Shih, JL. et al. (Eds.) (2023). Proceedings of the 31st International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Towards Automated Evidence Extraction: A Case Study of Adapting SAM to Real-World Educational Data

Kouki Okumura[a*], Izumi Horikoshi[b], Kento Koike[b] & Hiroaki Ogata[b]
[a]*Graduate School of Informatics, Kyoto University, Japan*
[b]*Academic Center for Computing and Media Studies, Kyoto University, Japan*
*okumura.kouki.27m@st.kyoto-u.ac.jp

Abstract: The demand for a shift from intuition- and experience-based to evidence-based education has been growing. A major challenge in realizing this transition is extracting evidence from real-world educational data. Conventionally, this data extraction process is performed by manually choosing classes for comparison. However, the selection requires expert knowledge of which classes should be compared with which indicators. In this study, we propose the use of deep learning algorithms to uncover inherent causal relationships within vast amounts of existing data. Specifically, we employ structural agnostic modeling, a causal search algorithm known for its exceptional performance on real-world data, to extract complex causal candidates in education. This approach has been referred to as "observational causal discovery." We evaluate the effectiveness of this method using real-world educational data and compare its advantages with those of conventional automatic comparison methods. Results demonstrate that the proposed method can identify various causal candidates, including those that are difficult for humans to discern, and even those without causal relationships. Different from existing methods, the proposed approach does not require selected or fixed comparative indices, thus potentially uncovering comparative indices that elude human comprehension. We anticipate that this research will enable the collection of substantial evidence from real-world educational data and promote evidence-based education.

Keywords: Evidence-based education, real-world evidence, SAM, causal search, causal analysis

1. Introduction

Evidence-based education (Davies, 1999) is expected to provide education that is not based on intuition or experience. *Evidence* is generally extracted from systematic reviews and randomized controlled trials (RCTs). Although these methods provide high-level evidence, collecting large amounts of evidence is difficult because doing so is often costly, unethical, or unfeasible (Slade & Prinsloo, 2013). Therefore, the concept of real-world evidence has been proposed in the medical field (Mahajan, 2015), and several attempts have been made to apply it to the educational field as well. Real-world evidence is evidence extracted from various real-world data, and although it is less reliable than that extracted from systematic reviews and RCTs, it is expected to be voluminous.

Our previous method compares a control group with a corresponding target group to verify the effects of an intervention (Nakanishi, 2021). This method is advantageous as it allows the effects of an intervention to be verified by setting target and control groups, thereby eliminating the need for experiments. However, it requires manually setting the classes to be compared, and its use in automatic extraction from many cases is difficult.

To address this issue, we developed a method that can automatically search for classes with similar contextual information, such as grade level, as a control group (Okumura et al., 2022). However, this method allows only one factor to be considered and compared. To further update this method, we adopted a method called "observational causal discovery."

Observational causal discovery is the process of hypothesizing causal relationships from observational data. This method has attracted attention from the machine learning community, thus reflecting the difficulty of conducting RCTs, which are the gold standard for establishing causal relationships (Lopez-Paz et al., 2015; Mooij et al., 2016; Peters et al., 2017). Observational causal discovery has been applied to many domains, such as in economics to understand and model the impact of monetary policy and in bioinformatics to infer network structures from gene expression data and prioritize exploratory experiments (Kalainathan et al., 2018, 2022). However, we have yet to find many applications of observational causal discovery in educational data. Therefore, we raise the following two research questions (RQs):

- RQ1: What do we obtain from the observational causal discovery method adapted to real-world educational data?
- RQ2: How different are the results of the conventional method and observational causal discovery?

To address the above RQs, we attempt to apply the observational causal discovery method, which has already been applied in other fields, to real-world educational data.

For RQ1, we apply the causal analysis algorithm called structural agnostic modeling (SAM), which has shown great performance in real-world biological data, to real-world educational data and verify its effectiveness. For RQ2, we compare our previously proposed method with the method using SAM and clarify their characteristics and differences. Through these RQs, we aim to present a new path for more effective evidence extraction from educational data.

