INTERVAL PATTERN RECOGNITION IN RELATION TO INFORMATION ABOUT THE TREATMENT OF COVID 19 IN PATIENTS WITH BLOOD DISEASES

[1] G. Tsitsiashvili, [2] V. Nevzorova, [2] A. Talko, [1,3] M. Osipova

•

[1] Russia, 690041, Vladivostok, Radio street 7, Institute for Applied Mathematics
Far Eastern Branch of Russian Academy Sciences
[2] Russia, 690002, Vladivostok, 2 Ostryakova Ave.,
Pacific State Medical University,
[3] Russia, 690002, Vladivostok, FEFU Campus 10 Ajax Bay,
Russky Island, Far Eastern Federal University,
guram@iam.dvo.ru, nevzorova@inbox.ru, talkang92@mail.ru, mao1975@list.ru

Abstract

Since the problem of coronavirus infection, especially in vulnerable groups, remains relevant in the healthcare system, it is necessary to recognize the risk of fatal events during the treatment of patients from COVID-19. Patients with hemoblastosis are exposed to a more severe course of coronavirus infection than the general population, as well as an increased risk of fatal events. The aim of our study was to determine the effect of signs on the risk outcomes of fatal events among patients with diseases of the blood system, both of a tumor and non-tumor nature. According to the values of the signs, it was necessary to recognize survivors and those who died during treatment from COVID-19. The method of interval pattern recognition was chosen due to the presence of big data, it is described in detail in the article. The patients were broken down by gender and age. The signs that significantly affect the forecast according to the developed algorithm were identified. This is especially evident in groups of men and women with malignant diseases of the blood system over the age of 60 years. In these groups, a positive outcome of treatment is detected due to the presence of a large set of uninformative signs. This phenomenon is closely related to the tasks of technical gerontology.

Keywords: interval pattern recognition, big data, informative signs, technical gerontology, coronavirus infection, hemoblastosis, tumor.

Introduction

According to world studies, it is known that patients with hemoblastosis are susceptible to a more severe course of coronavirus infection than the general population, as well as an increased risk of fatal complications [1, 2, 3]. In the era of the ongoing coronavirus pandemic, it is an urgent task for the population and the healthcare system to recognize the risk of fatal events during the treatment of COVID-19 patients with benign and malignant diseases of the blood system based on the results of their examination.

The initial sample contains data from medical records of the medical history of 221 patients with diseases of the blood system of Primorsky Krai who were treated for COVID-19 coronavirus infection in the infectious department of the KKB2 in Vladivostok, among which 48 patients had non-tumor (benign) diseases of the blood system, 173 - tumor (malignant) diseases of the blood system. Patients with blood tumor diseases were divided depending on the age category into groups: women at least 60 years of age - 60 patients, under 60 years of age - 31 patients, men at least 60 years of age - 53 patients, under 60 years of age - 29 patients. The use of medical data was made after the signing of informed voluntary consent from the patient.

As a result of this partitioning, the original rather heterogeneous sample was divided into five relatively homogeneous subsamples. Each of these samples was characterized by about 40 different characteristics. Some of these features have the character of Boolean variables of the yes - no type, and some are the results of laboratory studies. Despite the relatively small number of patients in each of the selected groups, the number of signs is quite large. This property allows the source information to be characterized as Big Data. The task was to investigate the effect of signs on the result of treatment in the selected groups. According to the values of the signs, it was necessary to recognize survivors and those who died during treatment from coronavirus infection. The presence of Big Data in the problem required solving by an algorithm with the lowest possible computational complexity. The method of interval pattern recognition was chosen as such a method. This method [4] was developed to predict outbreaks of tick-borne encephalitis by meteorological characteristics of the winter period and later extended to the analysis of extreme meteorological phenomena during last twenty years. The issues of medical information processing considered in the paper are closely related to the tasks of technical gerontology [5] – [7].

