

Data Merging for Learning Analytics in Learning Environments

ADELINA ALEKSIEVA-PETROVA AND MILEN PETROV

Problems

- how to merge and process data by using records from different systems, tools or services;
- how to protect the learner's anonymity.

APTITUDE Project

To design and implement a flexible platform which supports recommendation and adaptation of learning contents and activities in education based on learning analytics from different systems, tools and services.



<http://aptitude.w3c.fmi.uni-sofia.bg/>

Paper Contributions

The proposal and validation of a processing data method for merging with learners' privacy preservation and the definition of appropriate parameters for data analyses.

- Question 1. How can we combine data from two (or more) independent learning systems or tools?
- Question 2. Is it possible to analyze combined data and make conclusions based on them?
- Question 3. Is it possible to preserve the anonymity of the information across different learning systems, tools, and analytics engines?

Research Methodology

- Related Works
- A Data Merging Method for Learning Analytics
- Evaluation of the Data Merging Method
- Results

Related Works 1/2

Learning analytics:

- collect and analyze students' learning experience and behaviour data to predict the performance of the student by identifying and validating corresponding measures of student outcome activities and processes in learning environments.
- "a correlation between observed (assessments, content examination) and predicted (final results) values to improve students' performance" (Joshi, 2021).

The trainer could get information on the most preferred activities and topics in a given course based on the student's behavior:

- all kinds of activities their students performed during the learning process,
- the use and the comprehensibility of the learning content
- contextualized social interactions among students in the virtual learning environment.

Related Works 2/2

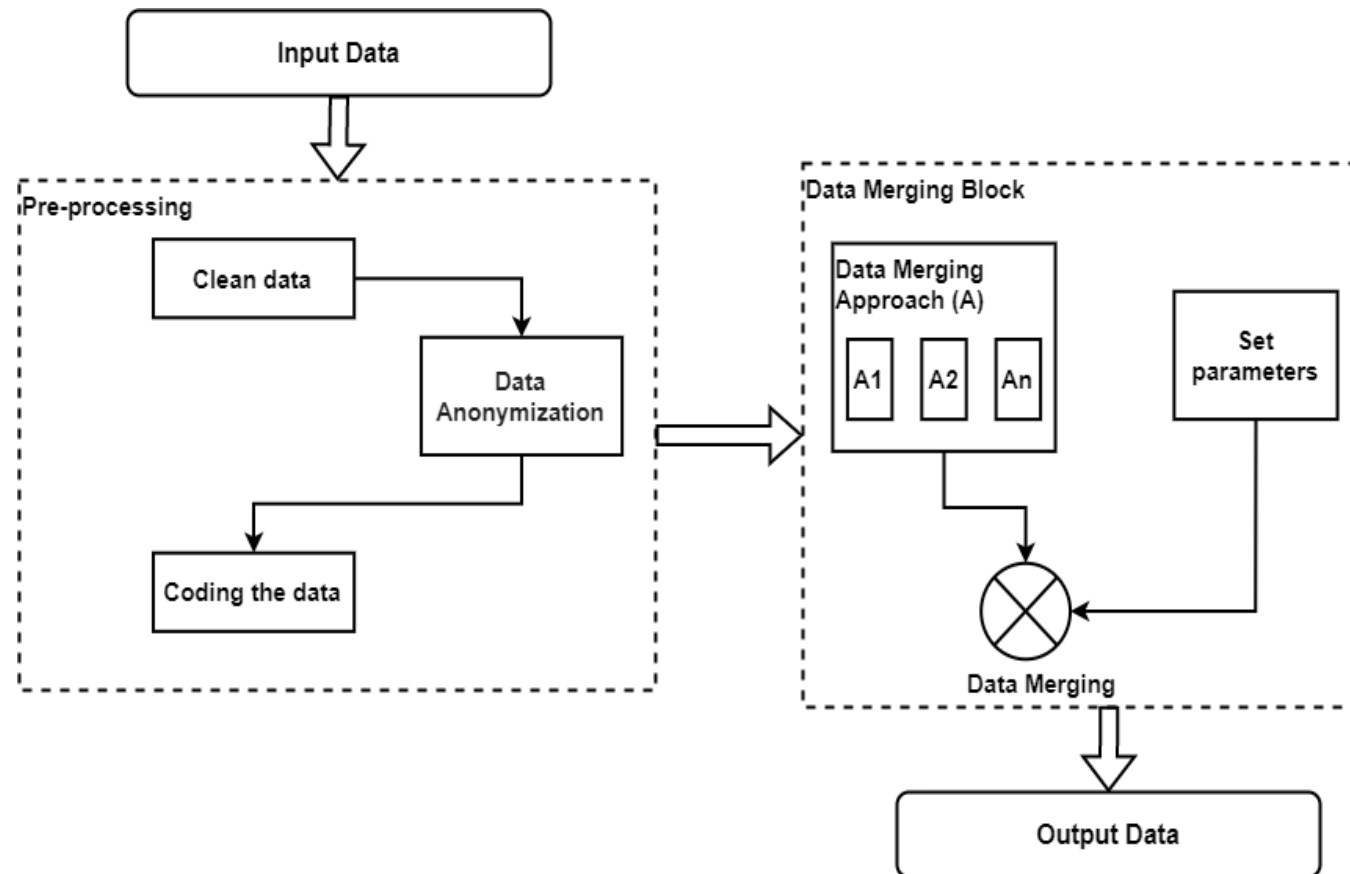
Common data collection methods:

