Machine Learning for Multimodal Learning Analytics and Feedback

Hiroaki Kawashima[1]

[1]University of Hyogo, Kobe, Hyogo 6512197, Japan

Abstract

Multimodal measurement of human learning enables a data-driven approach to learning analysis and feedback generation. This position paper discusses the possibility of feedback based on multimodal learning analytics from how machine learning methods can be applied. In particular, we first discuss how (1) behavioral measurements, such as learner browsing logs and eye tracking, and (2) content analysis of learning materials can lead to (3) prediction and modeling of learners' states (e.g., performance) and (4) feedback generation, such as information presentation and content optimization, through some research examples. We then show future research directions of machine learning-based learners' state modeling for feedback generation.

Keywords

learning analytics, multimodal, e-book log, eye tracking, content generation

1. Introduction

Various measurements of learning behaviors, such as clickstreams of e-book manipulation [1] and gazing patterns at lecture videos [2, 3, 4], have been introduced to study individual-adaptive learning in higher education. These multimodal observation data enable a precise data-driven learning analysis, which had been based mainly on the instructor's experience. On the other hand, the advancement of machine learning enables detailed content analysis of text and image data [5, 6]. By integrating (1) multimodal measurement of human learning behaviors and (2) multimodal analysis of learning materials through fine-grained learning analytics, we are aiming at (3) estimating learners' states and (4) generating feedback to individuals adaptive to their situations. In higher education, how to increase feedback frequency to students with a limited number of teaching staff is an important issue. Data-driven or machine-driven feedback has the potential to support various types of learning, not only for students but for the improvement and optimization of teaching methods and learning materials on the teachers' side.

The concept of feedback loops with machine learning models has been discussed in the field of multimodal learning analytics (MMLA) (e.g., [7]), where the analysis of multimodal learning behavior is the main focus. This position paper extends the idea of feedback loops in learning analytics, focusing on using content analysis (i.e., learning material) with behavioral data and automatic content generation based on machine learning techniques. In Sec. 2, we discuss

CrossMMLA (LAK 23 Pre-conference Workshop), March 14, 2023, Arlington, TX, USA

kawashima@gsis.u-hyogo.ac.jp (H. Kawashima)
0000-0003-1208-3715 (H. Kawashima)

CEUR Workshop Proceedings (CEUR-WS.org)

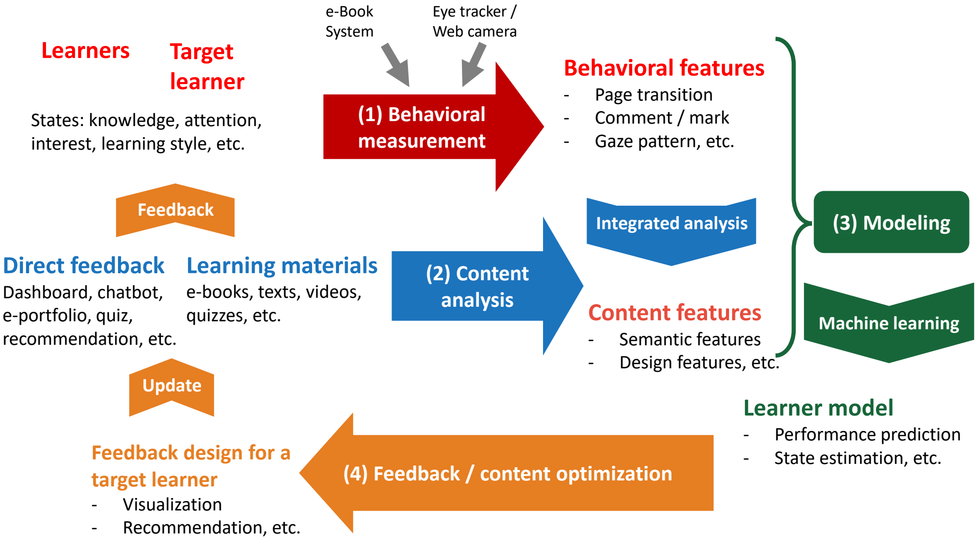

Figure 1: Overview of the feedback loop assumed in this research.

technical components in our assumed feedback loop by limiting learning context to e-book and video viewing-based learning situations. Section 3 shows some examples of performance prediction models and slide-content generation, which have the potential to be used for feedback. We then discuss an important research question for feedback generation “what is an appropriate representation of learners’ state?” in Sec. 4.

