresponding to the objects of one and the same class differ from each other more than descriptions corresponding to the objects of different classes.

In order to forecast this or that natural disaster using SPRL (e.g. floods, mud-slides, etc.) in the given year in the fixed region and the particular time period T_0 (let us call this period a zero block), the data for this previous 12 months of this very year should necessarily be known in advance. These data should be presented as a sequence i.e. description of parameter values of the corresponding natural event. In the indicated data must be implied the existing real data – parameter values (characteristic features). They can include characteristic features determining occurrence as well as non-occurrence of natural disasters.

In addition, before the system forecasts whether this natural event will occur or not in the given year in the fixed region and period T_0 (zero block), it should preliminarily elaborate the data about occurred and not occurred natural disasters in previous years in the same region and period T_0 . These data should correspond to the sequences of characteristic features determined on the basis of the data of the previous 12 months in each year in the same region and period T_0 (learning zero block) in case of occurrence and non-occurrence of the natural event. Meanwhile, these data should be given in respect to the learning zero block of the given natural event in case of occurrence and non-occurrence separately. It means that it is necessary to have the learning zero block and its corresponding learning descriptions in case of occurrence as well as non-occurrence of disasters. Elaboration of the learning descriptions is the aim of this article. We have elaborated the method which determines learning descriptions in correlation with the previous 12 months of the learning zero block. It will be included in the first model of the SPRL. Let us call the learning descriptions which correspond to such learning zero block in which a natural disaster has occurred, learning descriptions included in the first class; if a natural disaster has not occurred in the learning zero block - the second class. After determination of the learning descriptions, the best of them will be passed to the learning model. After transformation of the learning descriptions, the model will determine knowledge and data bases for each class separately.

Primary and formal additional parameters. Before learning descriptions are determined, first of all, on the basis of the data of the previous 12 months, for the corresponding natural event should be chosen its corresponding characteristic initial parameters. For example, in case of flood, we can presumably consider the following initial parameters in the fixed region and period: the average air temperature ranges from 12 noon to 12 midnight and from 12 midnight to 12 noon of the next day. From these time intervals, let us call the first of them the first part of the time interval and the second one – the second part of the time interval. Also, we should consider the mean values of air atmospheric pressure, air humidity and wind speed in the same region and in both time intervals. As these parameters are presumable, they can be specified later. For instance: maximum temperature, direct and scattered solar radiation, air relative humidity [2], etc. can be considered as initial parameters (the article deals only with 4 parameters listed above). Obviously, changing or adding the parameters, certainly, will not lead to serious changes in the elaborated method. The changes that will be provoked in correlation with the specificity of forecasting natural disasters, should be considered in the recognition model of the SPRL.

If we separately discuss 4 parameters corresponding to each part of the time interval, we will get the sequences consisting of 8 parameters in correlation with the learning zero block that corresponds to the previous 12 months of the given period and region. Let us call so defined parameters primary parameters. Under this learning zero block (period) we can imply any month of a year.

It is also possible that the values of the primary parameters chosen from the beginning (in our case of 8 parameters determined according to the first and second parts of the time interval indicated above) or the parameters even after their specification are not enough to determine informative learning descriptions and, consequently, to forecast natural disasters. Taking into consideration this fact, we considered it expedient to determine additional parameters. They look as follows:

$$P_1 = \max y_v + \min y_v; P_2 = \max y_v - \min y_v; P_3 = \frac{1}{m} \sum_{j=1}^m y_j = \overline{y}; P_4 = card\alpha;$$

$$P_5 = card\beta; \ P_6 = \sum_{j \in \{j/y_j < \overline{y}\}} y_j; \qquad P_7 = \sum_{j \in \{j/y_j > \overline{y}\}} y_j; \qquad P_8 = \frac{1}{m} \sum_{j=1}^m (\overline{y} - y_j)^2,$$

where y_j denotes the data given in j^{th} period, m – the number of data in j^{th} period. α corresponds to the set of data of above the average values of the data given in the j^{th} period, β corresponds to the set of data of below this average values. Let us determine four parameters based on the formulas given above. Let us mark them with P_9, P_{10}, P_{11} and P_{12} , and call them formal additional parameters.

$$P_9 = p_1 / p_2$$
, $P_{10} = p_3 / p_8$, $P_{11} = p_4 / p_5$ and $P_{12} = p_7 / p_6$.

