Learning Analytics Framework for Adaptive E-learning System to Monitor the Learner’s Activities

Salma EL Janati¹, Abdelilah Maach², Driss El Ghanami³
LRIE Laboratory, Mohammadia School of Engineers (EMI)
Mohammed V University, Rabat, Morocco

Abstract—The adaptive e-learning system (AE-LS) research has long focused on the learner model and learning activities to personalize the learner’s experience. However, there are many unresolved issues that make it difficult for trainee teachers to obtain appropriate information about the learner’s behavior. The evolution of the Learning Analytics (LA) offers new possibilities to solve problems of AE-LS. In this paper, we proposed a Business intelligence framework for AE-LS to monitor and manage the performance of the learner more effectively. The suggested architecture of the ALS proposes a data warehouse model that responds to these problems. It defines specific measures and dimensions, which helps teachers and educational administrators to evaluate and analyze the learner’s activities. By analyzing these interactions, the adaptive e-learning analytic system (AE-LAS) has the potential to provide a predictive view of upcoming challenges. These predictions are used to evaluate the adaptation of the content presentation and improve the performance of the learning process.

Keywords—e-Learning; adaptive e-learning system; learner model; learning analytics; business intelligence; data warehouse; content presentation

I. INTRODUCTION

In the last few years, there has been a growing interest in the adaptive e-learning system (AE-LS). It is a new approach that makes an e-Learning system more effective by adapting the content presentation to the learner in accordance with their preferences, knowledge, and behavior. The aim of AE-LS is to provide the appropriate information to the right learner at the right time. It is based on a learner model, which used to adapt learner’s interactions of the e-Learning system according to their needs [1]. However, AE-LS does not cover all learning aspects since it does not provide teachers and designers with the adequate tools, which allow them to monitor and access to learner’s activities [2].

Many teachers, and designers expend enormous amounts of effort to design their learning monitoring system for the e-learning system to maximize the value of learner’s interactions between tutors, learners, and content. These systems allow them to access and summary information of learning activities. These interactions comprise number of visited pages, the time spent on each page content, and the date of the first and last access to the application. This information is important to have an idea about learners’ activities, but not enough to analyse their behavior and perform their evolution. Therefore, the teacher has a serious problem to extract information and use it in a useful way.

To address this issue, a new business intelligence framework, named adaptive e-learning analytic system (AE-LAS), has been suggested. In this approach, we suggest an analytic framework for AE-LS using business intelligence tools. This approach consists of the extraction of data from the AE-LS and the transformation of these data to store it in data warehouse tools, in order to visualize the learner’s activities. The result of this visualization will be used to predict learners’ preferences and improve the performance of the learning process. This paper is structured as follows: The first section presents a literature review. The second section describes the suggested architecture of the AE-LS and details their feathers. The next section presents the case study and the result of the suggested approach. Finally, we conclude with conclusion and future work.

II. LITERATURE REVIEW

AE-LS and Learning Analytics (LA) are two complementary approaches to monitor learner’s activities. In the review, the decrease in usage of open data sets, providing of AE-LS is likely a reflection of the privacy issues that arise with each LA application. To note also that the number of LA applications that use data from AE-LS has significantly decreased as compared to early academic analytics studies, which focused more on improving learner retention, graduation rates and the adaptation of the course content to the learner.

A. Adaptive e-Learning System (AE-LS)

AE-LS is a set of technologies and approaches that combined to give learners an adequate content meeting their needs [3]. In adaptive e-learning approach, each learner has different characteristics, which not suitable for another type of learners [4]. AE-LS covers a variety of systems: E-learning Platform [3], Learning Management Systems (LMS) [5], Intelligent Tutoring Systems (ITS) [6], Adaptive Hypermedia Systems (AHS) [7], and Dynamic Adaptive Hypermedia Systems (DAHS) [8]. The adaptation in AE-LS is divided into four elements: adaptive content aggregation, adaptive presentation, adaptive navigation, and adaptive collaboration support as is presented in Fig. 1. Adaptation content aggregation consists of providing the learner with several content depending on his/her background, preferences, learning style, and his/her level [9]. While, the adaptive presentation attempted to adapt the content presentation to the learner’s need by introducing the adequate visual presentation (text, audio, video, etc.) [10].
ATTENTION: The answer to this question is not constructive, and it is not appropriate to include in the final output. Therefore, the answer is not included.