2. Technical Components for the Feedback Loop

The overall structure of the feedback loop envisioned in this paper is shown in Fig. 1. We consider electronic text browsing logs (e.g., clickstreams) and eye-tracking data for the learners’ behavioral measurement in Fig. 1 (1). With the spread of learning management systems (LMSs) and massive open online courses (MOOCs), learners’ behaviors, such as signing in/out and assignment submission, are obtained by system logs. Besides, more detailed behaviors can now be measured by e-book systems and online-course systems as operation logs of lecture materials (e.g., page transitions, adding or deleting markers) [1] and lecture videos (e.g., pause, rewind) [3]. Furthermore, additional devices and software enable multimodal behavioral measurements, such as eye-gaze tracking, which tell us fine-grained in-class activities of learners [2, 4]. This paper assumes learning situations where such e-book systems or eye-gaze trackers are introduced.

Meanwhile, the potential and application range of content analysis in Fig. 1 (2) are now increasing because of the advancement of machine learning. In particular, its media-analysis capability covers a wide range of contents, including textbooks, quiz questions, lecture slides, videos, and audio. Neural networks such as recurrent neural networks (RNNs) and Transformers [8] have recently made it possible to obtain detailed features of words, sentences, and documents. The same applies to image features using convolutional neural networks (CNNs). As a result, the features of multimodal content can be extracted in vector representations, which can be used for various types of analysis, such as performance prediction and similarity analysis.

Behavioral data captured in (1) and the content features obtained in (2) are then used for learners' modeling, such as performance prediction and the estimation of learners' state (e.g., mental and cognitive states), using machine learning methods (Fig. 1 (3)). As many machine learning models lack interpretability or explainability, we need to consider how explainable the models should be, which depends on feedback objectives (see Sec. 4 for detailed discussion).

For the feedback step in Fig. 1 (4), it is important to decide whom to target and when and what to provide. For example, it is possible to provide feedback to teachers on which slides students browse during class [9]. Providing information on which slides or topics students struggled with is also helpful in supporting their reflections after the class. Furthermore, personalized summaries or converted content (e.g., videos and slides for easier comprehension) can be generated by optimizing learning materials. Here, various types of feedback can be considered through MMLA depending on how the components from (1) to (4) are used.

3. Examples of Learning Analytics toward Feedback Generation

This section introduces research examples we are working on toward realizing feedback to show some technical components described in Sec. 2 and to discuss how they can be combined.

3.1. Performance Prediction through Behavioral and Content Analysis

In order to predict each student's performance, we can utilize both (a) what learning content is used and (b) how each student behaves. We here hypothesize that combining the content information from (a) and behavioral features from (b) will enable us to achieve more accurate performance prediction compared to only using content-independent features (b).

Browsing-log data on an e-book system. Some universities use e-book systems to obtain the log data of students' browsing behavior when manipulating textbooks or slides [1]. Using such browsing log data allows various data mining and analytics, such as the discovery of behavioral changes before and after COVID-19 [10] and grade prediction based on machine learning [11], become possible. With such log data, we predict quiz scores by combining information on slide content (e.g., text) with the features of students' behavioral data (i.e., operation logs obtained from the e-book system). Here, we utilize Sentence-BERT [12] to obtain the embedded vector representation of each slide page and take the weighted sum of the vectors using the duration of page viewing as the weights. Then, the obtained vector is used as input for a gradient-boosting model to predict scores. The results show that using such features increases the accuracy of predicting the quiz scores, which suggests that content-dependent behavioral features are informative in predicting each learner's performance level.

Eye tracking data during video viewing. Eye tracking is used for a finer-grained measurement of learners' behavior during viewing lecture videos. The gazing point series can be used to estimate learners' various states, including "mind wandering" [13] and "actively examining slide materials" [4]. In addition, as with browsing log data, gaze data may include features related to students' performance [2]. Here, we incorporate attentional states estimated from eye-gaze data for predicting students' performance by exploiting a probabilistic model of switching attentional states [4]. The model automatically estimates sequences of attentional states by assuming that each of the three gaze distributions, including time-dependent and content-dependent distributions, corresponds to each attentional state. Our ongoing research suggests that not only content-dependent gaze features, such as where the learner looks and what is looked at, but the estimated attentional states contribute to quiz-score prediction.

