Development of an Innovative Education System Mathematical Model

Krupa Tatiana

Abstract: With the development of the scientific progress in the different spheres the countries all over the world need the highly educated and talented young people who will be able to develop the science and to achieve big results in it. Now, the major task of the educational system is to develop and implement such educational process that will allow achieving high academic results for the students and giving them the opportunity to find out at which sphere they will achieve the best results and for the teachers to find out the sphere of the lack of the knowledge.

In the current article, there is given the way to develop the student mathematical model (as a subject of the learning trajectory). The model was developed with taking into account the external variables that reflected the students’ characteristics influencing on the learning process and getting the education.

The usage of the student mathematical model for the individual education will allow reducing the expenditures for this kind of education and at the same time will give all the advantages of the individual studies. Also, the usage of the neural networks to study and predict the properties of the educational trajectory will provide the new and effective research methods in the field of learning theory.

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

I. INTRODUCTION

Nowadays well-educated people should be considered as a national wealth of a country as they move the economy growth and country development ahead. The current state of the educational system is characterized by the search, development, creation and implementation of the educational conditions that will allow the students and pupils to show and improve their skills. The educational system should provide the development of a personality in the maximum range of his or her intellectual and psychological resources.

Traditional forms and methods of teaching are directed on the passive assimilation of necessary and unnecessary knowledge and demand only the perseverance from a child without development a desire for the self-realization and activity. It is absolutely clear that taking into account the individual peculiarities of a child the individual education is considered to be ideal for the formation of his/her complex of skills and self-development.

The stereotype educational system and the stereotype situations can break the positive school experience that may lead to the reduction of the academic performance and negative behavior of a child [2]. So, in this situation the individual education also has a big advantage.

The aim of the work is to find out the mathematical model of the student’s development for the further definition of the optimal educational trajectory.

The machine learning methods usage will allow reducing the individual learning cost for the formation of the personal learning trajectories due to the full or partial automation. It will influence positively upon the level of the educational results as many domestic and foreign pedagogical studies show.

II. LITERATURE REVIEW

Studying and analyze of the scientific work and researches has shown that the questions of individual learning trajectory (ILT) are represented by the following authors: Aleksandrova, Berezhnaya, Vdovina [1], Gardner, Maskaeva, Ulanovskaya, and Sysoev [2].

Boidovskiy, Vdovina, Klimov, Merlin, Surtseva, Yakimanskaya consider the individual learning trajectory as a notion that has a few ways of the realization: informative (the school teaching plans and educational programs, etc), active (special teaching methods and technologies) and procedural (the educational process organization).

The foreign scientists define 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].

At the same time along with the term “learning trajectory” it is used the term “learning progressions” that is more typical for the science education [4-6].

The development of the learning trajectories is connected with the constructivism, socio-constructivism, and/or socio-cultural views of learning and should be based on the cultural tools and the background of the children [7-10, 20, 21].

III. METHODOLOGY

In the given work, there are used the following methods: analysis of the scientific literature on the problem, deduction, induction, comparison and analogy methods to study the problem from the teacher to the children way and from the children to teacher direction to find out the mathematical model of the children’ development.

During the mathematical modeling there were used Jupyter Notebook and GNU Octave software.
Development of an Innovative Education System Mathematical Model

IV. RESULTS

All the data that is need to take into account during the calculation of the mathematical model of the learning trajectory can be divided into two types:
- the external variables;
- the internal variables.

The internal variables are the ones that describe the learning process themselves: school plans, educational programs, teaching methods and technologies, etc.

The external variables are the ones that describe the student’s properties that are not dependent on the step inside the learning trajectory. They are the gender, style of learning, geographical location and the age of a student, etc. The external variables and the methods of their value calculation are the mathematical model of a student as a subject of learning trajectory. The external data for developing the learning trajectory method were taken from the platform Global Lab and Dnevnik.ru.

To develop the mathematical model of the learning trajectory there we analyze all the external variables that could influence upon the learning process.

