ADA: A System for Automating the Learning Data Analytics Processing Life Cycle

Conference or Workshop Item

ADA: A System for Automating the Learning Data Analytics Processing Life Cycle

Dilek Celik[1], Alexander Mikroyannidis[1], Martin Hlosta[1], Allan Third[1], and John Domingue[1]

[1] Knowledge Media Institute, The Open University, Milton Keynes, UK
dilek@dcs.bbk.ac.uk
{alexander.mikroyannidis, martin.hlosta, allan.third, john.domingue}@open.ac.uk

Abstract. Learning analytics is an emerging field focusing on tracing, collecting, and analysing data through learners' interactions with educational content. The standardisation of the data collected to supporting interoperability and reuse is one of the key open issues in this field. One of the most promising routes to data standardisation is through the xAPI: a framework for developing standard 'statements' as representations of learning activity. This paper presents work conducted within the context of the Institute of Coding.[1] Additionally, we have developed a system called ADA for automating the learning analytics data processing life cycle. To our knowledge, ADA is the only system aiming to automate the turning data into xAPI statements for standardisation, sending data to and extracting data from a learning record store or mongoDB, and providing learning analytics. The Open University Learning Analytics Dataset is used in the test case. The test case study has led to the extension of the xAPI with five new methods: 1) persona attributes, 2) register, 3) unregister, 4) submit, and 5) a number of views information.

Keywords: Learning Analytics, Data Standardisation, xAPI

1 Introduction

There exists a rising interest in learning analytics (LA). LA focuses on collecting learners' data and analysing them using advanced technologies including Machine Learning (ML) to improve educational outcomes [1]. One of the key open issues in LA is the standardisation of the data collected to support interoperability and reuse [1].

Open University Analyse (OU Analyse)[2] also encountered with the data standardisation problem in LA. OU Analyse aims to provide early prediction of at-risk students building on their demographic data and interactions extracted from virtual learning environment (VLE) with the clicks of students to increase the retention rate at the OU and improve the quality of education [2]. To support research in this field, OU Analyse

[1] https://instituteofcoding.org/
[2] https://analyse.kmi.open.ac.uk/open_dataset

developed the Open University Learning Analytics Dataset (OULAD)[2] based on the courses presented at the OU [3]. The dataset contains demographic data of students and clickstream data of students' interactions in the VLE. Based on the demographic data and selected activity types of this legacy dataset, OU Analyse constructed four ML-based predictive models. To run these predictive models over other institutions learning dataset, data standardisation is required as they have various underlying data structure.

Unfortunately, there are yet no established standards of LA data usage on how to monitor, feedback and improve students' educational performances. Currently, the enormous amount of data generated through LA is processed in ad hoc and task specific ways that prevent interoperability and reuse.

In this paper, we introduce Automated Data Analytics (ADA) as the first attempt to address the above-mentioned challenge which is neglected in LA. ADA is a system that automates end-to-end execution of learning data analytics building upon the outcomes of the OU Analyse as part of the Institute of Coding (IoC). IoC is a UK Government's £40m+ initiative aims to transform the digital skills required by the 21st by innovative and industry-focused education in HE across the UK.

2 Pedagogical Background

ADA contributes to enhancing teaching and improving educational outcomes in HE through LA. LA is seen as an innovative pedagogy in the 21st century [4]. [4] highlighted that "our LA is our pedagogy" arguing that the ways followed for gathering data, interpreting them, and acting on them connect, enshrine, and show a role in pedagogy in action. LA's relationship with established pedagogic approaches is conceptualised in the literature [5]. Specifically, our proposed system provides analytical dashboards to the HE lecturers and course team about their students' progress. Since each student's interactions with VLE is recorded, it is easy to find the learning material that the student has missed. When a student identified as at-risk, the system recommends the resources that will be him/her back to track. Predictions of at-risk students and providing them right-support also promotes the personalised learning.

3 Technological Background

ADA is a system designed to automate end-to-end execution of learning data analytics. The flow of the ADA is presented in Fig. 1.

Fig 1. The flow of the ADA System

The system comprised of four main features: (1) turning data into xAPI statements; (2) storing data in mongoDB; (3) extracting data from mongoDB and preparing for analytics; and (4) providing learning analytics. ADA's main features are explained in the following paragraphs with details.

3.1 The System Features

Turning Data into xAPI Statements. ADA allows users to insert mass data in .csv file format or directly connecting to VLE and then turns the data into xAPI Statements. xAPI is a technical specification to standardise data about a learner or group's activities from various sources in a consistent manner. Creating or extending xAPI verbs are needed when there are not appropriate built-in properties of xAPI verbs.

