УДК 681.513.8

NAIVE BAYESIAN CLASSIFIERS FOR PURPOSES OF MEDICAL DIFFERENTIAL DIAGNOSTICS

Nina Kondrashova[1], Vladimir Pavlov[2]

[1] *International Research and Training Center for Information Technologies and Systems of the National Academy of Sciences and Ministry of Education and Science of Ukraine,*
[2] *National Technical University "KPI"*

NKondrashova@ukr.net, vpavlo@bk.ru

Відмінною особливістю даної роботи є ансамблювання наївних Байєсівських класифікаторів в схемі «один проти всіх» і використанні розширеного простору ознак. У первинній вибірці присутні метричні і категорійні змінні. Схема «один проти всіх» із застосуванням інших методів класифікації дає поліпшення на екзамені точності диференціальної діагностики порівняно з єдиним класифікатором, але не у випадку наївних Байєсівських класифікаторів. Отримані результати точності дозволяють порівняти їх з результатами інших методів розв'язання задачі класифікації: таких як МГУА і канонічний дискримінантний аналіз.

Ключові слова: метод групового урахування аргументів (МГУА), наївний Байєсівський класифікатор, "один проти всіх", медична диференційна діагностика.

A distinctive feature of this work is grouping naive Bayesian classifiers in the scheme of "one against all" and using the extended features space. Metric and categorial variables are present in the original sample. The scheme of "one vs. all" with the use of other methods of classification gives an improvement in the accuracy of the differential diagnosis on exam sample compared to a single Bayesian classifier, but not in the case of the Naive Bayesian classifiers. The obtained results allow us to compare methods accuracies with such as GMDH and canonical discriminant analysis in solution of classification problem.

Keywords: Group Method of Data Handling (GMDH), Naive Bayes classifier, "one vs. all", medical differential diagnosis.

Отличительной особенностью данной работы является ансамблирование наивных Байесовских классификаторов в схеме «один против всех» и использовании расширенного пространства признаков. В исходной выборке присутствуют метрические и категориальные переменные. Схема «один против всех» с применением других методов классификации дает улучшение на экзамене точности дифференциальной диагностики по сравнению с единым классификатором, но не в случае наивных Байесовских классификаторов. Полученные результаты точности позволяют сравнить их с результатами других методов решения задачи классификации: таких как МГУА и канонический дискриминантный анализ.

Ключевые слова: метод группового учета аргументов (МГУА), наивный Байесовский классификатор, "один против всех", медицинская дифференциальная диагностика.

Introduction

In order to quickly analyze and make decisions about prescribing drugs a computer programs are created to improve the quality of disease diagnosis. Specialists develop programs for processing of information, which helps the doctor to make a diagnosis, taking into account the individual characteristics of the patient and accumulated knowledge in the subject area. Automated technology of medical assistance should work in a shortage of time and resources for conducting expensive examina-

tions of each patient. Solution to the problem of computer diagnosis of blood diseases makes avoid lengthy and costly researches to establish the diagnosis and expedite the process of treatment and recovery of the patient.

The purpose of diagnosis is to understand what the diagnoses have the patients (to which classes they should be attributed) based on their observed clinical signs. Building classifiers are the main part of the diagnostic system. There are many approaches, methods and algorithms for this purpose.

The most well-known classification methods are a support vector machine (SVM) [1], clustering [2, 3], factor analysis [4], discriminant analysis [5], classification trees, statistical methods and neural networks. The book [6] in Russian includes three issues [3-5]. The SVM method is sensitive to noise and normalizing the data. Solving linear programming problem underlies SVM, and in the case when the classes are linearly inseparable there is no common approach to the automatic selection of nucleus [7]. In addition, it is slow learning [8]. In a cluster analysis union of similar objects in a group can be carried out by various methods. It is known at least eleven methods of cluster analysis, the most famous of which is *k*-means [9]. Factor analysis and the method of *k*-means work only with continuous data; *k*-means also requires pre-specifying the number of classes. Iteration according to the principle of *k*-means is extremely sensitive to poor initial partition, and it becomes even more complicated when the initial approximation is chosen randomly [10].

Features of the original data in our task are:

1) The mixed nature of the input and output variables: some of them are categorial and some metric;

2) The complexity of the introduction of a common metric for the dispersive criterion.