- trace data – a digital logs of learner actions undertaken in a different system type setting;
- self-report data - includes methods such as think-aloud, questionnaire, interview, or any other form where the learner is asked to reflect on an aspect of their learning.

Data merging

- “the process of combining two or more similar records into a single one”

A Data Merging Method for Learning Analytics



- Aggregate - combines records by grouping them based on parameter values;
- Join - combines data into new columns based on a matched condition between them;
- Union - combines records with the same data schema and structure.

Evaluation

Input data

Course "Web technologies", 2021-2022 academic year

- Moodle - the log file, 8.86 MB, CSV, 43,849 records:
time, user full name, affected user, event context, component, event name, description, origin and IP address
- The web-based system for peer assessment and review - system log file, 1.71 MB, a CSV format and 21,767 records
id, browser, device_type, os, started_at, type, description, user full name, email, IP address

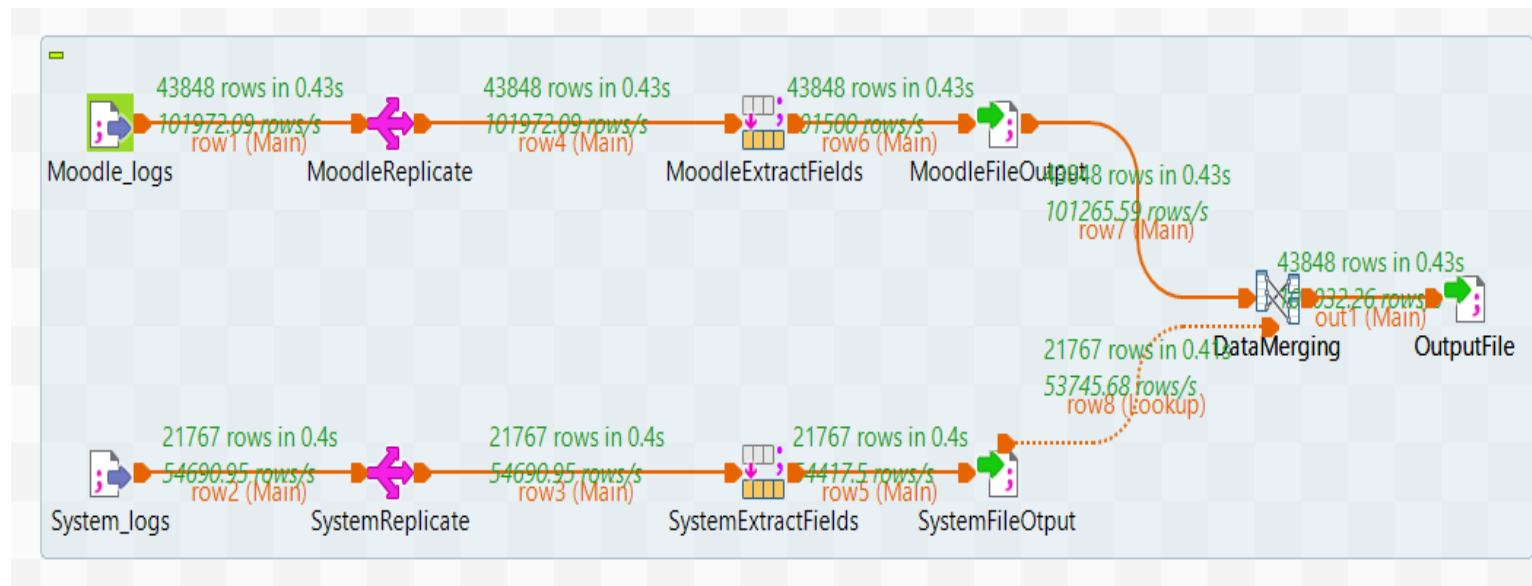
Evaluation

Pre-processing data

- any incomplete records are deleted
- All records from the Moodle log which affect the administrators and instructors are removed.
- The "User full name" field has been replaced with "AnonID", which uses the generated anonymous number
- The "IP address" column has also been anonymized by generating a unique number, called "AnnonIP".

Evaluation

Data merging



- Validation scenario - to track the execution of learning activities that depend on the OS in Moodle.
- Join approach - combines data into new columns

Results from the Data Merging Method

```
1 AnonID;EventContext;IP_ANON;device_type;os;browser
2 105;Course: Web technology;1;Desktop;Windows 10;Chrome
3 105;Course: Web technology;1;Desktop;Windows 10;Chrome
4 5;URL: Exercise 1 INTRO;2;Desktop;Windows 10;Chrome
