Adaptive Technology for Students' Knowledge Assessment as a Prerequisite for Effective Education Process Management

Vitaliy Snytyuk and Oleh Suprun

Taras Shevchenko National University of Kyiv, Intellectual and Information Systems Department, 24 Vandy Vasilevskoy Str., Kyiv, Ukraine
snytyuk@gmail.com, oleh.o.suprun@gmail.com

Abstract. Despite the rapid development of intellectual systems and their implementation in education process, the knowledge assessment technologies are almost the same as decades ago. This paper presents an intellectual system approach to the student's knowledge evaluation process, based on adaptive mechanisms development. It is an individual directed systems and it allows one to correct the questions complexity levels in real time, depending on the average students score and to reduce the examiner's influence to the evaluation process. Besides, the proposed system allows one to consider not only the testing result itself, but also different criteria, like time, used by a student to answer the question. According to this is idea, the assessment can demonstrate not only student's pure knowledge, but also his skills and ability to use obtained knowledge in different situations. The experimental results that show the system's effectiveness are presented.

Keywords: Intellectual Systems, Distance Education, Knowledge Evaluation.

1 Introduction

Over the past decades, intellectual systems usage and process automation in general can be found in the most diverse areas of human activity, and the education sector is no exception. Intellectual systems are used as a tool for creating mew courses, to plan students and teachers effective work, to evaluate various electronic sources and information resources, and to automate the assessment of students' knowledge [1, 2].

Standardisation of the student assessment results has become especially relevant within the Bologna Process, designed to create and spread the common education standards in different countries. But, at the same time, such systems use the testing as a way of primary evaluation of the trainee's knowledge, which does not always allow to evaluate his skills fully.

Comprehensive knowledge assessment is very important lately, because, due to the rapid science development and the widespread use of computer and information technology, the number of requirements for a specialist is steadily increasing. These requirements are far from being concrete and unambiguous, and the assessment of such specialists is made, taking into account various skill scores. For example, during the interviews in IT companies, students are often asked questions that do not require a specific answer, but which can show the way, the applicant thinks, like, "how many tennis balls fit into a school bus." Obviously, it is impossible to evaluate the answer for such a question using the classical testing system, but it is not possible to conduct a detailed conversation with each student during the exam.

Furthermore, distance education and self-education are gaining in popularity very fast and perfectly complement the classical education system in schools and universities, making it possible to acquire necessary knowledge, for example, for working people with the lack of free time, or residents of other countries. Assessing these student's knowledge is another problem, since the teacher's participation is minimized, or he is not present at all, and the trainee's skills can be assessed only according with his specific answers. It is also necessary to take into account the numerous courses conducted by large companies for their own employees [3]. It is much more profitable for a company to retrain a qualified specialist, than to search for a new one, spending money and wasting time, not to mention the risk associated with attracting a new employee with uncertain skills level.

Considering the speed of technological development, when a certain standard or methodology can be developed and become obsolete within several decades, or even years, it's impossible to verify the evaluation system correctness on a large number of students, the system should immediately react to the current student’s level, their knowledge and external requirements [4].

Thus, there is a need to develop an intellectual system, that would be able to consider not only the correctness of a given by a student answer, but also take into account different criteria of his answer [5]. Moreover, very important is the ability of such system to adapt itself, adjust the complexity of the questions in order to give the most objective assessment not only to the student's knowledge as such, but also the ability to apply this knowledge in practice and to use it in various situations.

The article presents the basis of such system, shows its ability to change the questions complexity and to take into account various criteria, for example, the time, used by the student for answering each question.

2 The Method for Adaptive Correction of Questions Complexity Scale

The necessity to objectively evaluate the complexity of the questions, used to assess students' knowledge, is obvious while creating the new education course, that can be unknown not only for students, but also for tutors [6]. According to the classical methodology, the complexity scale is set by an expert or examiner, that always has an element of subjectivity and requires a thoroughly revision using statistical data after repeated use of such a complexity scale [7]. However, considering the modern world realities, the results must be obtained very quickly, and the price of the mistake during the evaluation can be too high [8]. Not to mention the impossibility to collect a sufficient amount of statistical data, considering a large number of narrowly focused courses, primarily adapted to individual specialists training who must be ready to start their work immediately after graduation.

Adequate consideration of the questions complexity is also important during the study of classical fundamentals, for example, mathematics or physics. It is impossible to verify the student's knowledge for the entire course, it would take almost more time than learning itself, so the questions selection during assessment should be optimal.

