Mathematical Model of Development of an Innovative Education System

Tatiana V. Krupa
GlobalLab, LLC, Moscow, Russia. Email: t.krupa@globallab.org

Abstract
Due to in connection with the development of science in all countries there is a demand for young scientists who will be able to conduct successful research and discoveries. The educational system has the goal of organizing educational processes in such a way that students can show better academic performance, and teachers can identify the gap in the student's education and fill it.
This research provides a method for developing a mathematical model of the trajectory of the student learning the proposed model was developed taking into account external variables that reflected the students’ characteristics in turn influencing the learning process and the education. Such student mathematical model for the individual training will decrease the cost for this type of training and at the same time will give all the benefits of the individualized training. In addition, the use of the neural networks to research and predict the properties of the educational trajectory will make it possible to discover the new and effective research methods in the field of learning theory.

Keywords: education, learning trajectory, student, educational process, mathematical model

I. INTRODUCTION
Well-educated people promote the economic growth and development of the country and should be considered as a national treasure of a country. The present status of the educational system has such characteristics: research, development, design and realization of the educational conditions. The mentioned characteristics will allow the students and pupils to develop their skills. The educational system should be aimed at the full personality development within the limits of his or her cognitive resources.
What is more, it is crucial to enrich the system with information-communicative technologies (ICT) not only due their usage in every innovative field, especially connected with physics [23], but also to encourage the future highly educated individuals to use ICT for staying home in pandemic conditions [24].
Conventional format and teaching methods are focused on the passive interaction with the needed and extra material and require from the student perseverance for self-realization. This affects not only the individual’s development, but the social progress as a whole, and the level
and effectiveness of scientific innovations in particular, as is shown by the urgency the countries falling behind in the use of ICT are now introducing the national projects for scientific development [25]. It goes without saying that individual education considering the individual propensity of a child is considered as to be perfect for the creation of students’ set of skills and self-development.

The educational system that we have is sometimes the result of a negative learning experience in educational institutions. Among the consequences, low academic performance and deviant behavior of the child [2], as well as affect the child’s psychological health due to the raise of anxiety [22]. So, for such reasons the individual teaching is a more beneficial teaching method.

The purpose of the research is to develop the mathematical model of the student’s progress in order to determine the best educational trajectory.

As many pedagogical studies show, The use of machine learning methods through full or partial automation will help reduce costs for the formation of personal training trajectories and improve academic performance.

II. LITERATURE REVIEW

Problematic of individual learning trajectory (ILT) is the subject of research by scientists such as Aleksandrova, Berezhnaya, Vdovina [1], Gardner, Maskaeva, Ulanovskaya, and Sysoev [2].

In the works of such researchers as Boirdovskiy, Vdovina, Klimov, Merlin, Surtaeva, Yakimanskaya, the individual learning trajectory is understood as a concept that has a few ways of the implementation: cognitive (the school curriculum plans and educational programs, etc), active (special teaching methods and technologies) and procedural (the organization of educational process).

such scientists as Smith, Wiser, Anderson, Krajcik, Merzliakov and Podoliak give such a definition of a learning trajectory as “descriptions of successfully more sophisticated ways of reasoning within a content domain based on research syntheses and conceptual analyses” [3, 18].

For science education in the context of the term "learning trajectory" it is used the term “learning progressions [4-6].

The progress of the learning trajectories is tied with the constructivism, socio-constructivism, and/or socio-cultural views of learning and should be based on the cultural inputs and the children’s background [7-10, 20, 21].

III. METHODOLOGY

In this work, there are used such research methods as literature review on the problem, deduction, induction, comparison and analogy to study the problem from the teacher to the children way and vice versa to find out the mathematical model of the children’s development.

During the mathematical modeling the following softwares were used: Jupyter Notebook and GNU Octave.
IV. RESULTS
Among data that must be taken into account when developing the mathematical model of the learning trajectory two types can be distinguished:
– the external variables;
– the internal variables.
The internal variables characterize the learning process as such: school plans, educational programs, teaching methods and technologies, etc.
The external variables depict the student’s characteristics that are not related to the choice of the learning trajectory. These are such characteristics as gender, style of learning, geographical location and the age, etc. The external variables and the methods of their value calculation are the mathematic model of a student as a subject of learning trajectory. The external data for building a model were taken from the platform Global Lab and Dnevnik.ru.
To develop the mathematical model of the learning trajectory, all external variables that could affect the learning process were analyzed.

