UDC 378.147 DOI: https://doi.org/10.32782/1995-0519.2022.1.3

COMPREHENSIVE ASSESSMENT OF THE APTITUDES OF STUDENTS OF GEOENGINEERING EDUCATION

Igor Yeremeyev
DSc.,
Professor at the Department of Automated Process Control
Taurida National V. I. Vernadsky University, 33 Ivana Kudri str., Kyiv, 04000, Ukraine;
ORCID: 0000-0003-1968-0395

Alina Dychko
DSc.,
Professor at the Department of Geoengineering
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", 37 Peremohy Avenue, Kyiv, 03056, Ukraine, aodi@ukr.net;
ORCID: 0000-0003-4632-3203

The paper is devoted to the problem of students' knowledge assessment, because very fact of a grade of a student in a particular subject does not reflect his or her competences. The purpose of the research is the analysis of deviations, i. e. identification of data that differ in any parameters in the total mass. Usually, students become more critical during their education (from their point of view, subjectively or objectively) of the subjects they studied: something they do not need, and something – is extremely important. Accordingly, this is reflected in the estimates. It is proposed to use the ideas of spectral analysis, the method of nearest neighbors, plausible reasoning, fuzzy set methods and identification methods. The idea of spectral analysis is that all student assessments are located at fixed positions: each assessment of indicators in a predetermined place in groups of selected subjects. This rating can be the result of assessment of knowledge during the module control, assessment of the course project, test results, etc. When using distance learning methods and tools in the dialogue mode, the assessment is set by a program that monitors the assimilation of current material. Identification of a specific spectrogram can be carried out using the method of the nearest neighbor and the Euclidean metric as a measure of proximity. The presented appropriate algorithms of the student's self-preparation process includes the main aspects of education process. The proposed heuristics for decision making on the knowledge defines the presence of certain trend, comparing the situation where the estimate does not exceed the standard deviation and the result of the observation of the change in the estimate in the current and during the previous assessments. The developed approach helps to assess students' knowledge more objectively and to identify trends in the development of material, gaps in knowledge and the aptitudes of students.

Key words: students' knowledge, students' aptitudes, data mining, spectral analysis.

1. INTRODUCTION. DATA MINING IN STUDENTS' KNOWLEDGE ASSESSMENT

An existing methods of teaching and assessing students' knowledge and abilities have now become obsolete. The very fact of a high or low grade of a student in a particular subject says little. A systematic approach is needed to assess students' abilities and inclinations. This approach should be based on the intellectual analysis of data [1] on academic performance and other activities (participation in conferences, social activities) for the entire previous period of study in absolutely all disciplines relevant to the chosen specialty, both compulsory and optional. The purpose of such an analysis is to detect deviations, i. e. identification of data that differ in any parameters in the total mass. Usually, as students get older and gain new knowledge, they become more critical (from their point of view, subjectively or objectively) of the subjects they read: something they do not need, and something – is extremely important. Accordingly, this is reflected in the estimates. On the other hand, teachers emphasize students' importance of the subjects they read (usually giving examples of how a "trifle", that is not perceiving, and ignoring seemingly insignificant facts, can lead to a catastrophe or failure to achieve a goal) and thus, try to arouse interest in a particular section or subject in general. Here it is important to know – the observed ignorance is characteristic of all or only some listeners, what are the trends in different groups of students, etc. In addition, it is desirable to find relationships between individual indicators of general trends. Finally, it is very important to summarize the built model (template) for application to new data in order to predict the genesis of each of the listeners. All this, in fact, can be described by

the term of data mining. To implement the proposed approach, it is proposed to use the ideas of spectral analysis, the method of nearest neighbors, plausible reasoning, fuzzy set methods and identification methods. The idea of spectral analysis is that all student assessments (expressed in points, or rather, in the scale [0,00...1,00]) are located along the abscissa axis at fixed positions: each assessment of indicators in a predetermined place in groups of selected subjects. This rating can be the result of assessment of knowledge during the module control, assessment of the course project, test results, etc. In addition, when using distance learning methods and tools in the dialogue mode, the assessment is set by a program that monitors the assimilation of current material. After viewing the unit, the student takes a test and by its results he turns to the unit that is poorly mastered (and the number of iterations are taken into account as an indicator of the weakness of the assimilation of the material), or turns to the next subsection, which must be studied. The algorithm of this procedure is shown in Fig. 1–2.