Using statistical methods make the following data requirements:

1. Objects should not be correlated with each other.

2. The distribution of the objects should be close to normal.

3. Objects should satisfy the requirement of stability, which is understood as the lack of influence on their values of random factors.

4. The sample should be homogeneous.

In the paper in view of above remarks for solving the problem of classification we have tried to use the ideas contained in the well-known classification methods of a statistical approach [11]. In contrast to [12] described a common Bayesian classifier into four classes in this paper is used ensemble classifiers constructed on the principle of "one against all." In the construction naive (or raw) Bayesian classifiers an expanded space of features is used, viz the except a categorial is present the metrical variables in the original sample. First, classifiers "are trained" on a sample of U ($|U|$=70) of patients with known diagnoses, characterized by individual dataset. Then the diagnostic system for the patient from exam sample C ($|C|$=10) defines the diagnosis based on him observable clinical symptoms ($U \cap C = \varnothing$).

As the initial data about the disease mild form of coagulopathy and thrombocytopathy were used information about hemorrhagic signs inherent to patients (women aged 19-49 years). Hemorrhagic symptoms such as vaginal bleeding, nosebleeds, etc.

are connected with blood incoagulability. Ten categorial (of attributive) variables, called symptoms, and one metrical (quantitative) variable, viz the age are taken into account.

The set of studied diagnoses is presented by von Willebrand disease, disaggregation thrombocytopathy, coagulopathy and combined pathology of the hemostasis system. Each of the four diagnoses was established for initial sample patients in a clinical laboratory using modern and expensive reagents and using the latest advances in medical technology.

The idea of constructing a naive Bayesian classifier common for all four classes, which was built according to the formula of Bayes-Laplace, on the basis of which the most likely diagnosis was chosen, has already been considered in [11]. Precision was 30% at verification on the examination sample. We take into account that many statistical classification, clustering and recognition methods work on Bayesian decision rules. Bayes method has a number of possibilities and advantages [13]. The loss function of Bayesian strategy is minimal when changing the model parameters [14]. It should be noted that the method of expert estimates which widely used in medicine works on the "coarse" estimates of probability (frequencies). Two expert methods are offered and tested in [15, 16] to solve the problem of medical differential diagnosis.

In this paper, the following possibilities were considered:

1) Approach "one against all" unrealized in [12] for building an ensemble of classifiers, which helped improve the accuracy of classification in [17]; and

2) Biased estimates of the probability (frequency) by "Laplace smoothing" for taking into account the values of signs, which were not observed on the training set, but may appear on the examination sample. Laplace smoothing (not to be confused with *Laplasian smoothing*) is a technique used to smooth categorial data. It is historically known as the sunrise problem [18].

1. Problem statement

We start of the probabilistic nature of the observed sample. Practically the each patient received their disease independent of the others and that only the presence of common features can characterize this or that disease in the patients' specific group with the same disease.

We calculated the diagnosis probability according to the Bayes-Laplace formula of the form [19]

$$p(D_j \mid \mathbf{X}_s) = \frac{p(\mathbf{X}_s, D_j)}{p(\mathbf{X}_s)} = \frac{p(D_j)p(\mathbf{X}_s \mid D_j)}{\sum_{j=1}^{k} p(D_j)p(\mathbf{X}_s \mid D_j)} = \frac{p(D_j)p\left((\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_s) \middle| D_j\right)}{\sum_{j=1}^{k} p(D_j)p(\mathbf{X}_s \mid D_j)},$$

where features values $\mathbf{x}_i$ of the individual patient, for which are estimated the diagnosis probabilities, belong to the area of integer values $\Omega_{\mathbf{X}}$, $\mathbf{x}_i \in \Omega_{\mathbf{X}}$; a sign of age ($\mathbf{x}_1$) takes the values from the range: $\Omega_{\mathbf{x}_1} = \{19,20,...,49\}$, and most of signs are:

$\Omega_{\mathbf{x}_i \in \mathbf{X} \setminus \mathbf{x}_1} \in \{+1,0,-1\}, i = \overline{2,m}$; "dictionary" of clinical signs values: $\Omega_{\mathbf{X}} = \bigcup_i \Omega_{\mathbf{x}_i}$; where $|\Omega_{\mathbf{X}}|$ is the total number values of all the clinical signs in the "dictionary". Value of feature “+1” indicates its presence in the disease, “-1” is the absence of it; “0” means that there were no conditions for its appearance (for example, in some patients never removed the tooth and therefore not been possible to fix the bleeding while teeth removing);

