A Cognitive Model for Automatic Student Assessment: Classification of Errors in Engineering Dynamics

Dr. Jeffrey A Davis P.Eng., Grant MacEwan University

With degrees from both Civil and Mechanical Engineering, Jeff went on to obtain a PhD from the Institute of Energy Technology at ETH Zurich in 2004. His past research includes dispersion of pollutants in rivers, turbulent and multiphase flow modeling from a numerical perspective. Currently, Jeff is a first year engineering instructor at MacEwan University. With a passion for teaching, his focus on research has turned to understanding and automating student assessment techniques as well as looking at the socio-economic sustainability of educational institutions.

Dr. Shelley Lorimer P.Eng., Grant MacEwan University

Dr. Shelley Lorimer, P.Eng. is Chair of the Bachelor of Science in Engineering Transfer Program (BSEN) at Grant MacEwan University in Edmonton, Alberta. She teaches undergraduate courses in statics and dynamics, as well as courses in engineering professionalism. She is currently participating in a research project with Alberta Innovates – Technology Futures in the oil sands and hydrocarbon recovery group doing reservoir simulation of enhanced oil recovery processes. She has a Ph.D. in numerical modeling from the University of Alberta, also in Edmonton.

A Cognitive Model for Automatic Student Assessment: Classification of Errors in Engineering Dynamics

Abstract

The present paper focuses on the errors that students made on a first year engineering dynamics final exam. An error classification scheme, based on Action Theory, is used to classify errors as either mistakes or slips in logic. Simple rules are then developed for a computer to be capable of categorizing errors based on a priori indicators such as: the student's experience level and the topic's priority on the courses concept inventory as well as a posteriori indicators such as the frequency of the error within the exam. The assessment algorithm is then trained and optimized. The resulting algorithm is tested by comparing the results with a second, independent, expert. Results of the study are then discussed.

Introduction

The taxonomy of human errors is critical for professions that deal with public safety such as health care and engineering. Once errors are identified then preventive measures can be sought. Detection of errors is also important from both the teaching and learning perspective. When students make errors it is important to "pin down" why a specific error was made and then begin the process of correcting it by providing the student with individualized feedback and assignments. Although there have been many attempts at the classification of human errors, the results tend to be context specific[1],[2],[3]. One possible cause of this is the exclusion of any link of human errors to cognitive processes.

The starting point for the present research is based on Action theory. This is a goal-directed theory that assumes the existence of a conscious choice that guides a person's behaviour to some outcome[4],[5]. In this theoretical context, an error implies that through some intended action, the goal was not attained[1]. Rooted in this theory, two models of human error taxonomy commonly cited in the literature are: the Generic Error Modeling System model[1] and the Skill-Rule-Knowledge (SRK) model[6]. These models further classify errors as being either the failure of actions to go as intended (slips, lapse) or as being a failure of actions to achieve the desired consequences (mistakes)[6]. Based on cognitive levels, Rasmussen[6] suggested that mistakes be further categorized as being either knowledge, rule, or skill based. Reason rationalized that knowledge based errors occur because of a lack of expertise while rule based errors are a failure of expertise[1]. Reason noted that several further groupings could be: inattention and over attention (skill based), mis-application of good rules and application of bad rules (rule based) as well as others such as overconfidence, bias, halo effect, complexity, causality (knowledge based)[1].

Traditional assessment methods

Typically, in STEM courses, questions are developed such that the answer is of the form of a numeric response. A comparison of that answer with a known solution will tell whether the student obtained the correct answer or not. If the assessment is in the form of a multiple choice exam then the task of assessing the student is complete. Feedback can be provided to the student in the form of "Why the right answer was correct" but this offers no way of finding the root cause as to why the student chose that answer instead of the correct one. A better assessment method is to offer questions where not only the final answer is checked for correctness but also, the method the student used to get to the final answer is checked. For an experienced instructor, this task is fairly simple but it adds some degree of ambiguity on how to assign marks. Typically a rubric is introduced to assure consistency in marking, but it is not straightforward in the method to create one. In addition, less experienced markers are hired to grade assignments at the time the student is learning the material which is also the time feedback is most important. There exists a need to create algorithms which can aid in the automatic assessment of students. The long term goal of this research deals with the creation of software which will automatically assess student's work in STEM subjects. The vision is that students will eventually solve assignments on tablets with a stylus pen similar to the way that they solve it with paper and pencil. As the students work through the assignment, their handwriting goes through a recognition stage and is stored digitally. Their work is then submitted to a secure server where it goes through the assessment stage, marking the student's work and providing valuable feedback.

