Developing AI into explanatory supporting models: An explanation-visualized deep learning prototype

Chen, H.; Tan, E.; Lee, Y.; Praharaj, S.; Specht, M.; Zhao, G.

Publication date
2020

Document Version
Final published version

Published in
14th International Conference of the Learning Sciences

Citation (APA)
Chen, H., Tan, E., Lee, Y., Praharaj, S., Specht, M., & Zhao, G. (2020). Developing AI into explanatory supporting models: An explanation-visualized deep learning prototype. In M. Gresalfi, & I. S. Horn (Eds.), *14th International Conference of the Learning Sciences: The Interdisciplinarity of the Learning Sciences, ICLS 2020 - Conference Proceedings* (pp. 1133-1140). (Computer-Supported Collaborative Learning Conference, CSCL; Vol. 2).

Developing AI into Explanatory Supporting Models: An Explanation-Visualized Deep Learning Prototype

Haoyu Chen
University of Oulu
28 PUBLICATIONS **454** CITATIONS

Esther Tan
Delft University of Technology
24 PUBLICATIONS **309** CITATIONS

Yoon Lee
Delft University of Technology
7 PUBLICATIONS **13** CITATIONS

Sambit Praharaj
Ruhr-Universität Bochum
14 PUBLICATIONS **126** CITATIONS

Developing AI into Explanatory Supporting Models: An Explanation-Visualized Deep Learning Prototype

Haoyu Chen, University of Oulu, chen.haoyu@oulu.fi
Esther Tan, Delft University of Technology, e.b.k.tan@tudelft.nl
Yoon Lee, Delft University of Technology, y.lee@tudelft.nl
Sambit Praharaj, Open University of the Netherlands, sambit.praharaj@ou.nl
Marcus Specht, Delft University of Technology, m.m.specht@tudelft.nl
Guoying Zhao, University of Oulu, guoying.zhao@oulu.fi

Abstract: Using Artificial Intelligence (AI) and machine learning technologies to automatically mine latent patterns from educational data holds great potential to inform teaching and learning practices. However, the current AI technology mostly works as "black box" - only the inputs and the corresponding outputs are available, which largely impedes researchers from gaining access to explainable feedback. This interdisciplinary work presents an explainable AI prototype with visualized explanations as feedback for computer-supported collaborative learning (CSCL). This research study seeks to provide interpretable insights with machine learning technologies for multimodal learning analytics (MMLA) by introducing two different explanatory machine learning-based models (neural network and Bayesian network) in different manners (end-to-end learning and probabilistic analysis) and for the same goal - provide explainable and actionable feedback. The prototype is applied to the real-world collaborative learning scenario with data-driven learning based on sensor-data from multiple modalities which can assess collaborative learning processes and render explanatory real-time feedback.

Introduction

Harnessing the affordances of Artificial Intelligence (AI) and machine learning technologies to conduct Multimodal Learning Analytics (MMLA) is gaining increasing relevance and significance in the field of Computer-Supported Collaborative Learning (CSCL). However, although current AI, especially deep learning technologies can provide accurate predictions in different educational tasks, their "black box" mechanism becomes a huge obstacle to further support the analysis of the learning processes and outcomes (see Figure 1). In a collaborative learning scenario, a typical AI model always seeks for optimal prediction through black box models, but it most likely fails to go deeper to support the learning. An explanatory AI model is recommended in this work that it could provide interpretable and actionable insights. In this study, we present the possibility of developing AI technologies into explanatory supporting models by constructing a deep learning prototype that can offer and visualize interpretable insights for learning analytics to inform teaching and learning practices.

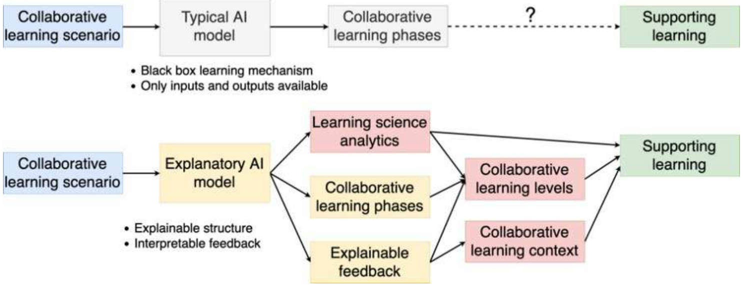
Collaborative learning scenario
Typical AI model
Collaborative learning phases
?
Supporting learning
• Black box learning mechanism
• Only inputs and outputs available
Learning science analytics
Collaborative learning scenario
Explanatory AI model
Collaborative learning phases
Explainable feedback
Collaborative learning levels
Collaborative learning context
Supporting learning
• Explainable structure
• Interpretable feedback

Figure 1. A typical AI model (top) and an explanatory AI model (bottom).