I. Interval pattern recognition and its properties

Let there be two sets of m-dimensional vectors characterizing some objects:

$$X^0 = \{(x^0_{11}, \dots, x^0_{1m}), \dots, (x^0_{n_0 1}, \dots, x^0_{n_0 m})\},$$

$$X^1 = \{(x^1_{11}, \dots, x^1_{1m}), \dots, (x^1_{n_1 1 1}, \dots, x^1_{n_1 m})\}.$$

We assume that the components of these vectors with numbers $k = 1, \dots, m' < m$ are Boolean variables, and the components with numbers $k = m' + 1, \dots, m$ are real variables. Each vector from the sets X^0, X^1 describes Boolean/real features of a separate object. Our task is to recognize whether objects belong to the set of X^0 or to the set of X^1 by the vector of signs characterizing these objects.

The essence of the interval pattern recognition method is as follows. An object defined by the vector $(x^1_{k1}, \dots, x^1_{km})$ is recognized as an element of the set X^0, if all inequalities are met

$$x_i^- = \min_{1 \le k \le n_0} x^0_{ki} \le x^1_{ki} \le \max_{1 \le k \le n_0} x_{ki} = x_i^+, \ i = 1, \dots, m. \quad (1)$$

Otherwise, the object characterized by the vector $(x^1_{k1}, \dots, x^1_{km})$ is recognized as an element of the set X^1, denote their number s (the number of correctly recognized objects of the set X^1). The quality of this method is determined by the ratio

$$\rho = \frac{n_0 + s}{n_0 + n_1}, \quad (2)$$

since the number of correctly recognized objects of the set X^0 is n_0.

Let's list some properties of the value ρ.

1. All objects of the set X^0 are correctly recognized.
2. The recognition quality of ρ increases with the number of m features.
3. The number of arithmetic operations for implementing interval pattern recognition depends linearly on m and on n.
4. The Boolean attribute i, such that not all numbers $x^0_{k,i}$ match, does not affect the quality of recognition ρ in any way.

Remark 1. *In multidimensional statistics, the recognition quality increases with the growth of n and with the decrease of m. In turn, in many real observations, on the contrary: n is small (about 30), and m is quite large (at least 5). This circumstance in conjunction with the properties {***2***,* **3** *makes it convenient to use the interval pattern recognition method when analysing Big Data.*

But to use this method to medical information, some adaptation is required. Particular attention should be paid to the signs determined by Boolean variables, since doctors consider these signs to be more reliable. Therefore, first a set of I_1 Boolean features is allocated (conditionally denote them $1, \ldots, m'$), according to which all objects of the set X^0 take the same values. Next is s_1 of objects, (conditionally denote them $1, \ldots, n_1^1$) of the set X^1, which are correctly recognized by the pain signs of the set I_1. Let X_1^1 be the collection of all such objects in the set X^1.

The next step is to build segments of the type (1) based on laboratory signs and check objects $X^1 \setminus X_1^1$ for the correctness of their recognition by signs $m' + 1, \ldots, m$. Denote s_2 the number of objects of the set $X^1 \setminus X_1^1$ correctly recognized by these signs, then the equality $s = s_1 + s_2$ is fulfilled. For each feature i, $i = 1, \ldots, m$,, the number of elements of the set X^1 correctly recognized by it, divided by the number of elements of this set n_1, is compared as its informativeness.