If we add these formal additional parameters to the sequence of 8 primary parameters, we will get the sequence consisting of 12 parameters.

Matrix 1 and Matrix 2. Using the primary and formal additional parameter values learning descriptions will be determined with the proposed method, on the basis of the previous 12-month data of the learning zero block (let us assume – January).

At first two matrices (matrix 1 and matrix 2) are made to determine these learning descriptions. The first matrix refers to occurrence of natural disasters and the second one - to non-occurrence.

In order to make these matrices (when January corresponds to the learning zero block), we should consider the following sequence of one-month period of the previous 12 months of the respective learning zero block: December, November, October, September, August, July, June, May, April, March, February, January (the latter refers to the next year). The months in the matrices are given in such sequence. In this case, each of these months should be divided into small periods with respect to days. Let us call them learning blocks because the fact of occurrence or non-occurrence of a natural disaster in the corresponding learning zero block is known in advance.

Since the number of days in each month does not coincide with each other, the abovementioned months in correlation with the days will be presented in the form of sequences of the following intervals (learning blocks): the months containing 31 days will be divided into the following intervals [31-24], [23-16], [15-8], [7-1]. The months containing 30 days will be presented in the following form [30-23], [22-15], [14-8], [7-1]; for February, if it belongs to a leap year, intervals will be presented in the following way [29-22], [21-15], [14-8], [7-1], but if the year is not leap, intervals will be presented in the following way [28-22], [21-15], [14-8], [7-1].

Thus, from the corresponding 12 months, each month will be divided into 4-4 learning blocks (short periods). From each matrix 48 learning descriptions will be determined in relation to the corresponding parameters, both parts of the time intervals, each month, and the learning blocks included in it.

In the first row of the matrix 1 is given the name of the parameters, in the next 48 rows are given the average values of the corresponding parameter values in correlation with the parts of the time interval and then in correlation with the learning blocks which is included in the month of the corresponding sequential number. Consequently, the matrix element A_{ii}^t is denoted by the average of

the values of i^{th} parameter which is determined at first with respect to the first part of the time interval (is implied data from 12 noon to 12 midnight) and then according to t^{th} learning blocks of j^{th} month.

 B_{ij}^{t} is determined analogically but in respect to the second part of the time interval (is implied data

from 12 midnight to 12 noon of the next day). Out of the 8 primary elements of the matrix 1, A_i , $i = \overline{1,4}$ and B_{i+4} of indices correspond to one and the same parameter but they have different meanings (loading) according to different parts of the time interval what is expressed with different markers *A* and *B*. The same applies to the elements of the matrix 2.

Out of the 48 aforementioned rows, each of which is a learning description, in the article are given only the first 4 rows. It refers to all periods divided into small periods of only the preceding 12th month of the corresponding learning zero block, with respect to parameter indices and the parts of the time interval.

The month, in which occurrence or non-occurrence of natural disaster should be forecasted in the next years, which is a learning zero block for the matrix, is denoted by $\mathbf{0}_k$. The index k indicates the

sequential number of the corresponding month of the learning zero block whose each month from the corresponding previous 12 months is divided into aforementioned 4-4 short periods (learning blocks).

In one upper line of the matrices is given the name of the preceding 12 months of the learning zero block, in the followed row – sequential numbers of these months, and in the top line of the matrices, along the name of the matrix (in brackets), in the same line is given the name of the learning zero block.

Dec. Nov. Oct. Sep. Aug. July June May Apr. March Feb. Jan. 12 11 10 9 8 7 6 5 4 3 2 1 Matrix 1 (learning zero block 0_1 - January)

 \overline{A}_{ij}^{t} and \overline{B}_{ij}^{t} are determined analogically in matrix 2, but non-occurrence of natural disasters is implied in this matrix. In matrix 2, as in matrix 1, after the first row out of 48 rows are given only the first 4 rows of the 12th month.