Notwithstanding CSCL is a relatively mature research field, we continue to witness increasing empirical research to explore the potential of a more efficient and deeper understanding of learning analytics in collaborative settings (Toh, Hung, Chua, He, & Jamaludin, 2016). On the self-same note, research on the assessment and evaluation of CSCL processes and outcomes are still lagging behind the theoretical concepts and models. We attribute this to several challenges as follows: i) the complexity of capturing and unpacking latent attributes of students among constant and random cognitive flow and interactions among individuals in the collaborative group makes it difficult to set objective standards (e.g., indicators) for assessing or measuring individual and collaborative learning; ii) although it has become a trend to utilize affordances of MMLA to deliver rich and diverse information for analyzing and understanding the processes and outcomes of CSCL (Di Mitri, Schneider, Specht, & Drachsler, 2018), it is extreme energy-consuming for teachers to teach and conduct the learning analytics simultaneously in real educational scenarios; the automatically mining of the useful latent attributes to obtain concise and effective representations with machine learning instead of human experts is desirable; and iii) deep learning based MMLA for CSCL has drawn wide research interests to optimize its potential for more superior performance (Fiacco, Cotos, & Rosé, 2019) , in particular, to render real-time feedback and automation. In sum, machine learning based MMLA could not, by itself, directly give constructive feedback to students and teachers (Fiacco, Choudhary, & Rosé, 2019). Meanwhile, an automatic analyzing model with the explanatory ability for supporting educational learning will involve interdisciplinary efforts such as computer science, education learning research and design learning.

To address the above-mentioned problems, with the efforts from multiple disciplines, this research provides provide an Explanation-Visualized Deep Learning (EVDL) prototype to facilitate MMLA for a specific collaborative learning form (Knowledge Building) in real-world educational scenarios. In this EVDL prototype (see Figure 2), data streams captured from multiple modalities (video modality: body movements and interactions, audio modality: talking content and voice properties, and text modality: forum discussion content) will be analyzed by explanatory deep learning models with the capability of giving interpretable feedback. Then the real-time feedback with visualized explanations is given to enable students and teachers to foster self-reflection, interactive, and representational learning. By implementing the EVDL, we aim to pave the way for analytics-based CSCL and mine the potential of MMLA to provide holistic insights into students' small group collaborative learning scenarios. Extended applications of this prototype can be like intelligent educational tutoring systems and adaptive human computer interfaces (Krakovsky, 2018).

In the succeeding sections, we first introduce Knowledge Building (KB), a well-established pedagogical model of CSCL used as a specific collaborative learning scenario in our prototype. We also discuss how previous work has attempted to integrate MMLA into CSCL: deriving various indicators from multiple modalities and delivering different feedback types. Next, we provide an overview of the EVDL prototype – how the multimodal data can be enacted to capture dispositional (motoric and physiological data) and discourse (audio & video data) attributes. We demonstrate the use of indicators to inform the quality of discourse from the dispositional and discourse analytics and showcase how the actionable insights for educators/teachers and learners/students can be visualized. Finally, we discuss the practical and scientific significance of this study and the envisioned directions of the larger research undertaking.

