Rodrigo, M. M. T. et al. (Eds.) (2021). Proceedings of the 29th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Learning Analytics Dashboard Prototype for Implicit Feedback from Metacognitive Prompt Responses

May Kristine Jonson CARLON & Jeffrey S. CROSS*
School of Environment and Society, Tokyo Institute of Technology, Japan
*cross.j.aa@m.titech.ac.jp

Abstract: Online learning can be challenging to learners as they need to have autonomous learning skills to succeed, and to instructors as direct observation and real-time communication with learners are limited. Learning analytics dashboards have been used to assist the learners in developing autonomous learning skills and the instructors in keeping track of the learners' progress. However, there is little information on systems supporting both learners and instructors in online learning environments. This paper builds on our previous work developing learners' metacognitive skills through open response prompts by using the learner inputs to create a dashboard that uncovers implicit feedback such as sentiments, misconceptions, and shallow learning. The instructor can consult the dashboard on-demand, and the input is from metacognitive prompts that only the individual learners see. Hence, the instructor can provide timely interventions based on inputs from learners who otherwise would not voice their concerns in more public channels such as discussion forums.

Keywords: Topic modeling, sentiment analysis, text similarity, metacognitive prompting, learning management system, learning analytics

1. Introduction

In a traditional classroom, learners have multiple avenues of interaction with the instructor available to them. When the learner needs something, they can raise their hand during class or approach the instructor after the class finishes or during office hours. Likewise, the instructor can gauge the learners' engagement by observing what is happening inside the classroom and with follow-up interactions outside of class. Even if the classroom interactions are minimal, there are still opportunities for the instructors and learners to exchange feedback during several activities, including homework, projects, and exam evaluation. In an online classroom, such communication can be more difficult. For those in synchronous classes, such as through teleconferencing, in most cases, the learners are just muted, and the instructor is not even sure if the learners are present. In other forms of the online classroom, such as in asynchronous classes like massive open online courses (MOOCs), learners and instructors can interact in the discussion forums. However, research has shown that in some cases, only about three percent of learners post on MOOC discussion forums (Onah et al., 2014). Another possibility for interaction is when learners provide feedback during course surveys. However, these surveys are frequently done at the start and the end of the course. Hence, there is not much opportunity for the instructor to adjust their teaching based on learner feedback. Additionally, most online classrooms are self-paced, so it is hard for them to see how learners are faring overall. An added complication is that some learners lack self-directed learning skills, which are essential to succeeding in online learning environments (Zhu et al., 2020).

An important thing to consider moving forward is that online learning has its strengths and weaknesses, and it is likely here to stay. The COVID-19 pandemic situation remains uncertain despite vaccination roll-out worldwide. We also have learned that some learners benefit from online platforms (Reich, 2021). For example, some learners have jobs with schedules conflicting with their classes, or some learners have disabilities that are better accommodated in an online learning platform. So, it is only fair to keep online classrooms or hybrid forms as an option for everyone who might need them now that we know that it is plausible for all levels of learning.

2. Related Work

Learner feedback on instructional style and materials is critical for improving teaching quality for future learners (Feistauer and Richter, 2017). Periodic feedback is also needed to prevent learners from losing motivation (Nicol and Macfarlane-Dick, 2006). However, student evaluation surveys, which are the most common source of instructor feedback, can suffer from a lack of validity, bias, and lack of utility (Marsh and Roche, 1997). An alternative source of feedback that recently has been gaining attention is learning analytics dashboards. While learning analytics dashboards can benefit learners in developing self-directed learning skills (Sedrakyan et al., 2020), they might prevent instructors from exercising creativity in instructional planning and decision making (Brown, 2020).

Self-directed learning can be handled separately by introducing metacognitive tutors in the learning environment (Gama, 2005). An example of this is the Personalized Online Adaptive Learning System (POALS) Metacognitive Tutor we developed, shown in Figure 1. This tool was designed to work either as a stand-alone web-based application or as an add-on to learning management systems (LMS) such as Canvas or Moodle. The POALS Metacognitive Tutor divides a learning activity into three phases - Preparation, Problem-Solving, and Evaluation – to allow the learners to reflect metacognitively using open-response prompts. The POALS Metacognitive Tutor has been instrumental in developing the learners' ability to regulate their knowledge, which is vital for self-directed learning (Carlon and Cross, 2021). Details about the utility of the POALS Metacognitive Tutor (e.g., how metacognition is measured, expected inputs to the prompts, and others) in metacognitive instruction is discussed in our other works.