2. Method

2.1 Overview

The methodology used in this study consists of two steps. The datasets used are listed in Table 1. The first step is for RQ1, and it involves examining whether SAM is applicable to educational data and whether it provides effective results. Specifically, we select dataset Y for analysis and apply SAM to it. The second step is for RQ2, and it involves comparing the SAM-based method with our previously proposed method. For this step, we apply SAM to dataset Z, which is from dataset X for the same activity as that in the previous study by Okumura et al. (2022).

In Japan, one tablet or a similar device is distributed to each student under the GIGA School Project. This project uses a learning analytics platform called the LEAF system (Ogata, 2018) and collects the log data accordingly. In the current study, we utilize the log data from the LEAF system.

The analysis is conducted as follows (Figure 1): First, the log data containing information on the students' activities in the platform, including what the activities are and when they are completed, is collected. Second, this information is summarized to show the number of operations with the digital textbooks in each class. Third, a causal search is conducted to determine whether a causal relationship exists. Fourth, the results are displayed as a table in the form of 0s and 1s according to a set threshold. Finally, the results are output as a directed acyclic graph (DAG).

Table 1. *Comparison of Our Approach and Conventional Studies.*

Dataset	Analysis Method	Data source	Publication
-	Manual Comparison	a junior high school	Nakanishi (2021)
Dataset X	Automated Comparison	a junior high school	Okumura et al. (2022)
Dataset Y	Automated Causal Exploration (SAM)	a university	This study
Dataset Z	Automated Causal Exploration (SAM)	a junior high school	This study

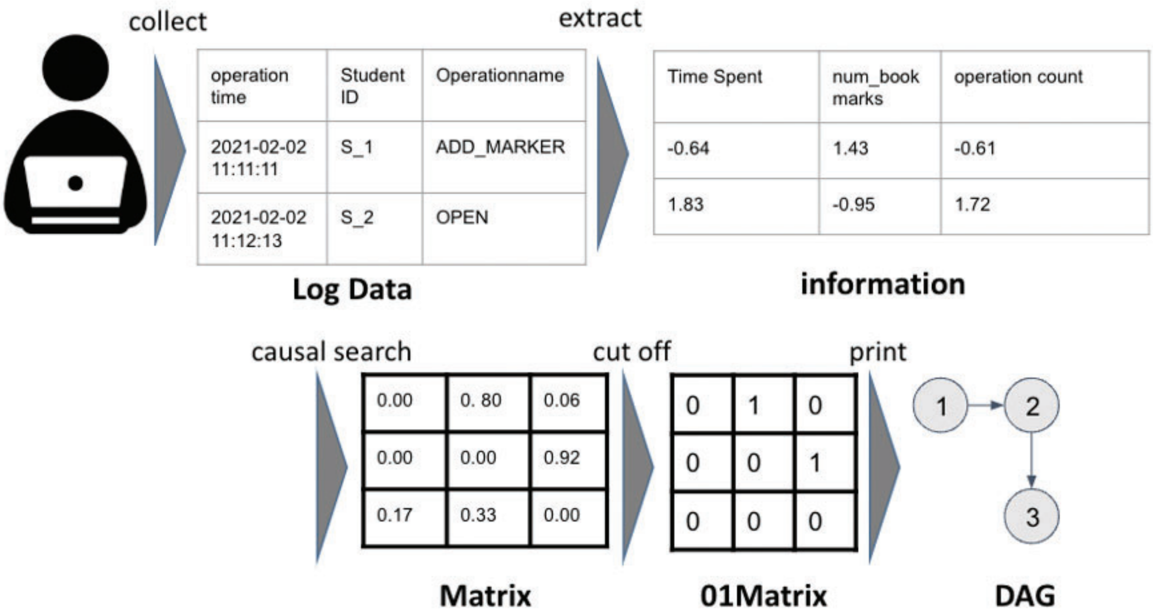

Figure 1. Method Overview.

2.1.1 Dataset from a University (Dataset Y)

We used the data collected from an educational data analysis contest conducted by the Council for Evidence-Driven Education Research (EDE, 2022).

Specifically, we utilized the educational data obtained with the LEAF system at a university. These data included digital educational material browsing behavior in four courses offered as Kyu data in the 2019 and 2020 academic years (Table 2). The logs were organized according to class time and content IDs, and data were produced for each indicator for each class. The indicators are listed in Table 3.

