ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Mathematical Support For The Formation Of Informative Signs Dictionary For The Probabilistic Estimates Calculation Of The Repeated Stroke.

Sudakov O, Gladskikh N, Alexeev N*, Bogacheva E, Kudashova E, Sudakov D, Kuzmenko N, and Fursova У.

Voronezh State Medical University named after N.N. Burdenko.

ABSTRACT

The article describes the stages of the formation of a dictionary of informative signs for calculating the risk of recurrent stroke. The limitations in the part of informativeness and measurability of the selected features are considered. Methods for minimizing informative redundancy are considered, and the properties of diagnostic coefficients are described. An algorithm for the formation of a dictionary of informative features is presented. Diagnostic coefficients and informative values for 20 signs associated with the development of ischemic stroke were calculated.
Keywords: repeated stroke, diagnostic features, mathematical model

**Corresponding author*

INTRODUCTION

At a preliminary stage, when calculating the probability of repeated stroke, it is necessary to obtain a list of the most informative signs. In medical research, the concept of informative character of a trait is associated with its diagnostic value in problems of differential diagnosis.

The formation of a vocabulary of signs used to calculate the probability of repeated stroke is an important and rather difficult task.

When developing a vocabulary of attributes, we have to face some limitations. One of them is that only those characteristics for which there is a priori information sufficient to describe classes in the language of these characteristics can be included in the dictionary. Another limitation is that some of the signs are inappropriate to include in the a priori dictionary because they are of little informative.

In the working dictionary, only those signs should be used which, on the one hand, are the most informative and, on the other hand, can be determined by available or specially created monitoring tools.

The dictionary of signs constructed in this way should be an informative basis for calculating the probability of repeated stroke.

The definition of a vocabulary of signs is possible using the following approaches:

1. The game approach to building a vocabulary of signs.
2. A method based on the comparison of a posteriori probabilities.
3. A method based on a comparison of probabilistic characteristics.
4. Method based on the determination of the amount of information.
5. A method based on the definition of Kulbak'sinformativeness.

In this study, an approach based on the determination of the Kullback attributes was made. This method, in comparison with other methods of minimizing informational redundancy, is the simplest and is available for algorithmization. Its machine adaptation is not time consuming and does not entail significant computational costs and resources.

The methodology for calculating the information content of the Kulbak attributes is based on the calculation of the diagnostic coefficients.

The diagnostic coefficient is represented as a logarithm of the ratio of the probabilities of manifestation of this characteristic in the main and control groups (p (Xij | A1) and p (Xij | A2), respectively) and multiplied by 100.

$$DC = 100 \lg \frac{P(x_j / A_1)}{P(x_j / A_2)} \quad (1)$$

Diagnostic coefficients are often two-valued or single-valued positive or negative numbers. They are positive if the probability p (Xij|A1) in the numerator is negative, if the probability p (Xij|A2) prevails. That is, the diagnostic coefficients with the plus sign indicate more plausibility of the hypothesis A1 (about belonging to the main group) with the sign "-" - about the greater likelihood of hypothesis A2 (belonging to the control group). Obviously, coefficients with a positive sign carry positive information, approximating the sum of the diagnostic coefficients to the threshold, which for A1 is positive. Coefficients with a negative sign, on the contrary, "give away" the sum from the threshold. For hypothesis A2, on the contrary, the coefficients with a negative sign approximate the sum to the threshold, and the coefficients with a positive sign-distance it from the threshold, since the threshold is negative.

It should be noted that the greater the value of the diagnostic coefficient, the more differential-diagnostic information, that is, the information on the prevalence of the probability of one of the diagnoses, it carries. However, the informativeness of each value of the characteristic also depends on the frequency with which this value occurs for each of the diseases, i.e., the values x_i^j of p (Xij | A1) and p (Xij | A2). If the diagnostic coefficient of the value of the sign is large, but patients with this value are relatively rare, then in the course of diagnosis the role of such a sign value is small.

To determine the information that the tag carries, you first need to calculate the amount of information that the characteristic values x_i give. For this, it is necessary to multiply the diagnostic coefficient obtained for a given sign of DC (x_i) by the difference in the probabilities of this characteristic when belonging to the main group (hypothesis A1) and to the control group (hypothesis A2):

$$\mathrm{DC}(x_i^j)[p(Xij|A1) - p(Xij|A2)] \qquad (2)$$

It should be noted that the difference [p (Xij | A1) - p (Xij | A2)] is positive if DC is positive. The difference (2)Will show how, on average, the sum of the diagnostic coefficients approaches the threshold as a result of the detection of a symptom in a patient.

