Informative predict system of educational activity of the first-year students*

Olga Marukhina[1], Irina Rogozhkina[2],
Margarita Bobrova[3]
[1,3]Dept. of Optimization Control Systems, [2]Dept. of
Engineering pedagogy
[1,2,3]National Research Tomsk Polytechnic University
Tomsk, Russia

Natalia Kozlova[4], Elena Berestneva[5]
[4]Dept. of Genetic and Clinical Psychology, [5]Education and
Research Center of Computer Science and Technology
[4,5]National Research Tomsk State University
Tomsk, Russia

Abstract—**The article shows the importance of educational activities and its prediction for the first-year students. Some computational methods of analysis and recognition of states were considered. The method of non-uniform sequential procedure of recognition of states was considered in detail as well. The developed software was applied to solve specific problems of the area of psychological research and diagnostics. Some universal algorithms were used in the most informative signs of physical nature and calculated with help of mathematics. The results of psychological tests can be performed as starting material for processing, medical data as well and data from other areas of research. It is possible to obtain a diagnostic assessment of projected state with the accumulation of a sufficient amount of information required for a certain type.**

Keywords—educational activities, students adaptation, motivation, ratio, quality improvement, informational content.

I. INTRODUCTION

There is a human with all his individual, distinctive features in the basis of any human activity. Educational process, learning process, in particular, in higher education – there is, first of all, the most complicated psycho-physiological process, experienced by an individual of studies in the process of his/her educational and cognitive activity in educational information acquirement of knowledge, knowledge formation, skills development, experience raise, disclosure of creative opportunities. The great teacher Ushinsky K.D. wrote: "If pedagogy wants to educate a man in all aspects, it should, first of all, learn him in all aspects, as well". Hence, there is a conclusion: if we want to teach a man and give him some high qualified education, we should, first of all, to have a clear understanding of objective laws and the mechanism possibility of mental processes of his cognitive activity.

The key objective at any university was and still exists is the quality improvement of students' process. The solution of this problem is a complicated process that requires not only teachers involvement, psychologists, deans and staff of the departments, but also experts in the field of information technology. Informational technology not only help to avoid many difficulties, but also improve the quality of education, allow to prompt students to active learning activities, allow to individualize the learning process, use studying time more effectively, and so on, that creates some conditions for improvement the learning process effectiveness in higher education.

It is very important to determine the level of development of psychological characteristics, cognitive and intellectual capacities of students as early as possible, namely, just in the first year in order to have a successful professional career in the near future. The research interest is in the study of the first-year students adaptation to educational activity in high school. Firstly, on the 1-st phase the Uniform state exam (USE) exam results are analyzed and show the level of an individual development. Secondly, conduct a quick diagnosis by a psychologist that helps to determine a social type of a student and his motivation to educational activity. The next stage - the first test point, its results indicate and show students who came to this profession consciously, with a real motivation to learn in this area. The second test point indicates the position of a fixed motivation, the intention to learn and get exactly this profession.

There is a big variety of motivations and its are really unlimited that is why it is impossible to describe a specific motivation that encourages the concrete individual to act, i.e. there is not the content that capable to motivate a person to study [1,2]. Therefore, in practice, in the process of study the motivational sphere, usually one allocates a limited set of features and the range of values and data that depend on the tasks that faces a researcher. By increasing the accuracy and meaningfulness of the required characteristics of a person increases the amount of tests, the complexity of the processing techniques, time spent on testing and analysis of the results. Therefore, it becomes an actual problem of information systems development for the automated collection and diagnostic analysis of psychological data. This system, implemented in the form of a computer program or software package, should be responsible for the collection, accumulation, storage, processing its according to the certain procedures and the issuance of diagnostic information.

As for pedagogy and psychology as well problem of motivation to educational activity and reasons in learning activities is significant. The way for effective learning lies through the understanding of a learner's motivation. If one has some diagnostic information about what reasons induce a student for activity you can try to develop an effective system of forms and methods of educational management.

When a psychologist studies the motivational sphere, then the main idea under the applied psychological analysis is to obtain a clear picture what motivates the given person or a group of people. This is a very difficult task, because the most human desires are dued to the different aspirations. For example, the note as a reason of learning activity can embody and need of a teacher's approval, and need to be at the level of his own self-esteem, and desire to gain class-mates' authority, and desire to facilitate the entry into university, etc. Obviously, you can not specify a particular reason for committing a single act, but you can select a group of symptoms, in a certain extent that are responsible for a person's behavior as a whole.