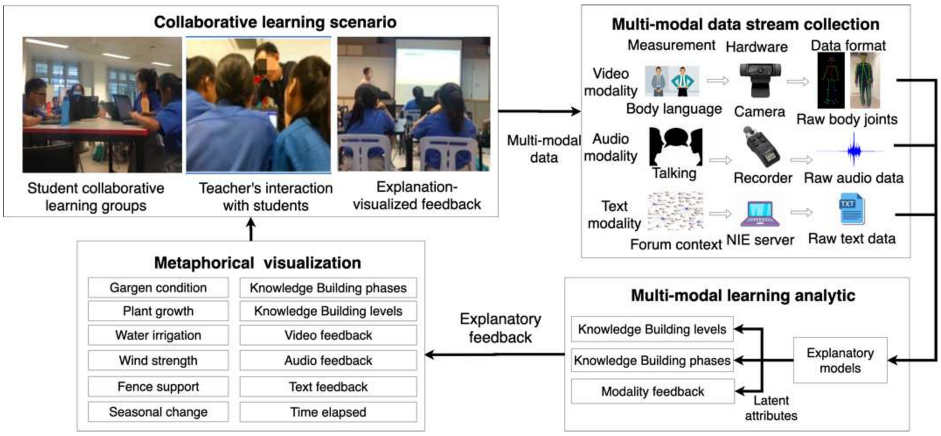

Figure 2. The overview of the EVDL prototype for CSCL with MMLA.

Related work

Knowledge Building: A well-established pedagogical CSCL model

Although CSCL is an extensively researched field in the education field (Toh et al., 2016), there is still a lack of systemic research on the assessment and evaluation of CSCL. One of the biggest challenges for CSCL is the assessment of collaborative learning processes and outcomes: quality of discourse (i.e., depth of discourse, critical and reflexive thinking), individual and group level learning gains. Owing to the constant and random cognitive flow and interactions between individuals in the collaborative group, it is crucial to search for a more accurate and objective depiction for measuring the quality of the collaborative learning process for the CSCL research (Miyake & Kirschner, 2014). Knowledge Building (KB) practice is introduced into CSCL as a novel pedagogical model for learning analytic (Scardamalia & Bereiter, 1994). In the KB learning setting, collaborative learning is premised on four core KB principles which foster collective cognitive responsibility (Scardamalia, 2002): “idea generation”, “idea connection”, “idea improvement”, and “rise above”. Students carry out collaborative tasks driven by real ideas, authentic problems. The KB practice is supported by Knowledge Forum (KF) - an online forum where students can develop, share, improve and advance their ideas in the various phase of KB. Thus, the collaborative learning can be presented by the process status of KB. Based on KB, different kinds of specific indicators (Zhang & Yuan, 2018) have been proposed to discriminate to allocate the KB phase (Tao, Zhang, & Huang, 2015). However, few efforts have been made to identify and visualize the KB discourse KB phases with AI, as current AI models cannot, by themselves, directly give constructive feedback in an explainable mechanism.

MMLA for CSCL

Recently, MMLA has been widely used for CSCL because it affords a more holistic understanding of collaborative learning (Blikstein & Worsley, 2016) from the diverse data streams. Di Mitri et al. (2018) presented the potential various sources of data ranging from students’ behavior cognition, emotion, to their motivations and beliefs. Currently, most of the MMLA approaches focus on manually exploring a richer set of indicators (e.g., body movements, facial expressions and human interaction) (Healion, Russell, Cukurova, & Spikol, 2017; Chen, Liu, Li, Shi, & Zhao, 2018). Hence, efforts to leverage the capability of machine learning especially deep learning to achieve automatic learning for the latent patterns remain desirable.

Visualized feedback for collaborative learning.

In previous works, various methods to visualize feedback have been proposed. For instance, Shapiro (2019) made use of interaction geography to interpret collaborative interaction across the physical environment. In another work (Praharaj, Scheffel, Drachsler, & Specht, 2018), the feedback for turn-taking and speaking time of the collaborative learning is visualized on a public dashboard in a real-time manner. A “groupgarden” is presented in which students are represented by corresponding plants (Tausch, Hausen, Kosan, Raltchev, & Hussmann, 2014), and different feedback is given by changing the status of the components in a metaphorical visualization manner. Also, the concept of explanatory learner models are proposed recently (Rosé, McLaughlin, Liu, & Koedinger, 2019) , which is to enable insight analytics in technology to enhance education learning. But a better utilization of the explanatory insights for real-time feedback is not presented in their work.

As illustrated in Table 1, we compare of our EVDL prototype to the above-discussed prototypes. Concerning developing deep learning into explanatory models for supporting teachers and students in the CSCL field, our proposed EVDL is a more comprehensive work compared with previous prototypes. In our EVDL prototype, the explainable feedback from deep learning based MMLA is visualized to students and teachers, which has not been realized in the compared work. In the next section, the overview if the EVDL will be introduced.