Matrix 2 (learning zero block 0_1 - January)

A_1	A_2	A_3	A_4	B_5	B_6	B_7	B ₈	$P_{9}, 1$	P ₁₀ ,	P ₁₁ ,	P ₁₂
$\overline{A_{112}^1}$	$\overline{A_{212}^{1}}$	A_{312}^1	A^{1}_{412}	B_{512}^1	$\overline{B_{612}^{1}}$	$\overline{B_{712}^1}$	$\overline{B_{812}^{1}}$	P_{912}^{1}	P_{1012}^1	P_{1112}^1	P_{1212}^1
$\overline{A_{112}^2}$	$\overline{A_{212}^2}$	$\overline{A_{312}^2}$	A_{412}^2	$\overline{B_{512}^{2}}$	$\overline{B_{612}^{2}}$	$\overline{B_{712}^2}$	$\overline{B_{812}^{2}}$	P_{912}^{2}	P_{1012}^2	P_{1112}^2	P_{1212}^2
$\overline{A_{112}^{3}}$	$\overline{A_{212}^{3}}$	$\overline{A_{312}^3}$	$\overline{A_{412}^3}$	$\overline{B_{512}^{3}}$	$\overline{B_{612}^3}$	$\overline{B_{712}^{3}}$	$\overline{B_{812}^{3}}$	$\overline{P_{912}^{3}}$	$\overline{P_{1012}^{3}}$	$\overline{P_{1112}^{3}}$	$\overline{P_{1212}^{3}}$
$\overline{A_{112}^4}$	$\overline{A_{212}^{4}}$	$\overline{A_{312}^4}$	$\overline{A_{412}^4}$	$\overline{B_{512}^{4}}$	$\overline{B_{612}^4}$	$\overline{B_{712}^{4}}$	$\overline{B_{812}^{4}}$	$\overline{P_{912}^{4}}$	$\overline{P_{1012}^4}$	$\overline{P_{1112}^4}$	$\overline{P_{1212}^{4}}$
771			•		1	1		1	1		`

Thus, two matrices are obtained: matrix 1 and matrix 2.

Any row from the second row of these matrices is the primary learning description. This is caused by the fact that these rows are given in the form of sequences of parameter values characterizing a natural event (these sequences include primary parameter and formal additional parameter values). In addition, at the same time, in their respective learning zero block the fact of occurrence of a disaster is known in advance in case of the matrix 1, and in case of the matrix 2 – the fact of non-occurrence of the same natural disaster. This means that according to the data from the previous 12 months corresponding to the learning zero block can be determined by the appropriate learning descriptions, also, the descriptions corresponding to the zero block, i.e. when the fact of occurrence or non-occurrence of the disaster in the zero block is not known.

Choosing learning descriptions. For each sequence (learning description) let us calculate the differences between each fixed i^{th} parameter value in respect to the both parts of the time interval separately and for t^{th} learning block of j^{th} month. Let us indicate these differences with d_{ij}^{Xt} , $d_{ij}^{Xt} = X_{ij}^{t} - \overline{X}_{ij}^{t}$, where under X is implied A or B primary $i^{\text{th}} = \overline{1,8}$ parameter values in correlation with the first or second part of the time interval consequently or $P i^{\text{th}} = \overline{9,12}$ – formal additional parameter values in correlations.

Thus, we get such sequences of differences that will allow us to choose the best sequences (consequently, learning descriptions from matrices). They will be chosen with the help of the following algorithm which comprises 4 stages:

1. For each of the above sequences, let us choose the maximum and minimum values from the values contained in it. Let us denote them with $\max d_{ij}^{Xt}$ and $\min d_{ij}^{Xt}$, and call them characteristics of their sequences (corresponding learning descriptions). By using them and the vector-optimization method of choice of the best variants [9], let us determine the best sequences i.e. such sequences which belong to Pareto set. Let us mark the set of such sequences with D_1 .

2. For each sequence, let us determine the average meaning of the values included in it. Let us determine the set of the best sequences the same way as given in the stage 1, but here the average value and maximal value are considered as characteristics of sequences. Let us mark this set with D_2 .

3. The procedures given in the stage 2 will be used in this stage, but minimal and average meaning will be used as characteristics (vector components) of sequences. Let us mark this set with D_3 . Thus, we will get 3 sets D_1 , D_2 , D_3 of the best sequences. Let us denote their united set with D. It should contain various differences (consequently, matrices should contain different learning descriptions).

