Understanding Student Answers In Introductory Physics Courses

C.W. LIEW [a,1] JOEL A. SHAPIRO [b] D.E. SMITH [c] and RYAN MCCALL [d]
[a] *Department of Computer Science, Lafayette College*
[b] *Department of Physics and Astronomy, Rutgers University*
[c] *Department of Computer Science, Rutgers University*
[d] *Department of Mathematics, Lafayette College*

Keywords. Mathematics and science education, Physics Algebraic Equations

1. Introduction

Many assigned problems in introductory physics courses require the student to formulate a system of algebraic equations that describe a physical situation. An Intelligent Tutoring System (ITS) designed to help must be able to identify which physical quantity each student variable represents, and which fundamental laws each of her equations incorporates. Many ITS's predefine the variables ([4,2]) or require the student to explicitly and precisely specify what each one represents (*e. g.* ANDES[3]). Our earlier work [1] relaxed this constraint for problems without explicit numerical constants. This paper describes how the technique has been extended to cover problems with numerical constants.

2. Problems and Issues

Determining the meaning of student variables and equations is difficult. To understand why, consider the following problem. A motorcycle cop traveling at 30 m/s is 400 m behind a car traveling at 25 m/s. How long will it take to catch the car? A student might use the relative velocity v_r and give $t = 400\ \mathrm{m}/v_r$, or she might match the positions in time, $d_m = v_m t$, $d_c = 400\ \mathrm{m} + v_c t$, $d_m = d_c$ and solve for t. These inputs need to be matched against a canonical set of variables and equations, which might well not include the relative velocity, and will not include numerical (given) quantities. Our approach to recognizing these variables begins with identification heuristics that list the possible physical concepts normally associated with the letter. Here v_m is likely to be a velocity, and t a time, though in other problems it could be a thickness. These possibilities are constrained by requiring dimensional consistency of all the student equations.

[1]Correspondence to: C.W. Liew,
Department of Computer Science, Lafayette College, Easton PA 18042
E-mail: *liew@cs.lafayette.edu*

Our algorithm attempts to assign numerical quantities used by the student to the canonical variable, so in the examples we would recognize 400 m as the given value of d_0. Numerical values not so identified are then compared to inverses of given values, combinations by addition and subtraction of given values of the same dimension, or the product or quotient of two given values. Had the first example used a numerical value for v_r, by writing $(5\ \mathrm{m/s}) * t = 400\ \mathrm{m}$, the 5 m/s would be found as the difference between v_m and v_c. Having replaced all the constants with the variables they represent, we are left with a purely algebraic system of equations with only dimensionless constants. This is the situation we addressed in our previous paper [1].

3. Evaluation

We evaluated our technique on data from three problems from the ANDES system used at the U. S. Naval Academy in 2001. All involve the use of physics principles from classical mechanics that do not employ trigonometric functions. Exkt4a is essentially the motorcycle cop problem described earlier, and Exkt3a gives the time taken for the shooter to hear a gong he hits with a bullet, given the speed of bullet and sound, and asks how far away the gong was. The last problem is

> *Exe5a*: A frictionless roller-coaster car tops the first hill whose height is 30.0 m above the base of the roller coaster with a speed of 3.00 m/s.
>
> What is the magnitude of the instantaneous velocity at the apex of the next hill which has a height of 10.0m?
>
> Choose the zero of gravitational potential energy at the base of the roller coaster.

The data we extracted from the ANDES logs on these three problems contained only correct equations. While the individual equations are correct, the student submission may be incomplete or contain redundant equations. Because ANDES immediately informs students if their equation is incorrect, and encourages the student to immediately fix it, the available logs did not provide full submissions including wrong equations which were appropriate to include in our test. Our earlier experiments [1] evaluated the ability of the algorithm to detect and analyze incorrect equations to generate useful feedback, but these evaluations were restricted to pure algebraic equations, *i. e.*, they did not contain dimensioned numerical values.