Table 2. *Summary of Dataset Y.*

Course Id	Term	Student	Log Count
A-2019	2019 Spring	50	129,358
B-2019	2019 Spring	164	352,467
A-2020	2020 Spring	62	147,452
B-2020	2020 Spring	93	197,593

Table 3. *Indicators from Dataset Y.*

Abbreviation	Description
CNT LP	Count of Lecture Period
SUM OP	Sum of Operations
AVG OP	Average Operations
SUM ATT	Sum of Attending Students
PG LP	Number of Pages Covered in Lecture Period
DV LP	Number of Devices Used in Lecture Period
SUM MRK	Sum of Markers
AVG MRK	Average Markers
NUM MAT	Number of Teaching Materials Used in Lecture Period
AFE ATT	Average Final Exam Score of Attending Students
ONLINE	Whether the Course was Online or Not

2.1.2 Dataset from a Junior High School (Dataset Z)

For Dataset Z, we analyzed the logs of the days when active reading activities were conducted at the same junior high school in school years 2021 and 2022. Active reading is "considered best suited for textbook reading, helping learners understand unfamiliar information in texts, and providing a structured approach to learning" (Toyokawa et al., 2023, p. 2). Active reading enables one to read quickly and understand important points, and it involves the use of multiple learning tools, such as e-books and learning analytics tools. The target class consisted of three sessions, namely, Day 1, Day 2, and Day 3. E-book memos, markers, and timers were used in the activities. In addition, quizzes were administered, and reading speed (words per minute, WPM) was measured before and after the class (pre- and post-class). The indicators computed from the log data of these activities are listed in Table 4.

Table 4. *Indicators of Dataset Z*

Abbreviation	Description
ARD D1	Active reading dashboard use on Day 1
ARD D2	Active reading dashboard use on Day 2
ARD D3	Active reading dashboard use on Day 3
MRK D1	Marker use on Day 1
MRK D2	Marker use on Day 2
MEM D1	Memo use on Day 1
MEM NC	Memo use on a day with no class
MEM D2	Memo use on Day 2
MEM D3	Memo use on Day 3
RDG D1	Reading operations on Day 1
RDG NC	Reading operations on a day with no class
RDG D2	Reading operations on Day 2
RDG D3	Reading operations on Day 3
TIM D1	Timer use on Day 1
TIM D2	Timer use on Day 2
TIM D3	Timer use on Day 3
QUZ PRE	Score in the pre-quiz
QUZ PST	Score in the post-quiz
WPM PRE	Words per minute before the active reading activity
WPM PST	Words per minute after the active reading activity
SUM SMR	Score of summary after active reading class

2.2 *Applied Algorithm: SAM*

SAM (Kalainathan et al., 2018, 2022) utilizes a type of deep learning called generative adversarial network (GAN). GAN is known for its ability to generate fake objects that are close to real objects by having two networks, namely, a generative network and a discriminative network, compete. SAM is an application of the GAN concept to causal analysis and is used to determine causal relationships among real-world educational data (Figure 2). Specifically, SAM quantifies the likelihood of causal relationships between indicators and examines whether causal relationships exist. The hyperparameters are the same as those used in the original study.