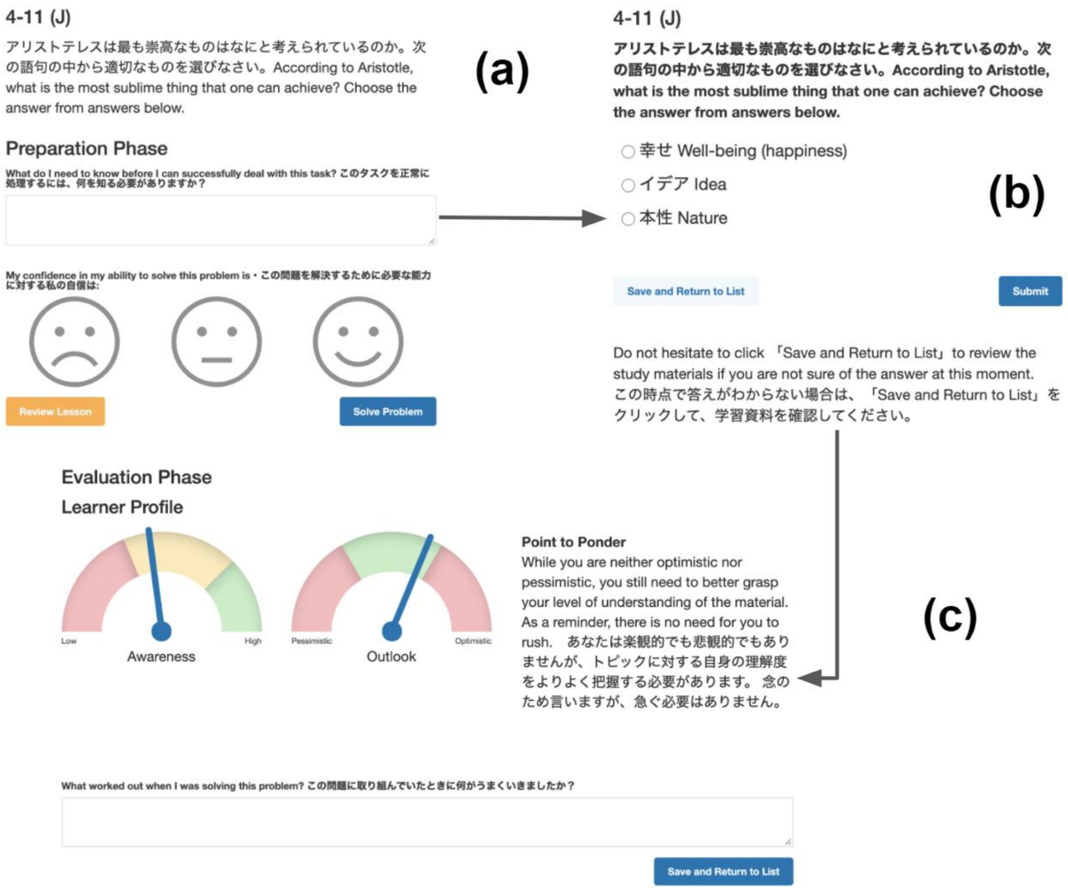

Figure 1. POALS Metacognitive Tutor with the Preparation (a), Problem-Solving (b), and Evaluation (c) screens in Japanese and English.

There are existing works on learning analytics dashboard for online learning environments. Some of these quantitative measures from learner performance but can be limited in their ability to enable qualitative assessment such as classroom climate. Those that allow qualitative evaluation through natural language processing (NLP) approaches typically rely on discussion forums and course surveys whose data may not be representative of everyone due to the problems mentioned earlier. By using the POALS Metacognitive Tutor which is a private channel between the educator and the learners and can possibly be made mandatory, we can have an analytics data source deriving inputs from everyone. Metacognitive instruction may also introduce new dynamics in the online classroom that is yet to be explored for analytics. This paper demonstrates how the POALS Metacognitive Tutor can be used as input for a learning analytics dashboard by constructing a prototype of the NLP-based POALS Analytics Dashboard. A user study was also conducted to evaluate the perceived usefulness and usability of the POALS Analytics Dashboard.