Overview of paper

The goal of this paper is to create a unifying scheme for classifying student errors rooted in cognitive theory. The scheme should have simple rules for differentiating each error as well as providing some indication of the severity of the error which can then be used for assessment.

Error classification

Beginning with the SRK model, errors are classified as being either slips or mistakes according to the schematic shown in Figure 1. To determine whether an error made on an assignment or exam was a mistake or slip in logic, the marker must make an inference to the student's state or intent. Some possible factors which weigh on this decision are the student's: health (mental, physical), personality (i.e. confidence, independence), current misconceptions, competency in different subjects (physics, algebra), competency in solving the specific type of problem, and frequency of error on the specific type of problem. In addition to individual factors, some group factors which can weigh in are: how much previous experience does the group have with the topic and the frequency of error with respect to the group of students. Finally, the instructor also affects the student through stressing the topic's importance in the classroom.
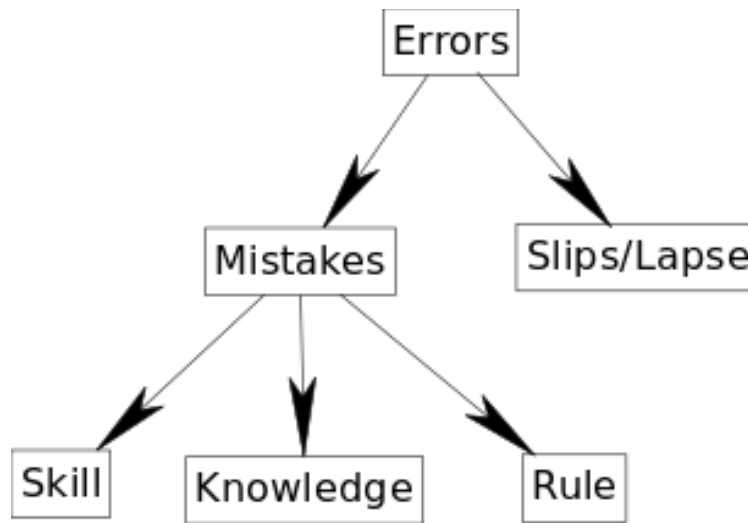
Figure 1. Flow Diagram of the classification scheme of errors.

The student's health and personality may play an important cognitive role in decision making. In addition, a student's ability and misconceptions surrounding a certain subject will play an important role. To use these factors as indicators, however, assessment exams such as a Math Advisory, Force Concept Inventory[7], or StrengthsQuest must be administered to the students prior to the course. One indicator which is more readily measurable is the current experience level of the class with the specific topic. Here, this indicator can easily be determined by looking at the course concept inventory, specific examples in class, and number of assignment questions given to the students throughout the semester dealing with that topic. Similarly, a readily measurable indicator is to determine whether the student made previous mistakes before in the class. Repeated errors generally suggest a mistake in logic whereas a single error (with a sufficient amount of practice) suggests a slip in logic. One can gauge the type of error by looking at the group as a whole. If a many students are making the same error then perhaps it is a mistake in the group's logic as opposed to a slip in logic. This is based on the assumption that slips are infrequent in nature which results in the probability that a group makes the same slip at the same time low. Finally, by stressing the relative importance of a topic in class, an instructor forces a student to pay closer attention to that topic and thus is more conscientious of their actions.

In summary, indicators determining the classification of an error into either a slip or mistake require that an instructor use criteria based on the student's previous learning experience. To accomplish this requires a better tracking of the student's achievements and progress during the semester as well as tracking of concept inventories in the classroom and assignments. Due to a lack of prior knowledge of the student's progress, the current study will only deal with three factors dealing with concepts closely related to the course: the student's previous experience level with the specific type of problem, the group's frequency of error on the problem, and the relative importance of the topic as stressed by the instructor.

