Trainable generator of educational content

Vladimir Rotkin
Independent Researcher, University of Haifa International School, Haifa, Israel

ABSTRACT
As the main problem of the research, the possibility of creating a universal educational platform that combines the possibilities of an online generation of educational content with the interface of the training process itself was considered. The methodology of the educational platform has been developed, in which the mass generation of content is carried out at random, based on simulation models of educational objects. A matrix interface is used, which allows performing custom operations by entering a sequence of typical operators. The system forms a reference base of operators, replenishing it from user solutions, which makes it possible to train and improve the system in order to provide methodological support to student users. An active demo layout of an educational content generator was created and tested, using the example of a specific problem from school mathematics. All methodological options function in the layout. There are three interface options: administrative, training and control. It was concluded that the approach based on the simulation of educational objects makes it possible to create a unified algorithmic platform that combines the functions of content generation with educational training. The system contains a unique option to teach yourself based on its interaction with students.

Keywords:
Content generator
Educational content
Information technology
Machine learning
Operator identification
Settlement operator

1. INTRODUCTION
The main feature of modern education, perhaps, lies in the presence of a significant gap between the capabilities of information and communication technologies (ICT) and their implementation in the educational process. On the one hand, there is a rapid penetration of the latest technologies into the field of education, from administration to the training process, many participants are motivated and ready for the widespread use of ICT. However, on the other hand, the “quantitative” advantages of ICT (speed and volume of information processing, the possibility of remote and joint work) are accompanied by incomparably small qualitative changes in the educational process itself. Mainly, traditional methods and didactics are used: old, possibly slightly modified, educational content retrieved from databases; video lectures from "talking heads"; the usual forms of training.

In all walks of life, ICTs have become ubiquitous institutions. The use of ICT in higher education contributes to the creation of a student-centred learning environment [1]. Benefits and challenges of using ICT: increasing student efficiency, reducing teacher time and effort, reducing costs and promoting higher order of thought. ICT also faced some difficulties such as inadequate technological infrastructure and insufficient computer experience of students.
A higher focus on digital learning is associated with improved attitudes towards change and more innovative behaviours. Aboobaker and Zakkariya [2] highlight the need to strengthen the role of digital orientation in teaching and learning in order to transform educational institutions that are sustainable in terms of preparing graduates, ready for change and innovative behaviour at work, in the context of the emerging digital economy.

Education is a process aimed at finding new knowledge, including finding alternative ways in the field of new technologies that serve to meet special educational needs [3]. Society requires these technological advances to solve problems and enable humans to work with greater ergonomics; a school, a social institution, also needs these resources so that all students can build functional and meaningful teaching and learning process. The education system offers an education that meets the educational needs of all students; new technologies are a way to support diversity.

Blended learning supports or increases access for most student cohorts and yields higher achievement rates for both minority and non-minority students. Characteristics that students consider important are associated with clearly establishing and progressing towards course objectives, creating an effective learning environment and effective communication between teachers [4]. If in their opinion, these three elements of the course are satisfied, they are almost guaranteed to rate their educational experience as excellent, regardless of most other considerations.

Transforming the learning environment is often synonymous with acceptance and continued attention to the potential benefits of online learning in the higher education sector. The blackboard learning management system was piloted and implemented using a top-down approach of integrated training for faculty, students, and support staff. Based on data from interviews with participants, the study [5] emphasizes the need to strengthen academic support for the design of online learning and increase the focus on the development of effective teaching practices among employees, while trying to understand how scientists perceive and interpret the role of online technologies in supporting effective teaching practices.

To meet the needs of a new generation of students, higher education institutions are increasingly using digital tools such as virtual learning environments (VLE) and social media (SM). Research-based on the theory of service productivity [6] finds that learning-oriented outcomes are most important even when digital technologies are not used, and these results are further improved when students use the VLE. Students tend to prioritize knowledge transfer results.

The use of inverted scenarios in the classroom, with increased attention to solving specific sets of problems, is presented on the example of a course in mechanical engineering [7]. The centrepiece of the course is the universities' own implementation of the moodle learning management system. On the one hand, it provides all the general information such as a detailed curriculum, organizational information, as well as a grading system for the course and an organizational discussion board. On the other hand, it contains all the thematic information. The necessary theoretical input is provided in the form of wiki pages and video lectures, and problem sets are available as exercises in Moodle. In addition, a discussion board is available for thematic issues.