$p(D_j \mid \mathbf{X}_s)$ is the conditional probability that a patient with a set of clinical signs $\mathbf{X}_s = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_s]$ belongs to class (diagnosis) D_j, $j = \overline{1,k}$, that's this, we need to calculate, where k is total number of diagnoses, equal to four, m is the total number of clinical signs, equal to eleven, $D_j \in \Omega_D$, $|\Omega_D| = k$;

$p(D_j)$ is unconditional probability the patient of class D_j in the whole sample;

$p\left((\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_s) \middle| D_j\right)$ is conditional probability of a patient having a specific set of symptoms $[\mathbf{x}_1 = x_{1j}, \mathbf{x}_2 = x_{2j}, ..., \mathbf{x}_s = x_{sj}]$ among all patients of class D_j;

$p(\mathbf{X}_s)$ is unconditional probability of a patient with a set of features $\mathbf{X}_s = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_s]$ in the whole sample.

We need the most likely diagnosis, i.e. it is necessary to calculate the probabilities for all classes and to choose the class that has the highest probability. Used for this purpose a Bayesian classifier based on a posteriori estimation to determine the most likely class

$$D^* = \arg\max_{j=\overline{1,k}} \frac{p(D_j) p(\mathbf{X}_s \mid D_j)}{p(\mathbf{X}_s)}$$

Denominator (the probability a specific patient) is a constant and can not affect the ranking of the classes, so we can ignore it

$$D^* = \arg\max_{j=\overline{1,k}} p(D_j) p(\mathbf{X}_s \mid D_j) \qquad (1)$$

As the volume of data is not sufficient ($|U|$=70, m=11), to take into account the interdependence of signs is not possible (not possible to calculate the conditional probabilities of all combinations of features when other combinations of signs present and specified diagnosis, i.e. $p(\mathbf{x}_1, ..., \mathbf{x}_v \mid \mathbf{x}_q, ..., \mathbf{x}_s, D_j)$).

Naive Bayesian classifier operates with a set of clinical features that conditionally do not depend on each other. Based on this assumption, conditional probability the patient to have a set of symptoms can be approximated by the product of the conditional probabilities of all clinical signs available at the patient

$$p(\mathbf{X}_s \mid D_j) \approx p(\mathbf{x}_1 = x_{1j} \mid D_j) \cdot \ldots \cdot p(\mathbf{x}_s = x_{sj} \mid D_j) = \prod_{\forall \mathbf{x}_i \in \Omega_{\mathbf{X}}} p(\mathbf{x}_i \mid D_j),\ j = \overline{1,k}. \quad (2)$$

The most likely diagnosis is calculated using posteriori probabilities when substituted (2) into (1) by the formula naive Bayesian classifier as:

$$D^* = \arg\max_{j=1,k}[p(D_j) \prod_{\forall \mathbf{x}_i \in \Omega_{\mathbf{X}}} p(\mathbf{x}_i \mid D_j)]. \quad (3)$$

If a patient is necessary to consider a sufficiently large number of clinical symptoms, will have to multiply the large number of very small numbers. In order to avoid arithmetic overflow below are commonly used property the product of the logarithms. Since the logarithm is a monotonic function, its application to both parts of expressions of the form (2) will only change their numerical values, and not the parameters at which it has the maximum (3). In the case when the logarithm of the number near zero, it is negative, but in absolute value significantly greater than the initial number. It makes the logarithmic probability values more suitable for analysis. Therefore, we rewrite our formula using logarithms. The base of the logarithm in this case does not matter. We will use the natural logarithm.

$$D^* = \arg\max_{j=1,k}[\ln p(D_j) + \sum_{i=1}^{|\Omega_{\mathbf{X}}|} \ln p(\mathbf{x}_i \mid D_j)]. \quad (4)$$

The assessment of probabilities $p(D_j)$ and $p(\mathbf{x}_i \mid D_j)$ are implemented on the learning sample. Probability of class can be estimated as $p(D_j) = \frac{n_{D_j}}{n_U}$, where n_U is the total number of patients in learning sample of 70; is the number of patients diagnosed. Estimating the probability of a unique value of the clinical feature in the class is held on the multinomial Bayesian model:

$$p(\mathbf{x}_i \mid D_j) = \frac{w_{ij}}{\sum_{\ell \in \Omega_{\mathbf{X}}} w_{\ell j}}, \quad (5)$$

where w_{ij} are the number times that a unique value of *i*-th clinical sign found in patients D_j class ; $\Omega_{\mathbf{X}}$ is the set of all unique values of clinical signs ("dictionary").