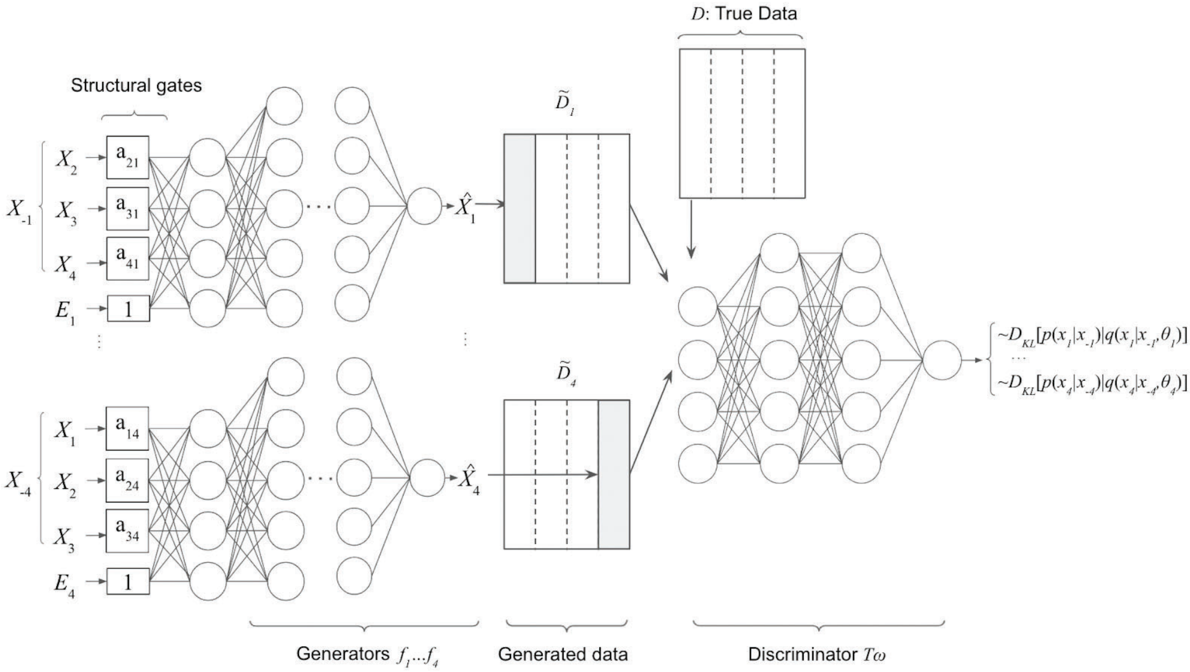

Figure 2. Algorithm of Structural Agnostic Modeling (SAM)
(Adopted from Kalainathan et al. (2018, 2022))

3. Results

3.1 *RQ1: What do we obtain from the observational causal discovery method adapted to real-world educational data? (Dataset Y)*

To answer RQ1, we examined whether SAM is applicable to educational data and whether it provides effective results. SAM was applied to dataset Y for analysis. Figure 3 shows the result of SAM for dataset Y.

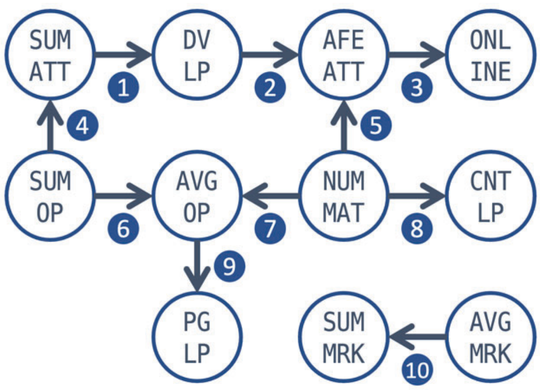

Figure 3. Result of SAM for Dataset Y.

(1) Intuitively understandable result
The analysis revealed a causal relationship between the number of teaching materials used during class (MV LP) and the average number of operations (AVG OP) (Figure 3, arrow 7). This result is intuitively understandable result: the more types of teaching materials available, the greater the number of operations.

(2) Counterintuitive results
However, a causal relationship was found between the number of device types used in class (DV LP) and the semester grade point average of the student attendees (AFE ATT) (Figure 3, arrow 2). This result counters the intuition as more types of devices lead to higher grades. If this relationship is actually a causal one, then it is a new educational technique worth trying. The causal relationship between whether the course was offered online (ONLINE) and end-of-semester grades (AFE ATT) (Figure 3, arrow 3) was also counterintuitive. The notion of academic grades increasing with the availability of online classes is difficult to envision. However, if such relation is true, then new educational possibilities will become apparent.

(3) Reversed direction causal relationship
A causal relationship was found between the total number of operations (SUM OP) and the number of participating students (SUM ATT) (Figure 3, arrow 4). However, this causal relationship was reversed. The number of operations, of course, did not increase the number of students. Instead, the greater the number of students, the more operations involved; this case shows a more natural causal relationship.

3.2 RQ2: How different are the results of the conventional method and observational causal discovery? (Dataset Z)

To answer RQ2, we applied SAM to dataset Z, which was from the same activity as that in the previous study (dataset X). Figure 4 shows the result of SAM for dataset Z.