3. The POALS Analytics Dashboard

The POALS Analytics Dashboard is based on the human-in-the-loop (HITL) design or interfaces that require human interaction. While HITL was previously attributed to increased problems in computer systems due to human errors in the past (Cranor, 2008), HITL is increasingly recognized to build fairer AI systems (Zanzotto, 2019). For the POALS Analytics Dashboard, the information is presented to the instructor in a digestible manner, and further actions were left for the instructor to decide (i.e., no recommendations provided). Since decision to action is still left to the instructors, POALS Analytics Dashboard does not aim to be a substitute for classroom interaction but as a supporting tool. The dashboard will be made up of three visualizations, each discussed separately, to uncover feedback that the learners might not provide to the instructors in an online environment.

The POALS Metacognitive Tutor described earlier was deployed in an electrical engineering course offered to first-year undergraduate students at Tokyo Institute of Technology (Tokyo Tech) over two academic quarters: from September 2020 to February 2021. Each quarter was made up of seven weeks: three weeks were delivered synchronously through teleconferencing and four weeks asynchronously as a small private online course (SPOC) on the LMS edX Edge. The resulting visualizations shown below are derived chiefly from metacognitive prompt responses completed by 29 students.

3.1 Sentiment Analysis

The online instructor might be interested to know how happy their learners are in their class. Subtle, elusive, unverbalized emotions are the basis of thought, meaning, and language, affecting perception and, eventually, cognition (Kanazawa, 2020). To address this, we conducted sentiment analysis on learner metacognitive prompt responses. Figure 2 (left) is a composite visualization of learner sentiments. The column graph shows the average absolute sentiment polarity score from the prompt responses. Its color changes to red when the sentiment is negative, orange when neutral, and green when positive. The ratio of learners with negative, neutral, or positive sentiments is shown through the half-donut chart.

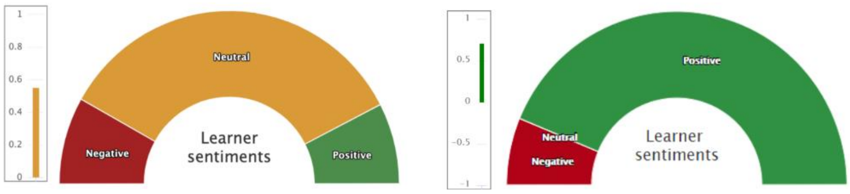

Figure 2. Ideal (left) and Resulting (right) Sentiment Analysis Visualization.

The learner responses in both the preparation and evaluation prompts were used as inputs for the sentiment analysis. All inputted text was presumed to be Japanese. The Bidirectional Encoder Representations from Transformers (BERT) was used as the pre-trained system for tokenizing, specifically bert-base-japanese-whole-word-masking. The bert-base-japanese-sentiment classifier was used as the pre-trained model, which labels the input as either positive or negative. In our case, it is crucial to see the percentages of learners who have polar sentiments; hence we need to establish a neutral range that is not predefined in our sentiment classifier choice. For both negative and positive cases, when the probability was less than 0.25, the label was changed to neutral. The cut-off is heuristically set and must be further investigated. These labels (positive, negative, neutral) corresponded to the green, red, and orange colors in the half-donut chart. The probabilities were multiplied by -1 when the original label is negative, and the values are summarized to get the value for the column graph.

Figure 2 (right) shows the result from the electrical engineering course, and it is noticeable that there are not many neutral responses. When the sentiment classifier was investigated, it was observed that even a very neutral sentence like “I live in Tokyo” gives a very positive score. A possible solution is to build a sentiment classifier specific to the task, but this requires extensive data collection that may not be easily scaled.

3.2 Topic Modeling

Another thing that the online instructor might think is, had the class been in-person, what topics will the learners be discussing? A word cloud like in Figure 3 (left) where the most prominent word in each topic tackled by the learners is shown and size is the probability assigned to the topic can be useful in answering this question.

Figure 3. Ideal (left) and Resulting (Right) Topic Modeling Visualization.

To get the topic information, we conducted topic modeling on the metacognitive prompt responses using Latent Dirichlet Allocation (LDA), where observations (e.g., words in responses) are used to explain latent or unobserved groups (i.e., topics) by looking where the observations are present (Blei et al., 2003). The presence matrix, called tf-idf, was constructed by matching up the word frequency (tf: term frequency) with the responses the said word appears (idf: inverse document frequency).

Figure 3 (right) is the result from the electrical engineering class after translation to English. While knowing the topics the learners are thinking of is interesting, learning about these topics can also help the instructor uncover misconceptions. To elaborate, if a topic from a particular module comes up, it may be intuitive to think that the class average for the said module may be high since enough learners have thought carefully about it. If an emerging topic ends up coming from a low-scoring module, it may be worth investigating if there is a misconception that the learners are repeating. Misconception is a prevalent problem in education, may it be in chemistry (Uce and Ceyhan, 2019) or any other field.