A. Geographical Location
As a whole, the quality of the knowledge getting by the students depends on the place of living and school location. In big cities, schools have better equipment and teachers are more qualified. However, the study of Manpower Demonstration Research Corporation (MDRC) has shown that the small schools in poorer areas perform better in student education than large schools in large cities, explain it with the fact of a more favorable environment in schools [11].
In order to analyze the geographical location data that can be used in the method developing there were considered such notions as the latitude and longitude of the place of living and the hierarchy of administrative units that was made on the data taken from the such geolocation services as Google Geocoding and OpenStreetMaps Nominatim. Such data criterias as the subject of the Russian Federation and the name of the municipality to which the educational institution belongs and the city where the educational institution is located were also used for the students who live on the territory of the Russian Federation.
The geolocation variable is one of the properties that are independent of the student’s current residence, and show some relative values. The study offered the following hypothetical external variables:
- $U_{\text{geoAdmLv1}}$ means and administrative level to which the student's place of residence, expressed as an integer: 1 is the capital city or big city that is informal capital of a region (for example: in the Russian Federation they can be Moscow, Saint Petersburg, Yekaterinburg and Novosibirsk, etc; in the USA – New-York, Washington, etc; in China: – Beijing, Shanghai and Chongqing, etc),
- 2 is a big city, capital of second level administrative division (the subject of the federation in the Russian Federation, state – in the USA, provinces – in China),
• 3 is a city that is the regional center of the administrative division of the third level city is the regional center of the administrative division of the third level (district center in the Russian Federation, county – in the USA, district – in China),
• is all the rest cities, towns and villages. In case if the geographical position of a student is not defined or to get the information about the administrative level is not possible, the variable takes on value 0;

1. \(U_{lat}\) means the latitude at which a student resides with the accuracy up to 0.1 degrees. Indirectly and imprecisely the latitude determines the climatic conditions in which a student lives. In addition if the other information is lack it can reflect the serious differences in the structure of the school year because the students may live in the different hemispheres. for whom the forecast of the optimal learning trajectory is made. In case if the geographical position of a pupil is not defined the variable takes on value 91;

2. \(U_{urb}\) means a sign of belonging of the settlement, where the student lives, to the type of urban settlements. By implication, this variable reflects the student’s access to cultural and educational infrastructure in the city: the theaters, museums and different centers of the additional education. Likewise, such variable marginally demonstrates the limited number of direct communication. The variable has the value 1 if a student lives in an urban settlement, 2 is rural settlement, 0 means that the geographical location of a student is not defined or it is not possible to get the information about the settlement type.

After data processing from GlobalLab and Dnevnik.ru sources there were found out the following geolocation records:
• UgeoAdmLvl – 44 thousand records;
• Ulat – 52 thousand records;
• Uurb – 39 thousand records.

B. Age and Gender

The scientists around the world offer various explanations how the age and gender on the students’ can determine their success in the education. Some researches show that at particular subjects the female students are more demonstrate higher performance than the male but so far there is no common conclusion about the relation between the gender and educational performance. The female success in the hard sciences can be explained by their thoughtfulness and persistence [12].

The school age is divided into three groups: primary (from 7 to 12 years), middle (from 12 to 15) and senior (from 15 to 18). The gender is divided into groups: male and female.

During the research the next supposed external variables were proposed:
1) \(U_{age}\) means the student’s age that he/she has had at the time of the variable value calculation and expressed as an integer. The neutral variable value is 0;

\(^1\) The value of a variable that indicates the unavailability of the source data is called neutral.
2) $U_{\text{agePrecis}}$ means the age accuracy. The variable gets the value 1 if the age is known with the accuracy up to 1 year; it means 2 if the age is known with the accuracy up to >1 year (the age has been calculated on the data about the school age). The neutral variable value is 0;
3) $U_{\text{sex}}$ means the student's gender.
1 means female;
2 means male;
0 is neutral value.

After data processing from GlobalLab and Dnevnik.ru sources the next age and gender records were found out:

- $U_{\text{age}}$ – 58 thousand records;
- $U_{\text{agePrecis}}$ – 58 thousand records (46 thousand records with the notation that is equal to 1, 12 thousand with the notation that is equal to 2);
- $U_{\text{sex}}$ – 24 thousand records.

C. Style and learning

The style of teaching and learning dictates the learning success of each student. To accomplish this goal, the style of learning should accompany the context of the material and must be suitable for each student. The learning style is the way how each student processes and incorporates received information [13].

For this research, the procedure of the data gathering about the learning styles of the students was developed based on the students’ interface interactions with the Global Lab platform.