A. Geographical Location

In general, the place of living and school location influences on the quality of the knowledge getting by the pupils. In big cities the schools are better equipped and the teachers have high qualification level. Still the research of Manpower Demonstration Research Corporation (MDRC) prove that the small size school in the poor areas show better results in pupils’ education than the big schools explain it with the fact of the better environment in the schools [11].

To analyze the geographical location data that can be used in the method developing there were taking into account 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 Google Geocoding or OpenStreetMaps Nominatim geolocation services. For the students who live on the territory of the Russian Federation there were also used such data criteria as the subject of the Russian Federation and the name of the municipality to which the educational institution belongs and the city where the student lives, to the type of urban settlements.

After processing the source data from GlobalLab and Dnevnik.ru there were found out the following the following geographic location records:

1)\(U_{geoAdmLv1}\) means and administrative level to which belong the place of student’s living expressed by the 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),
   4 is all the rest cities, towns and villages. In case if the geographical position of a pupil is not defined or to get the information about the administrative level is not possible, the variable takes on value 0;

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

3)\(U_{urb}\) means a sign of belonging of the settlement, where the student lives, to the type of urban settlements. Indirectly, this variable reflects the student’s access to urban cultural and educational infrastructure: the theaters, museums and variable centers of the additional education. Also, it may indirectly show the limited circle of student direct communication. The variable gets the value 1 in case if a student lives in an urban 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 processing the source data from GlobalLab and Dnevnik.ru there were found out the following the following geographic location records:

- \(U_{geoAdmLv1}\) – 44 thousand records;
- \(U_{lat}\) – 52 thousand records;
- \(U_{urb}\) – 39 thousand records.

B. Age and Gender

The scientists all over the world propose different theories about the influence of the age and gender on the students’ success in the education. The different tests show that at some subjects the females are more successful than the males but still there is no general conclusion about the connection between the gender and educational performance. They explain the female success in the exact sciences by their tentativeness and perseverance [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 there were proposed the following hypothetical external variables:

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;

2)\(U_{agePrec}\) means the age accuracy. The variable gets the value 1 if the age is known with the accuracy up to 1 year; it

The value of a variable that indicates the unavailability of the source data is called neutral.
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) $U_{sex}$ means the student’s gender.

1 means female; 
2 means male; 
0 is neutral value.

After processing the source data from GlobalLab and Dnevnik.ru there were found out the following the following age and gender records:
- $U_{age} - 58$ thousand records;
- $U_{agePrecis} - 58$ thousand records (46 thousand records with the notation that is equal to 1, 12 thousand with the notation that is equal to 2);
- $U_{sex} - 24$ thousand records.

### C. Style and learning

The success in education of each student depends on the style of the teaching and learning. To achieve this, aim the style of learning should coincide with the learning context and should be appropriate for each student. The learning style is the way how each and every student perceives, processes and integrates new information [13].

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

At the input, this procedure accepts many heterogeneous action messages about the actions of a definite user in the interface and, in the output, it looks like a vector that describes a user in the aggregated form. Many vectors of the different users in its turn have been used as the input data for clustering the algorithm of C-average and the output represents the user assignment to one of C clusters and defines the users by the behavior types.

For getting the hypothetical external variables there were defined several tasks for gathering the learning style data: 1) the session of creating a research project; 2) the research project idea creation session; 3) the session of filling out a research questionnaire and sending it to the project; 4) The session of acquaintance with additional non-interactive educational materials that accompanies the GlobalLab courses and projects (the text materials, videos and presentations, etc).

For each student there was done the vector calculation of $\vec{S}$ type based on each action having done by a student. Further 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).

$$\min \left[ \sum_{i} \sum_{c} \| \vec{s}_{i} - \vec{c}_{i} \|^{2} \right]$$

where $c_i$ is the centroid of the cluster $c_i$.