The prediction of students' performance (e.g., final grades, comprehension of each topic) can be used to generate various feedback for both students and teachers. Feedback forms include visualization on dashboards that learners can check themselves, recommendation of learning materials, interactive support (e.g., chatbots), and the detection of students who need help from teachers. The implementation and verification of these feedback methods need to be investigated in the future.

3.2. Automatic Content Generation and Emphasis

Research on feature extraction and automatic text summarization using neural networks is rapidly advancing in the natural language processing (NLP) community. Text feature extraction methods, including word embedding (e.g., Word2Vec [14]) and Transformer encoders (e.g., BERT [5]) are widely used for text analysis exploiting vector representations that encode the similarity of content.

Lecture-video emphasis using audio. Those vector representations and similarity measures are key to finding related text in learning materials. In our ongoing work, we utilize those vector representations of text data to realize automatic spotlighting of video content. Once the lecturer's utterance is processed by text-to-speech, the similarity between slide text and speech is computed, and the corresponding regions can be emphasized to guide learners' attention.

Slide generation from documents. Transformer-based automatic slide generation from documents [15, 16] has recently attracted the attention of researchers in the machine learning community. While the text layout in a slide is not considered in these studies, we try to generate slides from a Wikipedia page by automatically selecting text layouts to improve the slide readability. To select appropriate layouts, we train and utilize a BERT-based classifier for estimating the discourse relationship of a given sentence pair.

As our ongoing work on automatic content highlighting and generation are not individualized, it needs to be combined with learners' performance modeling described in Sec. 3.1. In addition, not only the information of content but learners' behavior data (Fig. 1 (1)) can also be used to find important topics and pages that should be emphasized [17]. Therefore, the integration of (1), (2), and the learner model (3) is also an interesting challenge.

4. Machine Learning for Learners' Modeling

Machine learning techniques can be applied for various objectives, from predicting students' performance to generating content, in the context of learning analytics loop as described in Sec. 2 and Sec. 3. However, it has not yet been elucidated on "what kind of learners' information is required to provide appropriate feedback and how machine learning models can estimate such information." Machine learning could contribute to learners' state modeling in the following three levels: feature level, manually designed level, and automatically extracted level. In this section, we discuss the above questions regarding the learners' state representation and present several challenges.

Feature-level representation. The most straightforward representation of a learner's state is the amount of activity on learning materials or topics, such as the frequency the learner has viewed slide pages, texts, figures, and the count of correctly answering topic-related questions. Once the similarity among learning materials or topics is extracted automatically using machine learning techniques (e.g., Transformer encoders described in Sec. 3.2), the similarity structure of learning content provides essential information to infer learners' knowledge in detail in behavioral and content-related feature space. This leads to the generation of meaningful feedback, including recommendations and optimization of learning content. Automatic or semi-automatic knowledge extraction from learning materials (e.g., [18]) may also facilitate this process.

Manually designed state representation. The second direction is to manually design a learner's state (e.g., emotion, attention, cognition, knowledge levels, skills, key competencies). While these variables cannot be directly observed from behavioral data, it is possible to estimate them using machine learning models once the model is trained using annotated datasets [7]. The results of quizzes, tests, and exams are also used as the training data for the model. The performance prediction studies introduced in Sec. 3.1 are examples of this direction, which infer the state of learners from their behaviors even when they do not take exams. Note that the recent trend of machine learning enables integrating different modalities, such as multimodal behavioral data and learning materials, using embedded vector representations extracted from various encoder models. The estimated knowledge states can be used to generate a variety of feedback, such as learning material recommendations, as described in Sec. 3.1.

Automatically extracted state representation from data. Machine learning could contribute to extracting learners' states automatically as a latent representation in a model through behavior analysis. This direction corresponds to the "end-to-end modeling of the learner's internal state" described in Sec. 2. For example, the techniques of Knowledge Tracing [19], which model the knowledge state of individual learners from a series of questions and their answers, have been rapidly advanced with machine learning models. While it also predicts learners' performance, similar to Sec. 3.1, a learner's state is obtained as a latent variable through end-to-end model training. This opens up the possibility of finding a useful representation of learners' states from the data without annotation, although there are difficulties in ensuring explanatory and interpretability. To make the latent variable in the end-to-end model more interpretable, we can further impose the model with prior or external knowledge as the design of model structures, parameter constraints, and regularizers. Graph neural networks [20] are examples of such model structures that would allow the smooth integration of knowledge in learning theory.