Table 1: A comparison of (multimodal) learning analytics prototypes

	MMLA	Auto	Coding	Explanatory	Visualization	Real-time
Praharaj et al., 2018	√		Hand-crafted	√	√	√
Shum et al., 2018	√		Hand-crafted	√		
Chen et al., 2019	√	√	Deep learning			√
Fiacco et al., 2019		√	Deep learning			
Rosé et al., 2019		√	Deep learning	√		
Authors	√	√	Deep learning	√	√	√

Note: “MMLA” stands for multimodality data, “auto” stands for the automation of the prototype, “coding” stands for the indicator coding methods, “explanatory”, “visualization” and “real-time” stands for the corresponding properties of the feedback.

Overview of the EVDL prototype

We apply the EVDL prototype to the problem of CSCL using multiple modalities in the real-world educational scenario. Specifically, we look at the problem of assessing a collaborative learning process and giving real-time explanatory feedback for small group collaboration in a KB classroom. We define the four KB stages as “idea generation", “idea connection”, “idea improvement” and “rise above”. The scenario we focus on has a student group undertaking a KB task and our prototype tries to infer and give feedback on the quality of collaborative discourse in the KB phases from the students’ gestures, movements, verbal and non-verbal discourse (KB forum).

The overall architecture of the EVDL (see Figure 2) consists of four components: 1) collaborative learning scenario, 2) multimodal data stream collection, 3) multimodal learning analytic and 4) metaphorical visualization. In the multimodal data stream collection component, the students' raw data from the cameras, audio recorders and the text in the forum are used to extract relevant features and transferred to the learning models. In the next MMLA component, we present two different kinds of models that can both generate explanatory feedback. The first one is with an end-to-end architecture that implements attention-based neural-networks. The explanation of the model is achieved by visualizing the weights in the neural networks. The other one is to design a probabilistic model with classifiers (experts) and evaluators (critics) that correspond to each of these modalities. It can naturally explain the concepts of experts and critics with the output of the neural networks. In the last metaphorical visualization component, the explainable feedback from the MMLA will be visualized in a metaphorical manner to both students and teachers.

Multiple modalities

There are three modalities used in the EVDL: video, audio and text. The data stream formats collected for real-time analysis are raw body joints, raw audio data and raw text data. Detailed settings of each modality can be seen in Table 2. The used hardware, APIs and libraries are listed as well.

Table 2: A detailed list of settings for each modality

Modality	Measurement	Hardware	API and library	Data format	Indicators	
					End-to-end model	Probabilistic model
Video	Body movement	Camera C920 HD	Pytorch Openpose	Raw body joints	Raw body joints	Spontaneous gestures, interactions with others, interaction with PC, sitting/standing postures
Audio	Talking	Zoom H6 recorder, lapel microphones	Google Cloud Speech-to-Text	Raw audio data	MMFC features of audio data	Duration, loudness, speech turns, pitches
Text	Forum context	Knowledge Forum, remote server	KBDEx Learning Analytics	Raw text data	Word embedding of the text data	Contribution pattern, cognitive scaffolds, idea building, pattern degree centrality

Data collection

The data collection is conducted among a group of 5-6 students in a room. The students are engaged in a KB task while collectively working in the online knowledge forum on their computers: 1-1 computing. Each student is able to search for the information on the Internet and make notes on Knowledge Forum. The discussion will be measured in real-time with the four KB phases. During the recording, three modalities are collected for real-time analysis: video, audio and text. Note that no restriction is given about the performance, and students are free to behave as they do every day in case they may inhibit or deliberately adjust their body gestures and behaviors.

Modality settings

For the video modality, we use cameras with a recording speed of 30 frames per second to capture the body movements of the students with the image resolution as 640*480. Then we adopt Openpose (1) (Cao, Simon, Wei, & Sheikh, 2017) which is an open-source library that can provide a pose estimator frame-by-frame to acquire human body landmarks. Each student in the group is tracked and designated with a unique ID number. Specifically, tracking is conducted on the coordinate information of upper-body with the Kalman Filter. The raw body skeleton coordinate data consist of tracks of 25 body skeleton joints (see Figure 2).