Other words, the numerator of the formula (5) shows how many times, *some* unique value of clinical sign found in a particular class of patients (including repetitions), and the denominator is the *total number* (with replays) of unique values of clinical signs in all patients given class.

If we meet the recognition stage (on examination sample) clinical sign, the value of which did not meet during learning (e.g., a certain value of patient's age $\mathbf{x}_1$), then this value is $w_{1j} = 0$, and hence probability $p(\mathbf{x}_1 \mid D_j)$ will be zero. This will lead to

what the patient with this clinical feature could not be recognized because it will have zero probability by to all classes

A typical solution to the problem of unknown values of clinical signs is additive smoothing (Laplace smoothing) [20].

$$p(\mathbf{x}_i \mid D_j) = \frac{w_{ij} + \alpha}{\sum_{\ell \in \Omega_{\mathbf{X}}} (w_{\ell j} + \alpha)} = \frac{w_{ij} + \alpha}{\alpha \left|\Omega_{\mathbf{X}}\right| + \sum_{\ell \in \Omega_{\mathbf{X}}} w_{\ell j}}, \qquad (6)$$

where $\alpha > 0$ is the smoothing parameter (α=0 corresponds to no smoothing).

Using the Laplace's rule of succession, some authors have argued that α should be 1 (in which case the term add-one smoothing [21, 22] is also used), though in practice a smaller value is typically chosen. We believe if α =1, then an essence (6) lies in the fact that we met the value of clinical signs at one time more and should add 1 to its frequency. Thus, the value of one of the clinical signs that we have not met during training model, gets, though small, but not zero probability. Naturally, this approach shifts the assessment of probabilities in the direction of less probable results. The greater number of values shall take signs and the greater the number of features, the less displacements of probability. Substituting the estimates (6) into (4), we obtain the final formula that will be Bayesian classification:

$$D^* = \arg \max_{j=\overline{1,k}} D_j = \arg \max_{D_j \in \Omega_D} [\ln \frac{n_{D_j}}{n_U} + \sum_{i=1}^{\left|\Omega_{\mathbf{X}}\right|} \ln \frac{w_{ij} + 1}{\left|\Omega_{\mathbf{X}}\right| + \sum_{\ell \in \Omega_{\mathbf{X}}} w_{\ell j}}].$$

2. Classification problem solution

To implement the Bayesian classifier, we need a learning sample, in which are put correspondences between the patients and the classes to which they belong. Then we need to collect the following statistics from the sample that will be used on the stage of classification and recognition:

- The relative frequencies of the classes in the sample, i.e. how often patients of a particular class appeared;
- The total number of clinical signs in patients of each class;
- The relative frequencies of the clinical signs within each class;
- Number of unique clinical signs in the sample.

The totality of this information, we will call the original data for constructing a classifier. Then, at step classification is necessary for each class calculate the value of the following expression

$$q_j = \ln \frac{n_{D_j}}{n_U} + \sum_{i \in Q} \ln \frac{w_{ij} + 1}{\left|\Omega_{\mathbf{X}}\right| + L_j}. \qquad (7)$$

and select the class D_j with the maximum value q_j. In this formula, all terms except the already mentioned plurality of values of clinical symptoms Q of the patient by

which they are classified (including repeats for counting w_{ij}) and the total number of values L_j clinical features of patients D_j class in learning sample.

Now, in order to say how likely a patient with a set of features $[\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_s]$ is diagnosed D_j, need for the values of logarithms of the probability to return to the values of the probabilities themselves. For this is necessary to do the inverse operation raising to the power and normalize the probability to get one in the sum form:

$$p(D_j \mid \mathbf{X}_s) = \frac{e^{q_j}}{\sum_{j \in \Omega_D} e^{q_j}}, \tag{8}$$

where q_j is assessment of the logarithm of D_j class appearance.

To find $p(D_j \mid \mathbf{X}_s)$, for each of the four classifiers is necessary to calculate the probability of hitting the patient to the class D_j, and the probability of a patient misses to class D_j (i.e. the probability of hitting the one of the other three classes). Thus, a classification scheme "one against all" is implemented.