3.3 Similarity Network

Finally, the instructor may be interested to know which topics the learners can remember well. Surface learning is the phenomenon where a learner picks up just enough knowledge to pass tests (Dolmans et al., 2016). The choice between surface learning or deep learning relies not just on the learners' motivation but also on the accompanying learning activities. The instructor must see if surface learning is widespread in their class, indicating concerns on how the learning activities are constructed instead of several individual motivations. The instructor may detect surface learning through a network graph, like the one shown in Figure 4 (left). Nodes represent modules, and related modules are connected; the weight of the lines expresses the degree of relatedness. Hovering a node shows the average learners' score in the exercises for the said module on a scale of 0 to 1. By inspecting the average learner scores on related modules, the instructor may detect surface learning.

The corresponding texts must first be converted into numerical representations to measure similarities between modules. The doc2vec algorithm available in the gensim library was used for vectorization. The resulting node weights then serve as the vector representation of the words. The vectors derived for each module might have different magnitudes (e.g., some modules having more words). The cosine similarity, which measures the angle between two vectors, was selected instead of the more popular Euclidean distance to reduce the effect of magnitudes.

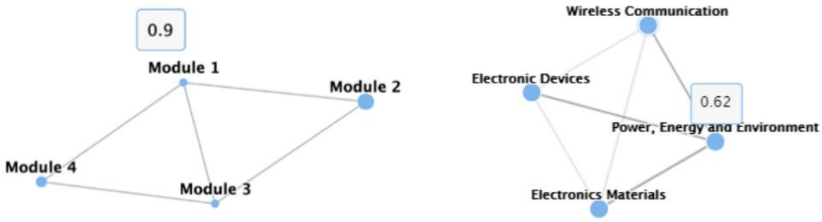

Figure 4. Ideal (left) and Resulting (Right) Similarity Network Visualization with Learners' Score.

Figure 4 (right) shows the result for the electrical engineering course. All similarity values are between 0.42 and 0.51. In this figure, we decided to make the nodes connected if their similarity score is above 0.4 due to proximity of similarity scores, so all nodes ended up being connected. This is a challenge with the proposed similarity network: what is the proper cut-off for similarity? We need to conduct similar experiments to have a better idea.

4. User Study

The POALS Analytics Dashboard was introduced to nine educators working in secondary schools, professional training, after-school support, and higher education (undergraduate and graduate) from Japan, the Philippines, United States, and Finland. They have experience in face-to-face and hybrid formats, and one has experience in a fully online format. They answered a questionnaire made up of four parts: a written interview to inquire about their experiences engaging with learners and their opinions about recent educational trends; an introduction to POALS Metacognitive Tutor and Analytics Dashboard; a Likert scale based on the Technology Acceptance Model (TAM) with a rating from 1 – Strongly Disagree to 5 – Strongly Agree (Davis, 1991); and a free-response form for further feedback.

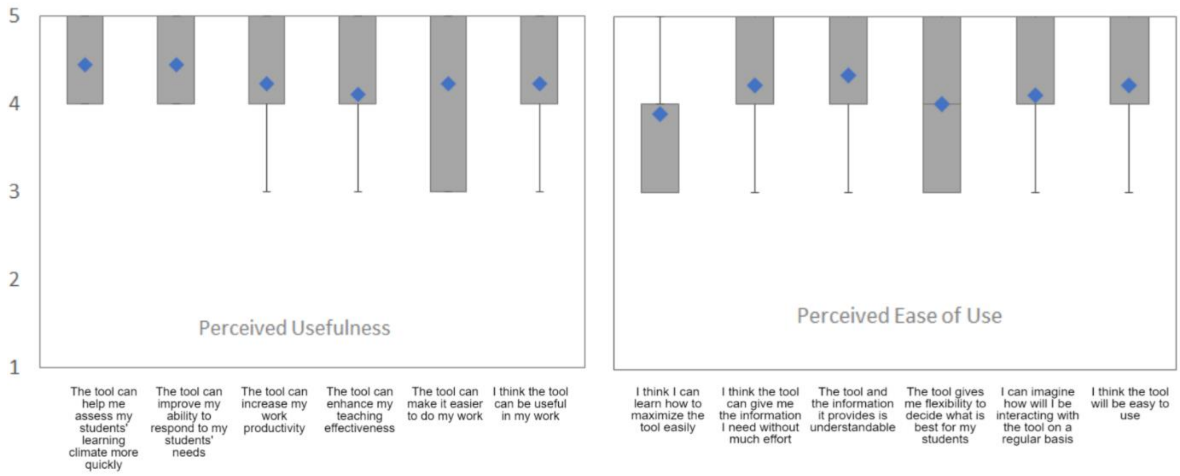

Figure 5. Boxplot of Modified TAM Results with Means Illustrated as Blue Diamonds.

