An Architecture for a Decentralised Learning Analytics Platform (Positioning Paper)

Audrey Ekuban[1,*], John Domingue[1]

[1]*Knowledge Media Institute, Open University, Milton Keynes, UK*

Abstract

Predictive Learning Analytics is a subfield of Learning Analytics that helps identify students who are at risk of dropping out or failing. However, the centralised approach to Predictive Learning Analytics raises privacy and ethical concerns, particularly in the area of data collection. In this paper emphasis is placed on a decentralised mechanism for collecting student consent. This mechanism is part of the EMPRESS framework, that combines self-sovereign data, Federated Learning, and Graph Convolutional Networks for Heterogeneous graphs to address these issues. EMPRESS allows data owners to control who has access to their data, processes data on their devices, and utilizes knowledge graphs for analysis. Course tutors can use the insights provided by the analyses to offer timely assistance to these students. In addition, this paper details how Heterogeneous graphs can be used in Predictive Learning Analytics.

Keywords

Learning Analytics, Federated Learning, Blockchain, Heterogeneous Knowledge Graph

1. Introduction

Learning Analytics is a socio-technical practice of collecting, measuring, and analyzing students' learning activity data to provide actionable insights that can be used to improve teaching and learning. Predictive Learning Analytics (PLA), a subfield of Learning Analytics, uses Machine Learning to predict students' likelihood of failing or dropping out of a course, enabling instructors to intervene and offer assistance to at-risk students [1]. This practice has grown in popularity and is becoming a "Business As Usual" activity in universities and other institutions. In its 2022 Horizon Report [2], EDUCAUSE, a nonprofit association whose mission is to advance higher education through the use of information technology, reported that institutions are now shifting from making "emergency" decisions, as a result of the COVID-19 Pandemic, to "long-term" planning. "Learning Analytics and Big Data" is reported to be one area where deployment is expected to rise, with the report paying some attention to Artificial Intelligence in Learning Analytics.

Big Data methodologies inherently bring about increased concerns regarding the protection, security, and management of the massive amount of data involved [3]. The core ethical and privacy issues in Learning Analytics include transparency, data ownership and control, accessibility, validity and reliability of data, institutional responsibility, communication, cultural values,

Woodstock'22: Proceedings of the ESWC 2023 Workshops and Tutorials, May 28-June 01, 2023, Hersonissos, Greece

*Corresponding author.

audrey.ekuban@open.ac.uk (A. Ekuban); john.domingue@open.ac.uk (J. Domingue)

0000-0002-2261-7218 (A. Ekuban); 0000-0001-8439-0293 (J. Domingue)

inclusion, consent, and student agency and responsibility [4]. There is an extra layer of risk present in PLA, as it necessitates connecting individual students' trace data to offer insights. This could potentially subject students to practices that are unethical or infringe upon their privacy.

One socio-technical concern not mentioned, is the issue of the use surveillance or "dataveillance" in education analytical tools. Some argue that Learning Analytics necessarily involves surveillance, and that there is no difference between Learning Analytics and a Learning Surveillance system [5], which might, for example, monitor a particular student's online attendance. This concern is one of the primary reasons for considering a decentralised approach to PLA.

1.1. Educational Mining with Privacy Rights and Ethics for Student Self-Sovereignty (EMPRESS)

As a solution to the above concerns, EMPRESS [6] is proposed. EMPRESS will be a Decentralised Machine Learning Pipeline containing several mechanisms to enable, as much as possible, the eight principles in The Open University's policy on the Ethical use of Student Data for Learning Analytics [7]. In the first instance, EMPRESS will pay particular attention to Principle 5 and Principle 6. Principle 5 focuses on transparent data collection. Principle 6 states that students should be engaged as active agents in the implementation of Learning Analytics, with informed consent being an example of this. Student engagement should seek to alleviate any mistrust that students have for Learning Analytics.

A Privacy by Design approach [8] will be adopted as follows:

- Proactive not Reactive: EMPRESS will eliminate the need to collect digital trace data.
- Privacy as the Default: In EMPRESS, a student would be presumed to not have opted into Learning Analytics until the student has decided to opt-in.
- Privacy Embedded into Design: EMPRESS is being designed with privacy as a key factor.
- Full Functionality—Positive-Sum, not Zero-Sum: EMPRESS will seek to demonstrate that it is possible for students to have both privacy and the benefits from PLA.
- End-to-End Lifecycle Protection: "Privacy by Default" and "Privacy Embedded into Design" will ensure that the Privacy by Design approach extends throughout the entire lifecycle of EMPRESS and any data involved.
- Visibility and Transparency ("trust but verify!"): The EMPRESS framework will utilise tools or techniques to implement Input and Output Verification, and Flow Governance [9]. It is anticipated that these tools / techniques will give assurance of any stated promises and objectives.
- Respect for User Privacy ("Keep it user-centric!"): In the case of the EMPRESS framework the individual is the student. EMPRESS will be inherently student-centric.