The adaptation of navigation is based on the links presented in the pages, these links allow the navigation between pages, personalize view in the page, and facilitate the adaptation process [11]. The last adaptation of collaboration support allows the communication between learners and the collaborative application using network-based communication [12].

These systems have suffered from a lack of analytical reporting tools using to track the learner performance, evaluate the visual content presentation, and improve the adaptation of the system. Most adaptive e-learning system applied potential Learning Analytics tools either to tell learners what to do next by automatically matching learning resources to the individual needs of the learner and recommending them different learning entities, based on their preferences of content, presentation, and navigation.

B. Learning Analytics (LA) in AE-LS

Learning Analytics (LA) has strong roots in a variety of fields. It is a multi-disciplinary field, which combines business intelligence, artificial intelligence, machine learning, data mining, and visualization.

LA has attracted a great deal of attention in AE-LS. It uses varied sources of educational data. These data sources of educational data fall into five categories: Open educational Data sets, E-learning Platform, Learning Management Systems (LMS), Intelligent Tutoring Systems (ITS), and Hypermedia Systems (HS). LA is oriented to be used by learners, teachers, intelligent tutors, intelligent mentors, instructors, and administrators. It is dedicated to measure, perform and analyse data based on the context of the learner in order to evaluate their learning process [13]. Learning Analytics study researches to use intelligent methods, procedures and models to improve the learning process for learners, teachers and instructors point of view. These researches are based on information science, including computer science, psychology, and pedagogy [6].

LA consists of four major steps organized as an iterative cycle: Data Collection, Pre-processing, Analytic Action, and Post-processing [14]. In the first step, LA collects data from various e-learning system and heterogeneous data sources to explore it in the next step. The Pre-processing allows the transformation data into a suitable format that can be used for a particular Learning Analytics method. The third step is based on the result of preparation and transformation data applied in the pre-processing. It includes analysis, adaptation, and visualization data. The last steep of the LA process allows compiling a new data, determining new indicators and metrics and modifying the variable in order to continue the improvement of the analytics system.

The main purpose of LA is to develop a model that attempts to predict learner’s behaviors and future performance, based on their current activities and accomplishments. This model tracks learner activities and generate reports in order to support decision-making by the institutors and teachers. This predictive model can then be used to provide an active intervention for learner who may need additional assistance or suffer from impairments.

To achieve these purposes LA uses different methods and techniques such as: Business Intelligence (BI), Statistics, Data Mining (DM), Information Visualization (IV), and Social Network analysis (SNA).

1) Business Intelligence (BI): In the literature, Business Intelligence is defined as a set of methods and concepts used to improve business decision making by a support system. It includes architectures, databases, methodologies, and applications [15].

BI is the set of tools for analyzing data for decision making using data extraction, transformation and loading (ETL) tools to consolidate and integrate data into Data Warehouse (DWH) or OLAP (On-Line Analytical Processing) [16]. DWH is considered the foundation of BI. Some research considers that DWH is the database of the system where data is stored and consolidate [17]. Others researches consider that the DWH is a business intelligence platform [18]. In the education field, BI is used to explore the learner’s activities under different perspectives, dashboard, and the report of the visualization. It is used to improve teaching and monitor learner’s activities using data log files of e-learning system [19].

2) Data Mining (DM): Data Mining is the process of analyzing and performing data from database, Web, text, image in several perspectives and summarizing it into useful information. This process allows to find correlations or patterns among dozens of fields in large databases and Data warehouses [20].

Several studies have demonstrated that Data Mining techniques could successfully be incorporated into e-learning. It can be used to resolve classification and prediction problems in e-learning. The most famous techniques used to solve these problems are: fuzzy logic networks, artificial neural networks, evolutionary computation, and association rules graphs and trees [21].