Figure 5 shows the boxplots of the modified TAM responses. All the respondents agreed that the POALS Analytics Dashboard can help them respond to their learners' unvoiced needs and assess learner progress. Most of them had shown in their open question responses a keen interest with sentiment analysis as knowing the learners' feelings is a challenge in online learning environments. The only item that had a mean below 4 – Agree is the perception of maximizing the use of the tool. The respondents should be given more time to use the tool in their classes to understand this better.

5. Summary and Future Work

The POALS Analytics Dashboard is a learning analytics dashboard grounded on instructor intuition and learning theories. It provides details that can be used for detecting poor sentiments, misconceptions, and surface learning while leaving the judgment whether interventions should be made or not to the instructor. User study indicates a positive outlook towards the concept; thus, it is worthwhile to develop further the POALS Analytics Dashboard (e.g., multiple language support, more interactivity, etc.).

Acknowledgments

This research was supported by the Tokyo Tech Electrical Engineering Literacy teaching team consisting of Professors Akira Chiba, Takayuki Iwasaki, and Khanh Tran Gia. The Japan Society for the Promotion of Science (JSPS) supported this work via the Grants-in-Aid for Scientific Research (Kakenhi) Grant Number JP20H01719. Furthermore, the authors are indebted to the Center for Innovative Teaching and Learning for use the open edX LMS on the TokyoTechX edge.edx.org server.

References

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *The Journal of Machine Learning Research*, *3*, 993-1022.

Brown, M. (2020). Seeing students at scale: How faculty in large lecture courses act upon learning analytics dashboard data. *Teaching in Higher Education*, *25*(4), 384-400.

Cranor, L. F. (2008, April). A framework for reasoning about the human in the loop. In *Proceedings of the 1st Conference on Usability, Psychology, and Security* (pp. 1-15).

Carlon, M. K. J. & Cross, J. S. (2021, March). Open response prompts in an online metacognitive tutor. In *Program of the JSET Spring Conference 2021* (pp. 565-566).

Davis, F. D. "Perceived usefulness, perceived ease of use, and user acceptance of information technology." *MIS Quarterly* (1989): 319-340.

Dolmans, D. H., Loyens, S. M., Marcq, H., & Gijbels, D. (2016). Deep and surface learning in problem-based learning: a review of the literature. *Advances in Health Sciences Education*, *21*(5), 1087-1112.

Gama, C. A. (2005). *Integrating metacognition instruction in interactive learning environments* (Doctoral dissertation, University of Sussex).

Feistauer, D., & Richter, T. (2017). How reliable are students' evaluations of teaching quality? A variance components approach. *Assessment & Evaluation in Higher Education*, *42*(8), 1263-1279.

Kanazawa, Y. (2020). Emotion is "deeper" than cognition: theoretical underpinnings and multidisciplinary lignes de faits to the Emotion-Involved Processing Hypothesis (EIPH). *国際学研究= Journal of International Studies*, *9*(1), 185-206.

Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: the critical issues of validity, bias, and utility. *American Psychologist*, *52*(11), 1187.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199-218.

Onah, D. F., Sinclair, J. E., & Boyatt, R. (2014, November). Exploring the use of MOOC discussion forums. In *Proceedings of London International Conference on Education* (pp. 1-4).

Reich, J. (2021). Ed tech's failure during the pandemic, and what comes after. *Phi Delta Kappan*, *102*(6), 20-24.

Sedrakyan, G., Malmberg, J., Verbert, K., Järvelä, S., & Kirschner, P. A. (2020). Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation. *Computers in Human Behavior*, *107*, 105512.

Üce, M., & Ceyhan, İ. (2019). Misconception in chemistry education and practices to eliminate them: literature analysis. *Journal of Education and Training Studies*, *7*(3), 202-208.

Zanzotto, F. M. (2019). Human-in-the-loop Artificial Intelligence. *Journal of Artificial Intelligence Research*, *64*, 243-252.

Zhu, M., Bonk, C. J., & Doo, M. Y. (2020). Self-directed learning in MOOCs: exploring the relationships among motivation, self-monitoring, and self-management. *Educational Technology Research and Development*, 1-21.