This group of researchers [8], [9] conducted a comparative analysis of conventional and electronic assessments in the educational process. The sample problems have been designed to accommodate a wide variety of inputs, from graphics to numeric and algebraic, and string input types. By implementing random variables, it is even possible to create an individual set of seed values for each participant. In addition, when working with complex problem examples, you need be aware of the transferred errors. To shorten the time it takes to give marks, the exam procedure consists of an e-assessment part and a classic paper and pencil part. The results of the electronic assessment and general examination were studied statistically, data were collected over several years. A clear correlation was found between the scores obtained on the electronic assessment and the classical one.

A new system for generating and modelling tasks in real-time is presented [10]. The use of modern principles of object-oriented programming and reflection-oriented programming allows real-time analysis to be divided into subsystems, where each such subsystem can be implemented as a runtime plugin that can be independently developed by different research groups. This method is intended to save a significant amount of time spent on validating results, as well as to provide peer reviewers with a more efficient review.

Abe et al [11] proposes a strategy to support the automatic creation and validation of tasks. The importance of supporting automatic task creation is to reduce teacher effort and personalize e-learning tasks for students, enhancing their understanding of the subject. In addition, an automatic verification strategy provides immediate feedback. The approach is based on the standardization of learning objectives by providing a formal definition of the structure of learning activities.

The creation of open datasets can accelerate the progress of research by allowing researchers to focus on developing and validating analytical methods rather than obtaining data. Open datasets also allow researchers to compare new analytical approaches with known standards and improve research.
reproducibility. It is proposed to use synthetic data generators to create open-access versions of student data [12]. Synthetic datasets take precedence over real datasets because private student data is protected by federal laws.

Personalization of online courses by context is always limited to existing teaching material; their creation is a laborious task. A conveyor for generating questions and correct answers based on educational texts, limited to actual questions for given sentences, is presented [13]. The methodology commonly used in bioinformatics is adapted to generate question and answer pairs. The system generates questions and related answers based on suggestions, 70% of which make sense. Teachers can suggest natural language corrections.

The system of intelligent formal reasoning and verification [14] has high efficiency due to the formal description of the formal proof and the regular matching algorithm after the introduction of the machine learning algorithm. Experimental results show that the system can check the correctness of logical reasoning of statements and reuse the results of logical reasoning of statements in order to obtain implicit knowledge in the knowledge base and provide a basic reasoning model for building an intelligent system.

Work in artificial intelligence has shown that rule induction is useful in gaining knowledge, but that induced rules can be difficult to understand and change. Terheyden and Chalcraft [15] describes a computer program for creating knowledge bases from examples in a form that can be interpreted either as a set of rules or as an inference network. The rules are easy to understand, so the structure can be changed by an expert into a form that the program will re-adjust to match the examples. This new data/knowledge analysis tool combines two very different methods, inductive and deductive, that are used in building expert systems.

The use of classification trees in various fields of application is presented [16]. Supplementing the direct use of induction with the use of forms of deductive (expert) knowledge is considered. Expert knowledge in the form of rules from human experts is used to improve the construction of a classification tree by supplementing inductive knowledge from examples when choosing the next node to add to the tree.

The idea of managing the learning process in the e-learning system is considered [17]. This study uses a personalized adaptive eLearning system that includes three developed theme sequences: teacher, student, or optimal theme sequences. The analysis showed that just over half of the students used the sequence of the teacher's topics; higher grades on topics were received by those students who chose the student or the optimal sequence of topics. This article proposes an algorithm for the development of the recommended learning path. Course topics and links between them are described using a weighted directed graph. The weight of each edge and vertex of the graph is calculated based on the values of the parameters describing the topic. Subsequently, it is assumed that the recommended learning path is the path with the least weight found in the weighted directed graph using search.

Through a verifiable experiment [18], significant statistics have been found that suggest that anthropomorphism in the user interface in the context of using online systems is more effective than non-anthropomorphic feedback. This will lead to better user interfaces by making them more user-friendly, more efficient, and more accessible to everyone.

There is a growing perception that academic institutions are not only providers of knowledge but also cultural agents. They must develop new skills in students. These include real-time problem solving, decision making, independent learning, knowledge synthesis and the daily challenges of an ever-changing New World, and the development of critical thinking and self-esteem. To stay relevant, the academic world must incorporate innovative content and learning paradigms that adapt to these changes, rather than sticking to traditional online learning methods. Schneider and Meirovich [19] describes the implementation of a unique student-centred teaching methodology that is studied and assessed digitally. The use of SGL teaching methodologies in targeting students on digital platforms allows for a significant degree of interaction between the interfaces - student-teacher, student-student, and student course content. This interaction provides a better learning experience and promotes safety and a digital tool experience.

