Evaluation of the Function that Detects the Difference of Learner's Model from the Correct Model in a Model-Building Learning Environment

Tomoya Horiguchi(✉) and Tetsuhiro Masuda

Graduate School of Maritime Sciences, Kobe University,
5-1-1, Fukaeminami, Higashinada, Kobe, Hyogo, Japan
horiguti@maritime.kobe-u.ac.jp, tetsu9988@gmail.com

Abstract. In science education, the model-building learning environment is one of the promising methods for promoting learners' ability to make appropriate models of various phenomena with scientific concepts. However, it isn't an easy task for most learners and some assistance is necessary. We have developed the function that detects difference between learners' models and the correct models and gives feedback about errors to learners. We conducted an experiment for evaluating the function which revealed that: (1) the degree of model completion increased by using the function, (2) the degree of final model completion was correlated with the frequency of using the function, and the correlation was clearer for easier tasks, and (3) the learners who preferred mathematical explanation of errors thought more deeply about why their models were erroneous.

Keywords: Science education · Model-building learning environment · Feedback to learners' errors · Difference between learners' models and the current models

1 Introduction

In science education, it is one of the most important purposes for learners to acquire the ability to make appropriate models of various phenomena based on scientific concepts. Learning with model-building learning environments (MBEs) [1, 2] is a promising method for promoting the ability. In MBE, learners are usually given a set of components of models and build their model by combining them. They can also simulate their model to see whether it behaves as they expected. If it doesn't, they modify the model and try simulation again. Through such process, learners become to be able to make models with which they can explain and predict various phenomena based on scientific concepts.

However, it isn't an easy task for most learners to make correct models in MBEs. In order for composed models to be calculable, the components usually correspond to some scientific/mathematical concepts and they must be combined in scientifically/mathematically right way. Learners who aren't familiar to such concepts and formulation often have difficulties. Therefore, some assistance is necessary.

S. Yamamoto (Ed.): HIMI 2017, Part II, LNCS 10274, pp. 40–49, 2017.
DOI: 10.1007/978-3-319-58524-6_4

In previous MBEs, various kinds of functions for assistance have been implemented such as *online help* (that provides general explanation of the usage of model components) and *syntax checker* (that detects and indicates formally illegal usage of model components) [3]. However, experiments revealed few learners initiatively used these functions when making models, and the effect of the functions significantly varied depending on the difficulty of tasks and learners' prior knowledge [4].

In this research, we evaluate the effect of the function that detects and explains the difference between learners' models and the correct models. The function was implemented as a module of a MBE we have developed [5]. We report about the experiment and its result that was conducted for investigating whether the function assists learners, what type of learners/tasks it helps for, and what type of explanation promotes learners' understanding.

2 Model-Building Learning Environment

2.1 Previous Model-Building Learning Environments

In order to facilitating learners' modeling activity, several model-building learning environments (MBE) have been developed [1–3, 6, 7]. In early MBEs, mathematical expressions were used as the modeling language and numeric results were derived [8]. However, the abstract concepts represented by mathematical expressions are relatively inaccessible to learners in elementary education. It is also difficult for them to interpret the results of numeric calculation. Additionally, mathematical expressions can't capture many crucial aspects of models, such as the conditions under which a model is applicable.

In contrast, in recent MBEs, ontological primitives of qualitative reasoning are used as the modeling language, which makes it possible to capture conceptual aspects of models' behavior, such as causality [2, 3]. These environments allow young learners to articulate knowledge using intuitive concepts.

The usefulness of MBEs has been verified through experiments in elementary and science/engineering education.

2.2 Evans: A MBE of Qualitative Modeling and Simulation

We have developed a MBE called *Evans* [5]. In Evans, learners can make qualitative models of dynamical systems by using components designed according to the vocabulary of qualitative reasoning. Learners can also see the qualitative behavior of their model by qualitative simulation.