At the input, the procedure takes a lot of diverse notifications on the actions of a certain specific user in the interface and, in the output, it looks like a vector that describes a user in the aggregated form. Many vectors of different users therefore have been used as the input data for clustering the C-average algorithm and the output is the assignment of a user to one of the C clusters and identifies users by their behavior.

To obtain the hypothetical external variables there were determined following steps for the learning style data gathering:

1) creating a research project;
2) the research project idea creation;
3) filling out a research questionnaire and sending it to the project;
4) acquaintance with additional non-interactive educational materials that accompanies the GlobalLab courses and projects (the text materials, videos and presentations, etc).

For each student the vector calculation of $\vec{S}$ type was developed based on each act committed by a student. Additionally, for the set of vectors $(\vec{S}^{(1)}, \vec{S}^{(2)}, \vec{S}^{(3)} \ldots \vec{S}^{(m)})$ for all m of the users, who have ever done some actions during the sessions of the given type there is done the division on the C clusters $C = (C_1, C_2, C_3 \ldots C_k)$, where in each of them it is minimized the total quadratic deviation of vectors $\vec{S}$ from the centroid of the cluster (1).

$$\min \left[ \sum_{i=1}^{\infty} \sum_{\omega} \left\| \vec{S}^{(\omega)} - \vec{c}_{i} \right\| \right]$$

(1)
where $c_i$ is the centroid of the cluster $c_i$

Each vector $\vec{s}$ meets the cluster serial number $c_i$ for which the expression value $\|\vec{s} - c_i\|^2$ is minimal. This match $c$ is denoted below as a function $k_{\text{means}}$.

The number of the C clusters was changed in the range from 2 to 7. Changing the range of cluster numbers is based on the thesis that a user behavior in the e-learning environment should be different for 3–5 learning styles. The maximum classification accuracy (the average for all the session types is 91.2%) was achieved during $C=5$. Consequently, it is a reasonable assumption that for each type of a session each user maybe assigned to one of the 5 teaching formats. Following this thesis in addition to neutral, it has been identified 4 variables of learning style:

1) $U_{\text{behPrj}}$ is the interaction class in a research project creation session (0 is neutral value, 1–5 is class);
2) $U_{\text{behIdea}}$ is the interaction class within the framework of the idea creation session;
3) $U_{\text{behData}}$ is the class of interaction in the framework of the research questionnaire filling session;
4) $U_{\text{behStd}}$ is the class of interaction in the framework of the educational materials work session.

Each variable value has got through (2):

$$U_{\text{beh}} = k_{\text{mean}}(\vec{s}_{\text{beh}}) \quad (2)$$

where $\vec{s}_{\text{beh}}$ is the vector $\vec{s}$ for the given session type.

For getting the variable $U_{\text{style}}$, indicating the student’s learning format, the clustering algorithm of the C-average was used on the variable vector $\{U_{\text{behPrj}}, U_{\text{behData}}, U_{\text{behStd}}\}$ with $C=4$ (3).

$$U_{\text{style}} = k_{\text{mean}}(k_{\text{mean}}(\vec{s}_{\text{behPrj}}), k_{\text{mean}}(\vec{s}_{\text{behData}}), k_{\text{mean}}(\vec{s}_{\text{behStd}})) \quad (3)$$

After data processing data from GlobalLab and Dnevnik.ru sources the following style of learning records were discovered:

- $U_{\text{behPrj}}$ – 26.8 thousand records;
- $U_{\text{behIdea}}$ – 10.2 thousand records;
- $U_{\text{behData}}$ – 27 thousand records;
- $U_{\text{behStd}}$ – 23.8 thousand records.

D. Academic performance (grades, test results)

The quality of educational processes in schools and universities affects the academic performance of students. Moreover, it gives a hint to the pupils what subjects they have learned better and under what conditions they obtained better skills to continue their learning process in the higher education institutions [14].

Exam grading systems and current grades differ from country to country. In this research the grades on the several school subjects of the pupils from different forms were used:

- Mathematics (Algebra for the higher school);
- Russian language and literature;
- Environment (for the pupils of 1-4 forms);
- History and Biology (starting from the 5th form);
– Physics (starting from the 7th form);
– Chemistry (starting from the 8th form);
– Technologies (for the pupils of 1-7th forms).

All the subjects were divided into three main groups:
1) The subjects of the humanitarian cycle – Russian language and literature, history;
2) The subjects of the natural science cycle and mathematics – the environment, biology, physics, chemistry and mathematics;
3) Technology.