One of the most significant features of e-education systems is the increased requirements for educational content. To implement an adaptive learning environment, not only large volumes of educational materials are needed, but also qualitative changes: greater variety, structuring by topic, complexity, and other characteristics, and, at the same time, the material must be methodically homogeneous. Meanwhile, existing content generation systems, as a rule, modify the traditional content retrieved from databases but do not produce qualitatively new content. To obtain such content, a fundamentally different approach is needed, based, for example, on imitation-ontological modelling and the creation of intelligent generators of knowledge [20]-[24].

Summarizing the above, it becomes clear that most approaches to the formation of e-education systems are of an anthropomorphic nature. This applies to all aspects of e-education: the content of training, the educational process itself, the concepts and technologies used. For example, intelligent systems are usually created on the basis of neural networks operating in the “black box” mode. It is considered necessary to develop audiovisual systems in natural languages using logical, semantic and other approaches. It is

Trainable generator of educational content (Vladimir Rotkin)
generally accepted that anthropomorphism unambiguously improves the quality of the system. However, the training materials obtained in this way often contain significant errors and are subject to additional selection. In addition, such approaches require an increase in computational resources. It should be assumed that the issue of the optimal level of anthropomorphism has not been sufficiently studied and requires a careful approach.

Considering that, unlike the field of scientific research, educational systems are repeatedly tested with stable educational material, it is possible to successfully apply deductive (analytical) types of AI-based on general mathematical models. This approach allows you to: get an unlimited number of diverse educational tasks, structure them according to any necessary criteria, and ensure methodological unity. Moreover, each operation is accurately identified by a unique set of calculated variables (parameters). In practice, this represents a simpler equivalent of the pattern recognition procedure in inductive artificial intelligence systems using neural networks and machine learning. This approach assumes the use of simple and intuitive matrix forms as an interface.

2. RESEARCH METHOD

The methodology of the formation of an educational system based on a deductive intellectual generator of knowledge is considered. The system assumes random generation of educational tasks, text input of solutions, and their verification through the assessment of the accuracy of the results. It also provides intellectual support for users and training the system by forming the element base of reference solutions.

2.1. Content generation

Training tasks are generated on the basis of a simulation model that determines the relationship between the main variables (parameters) of the general task X.

\[ F_1 (X_1 \ldots X_j \ldots X_J) = 0; \]
\[ F_K (X_1 \ldots X_j \ldots X_J) = 0; \]
\[ F_K (X_1 \ldots X_j \ldots X_J) = 0. \] (1)

Here J is the number of basic variables (X), K is the number of connections between variables. Accordingly, the dimension of the system is equal to J-K, that is, by specifying K variables in different combinations, as the initial ones, you can calculate the rest of the J-K calculated variables. Thus, combinations of J elements from K determine the number of possible particular problems based on the general model (problem). In turn, each particular problem can be generated in a variety of random variants if the values of its initial parameters are set randomly. The process of generating partial problems is based on the formation of an array of random values of the main variables as shown in Table 1.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sign</th>
<th>Operation</th>
<th>Data kinds</th>
</tr>
</thead>
<tbody>
<tr>
<td>NameR(i;1)</td>
<td>R(i;1)</td>
<td>RANDGEN(i;1)</td>
<td>Basic initial data</td>
</tr>
<tr>
<td>NameR(i;K)</td>
<td>R(i;K)</td>
<td>RANDGEN(i;K)</td>
<td>Basic calculated data</td>
</tr>
<tr>
<td>NameR(i;K+1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NameR(i;j)</td>
<td>R(i;j)</td>
<td>P_1-k(R(i;1);\ldots;R(i;K))</td>
<td></td>
</tr>
<tr>
<td>NameR(i;J)</td>
<td>R(i;J)</td>
<td>P_1-k(R(i;1);\ldots;R(i;K))</td>
<td></td>
</tr>
</tbody>
</table>

The initial data of the main array is generated by the random operator RANDGEN, in which the i index defines the current number of the value of the original variable in a random sequence. The corresponding values of the remaining variables are calculated using explicitly defined functions P. The resulting main array serves as a source for the formation of operational initial data for particular tasks, see Table 2. Using the RANDSEL (1; J) operator, a random selection of K non-repeating R variables from J variables of the main array is performed. The remaining J-K variables serve as the source of the main reference results (answers) for the respective tasks.
2.2. Matrix interface

In the process of solving the problem, the user interacts with the system through the interactive cells of the operational matrix. Operational calculated data \( Z \) (names of variables, calculation-and-logical formulas \( Q() \)) are sequentially entered by text form using standard operators.