It is possible to adjust the assessment as follows. If the grade exceeds, for example, 80 points (or 0.8), and the grades in related disciplines exceed 0.85, the following correction factor n can be offered:

$$n = (L - 1)/(2N + L)1/N\sum_{i=1}^{n} x_i, \qquad (1)$$

where L – number of grades in related disciplines exceeding 0.85, N – the total number of grades, x_i – grade of the i-th subject. In this case, the adjusted estimate is lower if $L = 0$, is not changed at $L = 1$ and increases with increasing L. The value of the estimate $x(i_{cor})$ is defined as $x(i\ cor) = n_{xi}$. The graphic interpretation of that is shown in Fig. 3.

Each student, as shown in Figure 3, corresponds to his spectral characteristics, and it is possible to formulate certain patterns

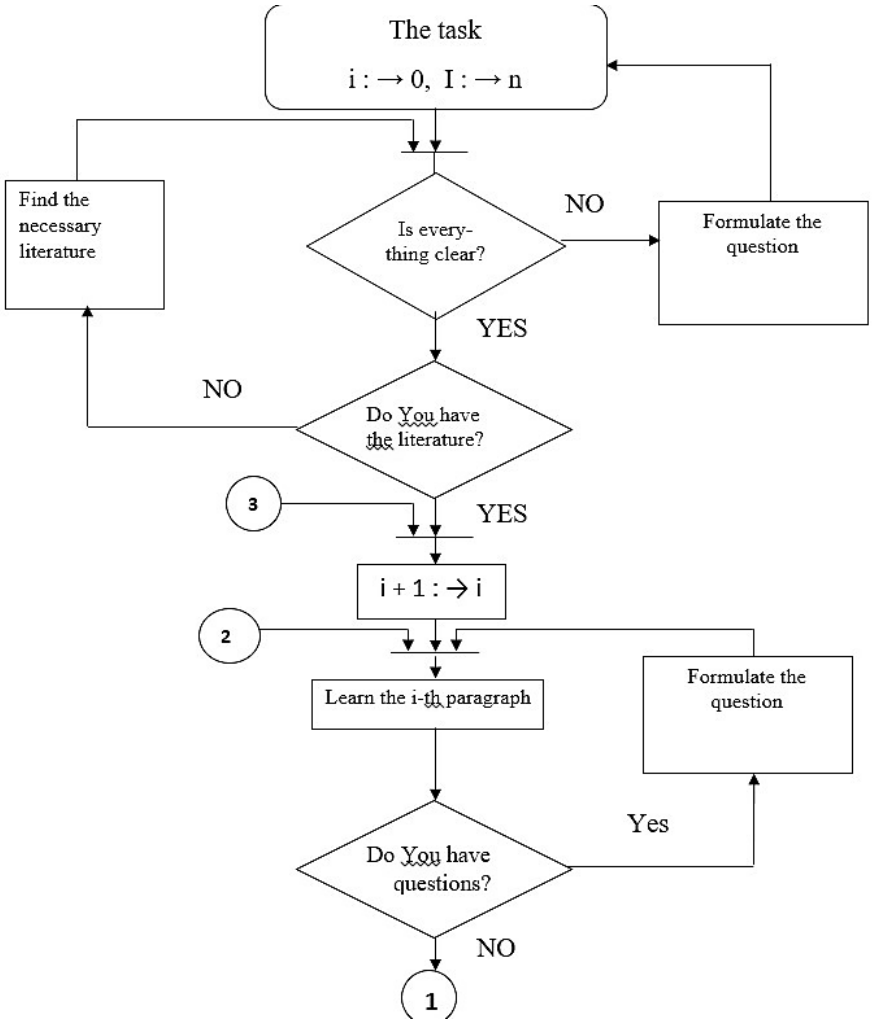

Figure 1 – A. Block diagram of the student's self-preparation process

that correspond to the assessment standards (classes) of the Bologna process.