Additionally, three hypothetical external variables were given:
1) $U_{\text{assessHum}}$ is the adapted to 100-point scale average grade on the humanitarian subjects;
2) $U_{\text{assessScience}}$ is the adapted to 100-point scale average grade on the subjects of the natural science cycle and Mathematics;
3) $U_{\text{assessTec}}$ is the adapted to 100-point scale average grade on the technical subjects.

The second part of the study of the academic performance was devoted to data collecting the test’s grades. For the research 60 various mini-tests with 10 questions on each subject were proposed. For the calculation of the result the time that a student used for passing a test was taken into account.

In this way there were found out the following hypothetical external variables:
1) $U_{\text{testHum}}$ is the adapted to 100-point scale average grade on the humanitarian subjects;
2) $U_{\text{testScience}}$ is the adapted to 100-point scale average grades on the subjects of the natural science cycle and Mathematics;
3) $U_{\text{testTec}}$ is the adapted to 100-point scale average grade on the technical subjects.

The neutral value for all the variables is 0.

The number of the pupils who took part in the tests was 27 thousand people.

E. Proficiency level in complex skills
To discover the proficiency level of the pupils’ complex skills the express method for evaluating the pupils’ skills in the field of project activities was implemented. In order for the cogneme experiment to be more suitable for pupils, it was conducted in a game form. The cogneme is a part of the knowledge which is in a fixed dynamic process of five parameters: value – path – word (sign) – area – function [15]. The brains perceive the information (word or sign) in the simplest way – by listening, touching, aromas and the other ways of sensing perception, learns the meaning and the area of its use and makes the operation of a body [16-19].

Before the experiment the pupils received a questionnaire with 100 stimulus words and then within 7-10 minutes, they completed them. The aim was to write the word that first came to mind as related to a word stimulus. The incentive words were selected from an area related to a variety of non-cognitive skills. For instance, the following words were chosen for the
cooperation: question, answer, plan, opinion, conflict and dispute, help and interaction, etc. Each stimulus word had its own number and the composition of the questionnaire was formed on the basis of the random number generator to avoid the situation of two same questionnaires. Upon completion the experiment the associative-verbal net was compiled for each student. It is a graph in which the nodes are the language units (LU) and the ribs are a fact of the LU speaker’s association. The of a rib length reflects the frequency of the association.

Using the C-average algorithm of the students were grouped by the mean connectivity property of their associative-verbal graphs with C=7. The user affiliation to one of 7 classes is indicated by the variable $U_{\text{noncog}}$, getting the value from 0 to 7, 0 is neutral value.

For some students, who have the teacher binding on the Global Lab platform, it was possible to find out the their teachers' level of the usage of innovative technologies. This level was expressed through the variable $U_{\text{noncogTutor}}$, getting the same values in the range 0–5, in accordance with the quantity of the innovation groups on which the teachers were divided. The neutral value for all the variables is 0.

Concluding all the hypothetical external variables that have been got during the research there were built the student mathematical model (4):

$$\bar{u} = \{U_{\text{geoAdmLvl}}, U_{\text{lat}}, U_{\text{age}}, U_{\text{assessHum}}, U_{\text{assessScience}}, U_{\text{assessTec}}, U_{\text{testHum}}, U_{\text{testScience}}, U_{\text{testTec}}, U_{\text{urb}}, U_{\text{agePrecis}}, U_{\text{sex}}, U_{\text{assessHum}}, U_{\text{assessScience}}, U_{\text{assessTec}}, U_{\text{testHum}}, U_{\text{testScience}}, U_{\text{testTec}}, U_{\text{urb}}, U_{\text{agePrecis}}, U_{\text{sex}}, k_{\text{means}}(\bar{S}_{\text{behPrj}}), k_{\text{means}}(\bar{S}_{\text{behData}}), k_{\text{means}}(\bar{S}_{\text{behStd}}), k_{\text{means}}(k_{\text{means}}(\bar{S}_{\text{behPrj}})), k_{\text{means}}(\bar{S}_{\text{behData}}), k_{\text{means}}(\bar{S}_{\text{behStd}}), U_{\text{noncog}}, U_{\text{noncogTutor}}\}$$

(4)

The vector $\bar{u}$ models a student as a subject of a learning trajectory, since it can describe him/her separately from the specific steps of the trajectory. The model can be considered as a linear as all the methods of modeling the user’s characteristics are linear too. Although the C-average method in the model was used that presuppose initialization of centroids by random variables the student’s model may be considered as a deterministic one. This result is explained by the fact that the formula (4) used after the C-average algorithm usage along the entire data array that causes getting always the same vector $\bar{u}$.