Storing Data in mongoDB. Another feature of ADA is sending the data that is in the form of xAPI statements into mongoDB. It is possible to send the data into mongo DB directly using xAPI verbs or using learning record store (LRS).

Data Extracting and Preparation for the Analysis. ADA extracts data from mongoDB and prepares it for the predictive analysis.

Analytics. ADA provides predictive analytics to the teachers using predictive models of OU Analyse with the available data. The analytical dashboard of ADA informs lecturers and course team about their students' progress to provide them right-support.

4 Test Case and Results

The OULAD is used in our test case. OULAD consists of information about 32,593 students, 22 courses, students' assessment results, logs of students' interactions with the VLE (10,655,280 entries). Our test case shows how the ADA system standardises data, processes data, and provides useful learning analytics. The test case led to the extension of xAPI verbs, which are presented in the following sub-sections.

4.1 xAPI persona attributes

The general form of xAPI statements is "[actor] [verb] [object]". In xAPI, each person's profile is named as a persona. A Persona describes an actual person with a compound of zero or more identifiers and attributes.

We proposed a schema for students' persona data according to data standardisation of Higher Education Statistics Agency (HESA). Table 1 illustrates the persona attributes used in OULAD and corresponding data items of the schema with definitions. The valid entries for data items are available in HESA[3].

Table1. OULAD persona attributes and their corresponding data items of HESA

OULAD Attribute	HESA Data Item	Definition
gender	SEXID	Gender of the learner.

[3] https://www.hesa.ac.uk/collection/c18051/index

age_band	BIRTHDATE	Date of birth of the student.
highest_education	QUALENT3	The highest qualification a student holds on entry.
region	TTPCODE	Postcode for the student's term-time address.
	TTACCOM	Student's living place during the current year.
imd-band	IMD	The official measure of relative deprivation.
disability	DISABLE	Type of disability that a student has.
final_result	OUTGRADE	The examination grade awarded to the student.

4.2 OULAD xAPI verbs

To define OULAD, the following verbs are extended or created, as there are no corresponding built-in properties of xAPI verbs: register, unregister, submit with banked and unbanked, and view.

For register and unregister verbs, an object which was the name of the course at the OULAD needed to be specified. However, existing xAPI does not provide objects for this. Therefore, we created an object to define a course with its type as follows.

```
"object":{"id": "http:\/\/kmi.open.ac.uk\/xapi\/verb\/course",
"definition":{
"type":"http:\/\/kmi.open.ac.uk\/xapi\/verb\/course\/"CourseName",
"name":{"en": "CourseName"}}}
```

"Submit" verb is extended with an object to make it specific for the assessment types of OULAD as presented in the example below. The object is also extended in two ways: isBanked and belongs to.

```
"object":{"id":"http:\/\/http://adlnet.gov/expapi/activities/assessment,
"definition":{ "type": "http:\/\/kmi.open.ac.uk\/xapi\/verb\/assess-
menttype\/"NameofAssessment", "name": {"en": "NameofAssessment"}
"extensions":{"http:\/\/kmi.open.ac.uk\/xapi\/verb\/isBanked":true,
"http:\/\/kmi.open.ac.uk\/xapi\/verb\/belongsto": "ModuleName"}}}
```

"Viewed" is a new verb created in OULAD case. This verb helps to define the number of clicks that are made on specific assessments at the VLE.

```
"verb": {"id": "http:\/\/kmi.open.ac.uk\/xapi\/verb\/viewed",
"display":{"en": "viewed"}}
```

References

1. Siemens, G.: Learning Analytics: The Emergence of a Discipline. Am. Behav. Sci. 57, 1380–1400 (2013).
2. Kuzilek, Jakub; Hlosta, Martin; Herrmannova, Drahomira; Zdrahal, Zdenek and Wolff, A.: OU Analyse : Analysing at - risk students at The Open University. Learn. Anal. Rev. LAK15-1. 1–16 (2015).
3. Kuzilek, J., Hlosta, M., Zdrahal, Z.: Open University Learning Analytics dataset. Sci. Data. 4, 170171 (2017).
4. Shum, S.B.: Our Learning Analytics are Our Pedagogy. In: Expanding Horizons 2012 Conference, Macquarie University (2012).
5. Knight, S., Buckingham Shum, S., Littleton, K.: Epistemology, pedagogy, assessment and learning analytics. In: Learning Analytics and Knowledge (LAK). pp. 75–84 (2013).