2. IDENTIFICATION OF THE STUDENTS' ABILITIES SPECTRUM

Comparison of characteristics with the standard (template) with simultaneous indication of acceptable tolerance in the scatter of characteristics 2 relative to the standard will allow to unambiguously identify them as belonging to a class, and the presence of "strange" spectrum data will show obvious flaws in student knowledge and the need course.

Graphical interpretation of the identification procedure based on the evaluation of the Euclidean metric, which characterizes the discrepancy between the compared spectrogram and the standard [2; 3] is presented at Fig. 4–5. Here 1, 2, 3, ... – disciplines S, relevant to the profiling course, Oc – the score in points (0.0–1.0), and the red line is the standard.

Identification of a specific spectrogram can be carried out using the method of the nearest neighbor [4], the essence of which is as demonstrated at Figure 6. Here, the Euclidean metric is used as a measure of proximity.

The test specimen (green circle) must be classified as a blue square (class 1) or as a red triangle (class 2). If k = 3 (solid circle), then it is classified as 2nd class, because inside the circle there are 2 triangles and only 1 square. If k = 5 (dotted circle), it is classified as the 1st class (3 squares against 2 triangles inside a larger circle).

The simplest algorithm of the method is as follows: 1. All points (Ai) of the data set are projected on one axis. 2. The test point (At) is also projected on this axis. 3. Find the two nearest points (to the right (Ap) and to the left (Al) from the point (At)). 4. Measure the distances between the two templates found in p. 3 and the test point (At) and identify which of the distances is shorter (Rd). In some cases, if it is needed to identify causal links and predict the dynamics of consolidation of student knowledge in repeated testing (which objectively contributes to better learning), a modification of the method

