Scientific Modeling Using Large Scale Knowledge

Sungeun An[1(✉)], Robert Bates[1], Jen Hammock[2], Spencer Rugaber[1], Emily Weigel[3], and Ashok Goel[1]

[1] School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA 30308, USA
sungeun.an@gatech.edu

[2] National Museum of Natural History, Smithsonian Institution, Washington, DC 20002, USA

[3] School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA

Abstract. The intelligent research assistant, VERA, supports inquiry-based modeling by supplying contextualized large-scale domain knowledge in the Encyclopedia of Life. Learners can use VERA to construct conceptual models of ecological phenomena, run them as simulations, and review their predictions. A study on the use of VERA by college-level students indicates that providing access to large scale but contextualized knowledge helped students build more complex models and generate more hypotheses in problem-solving.

Keywords: Agent-based modeling · Inquiry-based modeling · Ecology · College-level education · Science education

1 Introduction

Research on learning about scientific modeling has revealed the need for cognitive assistance of several kinds [2,4,9,10]. In particular, scientific modeling requires domain knowledge, e.g., relationships between variables describing the system being modeled, as well as mathematical skills [9]. Thus, the question our research addresses is how can we scaffold the acquisition of domain knowledge involved in scientific modeling?

Of course, large amounts of knowledge about many domains are now readily accessible on the internet. However, much of this general-purpose knowledge is not particular to any specific task and thus difficult to comprehend by many learners. Our research hypothesis is that contextualized acquisition of this knowledge may help students achieve deeper understanding about the domain and generate richer models. The Virtual Ecological Research Assistant (VERA) supports scientific modeling in the domain of ecology using large scale domain knowledge through Smithsonian's Encyclopedia of Life (EOL; [7]). Preliminary results from the experiment indicate that contextualized access to ecological knowledge from EOL helped the students build more richer models in problem-solving.

I. I. Bittencourt et al. (Eds.): AIED 2020, LNAI 12164, pp. 20–24, 2020.
https://doi.org/10.1007/978-3-030-52240-7_4

2 VERA: A Research Assistant for Ecological Modeling

VERA is a web-based system intended for large-scale use and supports scientific modeling in three ways [1]. First, it provides a visual language with a well-defined semantics to represent conceptual models clearly. Second, it automatically translates a conceptual model into an agent-based simulation suitable for the ecological domain without requiring any programming skills or mathematical expertise. Third, it provides access to large scale biological knowledge through EOL to help the students construct the conceptual models and set the simulation parameters.

While other recent modeling systems (Co-Lab and PROMETHEUS) use equation-based modeling that consists of a set of equations and executions to evaluate them [3,5], VERA uses a visual language to specify conceptual models that automatically generate agent-based simulations and leverages contextualized domain knowledge to assist the process of construction of the models of ecological phenomena.

VERA is built on our previous work [6] that integrated Component-Mechanism-Phenomenon (CMP) models and their agent-based simulations. VERA contains three types of components: *biotic abiotic, and habitat.* VERA's taxonomy of interactions among biotic components is based on the ontology of the interactions used by EOL, in particular, Global Biotic Interactions (GloBI) [8]. The specific interactions it uses are *produces, consumes, becomes on death, and affects.* It uses an off-the-shelf agent-based simulation system called NetLogo [11] because agent-based simulations are especially well suited for ecological modeling. Running the simulation enables the user to observe how system variables change over time and to refine their models through a generate-evaluate-revise loop.

