User Behavioral Data Enhancement and Vectorization in E-Learning Models

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

Abstract
The data acquiring mechanism on the criteria of users-interaction to the users-interface has been developed in the initial stage of this applied science research. The input system gets multiple inhomogeneous texts about the interface actions of a specific user, whereas the vector has been demonstrated by the results that describe the user in a consolidated manner. Moreover, the vector collection for various users has been employed as an input of clustering algorithm (i.e. K-means), the outcomes of behaviors of users are among the one of these (k-clusters) that differentiates the users by their behavioral types. Almost, 67.8% of the Interactional data of the interface of the user is accessible to GlobalLab platform users. However, for the electronic diary (i.e.Diary.ru) there is no corresponding data. Although not every user of the GlobalLab system takes the initiative to create the projects, ideas, questionnaire task, and educational resources, the proportion of learners have been 9.7 thousand whose value with all 4 variables varied according to a neutral one.

Keywords: Mathematical model, E-learning, Learning trajectory, Behavioral data, Education procedure

I. INTRODUCTION
The swift development of information technology has established the latest applications, including, e-health, e-commerce, and e-learning [26]. The exponential spread of internet-based technology has led to the introduction of various educational methods embodied in the use of e-learning systems. Almost every academic organization has already implemented web-based platforms for providing online courses, as the speed of e-learning is significantly growing. These also complement conventional approaches that allow learners to interact with their learning from any location, through different resources alongside or instead of face-to-face learning delivery. [27].

Educational frameworks, learning styles, and services for continuing education are based on a standardized approach, which may be unsuitable for the abilities of learners. Indeed, researchers can shift to quicker learning and more sustainable retention, through research in artificial intelligence and cognitive science, called Adaptive Learning.[28].

The question of providing an education adequate for the needs of the individual living in an information society, which would allow maintaining a satisfying quality of life [22], is of the utmost importance nowadays. This is a multifaceted problem concerning not only individual development and adaptivity but also the public trust in education, the level of which could be raised only by optimizing the sphere of education and science [24] and the progress of information society as a whole, including local innovations important for social and eco-friendly initiatives [25]. Moreover, it has been proven that the higher level of education the professionals have, the more successful they can be motivated [23].

Adaptive Learning Technology is one of the most promising solutions to the problem posed. It implements the “non-linear learning and debugging technique, relating to student interactions and illustrated level of performance, and then forecasting what kinds of information and materials students need to achieve success at some point”. It is an automated form of instructor. Furthermore, the learner characteristics are tracked and the educational atmosphere is sufficiently tailored to provide support and enhance the learning process. Such platforms are obtaining much attention as a result of their potential to
offer instructional information and evaluation by effectively adapting to the individual learner expectations and preferences [29]. The incredible aspect is how swiftly main publishing companies have adopted these innovations:

- Pearson worked with Knewton as a team, enabling Pearson to display his Mastering and MyLab transformation tools for a variety of topics, especially in science.
- McGraw Hill presented Smartbooks and ALEKS, the latest adaptive LearnSmart innovation.
- Macmillan recent projects has worked with Knewton and has approach to PrepU innovations..
- The collaboration of Snapwiz and Wiley has create a new WileyPlus product line. [1, 5, 9, 14].

There are numerous organizations and even institutes here:

- Organizations including ScootPad, CCKF, and SmartSparrow.
- Brightspace from D2L has been procured through the incorporation of Knowillage and adaptive LeaP learning technology.

In Europe, INTUITEL has been set up by around a dozen organizations to "transform the content of the e-Learning and Learning Management System (LMS) with services only given by human instructors so far." The Phoenix University has financed extensively in its technologies for adaptive learning.

Professors from Ohio State University have developed a module for adaptive learning (i.e. MOOCulus) that have been integrated for their MOOC calculation into the Coursera network.

Adaptive innovation has created the beach in higher education. The utilization of MyLab Pearson by Arizona State University and ELEXF ALEKS are important early initiatives. The reports on these programs are vague, as might be assumed while emerging innovations are just going to grow, but the scope of substantial priorities and resources aims to create this fundamental technology a feasible future.