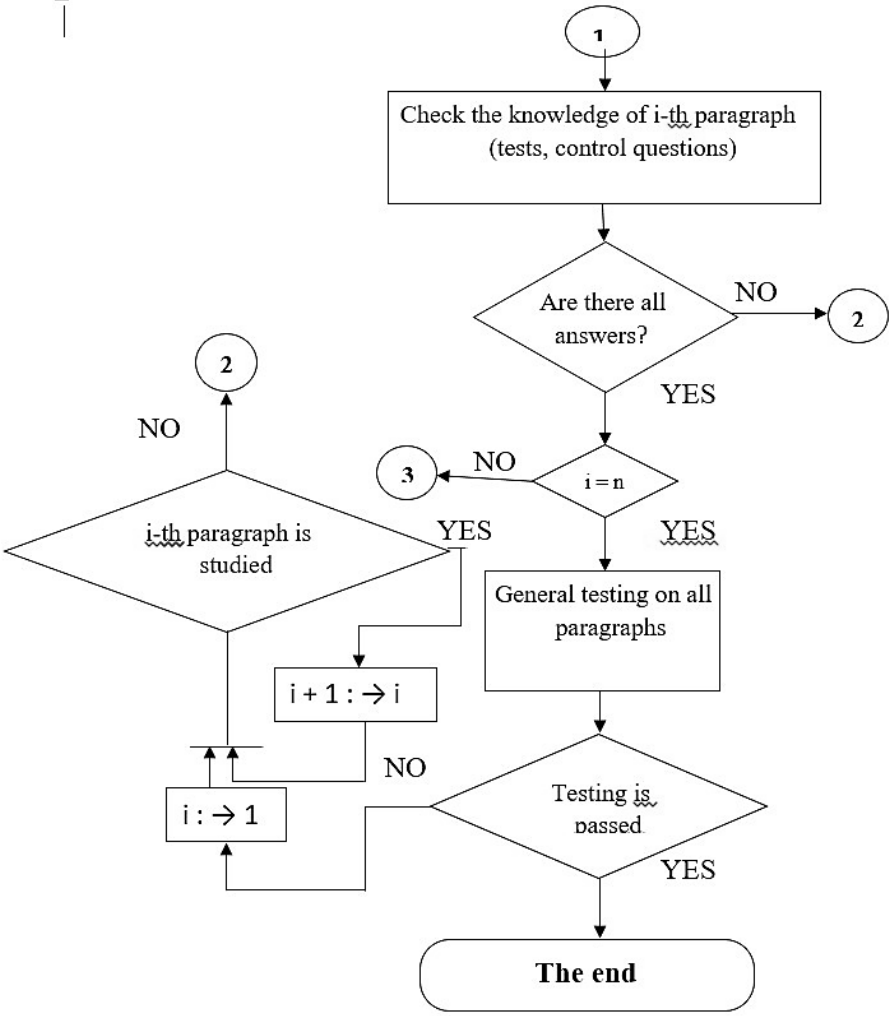

Figure 2 – B. Block diagram of the student's self-preparation process

of nearest neighbors may be used, which is based on plausible reasoning in assessing within 2 [4].

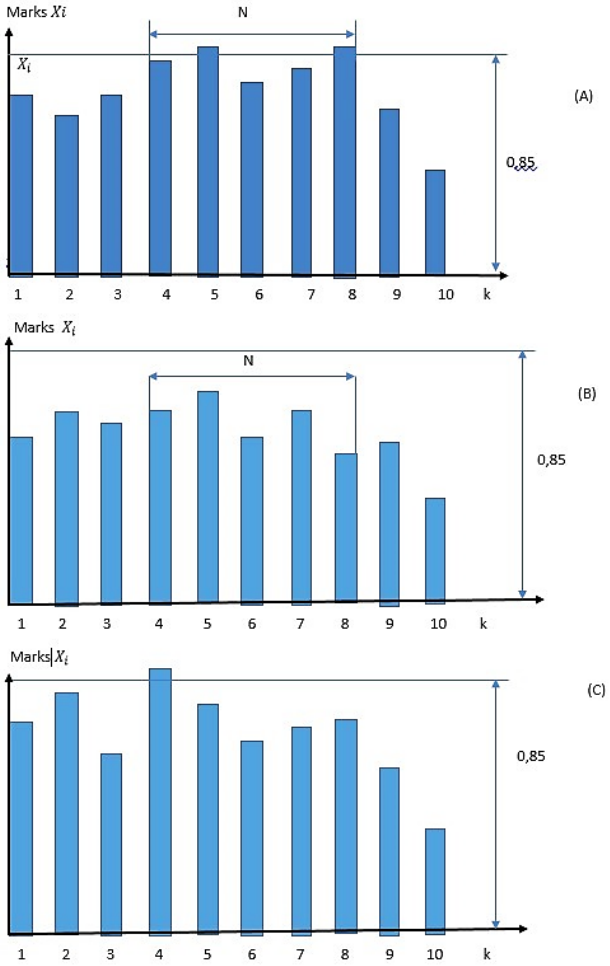

Figure 3 – Spectrograms for cases N = 5 and L = 2 (A), L = 0 (B) and L = 1 (C)