3 Contextualization of Domain Knowledge

EOL is the world's largest aggregated and curated database of species data with almost two million species and eleven million attribute records in the biological domain. VERA enables a learner to access EOL to find species of interest and automatically populate simulation parameters. VERA currently uses the following parameters specific to ecology from EOL: *lifespan, body mass, carbon biomass, respiratory rate, photosynthesis rate, assimilation efficiency, reproductive maturity, reproductive interval, and offspring count.*

3.1 Illustrative Example of Inquiry-Based Modeling Using VERA

In the following scenario, a learner wants to create a model of an observed food web to explore the predator-prey relationship between sheep and wolves. The learner begins by placing a biotic component into the conceptual model canvas, naming it "sheep," and clicking on "Lookup species on EOL." The system queries EOL for all matches to the scientific or common name "sheep" and checks for the existence of attribute records for each found species. The learner selects

"domestic sheep," and VERA extracts the species attributes from EOL that are relevant to the agent-based simulation (see Fig. 1). This provides the learner with valuable data that a student would be hard-pressed to locate and make sense of, reducing the cognitive load in model creation.

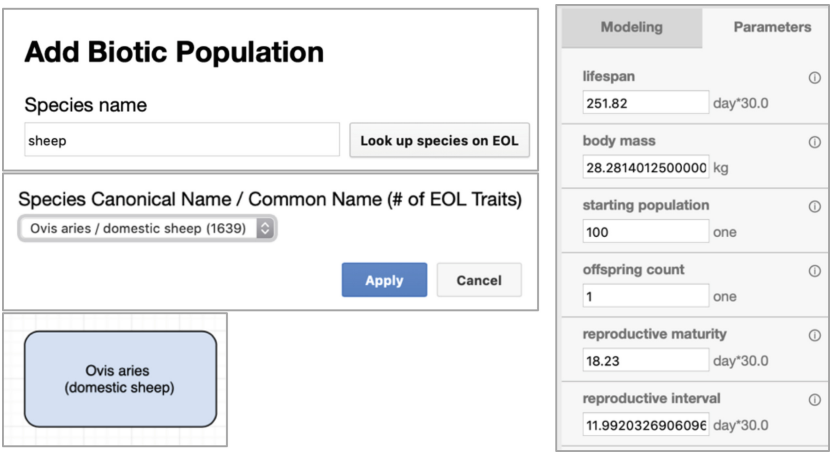

Fig. 1. Automatic filling of simulation parameters retrieved from EOL.

The learner carries on with the model construction to add a predator (wolf) and food source (grass) along with adding consumption interactions between the populations, leveraging EOL lookups with each component and interaction. In this way, our intrepid novice scientist has constructed a partial food web model revolving around the species of interest. VERA automatically spawns the simulation and displays the results as a set of graphs, for example, a graph indicating the changes in populations of various species over time (see Fig. 2).

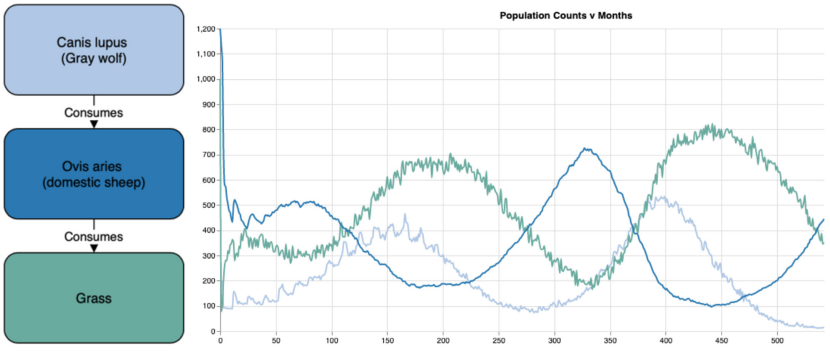

Fig. 2. A conceptual model of the relationships between species (left) and the simulation results generated from the conceptual model creating predator-prey cycle (right).

The learner may now experiment with different simulation parameters, revise the conceptual model, or generate an alternative hypothesis.

4 Lab Experiment

The goal of the study was to see if providing access to domain knowledge helps learners build richer models in problem solving. Fifteen self-selected students ($N = 15$) were recruited from a college-level general biology introductory course taught in Fall 2018. During the study, the students were asked to use VERA to explore multiple hypotheses to explain the decline in the sheep population. The students were also encouraged to actively use EOL when they needed information about a given species and later asked how often they used EOL while modeling.