D_1 is von Willebrand disease (VWD), D_2 is coagulopathy (CP), D_3 is disaggregation thrombocytopathy (DT); D_4 is the combined pathology of the hemostasis system (CPHS).

Observations distribution in the classes and the samples is following.

Disease VWD: there are 20 patients in learning sample U and 4 patients in examination sample C;

Disease CP: there are 15 patients in learning sample U and 2 patients in examination sample C.

Disease DT: there are 27 patients in learning sample U and 4 patients in examination sample C.

Disease CPHS: there are 8 patients in learning sample U and 0 patients in examination sample C.

Patient age encoded, so the presence in the initial sample patient the age $\mathbf{x}_1$ with value, for example, equal to 29, as $x_1(29)$. Hemorrhagic signs are encoded in the Latin alphabet from x_2 to x_{11} and its values in parentheses. For example, $x_2(-1)$ denote "in the absence of the patient's juvenile uterine bleeding", $x_2(0)$ "in the absence of the conditions for the appearance of clinical sign", $x_2(+1)$ means "in the presence of this sign". Values clinical signs are encoded. All designations used for clinical signs are shown in Table 1.

Table 1

Designations for clinical signs

Clinical sign	Designation
Patient age	$x_1(19) – x_1(49)$
Juvenile uterine bleeding	$x_2(-1), x_2(+1)$

Continuation of Table 1

Nosebleeds	$x_3(-1), x_3(+1)$
Bleeding gums	$x_4(-1), x_4(+1)$
Bleeding after teeth extraction	$x_5(-1), x_5(0), x_5(+1)$
Intra-and postoperative bleeding	$x_6(-1), x_6(0), x_6(+1)$
Post-traumatic hematoma	$x_7(-1), x_7(+1)$
Bleeding from superficial wounds	$x_8(-1), x_8(+1)$
Prolonged not wound healing	$x_9(-1), x_9(+1)$
After injection hematoma	$x_{10}(-1), x_{10}(+1)$
Postpartum bleeding	$x_{11}(-1), x_{11}(0), x_{11}(+1)$

Each patient has a unique set of clinical signs, so for each patient in the formula (7) the naive Bayesian classifier has its own set of frequencies w_{ij}. The same are indicators n_{D_j}, L_j, involved in the formula (7) in the limits of a given class, and $|\Omega_{\mathbf{X}}|$, n_U are common to all classes.

3. Results of classifiers synthesis

In tables 2, 3, 4 are statistics needed to compute the naïve Bayesian classifiers according to formula (7).

Table 2

The frequency of bleeding symptoms

Feature	w_{ij}			
	VWD	CP	DT	CPHS
$x_2(-1)$	0	2	4	1
$x_2(+1)$	20	13	23	7
$x_3(-1)$	2	5	3	4
$x_3(+1)$	18	10	24	4
$x_4(-1)$	6	6	12	4
$x_4(+1)$	14	9	15	4
$x_5(-1)$	1	3	3	1
$x_5(0)$	8	1	9	3
$x_5(+1)$	11	11	15	4
$x_6(-1)$	0	2	2	0
$x_6(0)$	15	6	18	3
$x_6(+1)$	5	7	7	5
$x_7(-1)$	10	5	10	3
$x_7(+1)$	10	10	17	5

Feature	w_{ij}			
	VWD	CP	DT	CPHS
x_8(-1)	10	7	12	2
x_8(+1)	10	8	15	6
x_9(-1)	10	10	18	5
x_9(+1)	10	5	9	3
x_{10}(-1)	19	13	26	7
x_{10}(+1)	1	2	1	1
x_{11}(-1)	4	5	11	1
x_{11}(0)	10	8	9	5
x_{11}(+1)	6	2	7	2