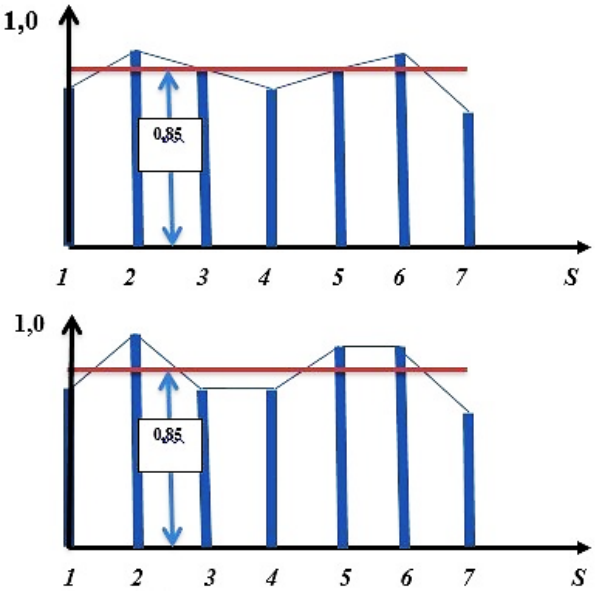

Figure 4 – A. Graphical interpretation of the spectrum identification procedure

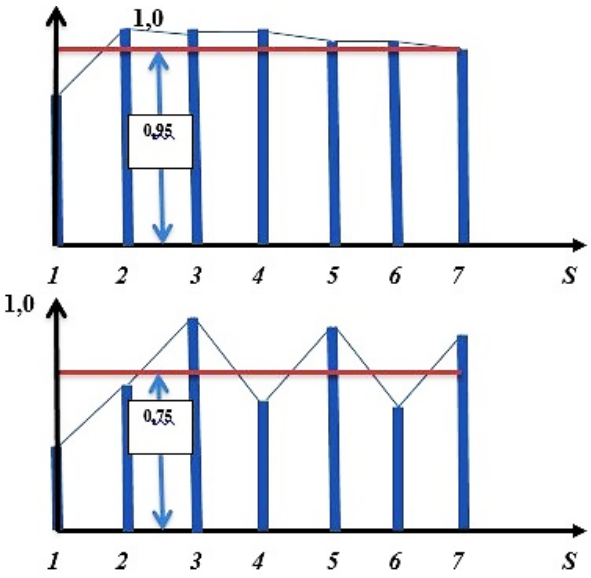

Figure 5 – B. Graphical interpretation of the spectrum identification procedure

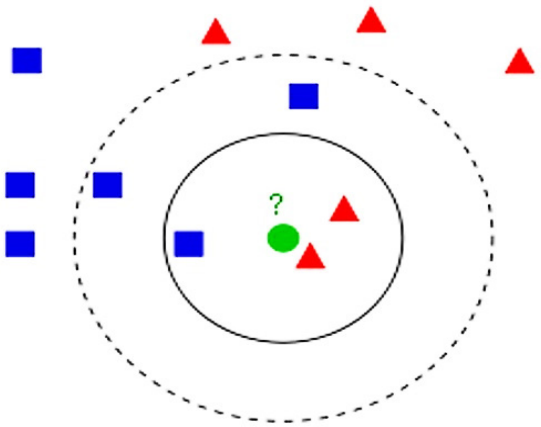

Figure 6 – Example of classification of k-nearest neighbors

The essence of the approach is that the spectrum data are compared at points of interest with the spectrum data at the nearest neighboring points. If there is a consistent dynamics of the amplitudes of the spectral lines at all adjacent points within the standard deviation of 2, it can be considered very plausible to have a certain trend that is credible. The following heuristics can be used to estimate the above trend:

IF ([STAB] AND [SIGNSTi±jEQ]), THEN [HOT], (2)

where STAB – situation where the estimate does not exceed the standard deviation, SIGNSTi ± jEQ – the result of the observation that the sign of the change in the estimate in the current assessment at all points to the left and right of *i*, i. e. at points *i* and *ij* and from *i* to *i* + *j*, is the same with respect to

the state at the same points during the previous assessment, HOT – there is a certain trend. Graphical interpretation of this approach is shown in Fig. 7.

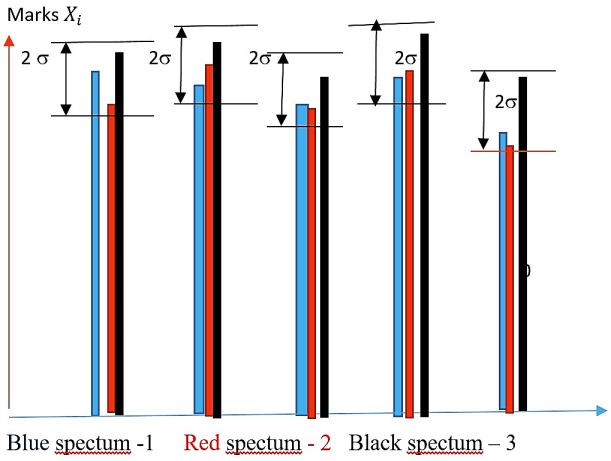

Figure 7 – Comparison of spectra within 2 according to heuristics (2)

CONCLUSIONS. The approach discussed above, thus, helps not only to assess students' knowledge more objectively, but also to identify trends in the development of material, gaps in knowledge and even the aptitudes of students to geoengineering education and sciences. This students' rating can be the result of assessment of knowledge during the module control, assessment of the course project, test results, etc. Identification of a specific spectrogram can be carried out using the method of the nearest neighbor.