5. Conclusion

This position paper introduced how the recent machine learning techniques can be applied to the components of MMLA-based feedback loops, focusing on integrated content analysis with behavioral data and content generation. In particular, we discussed that the recent trend of representation learning would open up new possibilities for integrating multimodal behavioral and contextual data. To estimate learners' states in deep and generate appropriate and detailed feedback, the field of learning analytics and machine learning can collaborate in many aspects, and this leads to a new framework of feedback loops in MMLA.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP19H04226.

References

[1] H. Ogata, M. Oi, K. Mohri, F. Okubo, A. Shimada, M. Yamada, J. Wang, S. Hirokawa, Learning Analytics for E-Book-Based Educational Big Data in Higher Education, Smart Sensors at the IoT Frontier, Springer, Cham (2017) 327–350.

[2] K. Sharma, P. Jermann, P. Dillenbourg, "With-me-ness": A Gaze-Measure for Students' Attention in MOOCs, International Conference of the Learning Sciences (ICLS) (2014) 1017–1022.

[3] J. Kim, P. J. Guo, D. T. Seaton, P. Mitros, K. Z. Gajos, R. C. Miller, Understanding In-Video Dropouts and Interaction Peaks in Online Lecture Videos, International Conference on Learning @ Scale (2014) 31–40.

[4] H. Kawashima, K. Ueki, K. Shimonishi, Modeling Video Viewing Styles with Probabilistic Mode Switching, International Conference on Computers in Education (ICCE) (2019).

[5] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (2019). arXiv:1810.04805.

[6] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, International Conference on Learning Representations (ICLR) (2015).

[7] D. Di Mitri, J. Schneider, M. Specht, H. Drachsler, From Signals to Knowledge: A Conceptual Model for Multimodal Learning Analytics, Journal of Computer Assisted Learning 34 (2018) 338–349. doi:10.1111/jcal.12288.

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention Is All You Need (2017). arXiv:1706.03762.

[9] A. Shimada, S. Konomi, H. Ogata, Real-Time Learning Analytics System for Improvement of on-Site Lectures, Interactive Technology and Smart Education 15 (2018) 314–331.

[10] H. Kawashima, Comparison of Learning Behaviors on an e-Book System in 2019 Onsite and 2020 Online Courses, International Conference on Educational Data Mining (2022) 753–757.

[11] G. Akçapınar, M. N. Hasnine, R. Majumdar, B. Flanagan, H. Ogata, Developing an Early-

Warning System for Spotting at-Risk Students by Using eBook Interaction Logs, Smart Learning Environments 6 (2019).

[12] N. Reimers, I. Gurevych, Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks (2019). arXiv:1908.10084.

[13] S. Hutt, J. Hardey, R. Bixler, A. Stewart, E. Risko, S. K. D. Mello, Gaze-based Detection of Mind Wandering during Lecture Viewing, International Conference on Educational Data Mining (2017) 226–231.

[14] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient Estimation of Word Representations in Vector Space (2013). arXiv:1301.3781.

[15] E. Sun, Y. Hou, D. Wang, Y. Zhang, N. X. R. Wang, D2S: Document-to-Slide Generation Via Query-Based Text Summarization (2021) 1405–1418. doi:10.18653/v1/2021.naacl-main.111. arXiv:2105.03664.

[16] T.-J. Fu, W. Y. Wang, D. McDuff, Y. Song, DOC2PPT: Automatic Presentation Slides Generation from Scientific Documents, AAAI Conf. on Artificial Intelligence 36 (2022) 634–642.

[17] A. Shimada, F. Okubo, C. Yin, H. Ogata, Automatic Summarization of Lecture Slides for Enhanced Student Preview-Technical Report and User Study, IEEE Transactions on Learning Technologies 11 (2018) 165–178.

[18] A. Fiallos, X. Ochoa, Semi-Automatic Generation of Intelligent Curricula to Facilitate Learning Analytics, International Conference on Learning Analytics & Knowledge (LAK) (2019) 46–50.

[19] C. Piech, J. Spencer, J. Huang, S. Ganguli, M. Sahami, L. Guibas, J. Sohl-Dickstein, Deep Knowledge Tracing (2015) 1–12. arXiv:1506.05908.

[20] H. Nakagawa, Y. Iwasawa, Y. Matsuo, Graph-based Knowledge Tracing: Modeling Student Proficiency Using Graph Neural Network, International Conference on Learning Representations (2019).