Solving these problems is the main goal of creating an adaptive knowledge evaluation system. Being inherently individually directed, it will allow to assess the knowledge of each student more accurate, while being free of the flaws such as subjectivity or predisposition. For example, it makes no sense to ask simple questions the student who gives excellent answers, and on the contrary – it is illogical to give complex tasks to the student who has shown his abilities to be below average. But first the proper scale of questions complexity, based on the students knowledge and their abilities, must be created.

The adaptive complexity correction system is based on the following simple principle – if the student gave the wrong answer to the question, this question should be considered to be more complicated, and the right answer reduces the question complexity. Of course, while implementing this system, many nuances must be taken into account, for example, the overall score or other indicator of the student's progress and knowledge, the statistics of answers to a particular question, and others.

Let the set $Q = \{q_i\}, i = \overline{1,m}$ be the set of questions, used for trainee’s knowledge evaluation, m the overall number of questions in the set. Then $p(q_i) \in (0,1]$ is the complexity of each question.

According to the classic methods, that don’t use intellectual or any other knowledge evaluation systems, the complexity of each question is determined by the tutor or examiner. Let the $p_i^0, i = \overline{1,m}$ be the predetermined questions complexity, according to which the tasks assignment and, later, the estimation is made. According to the adaptive assessment system main idea, this value is not a constant, but depends on students' answers in real time.

At the same time, the system should not contradict the following principles:

1. The correct answer to the question reduces its complexity, and on the contrary – if the answer is incorrect, the question is considered to be more difficult.
2. The corrected question complexity can not go beyond the previously established scopes, that is, it must always belong to the interval $(0,1]$.
3. The more students pass the test, the less influence each individual answer has on the question complexity.
4. 4. The question complexity should be adjusted depending on the total score of the student giving the answer. If a student, that has a comparatively high total score, gives the incorrect answer, the question complexity should increase by a bigger value than if the incorrect answer is given by the trainee with lower total score. And on the contrary – the correct answer of the student with a lower overall score

reduces the complexity of the question by a bigger value than the correct answer of the student with a high overall score.

According to these principles, after passing the test by each trainee, the complexity of each question should be recalculated according to the following formula:

$$p_i^j = p_i^{j-1} + f_i(p_i^0, m, d^{j-1}, Z), i = \overline{1,m}, j = \overline{1,n},$$

where n – the overall number of trainees, who pass the test, m – number of questions used to assess knowledge, p_i^j – the corrected question complexity, p_i^{j-1} – the question complexity, used during the previous trainee's testing, f_i – the function, used to correct each question complexity, according to the mentioned principles: p_i^0 – the predetermined questions complexity, amount of students m, d^{j-1} – the overall score of previous, $(j-1)$ -th trainee, who answered the question, Z – coefficient, that depends on the deviation of the given answer from the correct one.

According to the formula above, the questions complexity is recalculated in real time, after passing the test by each student that provides the ability of the system to work effectively without having a large amount of statistical data.

By assuming the presence of a large number of trainees, a conclusion can be made, that the value of the complexity level converges to a constant value.

At the same time, such adaptive system assumes the availability of initial data that are set by the examiner or expert, in particular, a preliminary complexity scale p_i^0. The situation when the initial questions complexity is not known must be considered. For example, when starting a new course, or if the examiner strives to objectify the assessing knowledge process. In this case, the students overall scores distribution, based on their previous education successes, or one of the additional heuristics can be used, and depending on this, the initial questions complexity can be distributed.

Obviously, students who passed the test first, can get a significant bias evaluation due to randomness factors. Thus it makes sense to recalculate their scores, using the previously acquired information about their correct answers, and applying a new questions complexity scale, obtained after passing the test by a certain number of students.

3 Multi-Criteria Knowledge Assessment System

Considering the modern world realities and the information society features in particular, specialists in many branches, especially those related to computer and computing technology, interacting with other people and working on large projects, are more often facing problems that require an integrated approach and the ability to apply skills from various branches of science. The knowledge availability is no less valued than the ability to apply them in practice in various situations. The requirements for such specialists can not always be clearly formalized [9], therefore, it is impossible or very difficult to assess their knowledge and level of their preparation by using the classical systems.

In practice, companies and employers solve the problem of adequate assessment by means of numerous interviews, during which not only the student's dry knowledge is analyzed, as it is done in universities, but also his capacity for logical thinking in general, the ability to find a non-obvious way out of complex problems and his ability to analyze some steps forward. It is obvious that it is often impossible to use such methods in universities or for any other education process, due of the limited amount time and resources, and the large number of trainees. Similarly, the use of widespread testing will can not guarantee the desired results.