One of the main authors' objective is to get numerical estimates of motivational characteristics and values while testing and construct the so-called "motivational profile of a person". To determine the motivational signs that were mentioned above a test was taken, based on a questionnaire by L.A. Groisman, a detailed description given in [3].

The second applied mathematical task is to predict progress. The following kinds of schemes are considered:

- «revealed desire to receive excellent marks";
- «no particular desire to study well";
- «this information is not sufficient for a particular classification."

The first two types are scales derived from the study of explicit "excellent students (study well)" or "poor students (study very badly)." The last scale is given when it is impossible to attribute a motivational profile of the given student to one of the classes: "poor students" or "excellent students". Typically, it is obtained while testing the "average" student. But here, too, though with less precision one can predict a good or satisfactory progress in the near future. Research task was to obtain a main rule for progress diagnostics. The following tasks required to solve:

1. With help of test "Study of motivational profile" to test the first-year students of Cybernetics Institute at TPU.
2. To highlight two classes based on the results of the first semester of freshmen: "excellent students" - students with marks "4" and "5" (max. marks for exam) for the exam session and "poor students" - students who did not pass a session (failed exams), or not admitted to it.
3. To calculate for each student of these groups the numerical values of the motivational signs.

4. To get the procedure that helps to determine the given student to one of these groups by means of analyzing the significance of the differences between these scales for classes "excellent students" and "poor students".

Initial data indicators of motivational sphere for the first-year students: good successful students (group A_1); bad successful students (group A_2). The study set includes the indicators of motivational sphere of 100 students of the Institute of Cybernetics at TPU. It is assumed the provisional recognition to what group (class) each student belongs.

The significance of differences in rates of motivational sphere in these groups was evaluated and constructed the main rule to predict students' progress on the basis of these indicators. The control sample of 50 people was set up to assess the effectiveness of the constructed rule.

On the basis of computer-based testing matrix of responses generated automatically results. The program also provides for the possibility of forming the matrix responses on the results of blank test. The total diagnostic evaluation are actually seven motivational scales constituting motivational personality profile, and two scales of emotional behavior, emotional components of the profile (CP). Turning emotional profile in the overall profile of the individual motivational dictated in scientific terms of internal motivation and common emotions, and practical - a significant diagnostic information.

The matrix of responses generated automatically based on the results of computer testing (Fig.1). Also possible to form the matrix by the results of the responses provided in the blank test program.
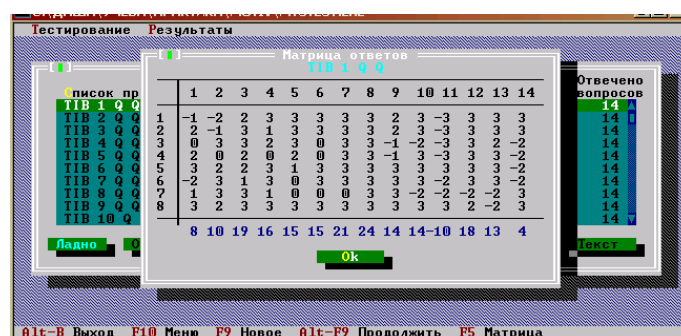

Fig.1. Kind of "answers Matrix" dialog box

III. INFORMATIONAL CONTENT DETERMINATION OF A SET OF SIGNS. HETEROGENEOUS CONSISTENT RECOGNITION PROCEDURE.

Informational content of sign depends entirely on how much it helps to differentiate the state. The sign will be more informative, the greater distance between the realizations of random variables. The most widespread information measure for such distances is by Kullback [4].

In 1948 was proposed a logarithmic measure to measure the amount of information by N.Wiener and K.Shannon independently of each other they proposed this formula that was recognized as quantitative measures of information [4]. There is one among similar measures that was introduced by Jeffreus in 1964 and was studied in detail as a measure of Kullback information measure $J(1,2)$ discrepancies between statistical distributions 1 and 2. For discrete distributions, this formula is as follows:

$$J(x_i/A_1, x_i/A_2) = \sum_j \lg \frac{P(x_{ij}/A_1)}{P(x_{ij}/A_2)} \left[P(x_{ij}/A_1) - P(x_{ij}/A_2) \right], \quad (1)$$

where A_1, A_2 – states classes; i – (sign) number sign; j – number of the i-th range sign; $P(x_{ij}/A_k)$ – probability of object hitting belonging to A_k class in the j range sign.