For the audio modality, each student is equipped with one lapel microphone attached to one audio recorder to capture the verbal data streams of the collaborative content. We adopt the Google API: Cloud Speech-to-Text (2) to convert the speech recordings into audio properties and text.

For the text modality, the content in the KB forum will be collected as the text modality data stream. The knowledge forum is used for students to take notes during collaborative learning. The text will be pre-processed by the NLTK (3) opensource library into raw text files as the text data stream. The pre-processing includes: removing punctuation, tokenization, stop words removal and stemming.

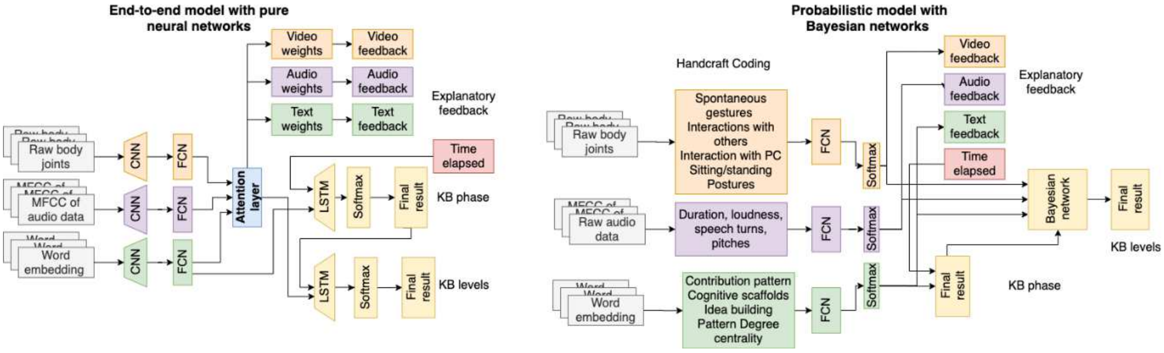

Figure 3. Two proposed explanatory AI models: the end-to-end model (left) and the probabilistic model (right).

Collaborative learning assessment with MMLA

In this section, we first discuss how to use the KB phases as the ground truth for measuring the collaborative learning level for EVDL. Next, we introduce two different AI models that can be used: an end-to-end neural network model and a probabilistic model to achieve the explainable assessment of the collaborative learning.

Table 3: A human coding scheme for evaluating the KB levels in each KB phase

	Level 1	Level 2	Level 3	Level 4
KB level	Fragmented discourse	Knowledge sharing	Knowledge construction	Knowledge building
Indicators	No knowledge advancement with isolated ideas	Low knowledge advancement with shared information	Medium knowledge advancement with the idea interaction	High knowledge advancement with idea development

Assessing collaborative learning

As afore-discussed, we embed Knowledge Forum (KF) into KB learning analytics as collaborative tasks. Since KB learning is a well-established model, there are various and reliable indicators for experts to assess and recognize the KB phase from the KF content. The KF notes is a first indication of the collaboration stages within the KB framework. Then, the four KB phases (from idea generation to rise above) are evaluated with the four KB levels (see Table 3). The coding indicators identify KB patterns at four levels: Fragmented discourse (level 1), knowledge sharing (level 2), knowledge construction (level 3) and knowledge building (level 4) which are adapted from Lin & Chan (2018). The evaluation of the KB levels in the KB phases is to mine the depth of the KB patterns for assessing and supporting the CSCL by harnessing the affordances of MMLA.

Human experts are enrolled to give measurements to the collaborative discourse (in the video clip) as ground truth. The experts will give a judgement of the level of collaborative KB at each phase according to the coding indicators (see Table 3). Meanwhile, data from self-report and focused group interviews of the students should also be conducted after the whole learning procedure. Students could provide an estimation of the level of their collaborations. This is used to analyze the subjective influence during collaborative learning. Then, two explanatory supporting models are introduced for real-time recognition of KB phases and levels with MMLA.