Solving this problem is possible, using the integrated or multi-criteria knowledge assessment system. It is based on the following principle – to assess the student's knowledge not only for each specific answer to the question, but also to consider his answers in general, and at the same time, to evaluate each answer from several points of view. For example, a student who failed to give a correct final answer to the assigned task, but who demonstrated the correct solving algorithm, will get a greater score than the one who gave the correct intermediate answer, but was guided by this erroneous opinion.

One of such criteria, in addition to the answer itself, can be the time used by the trainee to find the answer. In many situations, time is the decisive factor, and an incorrect or untimely decision can have serious consequences like material damage, financial losses, or even environmental or technological disasters. First of all, it applies to workers of rescue structures, power stations and other strategically important facilities. But, recently, the speed of decision-making is important for many office workers dealing with computing technique [10]. The best example of this is the attack of the virus “Petya”, when some large companies managed to repel a virus attack due to, first of all, the attention of one or two employees who noticed an abnormally rapid files distribution from a single source and blocked this source, which ultimately saved the company from huge financial losses. And, although the simulation of such a situation during the testing is quite difficult and expensive, it is quite possible to assess the speed of decision making or answering the question.

Designing such a system, the following factors should be considered:

- The answer to the question received from the trainee is correct, incorrect or partially correct;
- The answer was given within the prescribed time interval, with an excess, or it was not given at all;
- Answers received after the time expiration are considered to be incorrect, or taken with the application of appropriate penalties.

Considering the simplest version of the assessment system, taking into account the time spent on the answer, the following formula holds:

$$r = \frac{\sum_{k=1}^{m} p_k \cdot \chi(k) \cdot \delta(t^k \le T^k)}{\sum_{k=1}^{m} p_k},$$

where r – the overall testing result, the test consists of m questions, p_k – the complexity of k-th question, $k = \overline{1,m}$, $\chi(k)$ and $\delta(t^k)$ – coefficients, that depend on the answer correctness and time, used by trainee to give it, $T^k, i = \overline{1,m}$ – the maximal time, determined by the examiner to answer each question.

To simplify the model, assume that $\chi(k)$ can take only 2 values – either the answer is correct or incorrect:

$$\chi(k) = \begin{cases} 1, & \text{if the answer for the } k\text{-question is correct;} \\ 0, & \text{otherwise.} \end{cases}$$

According to the classical testing system, the answer given after the expiration of time is considered incorrect:

$$\delta(t^k \le T^k) = \begin{cases} 1, & \text{if } t^k \le T^k; \\ 0, & \text{otherwise.} \end{cases}$$

In this case, it is impossible to fully appreciate the student's knowledge, taking into account several criteria, for example, the time. Often the answer, which was given with time excess, but is correct, is worth considering. To do this, it is necessary to use a penalty function that reduces the amount of points received if the time limit is exceeded:

$$r = \frac{\sum_{k=1}^{m} p_k \cdot \chi(k) \cdot (1 - \beta_k \cdot \delta(t^k > T^k))}{\sum_{k=1}^{m} p_k},$$

where $\beta_k = \beta_k(p_k)$ – coefficient that determines the amount of penalty for an untimely but correct answer. Obviously, the penalty should depend on the question complexity – if the trainee answered with a delay to a simple question, then the penalty value should be bigger comparing to penalty when the time is exceeded with a complex question. For example, such penalty value can be used: $\beta_k = \eta_k \bullet (1 - p_k)$ where η_k – predetermined normalization coefficient. For the penalty coefficient correctness, it must satisfy the following condition: $0 < \eta_k < \frac{1}{1 - p_k}$.

Moreover, often it is important to consider, how far the time threshold was exceeded. Obviously, the more time the trainee has spent, the less points he will receive for the correct answer. For example, it can be done as follows: $\delta(t^k > T^k) = \frac{T^k}{t^k}$. As can be seen, with a large time limit excess, the student gets less points for the correct answer.

Thus the original formula takes the following form:

$$r = \frac{\sum_{k=1}^{m} p_k \cdot \chi(k) \cdot (1 - \beta_k(t^k)) \cdot \delta(t^k)}{\sum_{k=1}^{m} p_k},$$

where

$$\beta_k(t^k) = \begin{cases} 0, \text{ if } t^k \le \mathrm{T}^k; \\ \eta_k \cdot (1 - p_k), \text{ otherwise} \end{cases},$$

$$\delta_k(t^k) = \begin{cases} 1, \text{ if } t^k \le \mathrm{T}^k; \\ \frac{\mathrm{T}^k}{t^k}, \text{ otherwise} \end{cases}.$$

This formula makes it possible to evaluate the trainee's knowledge, taking into account not only the correctness of the answer, but also the time used to find it, even if the initial time limits were exceeded.