4.1 Results

The students in our study developed models in VERA to evaluate their hypotheses and used EOL to get information about the species being modeled. The complexity of a model was calculated by adding the number of components in the conceptual model and the total number of relationships among the components. We found that access to ecological data from EOL helped students build more complex models and generate more hypotheses. The students who answered that they used the EOL frequently were found to come up with multiple hypotheses and build more complex models (Pearson product-moment correlation coefficients; $r = 0.38$; $r = 0.26$). We conjecture that information about the relationships between predator, prey, and competitors in the EOL knowledge-base may have led to the construction of more complex models. Interestingly, building complex models was associated with generating more hypotheses ($r = 0.66$). This means that the students who build more complex models are likely to have generated more hypotheses.

5 Conclusion

The research question in this work is how might we scaffold the acquisition of domain knowledge for students engaged in scientific modeling? The research hypothesis is that the contextualized acquisition of domain knowledge will help students build richer models. VERA contextualizes EOL's large scale domain knowledge to support modeling of ecological systems. The study with college-level students using VERA confirmed that contextualized acquisition of domain knowledge helped them construct more complex models and more explanatory hypotheses.

Acknowledgements. This research is supported by an US NSF grant #1636848 (Big Data Spokes: Collaborative: Using Big Data for Environmental Sustainability: Big Data + AI Technology = Accessible, Usable, Useful Knowledge!) and Georgia Tech seed grants through the Brooke Byers Institute for Sustainable Systems.

References

1. An, S., Bates, R., Hammock, J., Rugaber, S., Goel, A.: VERA: popularizing science through AI. In: Penstein Rosé, C., et al. (eds.) AIED 2018. LNCS (LNAI), vol. 10948, pp. 31–35. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93846-2_6
2. Bransford, J.D., Brown, A.L., Cocking, R.R., et al.: How People Learn, vol. 11. National Academy Press, Washington, DC (2000)
3. Bridewell, W., Sánchez, J.N., Langley, P., Billman, D.: An interactive environment for the modeling and discovery of scientific knowledge. Int. J. Hum Comput Stud. **64**(11), 1099–1114 (2006)
4. Hogan, K., Thomas, D.: Cognitive comparisons of students' systems modeling in ecology. J. Sci. Educ. Technol. **10**(4), 319–345 (2001)
5. van Joolingen, W.R., de Jong, T., Lazonder, A.W., Savelsbergh, E.R., Manlove, S.: Co-lab: research and development of an online learning environment for collaborative scientific discovery learning. Comput. Hum. Behav. **21**(4), 671–688 (2005)
6. Joyner, D.A., Goel, A.K., Papin, N.M.: MILA-S: generation of agent-based simulations from conceptual models of complex systems. In: Proceedings of the 19th International Conference on Intelligent User Interfaces, pp. 289–298 (2014)
7. Parr, C.S., et al.: The encyclopedia of life v2: providing global access to knowledge about life on earth. Biodiv. Data J. (2), e1079 (2014)
8. Poelen, J.H., Simons, J.D., Mungall, C.J.: Global biotic interactions: an open infrastructure to share and analyze species-interaction datasets. Ecol. Inform. **24**, 148–159 (2014)
9. Sins, P.H., Savelsbergh, E.R., van Joolingen, W.R.: The difficult process of scientific modelling: an analysis of novices' reasoning during computer-based modelling. Int. J. Sci. Educ. **27**(14), 1695–1721 (2005)
10. VanLehn, K.: Model construction as a learning activity: a design space and review. Interact. Learn. Environ. **21**(4), 371–413 (2013)
11. Wilensky, U., Resnick, M.: Thinking in levels: a dynamic systems approach to making sense of the world. J. Sci. Educ. Technol. **8**(1), 3–19 (1999)