Similarly, other values of the same characteristic are calculated. The informativeness of the sign as a whole I (xi) will be equal to their sum:

$$I(xi) = \sum DC(x_i^j)[\,p(Xij|A1) - p(Xij|A2)] \qquad (3)$$

If we represent the value of the DC in its expanded form, then the formula (4/3) takes the form identical to the Kullback formula:

$$I(xi) = \sum 100 \lg \frac{\mathrm{p(Xij\,|\,A1)}}{\mathrm{p(Xij\,|\,A2)}}\ [\,p(Xij|A1) - p(Xij|A2)] \qquad (4)$$

The algorithm for forming a dictionary of informative features is shown in Figure1.

ISSN: 0975-8585

Formation of the main and control group of patients

↓

Calculation of probability of manifestation of a characteristic in the main and control groups $p(Xij|A1)$и$p(Xij|A2)$

↓

Calculation of diagnostic coefficients for traits DC(x_i^j)

↓

Calculating informativity for a given characteristic value DC(x_i^j)[$p(Xij|A1)$ - $p(Xij|A2)$]

↓

Calculating informativity$I(xi) = \sum\ DC(x_i^j)[\,p(Xij|A1) - p(Xij|A2)]$

↓

Selection of the most important signs $I(xi)$

Fig 1: The algorithm for forming a dictionary of informative features.

The advantages of this criterion over other, for example, non-parametric X2 or U criteria as an informative index is that it characterizes the average degree of approximation of the sum of diagnostic coefficients to the diagnostic threshold due to the feature in question.

Using the proposed methodology, diagnostic coefficients and values of informativity for each of the signs were calculated:

X1 - impaired consciousness
X2 -Hemianopsia
X3 - paresis in the hand
X4 - paresis in the leg
X5 - Sensitivity disorder (hemygipostezia)
X6 is a symptom of negation (anosognosia)
X7 - aphasia
X8 - heart rhythm disturbance
X9 - diabetes mellitus
X10 - blood glucose at the time of stroke
X11 - ultrasound dopplerography (UZDG)
X12 - age
X13 - floor
X14 - Blood pressure

X15 - cholesterol
X16 - CHD
X17 - Localization of the outbreak in the basins
X18 - frequency of subtypes
X19 - The severity of a stroke by Rankin
X20 - Barthel points

Examples of calculation of informativity for paresis in the hand (X3) and paresis in the leg (X4) are presented in Tables 1, 2.

Table 1: Results calculation of the informative value of the diagnostic trait paresis in the hand (X3)

p(x3i|A1)	p(x3i|A2)	DC_X3i	I(X3i)
0,39	0,64	-21,5115	5,377884
0,32	0,19	22,63964	2,943153
0,21	0,11	28,08266	2,808266
0,07	0,04	24,3038	0,729114
0,01	0,025	-39,794	0,59691
		I(X3)	12,45533

Table 2: Results of calculation of the informative value of the diagnostic sign paresis in the leg (X4)

p(x4i|A1)	p(x4i|A2)	DC_X4i	I(X4i)
0,52	0,69	-12,2846	2,088378
0,38	0,26	16,48102	1,977723
0,1	0,03	52,28787	3,660151
0	0,01	0	0
0	0,0125	0	0
		I(X4)	7,726252

Values of informativeness for signs X1-X20 are given in Table 3.

Table 3: Results of calculating the informative value of diagnostic features using the Kullback test

Value	Informativity
X6	0
X2	0,0175
X5	0,13086
X1	0,597

X19	0,822
X9	1,278
X20	1,42
X18	3,0178
X17	3,139
X8	4,295
X13	4,44
X10	5,40892
X7	5,9785
X12	6,39
X15	6,46
X4	7,7263
X16	12,08
X3	12,455
X11	15,557
X14	21,61

Thus, as a result of the analysis of the informativeness of the signs, a feature space was created that allows to fully identify the state of the modeling object. Several diagnostic significant signs were selected:

X3 - paresis in the hand
X4 - paresis in the leg
X7 - aphasia
X8 - heart rhythm disturbance
X10 - blood glucose at the time of stroke
X11 - ultrasound dopplerography (UZDG)
X12 - age
X13 - floor
X14 - Blood pressure
X15 - cholesterol
X16 - CHD
X17 - Localization of the outbreak in the basins
X18 - frequency of subtypes

The dictionary of signs constructed in this way is an informative basis for calculating the probability of repeated stroke.