The formula for question complexity level also must be modified, taking into account several parameters:

$$p_i^j = p_i^{j-1} + f_i(p_i^0, m, d^{j-1}, Z, T_i), j = \overline{1,m}, i = \overline{1,n},$$

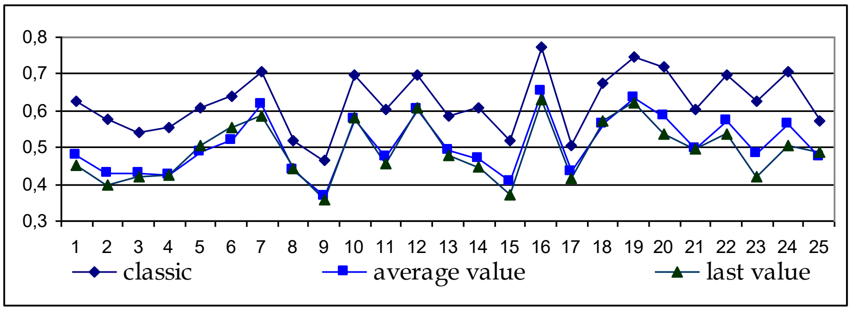
where m – the questions number, n – trainees number. In this formula also the time limit, which was set to answer the question, is used. If the answer is received in time, this parameter can be ignored, otherwise, if the answer is correct, but received with a delay, two options are possible:

- The question complexity should be reduced – if the correctness of the answer is more important than the time excess
- The question complexity should be increased – if time limit is more important than the correct answer.

Implementations of this approach results in more accurate control of the trainee's knowledge, and, more importantly, it gives opportunity to evaluate his ability to use the acquired knowledge.

4 Example of Adaptive Assessment Technology Application

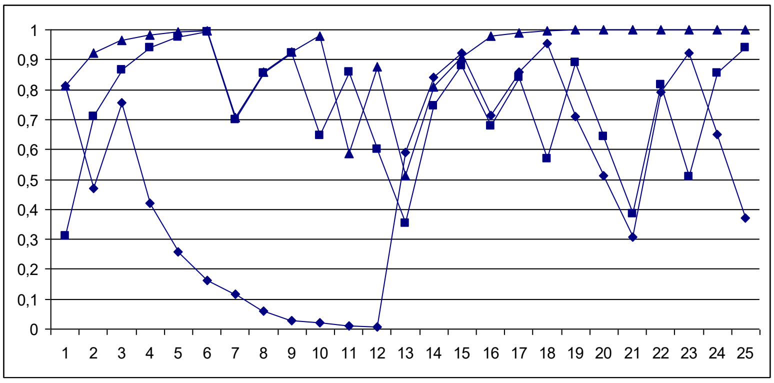
To test the method for questions complexity correcting in real time, an experiment was conducted with one group of 25 students. The test consisted of 100 questions. Before the test, the examiner had set the initial complexity of all questions to be equal $0,5$. The experiment results are shown in the figure below.



The top line corresponds to testing, using the classic knowledge assessment method, the average score was $0,623$.

The middle and bottom lines correspond to testing, using the proposed adaptive system with the initial questions complexity $0,5$. The complexity was changed according to the rules described above. After the testing was completed, the results of all the students of the group were recounted. Two variants were considered: with the final complexity (total score was $0,492$) and average complexity ($0,507$) during the testing time.

The questions complexity adaptation in real time on the example of three questions is shown in the figure below.



One of the main results of using the adaptive system is reducing the time, spent for assessment, which allows to increase the time of studying.

The proposed methodologies testing was made during assessment of six groups, it is presented below. According to the classic assessment system, the students had to answer all 50 questions. The optimized method, based on the adaptive ascending scheme [11], reduced the number of questions to 14-16 for each student. The time taken to pass the test for each of the 6 groups is shown in the table.

Group ID	Trainee's number	Classic assessment method, min.		Optimized method, min.		Relative deviation, %		Absolute deviation, point	
		A_1	A_2	B_1	B_2	C_1	C_2	D_1	D_2
1_1	29	60	62	24	27	6,7	4,2	0,201	0,126
1_2	30	77	78	28	31	4,2	5,6	0,126	0,168
2_1	32	70	75	27	25	5,1	5,8	0,153	0,174
2_2	36	75	75	25	27	5,3	4,9	0,159	0,147
3_1	36	72	76	28	30	6,5	6,1	0,195	0,183
3_2	37	80	80	28	32	6,2	6,3	0,186	0,189
Average	33,33	72,3	74,3	26,7	28,7	5,67	5,67	0,17	0,16

The average relative deviation of the estimates obtained using different schemes was $5,67\%$, or the score deviation $0,17\%$, according to the five-point rating scale.