II. Research results

For calculations from the initial medical information, the following signs were identified:
1. Phase of hematological disease at the time of diagnosis of COVID-19: 1 - remission induction, 2 - first remission, 3 - second and subsequent remission, 4 -relapse therapy, 5 - refractory course, 6 - no data;
2. Number of comorbidities (such as obesity, diabetes mellitus, cardiovascular disease, respiratory disease, liver disease, chronic kidney disease);
3. ECOG stage (from 0 to 4) at the time of admission (ECOG is a scale for assessing the general condition of a cancer patient);
4. ECOG maximum stage (that is, how the general condition of the patient has changed from the moment of infection with coronavirus to the moment of recovery/death);
5. Indicators of a clinical blood test: 5.1.hemoglobin (g/l), 5.2.leukocytes (g/l), 5.3. erythrocytes (g/l), 5.4. platelets (g/l); 5.5. ESR (erythrocyte sedimentation rate, mm/h), 5.6. eosinophils (%), 5.7. P / I (stab neutrophils, %), 5.8. c / i (segmented neutrophils, %), 5.9. lymphocytes (%), 5.10. monocytes (%);
6. Parameters of blood biochemistry: 6.1. ALT (IU/l), 6.2. Ast (IU/l), 6.3. GFR (glomerular filtration rate), 6.4. creatinine (mmol/l), 6.5. urea (mmol/l), 6.6.total protein (g/l), 6.7. total bilirubin (mmol/l);;
7. Indicators of the blood coagulation system: 7.1. APTT (activated partial thromboplastin time, sec), 7.2. PT (prothrombin time, sec), 7.3. fibrinogen (g/l).

The rest of the blood signs are not informative (few of the deceased have values in the table). Some of the Boolean features allocated to doctors do not affect the quality of recognition due to property 5.

Table 1. Quality of interval recognition

	n_0	n_1	$n_0 + n_1$	s	ρ
Benign blood diseases (group 1)	10	38	48	38	1
Malignant blood diseases, women at least 60 years old (group 2)	24	36	60	22	0,77
Malignant blood diseases, women under 60 (group 3)	15	16	31	11	0,84
Malignant blood diseases, men at least 60 years old (group 4)	28	25	53	15	0,81
Malignant blood diseases, men under 60 (group 5)	12	16	28	15	0,96

The interval pattern recognition algorithm described above (in the presence of Boolean signs) calculated the indicator ρ, characterizing the quality of interval recognition for all selected groups of patients (see Table 1). In our notation n_0 is the number of patients who died from coronavirus infection, n_1 is the number of patients who recovered from coronavirus infection, s is the number of recovered patients recognized as recovered (correctly recognized).

Table 2. Informative value of recognizing signs

feature number	Group 1	Group 2	Group 3	Group 4	Group 5
1	32/38	0	1/16	0	0
2	11/38	4/36	1/16	0	0
3	0	0	0	0	6/16
4	0	0	0	3/25	7/16
5.1	10/38	1/36	1/16	1/25	1/16
5.2.	33/38	2/36	2/16	1/25	1/16
5.3.	13/38	1/36	0	1/25	10/16
5.4.	18/38	0	0	7/25	3/16
5.5.	4/38	1/36	2/16	0	1/16
5.6.	15/38	4/36	1/16	2/25	1/16
5.7.	2/38	0	0	0	0
5.8.	35/38	0	0	0	3/16
5.9.	6/38	2/36	2/16	1/25	2/16
5.10.	1/38	0	6/16	0	0
6.1	7/38	4/36	2/16	0	1/16
6.2.	3/38	1/36	0	0	0
6.3.	2/38	1/36	1/16	1/25	3/16
6.4.	7/38	1/36	0	1/25	3/16
6.5.	28/38	2/36	0	0	2/16
6.6.	9/38	3/36	0	0	0
6.7.	6/38	1/36	0	0	2/16
7.1.	10/38	3/36	7/16	1/25	4/16
7.2.	4/38	0	0	0	0
7.3.	9/38	1/36	1/16	1/25	1/16

From Table 2, it follows that in group 1 of patients with benign blood diseases, three signs are distinguished with an informativeness greater than 0.70, the maximum of them is 0.92. In the group 2 female patients at least 60 years old with malignant blood diseases, the maximum informative value of signs is 0.11, but there are many signs with small values. In the group 3 female patients under 60 years of age with malignant blood diseases, the maximum informative value of signs is 0.38, however, there are many signs with zero informative value. In a group 4 male patients at least 60 years old with malignant blood diseases, the maximum informative value of signs is 0.28, but there are many signs with zero informative value. In a group 5 male patients under 60 years of age with malignant blood diseases, the maximum informative value of signs is 0.62. From a comparison of signs with maximum effectiveness, it can be seen that they are different in different groups (see Figure 1). In conclusion, it should be noted that due to its speed, the interval pattern recognition method proved to be convenient when processing big data in the treatment of coronavirus infection in various age groups of patients with diseases of the blood system. This is especially evident in groups of men and women with malignant diseases of the blood system over the age of 60 years. In these groups, a positive outcome of treatment is detected due to the presence of a large set of uninformative signs. The research was carried out within the state assignment for IAM FEB RAS (N 075-01290-23-00).