3) Statistics: Most existing e-learning systems implement reporting tools that provide basic statistics of learners’ interactions and e-learning system. Statistics in e-learning generate a simple statistical operation such as: standard deviation, evolution, average, and sum. For examples the number of visits course, the number of connections by day, time spent on the page, and frequency of learner’s replies [22].
4) **Information Visualization (IV):** Information visualization techniques provide a visualization dashboard for teachers and learners, so that they no longer need to drive blind. It represents the results of LA methods in a user-friendly visual form might in order to facilitate the interpretation and the analysis of the e-learning data [23]. There are different Information visual techniques such as: Bar chart, Box plot, Distribution plot, Combo chart, Gauge, Histogram, KPI, and Treemap, used to represent the information an understandable format.

5) **Social Network Analysis (SNA):** Social network analysis techniques have been applied in different Learning Analytics tasks. It is the quantitative study of the relationships between learner and e-learning system. Social network analysis is modeled by a graph \( G = (V, E) \), where \( V \) is the set of nodes representing actors, and \( E \) is a set of edges, representing a certain type of linkage between actors [24]. Fig. 2 presents the learning analytics process.

Moreover, LA is still in the early stages of research. However, LA tools suffered from several limitations using with e-learning system. In one hand, learners learn through the Internet, so their learning process cannot be collected by e-learning systems and teachers cannot know the progress of the course, the manner which learner learns, and learner preferences. All of these will lead to the gap between students and teachers. On the other hand, the pedagogical content is static and the teacher’s assignment and content given to every learner is the same content. While, learner shaves different background and the knowledge structure is dynamic.

For this reason, we should analyse learning habits, characteristics, knowledge structure, and improve the visualization of the analysis results to teachers to be more intelligible. Instead, more contextual information is needed to be captured in heterogeneous media and graphics. In order to solve these problems, we inspired by approaches previously presented, especially the Business Intelligence approach to build a new dimensional modeling of data warehousing to visualize the learner’s interaction and reconstruct the adequate learning process of learners. The most famous learning analytics monitoring tools used in e-learning are presented in Table I.

In this section, we present the architecture of the suggested learning analytics system used for AE-LS. The proposed architecture consists of three layers: ETL Process Layer, Data Warehousing Layer, and Restitution Layer. Fig. 3 presents the proposed learning analytics system for AE-LS.

### TABLE I. FAMOUS LEARNING ANALYTICS MONITORING TOOLS USED IN E-LEARNING

<table>
<thead>
<tr>
<th>AE-LS</th>
<th>LA Tools</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E-learning Platform</strong></td>
<td>MATEP [25]</td>
<td>It is a monitoring web interface tool used to help teachers to analyze and visualize the learner data. This tool uses information from log files and add the learner’s contextual information.</td>
</tr>
<tr>
<td><strong>Monitoring Virtual Classroom</strong> [26]</td>
<td></td>
<td>It is a virtual monitoring tool used to follow learner in e-learning system with similar visions of classical classroom.</td>
</tr>
<tr>
<td>Sinergo/ColAT [27]</td>
<td></td>
<td>It is a monitoring tool that offers interpretative views of the activity developed by learners in a group learning collaborative environment. It integrates the information from log files with contextual information.</td>
</tr>
<tr>
<td><strong>LMS</strong></td>
<td>CourseVis [19]</td>
<td>It is a learning analytics tool for tracking the learner data and extract it through the web log files of the LMS server to help teachers to follow learners and identify their needs.</td>
</tr>
<tr>
<td>GISMO [28]</td>
<td></td>
<td>It is a learner monitoring tool, designed to extract data from Moodle log files and represented it in graphics. These representations allowed teachers to perform learner’s behaviors.</td>
</tr>
<tr>
<td><strong>ITS</strong></td>
<td>Educational Data Mining Tool [29]</td>
<td>It is a monitoring tutorial tool, which use tutor logs directly into database to extract data, in order to summarize the learner’s interactions.</td>
</tr>
<tr>
<td>TADA-ED [30]</td>
<td></td>
<td>It is a tool, which integrates various visualization and data mining facilities to help teachers and instructors in the learning process.</td>
</tr>
</tbody>
</table>
III. SUGGESTED ARCHITECTURE

A. ETL Process Layer

The ETL architecture is organized into two areas: Staging Area and Operational Data Store (ODS). Each area consists of a separate database to store data provided by AE-LS.