This criterion makes it possible to draw conclusions about the differences of empirical images without special restrictions on the distribution of random variables forming an empirical way.

Symptoms, which are dominated by ranges of low informative, will have low informative value itself, i.e. will give with close frequency "right" and "wrong" diagnostic coefficients. That is why it is advisable that the signs in the diagnostic table were arranged in descending order of informativeness. And under this condition the speed of achievement the proper cutoff will be on the average the highest and the number of errors – the smallest.

In order to find out what contribution in average makes this given range of sign in the process of approximation to the correct diagnostic cutoff – and this is the information content itself of J (x), it is necessary to know not only the likelihood ratio (or its logarithms), but also take into account the probability of hitting in this range people who belong to classes A_1 or A_2.

Some probability will determine the number of diagnostic coefficients, which will zoom in response to the threshold, the right for the majority of the observations of this range. The likelihood of other factors will determine the number of diagnostic coefficients, which will remove the correct answer for the majority threshold.

All this will be taken into account, if the value $DK(x_i)$, i.e. value of logarithms of likelihood ratio, multiplied by the half – difference of "hit probability" in the given range of people who belong to classes A_1 or A_2, i.e., at the half of the same probabilities, that formed a likelihood ratio:

$$J(x_{ij}) = 10 \lg \frac{P_1(x_{ij}/A_1)}{P_2(x_{ij}/A_2)} 0.5(P_1(x_{ij}/A_1) - P_2(x_{ij}/A_2)), \quad (2)$$

Thus, the value $J(x_{ij})$, being always positive, will reflect the absolute value (module) of the contribution of this range in any approximation to the correct diagnostic threshold. It is necessary to summarize the information content of the ranges to calculate the information content of the whole sign:

$$J(x_{ij}) = 10 \sum_j \lg \frac{P_1(x_{ij}/A_1)}{P_2(x_{ij}/A_2)} 0.5(P_1(x_{ij}/A_1) - P_2(x_{ij}/A_2)) \quad (3)$$

The diagnostics problem decision have been built by the authors through the use of non-uniform pattern recognition procedures (NSP) [3, 6]. Measure of differences (1) is used in this procedure to rank the signs on its differential informativeness. Being one-dimensional, NSP does not require the use of complex multivariate statistical methods. It also does not require knowledge of the laws that empirical distributions are submitted, and is suitable for any form of distributions. In this sense, it can be evaluated as non-parametric.

As a result of NSP application one will get one of three responses: "belonging to the A_1 class", "belonging to the A_2 class", "there is insufficient information to make a decision with the planned level of reliability» – vague answer, often demonstrates not only a lack of information, but also the presence of an intermediate state between classes A_1 and A_2. One should compare the product of the ratio of signs probabilities with diagnostic threshold for one of these calculations. It is necessary to note that when one makes a diagnosis is possible to have errors of two kinds: the object of A_1 class can be attributed to A_2 class (the first kind of error, α). The opposite error is possible as well when an object from A_2 class assumed as A_1 class (the second kind of error, β).

Calculation of diagnostic factors was conducted by A.Vald's method according to the simplified Bayes's formula [4, 7]. Formula decision consistent with the diagnostic procedure in the case of diagnostic coefficients will take the form of the inequality at the correct of which the procedure is not interrupted:

$$10 \lg \frac{\alpha}{1-\beta} < Dk(x_1) + Dk(x_2) + ... + Dk(x_q) < 10 \lg \frac{1-\alpha}{\beta}, \quad (4)$$

where $Dk(x_i) = 10 \lg \frac{P(x_i/A_1)}{P(x_i/A_2)}$ – a diagnostic coefficient

(is positive in the case of the probability predominance of getting into A_1 class, located in the numerator, and negative in the case of the probability predominance of getting into A_2 class)

As soon as the inequality (2) is violated, the procedure is interrupted and one is taken out from the diagnostic decisions.

Indicators of motivational sphere are considered to be independent of each other and we consider each component as

a separate sign. Each sign was divided into 6 ranges (intervals): 1) an extremely high level of exponent (index), 2) a high level, 3) medium (positive) level, 4) medium (negative) level 5) a low level, 6) a very low level of exponent.

IV. Software implementation

The general scheme of the developed system shown in Fig. 2 [5]. In this Termwork module estimated the difference value of samples of each motivational sign according to the formula of Kullback, and calculates the coefficients of diagnostic and prepared diagnostic table. There is a student's progress prediction that is built on the basis of the main rule in the Diagnosis module.