4. If so determined set for D, cardD < 40, then, on the basis of the characteristics of the sequences of differences remained beyond set D for each difference, we will calculated vector lengths. According to the value of these lengths, we will fill set D so that is should contain 40 sequences. Consequently, 40 learning descriptions will be chosen from each matrix. As the statistics of using the SPRL has shown (in case of recognitions of objects) this quantity is enough to make knowledge and data bases for the learning model. Thus, such learning descriptions are obtained, that will be passed to the learning model of the SPRL. The learning model will transform them, increase informativity and define the knowledge and data bases in the process of machine learning [10].

To increase the informativity of learning descriptions, the learning model uses: from combinatorial mathemetics balanced and partially balanced incomplete block-designs and tactical configurations of (v,b,k,r,λ,μ) type [11,12], geometrical configurations [13] and the vector-optimization method of choice of the best variants. Namely with their help the learning model determined new artificial (formal) parameters, functions which show the internal hidden connections between the primary characteristic parameters of the natural event which really exist between them, but are not explicitly given in the primary learning descriptions. At the same time, thus defined parameters will increase this number in case of their small number and, but in case of large quantity of parameters) and their combinations which are characteristic to only one (i.e. each different) class. After this, learning descriptions will be recorded in the language (new codes) from which the learning model determines data bases for each class separately in case of occurrence of a disaster as well as in case of its non-occurrence.

The knowledge base contains all those formulas (functions), values which are used for transformation of learning descriptions and the language (new codes) in which learning descriptions will be recorded.

The data base which is determined on the basis of the above mentioned knowledge base contains the characteristic features of the both classes: single characteristic features, feature pairs, triplets and specific combinations (groups) of characteristic features. These groups contain combinations of characteristic and non-characteristic features of classes. The triplets as well as these combinations are determined using the aforementioned combinatorial schemes without exhaustive search. This fact significantly decreases the quantity of triplets and these combinations due to what their use becomes possible.

Using the method for determining learning descriptions, was discussed that case when occurrence or non-occurrence of natural disasters should be forecasted in January according to which matrix 1 and matrix 2 were determined, and on their basis – learning descriptions. The same method is used for each month separately.

For this purpose, the corresponding matrix 1 and matrix 2 will be made for each month (which we consider as a learning zero block). On the basis of these matrices, the same procedure will be used that was used for January.

Thus, to determine learning descriptions, besides the method given in this article, only the two models of the SPRL are used. After determination the learning descriptions to forecast disasters it is necessary to use the third (recognition) model of the SPRL, but it was considered only for object recognition. This is caused by the fact that in the SPRL elaborated by us is considered recognition of

only objects (satellite types, aircrafts, diseases, schedules, irises etc.) and is not considered the specifics of forecasting (recognition) of natural disasters, events.

Unlike recognition of any object, the specificity of recognition of a natural event (disaster) is as follows: a natural event is not forecasted by using such learning descriptions that correspond directly to the learning zero block and on the basis of these descriptions do not determine knowledge and data bases (as it happened in case of objects). In case of a natural event, based on the data of the previous 12 months of the learning zero block, learning descriptions (from which the control descriptions are separated by the first model) must be determined.

In this case at first the conditions of the previous period of this learning zero block should be studied the first model separates control descriptions from the initial learning descriptions from the very beginning. That is why it is implied that the fact of occurrence (non-occurrence) of a disaster in the learning zero block is not known to them. At the same time, it is obvious that control descriptions do not participate in making the aforementioned knowledge and data bases. Therefore, for them, the learning zero block has the role of the zero block, while the control descriptions play the role of new descriptions that are determined on the basis of the data corresponding to the previous 12 months of the zero block.

Therefore, at first, it is necessary to recognize the condition of the previous periods of this zero block. The learning model first transforms these learning descriptions in order to increase informativity. Then, with their help, it determines the knowledge and data bases separately using the data from all previous years (at least 5 years) on the basis of the determined learning descriptions. These bases determined in different years will be transferred to the recognition model. After this, in the learning process, control descriptions should be recognized by using different knowledge and data bases determined in the different previous years. Namely on the basis of the results of the recognition of these control descriptions will be recognized occurrence or non-occurrence of a disaster in the zero block, because these results of recognition show how correctly these bases correspond to the fact of occurrence or non-occurrence of a disaster, to what extent it is possible to recognize new disasters in the same region and period on the basis of the knowledge and data bases determined in the learning process. This leads to a number of changes to the recognition model (change of the decision-making criteria, etc.).