The advancement of the current learning system based on computer technology can provide distinct learning processes. Students as users may conduct different activities, such as downloading learning stuff, submission of assignments, performing tasks, attempting quizzes, etc. Besides, certain operations conducted by these users are referred to as user actions for e-learning. Each student has distinct abilities and expertise, so an educator is accountable for equally assessing the students in e-learning in the selection of discussion groups. However, there has been an issue, the teachers' difficulties in categorizing the discussion groups with unique skills and knowledge, since the discussion group in e-learning is typically performed based on each student's wishes or arbitrarily, without looking at the knowledge and skills data of the students.

Data mining is one of the methods of obtaining data among large datasets and is very well known today. The process for finding patterns in the data is known as data mining. This method runs automatically or runs semi-automatically more frequently. The patterns discovered should be relevant and productive [30].

For several aspects, namely classification, prediction, association, and clustering process, data mining can be used. This study has been applied the technique of clustering with the k-means algorithm. K-means is a non-hierarchical data clustering approach that aims to partition existing knowledge into one or more clusters. The clustering approach is a process for categorizing data according to certain desired characteristics by dividing data into groups, where the group identity of and data is not yet identified. With this technique, the group data is expected to be known and then the identity is given according to the issues reported.

**Behavior data analysis and management systems**

A significant aspect of the learning process is Learning Style (LS). In designing, creating, and integrating e-learning environments, learning styles should also be considered. Learning styles can be defined as an explanation of the behaviors and attitudes that assess the possibility of a way of individual learning. Every student studying at any university or college has their specific style of learning. According to the skill of the child, learning styles are independent and must meet the student's needs before he or she achieves the
desired level of understanding. Consequently, the automatic assessment of the learning style of a student may be an essential feature of an e-learning framework. [31].

An online learning platform will be incorporated into the behavioral data analysis framework, which will use the learner information acquired from the platform as well as the learner machine information extracted from the web pages on which the system was created. For each user on the network, the Learning Management Systems provide more information about the learner and some statistics regarding the usage of pedagogical resources.

On online learning sites, there is a lot of data about learners:

- Learners’ personal information: Age, gender, country
- Statistics of self-assessment outcomes
- Users’ behavioral data:
  - Number of Platform Accesses
  - Duration of access (morning, afternoon, or evening)
  - The average length of platform visits
  - Amount of visits to the person page
  - Length of visits to each site
  - Number of clicks on each page
  - Scrolling of mouse

Management systems of learning or e-learning platforms offer numerous services including content management, particularly through the creation, import, and export of learning items. All these resources that help to manage the teaching cycle and the interaction between users, such as access control services, synchronous and asynchronous communication tools, and user administration services, are represented in the range of tools accessible in the LMS.

Generally, management systems of learning are based on students, instructors, tutors, coordinators, and administrators. The users can access or download the materials downloaded by his teacher while following their academic progress; they can develop their learning activities. The teacher, who is accountable for one or more modules, will produce and maintain the educational materials that the platform wants to broadcast. Teachers may also create instruments to track the activities of learners. The tutor accompanies and tracks each learner, by supplying the instruments of communication and cooperation. Finally, the administrator is accountable for customizing the platform having the administration's rights extracted from it [32].

Comparison of e-Learning methods with conventional education is similar to e-books to paper books comparison

As the integrity of a hardform book could not be replaced by digital scripts, in the same manner e-learning could not eliminate conventional learning. However, it is always useful as long as you have your personal device available near to you. Usability and simplicity are only two of the key reasons that people like e-learning. Another reason is that, with certain advancements, such as augmented reality and virtual reality, you can optimize your educational experience. For example, by using e-learning tools students can learn new disciplines. It can be medical subject, technical knowledge, and as well engineering.

II. METHODOLOGY

Data has been used to make smarter choices, such as organizations using it for productive business processes and efficiency. Several other corporations are now using DM to make better strategic and operational decisions[33]. The data describes details about the educational procedure offers better insight into learning and enhancing its outputs. [34].

Education Data Mining (EDM) is a diverse disciplines research domain that is associated with producing techniques of research for the data that evolve in a process of learning. To investigate educational issues,
EDM uses quantitative methods to assess educational data [21]. The main motive of EDM is to evaluate education-based numerous forms of data to mitigate research problems in the education domain [35].