The implementation of the neural networks to research and prognostication the features of the learning trajectory will make obtainable the new and effective methods of research organizations and teams in the field of the educational theory: the modeling of cognitive process that is connected with the development of the different students’ skills; and the development of the more precise evaluation methods of the students’ academic performance.

With the further progression the model of the learning trajectory will help to adopt the educational system of any country for the individual needs of the students to have comprehensive access to the educational tools to enrich students’ skills in the general teaching and within the educational system. The cooperation of the students with the instructors within this learning trajectory could positively affect both sides of the educational process. Such fruitful cooperation will lead to the enhancement of the educational system, exploring the best methods of teaching the students that will help them to assimilate the information more
effective not only on the passive level but also will use it actively in the continuing their education.

V. DISCUSSION

The study shows, this list mentioned above was studied in terms of the prospect of synthesis certain experiments and annexing data extracted from the electronic diary of Diary.ru to them. To the second group of events from the data obtained from Diary.ru, an event of acquiring a rating was attached. Similarly, 2 types of events were added, reflecting the results of testing on the GlobalLab platform or completing tasks in GlobalLab courses. Such events are counted successful if the grade exceeds 65 of 100-point scale. Adding these 2 types of events allows to reflect changes in the student's academic performance in the mathematical model of the educational trajectory, which in turn leads to an increase in its accuracy and representativeness.

In the first group, the next events were combined:
- The user creates a school and the user creates a group;
- The user edits the school and the user edits the group;
- The user sends an invitation to another user, an invitation to a group, an invitation to school;
- The user comments on the blog post, group post, school post.

The differences between the School and Group entities on the GlobalLab platform are minor: frequently (~47% of cases) users use the School entity to simply bring together users who are not associated with the same school, and vice versa - use the Group to collaborate with users from the same class or school. This occasion allows make no distinction in the events connected with these entities, and combine them. Furthermore, such a combination allows to reduce the vector dimension representing the condition of the educational trajectory by 4 elements.

The event "The user selected an external link" was excepted from the composition of the events of the first group, since it is implicitly taken into consideration when data collecting on user interaction with the user interface.

According to the results of the changes described above, the final list of events that report the educational trajectory took the form presented in Table 1.

Table 1. List of event types taken into account by the mathematical model of the educational trajectory

<table>
<thead>
<tr>
<th>No.</th>
<th>Type of an event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>The user approves the project.</td>
</tr>
<tr>
<td>2</td>
<td>The user approves the idea.</td>
</tr>
<tr>
<td>3</td>
<td>User approves blog post.</td>
</tr>
<tr>
<td>4</td>
<td>User approves comment.</td>
</tr>
<tr>
<td>5</td>
<td>The user logged in to the system.</td>
</tr>
<tr>
<td>6</td>
<td>The user has activated.</td>
</tr>
<tr>
<td>7</td>
<td>User has registered.</td>
</tr>
</tbody>
</table>
The types of events form the Table 1 can be represented as types of steps of the educational trajectory. Therefore, any status of the user at each action of his educational trajectory is recorded only when one of the events in table 4 happens and, hence, the mathematical model of the educational trajectory considers only listed types of events.

The final list of events is taken into consideration by the mathematical model of the educational trajectory. The list of events makes it possible to raise the accuracy of the mathematical model of the educational trajectory with including in it a wide range of educational activities that are not commonly included in student models.

**VI. CONCLUSION**
During the study the student mathematical model (as a subject of learning trajectory) was elaborated. 43 variables that described the basic students’ qualities were considered.
Moreover, the proposed student’s model is functional and does not describe the whole complex of the complicated cognitive and other mental processes that occur in the student consciousness and conditioning and influencing on his/her characteristics. In this case the model describes the modeling object in a simplified form. However, the comparing to the students’ models that are historically used for the evaluation of their knowledge and skills (for example: the student’s model for defining the results of the unified state exam) is more complex. This is due to the consideration of many characteristics ranging from the geographical location to the compound skills development.

ACKNOWLEDGMENT

Applied research described in this paper is carried out with financial support of the state represented by the Ministry of Science and Higher Education of the Russian Federation under the Agreement #14.576.21.0100 of 26 September 2017 (unique identifier of applied research – RFMEFI57617X0100).

REFERENCES


particles”, Key Engineering Materials, 781: 3-7, 2018, DOI: 10.4028/www.scientific.net/KEM.781.3