When the initial information is given or it can be presented in the form of the learning descriptions, we can set the objective of forecasting natural disasters in terms of pattern recognition with learning. This, in its turn, conditions necessity of determining the learning descriptions corresponding to natural disasters to solve the aforementioned objective. This fact is caused by the fact that after determining such learning descriptions it is possible to use all three models of the SPRL but after modification of the recognition model (what is a separate objective). At the same time, the data in relation with the previous 12 months of the zero block should necessarily be given.

Conclusions. To forecast natural disasters (floods, mud-slides) with the system of pattern recognition with learning – SPRL (elaborated by us) in the given region and period T_0 , besides having the data of 12 months prior to this period, it is necessary to have learning descriptions - LDs. For their determination it is necessary to have the data of the previous 12 months in the same region and period T_0 of other years in case of occurrence and non-occurrence of disasters. Determining LDs based on the given last data is the aim of the article. For this purpose, the method which will be included in the model of preliminary elaboration of the initial information (the first model of SPRL) is elaborated. First of all, primary, additional and formal additional parameters are determined in the method. On the bases of these parameter values, two matrices are determined. The first of them corresponds to the fact of occurrence of disasters and the second one – of non-occurrence. The values of the parameters are given in these matrices. On the basis of them LDs are determined in correlation with the parts of the time intervals, each month of the previous 12 months and the learning blocks included in each month. Thus the determined LDs are passed to the learning model (the second model of SPRL) for further transformation and for making knowledge and data bases. For this purpose, the learning model uses balanced and partially balanced incomplete block-designs, tactical configurations of $(v, b, k, r, \lambda, \mu)$ type [11, 12], geometrical configurations and the vector-optimization method of choice of the best variants.

Thus, to determine learning descriptions, besides the method given in this article, only the two models of the SPRL are used. The learning model transfers the bases determined in different years to the recognition (third) model to forecasting a disaster only after this model is modified (what is a separate objective).

REFERENCES

- 1. V. F. Krapivin, I. I. Potapov, V.Yu. Soldatov. Natural Disasters and Natural Disaster Forecasting. Problems of Environment and Natural Resources. Review information № 1. M. 2017.
- 2. Ts. Basilashvili, M. Salukvadze, V. Tsomaia, G. Kherkheulidze. Catastrophic Flooding, Mud-slides and Avalanches in Georgia and Their Safety. Monograph. 'Technical University'. Tb. 2012.
- 3. Tkemaladze N. Theory of the System of Pattern Recognition with Learning and Its Application. Monograph. LAP LAMBERT. Academic Publishing. Norderstedt/Germany. 2017.
- 4. Tkemaladze N. System of Pattern Recognition with Learning and Its Theoretical Principles. Monograph. 'Technical University'. Tb. 2013.
- 5. N. Tkemaladze. On the problems of the automatized system of pattern recognition with learning. Journal of Biological Physics and Chemistry (JBPC). Vol. 2, #34, AMSI, CB, 12/2002.
- 6. Zhuravlyov Y. I. On Algebraic Approach to Solving the Recognition Problem or Classifications. Problems of Cybernetics, issue 33. 1978.
- 7. Verulava O., Khurodze R. Theory of Rank Relations Modeling of Recognition Processes. Georgian Technical University. Tbilisi, 2004.
- 8. Tu J., Gonzalez R. Principles of Pattern Recognition. Moscow: Mir, 1978.
- 9. Aizerman M. A., Malishevski A. B. Some Aspects of the General Theory of Choice of the Best Variants. Automation and Remote Control, #2, 1981.
- 10. Bishop Cr. Pattern Recognition and Machine Learning. NY, 2007.
- 11. Hall M. Combinatorics. Moscow, Mir, 1970.
- 12. N. Tkemaladze. Recognition, Classification, Estimation. Metsniereba. Tb. 1990.
- 13. Mason J. Metroids as Geometrical Configurations. Problems of Combinatorial Analysis. Moscow, 1980.