Figure 1 shows an example of a model made with Evans that represents the relation among a species' population, birth rate and death rate. It consists of the constraints such as *C1: the change rate is the difference between the birth rate and death rate*, *C2: the population is the integration (I+) of the change rate*, *C3/C4: the birth/death rate is proportional to the population* etc. Figure 2 shows the simulated behavior of this model. *QS0(0)* represents the initial state in which the population, birth rate and death rate are decreasing while the change rate is negative but increasing. This state instantly transfers to the state *QS1(1)* in which all variables are approaching to zero. Since their numerical values change every moment but their qualitative values don't in *QS1(1)*, their behaviors are aggregated into a qualitative state. Thus, qualitative modeling and simulation enables learners to focus on the important features of the system and its behavior.

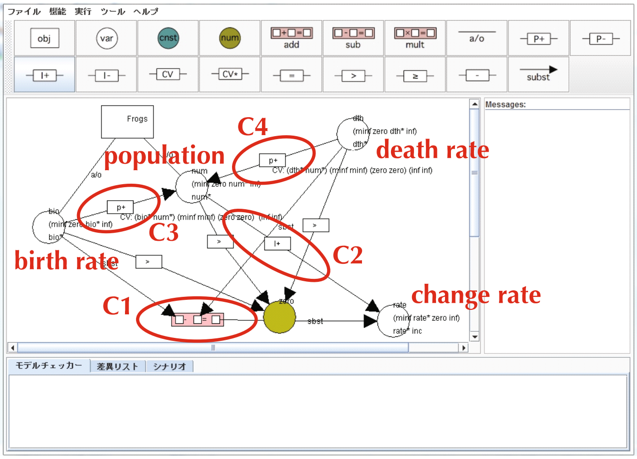


Fig. 1. Model of the change of a species' population

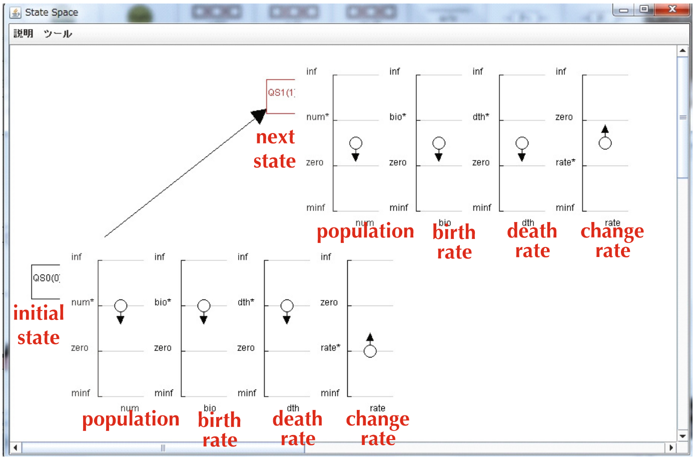
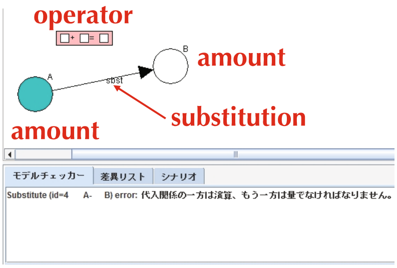


Fig. 2. Behavior of the model of the change of a species' population

2.3 Difficulty and Assistance in Modeling

Though qualitative terms are more intuitive and easier to use for learners than quantitative (i.e., mathematical) ones, it isn't still easy task for them to build qualitative models because components must be combined according to the formalism of qualitative reasoning language. In spite of visual and intuitive GUI, learners often make erroneous models. When a model made by learners includes syntactic errors, it can't be simulated.

