To appear in Procs. 19th International Conference on AI in Education, London, UK, June 27-30, 2018.

VERA: Popularizing Science through AI

Sungeun An[1], Robert Bates[1], Jennifer Hammock[2], Spencer Rugaber[1], and Ashok Goel[1]

[1] School of Interactive Computing, Georgia Institute of Technology, Atlanta GA 30308, USA
[2] National Museum of Natural History, Smithsonian Institution, Washington, DC 20002, USA
sungeun.an@gatech.edu

Abstract. Citizen scientists have the potential to expand scientific research. The virtual research assistant called VERA empowers citizen scientists engaged in environmental science in two ways. First, it automatically generates simulations based on the conceptual models of ecological phenomena for repeated testing and feedback. Second, it leverages the Encyclopedia of Life biodiversity knowledgebase to support the process of model construction and revision.

Keywords: Citizen Science · Conceptual Modeling · Ecology · Encyclopedia of Life · Simulation.

1 Introduction

General public scientists, sometimes called citizen scientists, participate in scientific research in part to contribute to and expand the impacts of any study [1–3]. Yet, the role of citizen scientists is often limited to data collection [4, 5]. We seek to develop computational techniques and tools to empower citizen scientists to play a more active role in scientific research, especially in environmental science and policy [6].

The scientific process often starts with the identification of an atypical or abnormal phenomenon such as unprecedented movement of a biological species into a new geographical region [7]. The scientist may then propose, elaborate, evaluate, revise/refine, and accept/reject multiple hypotheses for explaining the phenomenon. Professional scientists widely use multiple kinds of models of the system of interest including conceptual models and simulation models [8, 9]. Conceptual models are abstract and declarative representations of a system with components, relations, and processes. Simulation models can be executed to evaluate a hypothesis by calculating the real effects and courses of action under certain conditions of the system. Thus, a conceptual model can help express hypotheses that can then be evaluated and revised through simulations. The construction of a conceptual model requires relational knowledge that involves relationships between two species (e.g., predator-prey relationships), whereas model simulation requires quantitative information about a species (e.g., population, birthrate, lifespan). These are two areas in which many citizen scientists need much support.

Thus, our research question is how to develop a virtual research assistant that can help citizen scientists construct, test and revise their hypotheses about ecological phenomena. The Virtual Ecological Research Assistant (VERA; `http://support.dilab.gatech.edu/okuwiki/mantis/virtual_ecological_research_assistant_vera/start` empowers citizen scientists in two ways. First, it automatically generates simulations based on the conceptual models of ecological phenomena for repeated testing and feedback. Second, it leverages Encyclopedia of Life (`http://eol.org/`) biodiversity knowledgebase to support the process of model construction and revision. A couple of pilot studies show promising results.

2 VERA

Conceptual models of ecological phenomena in VERA are expressed in the Component-Mechanism-Phenomenon (CMP) language [10, 11]. CMP modeling of natural systems is an adaptation of the Structure-Behavior-Function modeling of technological systems [12, 13]. A CMP model consists of components and relationships between components. A component can be one of four types: biotic, abiotic, base population, and habitat. A relationship relates one component to another in a directed manner (e.g., component X consumes component Y). The allowed relationships vary based on the source, and destination components selected, but always are a subset of the relations ontology supported by EOL. The sixteen types of relationships presently implemented in VERA including "consumes," "destroys," "infects," and "spreads." Fig 1 illustrates the VERA system that includes MILA-S [10, 11] for automatic generation of agent-based simulations and EOL for retrieving traits of species.

2.1 Automatic Generation of Agent-Based Simulations

Following our earlier work on the ACT [14] and MILA-S system [10, 11], VERA uses an artificial intelligence compiler to automatically translate the patterns in the conceptual models into the primitives of agent-based simulation of NetLogo [15]. The running of the simulation enables the user to observe the evolution of the system variables over time, and iterate through the model-simulate-refine loops [16]. In this way, like MILA-S, VERA integrates both qualitative reasoning in the conceptual model and quantitative reasoning in the simulation reasoning on one hand, and explanatory reasoning (conceptual model) and predictive reasoning (simulation).

2.2 Integration with Encyclopedia of Life

VERA integrates MILA-S with EOL [17] to provide the user with access to knowledge about biological species, for example, data about the traits of a

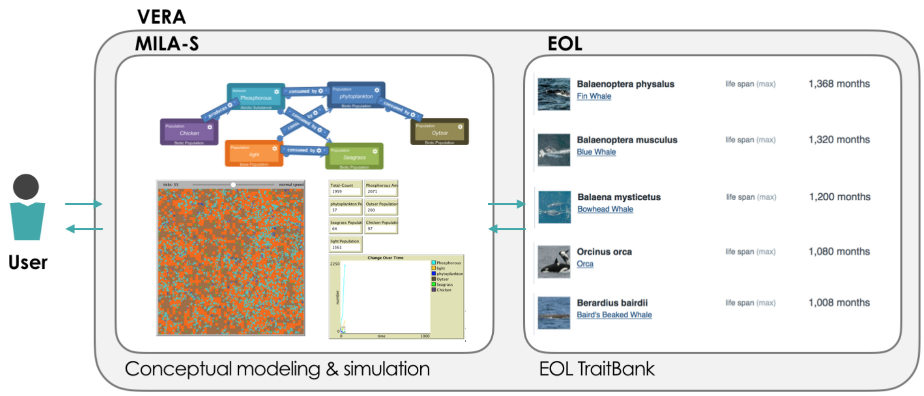

Fig. 1. The overall structure of VERA. VERA includes MILA-S to let users create conceptual models about the problems in ecological systems and execute simulations. In the meantime, VERA uses EOL TraitBank to scaffold the process of model construction.