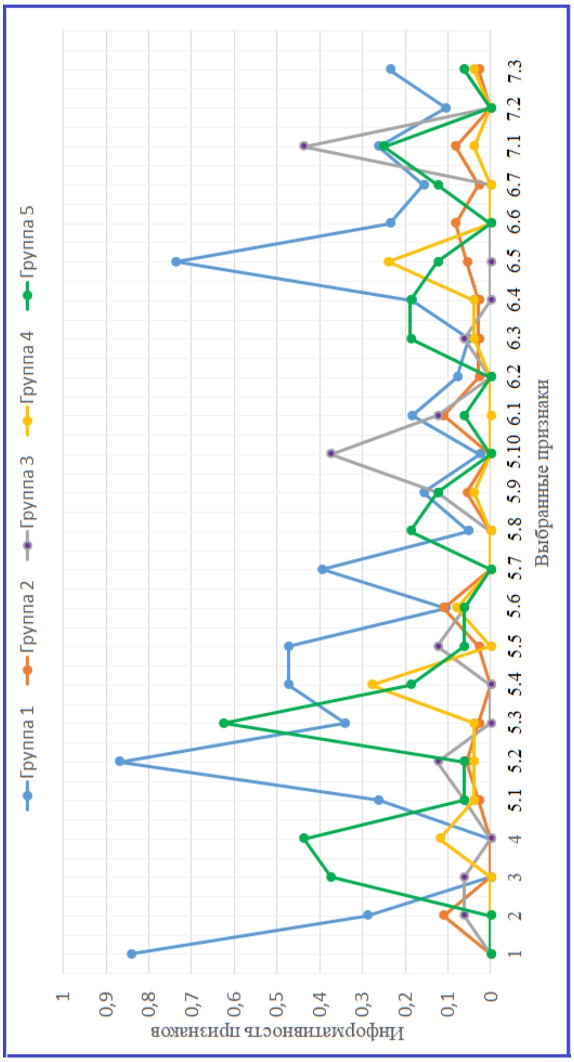

Figure 1: *Graphic representation of informative features in groups of patients*

References

[1] Pinana, J. L. et al. (2020) Risk factors and outcome of COVID-19 in patients with hematological malignancies. Experimental hematology and oncology, 9: 21–25. DOI: 10.1186/s40164-020-00177-z

[2] Gossell-Williams, M. (2020)

Faculty Opinions recommendation of Case Fatality Rate of Cancer Patients with COVID-19 in a New York Hospital System. Faculty Opinions. Post-Publication Peer Review of the Biomedical Literature. DOI:10.3410/f.737865566.79357670

[3] Kaprin, A. D. et al. (2020) The impact of the COVID-19 pandemic on cancer practice Siberian Journal of Oncology, 19(3): 5–22 (In Russian). DOI: 10.21294/1814-4861-2020-19-3-5-22.

[4] Tsitsiashvili, G. Sh.: editor. Efficient algorithms of time series processing and their applications. Nova Publishers Science. N.Y., 2009.

[5] Abramov, O. V. (2013) Algorithm for estimating and predicting the remaining resource of complex technical systems Reliability and Quality of Complex Systems, I, 5-?6. (In Russian).

[6] Kanade, T.: Guest Editor. (2012) Quality of life technology Proceedings of the IEEE, 100, 2394?-2396. DOI: 10.1109/JPROC.2012.2200555

[7] Schulz, R.et al (2015) Advancing the Aging and Technology Agenda in Gerontology Gerontologist, 55 (5), 724-?734. DOI: 10.1093/geront/gnu071