Table 3

Frequencies patients' age

Age	W_{ij}			
	VWD	CP	DT	CPHS
x_1(19)	2	2	2	0
x_1(20)	1	4	0	1
x_1(21)	1	0	3	1
x_1(22)	0	0	2	1
x_1(23)	0	2	1	1
x_1(24)	2	0	0	0
x_1(25)	1	0	1	1
x_1(26)	0	0	0	0
x_1(27)	1	0	1	1
x_1(28)	1	1	2	0
x_1(29)	1	0	1	0
x_1(30)	0	0	2	0
x_1(31)	1	0	0	0
x_1(32)	0	1	0	0
x_1(33)	0	0	3	0
x_1(34)	2	0	0	0
x_1(35)	0	1	0	0
x_1(36)	2	0	0	0
x_1(37)	0	0	1	0
x_1(38)	1	1	1	0
x_1(39)	0	0	2	1
x_1(40)	0	0	0	0
x_1(41)	0	0	1	0
x_1(42)	1	0	0	0
x_1(43)	0	1	0	0
x_1(44)	0	0	1	0
x_1(45)	1	0	0	0

Age	W_{ij}			
	VWD	CP	DT	CPHS
$x_1(46)$	1	0	0	0
$x_1(47)$	1	1	0	0
$x_1(48)$	0	0	0	0
$x_1(49)$	0	1	3	1

The probability of falling into the D_j -th class ("his" class) and the hitting into any of the other three classes (i.e., the probability of the miss to D_j -th class) is necessary to calculate using these data, for each patient. Then the probability that a patient belongs D_j -th class receives by formula (8). All these steps should be repeated for all four classes, and the comparing obtained probability, choose a maximum of them. The diagnosis (class) with the maximum probability will be diagnosed with a particular patient.

The tables 3 and 4 shows that "dictionary" symptoms $|\Omega_{\mathbf{X}}|$=54-$|\Omega_{\neg\mathbf{X}}|$=51, where 54 is the sum of all the unique signs of the first column in this group of patients. The set $\Omega_{\neg\mathbf{X}}$ is a set of attribute values that can appear in a patient in the examination sample. These are patients with age 26, 40 and 48, and the value of attributes will be $x_1(26)$, $x_1(40)$ and $x_1(48)$ respectively (the cells are denoted).

Data for "their" class will look follows:

	VWD	CP	DT	CPHS	Total, n_U
frequency classes, n_{D_j}	20	15	27	8	70
the total features number, L_j	220	165	297	88	

Data for not "their" class will have the form:

	$\Omega_D \setminus$ VWD	$\Omega_D \setminus$ CP	$\Omega_D \setminus$ DT	$\Omega_D \setminus$ CPHS	n_U
frequency classes, n_{D_j}	50	55	43	62	70
the total features number, L_j	550	605	473	682	

Accuracy was defined as the proportion of correctly classified objects to all objects in the class in each of the samples. For example, on the examination sample was

correctly recognized two of the four patients with a diagnosis of VWD, one of the four with a diagnosis of DT and neither of the two with a diagnosis of CP.

Table 4

Statistics for models calculation of (7) is:

		VWD	CP	DT	CPHS
"Their" class	n_{Dj}	20	15	27	8
	n_U	70	70	70	70
	$\|\Omega_{\mathbf{X}}\|$	51	51	51	51
	L_j	220	165	297	88
Another classes	n_{Dj}	50	55	43	62
	n_U	70	70	70	70
	$\|\Omega_{\mathbf{X}}\|$	51	51	51	51
	L_j	550	605	473	682

Table 5 shows the characteristics of the naive Bayes classifiers.

Table 5

Characteristics of the obtained classifiers

		Accuracy, %			
		VWD	CP	DT	CPHS
Whole sample	$U+C$	83,33	52,94	77,42	37,5
Learning	U	90	60	85,19	37,5
Examination	C	50	0	25	-

On average, the accuracy of the exam is: three from ten, i.e. 30% objects were correctly recognized. It is obvious that such a low accuracy is due to non-compliance with the demands to data for the application of statistical methods in this problem.

4. Conclusions

Solution of the diagnostic problem accomplished using a naive Bayesian classifier. This type of classification was chosen in view of the fact that it is the basis of many methods of statistical classification and can operate on a "one against all." This approach in conjunction with other methods gives improved accuracy of the differential diagnosis as compared with a single classifier for all classes. However, in the case with a naive Bayesian classifier, it does not allow to increase its precision at verification of the new (examination) data. The results obtained in the classification of the total sample (62.8%), teaching (68.17%) and examination (30%) allow us to compare these results with ones other approaches for solving the problem of classification: such as the method of expert estimates, GMDH and canonical discriminant analysis.