REFERENCES

[1] Esaulenko I.E.Genetic and Clinic-Pathogenetic Peculiarities of Prediction of Development and the Effects of Obesity at Young Persons/I.E. Esaulenko,T.N.Petrova, E.N.Kolesnikova, O.V.Sudakov //Reseaech

Journal of Pharmaceutical, Biological and Chemical Sciences. 2017; № 8(3): p. 1368-1374.

[2] Shiryaev O.Y. Psychosomatic Aspects of Chronic Heart Failure / O.Y. Shiryaev, V.L Yankovskaya., A.V. Budnevsky// International journal of biomedicine. 2017, № 7(3): p. 248-250.

[3] Tokmachev R.E.The Possibility of Nob-Pharmacological Methods in Increasing Clinical Efficiency of Treating Patients with Chronic Heart Failure and Metabolic Syndrome / R.E. Tokmachev, A.V.Budnevsky, KravchencoA.Y. // Research journal of pharmaceutical biological and chemical sciences. 2017,№ 8(6): p. 932-839.

[4] Shtankov S.I., Shiryaev O.Y., Sudakov O.V. and others Software Implementation the Methodology for Calculating Integral Indicators Rehabilitation Potential of Patientswith Schizofrenia / Research journal of pharmaceutical biological and chemical sciences. 2018, №9(1): p. 950-953.

[5] Mathematical, algorithmic and software support of the information support complex for making medical decisions in patients with diabetes mellitus and arterial hypertension / O.V. Sudakov, N.A. Gladskikh, E.V. Bogacheva, N.Yu. Alekseev, O.A. Androsova // System analysis and management in biomedical systems. 2015. P. 14. № 4. P. 815-819.

[6] Application of neural network modeling to support decision-making in the diagnosis of chronic heart failure / E.A. Fursova, E.I. Novikova, O.V. Sudakov // System analysis and management in biomedical systems.2009.T.8.№ 2.C.410-413.

[7] 11. Modeling of severe forms of acute toxic hepatitis / N.Yu. Alekseev, N.Yu. Kuzmenko, O.V. Sudakov // System analysis and management in biomedical systems.2012.T.11.№ 2.C.481-484.

[8] Construction of the model of the adaptation potential of patients with type 2 diabetes and arterial hypertension / O.V. Sudakov, T.P. Kuchkovskaya, A.V. Sviridov, G.M. Panyushkina, E.A. Studenikina // System analysis and management in biomedical systems. 2014. T..13.№ 2. 2. P. 447-452.

[9] Construction of an adaptive algorithm for the treatment of patients with diabetes mellitus and arterial hypertension / O.V. Rodionov, E.N. Korovin, O.V. Sudakov, E.A. Fursova, D.V. Sudakov // System analysis and management in biomedical systems. No. 3.S. 688-690.

[10] Assessment of the risk of recurrent stroke in patients with type 2 diabetes and hypertension / N.A. Gladsky, O.V. Sudakov, E.V. Bogacheva, N.Yu. Alekseev, E.A. Fursova // System analysis and management in biomedical systems. T.15.№ 1.C.C. 123-127.

[11] Construction of a mathematical model for choosing the type of coronary angioplasty in patients with coronary heart disease and diabetes mellitus / A.I. Borodulin, A.V. Sviridov, O.V. Sudakov, E.A. Fursova, N.Yu. Alekseev // Applied information aspects of medicine. 177. № 2.C.56-58.

[12] Esaulenko I.E. Information system for the selection of diagnostic signs of autonomic syndrome. I.E. Esaulenko, N.A. Gladskikh, S.I. Shtankov, M.A. Zheleznyakov // System analysis and management in biomedical systems. 2011. T. 10. № 4. Pp. 864 -869.

[13] Mathematical and theoretical models of diagnostics of vegetative syndromes. ON. Gladskikh, S.N. Shipilov, E.V. Bogacheva, S.I. Shtankov. // Applied information aspects of medicine. 2011. T. 11. № 2. P. 44-45.

[14] Calculation of diagnostic assessments of the effectiveness of the use of Mexidol in the provision of medical care to victims with TBMT in prehospital and early hospital stages / I.E. Esaulenko, V.L. Radushkevich, N.A. Gladskikh, E.V. Bogacheva // System analysis and management in biomedical systems. 2013. T. 12. № 4. P. 896-901.

[15] Gladskikh N.A. Development of methods of classification-prognostic modeling in the system of staffing of territorial health care / dissertation for the degree of candidate of technical sciences / VSTU. Voronezh. 2008, 211 pp.