1) Data source extraction: The proposed system uses multiple data sources of AE-LS such as Open Educational Data sets, E-learning system, LMS, AHS, and DAHS. To extract data from AE-LS database, the analytic system uses the Application Programming Interface (API) function, especially the Data Manipulation API and the Logging Data API. The data manipulation API is used to extract the learner data from the relational databases of AE-LS and import it in the Staging Area. While, Logging, Data API is used to record the data from AE-LS log files.

2) Staging area: Staging Area is created for every AE-LS data source that is activated. It has the same structure as their respective data source. It is also flagged with key fields for the request ID, the data package number, and the data record number. Staging Area is a copy of the data source that centralize data from heterogeneous sources to facilitate analysis operations. Its contents can be emptied after each load of the data warehouse.

In the staging area, the learner data is stored without cleaning or controls. Once the learner data is stored in the staging area, we can consolidate data in Operational Data Store (ODS) layer.

3) ODS: The integration of the data into the ODS implies a purge of redundant information and incoherent learner data. This stage is attempted to consolidate data, remove duplicate data, clean data, and change their format to the adequate format. All reconciliations and aggregations are stored in the ODS layer to obtain the desired structure for learner’s behavior analysis. The result of transformations is loaded in the Data Warehouse.

B. Data Warehousing Layer

In order to build an analytic model for AE-LS, we propose a model of DWH implemented in PostgreSQL database management system (DBMS). The proposed model covers all data and analysis issues covered by adaptive e-learning systems. The DWH is organized in star schema and hosted on a PostgreSQL server. It collects data from a variety of diverse and heterogeneous sources of AE-LS, which the primary purpose is to support analysis and facilitate the decision-making process. This DWH store a large amount of historical data and enable faster and complex queries across data and its derivatives or even in-memory. The database is used to store current transactions and allows quick access to specific learner’s transactions, including indexing technologies. It is usually standardized, which means that there is a single copy of each data.

The Data flow of this approach defines which objects are needed at design time and which processes are needed at runtime. These objects and processes are needed to transfer data from the AE-LS source to the ETL and DWH, to cleanse, consolidate, and integrate the learner’s data, so that it can be used for analysis, reporting and planning. The individual requirements of these processes are supported by numerous options for designing the data flow. Fig. 4 shows the data flow architecture for the proposed learning analytics system.

To ensure the quality of AE-LS data in a data warehouse, we opt to use a Master Data Management (MDM) solution. The MDM solution helps to increase their overall efficiency by better use of the analytics system. The implementation of the MDM solution in the DWH improves the process of data and allow a quicker and simpler management of the new activities of learners. It gives the possibility to eliminate unnecessary data and processes, make data reusable, improve overall data quality, and define reference tables.
C. Restitution Layer

The restitution layer receives the result data of the star schema from PostgreSQL DBMS and visualizes it with Qlik Sense tool. Qlik Sense allows to explore heterogeneous educational data, it makes it easy to choose data with a single click and offers a quick grip and setting. It is faster, more intuitive and less bulky than multidimensional OLAP solutions.

This solution is an end-to-end application, which contains the analysis objects, their setting, and the data to be analyzed. Qlik Sense uses high-performance technology that can process millions of recordings in memory; when constituting the data model, the AQL (Associative Query Language) technology proceeds automatically by homonymy, and indexes existing logical links in the data. This allows an exemption from the definition of relational joins when developing and using dashboards. Fig. 5 illustrates the Qlik Sense architecture.

Fig. 4. Data Flow Architecture for the Proposed Analytics System.

Fig. 5. Qlik Sense Architecture.
IV. CASE STUDY OF AN ANONYMOUS AE-LS DATA SET

In the case study, we use the anonymised Open University Learning Analytics Dataset (OULAD) [31]. It contains data about, learners, courses and their interactions with AE-LS. We extract data from the open dataset of AE-LS anonymously according to the ethical and privacy of the system. The AE-LS dataset contains data from years 2013 and 2014.

The data components of the system architecture include the data sources, which correspond to the anonymized Open University Learning Analytics Dataset. From this dataset, data are extracted, transformed and loaded into the data warehouse, where data is stored in a suitable format.

A. Talend Architecture Job

Talend is the ETL solution, which will be used. It not only guaranties the highest level of performance but also the most cost-effective solution available in BI. It is an open source desktop application offers intuitive drag-and-drop tools, smart guides and automated processing features. It allows users to explore, clean, enrich and combine data from different sources. It allows users to simplify and speed up the often time-consuming and laborious process of data preparation, management, manipulation and data analysis tasks using spreadsheets. Fig. 6 presents the ETL Job using Talend tool.