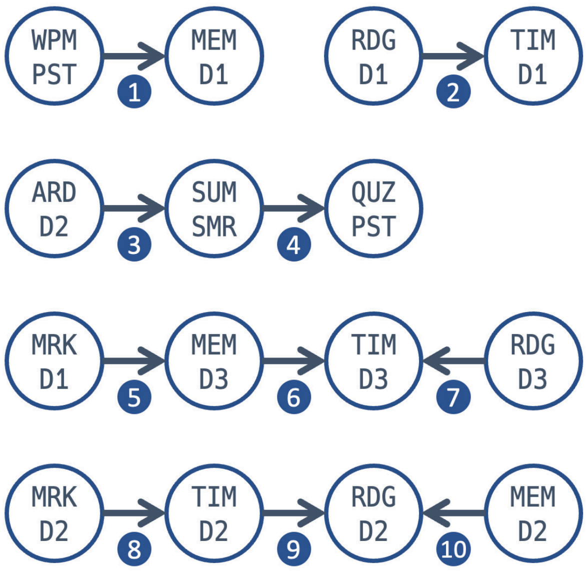

Figure 4. Result of SAM for Dataset Z.

(1) Effects of interventions
The analysis revealed the effects of the interventions we aimed to validate. For example, we found that active reading activities on Day 2 (SUM ATT) affected the quality of memo summaries (SUM SMR) (Figure 4, arrow 3c).

(2) Reversed direction causal relationship
In addition, we noted some reversals of the time axis, which suggested the possibility of causality. For example, we observed a relationship between WPM in the post-class activity (WPM PST) and memo use on Day 1 (MEM D1) (Figure 4, arrow 1). Given the order of the activities, the direction of the causal relationship must be reversed. However, if the direction was adjusted, the causal relationship between memo use and WPM cannot be easily found through human observation. Hence, AI that can process large amounts of data is at an advantage as it is able to suggest the existence of relationships that are not considered relevant by humans.

(3) Absence of causal relationship
The results also suggest not only the "possibility that a causal relationship exists" but also the "possibility that no causal relationship exists." For example, we did not observe an edge between active reading dashboard use on Days 1 and 3 (ARD D1 and ARD D3). This result indicated that the activities with the dashboard on Days 1 and 3 were not highly effective, whereas those on Day 2 might have been very effective. In future designs for active reading classes, an emphasis on the activities conducted on Day 2 could be considered. If effective, such designs will contribute to improving the learning effectiveness of learners while reducing the burden on teachers.

4. Discussion

4.1 Results of RQs

(*RQ1) What do we obtain from the observational causal discovery method adapted to real-world educational data? (Dataset Y)*

From these results, we confirmed the possibility of applying SAM to educational data and generating useful results. However, not all causal relationships shown in the DAG were true; therefore, care must be taken when interpreting them. An advantage of this method is that it can process large amounts of data and suggest causal relationships that humans may not be able to assume. Our results also implied the possibility of selecting the causal relationships that should be examined more deeply. In addition, the results indicated that SAM can be useful for educational data analysis. However, maximizing this method requires the proper interpretation and verification of the causal relationships in the output.

(*RQ2) How different are the results of the conventional method and observational causal discovery? (Dataset Z)*

We compared the automatic causal search method SAM with the previous automatic comparison method and found some differences. Specifically, SAM does not require the indicators to be fixed, whereas the previous method requires the indicators before comparison to be fixed. In this way, SAM may expand the exploration of factor candidates and lead to the possibility of finding comparative indices that humans would not have been able to consider. This feature is an advantage of the automatic causal search method and indicates its potential to open the door for new research.

4.2 Limitations

Addressing RQ2 involved extracting and analyzing the data with background information provided by a researcher who conducted an active reading class study using the LEAF platform. In fully automating the mining process without the need for any manually inputted information, the background information from which the data were obtained and that is tied to the data in some way must be stored to automate the extraction of evidence. We must note that DAG assumes that education is acyclic, although various and complex causal relationships can be hypothesized. SAM produces candidate causal relationships that have the time axis reversed and thus needs improvement so that it does not produce such relationships.

