Title
Functional Pursuit: A Model of Successful Induction in Mathematics

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 19(0)

Authors
Haverty, Lisa A.
Koedinger, Kenneth R.
Klahr, David

Publication Date
1997

Peer reviewed

Functional Pursuit: A Model of Successful Induction in Mathematics

Lisa A. Haverty (LH3C+@ANDREW.CMU.EDU)
Kenneth R. Koedinger (KOEDINGER@ANDREW.CMU.EDU)
David Klahr (KLAHR@ANDREW.CMU.EDU)
Carnegie Mellon University
Pittsburgh, PA 15213 USA

The induction of mathematical functions from data plays an important role in mathematics and in scientific discovery (Langley, et al., 1987). In this paper we present a computational model of the cognitive processes used by undergraduates in inducing mathematical functions from a series of (x,y) data pairs. Sixteen undergraduates were given two tasks in which they used a computer interface to collect data pairs in order to find a mathematical function consistent with the data. Participants could freely choose an X value and the interface responded with a Y value. For example, on Task A, a student might have collected the following data pairs {(10, 35) (3, 0) (7, 14) (4, 2) (5, 5) (6, 9)}. The reader is invited to try to discover the function behind Task A. The function behind Task B was x*(2x+1) and all but one student successfully discovered this function. In contrast, on Task A, only nine of the 16 students succeeded within the 25-minute time limit.

Such function finding tasks provide a simple experimental domain for investigating core skills of data collection, pattern finding, and hypothesis formation that are fundamental to reasoning and learning in math and science. Related prior work on the BACON model (Qin & Simon, 1990) identified five heuristic rules to characterize the discovery of Kepler's Third Law ($D^3 = cP^2$) from a table of data. The BACON project demonstrated that discoveries made by scientists can be captured in a computer model. BACON's five heuristics are computationally sufficient to succeed on Tasks A and B, yet the full variety of pattern finding and hypothesis generation activities observed in students' verbal reports are not captured in those five heuristics. Our goal is to extend this line of inquiry and to better understand the discovery skills used and learned by non-scientists.

Using verbal protocols, we contrasted the performance of successful and unsuccessful students by looking at how they differ on factors such as data collection strategies, the examination of intermediate quantities, the relative use of pattern finding versus hypothesis generating techniques, and the ability to symbolically describe patterns. We have created an ACT (Anderson, 1993) model based on the performance of successful participants. The model has several pattern detection capabilities as well as minimal algebraic skills, and these capabilities are based on processes observed in students.

Our model incorporates several productions which produce patterns independent of any hypothesis, such as the observed process of examining the differences between successive values of y in a table of (x,y) pairs. For example, using the data pairs for Task A {(3, 0) (4, 2) (5, 5) (6, 9) (7, 14) (10, 35)}, and taking the differences between the successive y-values of 0, 2, 5, 9, and 14, the model will find the pattern: {2, 3, 4, 5}. Our model uses the products of its pattern finding activity to generate potential components for hypotheses. Thus, the pattern {2, 3, 4, 5} can be expressed in terms of its corresponding x-values: {4, 5, 6, 7} with the expression "x-2". This skill of expressing a pattern in terms of x is crucial to the process of function induction.

We offer an example to demonstrate how our model captures the tendency of students to discover each of the components of a function independently (e.g., on Task B: "*2", "+1", "*x") and to then combine these components to produce the final function. In discovering the function y = x(2x+1), with data instances {(1, 3) (2, 10) (3, 21) (4, 36)}, the model first divides y by x, which produces the sequence {3, 5, 7, 9}=C. The model subsequently focuses on expressing C in terms of x. Here we describe the model's path for students who do not retrieve the fact that these odd numbers can be expressed as 2x+1. The model calculates the discrepancy between C and x, (C - x) which yields {2, 3, 4, 5}=D. Since D is not constant, the model again computes a discrepancy, this time between$D$ and x, which yields {1, 1, 1, 1}. Checking that this new quantity is indeed a constant, the model uses this constant = 1 to begin building its formal hypothesis. First, it proposes that x + 1 = D. Next, it uses the fact that $C = D$ + x, and produces C = 2x+1. Finally, it notes that y = C * x = (2x+1) * x.

We conclude that, as an accurate and detailed embodiment of the processes used by successful participants, our model of student performance is a first step toward identifying plausible instructional objectives for the teaching of inductive reasoning in mathematics.

References

Anderson, J.R. (1993). *Rules of the Mind.* Hillsdale, NJ: Lawrence Erlbaum Assoc.

Langley, P., Simon, H., Bradshaw, G., & Zytkow, J. (1987). *Scientific Discovery.* Cambridge, MA: MIT Press.

Qin, Y. & Simon, H.A. (1990). Laboratory Replication of Scientific Discovery Processes. *Cognitive Science, 14*, 281-312.