Classification algorithm

Given some error, ϵ, the probability that ϵ is a mistake, $P(M)$, is taken to be dependent on the variables x_i and is assumed to be a weighted sum of all of the independent contributions, i.e.,

$$P(M) = P(x_1, x_2, x_3, \ldots, x_n) = \frac{\sum_i \omega_i P(x_i)}{\sum_i \omega_i} \tag{1}$$

where ω_i are the weights of each of the independent variables x_i. An error is then classified as being either a mistake or a slip using the step function

$$\epsilon = \begin{cases} mistake; P(M) \geq \phi \\ slip; \ P(M) < \phi \end{cases} \tag{2}$$

where ϕ is a predefined variable describing the cutoff point between slips and mistakes. The separate functions, $P(x_i)$, are taken to be of the form of modified cumulative normal distributions, i.e.,

$$P(x_i) = a_i + b_i erf\left(\frac{\left(\frac{x_i}{e_i}\right) + c_i}{d_i}\right) \tag{3}$$

where a_i, b_i, c_i, d_i, e_i are coefficients of the model. As an example use of Equation 3, the probability that the error is a mistake based on the student's prior experience, the topic's priority as stressed by the instructor, and the group's repeatability of the error (on the specific assignment) is plotted in Figure 2.
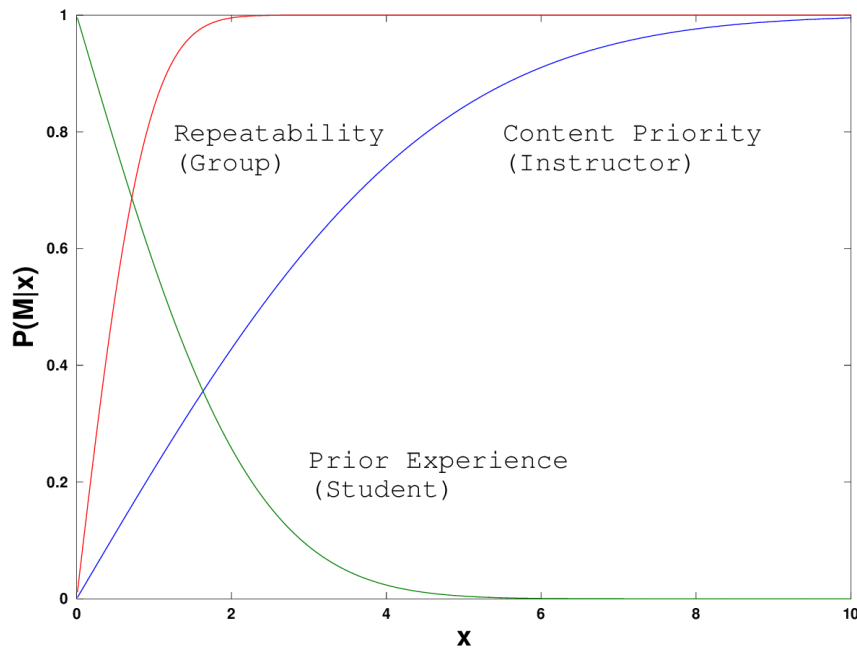

Figure 2. Plot of the probability distribution functions.

For the study, the values of the coefficients from Equation 3 and the weights, ω, from Equation 1 are shown in Table 1. These values were optimized by using the data discussed in the following sections.

Coefficients	Content Priority i=1	Prior Experience i=2	Repeatability i=3
a	0	1	0
b	1	−1	1
c	0	0	0
d	0.5	0.25	0.1
e	10	10	10
ω	2	1.5	1

Table 1. Table showing the coefficients used in the study.

Application

To test the application of the assessment scheme previously described, a total of 3 questions from the final exam of Engineering Dynamics were used. The questions were chosen based on the class average which served as an indication of the problem's difficulty. Sample statistics of the exams are provided in Table 2.

# of students	128		
Total # of questions	9		
Total time allotted	3h		
Exam average	61.7%		
Question	1	2	3
Average	76%	72%	60%
# of steps	11	7	6
Question worth	11%	11%	11%

Table 2. Statistics on the exam and individual questions.

Next, given each question, a list of the student's errors was compiled. Each error was then calculated and categorized as either a slip or a mistake using Equations 1-3 discussed previously.

Results

As an example, question 1 (see Table 2) was chosen and is shown in Figure 3. Here, the question is broken into separate parts which help to guide the students towards the solution. A diagram is also provided to help the students see the problem visually.