species. In particular, once having built a conceptual model of an ecological phenomenon, the user may need quantitative data about the biological species in the model to set up the agent-based model simulation. VERA enables the user to look up this kind of data in EOLs TraitBank [18]. For example, when creating a model that explains why a specific kind of starfish is dying off the west coast of USA, a user can search EOLs TraitBank for the birthrate and lifespan of the starfish and set the simulation parameters accordingly.

We are presently constructing additional tools for accessing knowledge from EOL to support citizen scientists. For example, when adding predator-prey relationships of starfish, a user may not know what starfish eat or what might eat starfish. To facilitate looking up this kind of information, we seek to use IBMs Bluemix services to search the EOL for an answer by entering a question such as "What do starfish eat?"

3 Conclusion

The VERA system helps citizen scientists in constructing and testing models of ecological systems in two ways. First, it automatically generates simulations based on the conceptual models of ecological phenomena for repeated testing and feedback. Second, it leverages the Encyclopedia of Life biodiversity knowledgebase to support the process of model construction and revision. Initial pilot studies indicate promising results. To contribute to environmental sustainability, citizen scientists can use the VERA system to model, analyze, explain, and predict problems in ecological systems.

Acknowledgement

At Georgia Tech, we thank David Joyner and Taylor Hartman for their contributions to MILA-S, and Abbinayaa Subrahmanian, Christopher Cassion and Pramodith Ballapuram for their contributions to VERA. At Smithsonian Institution, we thank Katja Schulz for her contributions to EOL. 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 the NSF South BigData Hub.

References

1. Bonney, Rick, et al. "Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. A CAISE Inquiry Group Report." Online Submission (2009)
2. Irwin, Alan. Citizen science: A study of people, expertise and sustainable development. Routledge, (2002)
3. Hand, E. People power. Nature, 466(7307), p.685-687 (2010)
4. Brossard, Dominique, Bruce Lewenstein, and Rick Bonney. "Scientific knowledge and attitude change: The impact of a citizen science project." International Journal of Science Education 27.9 (2005): 1099-1121
5. Cohn, Jeffrey P. "Citizen science: Can volunteers do real research?." AIBS Bulletin 58.3 (2008): 192-197
6. Couvet, Denis, et al. "Enhancing citizen contributions to biodiversity science and public policy." Interdisciplinary science reviews 33.1 (2008): 95-103
7. Nersessian, Nancy J. "Model-based reasoning in conceptual change." Model-based reasoning in scientific discovery. Springer, Boston, MA, 1999. 5-22
8. Clement, J. Creative Model Construction in Scientists and Students: The Role of Imagery, Analogy, and Mental Simulation. Dordrecht: Springer (2008)
9. Nersessian, Nancy J. Creating scientific concepts. MIT press, (2010)
10. Joyner, D, Goel, A., & Papin, N. 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). ACM. (2014)
11. Goel, A. & Joyner, D. Impact of a Creativity Support Tool on Student Learning about Scientific Discovery Processes. In Procs. Sixth International Conference on Computational Creativity, Park City, Utah, June (2015)
12. Goel, A., Gomez, A., Grue, N., Murdock, J., Recker, M., & Govindaraj, T. Towards Design Learning Environments - Exploring How Devices Work. In Proc. International Conference on Intelligent Tutoring Systems, Montreal, Canada, June (1996)
13. Goel, A., Rugaber, S., & Vattam, S. Structure, Behavior and Function Models of Complex Systems: The Structure-Behavior-Function Modeling Language. AIEDAM, 23: 23-35, April (2009)
14. Vattam, S., Goel, A., Rugaber, S., Hmelo-Silver, C., Jordan, R., Gray, S, & Sinha, S. Understanding Complex Natural Systems by Articulating Structure-Behavior-Function Models. Journal of Educational Technology & Society, 14(1): 66-81. (2011)

15. Wilensky, U. NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. (1999)
16. White, B., & Frederiksen, J. Causal model progressions as a foundation for intelligent learning environments. Artificial intelligence, 42(1), 99-157. (1990)
17. Parr, C., Wilson, N., Leary, P., Schulz, K. Lans, K., Walley, L., Hammock, J., Goddard, A., Rice, J., Studer, M., Holmes, J, & Corigan, R. The encyclopedia of life v2: providing global access to knowledge about life on earth. Biodiversity Data Journal 2 (2014), p. e1079.
18. Parr, C., Wilson, N, Schulz, K, Leary, P, Hammock, J, Rice, J, Corrigan, R. Trait-Bank: Practical semantics for organism attribute data. Semantic Web Interoperability, Usability, Applications 650-1860. (2014)