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

The number of the C clusters was changed in the range from 2 to 7. The range change of the cluster numbers is based on the hypothesis that a user behavior in the e-learning environment should vary for 3–5 learning styles. The maximum classification accuracy (the average for all the session types is 91.2%) was achieved during C=5. Thus, it will be reasonable to assume that for each type of a session each user maybe assigned to one of the 5 learning styles. Based on this assumption there were found out 4 learning style variables besides the neutral one:

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

Each variable value has got through (2):

$$U_{beh} = k_{mean}(\vec{S}_{beh})$$

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

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

$$U_{style} = k_{mean}(k_{mean}(\vec{S}_{behPrj}), k_{mean}(\vec{S}_{behData}), k_{mean}(\vec{S}_{behStd}))$$

After processing the source data from GlobalLab and Dnevnik.ru there were found out the following the following style of learning records:

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

### D. Academic performance (grades, test results)

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

The different countries there are different evaluation systems for the test and current marks. I this work there are used the marks on the several school subjects of the pupils from different forms:

- 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);
Development of an Innovative Education System Mathematical Model

- 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.

Further, there were given three hypothetical external variables:

1) \( U_{assessHum} \) is the adapted to 100-point scale average mark on the subjects of the humanitarian cycle;
2) \( U_{assessScience} \) is the adapted to 100-point scale average mark on the subjects of the natural science cycle and mathematics;
3) \( U_{assessTec} \) is the adapted to 100-point scale average mark on the subjects of the technology.

The second part of the research in the academic performance was dedicated to gathering the data of the test’s evaluation marks. There were proposed 60 different mini-tests with 10 questions on each subject. During the result calculation there was taking into account the time that a student used for passing a test.

In this way there were found out the following hypothetical external variables:

1) \( U_{testHum} \) is the adapted to 100-point scale average mark on the subjects of the humanitarian cycle;
2) \( U_{testScience} \) is the adapted to 100-point scale average mark on the subjects of the natural science cycle and mathematics;
3) \( U_{testTec} \) is the adapted to 100-point scale average mark on the subjects of the technology.

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 find out the proficiency level of the complex skills among the pupils there were used express method for assessing the skills of pupils in the field of project activities. The cogneme experiment had a game nature that was more acceptable by the pupils. The cogneme is a part of the knowledge that is in a constant dynamic process of five parameters: meaning – way – word (sign) – area – function [15]. The brains perceive the information (word or sign) in the easiest way – by listening, touching, fragrances and the other means of feeling perception, find out the meaning and the area of its usage and produce the functioning of a body [16-19].

During the experiment the pupils got the questionnaire with 100 stimulus words and filled them in during 7-10 minutes. The task was to write the very first associated word to a word stimulus. The incentive words are selected from an area associated with the various non-cognitive skills. For example, for the cooperation work the following words were chosen: question, answer, plan, opinion, conflict, and dispute, help and interaction, etc. Each incentive word had its number and the composition of the questionnaire was formed on the basis of the random number generator to avoid the appearance of two identical questionnaires.

Basing on the experiment results the associative-verbal net was composed for every student. It represents the graph in which the nodes are the language units (L.U) and the ribs are a fact of the L.U speaker’s association. The length of a rib reflects the association frequency.

By applying the algorithm of C-average the students were clustered 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 \( U_{noncog} \), 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 mastery level of the innovative technologies by their teachers. This level was expressed through the variable \( U_{noncogTutor} \) getting the same values in the rage 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 invented the student mathematical model (4):

\[
\bar{U} = \{ U_{geoAdmLvl}, U_{lat}, U_{age}, U_{assessHum}, U_{assessScience}, U_{assessTec}, U_{testHum}, U_{testScience}, U_{testTec}, U_{agePrecis}, U_{ageHum}, U_{assessHum}, U_{assessScience}, U_{assessTec}, U_{testScience}, U_{testTec}, U_{testHum}, U_{agePrecis}, U_{ageHum}, U_{assessHum}, U_{assessScience}, U_{assessTec}, U_{testScience}, U_{testTec}, k_{mean}(\bar{S}_{behPrj}), k_{mean}(\bar{S}_{behDat}), k_{mean}(\bar{S}_{behStd}), k_{mean}(k_{mean}(\bar{S}_{behPrj}), k_{mean}(\bar{S}_{behDat}), k_{mean}(\bar{S}_{behStd})), U_{noncog}, U_{noncogTutor} \} \quad (4)
\]