Two blocks at rest are shown in the Figure. A force $\vec{P}$ of magnitude 75 lb is applied horizontally to block A as shown. The weights of block A and B are indicated in the Figure. The coefficients of friction between block A and the plane are $\mu_s = 0.4$ and $\mu_k = 0.3$, respectively. Neglecting the mass of the pulleys and the effect of friction in the pulleys, use **Newton's equations of motion** to determine:

a) the tension in the cable,
b) the acceleration of block A,
c) the acceleration of block B,
d) the velocity of block A after 3 seconds,
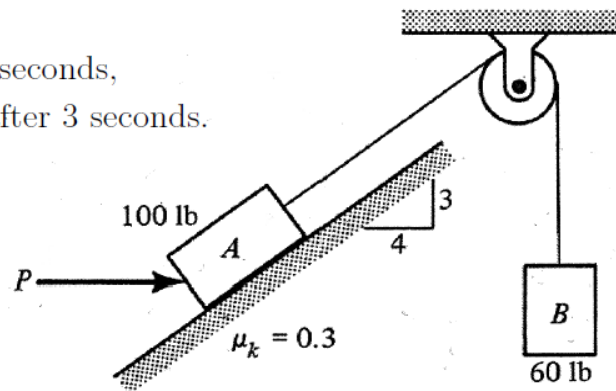e) the distance the block moved after 3 seconds.

Figure 3. Sample question for the study.

The solution of the problem can be broken into 11 steps:

1. draw FBD and MAD for block A
2. draw FBD and MAD for block B
3. determine angle
4. determine relationship between a_B and a_A
5. sum normal forces in block B to determine equation $a_B(T)$
6. sum normal forces in block A to determine N
7. sum tangential forces in block A to determine equation $a_A(T)$
8. solve for the rope's tension, T
9. solve for each block's acceleration: a_B, a_A
10. use kinematics equation to determine v_A
11. use kinematics equation to determine y_B

Breaking up the problem is useful for determining at which step the student made the error. The results of the errors made on this question are provided in Table 3.

Error	Count	Description
Sign error	33	a_T wrong direction
Incorrect balance of forces	27	FBD correct, equation incorrect
Arithmetic error in obtaining angle	21	Using $\cos(\theta)$ instead of $\sin(\theta)$
Incorrect solution of two equations	13	Correct eqns. Solved incorrectly
Missing acceleration of MAD	7	Setting $a_T = 0$
Missing direction/units	6	Final answer missing units
Mixing up units, weights	6	Using N instead of lb Using m instead of W
Transcription error	3	Re-writing equation introduces error
Missing forces on FBD	2	Missing Weight Term
Using wrong coefficients	2	Using μ_s instead of μ_k

Table 3. Errors made in question 1 from the ENPH 131 final exam.

Here, a clean separation can be seen between errors which were commonly made on the exam and those which were atypical. The error most commonly made was related to the student giving the acceleration the wrong direction on the acceleration term when balancing the equation $\sum F_T = ma_T$. Others errors commonly made include incorrectly writing the balanced equations, incorrectly solving the balanced equations, and incorrectly determining the angle or slope on which block A lies. Other less common errors include mixing up units (SI vs. FPS), using the wrong friction coefficient, and transcription errors where the student began with the correct equation but re-writing it introduced errors (a "+" became a "−" or a "4" became a "9" for example).

The results of the error classification algorithm are shown in Table 4.

Error Description	x_1	x_2	x_3	$P(x_1)$	$P(x_2)$	$P(x_3)$	$P(M)$	ϵ	ϵ_{exp1}	ϵ_{exp2}
Acceleration Sign	3	0	33	0.60	1.00	1.00	0.82	m	m	s
Balance of Forces	10	20	27	1.00	0.14	1.00	0.71	m	m	m
Angle	10	20	21	1.00	0.14	0.98	0.71	m	m	m
Substitution	10	10	13	1.00	0.46	0.85	0.78	m	m	m
Missing "a" in MAD	9	20	7	0.99	0.14	0.56	0.61	m	m	m
Missing dir./units	1	20	6	0.22	0.14	0.49	0.26	s	s	s
Mixing up units/weights	3	50	6	0.60	0.00	0.49	0.38	s	s	m
Transcription	1	50	3	0.22	0.00	0.26	0.16	s	s	s
Missing force on FBD	10	20	2	1.00	0.14	0.17	0.53	m	m	m
Wrong Coefficients	4	10	2	0.74	0.46	0.17	0.52	m	m	m

Table 4. Results of the classification of the errors from question 1.