In the Talend architecture jobs, we extract data from flat files of the anonymous dataset using File Input Delimited job and store data in Staging area using separate staging area databases. There are neither restrictions nor integrity key is used, the data are treated as if the tables were flat files. The second step is transforming the data in order to store them in ODS. Table II presents an example of transformations used in this approach.

<table>
<thead>
<tr>
<th>CSV Sources</th>
<th>Transformations</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>StudentRegistration</td>
<td>Join data together from Learner and learner_Info sources.</td>
<td>tJoin</td>
</tr>
<tr>
<td>StudentInfo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Courses Assessments</td>
<td>Remove duplicate data.</td>
<td>tUniqRows</td>
</tr>
<tr>
<td>Assessments</td>
<td>Select only certain columns.</td>
<td>tFilterColumns</td>
</tr>
<tr>
<td>StudentVLE</td>
<td>Extract Day, Month, and Year from date.</td>
<td>tMap: Day(date) /Month(date) /Year(date)</td>
</tr>
<tr>
<td>StudentInfo</td>
<td>Convert data into capital letters.</td>
<td>tMap: StringHandling.Upcase(region)</td>
</tr>
<tr>
<td>Courses</td>
<td>Delete spaces on the right.</td>
<td>tMap: StringHandling.TRIM(Label_courses)</td>
</tr>
<tr>
<td>VLE</td>
<td>Replace with '0' if null.</td>
<td>tMap: Relational.ISNULL(sum_click) ? '0' : sum_click</td>
</tr>
<tr>
<td>StudentInfo</td>
<td>Construct metrics and statistics of the learner’s activities based on values and calculations.</td>
<td>tAggregatorRows</td>
</tr>
</tbody>
</table>

Fig. 6. ETL Job of the Suggested Approach using Talend Open Studio.
B. Indicators Analysis and Measures Identification

The ODS layer of ETL process contains the most important transformations’ data, which be stored in ODS. This layer is intended to contain indicators, measures, and data of the finer granularity. In this approach, we mean by indicator, the aggregation of data, generally quantitative, called measures. Thus, we searched for indicators to measure learner’s activities. The particularity of these indicators is that are applied to study learner’s activities such as:

- Number of connections;
- Time of connection by day;
- Last access to the content;
- Number of visits courses;
- Time spend viewing learning content;
- Preferred content learning resources;
- Preferred learning resource;
- The manner which learners use the collaborative tools;
- Preferred learning activities of learners, etc.

These indicators and measures will be stored in the DWH, especially in the Fact table according to their dimensions.

C. Dimensions Identification

In the proposal, the dimension is an element constituting the context of the indicator. It is a structure that categorizes measures in order to enable teachers to answer pertinent questions about learner’s activities.

1) Assessments_DIM: Assessments_DIM contains the results of learner evaluations. If the learner does not submit the assessment, no results are recorded. In addition, submissions for the final exam are missing if the assessment result is not stored in the system.

This dimension includes the identification number and the name of the assessment, a status flag indicating that the assessment result has been transferred from a previous module and the learner’s score on this assessment.

2) Time_DIM: When a dimension is updated, the ETL keeps the memory of the old values in order to the old facts remain linked to the old values. To manage the history of dimensions’ values in the data warehouse, we associate dates with other dimensions and facts in order to know what dimension value is valid for what fact, using the Time_DIM dimension. It is structured as a hierarchy of minutes, hours, days, months, and years of activities. The suggested model of DWH, refresh data periodically.

3) Learners_DIM: Learners_DIM contains learners’ information, such as learner ID, the age band, gender, level of education, disability, etc.

4) Modules_DIM: Module_DIM contains the information about each module such as name, type, and levels of education in the module.

5) Courses_DIM: This dimension contains the list of all available courses to learners and their identification numbers.

6) Content_Presentation_DIM: Presentation_content_DIM contains information about the available content presentation materials in the AE-LS. Typically, these presentation materials are html pages, text, power point, audio, video, etc. Learners have access to these contents presentation online and their interactions with the materials are recorded. It includes the content presentation ID, the type of activity associated with the content presentation materials, and the learning resources.