The vector \( \bar{u} \) models a student as a subject of a learning trajectory because may describe him/her apart from the concrete steps of the trajectory. The model can be characterized as a linear because all the methods of modeling the user’s characteristics are linear too. Though there was used the method of C-average in the model that assumes initialization of centroids with random variables the student’s model may be characterized as a deterministic one. It is explained by the fact that the formula (4) is used after the C-average algorithm usage along the entire data array that causes getting always the same vector \( \bar{u} \).

The usage of the neural networks to study and predict the properties of the learning trajectory will make available the new and effective methods of making the work and studies for the research organizations and teams in the sphere of the educational theory, the cognitive process modeling connected with the formation of the different student competencies and the creation of the more accurate evaluation methods of the students’ academic performance.
With the further development the model of the learning trajectory will help to adopt the educational system of any country for the individual demands and needs of the students to have full and better access to the educational instruments to improve their skills in the general education and in the further specialized education as well. The cooperation of the students and the teachers within this learning trajectory will positively influence upon both parts of the educational process participants. The successful cooperation will lead to the improvement of the educational system, finding out the better methods of teaching the students with the help of which they will perceive the information better and will accept it not only on the passive level but also will use it actively in the continuation of their education.

V. DISCUSSION

As result, this abovementioned list was analyzed in terms of the possibility of combining some events and adding to them data extracted from the electronic diary of Diary.ru.

To the second group of events from the data obtained from Diary.ru, an event of obtaining a rating was added. Also 2 types of events were added, reflecting the results of testing on the GlobalLab platform or completing tasks in GlobalLab courses. These events are considered successful if the grade given to the 100-point scale exceeds 65. Adding these events allows you 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 adequacy.

In the first group, the following 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 minimal: often (~47% of cases) users use the School entity to simply pool users who are not connected by one educational institution, and vice versa - use the Group for collective user work from the same class or school. This circumstance makes it possible not to make distinctions in the events associated with these entities, and to unite them. Moreover, such a union allows us to reduce by 4 elements the dimension of the vector representing the state of the educational trajectory.

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

Based on the results of the changes described above, the final list of events that describe the educational trajectory took the form presented in Table 1.

<table>
<thead>
<tr>
<th>No.</th>
<th>Type of an event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The user approves the project.</td>
</tr>
</tbody>
</table>

The types of events listed in Table 1 can be represented as types of steps of the educational trajectory. Thus, any state of the user at any step of his educational trajectory is recorded only when one of the events in table 4 occurs and, therefore, the mathematical model of the educational trajectory takes into account only these types of events.

The final list of events is taken into account by the mathematical model of the educational trajectory. The list of events allows us to increase the adequacy of the mathematical model of the educational trajectory by including in it a wide range of educational activities that are not traditionally included in student models.

VI. CONCLUSION

During the research there was developed the student mathematical model (as a subject of learning trajectory). There were included 43 variables that described the basic characteristics of the students.
Development of an Innovative Education System Mathematical Model

It should also be noted that the proposed student’s model is functional and does not describe the whole complex of the complicated and currently less studied cognitive and other mental processes that take place in the student consciousness and conditioning and influencing on his/her characteristics. In this sense the model represents the modeling object in a very simplified form. Nevertheless, it is more complicated comparing to the students’ models that are traditionally used for the evaluation of their knowledge and skills (for example: the student’s model for defining the results of the unified state exam) because it considers many characteristics starting from the geographical location to the complex 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