Here, the probability of the error being a mistake, $P(M)$, is estimated using Equations 1-3. The values for x_1 are determined through the relative importance given to the topic by the instructor. Similarly, values for x_2 are determined by estimating the experience of the student with the topic from the number of times the student has attempted the problem from assignments and seminars. Next, before the results of the algorithm are known, one of the authors classified the error, ϵ_{exp1}, based on their experience. This allowed the calculation of the cut-off ϕ, from Equation 2, which was found to be $\phi = 0.5$. Thus if $P(M) \geq 0.5$ then $\epsilon = $ mistake (m) otherwise $\epsilon = $ slip (s). Then independently, without seeing the results of the algorithm, the second author categorized the errors as well. Results show that 10 out of 10 times the algorithm reproduced the expected results from training. In addition, 8 out of 10 times the algorithm and the two authors are in complete agreement. Differences in classification from ϵ_{exp2} are attributed to the relative importance and priority given to the topic by the instructor.

Similarly, results of the errors found in questions 2 and 3 of the exams are shown in Table 5. Here, for both questions, the classification of the error using the training from question 1 is found to match up for all results from ϵ_{exp1}.

Question 2				Question 3			
Error Description	ϵ	ϵ_{exp1}	ϵ_{exp2}	Error Description	ϵ	ϵ_{exp1}	ϵ_{exp2}
Incorrect Angle	m	m	m	Missing Angle	m	m	m
Setup of energy eqn.	m	m	m	Missing MAD	m	m	m
Setup of projectile eqn.	m	m	m	Sln. of energy eqn.	m	m	s
Setup/sln. of momentum eqn.	m	m	m	Spring Length	m	m	m
Sln. of quadratic eqn.	s	s	s	FBD missing force	m	m	m
Transcription error	s	s	s	Spring potential	m	m	m
Correct eqn. / wrong sln.	s	s	s	Radius of Curvature	m	m	s
Sign error	s	s	s	Angle	s	s	s
Angle error	s	s	m	Sign	s	s	s
				Transcription	s	s	s
				Normal Force	m	m	s

Table 5. Results of the classification of errors for question 2 and 3.

Comparison of the results of both questions show that results are in agreement for 8 out of 9 and 8 out of 11 errors for questions 2 and 3 respectively. The difference is attributed to the difference in priority of the concept given by different instructors. In addition, the differences from ϵ and ϵ_{exp2} correspond to errors which had $P(M)$ near the cutoff of 0.5. The results show that the algorithm is trainable to a specific instructor's concept of slip or mistake but the variation between instructors suggest that the classification of some types of errors requires more information such as a priori knowledge of the student's state. The inclusion of more factors into the model is straightforward but requires additional training of new coefficients and weights.

Conclusions and final thoughts

A simple algorithm was created which classifies errors into either slips or mistakes in logic based on readily available information: the student's experience level, the group's repeatability of the error, and the instructor's priority of the concept. This algorithm is a useful step towards the automatic assessment of students. The algorithm is easily extendible to account for more factors, however, requires additional training to determine model coefficients. Results of the algorithm were found to be promising as trained coefficients were found to be robust for errors on the same subject. Results show some variation between the classification of errors into slips or mistakes by various experts suggesting that more factors should be taken into account.

Future work on the algorithm will involve testing the algorithm in other subject areas, adding more factors into the model and extending the classification to include different types of mistakes.

Acknowledgements

Bibliography

[1] Reason, J., Human Error, Cambridge University Press, Cambridge, UK, 1990.

[2] Steif, P. and Dantzler, J. A Statics Concept Inventory: Development and Psychometric Analysis, *J. Engg. Ed.*,vol. 33, pp. 363-371, 2005.

[3] Sutcliffe, A. and Rugg, G. A taxonomy of Error Types for Failure Analysis and Risk Assessment, *Int. J. of Human Computer Interaction*, vol. 10, pp. 381-405, 1998.

[4] Hacker, W., Activity: A fruitful concept in industrial psychology, In: Frese, M. and Sabini (eds), Goal Directed Behavior: The Concept of Action in Psychology, Hillsdale, N.J., 1985.

[5] Zapf, D., Brodbeck, F., Frese, F., Peters, H, and Pruemper, J. Errors in Working with Office Computers: A First Validation of a Taxonomy for Observed Errors in a Field Setting, *Int. J. of Human-Computer Interaction*, vol. 4, pp. 311-339, 1992.

[6] Rasmussen, J., Information processing and human-machine interaction: an approach to cognitive engineering, North-Holland, New York, 1986.

[7] Henstenes, D., Wells, M. and Swakhamer, Force Concept Inventory, *The Physics Teacher*, vol. 30, pp. 141-158, 1992.