E-learning, as a significant outcome of the merging of education and technology, has appeared as a potent channel of education especially using online sources. The irrefutable influence of e-learning in education has resulted in a tremendous rise in e-learning courses and platforms providing various kinds of solutions. Therefore, the expansion of e-learning platforms is essential for providing effective use, successful delivery, and positive impacts on participants.

The process of online educational systems has been developed to collect detailed information about the users-behavior. This method is part of behavioral patterns, typologies, and algorithms, the session was categorized in section 3, such as the plants introduction by a collection of variables in the context of users-behavior Sx, and was clarified by the evidence that the session is kept in the warehouse. The primary goal of the proposed algorithm is a series of measures for any stage of physical actions (i.e. PhisEvent entity Storage). The session is a portion of the document (such as CM, VI & ESO session), and the duration of session variable have the value of Sx. [7, 13].

At the 1st session of command, observation of events and their utilization as an input form has been done. The user can be situated in a hierarchical system and the activity of the session. In the session parameters algorithm, The Tuple describes the behavioral patterns of the user Sx. Moreover, the session variables in the algorithm representing the behavioral patterns of the user Sx, a tuple is used to evaluate the UbehPrj for all users, and the clustering algorithm (i.e. k-means) has been used to determine Ustyle, UbehStd, UbehData. Therefore, the introduction duration of algorithm for the one session of a definite level of time-ordered series of events, and descriptive instance of the code of action and way of learning is a line in the of (tuple).

Fig. 1. Algorithm for session variables that represent the behavior of users [1]
III. RESULTS AND DISCUSSIONS
To demonstrate the working of the algorithm, the proceeding example of analyzing PhisEvent cognitive activities that appeared during the document-level session has been provided. The user performed the activity (i.e. PhisEvent) relating to those mentioned in (Table 4 of Section 9) of the PNI Report for the 1st step after accessing the website page [1]. All events had been managed by an algorithm of private users method that sought the proper rule for this kind of session. The given snippet of the session record will be acquired as a consequence of implementing this rule:

\[
\begin{align*}
S_{\text{scrollVector}} &= 0.15, \\
S_{\text{scrollLength}} &= 0.24, \\
S_{\text{scrollCoverage}} &= 0.87
\end{align*}
\]

The preceding example is quite simpler: it is presumed that users perform scrolling activities only during the session. Mouse clicking, selection of text, and on-screen scrolling are the sessions in real-life [8, 11, 18]. The rules for transforming and evaluating the relevant values of the Sx variables are investigated for each of these cases.

The clustering with session type records has been conducted by the algorithm (i.e. K-means), as shown in Section 1.

**Accommodation Modifications for Users-Behavior in E-Learning Systems**
In the first step, The structure of users-behavior information has been established that contains all the details that can be used for inputs and forecasts throughout the recurrent architecture training network (i.e. xMANN). [17, 19].

The scenario includes:
- Currently accessible data in the repository of GlobalLab;
- Users-interaction data with the users-interface;
- Data related to learning performance.

Furthermore, the corps contains information about student progress collected from the electronic diary (i.e. Diary.ru), which has been incorporated into the platform of GlobalLab. In the 2nd phase of the PNI, the structure of obtained users behavioral information in e-learning platform has been modified and updated as shown below:

1. The data that is present already in GlobalLab has been lowered in a suitable template with Model 1 of the input database for the algorithm to forecasting the efficient trajectory of learning [1, 16].
2. The data of training evaluation has been decreased in a adaptable template with Model 2 of the Input Archive.
3. The incorporation arrangement of electronic diary (i.e. Diary.ru's) has been restructured to deliver extra data about learners' academic institutes and scores.
4. The software has been established to execute the process for transforming outputs from the databases of sections 1 & 2 into the format of xAPI (section 3, input data storage).
5. The software has been established to transform user centric data to a session log throughout a session (i.e. PhisEvent entity) [1, 2].
6. The deployment of software for vectorization of event elements representing a phase of the academic trajectories and techniques of decreasing the dimensions of the vectors procured (Shown in PNI report's sections 4 & 5).
The process presented in the previous section, relying on empirical rules, on one side, is quite adaptable to construct a diversified range of simulated users systems, and on the other side, assure the originality of the acquired designs by analyzing particular arrangements in user's activity in the of EOS GlobalLab. It is evident that, because of the reduced insight of the rules of user's interaction in EOS, the more accurate technique to acquiring simulation based information has not been conceivable.