An end-to-end neural-network model with an explainable attention mechanism

In this model, we train an end-to-end neural network architecture to recognize the KB phases and levels (Figure 3 left). We introduced this model to explore the possibility of developing a typical "black-box" like based deep learning model for CSCL while it's endowed the explanatory ability. Firstly, we use Gaussian filters to process the raw body joints data, extract the MFCC (Mel frequency cepstral coefficient) features (Ainsworth, Popper, & Fay, 2004) of the audio data and embed the text data into word vectors. Next, convolutional neural networks (CNN) are stacked to the processed features of each modality to obtain diverse high-level features. Each CNN is connected with a fully-connected network followed by the 'sigmoid' activation function. To obtain explainable feedback from this model, we introduce an attention layer stacked to the extracted high-level features to give the weight of each modality. It not only enhances the training of the whole network architecture, but also provides posterior weights for each modality. The attention layer is formulated as:

$$Attention(Q,K,V) = softmax(\frac{Q^T K}{\sqrt{d^k}})V, \quad (1)$$

where Q, K, V are the corresponding matrixes of the query, key and value for the attention layer, d^k is the dimension of the key. Scaled dot-product attention is implemented in this model. Then we can exact the weight of the attention layer and get the posterior of each modality. The weights of each modality will be normalized and regarded as an explanatory output which serves as a feedback that can be visualized. At last, long short-term memory (LSTM) neural networks (a kind of neural networks that can process temporal sequences) (Greff, Srivastava, Koutník, Steunebrink, & Schmidhuber, 2015) are used to predict the KB phases and KB levels separately with the temporal input steps as 15 frames.

A probabilistic Bayesian network model with experts-based explanation

Besides the end-to-end model, we also design a probabilistic model that can naturally explain the concepts of experts and critics with a Bayesian inference network (Kapoor, Picard, & Ivanov, 2004). The neural networks serve as classifiers (experts) in the model and the Bayesian network can offer evaluation of the classifiers (critics) for predicting the KB phase.

In this model (see Figure 3 right), we firstly use the indicators proposed by experts to offer a KB phase prediction. Then, with the KB phase introduced as the conditional independence, a Bayesian network is used to combine the recognizing probabilities from the three modalities to give the recognition of the KB levels for that phase. The framework is formulated below. Consider a set of multimodal data X_M with the modality number $M = 3$, it is the objective evidence for inferring the ultimate collaborative learning level observed from a collaborative student group. The probability $p_{w_i \in C}(w_i|X_M)$ of different KB phases w_i is the ultimate target, for $i = 1, \ldots, L$. L is the number of KB levels $L = 4$. C is the set of KB levels from "fragmented discourse", "knowledge sharing", "knowledge construction" to "knowledge building". Then we have:

$$p_{w_i \in C}(w_i|X_M) = \sum_{\lambda=1}^{S} p(\lambda|X_M) \sum_{\widetilde{w_i}} p(\widetilde{w_i}|X_M, \lambda)\, p(w_i|\widetilde{w_i}, \lambda, X_M), \tag{2}$$

where λ is the KB phases inferred in the previous stages including "idea generation", "idea connection", "idea improvement" and "rise above, the term $p(\widetilde{w_i}|X_M, \lambda)$ corresponds to the classifiers (experts) for each modality. $p(\lambda|X_M)$ evaluates how confident each expert is on the input, the term $p(w_i|\widetilde{w_i}, \lambda, X_M)$ is the empirical distribution obtained by training the neural networks on the given data. As shown in Equation (2), we adopt a probability way to infer different KB levels with the given modalities. Based on the Bayesian rule we can further build a discriminative model for predicting the KB level as Equation (2). Also, it can naturally explain the contribution of each modality from the expert and critics aspects as the explanation feedback of this model.

Explanation-visualized metaphorical feedback in real-time

Metaphorical visualizations present the objects and scenes similar to real-life to transmit the information. In a collaborative learning based educational scenario, the user study (Streng, Stegmann, Hußmann, & Fischer, 2009) proves that metaphorical visualized feedback can better help self-regulation for students than that of diagrams.

Inspired by the successful use of metaphorical visualization in the work of Tausch et al. (2014), a metaphorical strategy is adopted to visualize the explanatory feedback. In the EVDL we design the visualization of the feedback from each modality with the corresponding metaphorical figures (see Figure 4). The visualization of the collaboration is achieved with a concept of plant growth which follows the four phases of the KB cycle.