Therefore, we previously implemented a *model checker* in Evans that detects formally illegal usage and combination of model components. Figure 3 shows an example of the assistance by the model checker. In Evans, *subst* link (substitution) should be used between an *amount* node (variable or constant) and *operation* node (addition, subtraction or multiplication), but used between two amounts in this example. Such syntactic errors are detected and indicated by the model checker.



```
Substitute (id=4    A-    B) error: 代入関係の一方は演算、もう一方は量でなければなりません。
```

Fig. 3. Model checker's message about a syntax error

Preliminary experiment revealed the model checker promoted learners' carefulness and correction of syntactic errors, and was effective in decreasing the number of such errors. However, the degree of completion of their model (correspondence to the correct model) wasn't necessarily high. Though the model checker was helpful in making models calculable, learners couldn't interpret the unexpected behavior of models and modify them appropriately. We, therefore, designed the function that detects and explains the difference between learners' models and the correct models.

3 Difference List

We have implemented a module of Evans that detects difference of learners' models and correct ones and gives feedback about errors to learners. This module is called *difference list*. It compares learners' model to the correct model made by a teacher and detects the following differences: (1) components that are in the correct model but aren't in learners' model (*lacking part*), (2) components that aren't in the correct model but are in learners' model (*extra part*) and (3) components that are in both correct model and learners' one but erroneously described in the latter (*erroneous part*). Difference list shows and explains each detected part as learners' error. There are two types of explanation: *structural explanation* and *mathematical explanation*. The former explains merely that some components are lacking/extra/erroneous. The latter explains the mathematical role of the components. For example, suppose the model shown in Fig. 4 is correct while the model shown in Fig. 5 is erroneous. In these models, the parts marked with red circles are the difference between them. In this case, structural explanation is as follows: *In your model, amounts num and dth are connected with integral link, but they aren't in* *the correct model. A necessary integral link isn't in your model. The direction of the greater-than link between amounts dth and zero is reversed.* (Fig. 6) Mathematical explanation is as follows: *In your model, when amount dth is greater/less than zero, amount num increases/decreases, but num doesn't in the correct model. A necessary integral link isn't in your model. Amount dth is less than zero.* (Fig. 7) Thus, both explanations clearly indicate how the erroneous model should be modified. On the other hand, the former explanation doesn't explain why erroneous parts are inappropriate, while the latter explanation suggests the reason by explaining the mathematical role of the components. (It isn't always sufficient. There are cases in which physical unnaturalness of the behavior of erroneous models should be explained. We previously proposed the framework for generating such explanation [5].) In Evans, learners can invoke difference list anytime they want, and check the part of the model correspondent to each error by pointing each explanation with mouse.

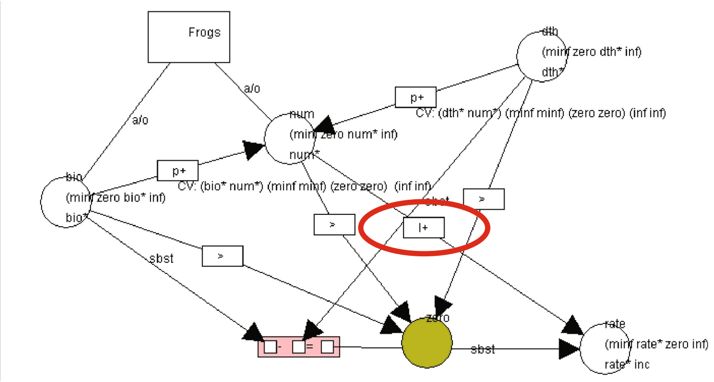


Fig. 4. An example of correct model (Color figure online)