2. EMPRESS - Technical Aspects

There are 3 parts to EMPRESS: EMPRESS Administrator, EMPRESS Processor, and EMPRESS Learners. These represent the Data Controller, Data Processor, and students, respectively.

2.1. EMPRESS Administrator

The proposed solution incorporates a consent mechanism, implemented by the EMPRESS Administrator, that uses incentivisation based on cryptocurrency. The cryptocurrency is stored in a wallet, which uses private and public keys. Token Contracts, manage the crypto tokens. These contracts run on a blockchain and have a set of rules implemented through code. The Token Contract has a contract owner, and both the contract and owner have public addresses, allowing for verification of transactions and the contract itself.

The above approach to a consent mechanism will facilitate storing a student's consent decision on a Blockchain, and reading the student's consent decision from the Blockchain. The default student consent will be "Opt-Out". It should be noted that only the student's crypto currency's Account Address, along with the consent choice should be stored on the Blockchain. An institution would need to implement the consent screen in a manner that is consistent with its internal processes.

Figure 1 shows the Use Case Diagram to implement the consent mechanism, which can be summarised as follows:

- Student signs in with a Crypto Wallet. Student is effectively signing a message with a private key to verify ownership of a crypto currency account.
- Student's consent (based on the account address) is read from the Blockchain.
- Student can elect to change consent.
- The institution signs the transaction with one of its private keys and sends the signed transaction to the Blockchain, by invoking a Smart Contract function.
- Student can then join a Processor, which could be a Secure Aggregator being used with Federated Learning.
- A processor can use the Blockchain to verify a student's consent
- Student can receive crypto tokens from the institution or the processor.

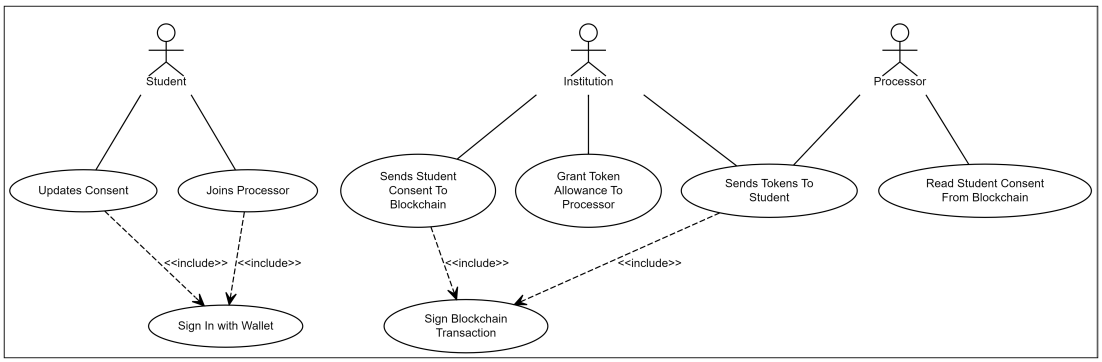

Figure 1: Use case diagram for consent mechanism

2.2. EMPRESS Processor and EMPRESS Learner

Based on current work, we have proposed the incorporation of a Federated Learning Data Aggregator [6]. Future work will address other privacy mechanisms, decentralised predictions and the ability of EMPRESS Learners to collaborate. As privacy mechanisms are added to EMPRESS, the blockchain can be used to indicate which are offered by an EMPRESS Processor.

For the Machine Learning model we focus on the Open University Dataset, OULAD, which contains anonymised data from a subset of Open University students that were registered in 2013 and 2014 [1]. OULAD, the output from Learning Analytics research, contains student demographic data, student aggregated Virtual Learning Environment (VLE) clickstream data and student assessment data. The data can be modelled as an Heterogeneous Graph. Heterogeneous Graphs have different types of information attached to nodes and edges. Figure 2 shows the node and edge features of OULAD. There are many techniques associated with heterogeneous graph embeddings [10], therefore some experimentation is required to determine which ones perform best for educational data. As student activity data is temporal, the incorporation of EvolvGCN [11], "which adapts the Graph Convolutional Network (GCN) model along the temporal dimension", requires some consideration.