The user independently determines the names of variables, their number and sequence, forming an operational chain, where each calculated variable is expressed through the previous calculated and initial variables. Data can be entered both in formulaic and numerical form, with an arbitrary combination of main, intermediate and additional variables, which provides high flexibility of interaction with the interface.

2.3. Data identification

Recognition and verification of the input data are carried out by converting the texts into a numerical format and comparing the obtained numerical values of \( Z \) with the corresponding values of \( Z_0 \) from the reference database. To ensure the reliability of this procedure, for each variable, a comparison is made of the sets of random values of this variable \( i = (1 \ldots I) \). Each such set \( \{Z(1;i) \ldots Z(I;i)\} \) acts as a unique identifier of the corresponding operator \( Q(i) \).

The modified root-mean-square value of its relative deviations from each reference value \( Z_0 \) is used as an indicator of the identification of the variable \( Z_0 \).

\[
\delta_{nj} = \sqrt{\frac{1}{I} \sum_{i=1}^{I} \left( \frac{Z(i;n) - Z_0(i;j)}{\max(\{Z(i;n)|Z_0(i;j)\})} \right)^2}
\]

Here \( I \) is the number of random values of \( Z \) and \( Z_0 \) in the compared sets. The difference between these values is divided by the maximum of the absolute values of these two values.

Of all the values of \( \delta_{nj} \), the minimum is selected, and if it does not exceed the specified limit value \( \delta_0 \), then the following are fixed: the reference variable \( Z_0(i;j) \), the value of the relative deviation \( \delta_n = \delta_{nj} \), and the values of \( Z \) and \( Z_0 \) (and the corresponding operators) are recognized as identical (equivalent).

2.4. Machine learning

Considering that users are given the opportunity to build unique calculation sequences using their own variables and parameters, it should be recognized that the array of basic variables is insufficient to identify custom operators. Obviously, the base of reference data needs constant replenishment. The natural source of updating this database is the solutions to the tasks performed by the users. If in solving the task the user correctly calculated the variables from the main array, but at the same time there are intermediate variables that do not have reference equivalents, then such variables can be attached to the reference database. To improve reliability, preliminary verification (moderation) of new reference data is required. Such a check can be carried out both automatically, by special algorithms, and manually. Thus, the system improves its ability to recognize the entered operators, perceiving new information from students (users).

2.5. Help and support

The presence of a replenished reference base of operators makes it possible not only to identify the entered operators and evaluate the user's actions but also to provide him with methodological support, providing feedback with the system. For this, an algorithm is used that, not only using the identification mechanism, selects a reference operator equivalent to the operator marked by the user, but also fixes the corresponding operational chain. The entire sequence of reference operators’ chain is represented in a special samples matrix. This matrix serves as an auxiliary interface for the methodological support of the user.
3. RESULTS AND DISCUSSION

The order and results of the implementation of the content generation methodology are considered in a specific example, followed by an analysis of the ways and prospects for expanding and modifying the generator model.

3.1. Implementation of a content generator using a tutorial topic as an example

The current layout of the educational content generator is implemented on a simple example on the topic from the school mathematics course "Vector in a flat coordinate system". The design model and the corresponding main array of variables are formed from the following data that determine the configuration options for the tasks as shown in Figure 1.

Initial data:

coordinates of point A - XA; YA;
length (modulus) of vector AB - AB;
the angle between the axis OX and the vector AB - α.