In this approach, we use Talend MDM to manage the reference data. The reference data is data used to categorize other data within applications and databases. It refers to this data as a look-up, a code or a value. This table is characterized by code and value such as the two reference tables of the proposed DWH: Language_DIM and Geography_DIM.

7) Language_DIM: Language_DIM is a reference table, which refers to categorize language who learner use, say or write when they are trying to communicate spontaneously with learners or with e-learning system.

8) Geography_DIM: Geography_DIM is a reference table, which refers to identify the geographic region, where the learner lived while taking the module-presentation.

D. Data Warehouse Modelling

In this section, we suggest a data warehouse modelling based on identified indicators, measures and dimensions previously presented. The data warehouse modelling consists of one fact table and eight dimensions previously mentioned: Learner_Activities_FACT, Time_DIM, Learner_DIM, Modules_DIM, Courses_DIM, Assessments_DIM, Content_Presentation_DIM, Language_DIM, and Geography_DIM.

1) Learner_Activities_FACT: The fact table, Learner_Activities_FACT, accumulates in the data warehouse; there is never any deletion or update (monotonous growth) to manage the history and traceability of data. It is linked to each dimension table with its foreign key. The fact table contains the learner activity ID, some attributes of learner’s activities, and the foreign keys of these dimensions. The main measure of the fact table are: number of visits courses, time spend viewing learning content, last access to the AE-LS, preferred learning resource, preferred language, preferred content presentation, preferred learning activities, the number times which the learner has attempted this module, the total number of credits for modules which the learner is currently studying, learner’s final result, the Index of Multiple Deprivation Band of the place where the learner lived during the module, and the number of the learner’s interactions with the materials in the AE-LS, etc. Fig. 7 presents the suggested data warehouse modelling for deploying analytics in AE-LS.

E. Results

Fig. 8 illustrates an example of dashboards using the suggested analytic system and the anonymised dataset of AE-
LS. The information used in this reports help teachers to have a visibility of the learner’s activities and know their preferences. Thus, it allows to improve the presentation of learning resources and the performance of learners taking into consideration the band of learner’s ages.

In the year 2014, 46.4% of learners prefer to participate in forums and virtual class, which allows them to interact directly with teachers and other learners. Around 23.1% of learners stay on the home page, which shows the obligation to improve the presentation of the home page to attract learners' attention and motivates them to attend courses and pass quizzes. In contrary, 3% of learners access to the content learning resources. For this reason, the e-learning system should use the interactive learning resources such as interactive videos, speeches and graphical simulations all over for learners who suffer from visual and auditory limitations that meet to their needs.

F. Limitations

This paper contributes architecture of analytics systems of AE-LS to monitor learner’s activities. This approach uses a model of data warehousing as a result of the process of the extraction, transformation, and loading data, which be explored in the restitution using Qlik Sense tool. The architecture is applicable for any e-learning data sources where the data types can be partitioned into numeric, date/time, and textual data types.

However, this approach is limited in the capability of handling unstructured data. Structured data has a fixed schema so that to fit it a relational model can be designed. On the other hand unstructured and semi structured data do not have a fixed schema. Nowadays in e-learning system, these types of data are becoming common. These data also hide prudent information. Therefore, the decision-making in the learning process would not be effective as these unstructured data are not suitable.

Fig. 7. Suggested Data Warehouse Modelling of Learning Analytics in AE-LS.
Fig. 8. Example of Dashboards using the Suggested Analytics System.
V. CONCLUSION

In this paper, we present a new approach to monitor and perform the learner’s behavior. The suggested framework tries to solve the lack of reports in an analytics system according to AE-LS. This approach provides teachers and administrators with detailed reports to assess, evaluate and perform learners’ activities. It allows instructors to visualize, analyze and discover learners’ behavior in order to improve the quality of teaching and take the suitable decision.

This study shows that learner spends little time on courses and tutorials and prefer to use collaborative activities, such as virtual classes and forums, rather than looking at learning resources. In future work, we will use a predictive model of data mining to help instructors to have an idea of future learner’s behavior, and propose to the learner an adaptive presentation content.

REFERENCES