5 5;Course: Web technology;2;Desktop;Windows 10;Chrome
6 63;Course: Web technology;3;Desktop;Mac OS X;Safari
7 63;Forum: Announcements;3;Desktop;Mac OS X;Safari
8 63;Forum: Announcements;3;Desktop;Mac OS X;Safari
9 63;Course: Web technology;3;Desktop;Mac OS X;Safari
```

- Output file - 2.58 MB; 43,848 records; form of a Moodle log.
- The resulting output gives information on the device (such as Desktop, Mobile), operating system (like Windows 10, Mac OS X) and browser (like Chrome, Safari), which are used by the students when an event is generated as a result of student activity.

Conclusion

1. How can we combine data from two (or more) independent learning systems or tools?

It is possible to combine data, but with the following prerequisite: that the data has to be related to the same course and to the same learners and activities.

2. Is it possible to analyze combined data and make conclusions based on them?

The simple answer is “yes”. We merged and analyzed data from both sources and made conclusions on the possibilities for the recommendation of content or learning activities.

3. Is it possible to preserve the anonymity of the information across different learning systems, tools and analytics engines?

With the anonymization of the user’s full name and IP address, we answer the third research question and we can conclude that it is possible to preserve the anonymity of the information across different learning systems, tools and analytics engines.

Acknowledgements

The research reported here was funded under a project entitled “An innovative software platform for big data learning and gaming analytics for a user-centric adaptation of technology-enhanced learning (APTITUDE)” by the Bulgarian National Science Fund with contract No: KP-06OPR03/1 from 13.12.2018.

Thank you!

- Adelina Aleksieva:
aaleksieva@tu-sofia.bg
- Milen Petrov
milenp@fmi.uni-sofia.bg