In the second stage, the PNI test procedure produced a trial creation of a collection of simulated information. The method that is developed in the 1st Stage has been deployed by the program as a utility for generation purposes which will be contained in the experimental sample in Stage 3. [3, 4, 10, 12]

**Ensuring Cloud Technology Implementation, Setup, and Management System That Enhances User Activity Housing In E-Learning Platforms**

In the first Phase, PNI, using the services of a commercial partner implemented, designed, and effectively managed cloud platform for array development of users behavioral data in e-Learning platforms (i.e. Infra-structure descriptions in section 12 of PNI Report for 1st & 2nd Phase) [1]. Due to the upgrade of the database of users activity in e-learning platforms, the implementation, setup, and administration of the framework has progressed. The overall range of services providing the task to modify the behavioral dataset of users in e-learning structures in the 2nd phase has been increased to 160 [15, 20].

The services have been established programmatically, which allowed relatively close integration with electronic diary (i.e.Diary.ru) via calls to Diary.ru API techniques. These services incorporate steps 1.1 to 1.9 of the algorithm's input for the effective learning trajectory prediction model [1, 20].

In the 2nd Phase, it transported from using the API service (i.e.Google Geocoding) to using the open-source framework Nominatim OpenStreetMaps, which needed this service to be deployed on a rental dedicated server. The step towards the free-ware OpenStreetMap Nominatim system implemented on your commercial server also enabled an advanced version of the Uurb, Ulat, and UgeoAdmLvl variable measurement system. Deployment of the procedures has been facilitated in such a manner that the range of applications for reverse and direct geocoding is presently boundless.

Therefore, the utilization of OpenStreetMap Nominatim services unlocks the extra possibilities for simulating the living environment of learners in the future as it enables:

- determine the quantity and range from the user of numerous cultural and educational resources;
- demography of the community in which the student lives;
- natural characteristics of the residence area of the learner (elevation, landscape type, estimated yearly wind Temperature and so on.)

Cloud infrastructure installation, setup, and system management have been executed on FastVPS EXV-10 and EX-4 dedicated servers. In the structure of the patents analyses, investigation for patents, analysis of the intellectual property situations, recent developments, and future opportunities in the area of computing services for the development, evaluation, and vectorization of the learner's academic trajectory has been carried out.

It also includes the information regarding interaction as part of machine learning (ML) to evaluate automatically numerous features of a learner's educational trajectories.

The innovations, utility designs, and specimens contiguous to the corporate entity have been considered for the research. As a consequence of scientific and technological researches, fourteen articles of computer course enrollment certificates and 5 patents have been selected. [1].

1475
The evaluation of patent records and certificates of online registration demonstrates that, to date, there are no developments that use particular ML techniques to simulate the learning trajectory properties. Often, the element of protection is not a particular model, but a technique of implementing it to a more common utility model. Thus, several approaches focused on the use of particular inputs in learning trajectory models and learners are also secured.

IV. CONCLUSION
The mechanism of extracting the additional behavioral data of users in phase 1 has been established, the algorithm has been developed to produce a users-behavior data array to automatically convert the individual physical interaction data into a cumulative EOS session of his kind of behavior and style of teaching. To upgrade the corpus of user's-behavior data in online learning platforms, applied scientific research performed all required implementation, setup, and system management worked for the cloud infrastructure. Moreover, updates have been made to the users-behavior data structure, both relevant to upgrading the repository with recent data and transforming the data to a model-compatible layout to store the input of the optimized learning trajectory algorithm. The simulated data regarding the users used to train the RNS has been developed to use the software establishment of a method established in the 1st step of the applied scientific research.

ACKNOWLEDGMENT
The conducted research presented in this study has been conducted with the financial support of the Government represented by the Russian Federation Science & Higher Education Ministry, under Agreement #14.576.21.0100 of 26/8/2017 (And unique identifier of applied research – RFMEFI57617X0100).

REFERENCES