	Exkt3a	%	Exkt4a	%	Exe5a	%
Successful analysis	210	**89.8%**	250	**96.2%**	135	**93.2%**
a) Correct Solution	83	35%	118	45.4%	118	81.4%
b) Partial Solution	110	47%	99	38.1%	11	7.6%
c) No nontrivial equations	17	7.3%	33	12.7%	6	4.1%
Unsuccessful analysis	24	**10.2%**	10	**3.8%**	10	**6.8%**
d) Non-derived equations	6	2.5%	0	0%	4	2.7%
e) Unknown constants	18	7.7%	10	3.8%	6	4.1%
Total	234	100%	260	100%	145	100%

Table 1. Results from ANDES Fall 2001 data

The results of the evaluation are shown in Table 1. Each column shows the number of submissions for each problem in each category and next to it is the percentage. Category

(a) are the submissions that were determined to be correct and complete solutions to the problem. Category (b) are the submissions that contained some number of correct equations but did not contain sufficient equations to solve the problem. Category (c) are the submissions where the student did not enter any equations beyond the the assignment of explicitly given values to variables. Our system successfully handled the submissions in these three categories, which represented 93% of all submissions.

The submissions that could not be fully analyzed fall into two categories. Category (d) are submissions that contain equations that could not be resolved because of limitations with the implemented technique, *e. g.*, common factors have been omitted. Category (e) are submissions that used numerical values that were **not** recognized by the algorithm for a variety of reasons. These reasons are discussed in the next paragraph.

The failure to recognize some numerical constants and certain student equations was due to three main causes: trigonometric functions, detection of implicit zeros, and the presence of common factors. Our algorithm does not handle trigonometric functions due to their numerous equivalent representations (*e. g.*, $\cos(x) = \sin(90 - x)$) and ambiguities in numerical evaluation. We do not handle implicit values of zero that manifest themselves as terms missing from equations (*e. g.*, in a momentum conservation problem with an object initially at rest). Because our method of reducing canonical equations [1] does not include inserting numerical givens, the momentum conservation equation without the zero term is not recognized. Enhancements to handle implicit zeros would not be difficult. Our implementation also fails to recognize equations in which a common factor has been removed, so the equation $g * h_1 + 0.5 * v_1^2 = g * h_2 + 0.5 * v_2^2$ is not recognized as the energy conservation equation $m * g * h_1 + 0.5 * m * v_1^2 = m * g * h_2 + 0.5 * m * v_2^2$.

4. Conclusion and Acknowledgments

We have described how a student's algebraic equations in physics can be analyzed to detect errors and provide useful feedback. Unlike previous approaches, our approach does not require the student to define the variables explicitly (a time consuming, tedious step) and can analyze problems with embedded dimensional numeric constants commonly used in introductory physics courses. The technique has been evaluated on three problems from the ANDES system and is successful over 90% of the time in determining (1) the mapping of the student's variables to specific physical quantities, (2) the correctness of each student equation, and (3) the completeness of the student's submission. We are grateful to Kurt VanLehn, Anders Weinstein and the Andes group for their help

References

[1] LIEW, C., SHAPIRO, J. A., AND SMITH, D. What is the student referrring to? mapping properties and concepts in students' systems of physics equations. In *Proceedings of the 12th International Conference on Artificial Intelligence in Education* (2005).

[2] `http://www.masteringphysics.com/`.

[3] SHELBY, R. N., SCHULZE, K. G., TREACY, D. J., WINTERSGILL, M. C., VANLEHN, K., AND WEINSTEIN, A. An assessment of the Andes tutor. In *Proceeding of the 5th National Physics Education Research Conference* (Rochester, NY, July 2001).

[4] TITUS, A. P., MARTIN, L. W., AND BEICHNER, R. J. Web-based testing in physics education: Methods and opportunities. *Computers in Physics 12, #2* (1998).