These results indicate a significant saving of time spent on testing, with a relatively small results deviation.

In the second experiment, a multi-criteria evaluation system was used. Within 50 questions proposed, 20 had a time limit, and the time criteria were predominant in comparison with the correctness criteria. The average estimation time (A_2) increased slightly. The adaptive system applying allowed to reduce the time (B_2) to 28 minutes in average, 16-18 questions were asked each student. The increase in the questions number by the system is due to a slight increase in the number of errors made by students.

As can be seen from the tests results, the proposed systems allows to significantly improve the knowledge assessing process.

5 Conclusions

Recently, more and more research are dedicated to intelligent systems application in the education process. Almost all of them are aimed at optimizing the learning process as such, helping to determine the set of necessary disciplines or to plan the use of time correctly, but they almost do not improve the systems of students' knowledge assessing. Only few of them are aimed at an adequate comparison of the results obtained during students testing, but they do not affect the testing process itself, leaving the questions selection and determining their complexity task for teachers and examiners. Moreover, classical testing system as such does not allow to comprehensively assess the students skills level, demonstrating only their knowledge, but not the ability to use them in a difficult situation, which is very important in the modern world.

The presented adaptive assessment system allows to objectify the knowledge assessment process by correcting the questions complexity of in real time, depending on the knowledge of specific student and students in general. At the same time, results are formalized, which allows to compare the results of different groups.

Applying the multi-criteria technique makes possible to assess not only the student's knowledge as such, but also the ability to use them, to take into account the various aspects of trainee's answers, which is necessary for a comprehensive evaluation of the training results. This system allows to analyze the student's knowledge fully, to indicate possible omissions in the training for their subsequent elimination.

Furthermore, the presented models can be used not only for knowledge assessment, but also as an element of self-educational programs. For example, a similar principle at the primitive level is implemented in mobile dictionary applications: the program offers the user a set of words to test his knowledge, if multiple correct answers are given, the word is removed from the control sample, or is offered again if an error is made.

It must be noted that applying such a system requires the implementation of a large amount of verification procedures, but, at the same time, it remains simple and can be easily upgraded according to specific requirements. Its use can reduce evaluation time, which is very important when working with a large number of trainees, or with distance education, and subsequent results analysis will help to improve the education process.

References

1. Hiltz, S.R., Murray, T.: Education goes digital: The evolution of online learning and the revolution in higher education. Communications of the ACM 48(10), 59-64 (2005)
2. Encheva, S., Sharil, T.: Application of association rules in education. In: Intelligent Control and Automation, Lecture Notes in Control and Information Sciences 344, Springer Berlin Heidelberg, 834-838 (2006)
3. Taylor, J.: Decision Management Systems: A Practical Guide to Using Business Rules and Predictive Analytics. Boston MA: Pearson Education (2012)
4. Gharajedaghi J., Ackoff R.: Toward Systemic Education of System Scientists. Systems Research 2, 21-27 (1985)
5. Snytyuk V., Gnatyenko G.: Optimization of the assessment process in conditions of uncertainty based on the structuring of the subject area and the axiom of unbiasedness. Artificial Intelligence 3, 217-222 (2008). (in Russian)
6. Diamond, R.M.: Designing and assessing courses and curricula: A practical guide. John Wiley & Sons (2011).
7. Encheva, S., Sharil, T.: Uncertainties in knowledge assessment. WSEAS Transactions on Advances in Engineering Education 5(4), 895-900 (2008).
8. Mreła A., Sokołov O.: Rankings of Students Based on Experts' Assessment of Levels of Verification of Learning Outcomes by Test Items. In: 13th International Conference on ICT in Education, Research and Industrial Applications (2016).

9. Popereshnyak S., Suprun O.: Tools and methods for intersubjective relationships in cyberspace forecasting. In: XIIth IEEE International Scientific and Technical Conference CSIT-2017, pp. 244-247, Lviv, Ukraine (2017).
10. Wick D.: Free and open-source software applications for mathematics and education. In: Proceedings of the twenty-first annual international conference on technology in collegiate mathematics, pp. 300–304, New Orleans (2009).
11. Snytyuk V., Yurchenko K.: Elements of adaptive type knowledge-oriented training systems. Kherson National Technical University 2 (38), 180-186 (2010).