Calculated data:

the cosine of the angle between the OX axis and the AB vector - cosα;
the sine of the angle between the OX axis and the AB vector - sinα;
the projection of the AB vector onto the OX axis - \(\mathbf{AB}_X = AB \cdot \cos \alpha\);
the projection of the AB vector onto the OY axis - \(\mathbf{AB}_Y = AB \cdot \sin \alpha\);
coordinates of point B -
\[X_B = X_A + AB \cdot \cos \alpha;\]
\[Y_B = X_A + AB \cdot \sin \alpha.\]

Figure 1. A vector in a flat coordinate system: random configurations
In the administrative (moderated) interface of the generator, see Figure 2, the names and designations of the main data are presented in the “glossary” table. On the left is a random geometric configuration of the object and the text of the educational task. This interface is intended to form the initial reference base of tasks. The administrator (moderator) enters the initial data into the working matrix, in the form of links to the names of the main variables from the ”glossary” table. Then the solution of the task is introduced in the form of a sequence of operators of calculated variables. If the reference base already contains the previous data, then the operations can be controlled by the values of the relative percentage deviations. Where there is no reference variable or the deviation does not fit into the limit value, a gap is fixed in the matrix.

Using the ”variant” counter, you can arbitrarily change the values of the initial data, controlling the corresponding deviations. In addition, the ”line” counter allows you to find operators in the reference database that are equivalent to the operator entered by the user in the specified row of the working matrix, and to display the corresponding calculated chains of reference operators in the samples matrix. After the initial data and calculated operators are entered and the solution is verified, the task is sent to the reference database by activating the ”save” option.

![Figure 2. Administrative interface "Moderator"](image)

The ”Instructor” training interface is formed by reducing the administrative interface (Figure 3). It differs in that the initial data of the tasks are not entered by the user, but are retrieved from the reference database using the “tasks” counter and are presented in numerical form. The rest of the interface works similarly to the ”moderator”. Problem solutions from the ”instructor” interface can also be forwarded to the reference database. To do this, open the task in the ”moderator” and after verifying the solution, use the ”save” button.

Further reduction of the interface, by the way of exception of the graphic image of the object and the samples matrix, gives a version for controlling the user's learning skills-the control interface "Student" as shown in Figure 4. To complicate the task, the ”Deviation” option can also be excluded.

The functional of the proposed training system based on a simulation content generator, presented, in particular, in the form of a demonstration layout [25], indicates that deductive non-anthropomorphic approaches in the creation of intelligent systems can compete with mainstream developments in the form of trained neural and logical-semantic networks, ontological systems, and so on. Such approaches make it possible to create full-fledged training systems with the necessary sets of intelligent options relatively simple means, without the formation of databases, with a minimum cost of computing and other resources.
3.2. Promising modifications of training content generators

The considered sample of the generator, although it is the simplest example of a training object, contains a significant set of capabilities: generating tasks with various object configurations and random variants of these tasks, the possibility of text input of solution components, recognition and verification of entered operators, interactive user support through the provision of sample solutions, elements of machine learning by replenishing the generator with components of new unusual solutions.

The methodological and technological solutions used make it possible to develop and improve educational generators in the following ways.

The simplest approach involves replicating generator versions by replacing the training object and, accordingly, its simulation model. This is done through the modification or replacement of some of the algorithmic modules. The accumulation of generator versions allows for the expansion of educational content.

The hierarchical approach is to move from modelling particular problems to more general ones. For example, in the plane, not one, but two vectors are specified, and then the following are determined: their sum, difference, scalar product, and other operations. You can go beyond the plane by going, for example, to...
vector or mixed products. This approach allows you to significantly expand the generation of both general and specific tasks.

Improvement of the feedback between the user and the system is needed. For example, in addition to the deviation value, the system must notify the user about the incorrect input of a typical operator (formula). The perspectives of machine learning are of considerable interest. At the initial stage, manual moderation of user solutions is applied before they are sent to the reference base. It is necessary to develop and improve special algorithms to automate moderation. The use of neural network AI systems is not excluded.

4. CONCLUSION

The deductive approach based on the simulation of educational objects allows you to create a unified algorithmic platform that combines the functions of a content generation with educational training. In comparison with anthropomorphic logical, ontological, semantic methods of content formation, simulating training generators demonstrate a number of advantages. They are capable of producing, online, an unlimited number of extremely varied tasks, together with solutions. These teaching materials are completely reliable and do not need any additional selection. The ability for the user to arbitrarily define the form and sequence of operations in solving problems gives the system high flexibility and variability. The representation of tasks in the form of sets of operators with numerical values, in combination with the random formation of the initial data, provides high accuracy of verification of user operations based on the comparison of integral numerical identifiers. Since the comparison is made with the reference solution base, the replenishment of this base from new individual solutions is a unique option for teaching the system based on its interaction with students.

REFERENCES