References

1. Cortes C., Vapnik V. Support-vector networks// Machine Learning . – 1995. – V. 20, Issue 3. – pp. 273-297.
2. Tryon Robert C. Cluster Analysis: Correlation Profile and Orthometric (factor) Analysis for the Isolation of Unities in Mind and Personality// Edwards Brothers. –1939.
3. Mark S. Aldenderfer, Roger K- Blashfield. Cluster Analysis //Quantitative Applications in the Social Sciences. Second Printing. : «Sage Publications, Inc.», 1985.
4. Jae-On Kim, Charles W. Mueller. Factor Analysis: Statistical Methods and Practical Issues //Quantitative Applications in the Social Sciences. Eleventh Printing. : Sage Publications, Inc. 1986.
5. William R. Klecka. Discriminant Analysis // Quantitative Applications in the Social Sciences. Seventh Printing. : Sage Publications, Inc. 1986.
6. Factor, Discriminant and Cluster analyses: ed. Yenyukov I.S. – M.: Finance and Statistics, – 1989. – 215. (in russian)
7. www.ccas.ru/voron/download/SVM.pdf – K.V. Vorontsov, Lectures on the method of support vector machines. – 2007.
8. http://uk.wikipedia.org/wiki/Support_vector_machines
9. Zhuravlev, Yu.I. Recognizing. Mathematical methods. Software system. Practical applications / Yu. I. Zhuravlev, V. V. Ryazanov, O.V. Sen'ko . - M .: Phasis, 2006. (in russian)
10. Blashfield, R.K. and M. Aldenderfer "Computer programs for performing iterative partitioning cluster analysis." Applied Psychological Measurement. – 1978. – N 2. pp.533-541.
11. Kondrashova, N. V., Pavlov, V. A., and Pavlov, A. V., Problem solution of medical diagnosis by linear discriminant analysis and GMDH, USiM, No. 2, 2013, pp. 79-88. (in russian)
12. Kondrashova, N. V. GMDH and probabilistic methods in construction of classifiers for medical differential diagnosis // Inductive modeling of complex systems. Col. of Sc. works. // Kyiv: IRTC IT&S, 2012. – Isse.4. – pp.102-113. (in russian)
13. Hand, DJ, & Yu, K. Idiot's Bayes – not so stupid after all?// International Statistical Review. – 2001. – Vol 69, – part 3. – pp. 385– 99.
14. Schlesinger M.I., Vodolazskiy E.V. Nearly Optimal Statistical Recognition and Learning)// Proc. of 4th International conference on Inductive Modelling (ICIM'2013). – Kyiv : IRTC ITS in Kyiv, 16-20 September, – 2013. – pp. 51-58.

15. Kondrashova N.V., Tomilin V.V. Solution of diagnosis problem diseases mild form coagulopathy and thrombocytopathy based on methods of expert ratings. System Technology, Regional Interuniversity collection of scientific papers, Issue. 6(71), – Dnipropetrovsk: 2010, pp.104-114. (in russian)

16. Kondrashova N.V., Pavlov A.V. Inductive and Probabilistic Approaches Comparison in Mild Cases Pathology Hemostasis Diseases Diagnosis // Proc. of 4th International Workshop on Inductive Modeling, IWIM 20011. Kyiv-Zhukyn, Ukraine. – C. 38-45. .

17. Kondrashova N.V., Pavlov A.V. Solution of classification problem with help of GMDH generalized relaxation iterative algorithm// International Journal of Applied Mathematics & Computer Science. (in preparation).

18. Chung, K. L. & AitSahlia, F. Elementary probability theory: with stochastic processes and an introduction to mathematical finance. Springer. – 2003. – pp.129–130.

19. Gnedenko B.V. Essays on the history of probability. 69 pages. (Gnedenko B.V. Course in Probability Theory: Tutorial – revised and expanded – 6th ed. – M: Nauka. – Chap. Eds. Physicomathematical. Lit. – 1988. – 448.). (in russian)

20. Manning, C.D., Raghavan, P. and Schütze, M. Introduction to Information Retrieval. Cambridge University Press. – 2008. – p. 260.

21. Russell, Stuart and Norvig, Peter. Artificial Intelligence: A Modern Approach (2nd ed.). Pearson Education, Inc. – 2010. – p.863.

22. Jurafsky, Daniel and Martin, James H. *Speech and Language Processing* (2nd ed.). Prentice Hall. – June 2008. – p.132.