4.3 Implications

Unlike the conventional method, the method in this study was not limited to a single indicator for comparison. Hence, the method proposed herein is able to find comparative indices that humans cannot assume, such as the relationship between memo use and WPM in Section 3.2 (2).
As shown in Section 3.2 (3), knowing specifically which day's activities are effective may reduce the burden on teachers and improve learning effectiveness.

5. Conclusion and Future Works

This study attempted to apply SAM to the automated extraction of evidence from real-world educational data and examined its effectiveness. The results showed that the causal relationships output by SAM do not necessarily indicate true causal relationships. However, such results are not necessarily meaningless, and the ability to automatically process large amounts of data presents the possibility of various causal relationships that are difficult to identify based on human assumptions.

In addition, through the proposed method, we showed the possibility of proceeding with the analysis even without fixing the indicators for comparison. This feature enables the identification of comparative indices that humans cannot assume. It is also an advantage of using AI and computers as it can provide new perspectives that go beyond those derived by humans.

Future studies should conduct causal inferences in addition to causal search and evaluate the magnitude of causal effects. In addition, they may compare our method with other algorithms for causal search. One of the goals of the current research was to implement the proposed method as a system for teachers and provide real-world evidence of learning. In achieving this goal, the method must be improved and developed while obtaining feedback from teachers based on the assumption that the tool will be used in an actual educational environment.

Acknowledgements

This study was supported by NEDO JPNP20006 and JSPS KAKENHI JP23H00505.

References

Davies, P. (1999). What is evidence-based education?. *British Journal of Educational Studies, 47*(2), 108-121.
EDE: Council for Evidence-Driven Education Research (2022, September). Educational data analysis contest. Retrieved from https://sites.google.com/view/ede-datachallenge-23/

Kalainathan, D., Goudet, O., Guyon, I., Lopez-Paz, D., & Sebag, M. (2018). Structural agnostic modeling: Adversarial learning of causal graphs. *arXiv preprint arXiv:1803.04929*. Retrieved from https://arxiv.org/abs/1803.04929

Kalainathan, D., Goudet, O., Guyon, I., Lopez-Paz, D., & Sebag, M. (2022). Structural agnostic modeling: Adversarial learning of causal graphs. *Journal of Machine Learning Research, 23*(219), 1-62.

Lopez-Paz, D., Muandet, K., Schölkopf, B., & Tolstikhin, I. (2015, June). Towards a learning theory of cause-effect inference. In *International Conference on Machine Learning* (pp. 1452-1461). PMLR.

Mahajan, R. (2015). Real world data: additional source for making clinical decisions. *International Journal of Applied and Basic Medical Research, 5*(2), 82.

Mooij, J. M., Peters, J., Janzing, D., Zscheischler, J., & Schölkopf, B. (2016). Distinguishing cause from effect using observational data: Methods and benchmarks. *The Journal of Machine Learning Research, 17*(1), 1103-1204.

Nakanishi, T. (2021). *Extracting evidence on lesson design using real world educational data*. [*Un published master's thesis*]. Kyoto University. (*In Japanese*)

Ogata, H., Majumdar, R., Akçapınar, G., Hasnine, M. N., & Flanagan, B. (2018). Beyond learning analytics: Framework for technology-enhanced evidence-based education and learning. In *26th International Conference on Computers in Education Workshop Proceedings* (pp. 493-496). Asia-Pacific Society for Computers in Education (APSCE).

Okumura, K., Horikoshi, I., & Ogata, H. (2022). Development of a "control group" search method for automated evidence extraction from real-world education data. *IPSJ Technical Report* (Vol. 2022-CLE-38, No. 2, pp. 1-5), Japan. (*In Japanese*)

Peters, J., Janzing, D., & Schölkopf, B. (2017). *Elements of causal inference: Foundations and learning algorithms* (p. 288). The MIT Press.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1510-1529.

Toyokawa, Y., Majumdar, R., Kondo, T., Horikoshi, I., & Ogata, H. (2023). Active reading dashboard in a learning analytics enhanced language-learning environment: Effects on learning behavior and performance. *Journal of Computers in Education*, 1-28.