Specifically, this drives real-time visualization dashboards for students with private displays and in the interface of the screen, students will see their own farmlands. Each student gets a different plant and the four stages of the plant growth reflect the four KB phases: "idea generation", "idea connection", "idea improvement" and "rise above". Feedback of the three modalities and time factor will be visualized in a metaphorical visualization manner. As shown in Figure 4, audio data will be represented as the wind (for audio and wind sharing the "sound" metaphor). The movement (caused by wind) of the plants is in proportion to the weight of audio modality given from the explanatory models. The same, the video data will be interpreted as water (for sharing the "dynamic" metaphor) and the text data from the knowledge forum will be represented by fences around the plants (for sharing the "developing" metaphor). Apart from the three feedback on students' learning progress, the collaborative time elapsed will be visualized on the screen as season changes from spring to winter (for collaborative time and season changes sharing the "time elapsing" metaphor).

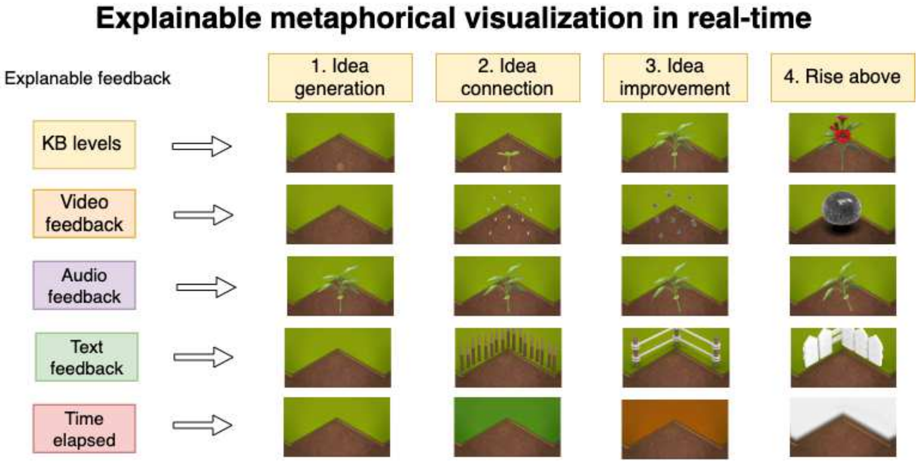

Figure 4. The metaphorical visualization of real-time explainable feedback with different figures.

Discussion

With the proposed EVDL prototype in this paper, we want to pave the way for analytics-based research to mine the potential of implementing AI and deep learning into CSCL with MMLA. Specifically, several directions that can be explored with the proposed prototype:

- The possibility of making deep learning-based learning analytics explainable and mining the potential in helping people to better understand the process and outcomes of collaborative learning;
- The learning capability of end-to-end neural network models for the task of recognizing relatively high-level semantic information, such as KB levels for learning analytics instead of pedagogical coding;
- The capability of the probabilistic deep learning model to make use of the human-coded indicators to evaluate the most valuable latent attributes; and
- The efficiency of metaphorically visualizing feedback with explanations from deep learning based MMLA for both students and teachers in the collaborative learning process.

Conclusions and future works

This study provides some first insights into the methodology to combine MMLA with deep learning methods which can be used for explainable inference of collaborative learning processes in real-world educational scenarios. As a prototype, it bridges the interdisciplinary efforts from AI, machine learning, statistics, linguistics, educational science, learning design, user experience design and art, etc. The contributions of EVDL include: 1) developing deep learning based prototypes for CSCL with MMLA by bridging the interdisciplinary efforts; 2) providing explanatory deep learning models with interpretability for educational data mining and learning analytics; 3) and using metaphorical visualization presentation to give actionable feedback for collaborative learning.

Moreover, we argue for a greater emphasis on designing for assessment and evaluation of collaborative learning environments in addition to designing for effective collaborative learning. Based on the EVDL, further research will explore how explanatory supporting models can contribute to the educational learning process by collecting large-scale datasets from real educational scenarios, and design unsupervised machine learning methods for mining more valuable latent attributes or patterns to unpack the complexity of individual and group knowledge embedded in collaborative learning processes from the large-scale dataset.

Endnotes

(1) https://github.com/CMU-Perceptual-Computing-Lab/openpose
(2) https://cloud.google.com/speech-to-text/docs/
(3) https://github.com/nltk/nltk

References

Ainsworth, W. A., Popper, A. N., & Fay, R. R. (2004). *Speech Processing in the Auditory System.* Springer New York.