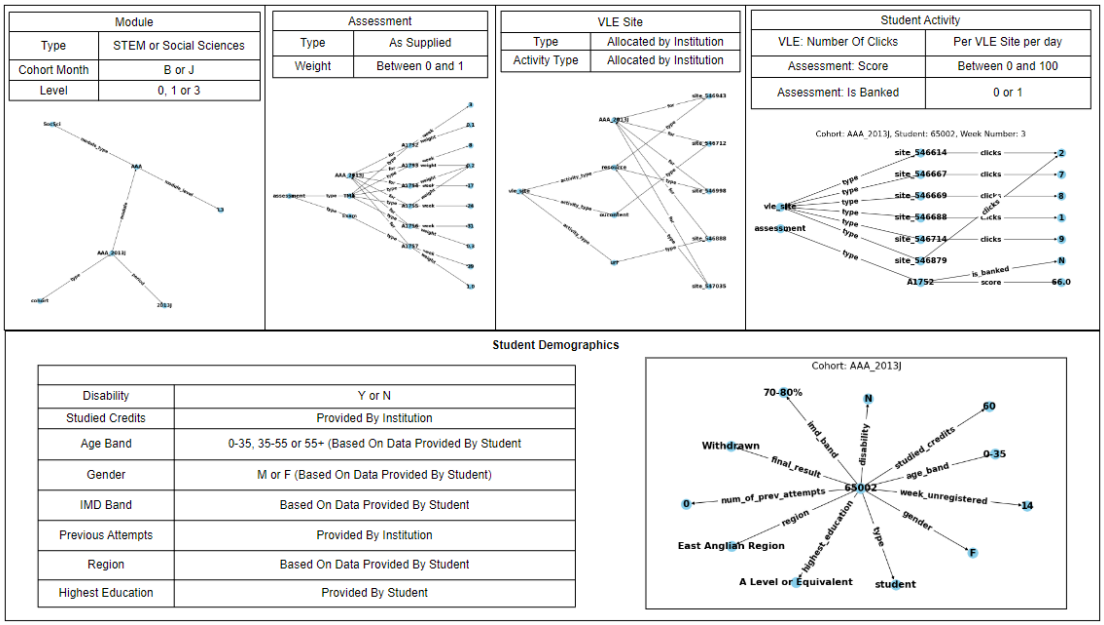

Figure 2: OULAD Heterogeneous Knowledge Graph example

3. Conclusion

This paper positioning introduced a way to gain consent for students in EMPRESS, a decentralised Machine Learning pipeline, with Blockchain technology as a supporting role. EMPRESS combines solutions for 1) Self-sovereign data, where students have ownership and control

over their data. 2) Federated Learning, where data scientists are able to build Machine Learning models using data that is not visible to them, and 3) Graph Convolutional Networks for Heterogeneous Knowledge Graphs.

Decentralised Predictive Learning Analytics aims to put students in control of their learning data and predictions. This approach reduces the risks associated with centralised predictive learning analytics, such as privacy concerns and the potential misuse of data by institutions. It also helps alleviate the perception that Learning Analytics is a surveillance tool by providing transparency and control to students. Ultimately, this methodology empowers students to take ownership of their learning and make more informed decisions about their academic progress. Additionally, the use of blockchain technology provides transparency to student consent decisions.

References

[1] J. Kuzilek, M. Hlosta, D. Herrmannova, Z. Zdrahal, J. Vaclavek, A. Wolff, Ou analyse: analysing at-risk students at the open university, Learning Analytics Review (2015) 1–16.

[2] K. Pelletier, M. McCormack, J. Reeves, J. Robert, N. Arbino, C. Dickson-Deane, C. Guevara, L. Koster, M. Sanchez-Mendiola, L. S. Bessette, et al., 2022 EDUCAUSE Horizon Report Teaching and Learning Edition, Technical Report, EDUC22, 2022.

[3] R. Rawat, R. Yadav, Big data: Big data analysis, issues and challenges and technologies, in: IOP Conference Series: Materials Science and Engineering, volume 1022, IOP Publishing, 2021, p. 012014.

[4] S. Slade, A. Tait, Global guidelines: Ethics in learning analytics (2019).

[5] M. Crosslin, Is learning analytics synonymous with learning surveillance, or something completely different?, 2019. URL: https://www.edugeekjournal.com/2019/10/30/is-learning-analytics-synonymous-with-learning-surveillance-or-something-completely-different/.

[6] A. Ekuban, J. Domingue, Towards decentralised learning analytics (positioning paper), in: Companion Proceedings of the ACM Web Conference 2023, 2023, pp. 1435–1438.

[7] Policy on ethical use of student data for learning analytics, 2014. URL: https://help.open.ac.uk/documents/policies/ethical-use-of-student-data/files/22/ethical-use-of-student-data-policy.pdf, accessed February 12, 2023.

[8] A. Cavoukian, Privacy by design: the definitive workshop. a foreword by ann cavoukian, ph. d, Identity in the Information Society 3 (2010) 247–251.

[9] A. Trask, E. Bluemke, B. Garfinkel, C. G. Cuervas-Mons, A. Dafoe, Beyond privacy trade-offs with structured transparency, arXiv preprint arXiv:2012.08347 (2020).

[10] X. Wang, D. Bo, C. Shi, S. Fan, Y. Ye, S. Y. Philip, A survey on heterogeneous graph embedding: methods, techniques, applications and sources, IEEE Transactions on Big Data (2022).

[11] A. Pareja, G. Domeniconi, J. Chen, T. Ma, T. Suzumura, H. Kanezashi, T. Kaler, T. Schardl, C. Leiserson, Evolvegcn: Evolving graph convolutional networks for dynamic graphs, in: Proceedings of the AAAI conference on artificial intelligence, volume 34, 2020, pp. 5363–5370.