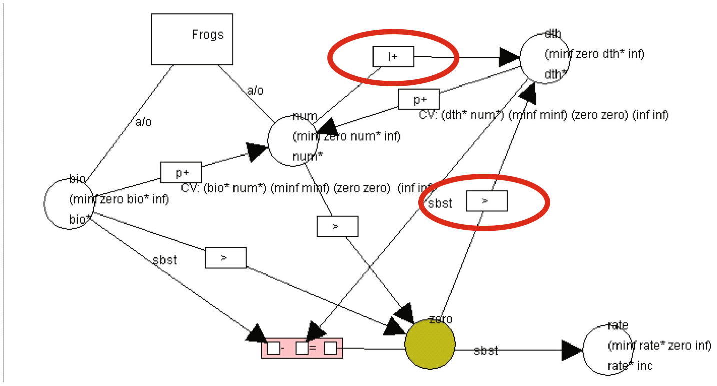
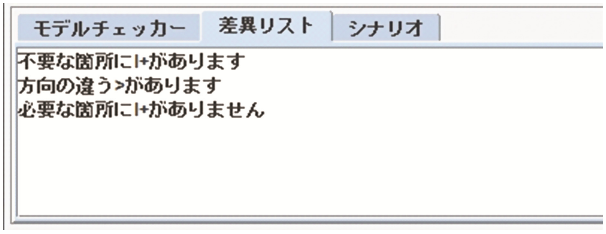
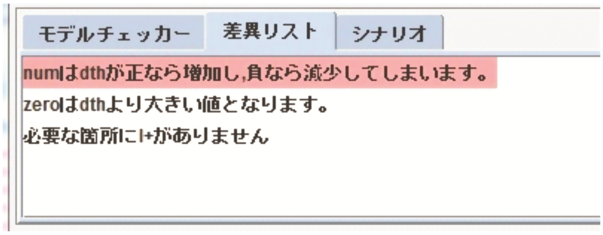


Fig. 5. An example of erroneous model (Color figure online)



不要な箇所にI+があります
方向の違う>があります
必要な箇所にI+がありません

Fig. 6. An example of structural explanation



numはdthが正なら増加し,負なら減少してしまいます。
zeroはdthより大きい値となります。
必要な箇所にI+がありません

Fig. 7. An example of mathematical explanation

4 Experiment

We conducted an experiment for evaluating the effectiveness of difference list in making models in Evans.

4.1 Design

Purpose. The purpose of this experiment is to clarify (1) whether the degree of model completion is increased by using difference list during making models, (2) whether there is a correlation between the frequency of using difference list and the degree of final model completion, and (3) whether the type of explanation by difference list influences learners’ model making.

Subjects. Six students whose major was engineering participated.
Instruments. The following materials were prepared and used.

- *Evans:* The model-building learning environment described in the previous sections. Difference list module was embedded.
- *A booklet for tutorial:* Basic usage of Evans to make models was explained with some examples.
- *Modeling tasks:* Task-1 dealt with a model of the water level of a bathtub with an inflow and an outflow. Task-2 dealt with a model of the population of frogs in a forest that increased by birth rate and decreased by death rate. Task-3 dealt with a model of the water flow between two tanks of which bottoms were connected with a pipe. Task-4 dealt with a model of the heat flow between two objects which touched each

other. Task-1 and Task-3 were isomorphic (i.e., though their domain was different, they had the same structure), and so was Task-2 and Task-4. In addition, Task-1 and Task-2 were relatively easy tasks (because some parts of their models were introduced in the tutorial), while Task-3 and Task-4 were advanced tasks.

- *Questionnaire:* The following questions were asked. (1) *Do you think the difference list was useful for making a model?*, (2) *How did you modify your model based on the feedback?* and (3) *Which explanation did you prefer, structural or mathematical?*

Procedure. In the first week, after a briefing of the outline of the experiment (5 min), the experimenter demonstrated the modeling with Evans by using the booklet for tutorial (10 min). Then, subjects worked on Task-1 with Evans, in which structural explanation of difference list was used (Session-1, 25 min). After that, subjects worked on Task-2 with Evans, in which mathematical explanation of difference list was used (Session-2, 25 min). In the second week, after a briefing of the outline of the experiment (5 min), subjects worked on Task-3 with Evans, in which structural explanation of difference list was used (Session-3, 30 min). After that, subjects worked on Task-4 with Evans, in which mathematical explanation of difference list was used (Session-2, 30 min). At last, subjects answered the questionnaire. In each task, necessary components were prepared by the system, so subjects combined them and inputted necessary values of parameters. During each session, subjects were instructed to save their model every five minutes. All operations by subjects during each session (i.e., the usage of difference list) were automatically recorded as operation logs.