REFERENCES

1. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. From Data Mining to Knowledge Discovery in Databases, *AI Magazine*, (17), 1996. P. 37–54.

2. Dychko, A., Yeremeev, I. Organization of environmental monitoring using fractal theory methods. *Complex Systems Development Management* (19), 2014.

3. Yeremeyev, I., Dychko, A., & Gulich, S. (2021). Intellectualization of Educational Resources and Analysis of Students' Training. *In 2020 3rd International Seminar on Education Research and Social Science (ISERSS 2020)*. P. 21–26.

4. Beyer, B., Goldstein, J., Ramakrishnan, R., Shaft, U. When Is "Nearest Neighbor" Meaningful? Database Theory. – *ICDB'99*, 1999. P. 217–235.

КОМПЛЕКСНА ОЦІНКА ЗДІБНОСТІ СТУДЕНТІВ ГЕОІНЖЕНЕРНОЇ ОСВІТИ

Ігор Єремєєв
доктор технічних наук,
професор кафедри автоматизованого управління технологічними процесами
Таврійський національний університет імені В. І. Вернадського, вул. Івана Кудрі, 33, Київ, 04000, Україна;
ORCID: 0000-0003-1968-0395

Аліна Дичко
доктор технічних наук,
професор кафедри геоінженерії
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», пр. Перемоги, 37, Київ, 03056, Україна, aodi@ukr.net;
ORCID: 0000-0003-4632-3203

Стаття присвячена проблемі оцінки знань студентів, оскільки сам факт оцінки учня з того чи іншого предмета не відображає його компетенції. Метою дослідження є аналіз відхилень, тобто виявлення даних, що відрізняються за будь-якими параметрами в загальній масі. Зазвичай студенти під час навчання (з їхньої точки зору, суб'єктивно чи об'єктивно) стають більш критичними до предметів, які вивчають: щось їм не потрібно, а щось – надзвичайно важливо. Відповідно, це відображено в оцінках. Запропоновано використовувати ідеї спектрального аналізу, методу найближчих сусідів, правдоподібних міркувань, методів нечітких множин та методів ідентифікації. Ідея спектрального аналізу полягає в тому, що всі оцінки студентів розміщуються на фіксованих позиціях: кожна оцінка показників у заздалегідь визначеному місці в групах вибраних предметів. Цей рейтинг може бути результатом оцінювання знань під час модульного контролю, оцінювання курсового проекту, результатів тестування тощо. Ідентифікацію конкретної спектрограми можна провести за допомогою методу найближчого сусіда. Розроблений підхід допомагає більш об'єктивно оцінити знання учнів та виявити тенденції освоєння матеріалу, прогалини в знаннях та здібностях учнів.

Ключові слова: знання студентів, здібності студентів, аналіз даних, спектральний аналіз.

BIBLIGRAPHY

1. Fayyad U., Piatetsky-Shapiro G., Smyth P. (1996). From Data Mining to Knowledge Discovery in Databases, *AI Magazine*, Volume 17, Number 3, pp. 37–54.

2. Dychko A., Yeremeev I. (2014). Organization of environmental monitoring using fractal theory methods. *Complex Systems Development Management*, (19).

3. Yeremeyev, I., Dychko, A., & Gulich, S. (2021). Intellectualization of Educational Resources and Analysis of Students' Training. *In 2020 3rd International Seminar on Education Research and Social Science (ISERSS 2020)*, pp. 21–26.

4. Beyer B., Goldstein J., Ramakrishnan R., Shaft U. (1999). When Is "Nearest Neighbor" Meaningful? Database Theory. – *ICDB'99*, pp. 217–235.

Стаття надійшла 14.03.2022