The developed software is easy to use, because it has a user-friendly interface and complete reference system. The program contains a context-oriented tip, that facilitates the work with it significantly. The developed program is versatile to work with any kind of data containing any number of subclasses (state groups), any number of attributes (signs). All these possibilities are realized in the menu item "Database". Some edit operations and view of already exist databases were implemented.

Item of calculating contains a sub-calculations of sign informational content and also a non-uniform implementation of the algorithm consistent recognition procedure. In subparagraph "Initial sample test" heterogeneous consistent recognition procedure works with the educational sample (selection). This makes it possible to assess the quality of the procedure. In sub-paragraph "Diagnosis" on the basis of constructed calculation one can carry out the assignment of a testing student to one of the specified classes.
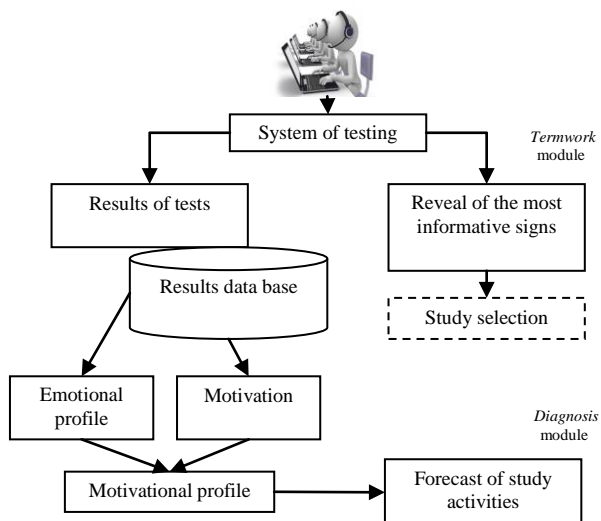

Fig.2. General scheme of the forecast system of educational activities

V. Conclusion.

As a result of the research some diagnostic coefficients were obtained and informational content of signs was calculated. The main rule was constructed on the basis of these data, which allows to predict the success of educational activity of the students of the first course. The work program on the control sample showed a fairly high level of recognition. The volume of educational sample was 194 people in accordance, the control -100 people. The results of the non-uniform sequence on the control sample recognition procedures at $\alpha = 0.05$ and $\beta = 0.1$ shown in Table 1.

TABLE I. WORK RESULTS OF NSP

% recognized correctly	% recognized incorrectly	% non-determined answers
78	15	7

Hence, the obtained main rule can be used to predict the students' progress on motivational indicators, removed as a result of testing.

Refer to the scientists' opinion who are studying the impact of motivational indicators of learning success, motivation to learn and study - it is the fundamental basis of the formation of the cognitive activity of a student, which contributes to the implementation of the educational activities by students from the awareness of the importance and necessity of this process, the understanding of its meaning and significance for further self-development, self-realization in professional life.

REFERENCES

[1] Mokina, E. Expert estimates in the informational support system of the university strategic plan// 8th Korea—Russia International Symposium on Science and Technology — Proceedings: KORUS 2004, — Vol. 3, — pp. 248—251.

[2] Genkin A.A., Medvedev V.I. Prediction of psycho—physiological states. Methodological issues and algorithmization. — L .: Nauka, 1973. — p. 96—110.

[3] Groisman A.L. Clinical psychology, psychosomatic medicine and psychoprophylaxis / A.L. Groisman. — M: Master-Press, — 2002. — p. 451.

[4] Gubler E.V. Computational methods of detection of pathological processes / E.V. Gubler. — L .: Medicine, 1970. — p. 319 .

[5] Marukhina O.V. Algorithms for data processing tasks in evaluating the quality of teaching the high school students on the basis of expert and statistical methods: theses of technical science candidate: 05.13.01. — Tomsk, 2003. – p.165.

[6] Berestneva, O.G., Pekker J.S. Simulation and evaluation of biological systems adaptive capabilities // Proceedings of 2014 International Conference on Mechanical Engineering, Automation and Control Systems, MEACS 2014, 15 December 2014, Article number 6986860.

[7] Berestneva O.G., Marukhina O.V., Ivankina L.I., Shukharev S.O. Modelling of Adaptation Strategies for Different Entities // The European Proceedings of Social & Behavioural Sciences (EpSBS). — 2016. — Vol. 7 : Lifelong Wellbeing in the World (WELLSO 2015), — pp. 252-258.