4.2 Results

As for each session, every saved model was scored based on the criteria for the degree of completion. That is, the score of a model by subjects was the number of correctly connected links (in this experiment, the correct model was uniquely defined). Figure 8 shows an example of the transition of scores of a model by a subject in Session-3 (polygonal line graph). In each interval (i.e., 5 min between two saves), the frequency of using difference list counted with operation log is also indicated (bar graph).

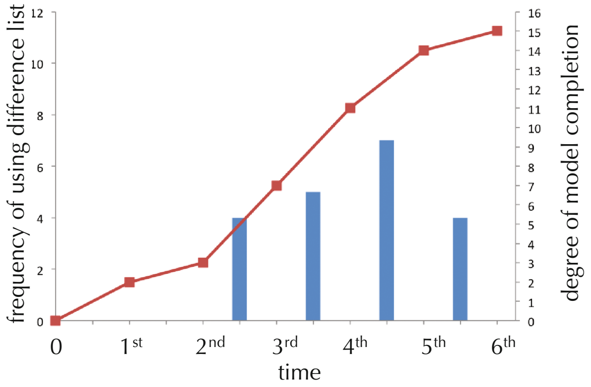


Fig. 8. An example of the transition of model completion and frequency of using diff list

Table 1 shows (1) the total number of intervals in which difference list was/wasn't used (called *dl-interval* and *non-dl-interval*, respectively), (2) the total increase of scores in dl-intervals and non-dl-intervals, (3) the average total increase of scores in dl-intervals and non-dl-intervals, that is, the total increase of scores divided by the total number of dl-intervals and non-dl-intervals, respectively (here, 'total' means the sum of data of all subjects). As for the total of all sessions, the average total increase of scores in dl-intervals is significantly greater than that in non-dl-intervals. This fact suggests using difference list was effective in the increase of model completion. In addition, in Session-1 and Session-2, the average total increase of scores in dl-intervals is equal or less than that in non-dl-intervals, while in Session-3 and Session-4, the average total increase of scores in dl-intervals is greater than that in non-dl-intervals. This fact suggests difference list was more effective in more difficult tasks. That is, since task-1 and task-2 were similar to the examples in the tutorial, subjects could work on them without difference list to a certain degree. On the other hand, task-3 and task-4 were unfamiliar to the subjects, and more complicated with more components than task-1 and task-2.

Table 1. Increase of scores in dl-intervals and non-dl-intervals

	Number of intervals		Total increase of scores		Average total increase of scores	
Diff list	Used	Not used	Used	Not used	Used	Not used
Session-1	10	20	10	27	1.00	1.35
Session-2	19	11	31	17	1.63	1.55
Session-3	20	16	43	28	2.15	1.75
Session-4	24	12	67	14	2.79	1.17
Total	73	59	151	86	2.07	1.46

Figure 9 shows the relation between the frequency of using difference list and the final model completion in Session-1. ANOVA, excluding the outliers (i.e., the data of subjects who rarely used difference list), revealed that these factors were significantly

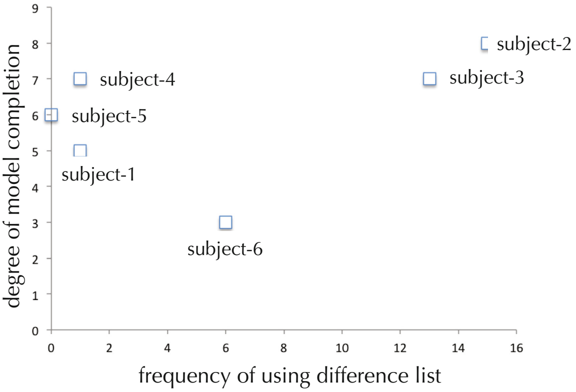


Fig. 9. Frequency of using diff list and final model completion (Session-1)

correlated (*F(1, 1) = 1875, p < .05*). As for Session-2 and Session-3, there was no significant correlation but the data tended to lie on a straight line that suggests positive correlation. As for Session-4, however, there was no significant correlation and any tendency of correlation wasn't suggested. This is because most subjects' final model completion indicated full marks (i.e., ceiling effect). The reason is supposed that task-4 was isomorphic with task-3 and that the subjects got experienced after the preceding three tasks.