Blikstein, P., & Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 220-238.

Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *Conference on Computer Vision and Pattern Recognition.*

Chen, H., Liu, X., Li, X., Shi, H., & Zhao, G. (2019). Analyze Spontaneous Gestures for Emotional Stress State Recognition: A Micro-gesture Dataset and Analysis with Deep Learning. *14th IEEE International Conference on Automatic Face & Gesture Recognition.*

Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018). From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning*, 34(4), 338-349.

Fiacco, J., Choudhary, S., & Rosé, C. (2019). Deep neural model inspection and comparison via functional neuron pathways. *the Annual Meeting of the Association for Computational Linguistics.*

Fiacco, J., Cotos, E., & Rosé, C. (2019). Towards Enabling Feedback on Rhetorical Structure with Neural Sequence Models. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge.*

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2015). LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*.

Healion, D., Russell, S., Cukurova, M., & Spikol, D. (2017). Tracing Physical Movement During Practice-Based Learning through Multimodal Learning Analytics. *In Proceedings of 7th International Learning Analytics and Knowledge Conference*, 588-589.

Kapoor, A., Picard, R. W., & Ivanov, Y. (2004). Probabilistic combination of multiple modalities to detect interest. *International Conference on Pattern Recognition.*

Krakovsky, M. (2018). Artificial (emotional) intelligence. *Commun. ACM*, 61(4):18–19.

Lin, F., & Chan, C. K. (2018). Promoting elementary students' epistemology of science through computer-supported knowledge-building discourse and epistemic reflection. *International Journal of Science Education*, 40(6), 668-687.

Miyake, N., & Kirschner, P. A. (2014). The social and interactive dimensions of collaborative learning. In *The Cambridge handbook of the learning sciences (2nd ed.)* (pp. 418-438). New York, NY: Cambridge University Press.

Praharaj, S., Scheffel, M., Drachsler, H., & Specht, M. (2018). Multimodal Analytics for Real-Time Feedback in Co-located Collaboration. *European Conference on Technology Enhanced Learning.*

Rosé, C. P., McLaughlin, E. A., Liu, R., & Koedinger, K. R. (2019). Explanatory learner models: Why machine learning (alone) is not the answer. *British Journal of Educational Technology*.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. *Liberal education in a knowledge society*, 97, 67-98.

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The journal of the learning sciences*, 265-283.

Shapiro, B. (2019). Integrative Visualization: Exploring Data Collected in Collaborative Learning Contexts. *International Conference for Computer Supported Collaborative Learning.*

Shum, S. B., Echeverria, V., & Martinez-Maldonado, R. (2018). The Multimodal Matrix as a Quantitative Ethnography Methodology. In *Advances in Quantitative Ethnography.*

Streng, S., Stegmann, K., Hußmann, H., & Fischer, F. (2009). Metaphor or Diagram? Comparing Different Representations for Group Mirrors. *In Proc. OZCHI, ACM*, (pp. 249–256).

Tao, D., Zhang, J., & Huang, Y. (2015). How did a Grade 5 community formulate progressive, collective goals to sustain Knowledge Building over a whole school year? *Proceedings of the 11th International Conference on Computer Supported Collaborative Learning.*

Tausch, S., Hausen, D., Kosan, I., Raltchev, A., & Hussmann, H. (2014). Groupgarden: Supporting Brainstorming through a Metaphorical Group Mirror on Table or Wall. *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational.*

Toh, Y., Hung, W. L. D., Chua, P. M. H., He, S., & Jamaludin, A. (2016). Pedagogical reforms within a centralised-decentralised system. *International Journal of Educational Management*, 30(7), 1247-1267.

Zhang, J., & Yuan, G. (2018). Cross-Classroom Interaction for Knowledge Building: A Design Experiment in Four Grade 5 Science Classrooms. *International Conference of Learning Science.*

Acknowledgements

We thank University of Oulu, Finland and National Institute of Education (NIE), Nanyang Technological University (NTU), Singapore for their advice on this manuscript. The first author is funded by the Academy of Finland for project MiGA (grant 316765), ICT 2023 project (grant 328115), Infotech Oulu, CSC - IT Center for Science, Finland and Chinese Scholarship Council.