In the questionnaire, three subjects who answered to question (2) *'I modified my model after considering why it was incorrect'* preferred mathematical explanation in question (3), while other three subjects who answered to question (2) *'I modified my model merely according to the message by difference list'* preferred structural explanation in question (3). As for the model completion, there was no significant difference between these two groups. That is, the learner who considered her/his model deeply preferred mathematical explanation while those whose purpose was merely to complete her/his model preferred structural explanation. This fact suggests though using difference list was effective in learners' model completion, it didn't necessarily promote their understanding (Fig. 10).

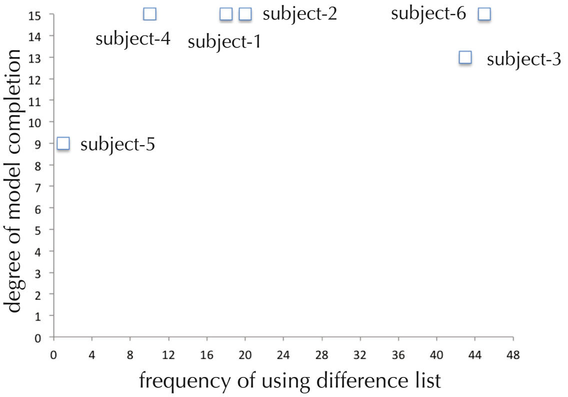


Fig. 10. Frequency of using diff list and final model completion (Session-4)

5 Conclusion

In this paper, we reported about the experiment to evaluate the effect of difference list that detects and explains the difference between learners' models and the correct models in MBE. The results revealed that using difference list was effective in learners' model completion, especially in difficult tasks. In addition, it was suggested that there was significant correlation between the frequency of using difference list and model completion, and that learners' strategy for modifying model correlates their preference of explanation by difference list. These results suggest the usefulness of difference list in MBE.

However, the following problem was also revealed. That is, in this experiment, the frequency of executing simulation of models was much less than that in the preceding preliminary experiment in which difference list wasn't implemented. This fact suggests using difference list makes learners focus on the elimination of listed differences, rather than the refinement of models by observing their behavior through simulation. Therefore, it is our important future work to find the method for appropriately combining the difference list and simulation to promote learners' activity and understanding in modeling.

References

1. Biswas, G., Schwartz, D., Bransford, J.: Technology support for complex problem solving-from SAD environment to AI. In: Forbus, K.D., Feltovich, P.J. (eds.) Smart Machines in Education, pp. 72–97. AAAI Press (2001)
2. Bredeweg, B., Linnebank, F., Bouwer, A., Liem, J.: Grap3 – workbench for qualitative modeling and simulation. Ecol. Inform. **4**(5–6), 263–281 (2009)
3. Forbus, K.D., Carney, K., Sherin, B., Ureel, L.: Qualitative modeling for middle-school students. In: Proceedings of QR 2004 (2004)
4. Bravo, C., van Joolingen, W.R., de Jong, T.: Modeling and simulation in inquiry learning: checking solutions and giving intelligent advice. Simulation **82**(11), 769–784 (2006)
5. Horiguchi, T., Hirashima, T., Forbus, K.D.: A model-building learning environment with error-based simulation. In: Proceedings of QR 2012 (2012)
6. (1985–2017). http://www.iseesystems.com
7. Jackson, S.L., Stratford, S.J., Krajcik, J., Soloway, E.: A learner-centered tool for students building models. Commun. ACM **39**(4), 48–49 (1996)
8. Costanza, R., Voinov, A.: Modeling ecological and economic systems with stella: part iii. Ecol. Model. **143**(1–2), 1–7 (2001)